Omnidirectional Vision Systems: 1998 PI Report *

Shree K. Nayar
Department of Computer Science
Columbia University
New York, N.Y. 10027
nayar@cs.columbia.edu

Terry Boult
Department of EE and CS
Lehigh University
Bethlehem, PA 18015
tboult@cortex.eecs.lehigh.edu

Abstract

Conventional video cameras have limited fields of view which make them restrictive in a variety of applications. Based on our previous work on omnidirectional imaging, we have developed a variety of new vision sensors, real-time algorithms for tracking and depth estimation and architectures for distributed surveillance. In this report, we briefly describe research projects that were completed during the 1997-1998 time frame.

1 Introduction

This report describes a variety of VSAM projects that are based on, or related to, omnidirectional imaging technology we have developed [Nayar, 1997], [Peri and Nayar, 1997], [Nayar and Baker, 1997], [Nayar and Boult, 1997], [Nene and Nayar, 1998], [Gluckman and Nayar, 1998a], [Baker and Nayar, 1998], [Boult, 1998a]. Broadly, our research projects can be divided into the following thrust areas:

- **Sensors:** We have made significant advances in the area of wide field of view sensing. Our new devices include: a very compact, self-contained, omnidirectional camera, a combined omnicamera and pan/tilt/zoom imaging system, a remote controlled car equipped with omnidirectional video and audio sensors onboard, a compact catadioptric stereo sensor, and a panoramic stereo sensor.
- **Algorithms:** We have developed new algorithms for correcting distortions in wide-angle lenses, fast generation of perspective video from omnidirectional video for remote reality applications, robust multi-body tracking, and real-time computation of both perspective and panoramic depth video.
- **Data Collection:** We have collected a large amount of omnidirectional video data from the Ft. Benning site. This footage has been used to develop and evaluate our multi-body tracking algorithms. The data has also been used to design a novel protocol for distributed target identification and surveillance.
- **IUE:** We have made significant contributions to DARPA's Image Understanding Environment (IUE). In particular, we have developed a new set of image processing classes, code generation processes and a system for quick creation of custom user-interfaces to the IUE.

We will now briefly summarize each of our research projects.

2 Small Omnidirectional Cameras

A compact version of our omnidirectional camera has been developed for autonomous navigation and remote visual exploration (see Figure 1). The camera was developed during Phase 1 of a "Tactical Mobile Robotics" project funded by DARPA. The complete system, including, a video camera, telecentric optics and curved mirror, is contained in a cylindrical enclosure that is 22 cm tall and 7 cm in diameter. A transparent (acrylic) tubing is used to attach the optical system to the mirror without obstructions. This omnicamera design is ideal for ap-

*This work was supported by the VSAM effort of the DARPA Image Understanding Program, under ONR contract No. N00014-97-1-0553. Some of the projects described in this report were also supported in parts by the DARPA/ONR MURI Grant N00014-95-1-0601, an NSF National Young Investigator Award, and a David and Lucile Packard Fellowship.

Figure 1: A compact omnidirectional camera that is 22 cm in height and 7 cm in width. This self-contained unit includes a video camera, telecentric optics and a curved mirror.

plications where the camera needs to be automatically protracted and retracted from a moving vehicle. We are presently pursuing further reduction in sensor size. In the near future, we hope to demonstrate omnidirectional cameras that are less than 12 cms tall and 5 cms wide.

3 Remote-Controlled OmniRover

A remote-controlled car has been mounted with an omnidirectional camera, a microphone, and speakers (see Figure 2). Using wireless communication, this compact vehicle enables a user to visually and acoustically interact with a remote location. The user can drive the car with a joystick, using the omnidirectional video produced by the on-board camera as visual feedback. The omnidirectional data is mapped to panoramic video for ease of visual interaction. A head-mounted display and the remote reality system developed at Lehigh University (see next section) [Boult, 1998b] can be used to "look around" the remote location. The on-board microphone and speakers are used to communicate with people located in the remote site. Among other applications, this system can be deployed for visual monitoring and exploration of a hazardous environment.

Figure 2: OmniRover: A remote-controlled car with an omnidirectional video camera, a microphone, a wireless video/audio transmitter, an FM receiver and speakers. A remote user can use the audio and video transmitted by the car to drive the vehicle, and communicate with people in a remote site.

4 Remote Reality System

Remote reality [Boult, 1998b] provides an immersive environment by combining an omnidirectional camera with a head-mounted display and head-tracker to produce perspectively correct images for the direction the user is facing (see Figure 3). We have gone through a series of designs for video acquisition hardware and image generation software. The system now provides 320x240 video at 30fps on a 233MHz PC. This system has been demonstrated to a large number of people with very positive feedback. We have developed a prototype for use with the radio-controlled vehicle described in the previous section. This project has been supported in part by our MURI project on autonomous vision systems.

5 Omnidirectional Pan/Tilt/Zoom System

We have implemented and demonstrated the combined wide-angle and narrow-angle imaging system shown in Figure 4. This system uses an omnidirectional camera to capture a hemispherical view of the scene and a conventional camera mounted on a pan/tilt/zoom (PTZ) unit to produce high-resolution narrow-angle views of parts of the scene. The omnidirectional video makes it possible for the user to always have a global view of the scene. Regions of

Figure 3: The remote reality system provides an immersive environment by combining an omnidirectional camera with a head-mounted display and head-tracker to produce perspectively correct images as the user looks around.

Figure 4: A combined wide-angle and narrow-angle imaging system. The omnidirectional camera provides a global view of the area of interest. Local regions are either automatically (via tracking) or manually (by user) selected from the global view and used as inputs to drive a pan/tilt/zoom system that produces high-resolution local views.

activity are either automatically tracked or manually selected and provided as inputs to the fast PTZ unit which provides detailed local views. The high optical resolution of a local view permits recognition of small objects and subtle activities.

6 Frame-Rate Multi-Body Tracking

In video surveillance applications, the ability to automatically track multiple moving objects is distinctly different parts of the field-of-view is highly desirable. We have made enhancements to our frame-rate multi-body omnidirectional tracking system, see [Boult *et al.*, 1998a] [Boult *et al.*, 1998b]. The system is designed for tracking in omnidirectional video, though it could be used with any video source. The system is based on background subtraction with variance testing. To better handle trees blowing in the breeze, the system learns two different "backgrounds" per pixel. Like most tracking systems, it adapts to the background by updating a reference image(s). However, our system adapts quite slowly; it takes a minute or more for targets to become parts of the background. This is important so that the algorithm can better handle targets approaching the camera as well as snipers or other slowly moving targets (see Figure 6). The tracking algorithm has a separate module to handle lighting variations. The system tracks up to 64 targets at 30fps at full 640x480 resolution on a 233MMX processor.

For high-contrast scenes the system can track targets as small as 10-12 pixels, which in our omnidirectional video corresponds to a human-size target at 50m. For targets wearing camouflage in the woods, it performs well for targets as small as 15-20 pixels, which corresponds to a person at 35m. Figure 5 shows an example where soldiers in a town square are tracked. We are now working on better techniques for maintaining target identity despite occlusions and for coordination of tracking between multiple omnicameras.

In addition to the vision aspects of the multi-body tracking project, we have been developing a new protocol for a distributed target tracking and surveillance system. One of the design constraints is the ability to scale to a large numbers of sensors, each with a large number of targets, without saturating the network. In the spring of 98, we collaborated with CMU on the design of the current VSAM protocol which incorporated many of our ideas. This protocol will serve as the basis for the main IFD VSAM demonstration as well as our distributed intelligent video scheduling demonstration.

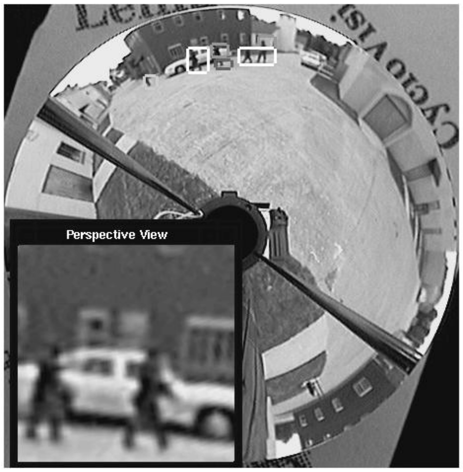

Figure 5: A robust multi-object tracking algorithm tracks soldiers moving in the woods at the Ft. Benning site.

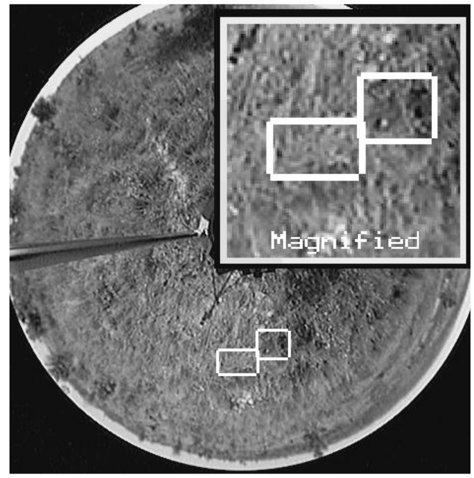

Figure 6: Tracking a sniper in the grass. The target is 12 feet away, though initial tracking began at 40 feet. Though it is hard to see the sniper, he *is* in the white boxes (and extends a bit outside them). Some of the image resolution is lost due to printing.

7 Ft. Benning Data Collection

As part of our VSAM project, we made three trips to Ft. Benning to collect omnidirectional data. This data will be used throughout the 1998-1999 timeframe to develop, tune and evaluate our omnidirectional tracking algorithms. Approximately 70 hours of omnidirectional video was collected. Seven cameras were used, with many data sets having multiple overlapping cameras. Four different camera signals were evaluated: progressive scanning NTSC; color S-video on NTSC hi-8; BW and color composite on NTSC hi-8; and S-Video on PAL hi-8. The data sets collected include scenarios such as individuals moving in the woods, small groups in camouflage moving in the woods, a sniper in a gilly suit, small groups of soldiers moving in and out of buildings and a large number of soldiers moving in the woods and in a town setting. The data includes both significant amounts of “targets” and empty scenes for false-alarm evaluation. Atmospheric conditions include light rain, partly sunny and windy to sunny with light breeze. Limited copies of the data are available upon request to tboult@eecs.lehigh.edu.

8 Distortion Correction for Wide-Angle Lenses

Wide-angle cameras are commonly used in video surveillance applications. Wide-angle lenses resort to non-perspective image projection models in order to map a large field-of-view onto a small planar image detector (CCD, for instance). The resulting images suffer from severe distortions that tend to increase with the field-of-view (see Figure 7). Images produced by wide-angle systems are used for two purposes: (a) monitoring by humans; and (b) visual processing by a machine. In both cases, it would be beneficial to eliminate the undesirable effects produced by distortion.

We have developed an efficient and accurate scheme for computing the distortion parameters of wide angle lenses [Swaminathan and Nayar, 1998] (in these proceedings). Our approach uses simple geometric scene primitives (such as straight lines) that have *unknown* locations and orientations in the scene. A wide-angle lens will, of course, map these straight lines to curves. We know that a perspective camera maps straight lines in the scene to straight lines in the image. This simple observation is used to compute the distortion parameters which would map our image curves (which correspond to scene lines) to perfect straight lines.

Our approach is similar in its spirit to that of Brown [Brown, 1971]. However, Brown’s work and subsequent ones suffer from one or both of the following limitations. First, none of the existing schemes have demonstrated resilience to noise, which is signifi-

Figure 7: A non-metric algorithm has been developed for computing radial, decentering, and translational distortion parameters. A severely distorted image captured using a typical wide-angle lens (top). The (perspective) image after distortion correction (bottom). The distortion parameters can also be used to generate other projections such as panoramic or spherical.

cant when using low-resolution video cameras and approximate scene lines (edges between walls, for instance). Second, almost all of the previous work is restricted to recovering only a few of the complete set of parameters at work. We have designed an objective function for estimating all the distortion parameters. This function explicitly takes into account noise [Swaminathan and Nayar, 1998]. Optimization of this objective function enables us to extract distortion parameters with high accuracy even when the (random) errors in selected image points exceeds 5 pixels. As reported in [Swaminathan and Nayar, 1998], the parameters recovered by the algorithm have been used to correct severe distortions in images and video (see Figure 7).

9 Real-Time Planar Catadioptric Stereo

By using two or more mirrored surfaces, multiple views of a scene can be captured by a single camera (catadioptric stereo). In [Nene and Nayar, 1998], we have described several different stereo configurations using planar, parabolic, elliptical, and hyperbolic mirrors, with a single camera. We have analyzed the epipolar constraints and the fields of view of each of these stereo systems.

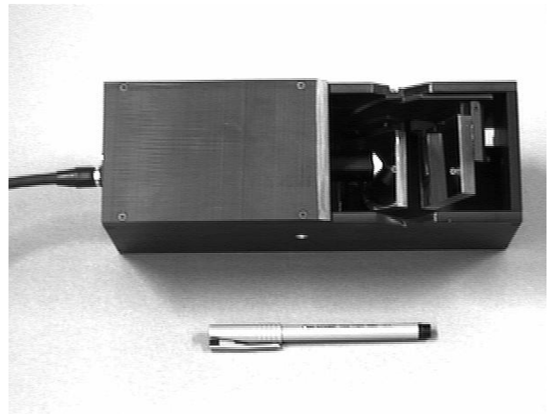

Figure 8: A compact catadioptric stereo camera that uses a single video camera and two planar mirrors. The positions and orientations of the planar mirrors can be adjusted by the user to alter the field of view based on the application. The planar motion constraint is used to robustly self-calibrate the device. A stereo algorithm has been developed that uses just a PC to produce 340x240 depth maps in real-time.

Single-camera stereo provides several advantages over traditional two-camera stereo, which we will briefly enumerate here. (a) Identical system parameters: Lens, CCD and digitizer parameters such as blurring, lens distortions, focal length, spectral responses, gain, offset, pixel size, etc. are identical for the stereo pair. Having identical system parameters minimizes the differences between the two views, thus facilitating stereo matching. (b) Ease of calibration: Because only a single camera and digitizer are used, there is only one set of intrinsic calibration parameters. This property simplifies the calibration process. (c) Data acquisition: Camera synchronization is not an issue because only a single camera is used. Stereo data can easily be acquired and stored with a standard video recorder without the need to synchronize multiple cameras.

With these advantages in mind, we have designed and implemented a compact catadioptric stereo system [Gluckman and Nayar, 1998b] (in these proceedings). Other researchers have designed stereo systems using planar mirrors to acquire stereo data in a single image (see [Gluckman and Nayar, 1998b] for a summary of previous work). With respect to previous work, our contributions are two-fold. First, have analyzed in detail the geometrical properties of planar catadioptric stereo. We show that, irrespective of the positions and orientations of the planar mirrors, the two views of the scene must satisfy the planar motion constraint. This observation has been used to developed robust self-calibrating algorithms that compute the fundamental matrix between the two views. Our second contribution is the development of a compact, portable real-time stereo sensor shown in Figure 8. A stereo algorithm has been developed that uses the video output of the sensor to compute a 320x240 depth map in real-time using no more than a PC (see [Gluckman and Nayar, 1998b]).

10 Real-Time Panoramic Stereo Camera

Today's video surveillance systems rely on brightness images to detect and track activities. It is highly beneficial to have depth measurements in addition to brightness. This would permit a user to define critical "volumes" in the scene that need to be secured. The intrusion of any such volume can be robustly detected only via the use of a real-time depth sensor.

Stereo provides a passive approach to real-time depth estimation. However, traditional stereo cameras have limited fields of view that restrict their use in many applications. We have developed a real-time stereo sensor (see Figure 9) that produces 360-degree panoramic depth video [Gluckman *et al.*, 1998] (in these proceedings). The sensor uses two omnidirectional cameras, each with a field-of-view that is larger than a hemisphere. The two omni-cameras are optically aligned so that epipolar lines are radial lines. Once the omnidirectional images are mapped to panoramas (cylinders), the epipolar lines become vertical and hence efficient to search along. Furthermore, each of our omnidirectional cameras has a single effective viewpoint. This permits fast mapping of computed disparities to three-dimensional scene coordinates with minimal calibration. With this panoramic depth sensor we are able to produce 360 degree depth maps in real-time.

Figure 9: A panoramic stereo sensor. The sensor comprises of two identical omnidirectional sensing units. The optical axes of the two sensing units are aligned to ensure that the epipolar lines are radial in the captured images. An efficient stereo matching algorithm, implemented on a PC, computes panoramic depth maps in real time.

11 Contributions to the IUE

The Image Understanding Environment continues to evolve. Over the past year we introduced a new set of classes for image processing. These new classes are IUE independent; they can be used with non-IUE classes and even used outside the IUE. The classes are highly templated and the types of items to be "filtered" are template parameters. With the support of a few helper functions almost any type, from IUE image classes to a raw C++ array, can be used as arguments to the filters. To make it more useful, the image processing model includes built-in support for block iterations over "images" too large to be loaded into memory at one time. This interface presumes the "image" class supports block-reading and block-writing but the user of a filter does not have to bother with those aspects of the implementation. A preliminary version of this library was developed for both the IUE and TargetJr. We are revising this version for use with the IUE3.0 release that is now in beta testing.

A second major IUE contribution we made is the development of code generation processes. Some users felt that the requirement of documenting classes/changes in a latex document before coding could deter them from using the IUE. Thus, we developed a set of macros that provide all the IUE required components of a class without the need for latex documentation. In January 1998, these macros were supplanted by a new version of the code generator that is based on PERL, that extracts the necessary information directly from the code files. This tool has since been enhanced by AAI and is a major feature of the new IUE3.0 release.

Our final IUE related effort has been the development of a system that allows easy creation of custom user-interfaces to the IUE by providing string-based "wrappers". This will greatly simplify the development of new interfaces, and will serve as the basis of our initial attempts at a CORBA interface for the IUE.

References

[Baker and Nayar, 1998] S. Baker and S.K. Nayar. A theory of catadioptric image formation. In *Proceedings of the 6th Internation Conference on Computer Vision*, Bombay, India, January 1998. IEEE Computer Society.

[Boult *et al.*, 1998a] T. Boult, Ali Erkin, Peter Lewis, Ross Michaels, Chris Power, Chen Qian, and Weihong Yin. Frame-rate multi-body tracking for surveillance. In *Proc. of the DARPA IUW*, 1998. (In these proceedings).

[Boult *et al.*, 1998b] T. Boult, Chen Qian, Weihong Yin, Ali Erkin, Peter Lewis, Chris Power, and Ross Michaels. Applications of omnidirectional imaging: Multi-body tracking and remote reality. In *Proc. of the IEEE Workshop on Computer Vision Applications*, October 1998.

[Boult, 1998a] T. Boult. Remote reality. In *Proc. of ACM SIGGRAPH 1998*, 1998. Technical Sketch.

[Boult, 1998b] T. Boult. Remote reality via omnidirectional imaging. In *Proc. of the DARPA IUW*, 1998. (In these proceedings).

[Brown, 1971] D.C. Brown. Close range camera calibration. *Photogrammetric Engineering*, 37(8):855–866, August 1971.

[Gluckman and Nayar, 1998a] J. Gluckman and S.K. Nayar. Egomotion and omnidirectional cameras. In *Proceedings of the 6th Internation Conference on Computer Vision*, Bombay, India, January 1998. IEEE Computer Society.

[Gluckman and Nayar, 1998b] J. Gluckman and S.K. Nayar. Real-time catadioptric stereo with planar mirrors. In *Proceedings of the 1998 DARPA Image Understanding Workshop*, Monterey, California, November 1998. (In these proceedings).

[Gluckman *et al.*, 1998] J. Gluckman, S.K. Nayar, and K. Thorek. Real-time omnidirectional and panoramic stereo. In *Proceedings of the 1998 DARPA Image Understanding Workshop*, Monterey, California, November 1998. (In these proceedings).

[Nayar and Baker, 1997] S.K. Nayar and S. Baker. Catadioptric image formation. In *Proceedings of the 1997 DARPA Image Understanding Workshop*, May 1997.

[Nayar and Boult, 1997] S.K. Nayar and T. Boult. Omnidirectional vision systems: PI report. In *Proceedings of the 1997 DARPA Image Understanding Workshop*, May 1997.

[Nayar, 1997] S.K. Nayar. Catadioptric omnidirectional camera. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition*, pages 482–488, June 1997.

[Nene and Nayar, 1998] S.A. Nene and S.K. Nayar. Stereo with mirrors. In *Proceedings of the 6th Internation Conference on Computer Vision*, Bombay, India, January 1998. IEEE Computer Society.

[Peri and Nayar, 1997] V. Peri and S.K. Nayar. Generation of perspective and panoramic video from omnidirectional video. In *Proceedings of the 1997 DARPA Image Understanding Workshop*, New Orleans, May 1997.

[Swaminathan and Nayar, 1998] R. Swaminathan and S.K. Nayar. Non-metric calibration of wide angle lenses. In *Proceedings of the 1998 DARPA Image Understanding Workshop*, Monterey, California, November 1998. (In these proceedings).