Omnidirectional Video Applications

T.E. Boult, R.J. Micheals, M. Eckmann, X. Gao, C. Power, and S. Sablak⋆

VAST Lab, Lehigh University
19 Memorial Drive West,
Bethlehem PA   USA

Abstract. In the past decade there has been a significant increase in the use of omni-directional video — video that captures information in all directions. The bulk of this research has concentrated on the use of omni-directional video for navigation and for obstacle avoidance. This paper reviews omni-directional research at the VAST lab that address other applications; in particular, we review advances in systems to address the questions " What is/was there?" (tele-observation), "Where am I?" (location determination), "Where have I been?" (textured-tube mosaicing), and "What is moving around me and where is it?" (surveillance). In the area of tele-observation, we briefly review recent results in both human factors studies on user interfaces for omni-directional imaging in Military Operations in Urban Terrain (MOUT). The study clearly demonstrated the importance of omni-directional viewing in these situations. We also review recent work on the DOVE system (Dolphin Omni-directional Video Equipment) and its evaluation. In the area of location determination, we discuss a system that uses a panoramic pyramid imager and a new color histogram-oriented representation to recognize the room in which the camera is located. Addressing the question of "Where have I been?", we introduce the idea of textured tubes and present a simple example of this mosaic computed from omni-directional video. The final area reviewed is recent advances on target detection and tracking from a stationary omni-directional camera.

1   Introduction

Omni-directional vision is becoming an important sub-area of vision research, and has now grown to the point of having its own workshop, e.g. the recent 2000 IEEE Workshop on Omni-Directional Vision. Omni-directional video processing has already been shown to have significant advantages for robotic applications, [Hon91, YY91, Mur95, YYM93, YYY95], with a very strong emphasis on its use for navigation and obstacle avoidance. However omni-directional sensing has many applications beyond computer controlled driving. For example, tele-observation, self-localization, "mosaicing" and surveillance. Our research on each of these application areas will be discussed. Our work uses the Paracamera designed by Shree Nayar and now commercially availible from RemoteReality (remotereality.com).

Because omni-directional imaging compresses a hemisphere field of view (FOV) into a small image, maintaining resolution and captured image quality is quite important, and takes careful design. Before we discuss applications we very briefly discuss some resolution issues and compare a paracamera image with a fish-eye image.

While the process scales to any size imager, our current systems use NTSC (640x480) or PAL (756x568) cameras. For a standard 640x480 camera we can compute the horizontal (vertical) resolution as the ratio of the number of pixels to the horizontal (vertical) FOV in degrees. For example an NTSC camera with a wide angle lens producing a $114°x85°$ FOV has a horizontal resolution of $\frac{640}{114} = 5.6$ppd (pixels per degree) and a vertical resolution of $\frac{480}{85} = 5.6$ppd. For a wider FOV lens, say 150 degrees, we get 4.2ppd.

Because the paracamera images the world in a circular-like pattern, computing its resolution is more difficult than for a standard camera. For horizontal resolution, we consider the direction tangent to the mirrors edge, (i.e. circles centered on the mirror), and for vertical resolution we use the normal direction. If we set the system so that the image of the mirror fills the image of the CCD we capture an FOV of approximately $360° \times 105°$. The horizontal resolution along the edge of the mirror, i.e. edge of the region of interest (ROI), is $\frac{240 \text{ pixels} * 2\pi}{360 \text{ degrees}} = 4.2$ppd. If we zoom in to fill the horizontal aspect of the camera (which limits the FOV to $215° \times 105°$), we increase resolution to 5.6ppd. From this we can see that near the mirror's edge, a paracamera with a $215° \times 105°$ FOV has similar resolution to a regular camera with a $114° \times 85°$ FOV. Since both are using the same camera, there must be a loss in resolution somewhere else. While it may seem counter intuitive, the spatial resolution of the omni-directional images is *greatest* along the horizon, just where objects are most distant. As targets move closer to the center of the mirror, the overall resolution drops by a factor of four.

Downsampled Fisheye

Full-resolution
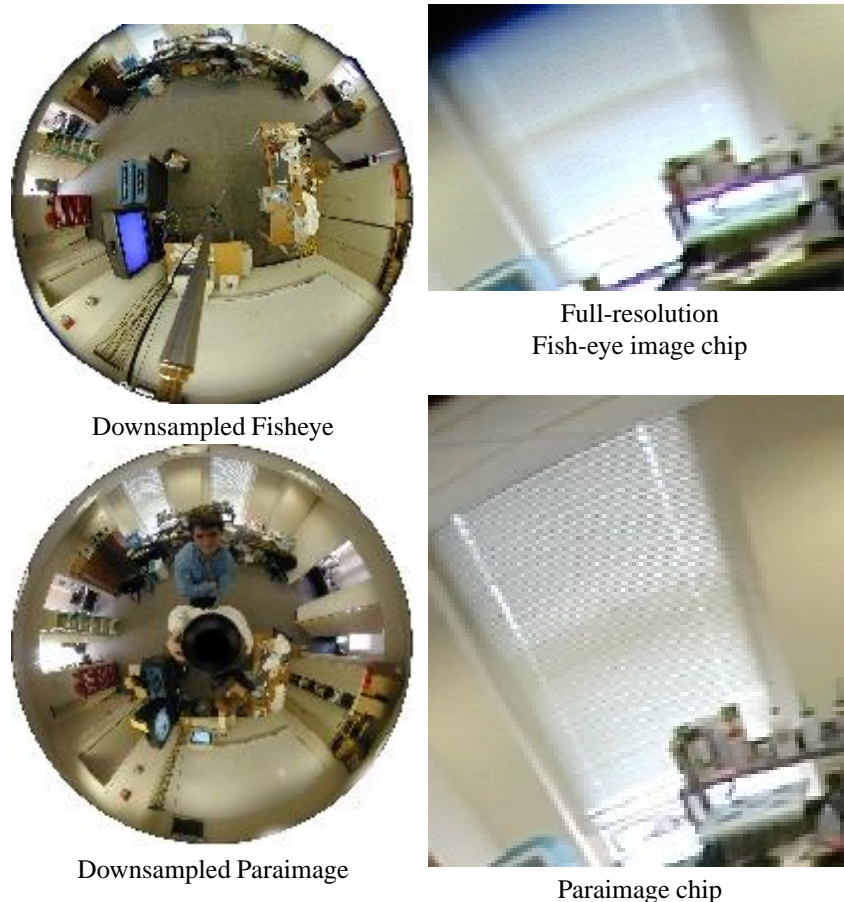Fish-eye image chip

Downsampled Paraimage

Paraimage chip

Fig. 1.: Left column shows a downsampled version of the 1280x960 fisheye image (top) and a paraiage (botoom). On the right is a full-resolution version of a small clip from that image (from about 11 O'clock in the room). The images were taken with the same camera from approximately the same location (though a few people are visible in the paraimage). The chips shown here are different in height because it takes different amounts of the image to show similar content. Details, such as gaps in the window blinds, are lost in the fish-eye image but visible in the paraimage. The ceiling is more visible in the paraimage (360×105 FOV) than the fisheye image (360×90 a.k.a. 180×180).

At this point we note the only way to get close to the paracamera's FOV without a catadioptric system, would be to use a "fish-eye" lens. These cameras also have a non-uniform spatial resolution. However, a fish-eye's resolution is worst along the edges of the image (and best in the center). For comparison, figure 1 shows images taken with a Nikon 360x90 FOV (a.k.a. 180x180 FOV) lens and with a 360x105 FOV Parashot camera. Even though the Parashot has a larger FOV, there are many details clearly visible in the paraimage that are lost in the fish-eye image.

While images captured by the Paracamera may look distorted, the underlying image has a single virtual viewpoint. This single virtual viewpoint is critical for our tele-observation software, as it permits a consistent interpretation of the world with a very smooth transition as the user changes the viewing direction. While there are other systems with large or even hemispheric fields of view, as shown in [NB97], *fish-eye lenses and hemispherical mirrors do not satisfy the single viewpoint constraint*. The single viewpoint also makes it simpler to back-project rays into the world for metrology or 3D target localization, e.g. [TMG+99].

2 Tele-observation: "What is there?"

An obvious application of omni-directional video is for tele-observation. The traditional role of cameras in this domain has been for remote driving. For example, Wettergreen et. al. [WBC+97] demonstrated the use of a panospheric imaging sensor via their long-distance teleoperation of the Nomad mobile robot in the Atacama Desert of Chile. Yamazawa et. al. [OYTY98] have developed and tested a system for teleoperation based on their hyperboloidal omnidirectional camera.

While omni-directional video has advantages for the driver, one of its more interesting properties is that it supports

observation by people other than the driver/camera operator.[2]. Non-drivers may watch the video system either in real time, or during later playback, and analyze it for items of interest. Because we are researching observation, rather than operation, we term this application tele-observation.

The use of omni-directional imaging has the advantage that the camera does not need to be accurately aimed. This observation led us to develop Dolphin Omni-directional Video Equipment (DOVE) — a system for operation by a marine mammal. Dolphins and whales have the natural ability to quickly navigate and locate potential targets of interest, even in very low visibility conditions. The idea of DOVE is to allow the mammal to carry a camera to record the items it finds and bring the video back to a human for analysis. While this could be done with a traditional camera, the limited FOV would require that the animal be more accurate in aiming the camera, and that it actually point the camera at all potential targets. By using an omni-directional camera we reduce the demands on precise operation and also allow the video capture of nearby, but unattended, targets. The tradeoff, of course, is that the targets are smaller in the omni-directional video and less clearly identified.

The system, pictured in figure 2 is described in more detail in [Bou00], which also includes a full description of the experimental analysis. In the experiment, runs were made with both an omni-directional and traditional camera, looking at both isolated and collections of targets. We analyzed the fraction of the time when targets should have been visible (based on the animal's location) and when targets were actually imaged. The omni-directional system maintained good viewing of all targets around 90% of available time, while the traditional wide-field lens only saw all targets less than 15% of the available time. We note that these animals were well trained to operate regular cameras. They always obtained good video of something, but when presented with multiple things that could be targets they did not capture all of them. These tests were done in water with 2-3 meter visibility. For tests in very murky waters, one would need to be very close to the target. Additionally, large targets require a wider FOV. For these experiments, the fractions of visible time on targets was significantly different. In addition, in the forward looking camera tests with two targets, some of the targets were imaged for under one second (briefly seen as the dolphin swam by). Thus, not only were often out of the FOV, some of them were barely visible.
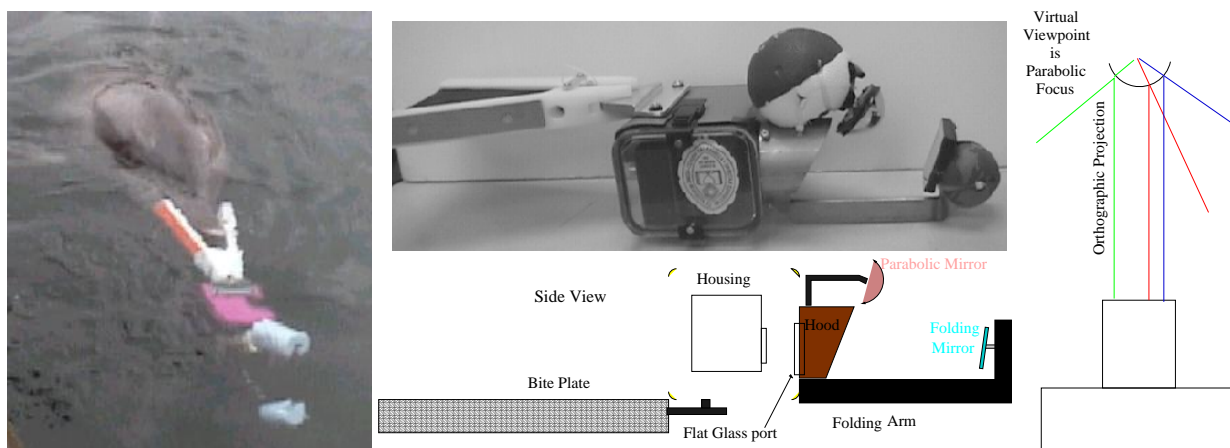

Fig. 2.: The DOVE system: A dolphin omni-directional video system.

A second interesting aspect of omni-directional video for tele-observation is that team members other than the driver can also view the video. For this type of operation there are many different interfaces one might use to view the omni-directional video, e.g. the raw omni-directional video, a head-tracked HMD that unwarps the video in the direction in which the user is looking[Bou98], or a panoramic unwrapping of the video. We have been evaluating[PB00], these interfaces and comparing them to a standard wide-FOV forward looking camera.

The experiments compare different interfaces using a target detection/recognition task. Each user was assigned an interface and given some time to practice with it. Their task was then to watch a pre-recorded video of a vehicle that drove through rooms of various complexity and clutter. The targets were a collection of colored boxes, luggage and people carrying toy weapons. Detection was determined by having the user approximately select the target using a mouse or centering it in the HMD's view.

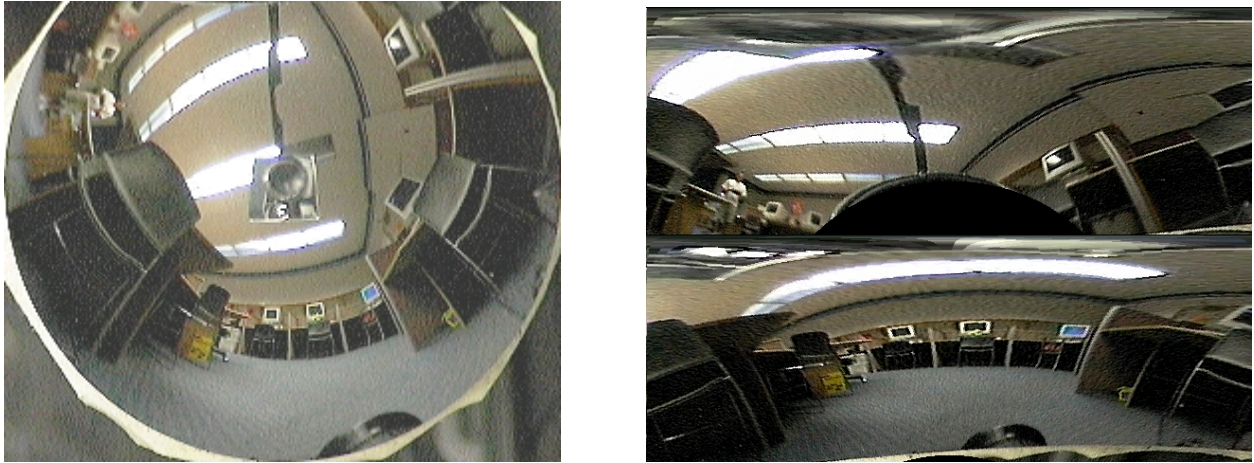---

[2] This deserves a formal study

Fig. 3.: Left: an Omnidirectional (Paraimage) taken from an tele-operated car. Right: a dual-panoramic display of a room. The top is the rear view (left-right reversed as in a mirror), and the bottom is the forward 180 view. This dual panorama is better suited to a the aspect ratio typical of a CRT.
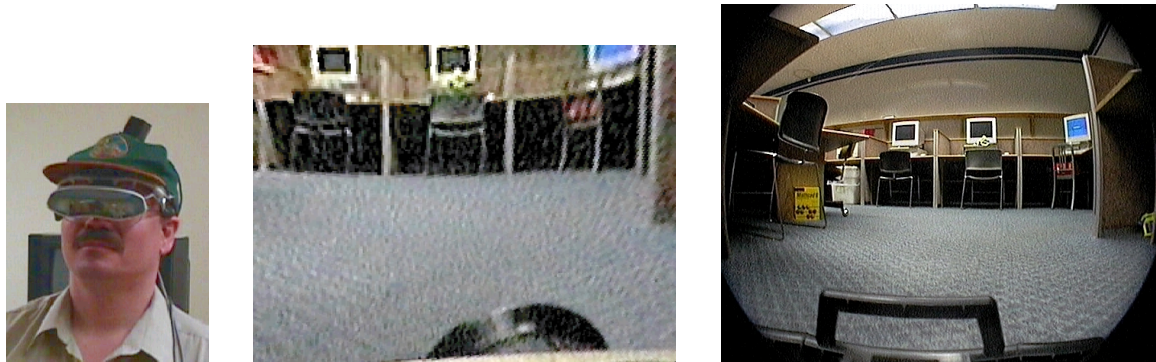

Fig. 4.: The immersive HMD display, and the view from inside the HMD interface. On the right is the raw forward looking camera view.

Results from two of the rooms are shown in figure 5. In these graphs, points closer to the origin are better. Ten subjects took part in this preliminary experiment. From the data for Room 1, we see that the raw omni images outperformed the other interfaces. Remote reality HMD both did very well. The dual panoramic interfaces were clustered tightly but did not perform as well. One subject using the standard camera did well, the others missed many targets and the average for the forward looking interface was the weakest overall. Based on resolution/FOV tradeoffs one might expect the forward looking performance we found: a low false alarm rate but a high miss detection rate. For Room 2, which had more clutter, there were targets that were never visible directly in front of the vehicle. For this environment, the dual panoramic interface was better, and the raw omni interface second best. Here the HMD performed better in total detections, but had more false alarms. Again, the subjects using forward looking cameras had the poorest performance. Even more surprisingly, they had slightly higher false alarm rates than the users of other interfaces.

The experiments are still ongoing, using a larger set of subjects. The preliminary data indicates that for tele-observation, omni-directional interfaces have strong advantages. The best choice between the various interfaces to the omni-directional video, however, depends on the level of clutter and (possibly) the user's experience. We are also extending these experiments to teams of observers. and to include tele-operation, i.e drivers, as well.

3   Textured tubes: "Where have I been?"

While the tele-observation is one way to summarize where the vehicle has been; it is also quite demanding on the human observer. Even at five-times normal playback speed, this is time and attention consuming. The total data size is huge. We have been developing an alternative for our ongoing work with mobile robots. The idea, which we call textured tubes, is to build a mosaic generated locally as though one is looking perpendicular to the wall. This is not a true panorama, but a type of orthographic-like strip mosaic, [RPF97, PJ97].
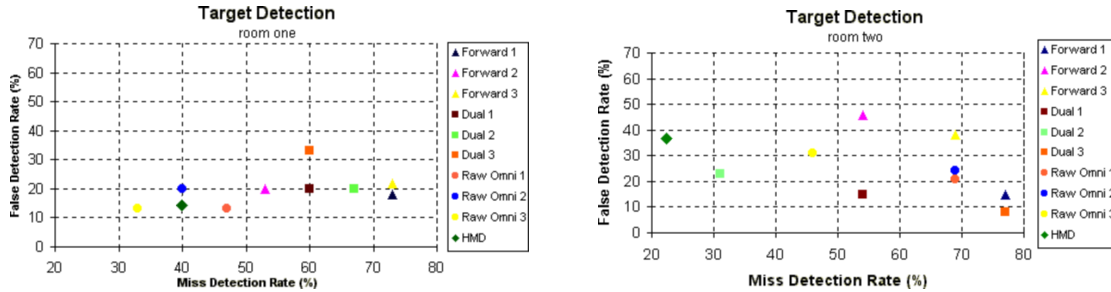
Fig. 5.: Left shows results for the first room, the lab. The right shows is the results for the second room, a highly cluttered office environment.

Fig. 6.: An original omni-directional image from a hallway and one half of the textured tube that results (from the left half of the hallway).

Imagine a robot is moving at constant speed down the center of the tubular world. Our goal is to recover the texture that would be on the walls of that tube. As the robot moves, the real-time video is captured. From this omni-directional video, we determine which strips map to the section of "tube" wall perpendicular to the robot's current location. These strips are then added to a mosaic (see figure 6). After processing, the robot has summarized the world as a mosaic textured-tube.

Of course the world is not a simple tube. In addition, the vehicle does not always move at a constant speed and may often be rotating as well as translating. Thus, building the textured tube is not just a straightforward mosaicing issue. The omni-directional video is important because it allows us to capture the needed slice independent of the vehicle's location. We also expect to exploit it to estimate the vehicle's ego-motion. Without knowledge of speed with respect to the environment, it is difficult to determine the rate at which the mosaic should be extended. For example, in figure 6 we see a "compression" of one of the doors in a region when the vehicle was moving faster than the algorithms estimate. We also note since the apparent velocity of a point depends on its distance so do does the proper sampling/update rate. For simple environments with planar walls, this is not too difficult. For general outdoor navigation, it is a challenging problem as the proper solution to this requires estimates of object distance. The following equation determines how often a sample should be taken.

$$F_s = (d + f)/(fr S_c) \tag{1}$$

where f is the focal length in meters, d is the estimated depth of object in meters, r is the resolution of the image in pixels per meter, (S_c) is the speed of the camera in meters per second, and (F_s), is the sampling frequency in seconds / pixel. If we want to take a five pixel width strip, then the sampling frequency is divided by five to determine how often a sample should be taken.

4   Location Recognition: "Where am I now?"

Another recent application of omni-directional imaging is location determination. For image-based localization, difficult problems include determining where to point the camera, and image/model registration. By using an omni-directional camera, pointed upright, we can produce a system that captures a consistent view variant only to rotation. However, an orientation insensitive technique and a reference map are still needed. Our solution was to treat this as an appearance-based recognition problem. We wanted to recognize the room we were in from a set of features computed from known rooms. We also were looking for a very compact representation that required minimal processing to support small mobile robots.

Color is a very important cue in extracting information from an image, and color histogram comparison has recently become a popular technique for image and video indexing[SB91], [SO95], [LD95], [NM95], [Pan96]. The popularity
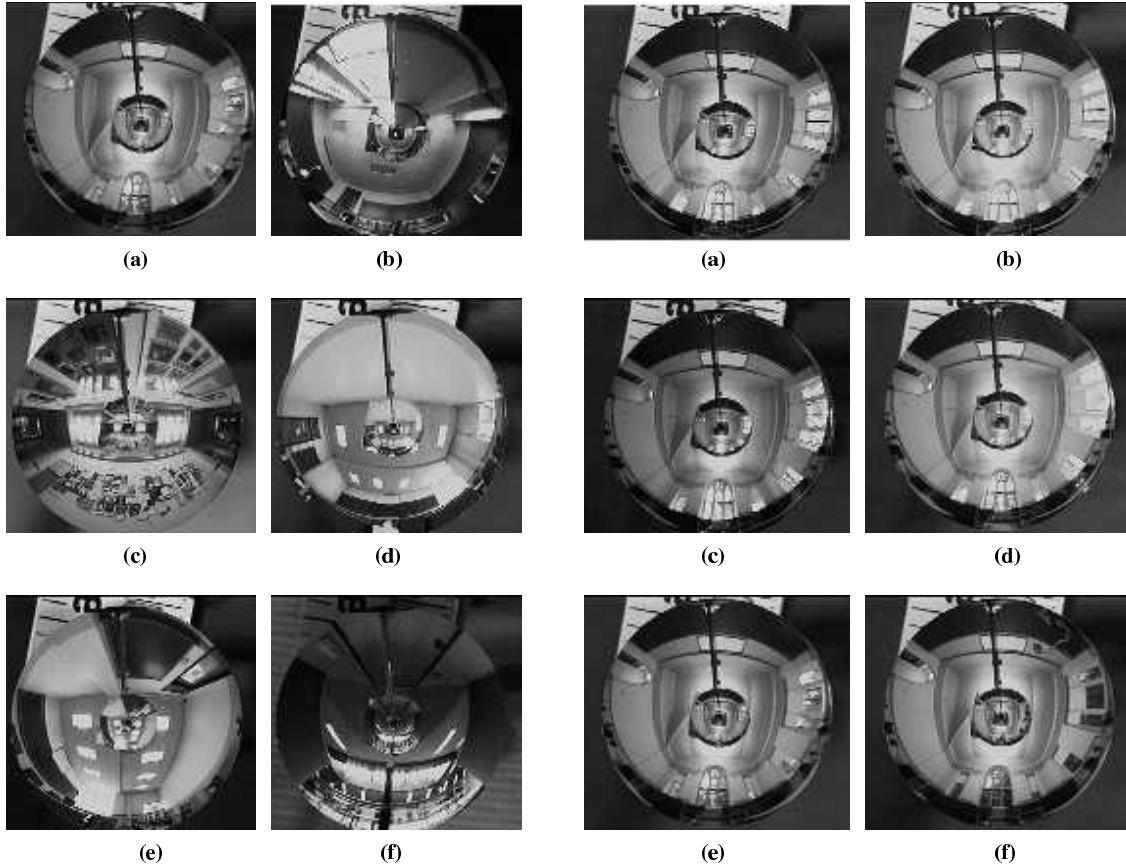
Fig. 7.: Left: Images from six different rooms. Right: Images taken in one room at different times.

of color as an index resides in its ease of computation and effectiveness[LM97]. Some papers suggest that color histograms are resolution independent. It is obvious that when actual bluring occurs, rather than sub-sampling, color histograms depend on the resolution, as high-frequency color textures blur together to form different colors. This is, after all, what color dithering in color printers depend upon. Thus, we include multiple resolutions in our analysis as this will capture a bit of the color texture information as well.

In this application we used images taken using a panoramic pyramid, [YB00], which allows us to generate a multi-resolution image using optics. We used a two layer pyramid, which allowed the system to capture textures at two resolutions but also kept a large FOV. The images were taken from about four feet off the floor and captured much of the walls as well as the ceiling, see figure 5.

Our representation is based on the location of peaks in the Hue (H) and Saturation (S) histograms computed from the panoramic pyramid images. By using the peaks instead of the whole histogram, we significantly reduce the size of the representation. This also makes it less sensitive to minor variations in lighting and scene composition. The latter is important since the lighting may change and the camera will probably not be in the same location as when the reference image was taken. The system currently uses only 7 peaks in H and S to represent the image. There are many details in the histogram peak detection system that cannot be discussed here because of lack of space in this paper. The interested reader should consult [SB99, Sab00] for details.

To evaluate the performance of the approach, we acquired a database using room images with large varieties, and also included a number of similar rooms. Rooms to be included in the database are taken under their normal illumination. For many rooms, the color distributions are similar regardless of changing camera location. For some rooms, especially larger ones, different locations that result in significant occlusion or disocclusion of colors are treated as separate entries in the database, but are labelled as the same room location. The database containeds the histogram peak representations of 205 distinct rooms. As an invariant indexing feature of omni-room image, the color histogram peaks were computed from an image captured at approximately 12:00 noon.

We then conducted a series of recognition experiments using the omni-room image database obtained from 205 rooms by using peaks. The histogram peaks from a test image are used to index into the database. We tested perfor-

mance on room images from different locations and different illumination conditions (using images taken at 9 AM, 11 AM, 1 PM, 3 PM, 5 PM and 7 PM). Figure 7 illustrates some example of omni-room images in our database. In our experimental setup, all images were obtained using a custom panoramic pyramid system. While obtaining the room images, people and all other objects were allowed to move freely in the room.

Overall testing with this database of 394 images from 205 rooms produced a recognition rate of 92 percent. Many of the failures occur at extreme lightning changes, confusing very similar rooms (as often occurs on a college campus), or with moderate variations in camera placement within the room. Details can be found in [SB99, Sab00].

5 Surveillance: "What is going on around me?"

For surveillance applications, especially against adversarials, targets may attempt to conceal themselves within areas of dense cover and sometimes add camouflage to further reduce their visibility. Such targets are only visible while in motion. The combined result of limited visibility and target visibility severely reduces the usefulness of any panning-based approach. As a result, these situations call for a very sensitive system with a wide field of view, and are a natural application for omni-directional video. We have recently developed and demonstrated a frame-rate surveillance/tracking system we call Lehigh Omni-directional Tracking System (LOTS); see [TMG$^+$99, BME$^+$98] for details.

The LOTS system builds on the basic omni-video property that the system can watch a large area without moving the camera. Thus the system is able to build a good background model. In LOTS, there are two backgrounds per pixel that are blended with new input images over time. This allows the system to handle certain natural motions that result in oscillitory disocclusions, such as trees swaying in the woods. The system has an explicit lighting renormalization process that is applied to each target. This normalization allows it to better handle shadows (This lighting normalization process also has access to a third background model that is never "blended"). The system uses a thresholding with hysteresis processes and a novel region connection process we call quasi-connected components. This process allows the system to detect and track small targets (six pixels on target) and have high sensitivity. For examples, see figure 8.
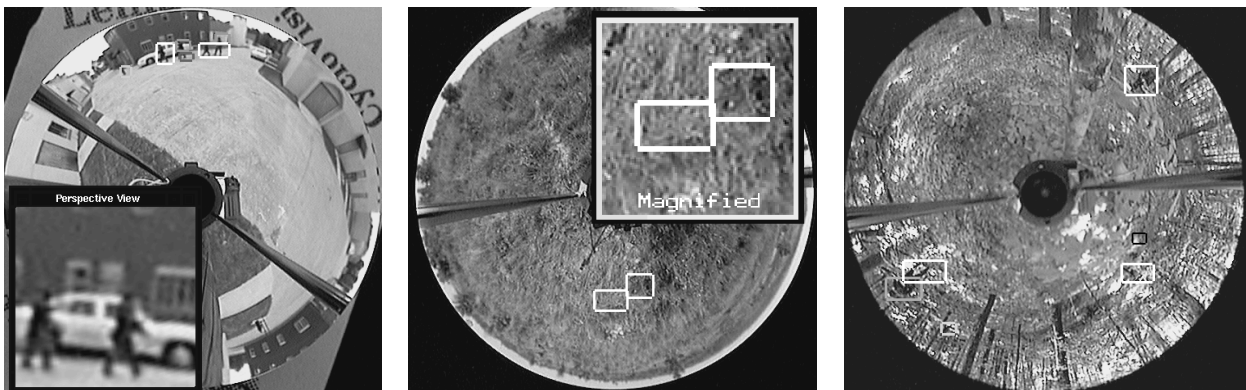

Fig. 8.: Left shows LOTS tracking targets system with a single perspective "target" window showing the most significant target unwarped. (With a left-right reversal because of mirror.) The middle shows the tracking of a sniper moving in the grass. The sniper's camouflage is quite good, but careful background differencing allows LOTS to detect the motion. Frame-to-frame motion is small; a good sniper may crawl at under 0.5 meters per minute and be motionless for minutes at a time. The right shows tracking of soldiers moving in the woods at Ft. Benning GA. Each box *is* on a moving target. The multi-background modeling and thresholding with hysteresis are important parts of the system and allow it to ignore the many moving trees in this scene, and also help connect the soldiers in spite of their camouflage.

After tracking the targets the system uses its single-viewpoint model, and a local model of the ground plane, to back-project and locate the 3D position of the target. The system can track multiple (up to 64) targets simultaneously, and maintains 3D tracks of their motion. There are a number of heuristics that estimate target confidence. Using this the system unwarps the top N targets. In ongoing work we are extending this to coordination of multiple omni-cameras with a very long baseline (10-20 meter) stereo.

Evaluation of this type of system is non-trivial and somewhat subjective. LOTS has been demonstrated numerous times, usually "demonstration" sessions are informal and have cooperative targets that are easy to track. For the intended applications there is significant occlusion and camouflaged targets. In these situations it is often hard to say

if a target should be visible or not. It is also not clear when something is a false alarm as compared to a previously unseen animal/insect or a new motion pattern for brush that might be worth investigating. An external evaluation of our system was done in conjunction with researchers at the Institute for Defense Analysis, where their goals were to see how well video surveillance and monitoring could be used to support small unit operations. The scenarios evaluated included: a short indoor segment, two urban/street scenes, a town perimeter (town edge and a nearby tree-line), two different forest settings, and a sniper in a grass field. For the forest and field scenes the evaluation was limited to a 2-4 batch minute "learning" phase for acquiring the multiple-backgrounds; the others had at most 30 seconds of learning. No "learning" based on feedback on false alarms was allowed.

Scene type	Certainty > 0		Certainty > 1	
	Detection Rate	FAR	Detection Rate	FAR
Indoor 1	100%	0.18	100%	0.0
Intersection 1	89%	0.33	89%	0.0
Intersection 2	87%	0.83	62%	0.0
Town Edge/Field	95%	0.90	92%	.34
Forrest 1 (1min train)	92%	1.71	NA	NA
Forrest 2 (4min train)	100%	0.60	76%	0.0
Field (sniper)	100%	1.06	82%	.10
Mean	95%	0.80	NA	NA
Std.Dev.	5%	0.50	NA	NA

Table 1.: Left shows an example from the DARPA VSAM IFD demo with 3 perspective windows tracking the top 3 targets. In the paraimage targets have trails showing their recent "path." Right is a table from the first evaluation. The chart shows frequency of detection and False Alarm Rate (FAR) per minute of the basic LOTS tracker as of Aug. 1998 (before lighting algorithms) and without adaptive feedback. Main sources of false alarms were about 60% "uninteresting" motions (e.g. leaves and bugs) and 30% lighting& shadows.

The summary analysis is shown in table 1. Almost all detections were considered "immediate," with only the most difficult cases taking longer than a second. In the forests and field, most of the missed detections were targets with low contrast moving in areas where there were ancillary motions (i.e. where the system's multiple backgrounds entered a state that reduced sensitivity). In the intersection scenes, most of the missed targets were either too small (but with enough contrast that the human could see them), or in areas with ancillary motion and multiple backgrounds. The main False-Alarms in the town scenes were lighting/shadow effects while branches, animals and bugs were dominant in the forests and fields.

At the time of this initial external evaluation, the only region cleaning phase was area based. A large fraction of detected false alarms were small to moderate sized regions with lighting related changes, e.g. small sun patches or shadows. In a wide field of view, many of these lighting effects can produce image regions that look like a person emerging from occlusion or a moving low-contrast vehicle. The "ghosting" of targets was also noted in their report, wherein a target that is still for a while leaves a false-target in the region that it disoccludes. This feedback lead to additional cleaning phases, in particular the new lighting normalization testing. Our updated system is a component in a SUO/SAS (in a project lead by CMU) that is being installed at Ft. Benning and will be evaluated in field operations sometime in 2000.

Readers can find videos as well as raw data for testing at http://www.eecs.lehigh.edu/~tboult/TRACK/. Note for effective transmission on the web the "results" are MPEG files that have sacrificed some image quality for the sake of compression.

6   Conclusions and Future Work

Omni-directional imaging systems are, quite literally, changing the way we see the world. They have many properties that are candidates for exploration by vision systems, and the applications presented here highlight a few of those properties. The wide FOV means that camera orientation is not critical, which allows for less sophisticated camera operators (dolphins), a simplified room recognition process, and the ability to generate textured-tube representations. Combining its single view point imaging model with its hemi-sphere FOV allows for immersive video systems, tele-observation and also support surveillance systems. We are continuing to expand our research in each of these directions as well as developing new applications. Of particular focus is refining/extending the textured tube technology, combining the

textured tubes and histogram peak technique to provide a verification stage and possibly an online "control" algorithm, more sophisticated human interface experiments, high resolution panoramic pyramids and new sensor platforms.

References

[BME$^+$98] T.E. Boult, R. Micheals, A. Erkan, P. Lewis, C. Powers, C. Qian, and W. Yin. Frame-rate multi-body tracking for surveillance. In *Proceedings of the 1998 DARPA Image Understanding Workshop*, volume 1, pages 305–313. DARPA/ISO, Morgan Kaufmann Publishers, Inc., November 1998.

[Bou98] T.E. Boult. Remote reality via omnidirectional imaging. In *Proc. of the DARPA IUW*, 1998.

[Bou00] T.E. Boult. Dove: Dolphin omni-directional video equipment. In *Proc. of the Inter. Association for Science and Technology Development, Robotics and Automation Conference*, August 2000.

[Hon91] J. Hong. Image based homing. In *IEEE Conf. Robotics and Automation*, May 1991.

[LD95] J. Lee and B. W. Dickinson. Multiresolution video indexing for subband coded video databases. *Proceeding of SPIE Storage and Retrieval for Video Databases*, 2185:321–330, March 1995.

[LM97] R. Lenz and P. Meer. Illumination independent color image representation using log-eigenspectra. Technical report, Department of Electrical Engineering in Linkoping University, October 1997.

[Mur95] J. R. Murphy. Application of panoramic imaging to a teleoperated lunar rover. In *Proceedings of the IEEE SMC Conference*, pages 3117–3121, October 1995.

[NB97] S. K. Nayar and S. Baker. Complete Class of Catadioptric Cameras. *Proc. of DARPA Image Understanding Workshop*, May 1997.

[NM95] B. M. Mehte M. S. Kankanhalli A. D. Narasimhalu and G. C. Man. Color matching for image retrieval. *Pattern Recognition Letters*, 16:325–331, March 1995.

[OYTY98] Y. Onoe, K. Yamazawa, H. Takemura, and N. Yokoya. Telepresence by real-time view-dependent image generation from omnidirectional video streams. *Computer Vision and Image Understanding*, 71(2):154–165, August 1998.

[Pan96] M. K. Mandal T. Abdulnasir S. Panchanathan. Image indexing using moments and wavelets. *IEEE Transactions on Consumer Electronics*, 42(3):45–48, August 1996.

[PB00] C Power and T. E. Boult. Evaluation of an omnidirectional vision sensor for teleoperated target detection and identification. In *Proceedings of the ICRA Vehicle Teleoperation Workshop*, San Francisco,CA, April 2000.

[PJ97] S. Peleg and J.Herman. Panoramic mosaics by manifold projection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 338–343. IEEE, June 1997.

[RPF97] B. Russo, S. Peleg, and I. Finci. Mosaicing with generalized strips. In *Proceedings of the DARPA Image Understanding Workshop*, pages 255–264. DARPA, May 1997.

[Sab00] Sezai Sablak. *Multi-level Color Historgram Peak Representation and Room Recognition System*. PhD thesis, Dept. of EECS, Lehigh University, Jan 2000.

[SB91] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[SB99] Sezai Sablak and T.E. Boult. Multilevel color histogram representation of color images by peaks for omni-camera. In *Proc. of the Inter. Association for Science and Technology Development*, October 1999.

[SO95] M. Stricker and M. Orengo. Similarity of color images. *Proceeding of SPIE Storage and Retrieval for Image and Video Databases III*, 2420:381–392, 1995.

[TMG$^+$99] T. E. Boult, R. Micheals, X. Gao, P. Lewis, C. Power, W. Yin, and A. Erkan. Frame-rate omnidirectional surveillance and tracking of camouflaged and occluded targets. In *Second IEEE International Workshop on Visual Surveillance*, pages 48–55. IEEE, 1999.

[WBC$^+$97] D. Wettergreen, M. Bualat, D. Christian, K. Schwehr, H. Thomas, D. Tucker, and E. Zbinden. Operating noman during the atacama desert trek. Technical report, Intelligent Mechanisms Group, Nasa Ames Research Center, 1997.

[YB00] W. H. Yin and T. E. Boult. Physical panoramic pyramid and noise sensitivity in pyramids. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, June 2000.

[YY91] Y. Yagi and M. Yachida. Real-time generation of environmental map and obstacle avoidance using omnidirectional image sensor with conic mirror. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 160–165, June 1991.

[YYM93] K. Yamazawa, Y. Yasushi, and Y. Masashiko. Omnidirectional imaging with hyperboloidal projection. In *Proceedings of the 1993 IEEE International Conference on Intelligent Robots and Systems*, Yokohama, Japan, July 1993. IEEE.

[YYY95] K. Yamazawa, Y. Yagi, and M. Yachida. Obstacle avoidance with omnidirectional image sensor HyperOmni Vision. In *IEEE Conf. Robotics and Automation*, pages 1062–1067, May 1995.