Alternative "Vision": A Haptic and Auditory Assistive Device

Dan Morris and Neel Joshi
Computer Science Department
Stanford University
353 Serra Mall, Stanford, CA 94305
{dmorris,nsj}@cs.stanford.edu

ABSTRACT

We have used two cameras and a SensAble Technologies "Phantom" force-feedback haptic display to haptically render a three-dimensional surface that represents key aspects of a visual scene. In addition to rendering depth and contour information with the Phantom, we capture optic flow and present this to the user using sound cues. We propose that with further development, this system could be used as an assistive device that would allow visually impaired individuals to explore the "visual" world.

Keywords

Haptics, assistive devices, real-time vision

INTRODUCTION

The current trend toward miniaturization of powerful computers will allow algorithms in computer vision and haptics – fields that have traditionally been constrained to high-end desktops – to run on laptops and handhelds. Consequently, advances in these areas will soon be able to contribute to a traditionally "low-tech" field: assistive devices for the blind. The goal of this project is to develop a system that extracts information from a moving camera and presents important or especially salient visual features to the user's non-visual senses.

We have used two cameras and a SensAble Technologies "Phantom" force-feedback haptic display [3] to visually and haptically render a three-dimensional surface that represents key aspects of a visual scene. In particular, our system can be used to "feel" a depth map of the visual scene or to "feel" the contours defined by edges in the visual scene. The system runs in real-time; the user presses a button to trigger the capture and rendering of either a depth map or an edge-detected image. Once an image has been rendered, the user can use the Phantom to explore the entire surface.

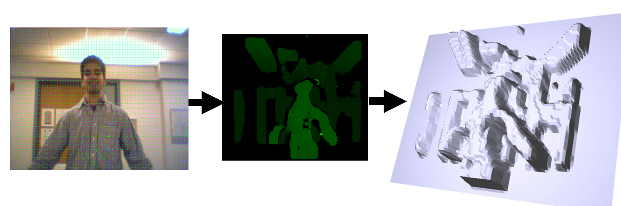
In addition to rendering depth and contour information with the Phantom, we capture optic flow in real-time and present this to the user using sound cues: a tone pans from left to right according to the horizontal location of maximum optic flow and scales in volume according to the magnitude of optic flow.

CHI 2003, April 5–10, 2003, Ft. Lauderdale, Florida, USA.
ACM 1-58113-630-7/03/0004

METHODS

Haptic Representation of Depth

Depth information is critical both in avoiding potential obstacles and in object recognition. We use a stereo camera to create a depth map of a visual scene based on disparity between the images captured at the two camera heads. The depth/disparity map is then converted to a height map, which is displayed here as an OpenGL rendering. The model displayed here contains precisely the same polygons as the model that is displayed haptically using the Phantom.



Haptic Representation of Contours

The human visual system is particularly good at recognizing edges in images; in many cases, the nature of key objects in a scene is apparent based only on the location of strong edges in the scene. Edge-detection thus represents a significant reduction in the complexity of a scene, while maintaining a disproportionate amount of information, making it ideal for haptic rendering.

As is shown below, when the user triggers an image capture and requests a counter rendering, we process the image using a low-pass filter and a Canny edge detector. We then convert the resulting gradient map to a height map, which is displayed here as an OpenGL rendering. The model displayed here contains precisely the same polygons as the model that is displayed haptically using the Phantom.

Representing Optic Flow Using Sound

Optic flow (the relative motion of pixels at different locations in the visual field) has been used successfully to guide robots along hallways [1], and it has been shown that humans use optic flow as an important navigational cue [5]. Our system represents optic flow using sound: the left-right pan of a sound indicates which side of the visual scene has a larger net flow, and the volume of the sound is used to represent the magnitude of flow.

It may be possible for a user to balance optic flow by balancing the pan of the tone, providing a navigational aid to the visually-impaired user. High-volume sound can serve as a warning that a moving object is close to the user (e.g. a person is walking in front of the camera). Furthermore, the presence of optic flow suggests to the user that motion has occurred, possibly necessitating a re-rendering of the scene for further exploration.

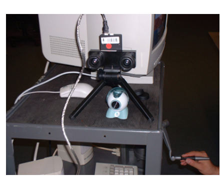
System Description

Our system currently runs on a PC under Windows 2000; our test machine is a dual-processor 850MHz desktop with 512MB of RAM. Images are acquired for optic flow and edge-detection via a low-cost USB webcam and processed using Intel's OpenCV [2] computer vision library.

Images are acquired for depth-mapping via a Videre Design stereo camera. Disparity maps are generated using SRI International's Small Vision System [4]. Haptic rendering is performed using the SensAble Ghost API and the SensAble Phantom force-feedback haptic display. We have developed software in C++ to capture images from both camera systems and convert the results into a three-dimensional triangle mesh suitable for rendering using OpenGL and the Ghost API. Our software also computes global optic flow information from the local flow vectors provided by OpenCV, and generates tones accordingly.

The program displays the real-time video stream from both cameras including disparity and optic flow information, and provides an OpenGL representation of the rendered scene, which is simultaneously displayed on the haptic device. The user can press a key at any time to capture either a depth-map or an edge map. The user can then explore the scene using the Phantom. The user can press a button on the Phantom to enable translation and rotation of the scene.





Future Work

Haptic devices are available or will soon be available that add a "grip" to the one-point force-feedback that is available using the Phantom. We feel that giving the user even a single force-feedback grip will provide a much more convincing representation of the world, since humans tend to explore a haptic environment with more than a single finger. Similarly, the Phantom's API allows for multiple Phantoms; we would be interested to experiment with two-handed exploration or with the use of a Phantom in one hand and a tactile display in the other.

The Phantom has an update rate of 1kHz, which is sufficient to convey texture with astounding accuracy. Our images currently have numerous flat surfaces that could be layered with texture; we expect that texture might be used to represent visual properties that are not available from shape alone (color would be a logical choice). Similarly, haptic devices are available that present tactile information to the user; such devices could be used to directly represent image texture or to represent other aspects of the captured image.

Most importantly, we would like to present the device to a visually-impaired subject and consider their suggestions very seriously. It is very difficult for us to conceive of the how information that is provided by the system could really be used without visual feedback, since our mechanisms for exploring a scene are fundamentally visual.

ACKNOWLEDGMENTS

We thank the Stanford haptics group for equipment and support.

REFERENCES

1. Coombs, D., Herman, M., and Nashman, M. Real-time Obstacle Avoidance Using Central Flow Divergence and Peripheral Flow, presented at the *International Conference on Computer Vision*, 1995.
2. Intel® Open Computer Vision Library. http://www.intel.com/research/mrl/research/opencv/
3. Salisbury, J. K. & Srinivasan, M. A. Phantom-based haptic interaction with virtual objects. IEEE Comput. Graph. Appl. 17, 6-10 (1997).
4. SRI International Small Vision System. http://www.ai.sri.com/software/SVS
5. Warren, W.H. and Hannon, D. H. Direction of self-motion is perceived from optical flow. *Nature 336*, 162-163 (1988).