

Worldlets - 3D Thumbnails for Wayfinding in Virtual Environments

T. Todd Elvins
todd@sdsc.edu

San Diego Supercomputer Center
P.O. Box 85608
San Diego, CA 92186-9784, USA

David R. Nadeau
nadeau@sdsc.edu

David Kirsh
kirsh@cogsci.ucsd.edu
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0515, USA

ABSTRACT

Virtual environment landmarks are essential in wayfinding: they anchor routes through a region and provide memorable destinations to return to later. Current virtual environment browsers provide user interface menus that characterize available travel destinations via landmark textual descriptions or thumbnail images. Such characterizations lack the depth cues and context needed to reliably recognize 3D landmarks. This paper introduces a new user interface affordance that captures a 3D representation of a virtual environment landmark into a 3D thumbnail, called a *worldlet*. Each worldlet is a miniature virtual world fragment that may be interactively viewed in 3D, enabling a traveler to gain first-person experience with a travel destination. In a pilot study conducted to compare textual, image, and worldlet landmark representations within a wayfinding task, worldlet use significantly reduced the overall travel time and distance traversed, virtually eliminating unnecessary backtracking.

KEYWORDS: 3D thumbnails, wayfinding, VRML, virtual reality.

INTRODUCTION

Wayfinding is "the ability to find a way to a particular location in an expedient manner and to recognize the destination when reached" [13]. Travelers find their way using *survey*, *procedural*, and *landmark knowledge* [5, 13, 14, 9]. Each type of knowledge helps the traveler construct a *cognitive map* of a region and thereafter navigate using that map [10, 11].

Survey knowledge provides a map-like, bird's eye view of a region and contains spatial information including locations, orientations, and sizes of regional features. Procedural knowledge characterizes a region by memorized sequences of actions that construct routes to desired destinations. Landmark knowledge records the visual features of landmarks, including their 3D shape, size, texture, etc. [2, 9]. For a structure to be a landmark, it must have high *imagability*: it must be distinctive and memorable [10].

Permission to make digital/hard copies of all or part of this material for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication and its date appear, and notice is given that copyright is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires specific permission and/or fee.

UIST 97 Banff, Alberta, Canada

Copyright 1997 ACM 0-89791-881-9/97/10...\$3.50

Landmarks are the subject of landmark knowledge, but also play a part in survey and procedural knowledge. In survey knowledge, landmarks provide regional anchors with which to calibrate distances and directions. In procedural knowledge, landmarks mark decision points along a route, helping in the recall of procedures to get to and from destinations of interest. Overall, landmarks help to structure an environment and provide directional cues to facilitate wayfinding.

Landmarks also influence the search strategies used by travelers. With no a priori knowledge of a destination's location, a traveler is forced to use a *naive*, exhaustive search of the region. Landmarks provide directional cues with which to steer such a naive search. In a *primed* search, the traveler knows the destination's location and can move there directly, navigating by survey, procedural, and landmark knowledge. In practice, travelers use a combination of naive and primed searches. The location of a curio shop, for instance, may be recalled as "near the cathedral," enabling the traveler to use a primed search to the cathedral landmark, then a bounded naive search in the cathedral's vicinity to find the curio shop.

In city planning, the *legibility* of an environment characterizes "the ease with which its parts can be recognized and can be organized into a coherent pattern" [10]. Legibility expresses the ease with which a traveler may gain wayfinding knowledge and later apply that knowledge to search for and reach a destination. For instance, a city with distinctive landmarks, a clear city structure (such as a street grid) and well-marked thoroughfares is legible.

In virtual environment design, the use of landmarks and structure is essential in establishing an environment's legibility. In a virtual environment lacking a structural framework and directional cues, such as landmarks, travelers easily become disoriented and are unable to search for destinations or construct an accurate cognitive map of the region [5]. Such a virtual environment is illegible.

Real and virtual world travel *guidebooks* describe available landmarks and tourist attractions, highlighting regional features that enhance the environment's legibility. Guidebook descriptions facilitate wayfinding by priming a traveler's cognitive map with landmark knowledge, preparing them for exploration of the actual environment.

Similar to travel guidebooks, virtual environment browsers facilitate wayfinding by providing menus of available destinations. Selection of a menu item "jumps" the traveler to the destination, providing them a short-cut to a point of interest. Systematic exploration of all destinations listed on a menu enables a traveler to learn an environment and prime their cognitive map with landmark knowledge.

Whereas a traveler's landmark knowledge characterizes a destination by its 3D shape, size, texture, and so forth, browser menus and guidebooks characterize destinations by textual descriptions or images. This representation mismatch reduces the effectiveness of destination menus and guidebooks. Unable to engage their memory of 3D landmarks to recognize destinations of interest, travelers may resort to a naive, exhaustive search to find a desired landmark.

This paper introduces a user interface affordance to increase the effectiveness of landmark menus and guidebooks. This affordance, called a *worldlet*, reduces the mismatch between a traveler's landmark knowledge and the landmark representation used in menus and guidebooks.

LANDMARK REPRESENTATION LEGIBILITY

Analogous to virtual environment legibility, the *legibility* of a landmark representation technique expresses the ease with which it may be used to facilitate wayfinding. As a basis for comparing landmark representations, we propose the following legibility criteria:

- *imagability*: A landmark representation has good imagability if it provides a faithful rendition of a landmark, preserving the landmark's own imagability. Key landmark features recorded within a traveler's landmark knowledge, such as 3D shape, size, and texture, should be expressed in the landmark representation.
- *landmark context*: In addition to the landmark itself, a landmark representation should include portions of the surrounding area. Such context supplies additional visual cues and enables a person to understand the larger configuration of the environment [6, 7, 13].
- *traveler context*: Where landmark context expresses the relationship between a landmark and its surroundings, traveler context expresses the relationship between the landmark and the traveler. Travelers are better at recognizing a landmark when it is viewed from the direction in which they first encountered it along a route [1]. Traveler context expresses this notion of an expected view of a landmark, such as a view of a prominent skyscraper from street level.
- *multiple vantage points*: While traveler context provides a typical vantage point of a landmark, additional vantage points enable a more comprehensive understanding of a landmark and its context [10].

In addition to satisfying these criteria, a good landmark representation technique should be efficient to implement and have broad applicability.

RELATED WORK

Landmark representations are used to characterize destinations listed within the user interface of virtual environment browsers and within virtual environments themselves. A browser may, for instance, list available destinations within a pull-down menu or in an on-line travel guidebook. A virtual environment may provide clickable *anchor* shapes distributed throughout the environment. Clicking on a door anchor shape in a virtual room, for instance, may select and load a new virtual environment presumed to be behind the door.

Landmark representation use may be classified into two broad categories:

- *World selection*: A virtual world is an independently loadable destination environment with its own shapes, lights, structural layout, and internal design themes. Browser world menus, guidebooks, or virtual environment anchors provide a selection of destination worlds that, when clicked upon, load the selected world into the traveler's browser.
- *Viewpoint selection*: A viewpoint is a preferred vantage point within the currently viewed virtual environment. Viewpoints are characterized by a position and orientation. Browser viewpoint menus, guidebooks, or virtual environment anchors provide a selection of vantage points that, when clicked upon, jump the traveler to the selected destination.

Using the landmark representation legibility criteria above, we consider each of several representation techniques used for browser destination menus and guidebooks, or in virtual environments themselves.

Textual Descriptions

Textual descriptions are the dominant method used to represent virtual environment landmarks in viewpoint and world selection user interfaces. HTML pages, for instance, often provide lists of available Web-based virtual environments (such as those authored in VRML, the Virtual Reality Modeling Language [3]), each one characterized by a URL, an environment name, and/or a brief description. Within VRML worlds, textual descriptions characterize viewpoints and describe destinations associated with clickable anchor shapes.

In terms of our landmark representation legibility criteria, textual descriptions provide poor imagability, landmark context, traveler context, and support for multiple vantage points. The subjective, and often brief nature of textual descriptions limits their ability to express important visual characteristics of a landmark and its context. The complex 3D shape of a distinctive building, for instance, may be difficult to describe. The 3D position of a traveler in

relation to a landmark is often omitted from textual descriptions, providing little support for traveler context. When traveler context is present in a textual description, it characterizes the author's traveler context, and not necessarily that of other travelers. Finally, the need to keep textual descriptions relatively brief prevents a description from providing descriptions for more than a few vantage points. Overall, textual descriptions provide a relatively illegible form of landmark representation.

Images and Icons

Clickable icons, thumbnail images, and image maps provide common visual wayfinding aids. In a 3D context, games often provide "jump gates" onto which images of remote destinations are texture mapped. Stepping through such a gate jumps the traveler to the destination depicted on the gate.

In terms of our legibility criteria, images provide improved imagability, landmark context, and traveler context, compared to textual descriptions, but do not support multiple vantage points. An image capturing a canonical view of a landmark can show important visual details difficult to describe textually. For complex 3D landmarks, or for landmarks placed in complex contexts, a single image may be insufficient. Overall, image-based descriptions provide an improved, but somewhat limited form of landmark representation.

Image Mosaics

An image mosaic groups together multiple captured images into a traversable structure. Apple's QuickTime VR, for instance, can use images captured from multiple viewing angles at the same viewing position [4]. By ordering images within a traveler-centered cylindrical structure, QuickTime VR can provide a traveler the ability to look in any direction through automatic selection of an appropriate image from the structure. By chaining multiple mosaic structures together, the content author can create a walk-through path that hops from vantage point to vantage point. Similar image mosaics can be used to create zoom paths, pan paths, and so forth.

Using our landmark representation legibility criteria, the inclusion of multiple images within an image mosaic improves imagability, landmark context, and traveler context compared to that of a single image. Mosaics also offer multiple vantage points, but only those authored into the mosaic structure. In a typical use, a QuickTime VR cylindrical mosaic provides multiple viewing angles, but only a single viewing position. Such a mosaic structure may not provide sufficient depth information to facilitate recognition of complex 3D environments. Overall, mosaic-based descriptions provide increased landmark representation legibility, but are still limited in the range of vantage points they support.

Miniature Worlds and Maps

Most 3D environment browsers enable the traveler to zoom out and view the world in miniature, thereby gaining survey

knowledge. Stoakley et al extend this notion by creating a world in miniature (or WIM) embedded within the main world [15, 12]. The miniature world duplicates all elements of the main world and adds an icon denoting the traveler's position and orientation. Held within the traveler's virtual hand, the traveler can reach into the miniature and reposition world content or themselves. Simultaneously, the outer main world is updated to match the altered miniature, automatically adjusting the positions of shapes, or the traveler.

Similarly, 2D and 3D maps are frequently found as navigation aids within virtual environments. 3D games, for instance, often provide a 3D reduced-detail map in which an icon denotes the player's location. Such maps can be panned, zoomed, and rotated to provide alternate vantage points similar to that possible with miniature worlds.

Using our legibility criteria, miniature worlds and 3D maps do a good job of supporting imagability, landmark context, and multiple vantage points. Complex 3D landmarks, and their context, are accurately represented. The dominant use of a bird's eye view of the miniature or map, however, somewhat limits the range of vantage points available and reduces support for traveler context. For instance, a landmark typically viewed and recognized at street level may be unrecognizable when viewed in a miniature from above.

The WIM approach is primarily designed to support a map view of a region within an emersive environment. This special-purpose implementation has a few drawbacks. A WIM is held within the traveler's virtual hand, occupying space in the main world and moving as the traveler moves. This implementation doubles the world's rendering time and requires that the traveler maintain adequate space in front of them to avoid collision between the WIM and main world features.

Additionally, the presence of the WIM within the main world may clash visually, affecting the environment's stylistic integrity. A WIM of a mountain landscape hovering within the cockpit of a virtual aircraft simulator, for instance, would look out of place.

WIMs appear best suited within bounded environments, such as virtual rooms with walls and floors. In an unbound environment, such as one for a galaxy simulation, the similarly unbounded miniature may be indistinct and become easily lost in the background of the main world in which it hovers.

Overall, a miniature 3D representation of a virtual world landmark provides improved legibility over that available with textual descriptions, images, or image mosaics. WIMs illustrate a special-purpose approach to using 3D representations within an emersive environment. This paper introduces a general-purpose technique for creating 3D landmark representations.

WORLDLETS

A *worldlet* is a 3D analog to a traditional 2D thumbnail image or photograph. Like a photograph, a worldlet is associated with a viewing position and orientation within a world. Whereas a photograph captures the view of the world as projected onto a 2D film plane, a worldlet captures the set of 3D shapes falling within the viewpoint's viewing volume. Where a photograph clips away shapes that project off the edges of the film, a worldlet clips away shapes that fall outside of the viewing volume.

Like a thumbnail image, a worldlet provides a reduced-detail representation of larger content. Whereas a thumbnail image reduces detail by down-sampling, the worldlet reduces detail by clipping away shapes outside of a viewing volume.

In typical use, the worldlet's viewpoint is aimed at an important landmark, and the worldlet's captured shapes reconstruct that landmark and its associated context. When viewed within an interactive 3D browser, a worldlet provides a manipulatable 3D thumbnail representation of the landmark.

We have developed two types of worldlets:

- A *frustum worldlet* contains shapes within a standard pie-shaped viewing frustum, positioned and oriented based upon a selected viewpoint. When viewed, a frustum worldlet looks like a pie-shaped fragment clipped from the larger world.
- A *spherical worldlet* contains shapes within a spherical viewing bubble, positioned at a selected viewpoint with a 360 degree field of view. When displayed, a spherical worldlet looks like a ball-shaped world fragment, similar to a snow globe knick-knack.

For both worldlet types, hither and yon clipping planes restrict the extent of the worldlet, insuring that the worldlet contains a manageable subset of the larger world. Worldlet shape content is pre-shaded and pre-textured to match the corresponding shapes in the main world. Though the main world may have content that changes over time, the captured worldlet remains static, recording the content of the world at the time the worldlet was captured.

Figure 1 shows a virtual city containing buildings, monuments, streets, stop lights, and so forth. Figure 1a shows the world from a viewpoint aimed at a landmark. Figure 1b shows a bird's eye view highlighting the portion of the world falling within the viewing frustum anchored at the viewpoint in Figure 1a. Figures 1c through 1f show several views of the same frustum worldlet captured from this viewpoint.

Figure 2a provides a bird's-eye view of the same virtual city, highlighting a spherical portion of the world falling within a viewing sphere anchored at a viewpoint. Figure 2b shows a spherical worldlet captured at the viewpoint.

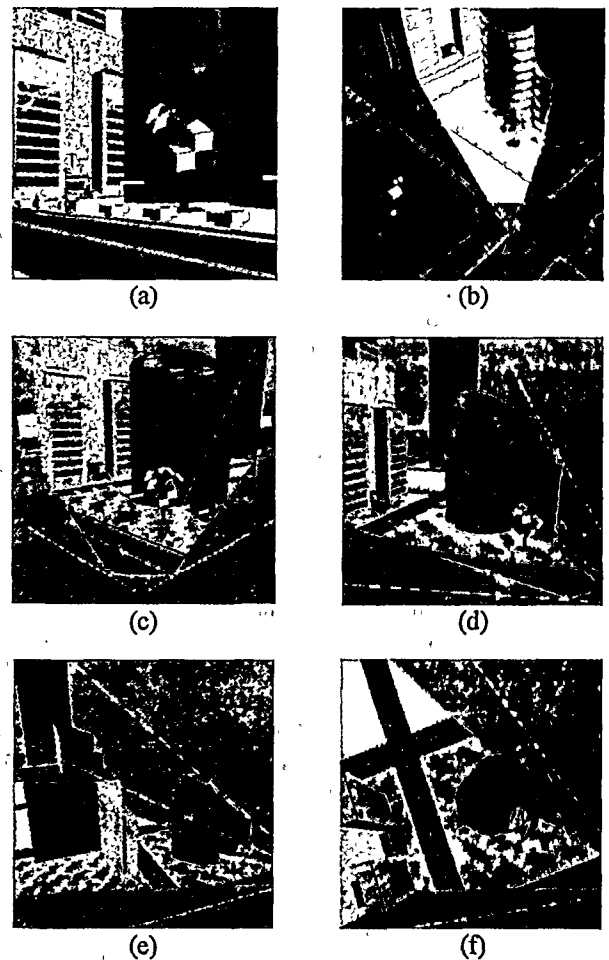


Figure 1: A virtual city landmark (a) viewed from a vantage point, (b) showing the viewing frustum from above, and (c-f) captured within a frustum worldlet.

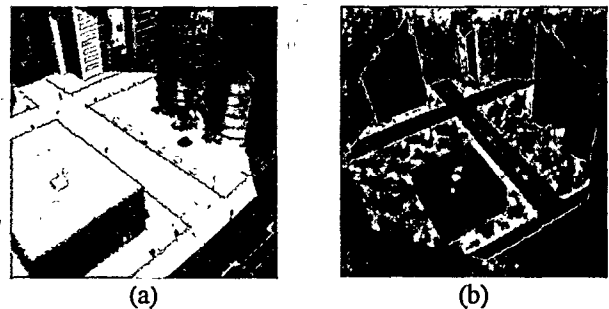


Figure 2: A virtual city landmark (a) showing a viewing bubble from above, and (b) captured within a spherical worldlet.

Using our landmark representation legibility criteria, a worldlet provides good imaginability, landmark context, traveler context, and support for multiple vantage points. The 3D content of a worldlet preserves a landmark's 3D shape, size, and texture, facilitating a traveler's use of landmark knowledge to recognize a destination of interest. The frustum or spherical capture area of a worldlet insures that landmark context is included along with a landmark.

To support a notion of traveler context, a worldlet is typically captured from a traveler-defined vantage point, such as street level within a virtual city. The traveler-defined vantage point insures that the landmark representation expresses what the traveler saw, while the 3D nature of the worldlet enables the traveler to interactively explore multiple additional vantage points.

WORLDLETS IN THE USER INTERFACE

We have incorporated worldlets into the user interface for a VRML browser. The browser provides features to select among world viewpoints and among previously visited worlds on the browser's history list.

Selecting Viewpoints

Traditional VRML browsers provide a viewpoint menu offering a choice of viewpoints, each denoted by a brief textual description. We have extended this standard feature to provide three experimental viewpoint selection interfaces, each using worldlets. All three present a set of worldlets, one for each author-selected viewpoint in the world. The browser also supports on-the-fly capture of worldlets using the traveler's current viewpoint.

- The *viewpoint list window* provides a list of worldlets beside a worldlet viewer. Selection of a worldlet from the list displays the worldlet in the viewer where it may be interactively panned, zoomed, and rotated. A "Go to" button flies the main window's viewpoint to that associated with the currently selected worldlet.
- The *viewpoint guidebook window* presents a grid of worldlet viewers, arranged to form a guidebook photo-album page. Buttons on the window advance the guidebook forward or back a page at a time. Selection of any worldlet on the page enables it to be interactively examined. A "Go to" button flies the main window's viewpoint to that of the currently selected worldlet. Figure 3 shows the viewpoint guidebook window.

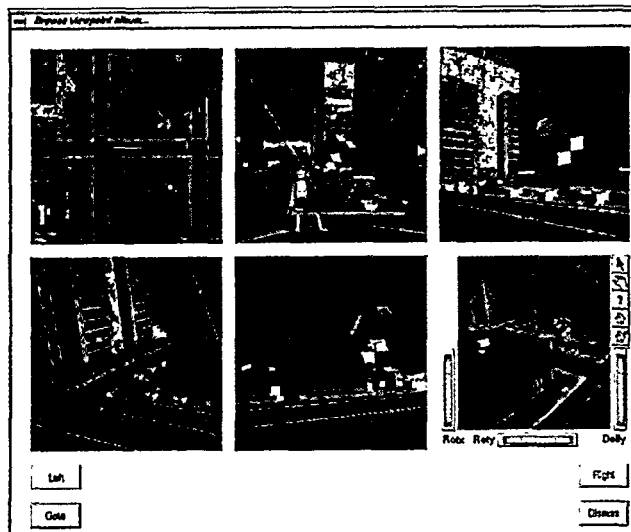


Figure 3: The viewpoint guidebook window.

- The *viewpoint overlay window* enables the traveler to select a worldlet from a list, and overlay it atop the main window, highlighted in green. This worldlet overlay provides a clear indication of the worldlet's viewpoint position and orientation, along with the portion of the world captured within that worldlet. Figures 1b and 2a, shown earlier, were each generated using this overlay technique.

Selecting Worlds

Traditional VRML browsers provide a history list of recently visited worlds, each denoted by its title or URL. We have extended this standard feature to provide two world selection interfaces, each using worldlets.

- The *world list window* provides a list of worldlets beside an interactive worldlet viewer, similar to the viewpoint list window discussed earlier. One worldlet is available for each world on the browser's history list. A "Go to" button loads into the main window the world associated with the currently selected worldlet.
- The *world guidebook window* uses the same guidebook photo-album layout used for the viewpoint guidebook window discussed earlier. One worldlet is available for each world on the history list. A "Go to" button loads the world associated with the currently selected worldlet. Figure 4 shows the world guidebook window.

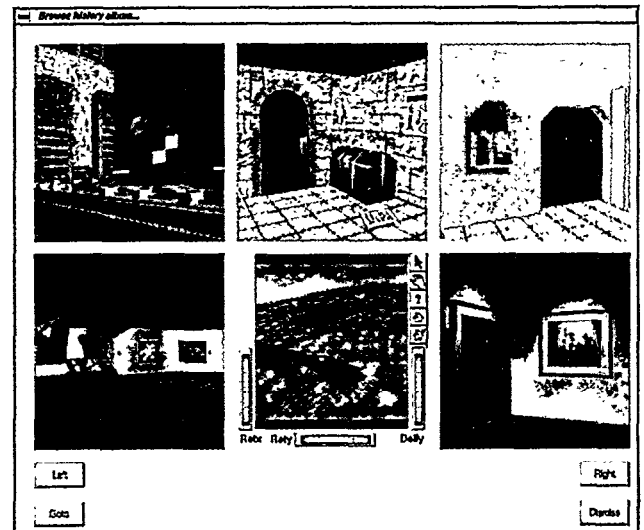


Figure 4: The world guidebook window.

Creating Worlds of Worldlets

A "Save as" feature of the VRML browser enables the traveler to save a worldlet to a VRML file. Using a collection of saved worldlets, a world author can create a VRML world of worldlets. Such a world acts like a 3D destination index, similar to a shelf full of snow globe knick-knacks depicting favorite tourist attractions. When cast as a VRML anchor shape, a worldlet provides a 3D "button" that, when clicked upon, loads the associated world into the traveler's browser

Figure 5 shows such a world of clickable worldlets. Figure 5a shows a close-up view of a world “doorway” and a niche containing a worldlet illustrating a vantage point in that world. Figure 5b shows a wider view of the same world and multiple such doorways.

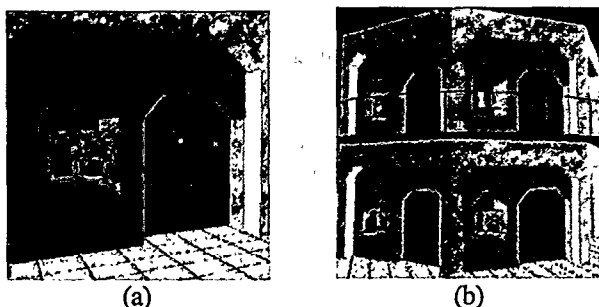


Figure 5: A world of worldlets that (a) associates a worldlet with each doorway (b) in an environment containing multiple such doorways. Each doorway leads to a different world.

Summary

The viewpoint selection windows enable a traveler to browse a world’s viewpoint set using worldlets. Each worldlet represents a 3D landmark and its context, facilitating the traveler’s recognition of a desired destination. The use of viewpoint animation to fly between selected viewpoints helps the traveler understand landmark spatial relationships and build up procedural knowledge for routes between the landmarks.

World guidebook windows and worlds of worldlets both enable a traveler to examine landmark worldlets in a set of available worlds. Worldlets provide visual cues that help a traveler recognize a destination of interest.

In contrast to WIMs, the browser’s viewpoint and world selection features display miniature worlds outside of the main world. No reserved space is required in the virtual environment between the traveler and collidable 3D content. No stylistic clash or confusion with unbounded environments occurs. The separate display of worldlets and the main world avoids impacting rendering performance. The use of separate worldlet display windows also enables the simultaneous display of multiple worldlets, including those for worlds different from that currently being viewed in the main viewer window.

An effect similar to WIMs can be created by including a worldlet within a world, like that shown in Figure 5. A worldlet can remain stationary in the world or move along with the traveler, as in a WIM. In this regard, WIMs are a special-purpose implementation of the more general worldlet concept.

IMPLEMENTATION

The VRML browser used in this work maintains virtual environment geometry within a tree-like *scene graph*. Worldlets are also stored as scene graphs, together with

additional state information. To capture a worldlet or display a worldlet or virtual environment the VRML browser traverses the associated scene graph and feeds a 3D graphics pipeline.

Worldlet Capture in General

Any 3D graphics pipeline can be roughly divided into two stages: (1) transform, clip, and cull, and (2) rasterize [8]. The first stage applies modeling, viewing, perspective, and viewport transforms to map 3D shapes to the 2D viewport. Along the way, shapes outside of the viewing frustum are clipped away and backfaces removed. The second stage uses 2D shapes output by the first stage and draws the associated points, lines, and polygons on the screen.

Worldlet capture taps into this 3D graphics pipeline, extracting the transformed, shaded, clipped, and culled shape coordinates output by the first stage prior to rasterization in the second stage. An extracted coordinate contains X and Y screen-space components, a depth-buffer Z-space component, and the W coordinate. Each extracted coordinate has an associated RGB color and texture coordinates, computed by shading and texture calculation phases in the first pipeline stage.

To create a worldlet, these extracted coordinates are *untransformed* to map them back to world space from viewport space. The inverses of the viewport, perspective, viewing, and modeling transforms are each applied. Coordinate RGB colors and texture coordinates are used to reconstruct 3D worldlet geometry in a worldlet scene graph.

Display of a worldlet passes this 3D geometry back down the graphics pipeline, transforming, clipping, culling, and rasterizing the worldlet like any other 3D content.

Frustum and Spherical Worldlets

A frustum worldlet is the result of capturing 3D graphics pipeline output for a single traversal of the scene graph as viewed from the traveler’s current viewpoint. The shape set extracted after the first pipeline stage contains only those points, lines, and polygons that fall within the viewing frustum. The worldlet constructed by the browser from this geometry looks like a pie-shaped slice cut out of the world.

A spherical worldlet is the result of performing multiple frustum captures and combining the results. The VRML browser captures a spherical worldlet by sweeping out several stacked cylinders around a viewpoint position, generating a set of frustum worldlets each using a different viewing orientation. Additional captures aimed straight up, and straight down complete the spherical worldlet. The resulting set of capture geometry constructs a 360 degree spherical view from the current viewpoint.

When displayed, the spherical worldlet’s geometry looks like a bubble cut out of the virtual environment. A close yon clip plane keeps the bubble small, insuring that it captures only landmark features in the immediate neighborhood, and not the entire virtual world.

Worldlet Capture in OpenGL

To take advantage of the rendering speed offered by the accelerated 3D graphics pipeline on high-speed workstations, we implemented worldlet display and capture using OpenInventor and OpenGL graphics libraries from Silicon Graphics. Scene graph construction and display traversal is managed by OpenInventor. To capture worldlet geometry, the VRML browser places the pipeline into *feedback* mode prior to a capture traversal, and returns it to *rendering* mode following traversal.

While in feedback mode, the OpenGL pipeline diverts all transformed, clipped, and culled coordinates into a buffer provided by the browser. Upon completion of a capture traversal, no rasterization has taken place and the feedback buffer contains the extracted geometry. By parsing the feedback buffer, the VRML browser reconstructs worldlet geometry, applying appropriate inverse transforms.

OpenGL feedback buffer information includes shape coordinates, colors, and texture coordinates, but does not include an indication of which texture image to use for which bit of geometry. To capture this additional information, the VRML browser uses OpenGL's *pass through* features to pass custom flags down through the pipeline during traversal. To prepare these pass through flags, the browser augments the world scene graph prior to traversal, assigning a unique identifier to each texture image. During a capture traversal, each time a texture image is encountered, the associated identifier is passed down through the pipeline and into the feedback buffer along with shape coordinates, colors, and texture coordinates. During parsing of the feedback buffer, these texture identifiers enable worldlet geometry reconstruction to apply the correct texture images to the correct shapes.

PILOT STUDY

A pilot study was conducted to evaluate landmark representation effectiveness within a wayfinding task. Subjects in the study were asked to use an on-line landmark guidebook and follow a sequence of landmarks leading from a starting point to a goal landmark. Guidebook entries providing landmark descriptions were offered in three ways: in textual form, as 2D images, and as 3D worldlets.

The pilot study used five subjects, three female and two male. All subjects were computer-literate, but had varying degrees of experience with virtual environments. Subject occupations were student, programmer, ecologist, molecular biologist, and computer animator.

Virtual Environment Design

Six different virtual city environments were created for the study. Each city was composed of a street grid, five blocks by five blocks, with pavement roads and sidewalks between the blocks. Each block contained 20 buildings, side-by-side around the block perimeter. Using a cache of 100 building designs, buildings were randomly selected and placed on city blocks. Buildings were colored using texture images derived from photographs of buildings in the San Francisco

area. Typical building photographs were of two-story houses, office buildings, shops, and warehouses.

Three of the six cities were used for training subjects, and the remaining three used for the timed portion of the experiment. The timed experiment required that subjects make their way from a starting point to a goal. Timed experiment cities, therefore, contained a starting point, an ending goal, and three intermediate landmarks. The distance between any adjacent pair of these varied between one and two blocks. The total distance from the starting point to the ending goal was six blocks. The intermediate landmarks included two buildings and one non-building (mailbox, fire hydrant, or newspaper stand). The ending goal was a distinctive six-sided kiosk marked "GOAL". The starting point was unmarked.

Training cities were structurally equivalent to cities used in the timed experiment. However, subjects were given a starting point, only a single intermediate landmark, and the goal kiosk. The landmark in each training city differed from landmarks used in the timed cities.

Software Design

The VRML browser user interface was modified for the study. A main city window displayed the city. Keyboard arrow key presses moved the subject forward and back by a fixed distance, or turned the subject left or right by a fixed angle. Subjects were instructed to press a "Start" button to begin the experiment and press a "Stop" button when they reached the goal. Between the two button presses, data describing the subject's position and actions was automatically collected at one second intervals.

A "Guidebook" button on the main window displayed a full-screen guidebook photo-album window with textual, image, or worldlet landmark descriptions. A "Dismiss" button on the guidebook window removed the window and again revealed the main city window. The subject could not see the main city window without dismissing the guidebook.

The study used a within-subject randomized design. Each subject visited three virtual cities in a random order. For each subject, one city provided a guidebook with textual landmark descriptions leading to the goal, one provided image landmark descriptions, and one provided worldlet landmark descriptions. In cities using textual and image landmark descriptions, the guidebook contained static textual and image information. In the city using worldlet landmark descriptions, the guidebook contained interactive worldlets, each of which could be explored using the same arrow key bindings as the main city window.

For each landmark, the landmark and a fifteen meter radius around the landmark, were expressed in the description. Textual descriptions described both the landmark and the immediate surroundings. Image landmark descriptions showed portions of the neighboring buildings. Worldlet descriptions included a spherical bubble with a fifteen meter radius centered in front of the landmark.

Procedure

Prior to beginning the experiment, instructions were read to each subject and an image shown of the goal kiosk. Each subject was shown the user interface and taught use of the arrow keys, both for city movement and worldlet movement. Subjects were allowed to spend as much time as they needed practicing in three training cities, each with guidebook landmark descriptions in either text, image, or worldlet form. When subjects felt comfortable with each interface, the timed portion of the experiment was begun.

During the timed portion, subjects were asked to navigate from the starting point to the goal kiosk as quickly as possible.

Results

The independent variable in the study was the type of landmark description used: text, image, or worldlet. Dependent variables include the time spent consulting the guidebook, the time spent standing still within the city, the time spent moving forward over new territory, the time spent backtracking over territory previously traversed, the distance traversed moving forward, and the distance traversed while backtracking. Table 1 includes the mean values for subject data collected for each of the dependent variables. Travel time is measured in wall-clock seconds while travel distance is measured in meters within the virtual environment. Mean overall travel times and distances are also listed in the table.

Table 1: Mean times and distances traveled.

Mean Times (seconds)	Text	Image	Worldlet
Consulting guidebook	47.6	44.6	91.0
Standing still	179.2	141.6	58.6
Moving forward	155.0	156.4	91.0
Backtracking	86.2	78.0	0.4
Overall	468.0	420.6	241.0
Mean Distances (meters)			
Moving forward	684.6	739.0	421.6
Backtracking	409.0	371.4	2.2
Overall	1093.6	1110.6	424.0

In the table above, *Consulting guidebook* values indicate the time subjects spent with the guidebook window on-screen. City movement could not occur while the guidebook window was displayed.

Standing still values indicate the time subjects spent standing at a single location, looking ahead or turning left and right.

Landmarks in all three cities were arranged so that at no time would a subject be required to traverse the same block twice to reach the goal. *Moving forward* times and distances record movement through previously untraversed territory. *Backtracking* times and distances measure unnecessary travel over previously traversed territory.

In a post-study questionnaire subjects were asked to rank each landmark representation technique according to how easy it was to use. Table 2 summarizes subject rankings for the five subjects in the pilot study.

Table 2: Rankings of landmark representations.

	Text	Image	Worldlet
Very easy	0	0	3
Easy	2	2	2
Doable	1	1	0
Difficult	2	1	0
Very difficult	0	1	0
Median	Doable	Doable	Very easy

Analysis

A one-way analysis of variance (ANOVA) was performed for each of the dependent variables and the overall times and distances. The within-subjects variable was the landmark description type with three levels: text, image, and worldlet. Post-hoc analyses were done using the Tukey Honest Significant Difference (HSD) test. We adopted a significance level of .05 unless otherwise noted. Table 3 summarizes these results.

Table 3: F-test values for F(2,8) and $p < .05$.

Mean Times	F(2,8)
Consulting guidebook	5.78
Standing still	5.80
Moving forward	8.20
Backtracking	5.40
Overall	5.46
Mean Distances	
Moving forward	7.09
Backtracking	5.82
Overall	6.79

Post-hoc analyses of each of the dependent variables revealed:

- Time spent consulting guidebook: text and image times were not significantly different, but image times were significantly less than for worldlets.
- Time spent standing still: text and image times were not significantly different, but text times were significantly greater than for worldlets. Image and worldlet times were not significantly different.
- Time spent moving forward: text and image times were not significantly different, but both were significantly greater than for worldlets.
- Time spent backtracking: text and image times were not significantly different, but both were significantly greater than for worldlets.
- Overall time: text and image times were not significantly different, but text times were significantly

greater than for worldlets. The difference between image and worldlet times approached significance ($p = .08$) with image times greater than those for worldlets.

- Moving forward distance: text and image movement distances were not significantly different, but both were significantly greater than for worldlets.
- Backtracking distance: text and image backtracking distances were not significantly different, but both were significantly greater than for worldlets.
- Overall distance: text and image movement distances were not significantly different, but both were significantly greater than for worldlets.

Discussion

Figure 6 plots mean times for each type of landmark description for the time used consulting the guidebook, standing still, moving forward over new territory, and backtracking over previously traversed territory.

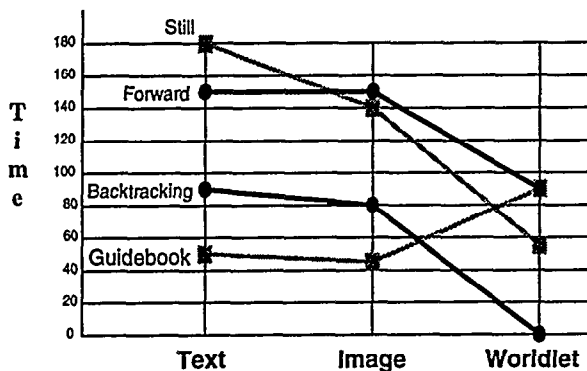


Figure 6: Mean times.

Subjects spent more time on average consulting worldlet descriptions than consulting either text or image descriptions. This extra consultation time was more than compensated for by reductions in time spent standing still, moving forward, and most dramatically in time spent backtracking.

A natural conjecture is that subjects spent the additional time with worldlets creating a more comprehensive cognitive model of the landmark region which enabled them to spend less time searching for landmarks or landmark context. This is reflected in the reduced total travel times. The striking reduction in backtracking time, bringing it virtually to zero, indicates that worldlets enabled subjects to do less wandering and to move more directly to the next landmark.

Figure 7 plots mean travel distances for each type of landmark description. As with travel time, forward and backtracking travel distances also were reduced when using worldlets.

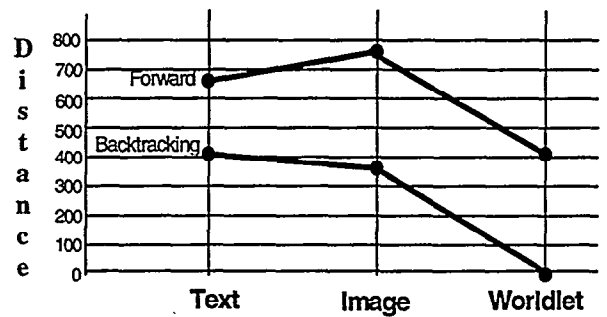


Figure 7: Mean distances.

CONCLUSIONS

Wayfinding literature provides clear support for the importance of landmarks in navigating an environment, whether real or virtual. Landmarks anchor routes through an environment and provide memorable destinations to return to later. Landmarks help to structure an environment and supply directional cues used to find destinations of interest.

Whereas a traveler's landmark knowledge characterizes a destination by its 3D shape, size, texture, and so forth, the menus of today's virtual environment browsers characterize destinations by textual descriptions or thumbnail images. This representation mismatch reduces the effectiveness of landmark descriptions in destination menus. Unable to use their memory of 3D landmarks to choose among menu items, travelers may resort to a naive, exhaustive search to find a desired landmark.

In a wayfinding task, textual or image guidebook landmark descriptions fail to engage the full range of 3D landmark characteristics recognized and used by travelers to find their way. Unable to extract sufficient landmark knowledge from textual or image descriptions, travelers move through an environment with less comprehensive cognitive models, spending more time standing still and looking around, moving in incorrect directions, and backtracking over previously traversed territory.

This paper has introduced a new user interface affordance to increase wayfinding efficiency. This affordance, called a *worldlet*, captures a 3D thumbnail of a virtual environment landmark. Each worldlet is a miniature virtual world fragment that may be interactively viewed in 3D. By encapsulating a 3D description of a landmark, worldlets provide better landmark imaginability, landmark context, traveler context, and multiple vantage point support than text or image representations. Displayed within a browsable landmark guidebook, worldlets facilitate virtual environment wayfinding by enhancing a traveler's ability to recognize and travel to destinations of interest. When used to provide guidebook descriptions in a wayfinding task, worldlets significantly reduced the overall travel time and distance traversed, virtually eliminating backtracking.

FUTURE WORK

Development of worldlets and the VRML browser revealed issues requiring further study:

- To insure that spherical worldlets capture only the traveler's immediate vicinity, the yon clip plane is automatically placed relatively close to the traveler's viewpoint. The current approach sets the yon clip plane distance to a fixed value. However, this distance should vary with traveler avatar characteristics, the environment being viewed, or the landmark capture intended. A general-purpose, automatic yon clip plane selection algorithm is needed.
- VRML provides features that describe world characteristics that do not reduce to points, lines, or triangles, and thus do not show up in a captured worldlet. These features include background color, sounds, behaviors, and shape collidability. Worldlets constructed without capture of these features may not look and act like the main world from which they were captured. A mechanism to capture this additional information is needed.

In addition to these issues, future work will include a more extensive user study. The pilot study's finding that backtracking was practically eliminated was unexpected and deserves further attention.

ACKNOWLEDGEMENTS

The San Diego Supercomputer Center (SDSC) is funded by the National Science Foundation (under grant ASC8902825), industrial partners, and the State of California. This work was also partially funded by the San Diego Bay Interagency Water Quality Panel. Suzanne Feeney of the University of California, San Diego (UCSD) Psychology Department and Rina Schul of the UCSD Cognitive Science Department were instrumental in developing the pilot study. Special thanks to John Moreland for assistance in developing the software, and to Mike Bailey, Andrew Glassner, Allan Snavely, and Len Wanger for their input on the project. Thanks also to John Helly and Reagan Moore for their support.

REFERENCES

1. Allen, G.L., Kirasic, K.C. Effects of the Cognitive Organization of Route Knowledge on Judgments of Macrospace Distances. In *Memory & Cognition*, 1985, 3, pp. 218-227.
2. Appleyard, D.A. Why buildings are known. In *Environment and Behavior*, 1969, 1, pp. 131-156.
3. Bell, G.; Carey, R.; Marrin, C. The Virtual Reality Modeling Language, version 2.0, 1996. At <http://vrm1.vag.org/VRML2.0/FINAL>
4. Chen, S. E. QuickTime VR - An Image-based Approach to Virtual Environment Navigation. In

Proceedings of the ACM SIGGRAPH 95 Conference, August 1995, Los Angeles, CA. pp. 29-38.

5. Darken, R. P., and Sibert, J. L. Wayfinding Strategies and Behaviors in Large Virtual Worlds. In *Proceedings of the ACM CHI 96 Conference*, April 1996, Vancouver, BC., pp. 142-149.
6. Downs, R. J., and Stea, D. Cognitive Maps and Spatial Behavior. In *Image and Environment*, Chicago: Aldine Publishing Company, 1973, pp. 8-26.
7. Evans, G. Environmental cognition. In *Psychology Bulletin*, 1980, 88, pp. 259-287.
8. Foley, J., van Dam, A., Feiner, S., and Hughes, J. *Computer Graphics Principles and Practice*, Addison-Wesley, 1990.
9. Goldin, S.E., Thorndyke, P.W. Simulating Navigation for Spatial Knowledge Acquisition. In *Human Factors*, 1982, 24(4), pp. 457-471.
10. Lynch, K. *The Image of the City*, M.I.T. Press, 1960.
11. Passini, R. *Wayfinding in Architecture*, Van Nostrand Reinhold, NY, second edition, 1992.
12. Pausch, R., Burnette, T., Brockway, D., Weiblen, M.E. Navigation and Locomotion in Virtual Worlds via Flight into Hand-Held Miniatures. In *Proceedings of SIGGRAPH 95*, 1995, pp. 399-400.
13. Peponis, J., Zimring, C., and Choi, Y.K. Finding the Building in Wayfinding. In *Environment and Behavior*, 1990, 22 (5), pp. 555-590.
14. Satalich, G. A. *Navigation and Wayfinding in Virtual Reality: Finding the Proper Tools and Cues to Enhance Navigational Awareness*. Masters Thesis, Department of Computer Science, University of Washington, 1995.
15. Stoakley, R., Conway, M. J., and Pausch, R. Virtual Reality on a WIM: Interactive Worlds in Miniature. In *Proceedings of the ACM CHI 95 Conference*, pp. 265-272.