# A MULTIMEDIA SYSTEM FOR ROUTE SHARING AND VIDEO-BASED NAVIGATION

*Wen Wu, Jie Yang*

School of Computer Science, Carnegie Mellon University
Pittsburgh, PA 15213 USA

*Jing Zhang*

General Motors R&D
Warren, MI 48090 USA

## ABSTRACT

Trip planning and in-vehicle navigation are crucial tasks for easier and safer driving. The existing navigation systems are based on machine intelligence without allowing human knowledge incorporation. These systems give turn guidance with abstract visual instruction and have not reached the potential of minimizing driver's cognitive load, which is the amount of mental processing power required. In this paper, we describe the development of a multimedia system that makes driving and navigation safer and easier by offering tools for route sharing in trip planning and video-based route guidance during driving. The system provides a multimodal interface for a user to share his/her route with others by drawing on a digital map, naturally incorporating human knowledge into the trip planning process. The system gives driving instructions by overlaying navigational arrows onto live video and providing synthesized voice to reduce the driver's cognitive load, in addition to presenting landmark images for key maneuvers. We describe our observations which had motivated the development of the system, detailed architecture and user interfaces, and finally discuss our initial test findings in the real-road driving context.

## 1. INTRODUCTION

Navigation is an important driving task. Navigation user interfaces have changed dramatically over the last few years due to the availability of electronic maps and GPS devices. With increasing popularity of GPS hardware and the Internet, travel by driving has become much more convenient in terms of trip planning and navigation. Most drivers rely on map services on the Internet for trip planning, and simple turn-by-turn guidance (turn instruction symbols and voices) for navigation during driving. However, most of the current systems emphasize machine intelligence without paying much attention to taking advantage of human knowledge and minimizing the driver's cognitive load, the mental process required. Map systems usually plan a route by optimizing based upon one or two criteria, which may not reflect dynamically changing traffic and human knowledge about a local street. During driving, a driver has to map abstract driving instructions, e.g., an arrow indicator on a map, to real world coordinates, which adds extra cognitive load to the cognitively intensive driving task. Therefore, new navigation technologies are not necessarily effective.

Driving is a focus-attention multi-task process. The driver needs to distribute attention among different aspects of the process. First of all, the driver needs to pay attention to issues directly related to driving, including the surrounding traffic, dashboard displays, and other influx of information on the road such as traffic lights and road signs. In addition, the driver may choose to talk to the passenger(s), listen to the radio, and talk on the cell phone. The limiting factor for information flow during driving is the driver's cognitive load. Although there are many ways to potentially reduce the cognitive load of a driver, in this paper we present two concepts, *route sharing* and *video-based navigation*, to enhance navigation effectiveness and consequently driving safety. Route sharing techniques provide a way to add human knowledge into trip planning. The shared landmark images and video-based navigation techniques directly reduce the driver's cognitive load during driving.

People usually have a good degree of knowledge about their local driving environment. They know the best way to a certain destination (school, company, shopping mall, park, etc). When people share their route face-to-face, they usually draw a rough map with street names on the route and directions. In particular, they may highlight complex intersections and prominent landmarks on the route (e.g., church, traffic lights, gas stations) to help others to follow the route. However, to the best our knowledge, the existing map systems provide no way for a driver to share his/her knowledge. In fact, there are many ways to incorporate human knowledge into a map system. A simple way is to use a multimodal interface that allows a user to draw a route on a map or modify an automatically generated route. An alternative way is to allow a driver to save GPS data with respect to a route and share it with others. Technically, the two methods are equivalent, provided that there is a way to convert the GPS data into graphical coordinates of the map system.

In this project, we propose a driver centered approach for enhancing driving safety. We carefully design and develop the system to fulfill two goals: i) to incorporate human knowledge into trip planning through route sharing; and ii) to reduce the driver's cognitive load on the road by providing additional visual cues to the traditional driving instructions. We develop a multimedia system, called NavStar, which offers a number of tools for route sharing in trip planning and video-based navigation during driving. It provides a multimodal interface for a driver to share his/her knowledge about the route with others by drawing on a map and sharing landmark images. When the driver goes on the road, the system automatically gives driving instructions in three ways: a) turn arrows overlaid on live video, b) synchronized voice directions and c) shared landmark images. The system has been integrated with an electronic map system.

## 2. RELATED WORK

There has been much research and applications on in-vehicle navigation. Early research focused on human factor tests in navigation displays [7]. Green et al. surveyed human factors for driver information systems [6]. They described objectives, principles and guidelines for the design of in-vehicle devices. Dale et al. studied the problem of generating natural route descriptions for navigational assistance [4]. Recently, Lee et at. focused on how to provide the situationally appropriate map information to drivers [8]. In contrast, there is few research on route sharing for trip planning since it is quite a new problem. Nevertheless, landmarks have great potential in both route sharing and driving contexts. Burnett et al. studied which landmarks are valued for driving navigation and their salient characteristics [2]. They found the significance of so called 'road furniture' landmarks, such as traffic lights and petrol stations. Landmarks can also support pedestrian navigation [5]. Combining visual cues with voice instructions was discussed in [1]. The remainder of the paper is as follows. Section 3 describes the system architecture and user interfaces; and Section 4 discusses issues on evaluation and Section 5 draws conclusions and points to ways forward.

## 3. SYSTEM ARCHITECTURE AND USER INTERFACES

### 3.1. Route Sharing

As noted above, the existing map systems provide no way for a driver to share his/her knowledge about a route. Inspired by the way that people share a route in the face-to-face condition, we develop a multimodal interface that allows a user to electronically share the route as s/he does on the paper. The system offers *route sharing* for trip planning through drawing on a map, showing landmark images and giving synthesized voice instructions. It naturally combines the human knowledge with machine intelligence to tackle the route planning task. The system is built on top of the Microsoft MapPoint map system on a Tablet PC and uses text-to-speech engines from SAPI 5.1 for the synthesized voice.

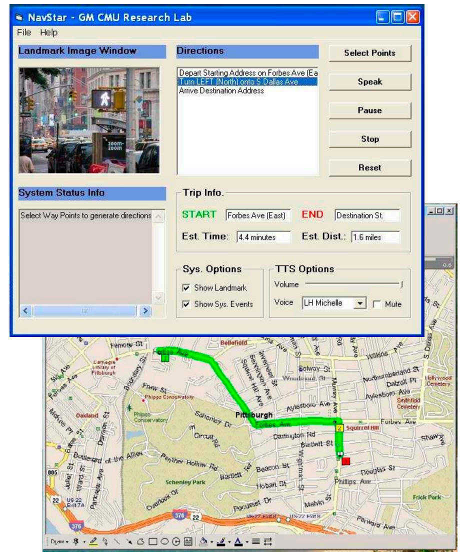


**Fig. 1**. User interfaces for *route sharing*. Input: text entry and touch screen; output: text directions, visual illustrations, landmark images and synthesized voice.

Fig.1 shows the route sharing interface which includes map and control two panels. Using the map panel, the user can choose one of different kinds of input to share a route of interest. For example, Henry shows Andrew how to drive from the campus to the cinema while avoiding the traffic. The easiest way is to use the keyboard and just enter the campus and cinema addresses. Or Henry draws the route on the map panel directly. If the route is too complex to draw once, Henry can select waypoints on the route and the system will automatically generate it. If some errors are found on the route, he can modify it by changing the destinations or adding and deleting some way points. This is an example of the combination of human knowledge and machine intelligence.

Furthermore, the control panel in Fig.1 uses different output modalities to present the route to the user. First, as the current map presentation, the user sees the turn-by-turn directions in the Directions window in the control panel and visual illustration of the route in the map panel, and also trip summary, like total time and distance. Second, s/he can get to know every intersection on the route through its corresponding landmark image. He clicks on one direction item, and then the landmark image is shown in the Landmark Image Window. The landmark images can be provided by the sharer or other sources. Third, s/he can also listen to vocal directions of the route. The synthesized voice will read out the landmark based directions for the user. For example, it says, turn right immediately after the BP gas station and then turn left at the first traffic light. The system also allows users to share the GPS data by copying from the USB flash drive.

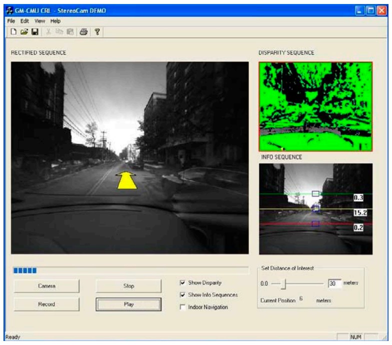

**Fig. 2**. User interface for *on-road navigation*. The left window shows overlaying the navigational arrows (e.g., go-ahead arrow) on live video. The right windows show the disparity sequence and real-time depths computed at three locations.

### 3.2. Video-based In-vehicle Navigation

Most current in-vehicle navigation systems give driving instructions in the form of map and synthesized voice. However, the lack-visual-cue instruction is indirect and sometimes attention demanding to the driver. In particular, it can cause problems when driving to an unfamiliar area or the driver's attention gets disturbed by other tasks, like receiving a phone call and talking. Sometimes drivers just do not have enough time to map from the driving instructions to real world situations. Then they get lost. Since the voice and paper instructions are not good enough and even not helpful in some cases, can we add additional visual cues for in-vehicle navigation?

Studies have shown that that users' performance can be significantly improved using a combination of directional arrows and photographs in navigation tasks, and also that this combination was highly preferred by users [3]. Following the same rational, we introduce the *video-based navigation* concept to in-vehicle navigation to reduce the driver's cognitive load and further enhance driving safety. Our system combines visual cues with voice instructions. It overlays navigational arrows on the road in live video and gives landmark images at each turning intersection. The live video shows the scene in front of the car from time to time and is captured by a video camera. The driver can still navigate the route by looking at the display even after missing the voice instructions.

Fig.2 shows the user interface for the video-based navigation. The system knows the vehicle's current location from GPS and the distance, $K$, from the next turning intersection from the map system. If $K \geq \lambda$, the system shows the turn arrow, as in Fig.2; otherwise, the system shows the turning arrow (left or right) based on the direction. $\lambda = 30m$ for outdoor and $\lambda = 6m$ for in-door. The task involves two steps: a) measuring the distances from the vehicle to its surrounding objects; and b) overlaying on the road in video the turn arrow with proper perspective. We will next briefly describe these two steps.

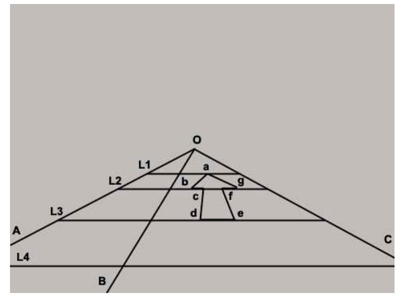


**Fig. 3**. Geometry template to depict the perspective arrow.

We adopt the stereo vision technique for measuring the distance because of its cost advantages over other devices, like radar and lidar. We use the Point Grey's Bumblebee, a two-lens stereo camera, in this project. Stereo processing is a three-part procedure. First, it establishes correspondence between image features in two views of the scene. Second, it calculates the relative displacement between feature coordinates in each image. The obtained displacements for every pixel constitute the disparity image. Third, it determines the 3D location of each feature point relative to the camera by knowing the camera geometry. Since we only need the depth for each point,$x_i$, and it can be computed as, $Z = \frac{Bf}{d}$, where $B$ is the distance between the optical centers of two stereo cameras, and called the baseline; $f$ is the focal length; $d$ is the disparity at $x_i$ and $Z$ is the distance (depth) between $x_i$ to the stereo camera center. $B = 12cm$ and $f = 2mm$.

Once we know where to mark, the next step is how to mark the navigational arrow *perspectively* on the road. Take the go-ahead case as an example. Fig.3 illustrates the geometry template we design to draw the perspective go-ahead arrow. $OA$ and $OC$ represent the left and right road borders, $O$ is the vanishing point and $OB$ shows the dividing line. $L1, L2, L3$ show the three horizontal lines controlling the vertical positions of the overlay arrow, and $L4$ represents the upper boundary of the in-vehicle portion. Furthermore, $a - g$ specify the seven key points of the arrow sign. We estimate the coordinates of $O, A, B, C$ from the data empirically by assuming the known car and camera geometry. Other vanishing point and lane border detection algorithms can be applied if high accuracy is desired. Coordinates of $a - g$ are further determined according to the template based on the estimations of $O, A, B, C$. The same technique also applies to depict turning arrows and we cannot elaborate it here due to limited space.

## 4. USER FEEDBACK AND DISCUSSIONS

Performing a rigorous and comprehensive evaluation of the system of this kind is an extremely costly work. Following the evaluation methodology of Dale et al. in [4], we have performed a small-scale expert evaluation in a task-based con-

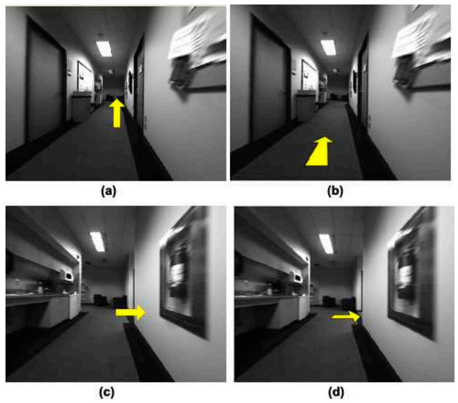


**Fig. 4**. Navigational arrows for the *on-road navigation* user interface in indoor context. (a)&(c): icon based go-ahead and turn-right arrows; (b)&(d): perspective go straight and turn right arrows.

text. Our user study group includes five users, one vehicle manufactory designer, two experienced drivers, and another two junior drivers. The goal of the evaluation is to obtain user feedback on the functions and user interfaces of the system. The experiment is carefully designed to minimize the many factors which can influence the feedback of the users. For the *video-based navigation*, we test the system in both indoor and outdoor contexts. For the indoor case, we run the system on a laptop, put it on a trolley and move in hallways. For outdoor case, we choose LCD as the in-vehicle display due to the limited amount of vehicles which include reconfigurable HUD displays. Originally, we design the icon-based navigational arrows and later refine the system with the current perspective arrows. Fig.4 shows four examples. All five users prefer the perspective arrows over icon ones. For the *route sharing*, we compare user feedback on two types of input modality, i.e., drawing a route on the map or selecting way points. Among five users, three prefer selecting way points while other two choose drawing the route.

Video-based navigation reduces the cognitive load of the system, but providing visual cues also has limitations. Specially, uncontrollable outdoor lighting conditions can degrade the accuracy of the stereo algorithm and even increase the cognitive load. For example, the cognitive load from watching the video of a night route (with all the details hidden because of the low-light conditions and the size of the screen) would be higher than reading a map. Combining video-based navigation with map systems would be a solution to the problem.

## 5. CONCLUSION AND FUTURE WORK

We have presented a driver centered multimedia system that offers tools for trip planning and in-vehicle navigation. We have proposed two new concepts, *route sharing* and *video-based navigation*, to enhance driving safety. Unlike existing in-vehicle navigation systems, our system provides multimodal interfaces to combine human knowledge with machine intelligence. It facilitates trip planning through electronic route sharing and reduces the cognitive load during driving by providing video-based navigation. Our initial findings in evaluation have given us encouraging user feedback. In the future work, we will conduct further comparison between the existing navigation systems and our approach and study the use of human knowledge sharing in the navigation context.

## 6. ACKNOWLEDGEMENT

This work was supported by the General Motors & Carnegie Mellon University CRL. We thank Jiazhi Ou for helpful discussions on SAPI 5.1 and Joy Zhang for video collection. We also thank Mon-Chu Chen for suggestions on this work and anonymous reviewers for their helpful comments.

## 7. REFERENCES

[1] G. Burnett. Usable vehicle navigation systems: Are we there yet? In *Vehicle Electronic Systems 2000 - European conference and exhibition*, June 2000.

[2] G. Burnett and D. Smith. Supporting the navigation task: Characteristics of ’good’ landmarks. In M. Hanson, editor, *Proceedings of the annual conference of the ergonomics society*, 2001.

[3] L. Chittaro and S. Burigat. Augmenting audio messages with visual directions in mobile guides: an evaluation of three applications. In *MobileHCI*, September 2005.

[4] R. Dale, S. Geldof, and J.-P. Prost. Using natural language generation in automatic route description. In *Journal of Research and Practice in Information Technology*, volume 37 of *1*, February 2005.

[5] J. Goodman, S. Brewster, and P. Gray. How can we best use landmarks to suppport older people in navigation? In *Behaviour and Information Technology*, volume 24 of *1*, pages 3–20, January-February 2005.

[6] P. Green, W. Levison, G. Paelke, and C. Serafin. Preliminary human factors design guidelines for driver information systems. In *UMTRI-93-21*, 1995.

[7] P. Green and M. Williams. Perspective in orientation/navigation displays: A human factors test. In *the Third International Conference on Vehicle Navigation and Information Systems (VNIS)*, 1992.

[8] J. Lee, J. Forlizzi, and S. E. Hudson. Studying the effectiveness of move: A contextually optimized in-vehicle navigation system. In *Conference on Human Factors in Computing Systems (CHI)*, 2005.