DOVE: Dolphin Omni-directional Video Equipment

Terry Boult* tboult@eecs.lehigh.edu
Vision and Software Technology Lab, EECS Department, Lehigh University

Keywords: Sensors, Teleoperation,Human/Machine Interface, Virtual Reality

Abstract

Visual exploration and reconnaissance underwater, whether by humans or unmanned vehicle, presents many challenges. One of these challenges is the limited visibility distance in much of the worlds oceans. When visibility is limited there are significant advantages, as natural evolution has shown, to a very wide field of view. While traditional "Fish-eye" lenses have been around for years, they distort the image in a way that is very difficult to convert back into a perspectively correct image. This paper examines the use of recent designs in omni-directional cameras and their use for underwater exploration. Not only is the wide-field of view important in low-visibility, but with a more than hemispherical field of view, it also results in a system which places minimal "aiming" requirements for the camera operator.

Marine mammals, in particular dolphins and whales, have a natural ability to navigate and locate targets even in near-zero visibility conditions and having their help with exploration or reconnaissance has many potential advantages. This paper discusses an omni-directional camera designed to be carried by marine mammals. The system was tested with both a dolphin and a Beluga whale. This paper briefly discusses both system issues for collection and interfaces for human viewing of the resulting omni-directional video.

I. Introduction

Underwater is a very challenging environment for video systems. In most of the world's near-shore water ways, visibility is limited to under 8-10 meters on a good day and often only 1-3 meters. In addition, the true 3D nature of an underwater world adds an additional challenge—it is not clear in which direction one should look.

For remote vehicle operation there are many choices for imaging systems including remote pan and tilt units. For this project, however, the imaging system was intended to be carried by a marine mammal. The underlying goal was to take advantage of the natural echo location abilities of these animals to navigate to targets of interest in murky waters, and then bring back video for analysis by human operators. This would have value in search and rescue, salvage and intelligence operations. There is an inherent added difficulty — the animal operator does not necessarily know what is of interest to the human handlers to whom they will return the camera. Other video systems have been designed for dolphins to carry, but they have limited views and mobility. For this environment it is natural to seek a very wide-field-of-view sensor, ideally an omni-directional video sensor. In this way if the animal gets close to the target, video of it will be collected. Our system is shown in figure 1.

Fig. 1. The DOVE system: A dolphin omni-directional video system.

Previous work has either accepted whatever video the animals have collected, or have used the animal in a cooperative mode. In the latter, the dolphins were trained to respond to auditory signals and could both aim and pan the camera under guidance from a remote handler who was watching a live feed[1]. This did take advantage of the animal's ability to locate targets, but placed significant limits on how far/fast the animals could go.

While our primary goals herein are related to an imaging system to be carried by a marine mammal, much of it applies equally well to underwater remote operated vehicles (ROVs), especially if the ROV is autonomous.

In section II we present an overview of the Paracamera design and its advantages for underwater use. In section III we examine the resolution issues for the camera and compare it to a traditional camera and a fish-eye lens. In section IV we discuss display issues and different ways of viewing the data. In section V we show various images taken with underwater cameras and discuss our testing and overall experiences. We end with a discussion of future work.

II. Paracamera Designs

The ability to generate omni-directional video has been around for years, e.g. see [2], [3], [4], but it has seen limited usage. What has changed recently, and is driving a growing interest, is the combination of simultaneous decreased size

* This work supported by the ONR MURI program contract #N00014-95-1-0601. Also, special thanks to the many people at the SPAWARS center, especially Randy Brill, and the handlers John, Debbie and Cindy.

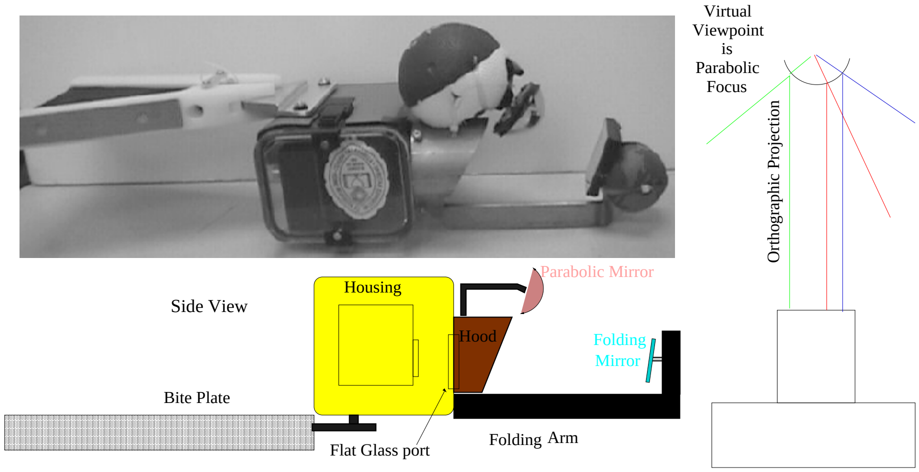

Fig. 2. Components of the marine mammal carried Paracamera. Note basic design is straight imaging of a mirror. The folding does not change the basic imaging system, it just folds the optical path making it more compact.

and increased quality in collection systems, coupled with low-cost means of presenting/processing this data to provide undistorted perspective images.

Our omni-directional work is based on the basic "Paracamera" which is an omni-directional camera designed by Shree Nayar [5].[1] This camera directly captures a full hemisphere (or more) while maintaining a single perspective viewpoint allowing it to be used for full motion video. Furthermore, placing two of these Paracamera systems back-to-back allows a true viewing sphere, i.e. 360 x 360 viewing. Unlike fish-eye lenses, each image in the Paracamera system can be processed to generate *geometrically correct* perspective images in any direction within the viewing hemisphere.

As shown in figure 2, our custom variations of the Paracamera omni-directional imager combines a orthographic lens and a parabolic mirror. To make it easier to operate, a small flat "folding mirror" is added to fold the optical path such that it is parallel to the axis of the parabolic mirror while keeping the actual camera out of the mirror's field-of-view.

The orthographic lens results in the rays entering the camera being parallel. Rays parallel to the optical axis reflect off a parabolic surface at an angle such that they virtually intersect at the focus of the parabolic surface; the focus of the parabolic surfaces provides a single "virtual" viewpoint. This is similar in spirit to the use of parabolic surfaces in satellite dishes, but dishes use the "inside" of the parabolic surface to collect the energy from parallel rays from a distant radio source.

We note that the use of an orthographic lens, just before the camera, also has an added benefit for underwater work. Because the rays entering the camera are parallel, the system uses a flat glass portal between the glass and the water, while the imaging mirror is in the water. Because the rays of interest are parallel there is not an issue of refraction and chromatic aberration as we cross the air/glass/water boundaries. The curved mirror is made of polished stainless steel, and since it is operating via reflection rather than refraction, the aberrations are small.

Not only does the flat glass portal reduce issues of refraction artifacts, it is also significantly cheaper to build, especially for deep water operation. If one wanted to add a portal to a very deep submersible vehicle, this would be ideal as a small (1-2") flat glass portal is all that is needed to provide a more than hemi-spherical FOV.

Fig. 3. Example Paraimage from an underwater scene. The divers are all around the dive master, who is playing around by putting an octopus on his face.

The images captured by the Paracamera have a doughnut-like shape, e.g. figure 3 shows an underwater scene near a cave containing a shark. While it may look distorted, the underlying image has a single virtual viewpoint. This single virtual viewpoint is critical for our Remote Reality 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 show in [6], *fish-eye lens and hemispherical mirrors do not satisfy the single viewpoint constraint*. The single viewpoint also makes it simpler

[1] RemoteReality Inc (formally Cyclovision Inc) has exclusive rights to this patented design.

to “back-project” rays into the world for metrology or 3D target localization, e.g. [7].

III. Resolution

Because omni-directional imaging compresses a hemisphere FOV into a small image, maintaining resolution and captured image quality is quite important, and takes careful design. While the process scales to any size imager, the 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 $114° \times 85°$ FOV has horizontal resolution of $\frac{640}{114} = 5.6$ppd 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 approximately $360° \times 105°$ FOV.[2] The horizontal resolution along the edge of the mirror (i.e. edge of the ROI) is $\frac{240 \text{ pixels}*2*\pi}{360 \text{ degrees}} = 4.2$ppd. At the mirror's edge the vertical resolution is, as in the standard camera case, the same as the horizontal. 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 one targets move closer to the center of the mirror the overall resolution drops by a factor of 4.

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 packing of pixels into the image array. However, a fish-eye's resolution is worst along the edges of the image (and best in the center). For comparison figure 4 shows a small area of a raw taken with a Nikon 360x90 FOV (a.k.a. 180x180 FOV), one using a Nikon fish-eye lens and the other with a 360x105FOV Parashot cameras. Even though the Parashot has a larger field-of-view, there are many details clearly visible in the para-image that are lost in the fish-eye image.

[2] Unfortunately, terminology for describing large fields of view is not always consistent. We are using the notation that FOV is measured with 2 angles. We would say a hemispherical FOV is $360° \times 90°$ and a full spherical view would be $360° \times 180°$. One could also refer to a hemisphere FOV as $180° \times 180°$, however this format is difficult to use for fields of view greater than a hemisphere. It is also common for the angles to be use so that overlap, e.g. a hemispherical field of view described as 360×180 and a sphereical view as 360×360.

IV. User interfaces for Omni-directional video

Given the more than hemispherical field of view, there is the question of how to present this data to the user. We have developed a few different interfaces, and in [8] we discuss experiments to measure the effectiveness of the different interfaces for the task of building interior reconnaissance. The interfaces are discussed in [9], and can be broken down into three groups:

- highly immersive: giving the user the impression they are at the remote location; hence, we call it *Remote Reality*.
- informative: giving the user access to remote “information” in any or all directions, while still maintaining the user's local situational awareness.
- augmentive: enhancing either of the above interface with overlayed contextual information. This reduces immersion and adds complexity to the system, but it can increase situational awareness.

The first two are briefly described here.

A. High Immersion: Remote Reality

Our first interface is immersive, like in many virtual reality system, but because it provide video access to a remote location we refer to it as *Remote Reality*. This interface uses a bi-ocular HMD with a head tracker, see figure 5. The head tracker provides orientation information that is used to determine the viewing direction for the unwarping map. As the HMD turns (or if the users request a software “zoom”) the virtual viewpoint is stationary; only the direction of the virtual “imaging array” is moved. We note that use of an HMD makes this ideal for shipboard interfaces, as HMDs are viewable even direct sunlight.

While this type of interface could be built with a fish-eye or other panoramic image generation process, there are technical difficulties with doing so. In particular if the viewpoint is not constant (or at least constrained to be in a very small volume), the result is a lurching or bending in the images as the HMD changes orientation. Such artifacts significantly reduce the immersion and increase occurrences of HMD-induced motion sickness.

We note that some graphics/VR-oriented professionals might be quick to dismiss the remote reality interface as inadequate when they hear about the output resolution, 320x240@16bit color. However, as an informal point on the “quality” of this interface, we note that the initial system has been demonstrated to a large number of people ($over 1000$), e.g. see [10], [11] and [12], with very positive feedback from most of them. Even the “skeptics” who have tried it admitted they were surprised at the quality. While the resolution is far from that of high end graphics systems, the naturalness of objects, fluidity of motion and the complex/subtle textures (even at low-resolution) of the video seem to make up for the pixel loss.

B. Informative

For other situations it may not be acceptable for the user to be completely immersed, or the use of a head-tracked HMD is unacceptable. Thus we have been investigating different types of informative, but minimally invasive, interfaces. The display might be via a small unobtrusive monoc-

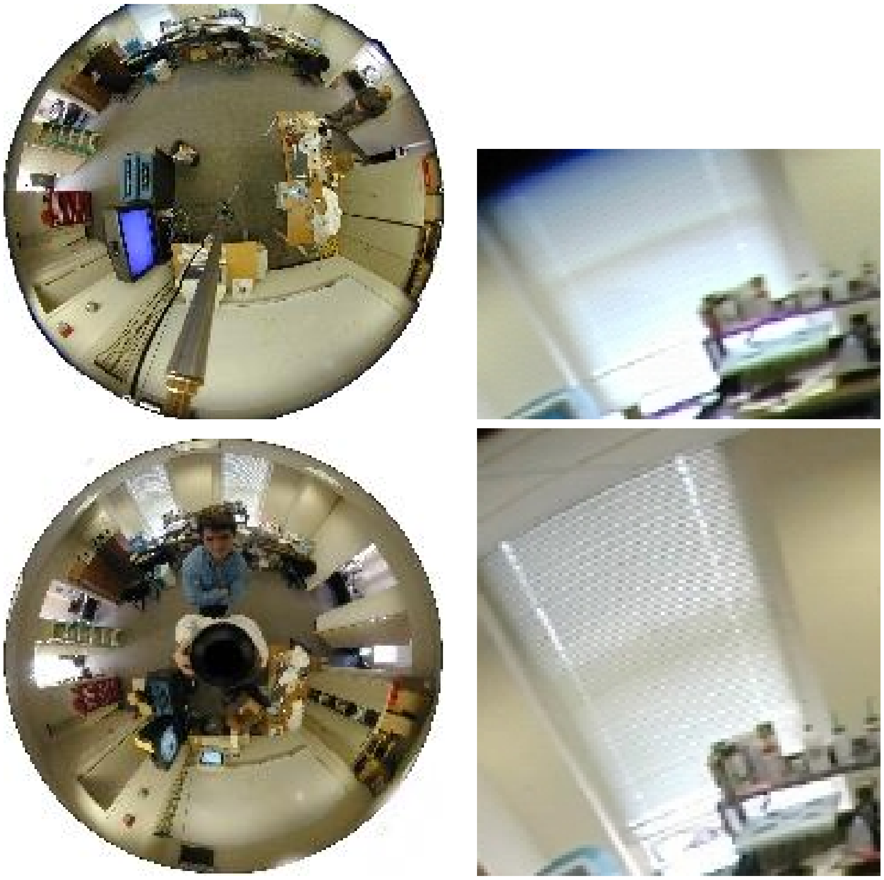

Fig. 4. Top row shows small version of the 1280x960 fisheye image, with a blow up of a small clip from that image (from about 11 O'clock in the room). The bottom row shows the original paraimage and a similar clip taken from that paraimage. The images were taken with the same camera from approximately the same location (though a few people are visible in the paraimage). The images shown here are different sizes because it takes different amounts of the image so show similar content. Details, such as gaps in the window blinds, are lost in the fish-eye image but visible in the para-image. The ceiling is visible on the right because the paraimage has a larger FOV (360x105 FOV) than the fisheye image (360x90 a.k.a. 180x180).

ular HMD, see figure 5, or a computer screen or even a hand-held device such as the portable TV. Without the head-tracking we must either provide a means to choose a viewing direction, or somehow provide an interface that provides information in all directions at once.

Fig. 5. Left is an immersive interface: Remote Reality head-tracked HMD and right is an informative monocular display with (a track-ball pointer).

The "simplest" interface, is to view the paraimage on a display device. This approach has three primary advantages:

1. There is need for the user to "point", as the display shows all directions at once.
2. There is no added computational requirements.
3. The direction within the image is the actual direction from the camera to the object of interest.

The primary disadvantage is that the interpretation of the image is not as intuitive. As can be seen in figure 3, the lower part of the image is relatively easy to understand (front of vehicle or animal), but objects behind the vehicle are upside down. With a little training, however, it becomes quite understandable. This is now the preferred interface by our group for teleoperation operations in complex environments. If upside-down viewing is a problem, hand-held displaces can be rotated if needed, or inexpensive video flippers could be used. This interface was also found to be the best in our building clearing experiments.

Another option is to provide the user with some type of pointing device, e.g. the belt-worn track-ball in figure 5, or a mouse on a laptop computer, where the pointing device is used to choose a viewing direction. The advantages of this is that they can maximize the displayed resolution, and, when needed, can choose new viewpoints. The disadvantage is that choosing a view requires a free hand and some practice with the interface. This approach can be effective for team operations where someone is assigned a task to watch a particular direction.

An obvious alternative is to use a panoramic view. Unfortunately the aspect ratio of the panorama from our images is $\approx 1500 \times 240$, and is far from that of most display technologies and direct display would result in very poor visible resolution. There is also the question of the type of panorama to show (spherical, cylindrical, or some custom version). To help with the resolution issues we display the scene in a split view panorama. This approach has a panorama for the forward (with respect to vehicle) and one for the rear-view (with left-right reverse as in a rear-view-mirror of a car). These are then stacked to provide full coverage in a 4x3 aspect ratio display. Note that this interface requires very little training and no user interaction, but places the highest demands on the computing and I/O subsystem (we warp the full 640x480 image) and display resolution.

C. Systems issues

The first prototype immersive system strove to minimize cost while maintaining acceptable quality. Thus the system uses COTS parts. Our current data collection system cost approximately $4K (+$1K for underwater housing) and the computing/HMD play-back system cost about $4K. The system uses a 233Mhz K6 CPU (running Linux) and a $300 video capture card. The system computes biocular 320x240 30 fps NTSC video. This resolution is reasonably matched to low-end HMDs such as Virtual I-O glasses, though our current display is using a Sony Glasstron and an Intersense head-tracker. Better HMD's and head trackers are commercially available, at costs ranging from $2K to $10K. We have recently ported the system to run on a laptop, using AVI/MPEG files, and expect to demo a version of this system at the IASTED meeting.

V. Experiments

We have built three different underwater prototypes. The first two were for human operators and the final was designed for operation by a dolphin or whale.

The human operated cameras have been used in many settings including Hawaii, the Florida Keys, Curcao and Monteray Bay, CA.[3] One of the most interesting of those data collections was a dive with 4 divers in Monterey Bay, in November 1998. The dive was in a kelp bed near a collection of timid sea-lions. While we were on the 50 minute dive the divers though only 4 sea-lions were ever near us and then only when they came into the middle of the group. The omni-directional camera was collecting video for most of the dive, facing in various directions including facing down about 50% of the time. In the after dive review in the upward or forward looking video we saw 4 approaches we expected, but also found 25 different "fly-bys" of the sea lions, such as the one in a figure 6. It was, in fact, this observation that convinced us of the paracamera's potential for animal operation.

[3]No government funds were used for any of the dive trips; I did not even take them as a tax deduction.

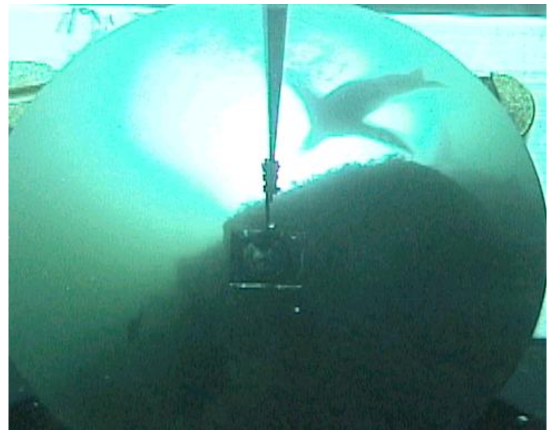

Fig. 6. A sea lion on a fast fly-by of our dive group. The sea lion was visible in the video only for 2 seconds. We never saw it but the omni-directional video captured it :-)

For the experiments we worked with the researchers at the SPAWARS marine mammal program in San Diego, CA. They provided a bite-plate for the animals, and we designed the camera system to mate with it. The camera was a Canon Elura with a custom designed housing (see figure 2). The housing was designed more ruggedly than the human operated system, but was still only 4kg when loaded. The basic system was negatively buoyant and flotation material was added to make it slightly positive. Example videos, and some of the logs of events can be found at `www.eecs.lehigh.edu/~vast/DOVE`.

While an unsupervised "exploration" by the animal is the long term goal, actual experiment was more controlled. The animals were to take the camera over to a second boat and collect video near a set of targets. The animals that took part in the experiment were a Bottle-nosed dolphins, named Buster, and a Beluga whale, named Muk-Tuk. These animals had taken part in many studies including a study using auditory-controls to aim/pan a tethered video camera. In the two weeks before we began the experiments, the handlers had trained the animals to carry the bite plate to a target and point at it until they heard the bridge (an auditory signal).

The data collection was done in August 1999. Data was collected just off the Marine Mammal docks at the SPAWARS center. The visibility in the local area was approximately 1-2 meters. The lighting was broken cloud cover with a few periods of sun. The water temperature was around $60°F$ and the air was $80°F$. Fogging of the camera housing was an issue and required a few breaks to clean the housing/lenses.

After familiarizing the animal with the equipment, largely to get them used to the noises of the camera, the animal was taken into an open bay for testing. The handler, in a zodiac, placed the camera into the mouth of the animal and signaled it to start. The animal would then submerge and go find the second boat (which was 5-15 meters away) were a number of underwater targets could be placed. The target boat also included a handler that could see down to the targets, and would signal bridge when the task was complete, at which time the animal would swim back to the zodiac to return the camera. After returning the camera, and at various other times throughout the experiments, the animals received a part of the animals normal daily allotment of fish.

Data was collected with both the paracamera and a wide angle view. The Elura, in its widest field-of-view, provides a FOV of approximately 90 degrees. When not using the paracamera attachment, the unit was shorter and lighter. Three targets were used: a 6inch stainless steel sphere, a 9inch orange rubber sphere with white tape surface markings, and a mock-mine shape which was approximately 38 inches long, 14 inches high and 12 inches wide. Experiments were done with each target individually and then with a collection of all three. When in the collection, there was a spacing of approximately 1 meter between targets. There were between two and 8 separate runs for each target group for each sensor. The analysis is broken into approaching, viewing, and returning stages. In what follows we use percentage of "appropriate time" during the run i.e. fractions of time when we should have been able to see the targets. These are somewhat subjective measurements since it a little difficult to tell when the animal was approaching, scanning or returning. The animals' tendency, possibly a result of their prior training, was to get the camera very close to the target. Although they were trained to never to hit the target with the camera, they did 10% of the time with the omni-directional system, and 30% of the time with the regular camera.

We first begin the discussion with isolated target experiments in which the animals always collected video of the target. This is followed up with the collection experiments.

A. Isolated targets

For the omni-directional video, the full target, including the whole mock target, was seen 100% of the viewing time[4]. We consider the target not visible if greater than 30% of it is occluded or out of the field of view. For the single target omni-directional experiments the target was visible approximately 93% of the time. Mounting the camera at a higher angle would have prevented this, but it was mounted at this angle to maximize the the view for a forward swimming animal. To my surprise, in most of the experiments the animals stop and change their position in the water and adjust their position to point directly at the camera (again possibly a result of their earlier training of which I was unaware when I designed the camera). The targets were always visible on approach, usually at 2 meters for the mock target and 1 to 1.5 meters for the spherical targets. Unfortunately the single target traditional video was lost in a camera drowning accident.

B. Collection targets

The more interesting experiments simultaneously presented with two or three potential targets. For the three target collection the targets were ordered smallest to largest (6", 9" , mock).

In all runs, the Paracamera captured video on all three targets both during approach and during viewing. However, when the focus of attention was to the right of the mock target, the 6" target (about 2 meters away) was poorly imaged to the point that identification of the item would have been difficult. Blockage of part of one of the targets by the folding mirror was still present but the system maintained good visibility 90% of the time. All the targets were unoccluded for at least some part of the viewing.

For the traditional camera, we had two of four runs where only two of the three targets were seen during approach or viewing, i.e. even a count of the targets would have been wrong. Overall there good visibility of the collection only 11% of the time. In general during the viewing only one target was visible (usually the small metal sphere) largely because the animal would focus on only one of the targets. Because of the limited FOV, the other targets were only briefly seen as the animal toward the target area. On the positive side, the larger size of the single visible target made it much clearer in the video.

[4] Except for its self-occlusions, of course

VI. Conclusion and Future work

The experiments with animal controlled cameras clearly show the need for an extreamly wide FOV. The omni-directional system maintained good viewing of all targets around 90% of availible time, while the tradational wide-field lens only saw all targets less than 15% of the avaible time. However, given the level of training of the animals, it is likely for isolated targets that a $180°$ FOV fish-eye lens would provide better imaging; the animals were surprisingly good at panning the camera over an object and eventually getting the isolated target in the center of the FOV. If, however, targets are not isolated, not well localized, or large, the extra wide field of view, and the extra resolution when the target is farthest from the camera, give an edge to the omni-directional system.

While not directly tested on ROV, the results suggest that the choice of imaging system would be largely driven by where one wants the resolution and the number/distribution of targets. If there is a natural direction of interest and the camera can be assured of pointing in that direction, then a fish-eye camera, which packs its highest resolution in the center of the field-of-view, should be used. If however we are exploring with an autonomous ROV or have a teleoperated ROV and want/need to watch all directions, the Paracamera design has the advantage. The Paracamera designs are usually mounted with the optical axis vertically aligned, thus they watch the horizontal with great accuracy and have limited resolution directly above (or below) the vehicle. With paracameras above and below (or on the left/right) of the vehicle, the full spherical FOV can be captured.

Future work may include exploring the use of this type of video collection for a more exploratory projects where the animal is not swimming to a target, but rather swimming over an area, e.g. an area where unexploded ordinances might be scattered on the bottom. We are also seeking partners to test it with an operational ROV.

[1] R. Brill, 1999. Personal Communication, while visiting SPAWARSYSCEN.

[2] D. W. Rees, "Panoramic Television Viewing System," *United States Patent*, April 1970.

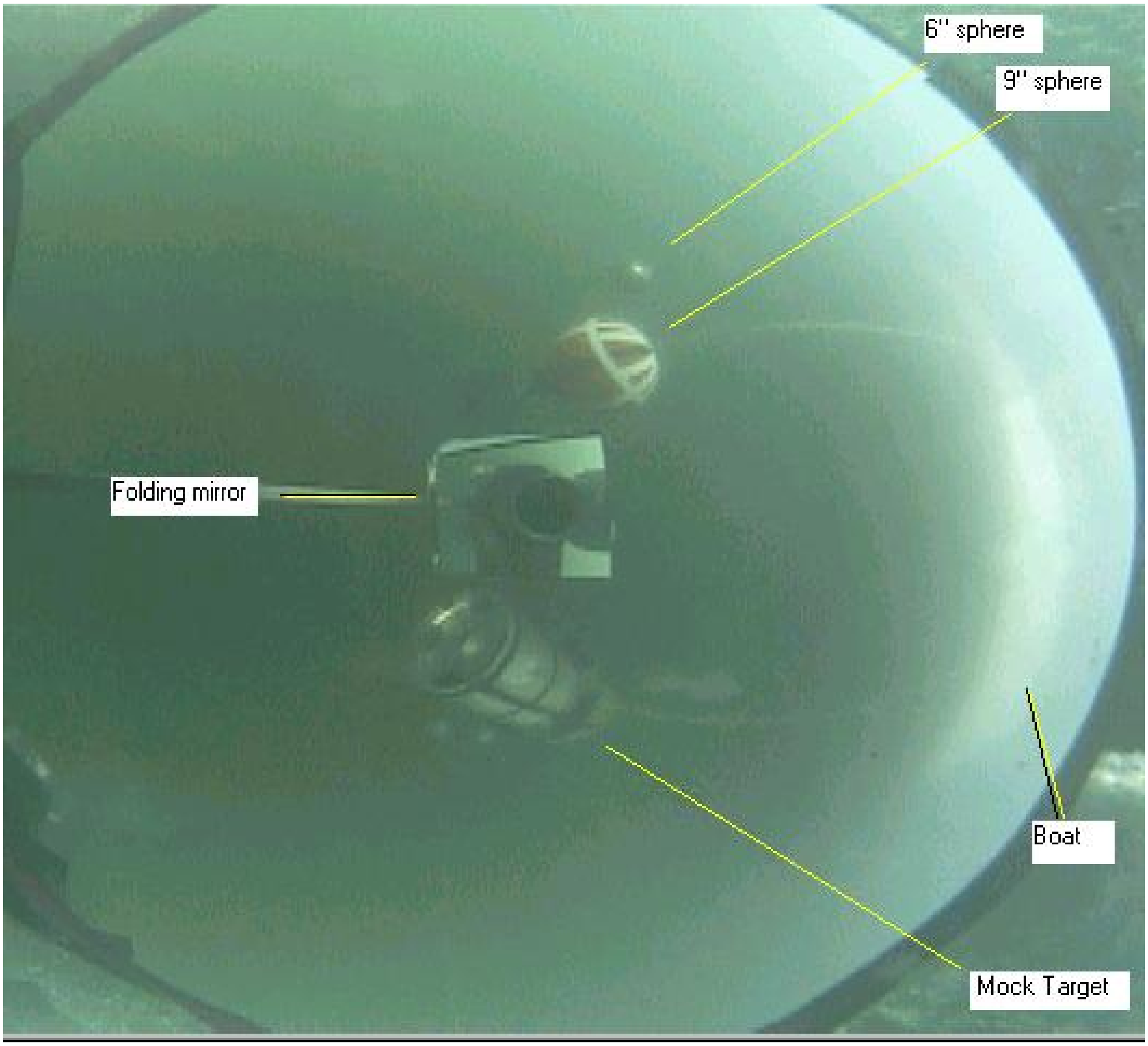

Fig. 7. Raw Paracamera image of collection of 3 targets

[3] G. R. Rosendahl and W. V. Dykes, “Lens Systems for Panoramic Imagery,” *United States Patent*, July 1983.

[4] J. R. Charles, “Portable All-Sky Reflector with ”Invisible” Axial Camera Support,” *United States Patent*, November 1990.

[5] S. Nayar, “Catadioptric omnidirectional camera,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 482–488, June 1997.

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

[7] 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*, pp. 48–55, IEEE, 1999.

[8] C. Power and T. Boult, “Evaluation of an omnidirectional vision sensor for teleoperated target detection and identification,” in *Proceedings of the ICRA Vehicle Teleoperation Workshop*, April 2000.

[9] T. Boult, “Personal panoramic perception,” in *Proc. Int. conf. on imaging science, systems and technology*, pp. 383–390, World Sci. Eng. Soc., July 1999.

[10] T. Boult, C. Qian, W. Yin, A. Erkin, P. Lewis, C. Power, and R. Micheals, “Applications of omnidirectional imaging: Multi-body tracking and remote reality,” in *Proc. of the IEEE Workshop on Computer Vision Applications*, Oct. 1998.

[11] T. Boult, “Remote reality demonstration,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1998. Technical Demonstration.

[12] T. Boult, “Remote reality,” in *Proc. of ACM SIGGRAPH 1998*, 1998. Technical Sketch.