

Integrated Videos and Maps for Driving Directions

Billy Chen
Microsoft
One Microsoft Way
Redmond, WA 98052
bill.chen@microsoft.com

Boris Neubert
University of Konstanz
78457 Konstanz, Germany
neubert@inf.uni-konstanz.de

Eyal Ofek
Microsoft
One Microsoft Way
Redmond, WA 98052
eyalofek@microsoft.com

Oliver Deussen
University of Konstanz
78457 Konstanz, Germany
deussen@inf.uni-konstanz.de

Michael F. Cohen
Microsoft Research
One Microsoft Way
Redmond, WA 98052
mcohen@microsoft.com

ABSTRACT

While onboard navigation systems are gaining in importance, maps are still the medium of choice for laying out a route to a destination and for way finding. However, even with a map, one is almost always more comfortable navigating a route the second time due to the visual memory of the route. To make the first time navigating a route feel more familiar, we present a system that integrates a map with a video automatically constructed from panoramic imagery captured at close intervals along the route. The routing information is used to create a variable speed video depicting the route. During playback of the video, the frame and field of view are dynamically modulated to highlight salient features along the route and connect them back to the map. A user interface is demonstrated to allow exploration of the combined map, video, and textual driving directions. We discuss the construction of the hybrid map and video interface. Finally, we report the results of a study that provides evidence of the effectiveness of such a system for route following.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

General terms: Design, Human Factors

Keywords: video, map, routes, driving directions.

INTRODUCTION

Finding one's way to a destination has for centuries depended on the use of maps. Today, we are offered interactive maps that can draw a route, generate textual descriptions of routes, and provide interactive panning and zooming. Yet, even with these resources, the first time we navigate our way to a new destination we are uncertain of our way, having no visual

memory to rely on. Once we have driven a route, the next time is invariably much easier as our visual memory both gives us confidence and provides visual cues about what is approaching and where we should turn.

Companies such as Google and Microsoft have now driven a large number of the cities streets capturing 360 degree panoramas every few meters. These are delivered in, for example, Google's Streetview application. This data now provides new opportunities for almost any location to add visual previews to maps and textual descriptions. We show how the collected images can also provide a proxy visual experience of driving the route for the first time in the form of a video. Coupling this experience with more traditional maps and textual instructions leads to a new interface which we demonstrate.

This interface integrates an online map with an automatically generated summary video of the driving route. The goal of the interface is to enable the user to better understand his/her driving route, by interacting with the map and video. The integrated interface and the summary video are outputs of the system.

Although providing a literal video of the route would be trivial, few people would want to sit through a 20 minute video of a 20 minute drive. Simply speeding up the video 20 fold to create a one minute video would result in a useless blur. We demonstrate a method to automatically vary the speed of travel along the route to highlight turns and landmarks.

A video that always looks straight along the travel direction would miss landmarks as they pass by and would be jarring at turns. We demonstrate automatically generated videos that vary the field of view, aspect ratio, and view direction to provide a useful proxy for driving the route (see Figure 1). Finally, we integrate the video within an interactive mapping application that includes graphical and textual routes interspersed with video cliplets and the route video itself. We also report the results of a study of the effectiveness of our system.

We realize that in vehicle GPS and navigation systems may

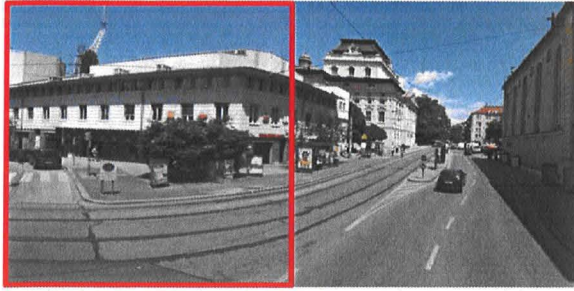


Figure 1: One frame of the route video. The field of view is automatically expanded to the left to emphasize the building landmark at this turn. The visual proxy is highlighted.

somewhat obviate the need for maps and the technology we introduce here. However, it will be some time before all cars have GPS systems. Future navigation systems might also incorporate similar rehearsal videos that depict approaching important landmarks and turns.

Our specific contributions thus include:

- Generation of a route video from a sequence of panoramas
- Automatic speed control based on the route parameters
- Automated orientation of the video to anticipate the route
- Visualization of landmarks and key images at turns by varying the field-of-view
- Integration of video into a mapping application
- A user interface providing video cliplets within the textual descriptions and along a route

A final contribution is the report of a user study that lends credence to the usefulness of a video preview. After a brief discussion of previous work, we describe and demonstrate each of the above aspects of our work.

RELATED WORK

Intuitively, visual memory of a route aids in wayfinding a second time a route is navigated. This intuition was confirmed in *To See But Not to See: A Case Study of Visual Agnosia* [4], where Humphreys and Riddoch discuss a patient who, because of an inability to form visual memories, has great difficulty finding his way along routes he commonly takes, even though he can draw a fairly accurate map of the route.

There are certainly differences across individuals with respect to how one performs wayfinding tasks. Jones and Healy [6] show that women orient themselves more by remembering landmarks while men use more abstract features such as maps. Combining both elements should optimize the recognition for both.

There is an extensive literature on wayfinding in the psychology literature related to maps, signage, etc. There have also been a number of studies on wayfinding in virtual worlds with and without using virtual reality (heads up) hardware or using driving simulators [1, 10]. If any conclusion can be drawn from such work, it is that different individuals use different strategies for wayfinding. Some are very visually ori-

ented (remembering the way a route looks), while others are more spatially oriented (mentally building and remembering the 2D structure of the route). Adding a video preview to a map should help both categories of people since the spatial information is amended by visual aids.

A seminal work in creating flexible video views of routes was presented almost 30 years ago in *MovieMaps* [9]. We have not found subsequent work related directly to video previews of driving routes, most likely since this type of data has not been widely available until recently.

There has been significant work studying the relationship between mapped representations of routes vs textual representations [11, 3] which have generally found these types of descriptions to be complementary. Lee, Tappe, and Klippel [8] discuss the use of icons to represent landmarks within maps to help in route following.

We automatically adapt the video's speed to the salience of the local portion of the route. This was inspired by *Linedrive*, by Agrawala and Stolte [2] which renders geographic route maps in a schematic style to provide both detail near the start and end point of a route while maintaining a consistent layout of the complete route. The system uses an optimization function to select and score the elements of a given route. The recent MOVE system [7] uses a similar approach for in-vehicle navigation systems and demonstrates that with such an approach users can more quickly comprehend the local region. We use a similar mechanism to assign more time and space in the video for important visual features along the route.

SYSTEM DESIGN

A vast amount of panoramic imagery is currently being commercially captured every few meters on every street in many cities and towns worldwide. A map database and related planning software create a route from which the relevant imagery can be extracted to create a video of the route. At the time of writing this paper we had only two routes of imagery available in Graz, Austria, which we used for our prototype system and experiments.

Landmarks along the routes are chosen, currently by hand, by indicating points on the map. A typical landmark is a building corner near a turn. A route video is then automatically generated. This video is finally combined with a specialized mapping software that shows the video along with the path on the map. Figure 2 shows the outline.

It is important to note here that our system takes as input a driving route and landmarks, currently chosen by hand. In the future we expect that landmarks will be automatically chosen, based on proximity to important turns, known sites of importance, visual importance, etc. Our system then automatically creates a summary video of the driving route. This video, integrated into a map application is presented to the user.

Results of a user study to study the effectiveness of combining video and maps is then reported. We close with some thoughts on the future impact of our system.

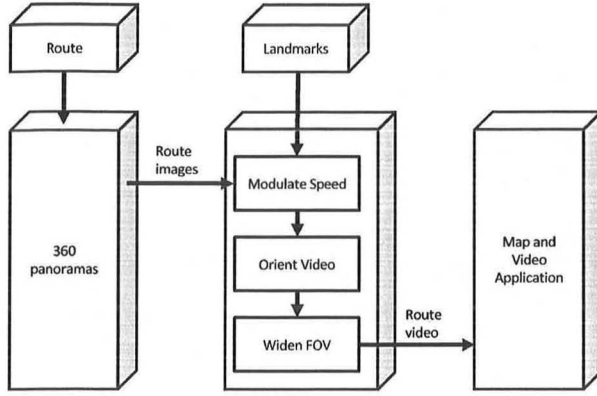


Figure 2: System design for our integrated map and video player.

CREATING THE ROUTE VIDEO

The route video is constructed from a series of 360 degree panoramas. These panoramas are unevenly spaced varying from 1 to 4 meters apart along the route (see Figure 4 (a)). Each is marked with latitude and longitude coordinates placing them on the route map. From this data, we first construct a path length parameterized list of panoramas. $X(s)$ is the position (lat/long) parameterized by path length, s . X_i and S_i represent the absolute position and path length position of the i^{th} panorama along the route. We also have a set of landmarks, L , with given positions, $X(L)$. We manually establish one landmark near each turn plus at any other notable points near the route. These landmarks typically represent building corners or facades of salient signs. Thus the $X(L)$ do not land exactly along the route's path, however, for each landmark, L_j , we can find the closet point along the path, S_j .

Modulating the Virtual Speed Along the Route

We wish to create a video that compresses the amount of time a real drive would take as much as possible while still retaining sufficient visual information. We modulate the speed of travel along the route such that long straight sections of the route have the highest apparent speed, shorter segments are traversed slower and sections related to turns and landmarks are the slowest.

More formally, we thus wish to construct a velocity profile, $v(s)$, which represents the speed along the path. We break the route up into segments between consecutive landmarks, e.g., the path length, D_j between landmarks j and $j+1$ is $S_{j+1} - S_j$. We set a *target velocity* for that segment, v_j^* to be,

$$v_j^* = a \cdot \log\left(\frac{D}{c} + 1\right) + \frac{2D}{c} + b$$

where D is in meters, and a , b , and c are parameters we have set to 200, 20, and 500 respectively. On shorter segments the constant third term, b , dominates. As the segments increase in length, the first term increases the speed. At the longest lengths the second term dominates, increasing the speed linearly, thereby keeping the time to complete a segment near some constant maximum.

We also set a slower target velocity around the landmarks,

$v_L^* = 10$ meters per second. This target speed covers the 40 meters before and after each landmark position, S_j .

We now have a discontinuous target velocity, $v^*(s)$ defined at all points along the path. The actual velocity is computed iteratively starting from $v_0 = 0$ and $s_0 = 0$, and proceeding at each frame, f , of the video.

$$v(s_f) = \alpha \cdot v(s_{f-1}) + (1 - \alpha)v^*(s_f)$$

and

$$s_f = s_{f-1} + v(s_{f-1}) \cdot \Delta t$$

where Δt is a frame time, e.g., one thirtieth of a second, and we have set α to be 0.25. The hysteresis term, α , provides the acceleration and deceleration around the landmarks and turns. Figure 3 shows an example velocity profile.

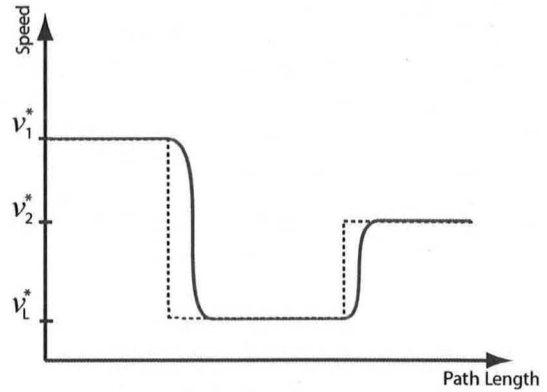


Figure 3: Predefined velocity for different legs of the route (dotted black) and final velocity profiles (solid blue). The hysteresis smooths out the velocity profile to produce the final velocity.

This results in a sequence of path positions, s_f , for each video frame, f (see Figure 4 (b)). The panorama located at the s_i closest to s_f is selected to pull the imagery for each frame of the video. Note that at faster speeds panoramas may be skipped while at slower speeds some panoramas may be used two or more times in a row depending on the exact spacing of the captured panoramas. This effectively decouples the panorama spacing from the apparent driving speed along the route.

Field of View

At each frame in the video we have a full 360 degree panorama from which to render. Although, a 50-60 field of view (FOV) is often considered the most "natural", it creates a very constricted view of the world not unlike looking through a very small windshield. We have found from informal experimentation that a 90 degree FOV achieves a good balance between providing a wide enough FOV to provide side to side context while exhibiting only mild wide angle distortion (see Figure 4 (c)). This field of view is further extended when approaching landmarks as described later.

Projection

We also have a choice of projection between normal linear perspective and a cylindrical (or other curved) projection. Curved projections minimize wide angle distortions

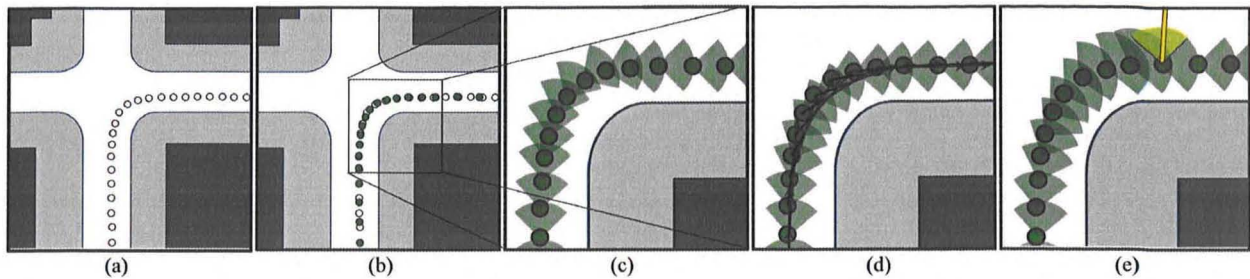


Figure 4: (a) Original spacing of panoramas. (b) Final spacing of video frames. (c) Straight ahead orientation. (d) Final orientation with look ahead. (e) Widening the view and freezing the landmark thumbnail.

while perspective projections keep straight lines straight. We have found cylindrical projection to be the slightly preferred mode, and more importantly, this integrates well with how we will depict landmarks.

Orienting the Video: Lookahead at Turns

Finally, we have a choice of how to orient (center) the view. The intuitive choice is to center the 90 degree window on the direction of travel. However, this produces rapid, disconcerting swings in orientation at turns. In a real car, one typically looks ahead in a direction to a point one expects to be at in the near future. We do exactly this by always orienting the view towards a point on the path 25 meters ahead. In other words we begin to turn our view anticipating the impending change of direction. This smooths out changes in view orientation in a natural feeling way as can be seen (in the accompanying video, http://research.microsoft.com/~cohen/vdd_webpage and) in Figure 4 (d).

Improving Recognition of Landmarks

Many people rely on visual landmarks to orient themselves. In particular, we want to make an added visual impression at turns and other important points along the route. The hope is that such an impression will aid in navigating the route after viewing the video. We thus modify the video to enhance the recognition of landmarks and turns.

As stated earlier, landmarks are currently identified manually by marking a point on the map. They may represent the corner of a building at an intersection or some other visually salient feature.

To emphasize the visual impression of the landmark we gradually widen the field of view as we proceed forward. The widening occurs only on the side where the landmark appears. For example, if the landmark is on the left (i.e., the route passes to its right) then we will open the left side of the video's frame, leaving the central 90 degree FOV in place and intact. We begin the FOV opening 40 meters before the landmark and continue until we are at the point on the route closest to the landmark. The landmark now sits orthogonal to the direction of motion. The total field of view at this position includes the original 90 degrees plus enough to center a 90 degree field of view looking to the left (or right) with the landmark point in the center (see Figure 4 (e)).

At this point two things happen. The FOV ceases to widen, and the 90 degree portion of the frame containing the land-

mark freezes and is framed, i.e., it *peels off* from the moving video. The central portion of the video continues forward, albeit slowly based on the speed profile. The frozen peel-off of the landmark remains for 3 more seconds and then disappears leaving the original 90 degree video to proceed. The frozen image of the landmark is the same one that appears on the map and in the annotated textual directions as described in the next section.

Figure 5 compares frames from three videos: 1) a video with a straight ahead orientation, 2) a video with a look ahead orientation, and 3) a video with widening FOV for encompassing a landmark. We demonstrate each of the above effects in the accompanying video.

INTEGRATING THE MAP AND VIDEO

We have designed an application that embeds the route video with a map and driving directions. In doing so, the map provides a geographical context for the video. Based on the map, viewers can understand the route's shape and proximity to nearby landmarks, occluded in the video. In addition, the map and driving directions enable interactive, non-linear browsing of the video by clicking on events in the directions or landmarks on the map.

User Interface

Figure 6 is a screenshot of the application. The application has 3 components: 1) the video, 2) the map, and 3) the driving directions. The video component is the route video as described in the previous section. It has a dynamic frame, extending when necessary to highlight landmarks. It also allows the user to linearly browse the video via standard Play/Pause and Stop buttons. Pressing Play/Pause stops or continues the video. Pressing Stop resets the video to the starting position.

Surrounding the video is a map which can be panned or zoomed. We draw the route representing the video as a semi-transparent green polyline. The polyline is composed from the latitude and longitude coordinates associated with each frame. Initially, the map is zoomed out so that the entire route is visible. On the route we draw a red dot representing the current position in the video. This dot moves along the green route as the video progresses. Along the route, we place thumbnails of landmarks. These thumbnails are the same images that peel-off in the route video. We place the thumbnails near the position where the landmark appears on

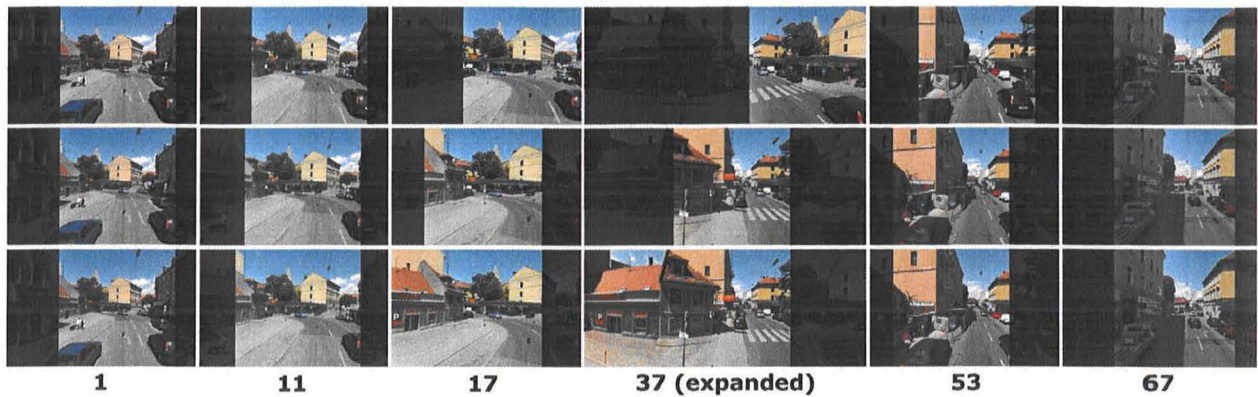


Figure 5: Comparing three different videos at a turn. Each column represents a frame. Each row is from a video. The highlighted portion of each frame is the visible region of the panorama in the video. The top row shows frames when looking straight ahead through a turn. This causes a rapid, disconcerting swing in orientation. The middle row shows frames when looking ahead into the turn, smoothing out the orientation swing. The bottom row shows the same frames where the field of view expands to encompass the landmark.



Figure 6: Integrated video and mapping application.

the route. Clicking on a thumbnail resets both the current position on the map and the temporal position in the video, enabling non-linear browsing of the video by trading time for space.

Along side the map, we add a list of driving directions. Each direction has a textual description and an accompanying thumbnail. These thumbnails are also clickable, which enable setting the current position on the map and the temporal position in the video.

Interaction

Using this interface, there are 2 interaction scenarios: 1) play the route video and 2) interactively select and play cliplets of the route. In the first scenario, the viewer receives a quick summary of the entire route. In the second scenario, the viewer dives into more detail about individual route segments.

To watch a summary of the route, the user presses Play on the video component. This action brings the map from its current location to the start of the route and zooms in appropriately. We postpone the discussion on zoom level selection to the following section. After returning to the start position, the video begins to play, with the icon travelling along the route. While travelling, the map pans and zooms appropriately to balance centering over the icon and panning to provide current and future context on the map.

As the video approaches a landmark, it dynamically extends its frame, as shown in Figure 7. This frame is held for 3 seconds to provide the viewer with a snapshot of the landmark. Meanwhile, the appropriate landmark on the map and driving directions flashes to connect this landmark to the map.

In the second scenario, the viewer dives into more detail on individual video segments. She clicks on thumbnails on the map or driving directions. This activates the map pan and zoom mechanism to smoothly transition from its current position and zoom level to the target one. The following section discusses how we automate panning and zooming on the map based on the user interactions.

Automated Map Controls

As one explores the map and the associated video of the chosen route, a choice has to be made about how to position the map by panning, and the area to focus on by zooming. The simplest choice would be to select a static zoom level and position that shows the full route. The user could then manually adjust the pan and zoom to his or her perceived needs. Although our interface does allow manual settings, we also provide an automated pan/zoom based on the users actions as described below.

Note that there is a one to one correspondence between each frame in the video and a position along the route, and we also have a velocity associated with each point. We will leverage these in deciding on both the pan and zoom settings.

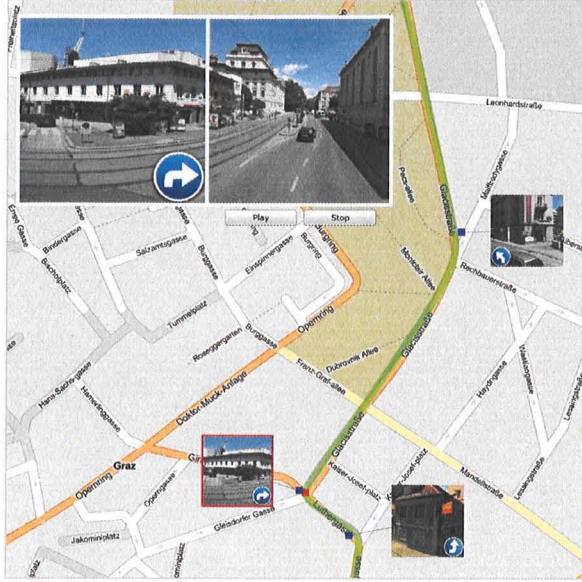


Figure 7: Map with extended video frame. The extended frame is shown to the left of the video. The associated thumbnail is highlighted in red on the map and on the driving directions (not shown).

As described above, there are two ways the current map position in the video changes. Simply playing the video moves the position along the route, or the position jumps from one point in the video to another when the user requests to examine a particular part of the route. We discuss each of these in turn.

Pan and Zoom during Video Play Our goals during the video playback is to provide information about the location of the current frame, as well as to provide context around this location. The first goal is achieved, by placing a moving red icon representing the current location depicted in the video.

The amount of context revealed around this point by choosing the zoom level is inspired by Igarashi and Hinckley's work [5] on speed dependent zooming in browsing text documents. As in their work, we set the zoom level to cover a larger area when the velocity (the apparent speed of the video) is higher and zoom in when the video slows down for short stretches and at turns and landmarks.

In particular, the target zoom level, z^* , is set to be the lower of a fixed maximum, or a level such that at the current velocity the video's location would traverse the window in 5 seconds. One exception is that at the start of the video, the target zoom level is set to reveal the full route. The actual zoom level, z , is smoothed in time through hysteresis

$$z_f = \alpha_{map} z_{f-1} + (1 - \alpha_{map}) z^*$$

where $\alpha_{map} = 0.25$.

We also have freedom in panning the map to keep the current position in view. One simple solution would be to pan the map to hold the current position in the center of the map

region below the video. This would essentially slide the map under the current position. We modify this idea to create a desired position, p^* , that will reveal more of the future route than the past. We do this by setting p^* to lie on a circle with radius one quarter the width of the map, and at a point on the circle opposite the momentary velocity. As with the zoom level we apply the same hysteresis, thus

$$p_f = \alpha_{map} p_{f-1} + (1 - \alpha_{map}) p_f^*$$

The result is a map that always show the current location and a sufficient amount of the future to understand the context of the current scene shown in the video.

Jumping Between Points When the user requests to jump from one point in the video to another we are presented with an instantaneous large change in the desired position, p_f^* , and a similar jump in desired zoom level. We accommodate this jump by panning and zooming over a fixed interval set to 2 seconds. We also compute another zoom level, z_{max}^* that is either a level just enough to show both the start and end positions of the jump, or if the points are close enough together then z_{max}^* is simply the minimum of the two endpoint zoom levels.

The position, p^* , shifts over the two seconds from the starting point to the ending point using a standard sigmoid function. This function provides a smooth acceleration between the points, as compared to linear interpolation. The target zoom, z^* , moves from the initial zoom through z_{max}^* at the one second mark to the final position's associated zoom level at two seconds. This interpolation uses a cosine curve to smooth the changes.

The result for long jumps is that the map smoothly zooms out to show both the starting and end points of the jump and then back in as the map position smoothly moves from the start to the end of the jump. For shorter jumps, the zoom level changes little, while the map repositions itself to the new location.

USER STUDY

We conducted a user study to measure the effectiveness of the video and map interface (the *Videomap* interface, Figure 6). We compared the *Videomap* to an interface consisting of only photos and a map, which we call *Photomap* (See Figure 8). In the *Photomap*, the user can see thumbnails at each turn of the driving route. These thumbnails are shown embedded in the driving directions and on the map.

Given either the *Videomap* or *Photomap* interface, users have 5 minutes to familiarize themselves with a driving route by using the interface. For example, with the *Videomap*, the user can play through the video, navigate the map, zoom into details, etc. In the *Photomap*, the user may click on different photos, zoom in on them, pan the map, etc. During this time we also provide the participant with printed driving directions for consultation.

To measure how much of the driving route is retained, ideally we would ask the participant to drive the actual route, while measuring the number of successfully navigated turns.

Unfortunately, if a participant gets lost along a segment of the route, it is difficult to measure retention in the remaining portion. Additionally, our datasets are captured in Graz, Austria, which is prohibitive for conducting the experiment in person. For these reasons, we chose to conduct a more controlled experiment in which the participant performs a virtual drive along that route (explained below). During this drive they also have printed driving directions for consultation. We measured users' driving performance in terms of the percent of successful decisions at intersections.

Since needing to consult a map while driving may be a distraction that should be avoided in a real world setting, we also record the number of times the user consulted the printed driving directions per decision point. After the driving test, users are asked to fill out a questionnaire about the technique to gather intuition about the interfaces from qualitative impressions.

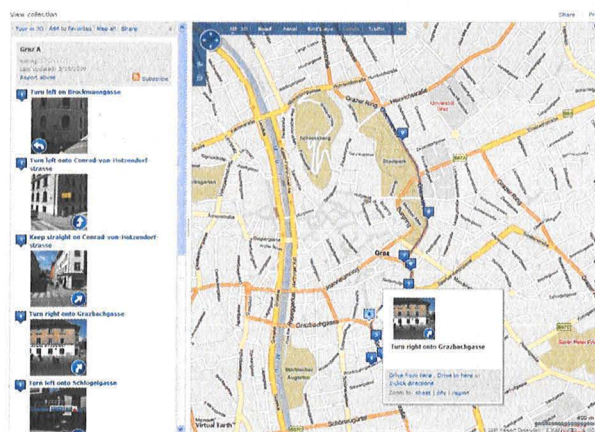


Figure 8: The Photomap interface. We create a collection of photos using Microsoft Virtual Earth. Photos are placed at turns in the driving route. These photos appear as thumbnails on the left and on the map. Clicking on the pushpins on the map reveals the associated thumbnail. The route is also drawn on the map.

Apparatus and Participants

We used PCs capable of streaming and rendering digital video with 800 x 800 resolution at 30 frames-per-second. The participants consisted of 6 females and 14 males, recruited within our department. All users spent more than 6 hours on a computer per day. 4 users were familiar with online maps that utilize photos. All users, save 1, drive regularly to work. All users were familiar with maps. Regarding previous knowledge of in-car navigation systems, 5 users owned and used such systems, 8 users were familiar with, but did not own a system, and 7 users had not used them at all.

Task and Stimuli

Each task consists of a participant learning a route by using either the Videomap or the Photomap interface. The participants are given 5 minutes to familiarize themselves with the route by using the interface and are also given a set of printed driving directions. The printed driving directions are given

to the user at the same time they are presented with the interface. In the case where the interface is completely new to the participant, we allowed an initial learning period on a different dataset.

After the 5 minutes for the participant to learn the route, the user performs a virtual drive. The virtual drive consists of a large 120 degree FOV straight-ahead video of the route without any of the variable speed and orientation enhancements in the Videomap interface. The wider FOV roughly matches the view as seen through a car's windshield. In other words, the virtual drive is simply a video driving along the route specified by the driving directions.

As intersections are approached, the video graphically depicts options (e.g., left, straight, right) for 5 seconds. Subjects are asked to verbally state which way they think they should go to follow the route. Figure 9 illustrates one frame of the test video, showing possible turn options at that intersection. We ensure that the turn options are shown (and removed) before the video implies any turn direction (i.e. moving to the left lane for a left turn).

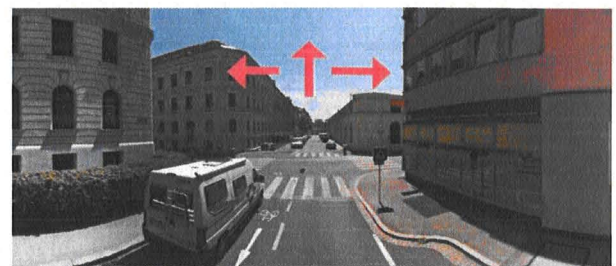


Figure 9: One frame of the virtual drive. The turn options are shown in red.

During the driving simulation, we record the number of correct answers at each intersection and the number of times the participant glances at the map to consult the printed driving directions.

Procedure and Design

We used a 2 Technique x 2 Datasets within-participants design comparing the Photomap to the Videomap technique. Each participant is presented with two datasets, A, and B, representing two different driving routes. The participant uses one of the techniques (e.g. Videomap or Photomap) on dataset A, and the other technique on dataset B. The dependent variables we are measuring are Success Rate and Looks. The success rate is the percent of correct choices at intersections. For example, if Ann chose correctly at 6 out of 12 intersections in the driving simulation, then her success rate is 0.5. The Looks is the number of times the participant consulted the printed driving directions, per decision point. The presentation order was counter-balanced across participants.

Results

The study took an average of 45 minutes per user. We performed a two-way repeated measures analysis of variance (RM-ANOVA) on Success Rate and Looks. We found a significant effect for technique on Success Rate ($F(1,36) = 4.113165$; $p = 0.002$). The results show that on average users

were more successful in navigating the driving turns after using the Videomap interface, as compared to using the Photomap interface. Figure 10 shows a plot of the mean success rate for both techniques.

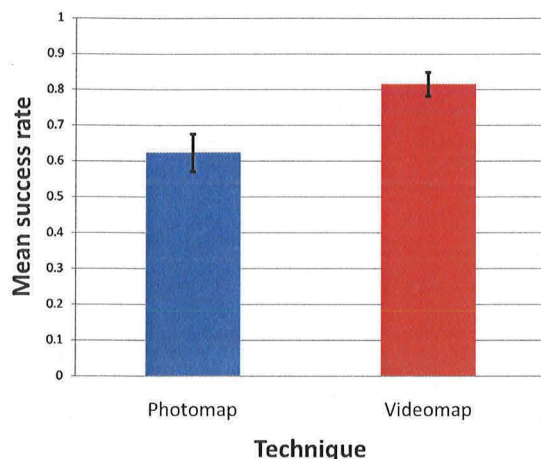


Figure 10: Plot of mean success rates for the two techniques. 95% confidence intervals are shown.

The difference in mean success rate was not statistically significant across the two datasets ($F(1,36) = 4.113165$; $p=0.99$). This means that the differences between route A and B did not affect the mean success rate.

We also found a significant effect for technique on the number of Looks ($F(1, 24) = 4.25967$; $p = 0.00177$). The results indicate that on average Videomap users tended to rely less on the printed driving directions during the driving simulation. Figure 11 plots the number of Looks per decision point.

After performing each test, participants were asked to answer a questionnaire. When asked which technique he/she preferred, 17 out of 20 participants preferred the Videomap over the Photomap. Participants were also asked the following two questions for each technique: 1) When driving the video simulation, I needed the printed map and directions to navigate the turns. 2) When driving the video simulation, I needed more time to decide each turn. The questions were answered using a Likert scale (1 = strongly disagree, 5 = strongly agree). In both techniques, printed maps and driving directions were needed. However, when asked about needing more time, Photomap users on average agreed while Videomap users on average neither agreed or disagreed ($F(1,36) = 4.113$; $p = 0.026$). Figure 12 visualizes the mean and 95% confidence levels of the participants answers.

The responses to the questionnaire resonated well with their performance using both two techniques. The majority of the participants preferred the Videomap over the Photomap. The main reason is that the Videomap provides a visual context that is absent in a single image. This idea is exemplified well in several user comments. One participant noted, “[The] photos are too close to the intersections, forcing me to turn late. Videomap seemed easier, and I have years and years of maps experience.” When looking at the thumbnails in the Photomap, another user said, “Which way am I coming

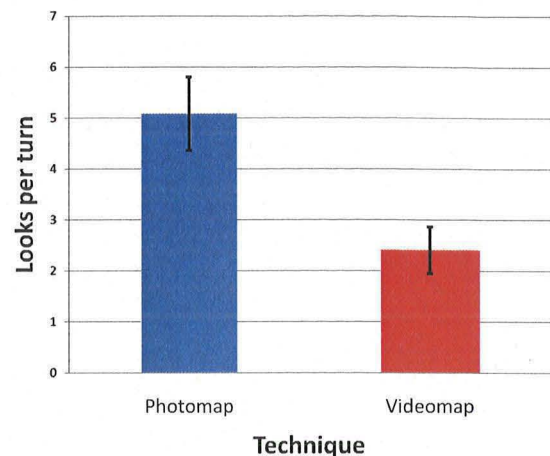


Figure 11: Plot of looks per turn for the two techniques.

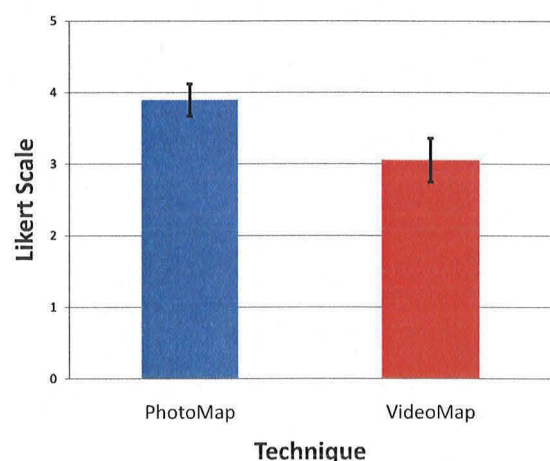


Figure 12: Plot of average user responses to the statement “When driving the video simulation, I needed more time to decide each turn.” A “1” response means “strongly disagree” while a “5” response means “strongly agree”.

from?” However, when using the Videomap, the same user said, “That’s cool, it [the video] leads up to the waypoint.” This comment shows that the Videomap provides a context to the turns, which is lacking in the Photomap.

While the study indicates quantitatively and qualitatively that the Videomap increases driver performance over the Photomap, the interface itself could also be improved. One concern that surfaced was that the Videomap had too many dynamic elements. Since the video and the map can both move, the users attention can be split between them.

Caveats and Summary

We realize there are many aspects of the user study that are not ideal for assessing the real world performance. The fact that the test was given immediately after the training may not reflect what would occur in a real application. The test itself also differs from having to drive and navigate in a real

automobile, but creating such a “real” test was beyond our reach for this initial prototype. In addition, ideally, the video used for the test should show different lighting conditions than the one in the Videomap. However, due to limitations in data, we use the same visuals although presented as a simple straight forward video.

Caveats aside, the user study confirmed our hypothesis that the Videomap improves driving performance over the Photomap. Surprisingly, even though photos do help in understanding the turns, if placed too close or too far from the intersection, the photos can be detrimental to driving performance. In contrast, a video of the turn alleviates the problem of finding the “perfect” photo position and also provides visual context. Quantitatively, users were more successful navigating the turns after using the Videomap. Qualitatively, Videomap users felt they needed less time to decide where to turn.

FUTURE WORK AND CONCLUSIONS

Newly acquired imagery from the world’s streets is opening up many new applications. Here, we have demonstrated for the first time, a prototype that combines maps with imagery assembled into a video to depict driving a route to help a user in wayfinding along that route. A user study provides evidence that such a system is useful by enhancing the user’s visual memory of the route.

Although we demonstrate an interface to leverage the new visual data, this work has created as many questions as answers. For example, we do not know:

- How long does the visual memory from the application last?
- How will lighting differences effect the outcome?
- How will the differing modalities between viewing a video and actually driving in a car effect the usefulness of the video?
- How long a route can one retain in memory?

We are formulating how to best answer these questions.

Even with leaving these questions unanswered, we feel we have provided a starting point to explore how to use this new wealth of visual information in the task of wayfinding.

There are many other practical issues we are currently exploring. The data collection, even when it covers all roads may not include images collected moving in both directions. Thus, we will need to piece together a route from panoramas sequences that run backwards in time. This will give an impression of driving on the opposite side of the street and will exhibit other vehicles and pedestrians driving/walking backwards. In addition, the assembly will require clips from different points in time, and from which the capture vehicle did not turn at turns in the route. Thus there may be sudden lighting changes and possibly inconsistent positions. We do not know what the effect of such distractions might be on the efficacy of the application.

We also are exploring ways to automate the selection of appropriate landmarks. Clearly, salient features at turns are important, but one would also like to include memorable objects along the way, such as gas stations, signs, etc.

Moving forward, we expect to see many other uses of the visual data that has been and is being captured. We look forward to continuing in this endeavor and hope we have inspired others to work in this area.

REFERENCES

1. Vlada Aginskya, Catherine Harris, Ronald Rensinka, and Jack Beusmansa. Two strategies for learning a route in a driving simulator. *Journal of Environmental Psychology*, 17:317–331, December 1997.
2. Maneesh Agrawala and Chris Stolte. Rendering effective route maps: Improving usability through generalization. In *Proceedings of ACM SIGGRAPH 2001*, Computer Graphics Proceedings, Annual Conference Series, pages 241–250, August 2001.
3. M. Denis. The description of routes: A cognitive approach to the production of spatial discourse. *Current Psychology of Cognition*, (16):409–458, 1997.
4. Glyn W. Humphreys and M. Jane Riddoch. *To See But Not to See: A Case Study of Visual Agnosia*. Psychology Press Ltd., 1987.
5. Takeo Igarashi and Ken Hinckley. Speed-dependent automatic zooming for browsing large documents. In *UIST '00: Proceedings of the 13th annual ACM symposium on User interface software and technology*, pages 139–148, New York, NY, USA, 2000. ACM.
6. C. Jones and S. Healy. Differences in cue use and spatial memory in men and women. *Proceedings of the Royal Society*, 2006.
7. Joonhwan Lee, Jodi Forlizzi, and Scott E. Hudson. Studying the effectiveness of move: a contextually optimized in-vehicle navigation system. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 571–580, New York, NY, USA, 2005. ACM.
8. Paul U. Lee, Heike Tappe, and Alexander Klippel. Acquisition of landmark knowledge from static and dynamic presentation of route maps. *KI*, 16(4):32–34, 2002.
9. Andrew Lippman. Movie-maps: An application of the optical videodisc to computer graphics. *SIGGRAPH Comput. Graph.*, 14(3):32–42, 1980.
10. Glenna Satalich. Navigation and wayfinding in virtual reality: Finding proper tools and cues to enhance navigation awareness. Master’s thesis, HitLab, University of Washington, Seattle, WA, 1995.
11. Barbara Tversky and Paul U. Lee. Pictorial and verbal tools for conveying routes. In *COSIT '99: Proceedings of the International Conference on Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science*, pages 51–64, London, UK, 1999. Springer-Verlag.