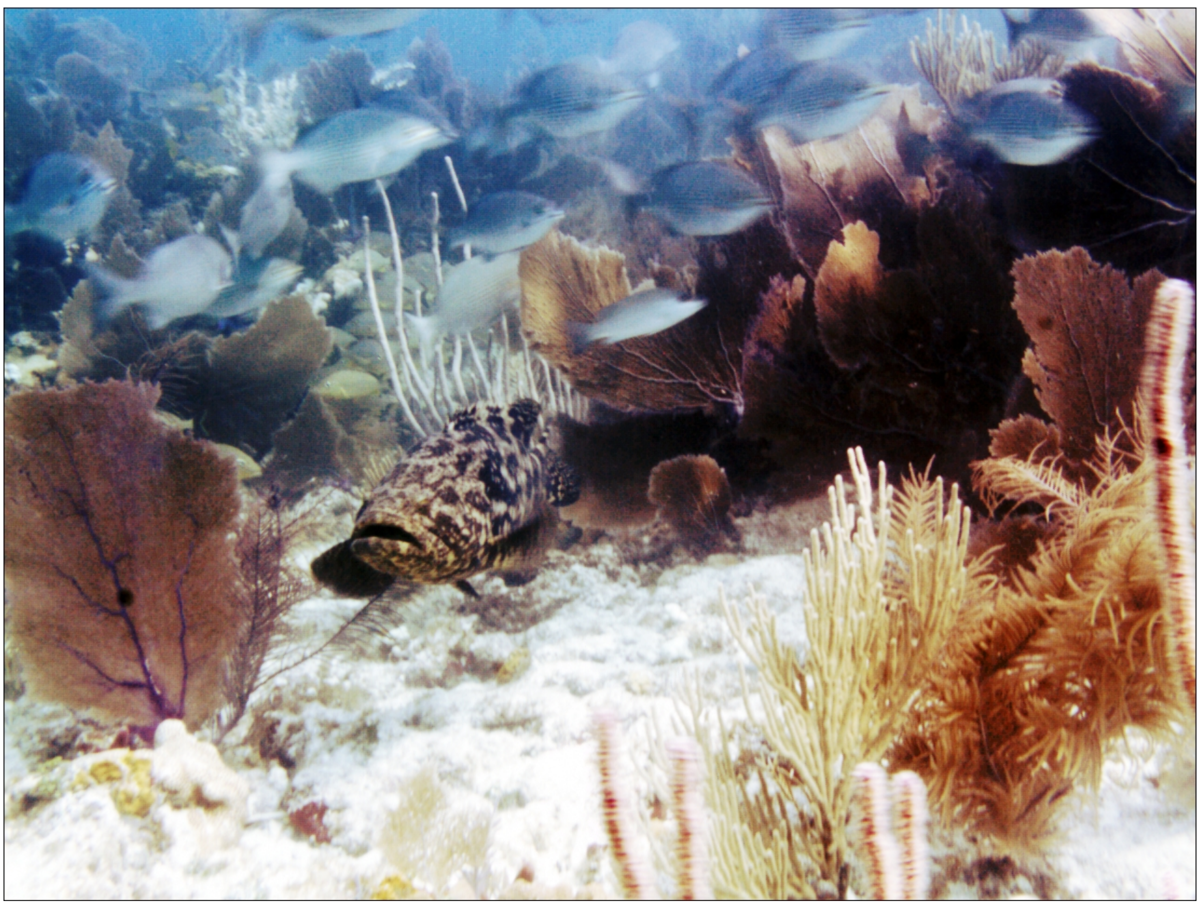
An Underwater Digital Stereo Video Camera for Fish Population Assessment

Authors:

G. Chris Boynton and Kenneth J. Voss
Physics Department
University of Miami
Coral Gables, Florida.



Final Report NOAA/NMFS Contract #GE133F04SE1220:SEFSC Underwater Stereo-Video Camera Research.
Final Report for NOAA Fisheries Coral Reef Program (Grant #NA17RJ1226 SB33).

Abstract

The practicality of using stereo imaging for improved accuracy of fish surveys is examined through the development and testing of an underwater, diver operated, digital stereo video imaging device. The disparity data in stereo image pairs can always be used to extract range and size information given the geometry of the optical system (lens focal length and lens pair separation). The performance of off the shelf, Small Vision Systems, automatic range data extraction software package is examined. Issues of hardware and software design, the reliability of the camera system in field work, and the recovery of size data are also examined.

Cover Photo: Right image of stereo pair showing a small Goliath Grouper nearly head on towards camera, 1280x960 resolution, 02Sep2005, Grecian Reef, Key Largo Florida. See section on Color for complete analysis.

Table of Contents:

Figure Table:

Introduction:

Some nomenclature can be confusing when describing this camera system. The stereo camera system is composed of two separate cameras. To keep the references to the different cameras clear, we have tried to adhere to the following convention – we refer to the stereo camera system as the camera, and we refer to the individual cameras that go into the stereo camera system as the imagers.

The feasibility of stereo video for accurate underwater fish size and population assessment has been well demonstrated by previous investigators, [Harvey and Shortis, 1996], [Van Rooij and Videler, 1996].[Woods, Docherty, and Koch, 1994] This is a report on an underwater stereo video system which builds on the knowledge and difficulties discovered by some of the previous investigators. We have built a system from as many off the shelf parts, hardware and software, as possible. The system uses available non-custom software libraries for data acquisition in a flexible and programmable manner. (The main camera controlling program for data acquisition and exposure control is a customized version of an example program provided by the authors of the libraries that come with the image sensors.) Our goal was to demonstrate an improved practical system that other investigators without high end technical facilities could build and use themselves.

The primary motivation for using stereo imagers is that the information provided by a stereo pair of images of an object allows the determination of the objects range (distance) and therefore its size. Accurate visual sizing of objects underwater is not trivial. Distortions of perception of distance and size in the visual field caused by the air/water refractive interface are difficult to overcome without sufficient repeated training. A method that could passively determine fish size could be of great help in increasing the accuracy of fish population surveys.

After a review of the literature on what techniques had been tried, what worked and what did not, we found that the general method of stereo imaging underwater had already been successfully applied for fish sizing. We examined their methods and any reported remaining issues and determined that a system which automatically collected simultaneous digital images would be the most desirable. Simultaneity of the image pairs is essential for accurate extraction of range and size information.

Two of the issues we gleaned from the previous investigators that we felt we could improve on were the need to provide some automatic built-in method of synchronizing the image acquisition [Harvey and Shortis, 1996]. Also, although it was not mentioned as a problem, it was clear that the requirement of digitizing the video for data analysis was time consuming and added extra expense. We decided that a digital computer controlled camera system with digital sensors and digital storage would be a more convenient and practical evolution of the previous designs. We hoped that allowing for direct storage, retrieval and analysis of the data in a fully digital environment would ultimately allow for more rapid data analysis.

Our camera system is capable of either single directed shots or taking “continuous video”. Single directed shots are nice because the images of interest are essentially pre-selected. Video has the disadvantage of producing a large proportion of insignificant imagery that must be sifted through looking for the images of interest. None the less, video has some unique advantages that should be mentioned. Video is worthwhile because it allows you to record spontaneous events without thinking about the need for a series of shot, one can then go back and select the best shot out of a series of frames, we refer to this feature of video as “over sampling”. The second and less frequently expressed advantage of video is the ability to use motion filtering. The human eye-brain system is very good at detecting moving objects in complex visual fields. This enables us to identify otherwise cryptic objects in our environment. This capability cannot be utilized in a system where there are only infrequent single directed shots of the environment.

Finally a comment on what is meant by “video” and its relationship to frame rates. All video is a “rapid” sequence of still shots or frames that when played back at the same frame rate gives an illusion of continuous coverage and smooth natural motion. Consumer video (NTSC) is shot at 30 full frames a second (there is a subtlety with interlacing). And standard film movies use 24 frames a second. Our camera’s frame rate at the highest resolution of 1280x960 (which has about 4 times normal video resolution) is only 5 frame pairs a second in continuous “movie” mode and thus is not technically what most people think of as video (the camera system has a burst or “buffer” mode which has a slightly higher frame rate of 7 frames a second). At these frame rates smooth continuous normal video is not possible, but the resulting “video” is still very effective at providing the benefits of over sampling and allowing motion filtering, mentioned previously, that make video preferred over directed single shots.

Assembled system:

Figure **1** - Complete underwater digital stereo camera system with battery

Guiding principles:

We set about finding as much off the shelf hardware and software as possible so as to provide rapid development for experimentation and lowered costs. We also wanted to provide a system guide to other potential users which they could readily implement.

Digital

We chose to acquire the images digitally since the analysis of the data would occur in the digital domain. Previous researchers had used off the shelf video cameras and then imported the video data into a computer for digital analysis using computer controlled video playback decks and frame grabbers. Digital imagers also fit in with our plan for using a digital computer to control the imagers.

Another feature of our digital system is that the data is uncompressed, or in any other way processed at acquisition. We did not originally plan this as a feature, but our desire to have a system which will provide data that can be automatically analyzed has caused us to prefer uncompressed data. Lossless image compression would be acceptable (but time and processor intensive), allowing the user to recover all the information in the original images, but any of the standard image or video compression techniques are lossy, the user cannot recover the complete set of original information in the image, or worse, these compression schemes introduce artifacts which will render automatic image recognition schemes ineffective. DV, or Digital Video, is notorious for introducing these types of artifacts into the compressed image stream – under the some circumstances with a complex visual field the viewer will notice a variety of peculiar effects, variously called "mosquitoes", "quilting" and "motion blocking".

Simultaneous Image Pairs

Previous investigators knew that simultaneous images are the only effective stereo image pairs for moving objects. Range data extraction from stereo image pairs uses the image shift, also called disparity, in objects of interest that is due to the unique viewpoint of each imager. No part of this image shift can be caused by the object or camera moving between non-simultaneous shots from the left and right imagers. The image shift or disparity must be exclusively caused by the geometry of the pair of imagers; otherwise the range data cannot be accurate. Previous investigators had mentioned the problems they had with achieving synchronous imaging with two independent video camera imagers and solving this problem was a significant part of their success in getting this technique to work. Due to a publication by [Harvey and Shortis, 1996] which explicitly detailed this problem we recognized the essential need for simultaneous stereo image pair acquisition. The authors utilized a clever clock like object that would always be present in the visual field of each video camera. Then in the video post processing phase, the pair of video tapes would be manually jogged until both were synchronized. Synchronization of all subsequent images from the video tape was maintained by computer control of the separate frame addressable video tape decks. Our idea was to use computer controlled imagers to which we could issue simultaneous command triggers and thus acquire simultaneous images. In our search for these kinds of imagers we found researchers in robotic vision (Videre Design Systems) that produced a stereo camera system that was designed to always produce simultaneous image pairs.

Programmable Flexibility

Given that the initial goal was to show the general feasibility of the idea of using stereo imaging to aid in fish population surveys, we wanted a system that would be flexible enough to accept different imagers, different control programs, different resolutions, etc.

This kind of flexibility is very desirable in a research/prototype instrument, and it is most easily achieved by using a programmable digital computer to control the imagers, if the control program is found to be inadequate or some other control aspect of the system can be improved then, with some limitations, these improvements can be implemented in software.

Networkability

Another advantage of using a "PC" type computer for imager control, data acquisition and storage is that a standard Ethernet TCP/IP type network interface can be provided that will allow various "remote" activities. Data can be off-loaded remotely, programming changes can be made remotely, and remote control is also possible. Our choice of the Linux operating system was seen as an advantage in providing this type of capability without any additional software. Linux provides all this functionality built in.

Physical Components:

External: Anodized machined aluminum case, approximately 14"x 9"x 8", split between front and rear, with O-ring seal and 6 stainless steel adjustable latches:

Figure **2** - Front has two round flat BK7 optical glass ports with center separation of 21cm. The scratches and dings on the metal rings surrounding each window indicate their service as scratch protection for the optical quality windows.

Figure **3** - Rear large Plexiglass window for view of internally mounted 7 inch LCD display, armorized keypad for underwater control by a diver.

We feel the rear window is probably the weakest point with respect to pressure from depth, and therefore it is the component which limits the depth. We estimate that the 0.5 inch thick plexiglass will maintain integrity to 200feet DSW, but we have so far only tested it to 55 feet DSW

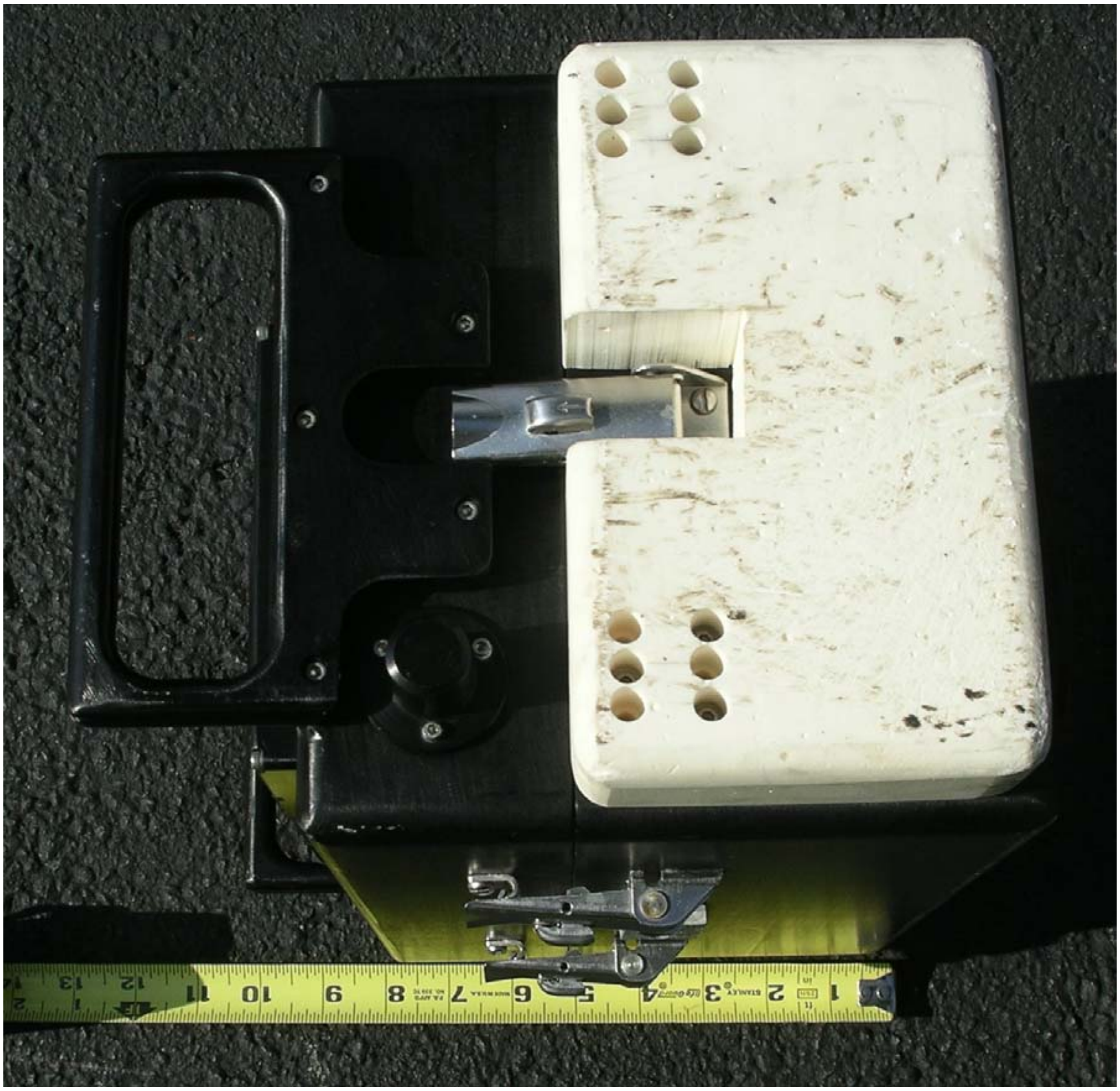


Figure **4** - View of the right side of the camera. Just below and to the left of the center of the image is a small black cylindrical cover held in place with 3 hex-head screws. It covers a Schrader valve for filling the internal volume with dry nitrogen, and providing a small positive pressure with respect to the atmosphere for seal integrity detection. The metal rectangular loop on the left side of the photo is one of the two handles.

Figure **5** - Left side (from rear towards front), extra flotation for providing optimal neutral buoyancy. (the camera has a slight negative buoyancy) Two round “ports” with covers are top: power connection, black cover; and bottom, network connection, gray cover.

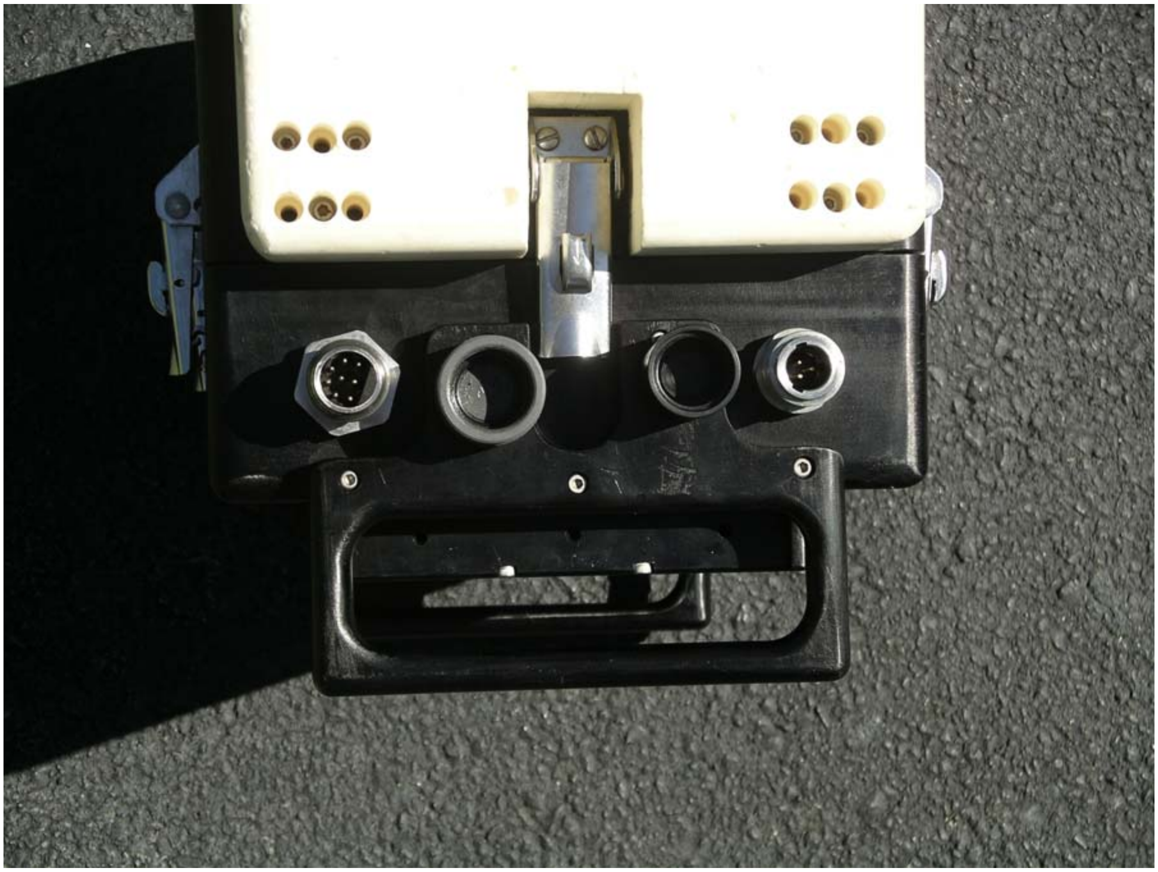


Figure **6** - Photo of left side showing network (left, gray cover) and power (right, black cover) connectors.

Figure **7** - Photo of top of Stereofish camera, showing power cable strain relief, rear LCD and Keypad, handles, flotation on both sides, and the covered power connector on the left side.

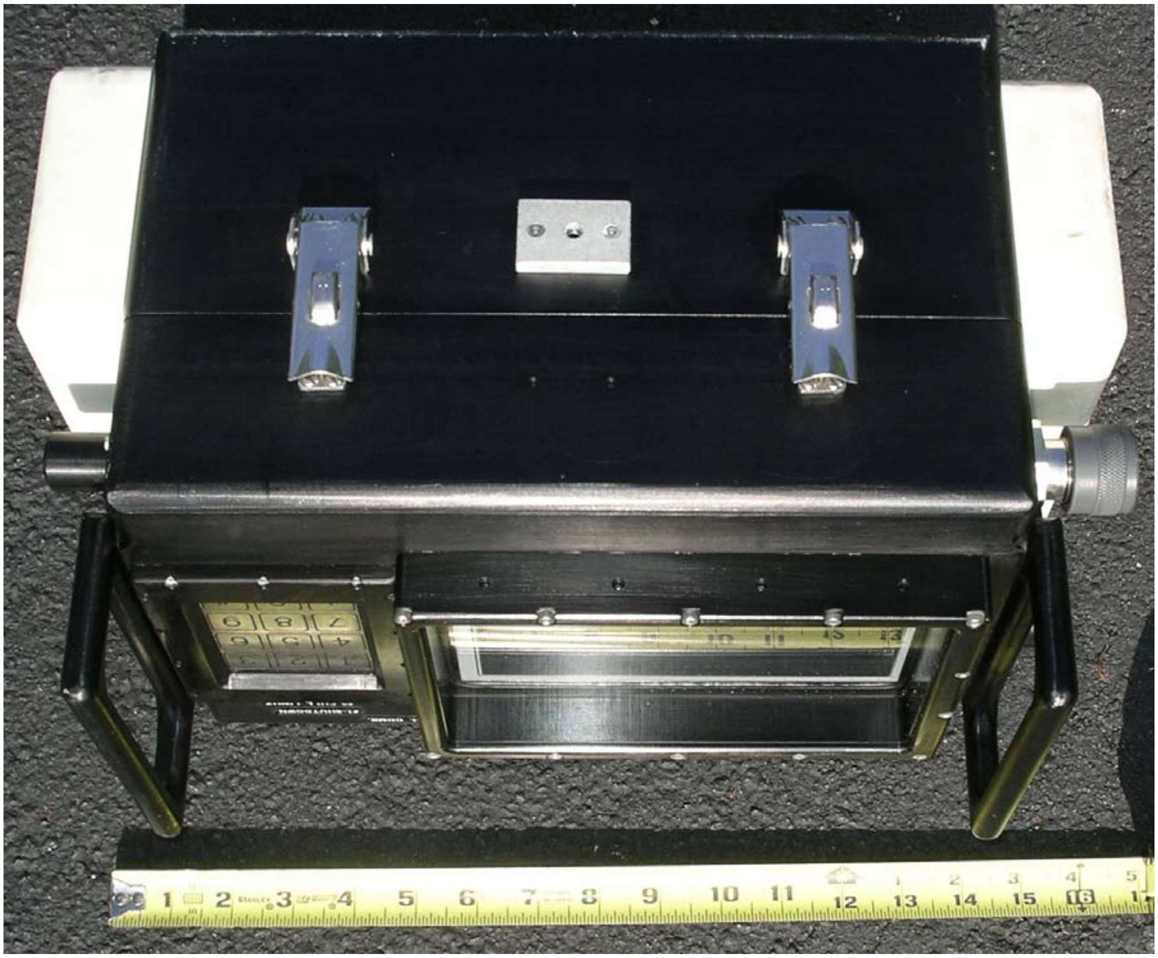


Figure **8** - Bottom of Stereofish camera. Note Schrader valve cover (black cylindrical object) on left, and network port with cover (gray) on right.

Internal:

The camera system underwent a hardware upgrade during its development. We changed the imagers from black and white imagers to a pair of color capable imagers and upgraded the capacity of the internal hard drive used for the control program and data storage. We refer to the first system (black and white imagers) as the Mark I and the upgraded color capable camera system as the Mark II. This hardware upgrade also allowed us to upgrade various software components which we will describe in a later section.

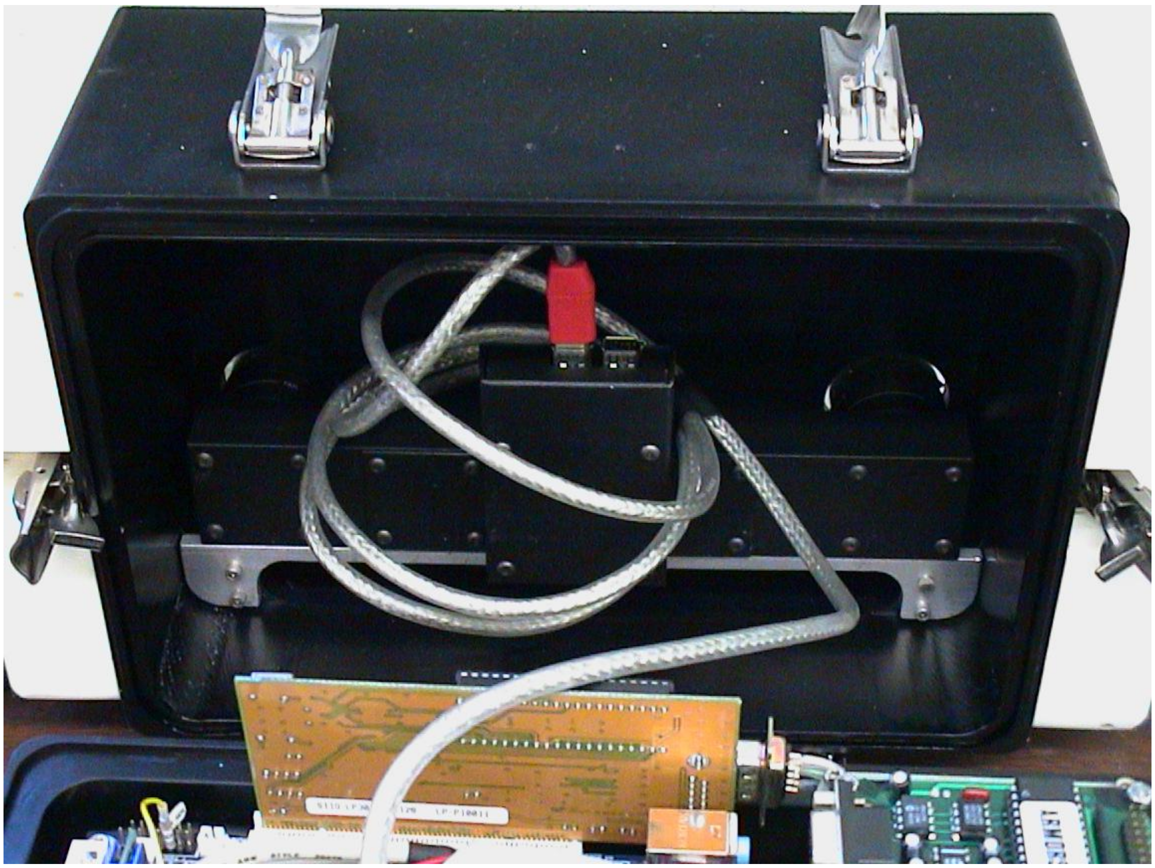


Figure **9** - Front section containing imaging sensors (Videre Designs B&W Mega-D Wide) Thin arcs of light can be seen around the lenses close to the optical windows.

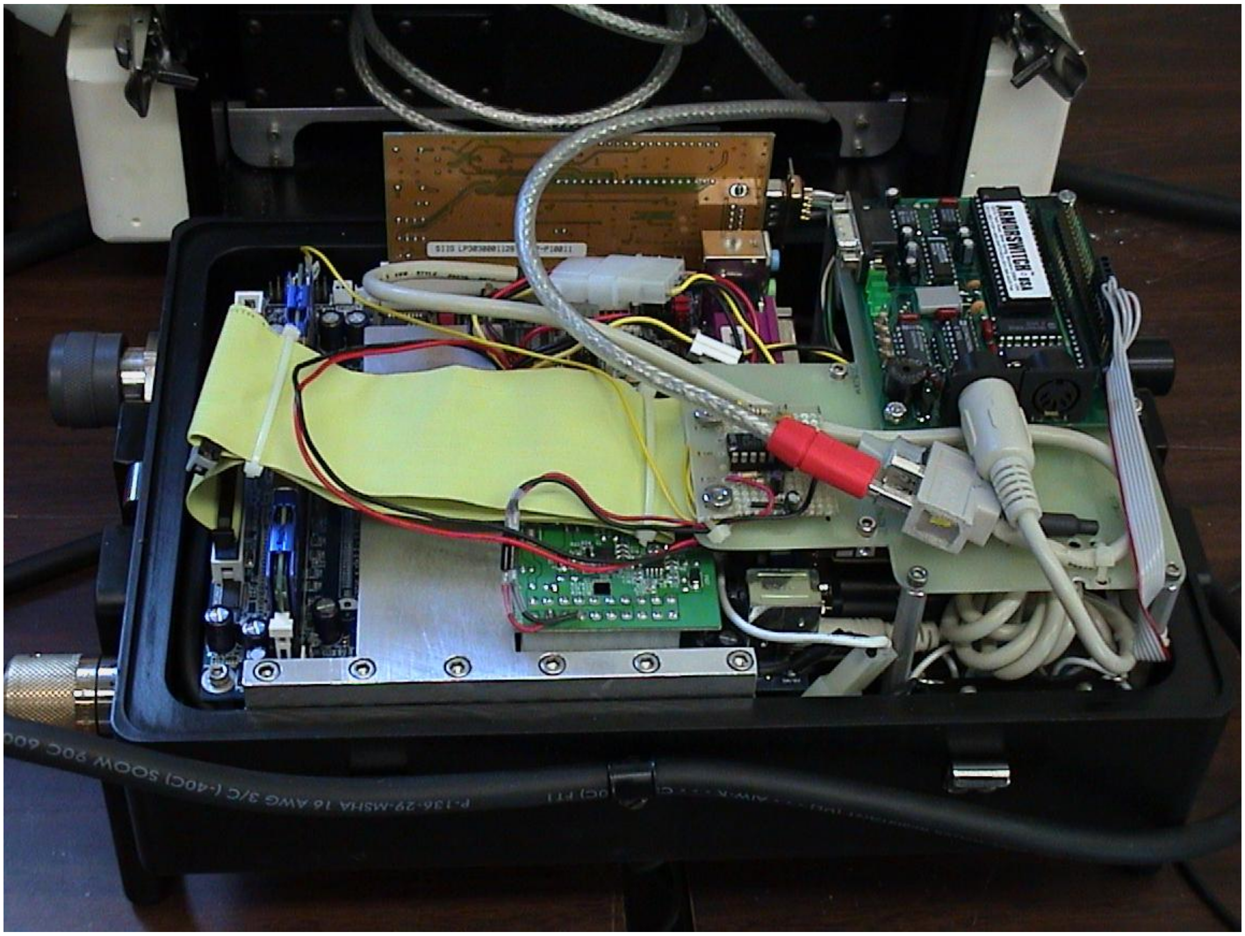


Figure **10** - Rear section, containing computer (motherboard, ram, laptop format hard drive), power conversion electronics, keypad, keypad electronics, and LCD display. Aluminum bar mounted to the side of the case and extending in over the electronics (computer motherboard) is a “heat pipe” designed to conduct heat between the CPU and support chips and the thermal environment of the case.

An outline of internal hardware components:

Mark I

Videre Designs “Mega-D” Wide imager,
- Simultaneous exposure and synchronized readout
- 1280x960 pixel resolution, Adjustable resolution
- Maximum of 1280x960 (2X video resolution, 4X pixels)
- Includes SmallVisionSystems Stereo Pair Capture, Calibration and Analysis Software version 2.3

7mm C mount lenses, manual focus, manual aperture control

VIA EPIA Mini ITX board with 600MHz Eden C3 processor,
- 512MB RAM,
- Integrated IEEE1394 “firewire”,
- Integrated video graphics adapter.

Xenarc 7”LCD display, 16:9 “wide” format.

External Battery power supply -24VDC (2x 12V sealed lead acid in), supplies a 12V and a 5V inverter for supplying power to the internal electronics.

Mark II (hardware changes to above spec)

Videre Designs MDCS2 Variable baseline imager

80GB 5400rpm laptop format hard drive

Figure **11** - Battery and waterproof container.

The power for the system is provided by two 12V sealed Lead-Acid batteries that are connected in series to provide 24V to the camera system. The camera contains several DC to DC electronic power supplies that supply the needed combinations of 5V and 12V to the computer motherboard, the imagers (via IEEE 1394 connections to the motherboard), the hard drive, the keypad and LCD display. The system draws 1.5 – 2A, and one fully charged battery pack will provide over four hours of runtime.

Figure **12** - View of the rear section of the camera case separated from the front.

Software Components:

In keeping with our off the shelf components philosophy we utilized imagers from Videre Designs that were supplied with software from Small Vision Systems (SVS) for acquisition, calibration and analysis of stereo pairs. This software was available for both the Microsoft Windows and Linux Operating Systems. We picked the Linux OS primarily because it is robust, provides excellent remote operation facilities built in, and it is very low cost (free).Linux is developed under an Open Source Software model which allows the users to examine and modify the code. We felt that access to the source code in Linux could be a great help if we needed to modify some behavior of the Operating System for our particular hardware and “real-time” environment. Microsoft Windows at the time did not guarantee the ability to run a real time data acquisition system. The choice of Linux for the Operating System turned out to be a very good one for another reason we had not anticipated. The SVS based libraries and software was significantly more robust on a Linux based system. We use the SVS software on both Linux and Windows based systems and the reliability of the software on the Linux based systems was clearly superior when we first started. The reliability gap has shrunk, but we still prefer using the SVS software packages on a Linux based system.

As we described in the hardware section, the camera system underwent a hardware upgrade during its development. We changed the imagers from black and white imagers to a pair of color capable imagers. We refer to the first system (black and white imagers) as the Mark I and the upgraded color capable camera system as the Mark II. This hardware upgrade also allowed us to upgrade various software components. A major improvement in the software was upgrading the OS from RedHat Linux 7.0 to RedHat Linux 9.0 and upgrading the SVS software libraries from version 2.3 to version 4.1. The OS upgrade introduced a journaling file system which eliminated one of our last issues with reliability. Previously small power glitches or operator errors with turning the power off without properly shutting down the camera system would result in disk errors that could on occasion be very severe, requiring a complete re-imaging of the hard drive inside the camera (requiring the opening of the case, extra power supplies, attachment of floppy drives, etc). The new OS allows for robust automatic repair of damaged file systems and since the upgrade we have not had any further issues with hard disk corruption despite occasional improper shutdowns of the camera system.

Another software component of the camera system is a modified version of the Small Vision Systems capture program. The modifications were made by Marco Monti. The modifications involved displaying various camera system parameters and methods for allowing the diver to modify those settings using only a numerical keypad. When we upgraded the system (Mark II) to use the color imagers we found that the standard SVS software library calls to store the images was much too slow and used 8 times more storage space as the same resolution images from the black and white imagers. We discovered the SVS software was getting the raw data from the color imagers and then converting them to both black and white and color “bmp” format images (using the color imagers Bayer color filter pattern), then storing both the color and black and white images in bmp form. Marco Monti figured out a way to circumvent the SVS storage system libraries and directly store the raw unprocessed images from the imagers. This greatly increased the performance of the color camera system, reducing the storage required by a factor of 8 and increasing the acquisition and storage speeds by a combined factor of about 5 due to the reduced amount of data written and the reduced processing time for raw to bmp format conversion. The only disadvantage of this modification is that the images are now stored in the raw format and requires some post processing before stereo pair analysis can proceed. We have written a post processing program in the IDL data and image processing language to do this.

An outline of the original and upgraded camera systems software follows:

Mark I

- Linux RedHat Version 7.0 (kernel 2.2.x)
- Small Vision Systems version 2.3
- “stereo” camera system control program version 1.1 written by Marco Monti

Mark II

- Linux RedHat Version 9 (kernel 2..4.20x),
- SVS 4.1
- “stereo” camera system control program version 1.6 written by Marco Monti.

Calibration:

Calibration of the stereo system to account for various optical distortions and other parameters of the system requires multiple shots (initially 6, now 10) of a know calibration target in different orientations. The calibration target is a simple black and white checkerboard pattern (9x7) of 54mm squares. The calibration target needs to be photographed in five different orientations and at 2 different distances. We found it useful to shoot at least 2 or three images at each of the 10 combinations of position and orientation. The five orientations can be described as a series where the calibration target is tilted with respect to the camera. The tilt of the calibration target with respect to the plane perpendicular to the optical axis should be between about 30 to 40degrees and should not be greater than 45 degrees. The orientations of the calibration target can be described as follows:

1) perpendicular to optical axis
2) left edge closer than right edge
3) right edge close than left edge
4) top edge closer than bottom edge
5) bottom edge closer than top edge.

We discovered that the series of close images tend to be more difficult to acquire because you want images of the calibration target that fill most of the field of view of the camera. One needs to be careful to get the complete target in both the left and right hand imagers' view.

Figure **13** - A "far" image of the calibration target held at perpendicular orientation. (pic-0-L.bmp, 1MP)

After acquiring the calibration images, and moving them off the camera and onto the "data analysis" computer, one runs a special version of the SVS (Small Vision Systems) software called smallvcal supplied with the Videre Designs imagers. This program allows you to select the 10 stereo image pairs you want to use for calibration purposes. After verifying that the software can recover the known features in the calibration target from each of the set of images, one runs the calibration routine and the software develops a set of parameters that can then be saved and used to recover range and size information for any set of images acquired with these same optical settings and components. I mention the optical settings because we adjusted the focus and aperture settings on the lenses repeatedly to try to achieve the optimum combination of depth of field, image sharpness in the target range and maximum brightness. The adjustment of focus has an influence on the calibration parameters, and the initial focus setting at infinity was not sufficient for achieving sharp images at close distances.

Stereo Image Pair information:

Disparity – range relationship:

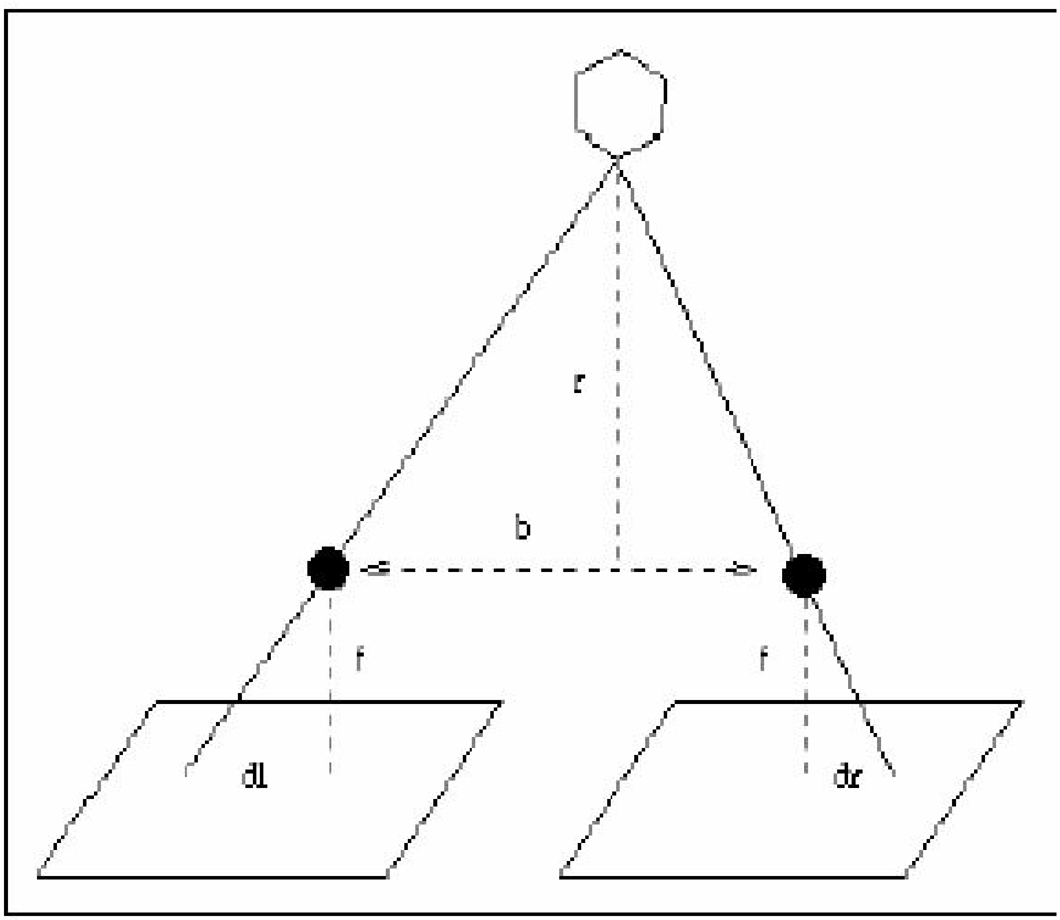


Figure **14** - relationship between disparity, range, f ocal length, and baseline separation. Diagram supplied by Small Vision Systems [Konolige and Beymeyer, 2005].

Where:
r = distance or range,
b=baseline separation of lenses,

f=focal lengths of lenses,
d=(dr-dl) is the disparity.

The fundamental equation which relates the disparity, the baseline separation, the focal length and the distance is:

$$d = b*f/r$$

Notice that the disparity, d, increases with decreasing distance, r.

Theoretical Performance:

We have designed the system to deliver adequate size resolution from 1 meter to 5 meters, when used in the 1MegaPixel mode (1280 pixel horizontal resolution), the size resolution varies as a function of range and range resolution. The range resolution also varies as a function of range. The author of the SVS software claims 1/16 of a pixel resolution for their software. We consider this over optimistic. One pixel disparity resolution seems the most reasonable and best value to use, and it is what we use in our calculations.

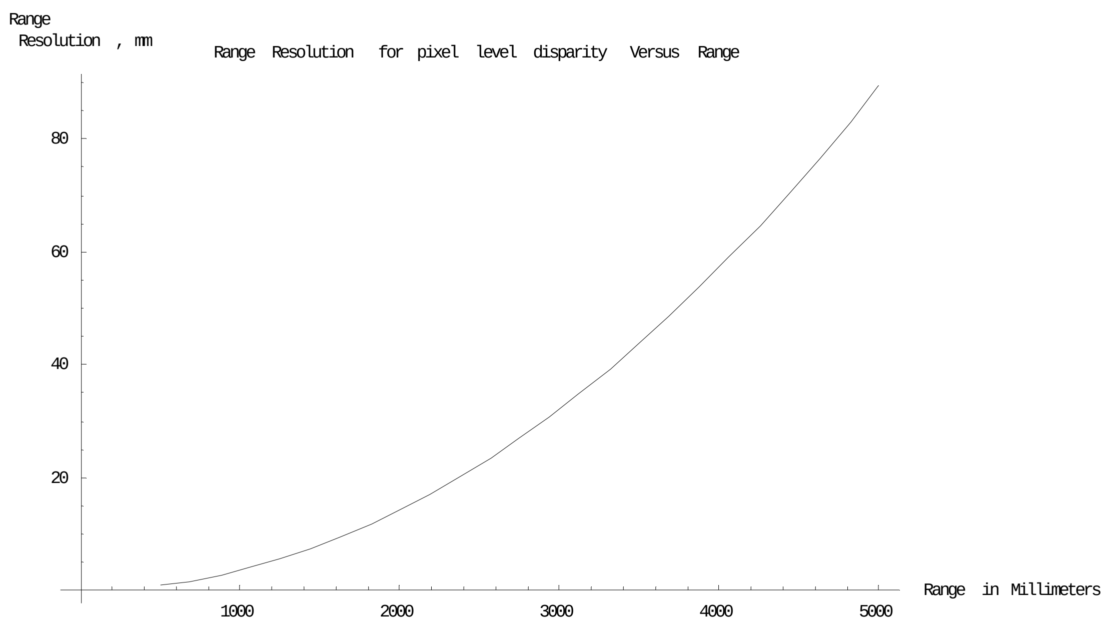


Figure **15** - Range in millimeters versus range resolution in millimeters for single pixel disparity. Notice that as the range increases, the single pixel disparity corresponds to an increasingly large inaccuracy in range resolution.

The range resolution at 5 meters for a one pixel disparity is 89.3mm.One could also say that a one pixel disparity error at 5meters will result in a range error of 89.3mm. The range resolution accuracy varies linearly from 0.18% at 0.5 meters to 1.8% at 5 meters.

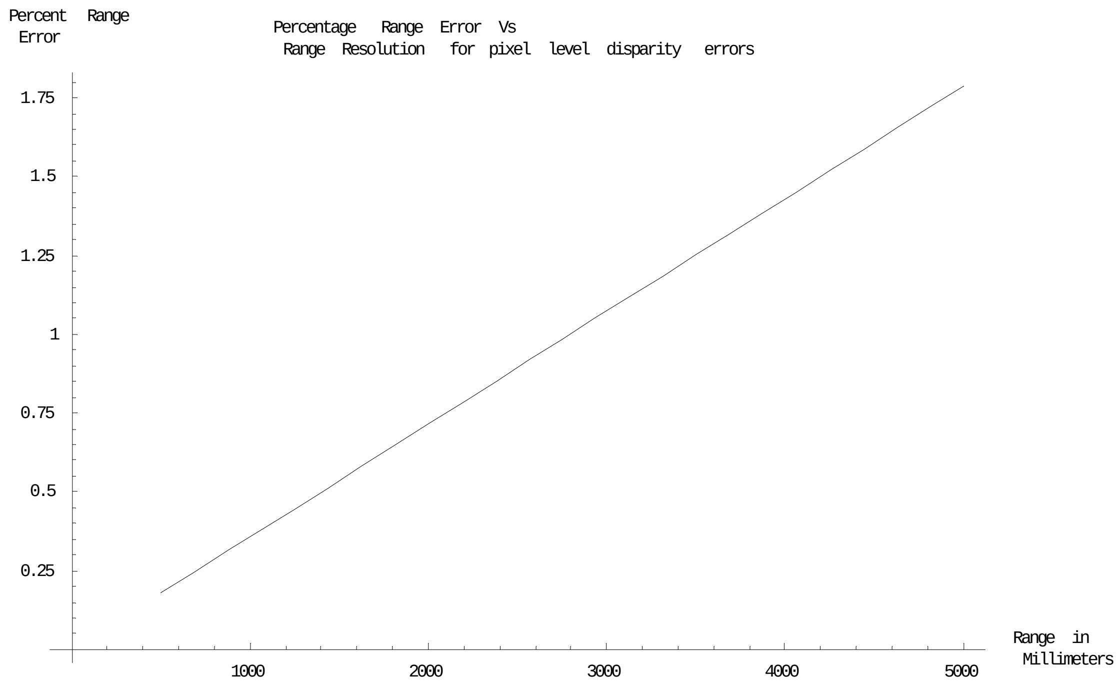


Figure **16**. - Percent Range Error versus Range in Millimeters for 1 pixel differences.

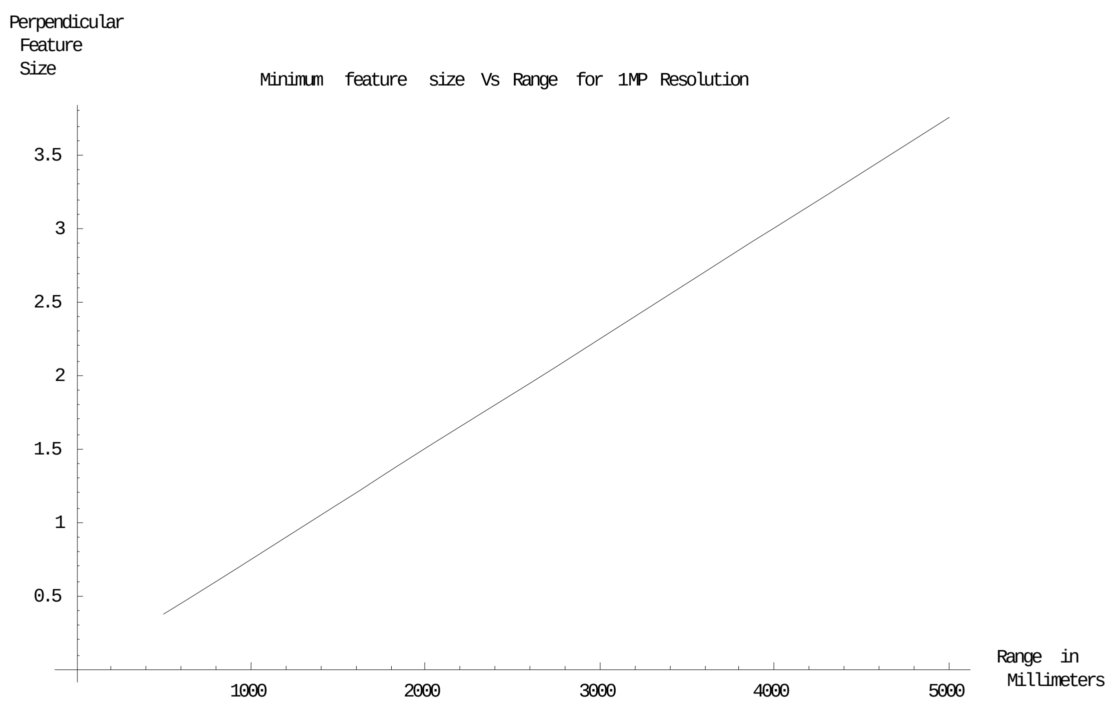


Figure **17**. - Minimum feature size versus range for Stereofish lens and imager combination underwater. Range and features size is in millimeters.

The size resolution accuracy is a function of both the range accuracy and the angular resolution accuracy. The extent of the smallest feature measurable by the imagers corresponds to 1 pixel. For the MegaD B&W imagers at their highest resolution of 1280x960, the pixels are 7.5 micrometer squares. This smallest linear size translates to a smallest angular size measurement with the use of a lens. Underwater, the effective focal length of the lenses is 1.33*7.5mm = 10mm due to the additional refraction at the air-water interface. This smallest angular feature for this imager and lens combination is 7.5E-04 radians, or 4.3E-02 degrees, or 2.6 arc-minutes. This angular size translates to a perpendicular feature size that varies linearly as a function of range. At 5000mm range, the 1 pixel resolution of our imager and lens combination corresponds to a perpendicular feature size of 3.75mm, at 2500mm the feature size would be 1.875mm and at 500mm the perpendicular feature size would be 0.375mm, or one tenth of the feature size at 5000mm.

The total size error at a particular range is a combination of the three size errors, one in range and the other two in the perpendicular feature size.

It is useful to introduce a 3 dimensional Cartesian coordinate system at this time. In keeping with the convention used by Small Vision Systems, the three coordinates of the space in front of the camera are labeled x, y and z. All the coordinates are defined with respect to the stereo camera systems optical systems. The z direction is the distance from the optical center of the left lens, along the optical axis, with z being positive in front of the imager. The z axis is also perpendicular to the line between the centers of the optical systems of the imagers. This line between the centers of the stereo pair of optical systems is the x axis, the origin is on the z axis, and positive values of x are to the right when looking along the z-axis. Finally the y axis is the line perpendicular to both the x and z axes, passing through the origin of the x and z axes. The origin of the y axis is at the same location as the origins of the x an z axis (the optical center of the left lens) and when the camera system is oriented “normally” with the x axis horizontal, then the y axis is vertical, and positive going up.

The z coordinate is essentially equivalent to the range, and an error in range can be given the value Δz. The minimum perpendicular feature size discussed above would be in the x-y plane, and its dimensions would be both Δx and Δy. This defines a volume element, or voxel, of dimensions, Δx, Δy and Δz. Thus the 1 pixel resolution of the imagers corresponds to a minimum volume element in the space being measured. The maximum size error possible would be along the diagonal of length $[\Delta x^2 + \Delta y^2 + \Delta z^2]^{1/2}$. We will call this the error voxel size and since each component is dependent on the range, the error voxel size is dependent on range.

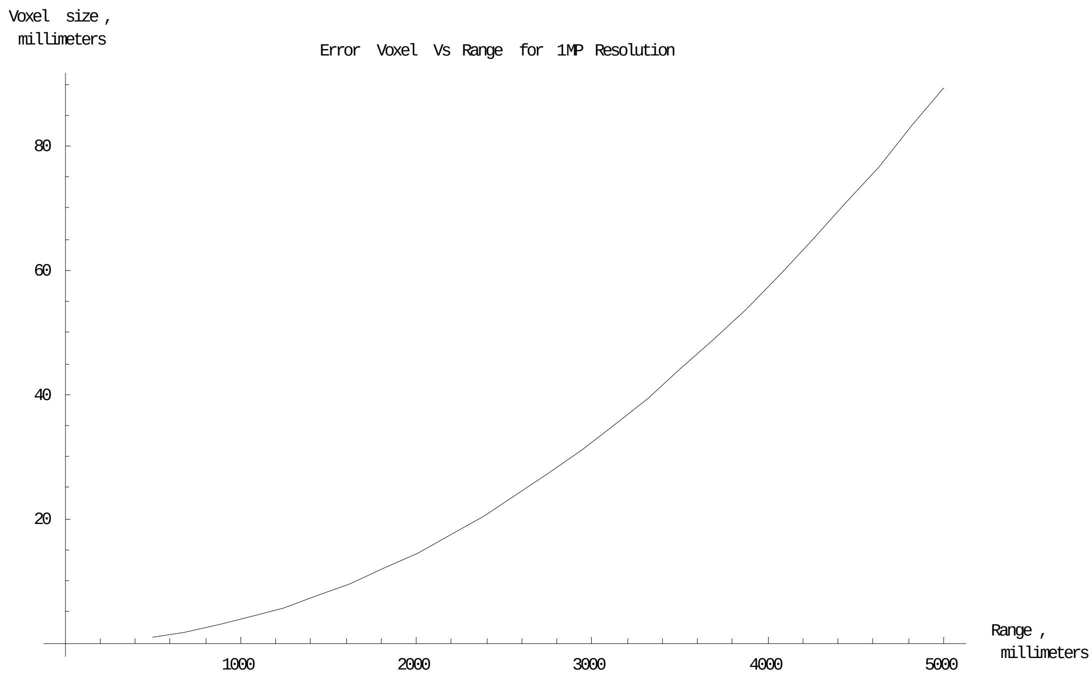


Figure **18**. - Error Voxel size versus range. for 1280x960 resolution

Three representative values of the error voxel size are 89.5mm at 5000mm range, 22.5mm at 2500mm range, and 1mm at 500mm range. Notice that these are not very different from the range resolution error. This is consistent with the order of magnitude difference in the two sources of error. The range resolution error is the dominant term in the error voxel size, and it is therefore the dominant source of error in size measurement.

The imagers can operate in different resolution modes. Another resolution we tried was half the resolution in each direction, 640x480 with one quarter the number of pixels. One can repeat the above error voxel size calculation for the VGA resolution mode easily. At the resolution of 640x480 each dimension of the pixel is double the dimension of the imagers at the 1280x960 resolution mode. The perpendicular feature size error is doubled in each direction, and close examination of the range resolution error formula reveals that term is also doubled. The end result is that the error voxel size is doubled at the 640x480 resolution mode of the imagers.

Measured Performance:

We performed a series of field experiments to determine the overall feasibility of using this stereo camera and software for fish surveys. The field experiments were designed to address 3 specific issues, reliability of the camera system, auto-exposure issues, and measurement accuracy. Indirectly the field experiments also addressed the usability of the system and other unexpected issues.

Range data recovery and image pattern matching:

Stereo imaging will always work for recovery of range and size information as long as the resolution of the imaging system and the baseline separation of the imagers are sufficient to produce a measurable disparity. As long as the identical point on an object can be identified in each of the pair of images and their coordinates within the image measured, the disparity of the point in the image pair can be easily calculated. This disparity along with knowledge of the effective focal length of the optical system will yield the range directly. It is not much more complicated to do this for two separate points on an object and then also recover the physical distance between those two points yielding the dimensions of interest. The real issue lies in how one correlates identical points on an object. One can perform the correlation manually or one can rely on computer algorithms to automatically match image points between a pair of images.

Example size measurement:

To illustrate the issues and solutions, let‘s use an example. The example stereo pair below should serve as a good example for an explanation of the stereo matching algorithm. The current SVS algorithm is based only on patterns of light and dark, or black and white.

Figure **19** - A sample stereo pair – the diver is holding an object of known size, a meter stick approximately 2 meters from the camera. Notice the large disparity between the two images, the right end of the meter stick is almost on the right of the background pillar in the left hand view of the pair, but in the right hand view of the pair, the right end of the meter stick is on the left side of the pillar.

Image Rectification:

To facilitate disparity calculations, both images need to have their individual optical distortions removed (lenses are not perfect pinholes, nor are they perfectly identical). The parameters derived from the calibration sequence allow the software to remove these distortions by rectifying the original images. Below is the same image pair as above, but with the optical distortions removed.


Figure **20** - Rectified image pair with optical distortions removed

Manual Pixel Pattern Matching method:

Manual correlation of corresponding pairs of points in each image will always work as long as the point is visible or identifiable in each of the images. It may be tedious, but a user can always manually recover the disparity and therefore the range of a human correlated pair of points from each image. This is essentially the method that previous investigators used. The SVS software is not designed to easily recover the range information for a pair of manually matched points but it still possible with some extra work. We did not pursue developing a software interface that would facilitate using the manual matching method.

Automatic Pixel Pattern Matching (APPM) method:

Our main interest was in the performance of the Small Vision Systems supplied computer program at pixel pattern matching and subsequent disparity and range recovery.

As discussed in the section on stereo image pair information, the keys to measuring the actual size of an object from an image is in knowing the distance to the object, the focal length of the lens, and the size of the image of that object. The focal length of the lens and the size of the image are fairly straightforward to know or measure. Measuring the distance to the object uses the stereo image pair information, or more specifically the disparity or image shift of the object in a pair of images. This is effectively triangulation. The image shift of an object viewed from two separate positions is referred to as the disparity. Making disparity measurements seems quite simple – match objects, or parts of an object, or more specifically points on an object in each of the stereo image pairs, measure the pixel disparity, then use the optical parameters to calculate the disparity between them. If the points on the objects are matched this is a triviality. The problem comes in matching points on an object. Naively this does not seem so difficult. The problem is constrained to one dimension and the disparity shifts of the images occur only along the direction between the two lenses or optical systems. Using the conventions from Small Vision Systems we pick this direction to be the x direction. The y direction is in the plane of the image and is perpendicular to the x direction. It would appear that a pixel pattern matching algorithm would only need to loop through a search range where it would effectively shift a small group of pixels from one image of the pair in the x direction and compare them to the same size group of pixels in the other image at the same y coordinate. When a match is found then the disparity is known. Unfortunately there are several complications which get in the way of complete success with this simple minded approach. To best understand some of the complications in the process we will continue to use our example stereo image pair.

After rectification, the software can then be asked to generate a disparity map where the software searches through the entire image to find the matching set of pixels in the corresponding image pair. We naturally see objects in images, the pixel pattern matching algorithm does not, it is only looking for matching patterns in a (small) group of pixels called a window (the size of which is adjustable). Fortunately, because of the optical geometry of a stereo system, the disparity only occurs in the one dimension along the line between the two imagers, so the pixel pattern match search is only along one dimension. (Close to the horizontal direction in all our example pairs.)

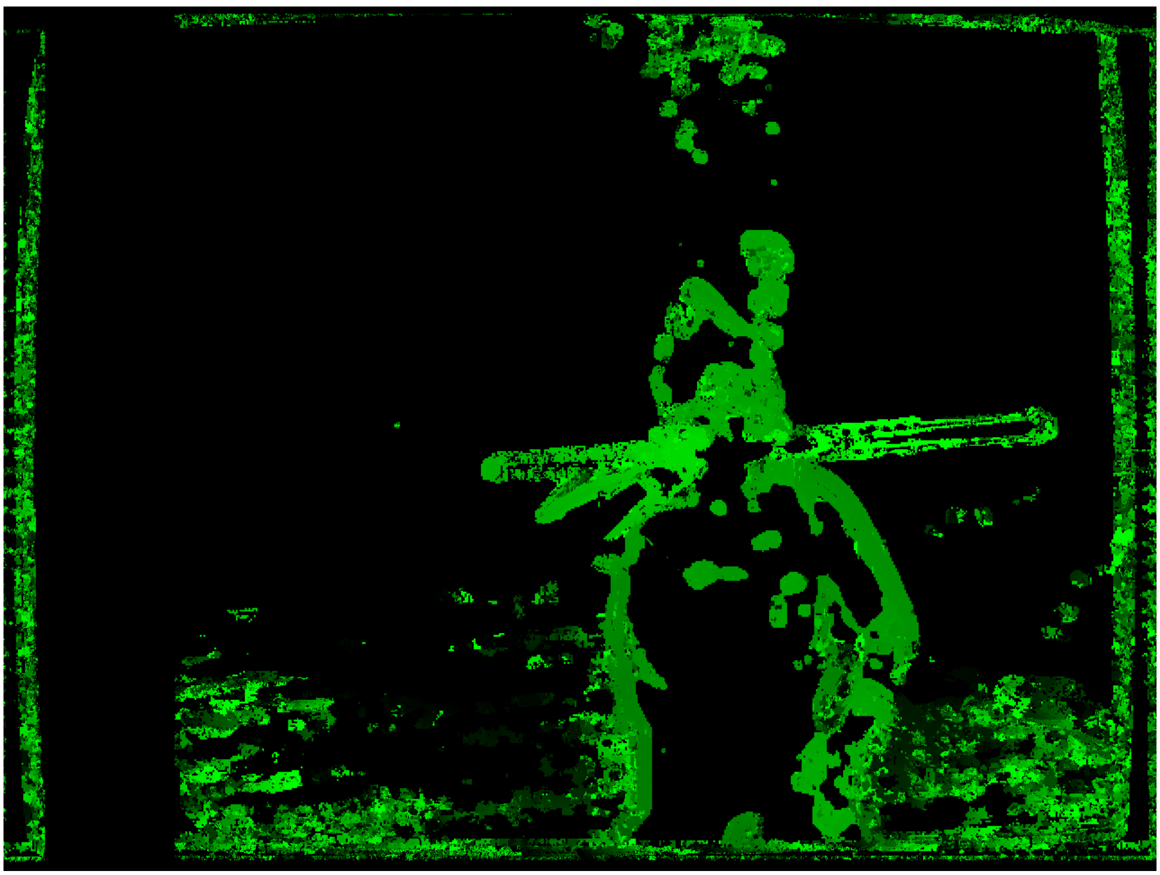


Figure **21** - Disparity map for the above example stereo image pair. Disparity can only be computed when there is a correct match between pixel patterns from both images of the pair .The brighter or higher intensity represents a closer point (or larger disparity)

Examination of the above disparity map reveals many features where the automatic image disparity detection program succeeds and fails. Where the image is black there was no successful automatic pixel pattern matching and therefore calculation of disparity cannot be done. For the featureless background this makes sense, but there are several locations that are black where one might have naively assumed the algorithm should have detected a disparity. The key to understanding where the algorithm will succeed in automatically matching the pixel patterns is realizing that the pixels must have a detectable distinctive pattern. Featureless objects will present a problem for pixel pattern matching. In general, hard, distinct or contrasty vertical edges are the features in the image that will experience the most reliable automatic disparity recovery. Thus the divers outline in the above figure is successfully detected, but the essentially featureless torso is not detected.

Another feature the above disparity map illustrates is that for successful disparity recovery the object or pixel patterns must be in both images. Both edges of the disparity map will display vertical areas where the disparity recovery has failed.

Horopter:

A more subtle cause of accurate disparity recovery failure is related to the limited search range of the algorithm. We feel this is the most dangerous form of error in the disparity recovery algorithm because it often results in an incorrect or false disparity recovery instead of no disparity recovery. The algorithm that matches pixel patterns has a limited search range. This search range can be modified in multiples of 2 from 16 to 128 pixel disparities. Furthermore, the disparity search range can be offset with the Horopter offset control. The disparity search range is referred to as the horopter. The limited disparity search range is equivalent to a limited depth or range recovery distance. Only object features at distances that produce a disparity within the disparity search range will produce correct range (and therefore size) results. The default configuration is to search from 0 disparity (infinity) through to some maximum disparity, such as 128 pixels which corresponds to some distance away from the front of the camera, called the near point. Only object features between the near point and infinity will have a chance of being successfully ranged. For our camera systems combination of imagers, lenses and baseline, at 1280x960 resolution the 128 pixel disparity point is close to 2200mm distant. With no further adjustment of the horopter, the software will impose a search range from 2.2meters to infinity, and it will only successfully range object features in that range. Realizing that seeing object features at large ranges underwater is very difficult, we can adjust the horopter so that the far point of the algorithm is at a closer distance. Our interest is in the distance range closer than 5 meters, which corresponds to a pixel disparity of nearly 56 for our camera system at 1280x960 resolution. This will bring the near range in closer as well; the pixel disparity range that will be searched is 56 to 128+56=184. A pixel disparity of 184 in our camera system at 1280x960 corresponds to a range of 1525mm. Therefore at a horopter offset of -56 all object features within the range from 5000mm to 1525mm could be ranged. The horopter offset control allows the user to adjust this range, and if the user wants to range an object that is closer than this near point at 1525mm, they can increase the horopter offset. Likewise, if the user wants to range an object further than the 5000mm far point they can decrease the horopter offset. The consequence of this information about the horopter is to realize that there is a limited search range for the disparity algorithm and therefore object features closer than the near point or farther than the far point will not be ranged correctly.

In the disparity map image above, notice the regions on the bottom on either side of the diver. The disparity map in these areas appears noisy, with small bright and dark areas adjacent to each other in an apparently random mottling pattern. It is very important for the user to look at the disparity map in the region of interest and ask themselves if the pattern make sense. On the bottom in this image one should see an essentially smooth gradation of disparity or distance, with increasing range as one goes up. (The “floor” is closer in the bottom of the image and recedes as one moves up in the image)

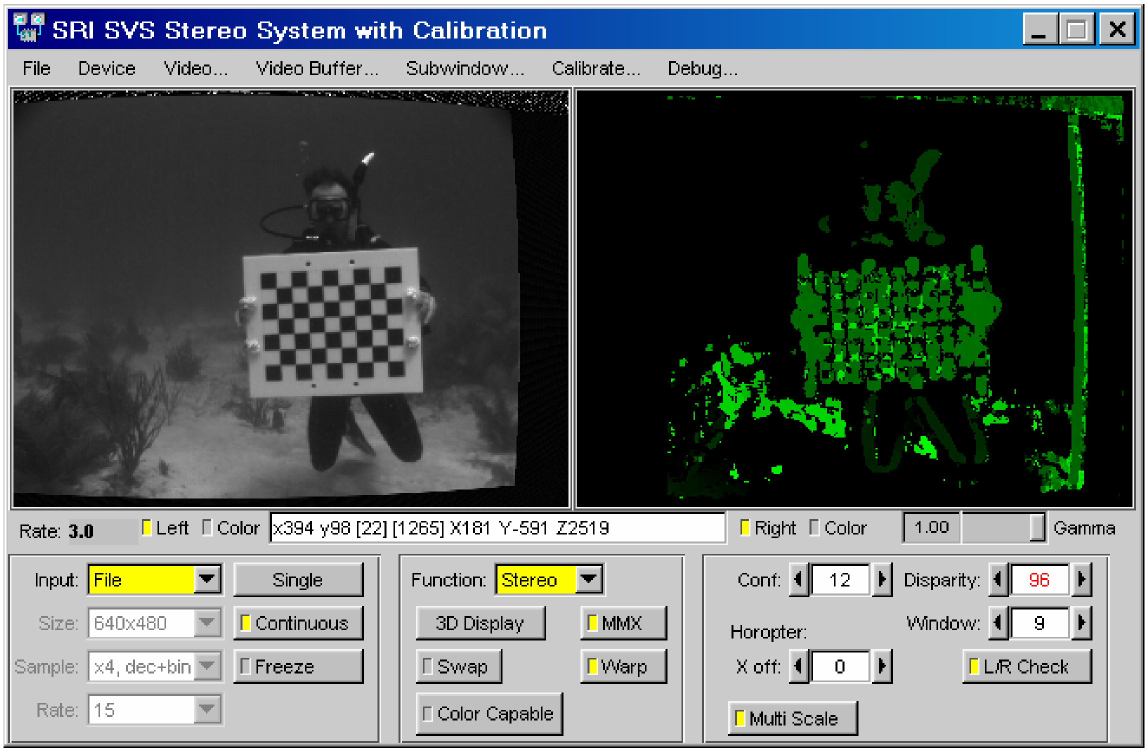


Figure **22**. A screen capture of the Small Video Systems stereo image pair analysis program. This is SVS version 2.3 running on a Microsoft Windows platform. The program has automatically analyzed a stereo image pair and produced a disparity map.

In the above figure another stereo image pair has been analyzed. In particular notice the bright regions in the disparity map for the area of the calibration target. These are regions where the software has detected a pixel pattern match, but has made an error. This is typically because the value of the horopter offset has not been considered correctly.

Focus, Depth of Field, and aperture setting:

To achieve the highest number of disparity recoveries, the pixel matching algorithm needs an image rich in clear sharp features. The small scale detail of the image needs to be present for the matching algorithm to be successful. In order that the image pairs have this “maximum” detail, the imaging system needs to provide a good exposure with maximum contrast, the images need to be well focused, and if the objects are moving, the exposure time must be sufficiently short to “freeze” the objects and prevent blur.

Our system currently uses fixed focus lenses so we need to close the aperture on the lens (reducing the light falling on the sensors) to achieve sufficient depth of field to have the image remain sharp enough over the range of interest from 1m to 5meters. Achieving the optimum focus setting of the lenses over this range underwater was an iterative process.

Automatic Disparity/Range Software Performance

On several dives we carried an object of known size (a meter stick) and shot video of the object at approximately 1, 2, 3, 4 and 5 meter distances. The following are the results from a dive on 09 Feb 2005. The camera was in 1280x960 resolution mode.

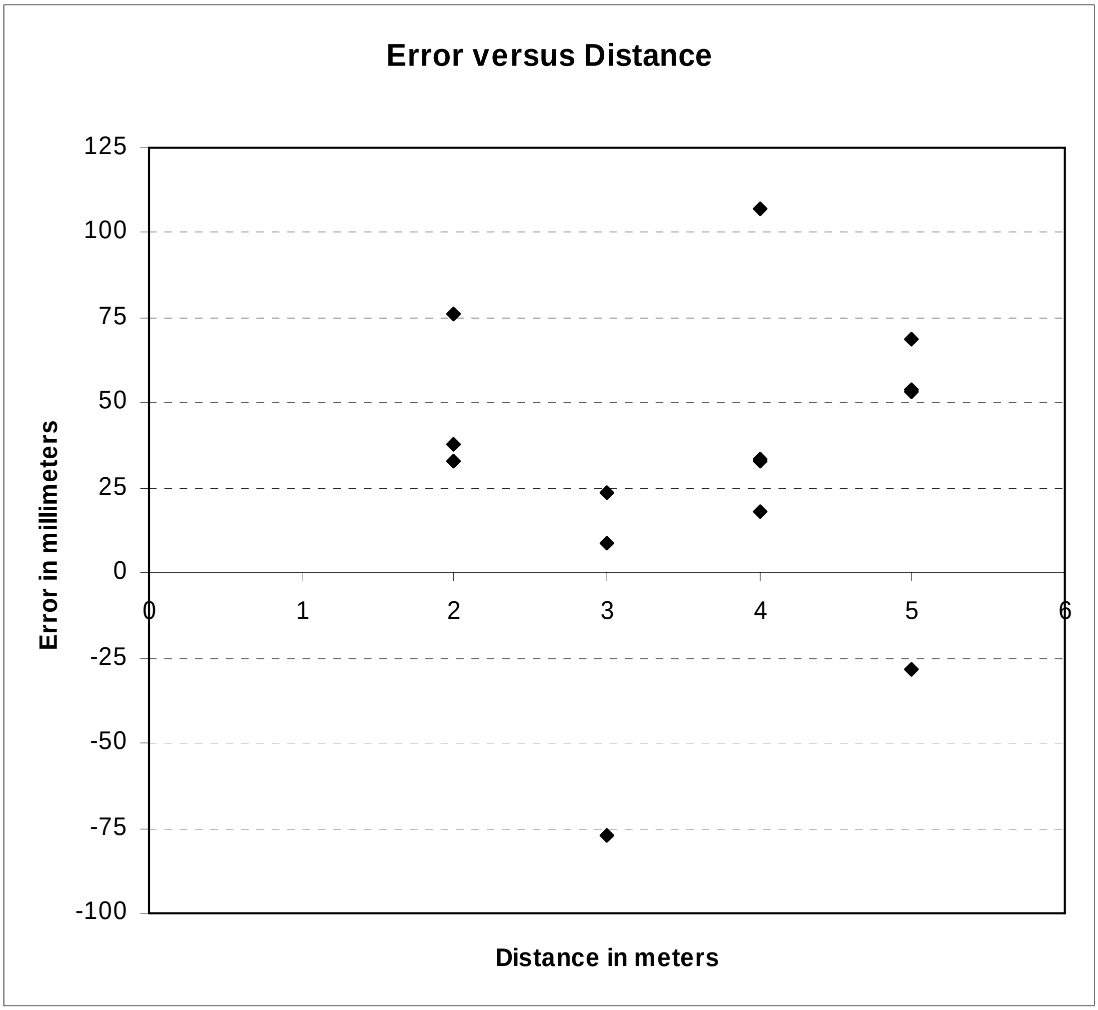


Figure **23**. - Error in measurement of length of 1 meter object at multiple distances. Error is actual length minus measured length.

The errors are predominately short; the stereo pair derived measurements are more commonly less than the actual length. This may make some sense if one considers how the automatic pixel pattern matching algorithm works. Due to uniqueness (discussed below), the exact edges of an object may not be ranged and when the user clicks on the points closest to an edge they may not receive a range until they are well inside an edge, leading to shorter length measurements. The percent error can be as large as 10.7%, but the average error is 3.16%, and the average absolute error is 4.66%.

Uniqueness:

An issue of automatic pixel pattern matching in stereo pairs can be described under the label of uniqueness. Disparity, or the image shift due to the different viewpoints of the pair of imagers is the information that stereo image pairs provide that enable one to calculate range and therefore size. However, this disparity between objects in an image pair due to the different or unique viewpoints can also be a problem for automatic pixel pattern matching in two ways. These are issues that our own visual processing system deals with automatically that we are not typically aware of.

One problem occurs when the background is not featureless and is at a different distance than the object of interest. In this fairly normal situation, the foreground object of interest obscures different parts of the background pattern. More significantly but equivalent, in the image areas just adjacent to the foreground object the unique viewpoints of each imager reveals different parts of the background not seen in the other image of the pair. These unique areas of the different images will occur near right and left edges of objects (assuming right and left orientation of the stereo imagers). This means that the pixel matching algorithm cannot find a matching pattern in these areas of the images. However these are very close to the areas where the pixel pattern match is needed. If the foreground object of interest has sufficient image features, then the automatic pixel pattern matching technique will be able to match the pixel patterns in the areas of the image on the object, but if the object does not have much detail, then the pixel pattern matching technique will attempt to match on a window of pixels around the edge of the object that will include these unique patterns. Thus these unique areas can be a problem for the automatic pixel pattern matching technique.

The other problem related to the unique view of the object by each of the imagers, is that the object itself may appear slightly different from each view point. An excellent example of this occurs in the images of the Goliath Grouper further down in the section on color. In this pair of images one can see that each view of the Goliath Grouper is seen from a unique perspective. Again, near the right and left edges of the object of interest, but now on the object itself instead of in the background, there are unique sections of the object that each imager sees and the other does not.. As in the first example of uniqueness in the background areas, these unique areas of the image cannot be matched with the corresponding area in the other image of the stereo pair. Again this can be a problem if one wants to size an object by measuring it from edge to edge. To give this a distinctive name from the previous or “obscuration” uniqueness, we would call this type “Cubist” uniqueness to associate Picasso’s ideas behind his Cubist style with this form of uniqueness.

Finally, we approached this area of the project with overly high expectations and as a result the performance of the automatic pixel matching and range recovery software was initially our greatest disappointment. Our subsequent analysis of its successes and failures has led to a greater understanding of what the SVS supplied software is doing and what it is or isn’t capable of. We have also seen substantial improvement in the stability and performance of the software in calibration and automatic disparity recovery but, in our experience, it is still not (current version 4.2e) stable enough for production work in either the Windows or the Linux environment. We have found it is generally more stable in the Linux environment.

Auto Exposure

When we purchased the camera sensors, the documentation stated that the camera control and acquisition software could provide an Auto-Exposure function. Testing revealed this function did not work at all, and subsequent conversations with Videre Systems revealed that this function would not work correctly with these B&W Mega-D sensors due to some unforeseen issue with their hardware supplier. When we started the field testing stage of the project we proposed a few experiments to determine what the best exposure settings was for image pair data reduction to range data so that we might be able to develop our own method of exposure control. We implemented an exposure ramp program in the software to continuously ramp the exposure values in a cyclic pattern to gain information on which exposures provided the most effective range recovery. We ran several experiments where we set the camera up on a tripod underwater and ran this exposure ramp program (called ramp-o-matic). We then subjected the data to automatic image pair data analysis for range recovery. Unfortunately the non-constant scene due to uncontrolled scenery and the natural variability of the lighting underwater rendered these experiments almost completely useless. At the extreme ends of the exposure ranges, highly underexposed and highly overexposed, the automatic range data reduction from the stereo image pairs fails completely. The best exposures are where the objects of interest are rendered with the greatest contrast or difference in brightness, also referred to as with the greatest dynamic range. Unfortunately this statement is too simplistic a formulation since one also wants to prevent areas of saturation (full intensity) or complete blackness (zero intensity). Where the pixels are either maximally or minimally exposed is essentially featureless as far as the pixel pattern matching algorithm is concerned, and furthermore no amount of digital image processing will be able to reveal any detail or features in these regions.

Turbidity, Contrast and Contrast enhancement:

Turbidity is a not uncommon condition in underwater environments. Turbidity reduces contrast, and in some cases severely reduces the diver's ability to visually survey a site. We hoped that a few of our dives might provide some data on the ability of the Stereofish camera system to overcome a turbid environment and the reduced contrast. It is obvious, but should be stated none the less, that no image/pixel matching system is going to work on images where there are no features. If a human cannot make out features in the image, processed or not, the pixel matching algorithm is not going to work. On 06 December 2004 we inadvertently set up a test at such a site. The bottom had very fine sediment that was easily re-suspended into the water column. Within minutes of setting up the equipment we noticed that we had turned our test site into a turbid water test site.

Figure **24** - Turbid environment underwater. Can digital image processing techniques improve image quality sufficiently to allow automatic disparity matching?

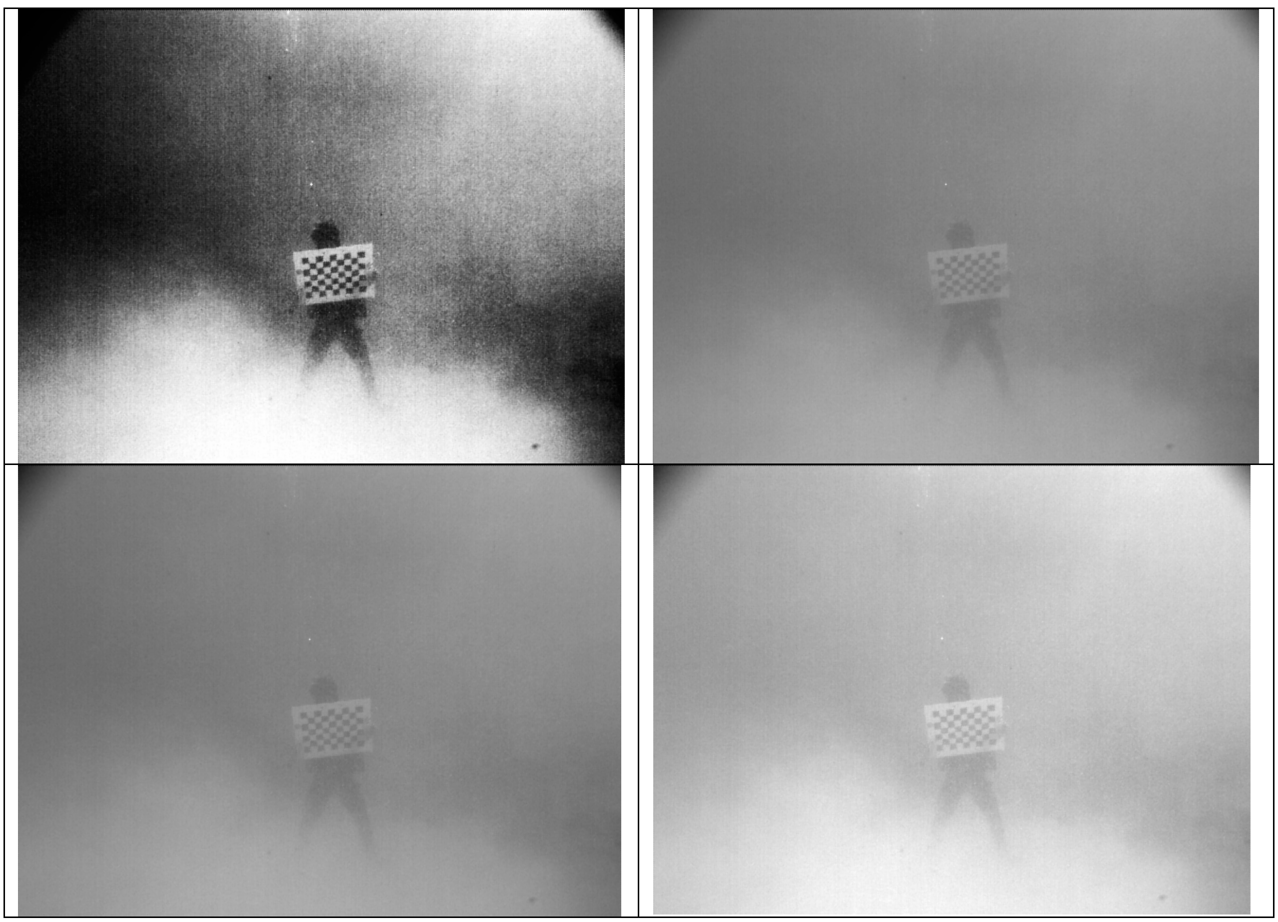


Figure **25** - Automatic image enhancement techniques applied to the original left image in the previous figure. Top left: auto equalization. Top right: auto normalization. Bottom left: auto contrast stretch, a. Bottom right: auto contrast stretch, b

Of the four automatic methods, three; auto-equalization, auto-normalization, and auto-contrast-stretch-a; are available as automatic methods within the popular open source image processing program, GIMP (GNU Image Manipulation Program), available for both Linux and Microsoft Windows environments. The fourth method, auto-contrast stretch-b is one of the authors own methods implemented in the IDL programming language. The idea is to maximize all the available information in the image and minimize the loss of any information already in the image. This is done by examining the histogram, finding any gaps either at the bottom or top of the intensity range, and then re-mapping the exposed values into the full range of available intensity values. Another constraint is to do this linearly, or with gamma=1. This prevents the situation produced by the auto-equalization method (top left) where the contrast is stretched by mapping the middle range of old intensities into a larger number of new intensities (range expansion) thus requiring the old lower and higher range intensity values to be mapped into fewer new intensity values (range compression).The result is very contrasty but also unnatural. Auto-contrast-stretch-b can be simulated manually in the GIMP through the levels control leaving gamma=1. The performance of the auto-contrast stretch-b method was superior to the other three image contrast enhancements. It was the only method that provided feature recovery by the calibration routine and also allowed reliable range and size data recovery.

Auto contrast-stretch-b:

The SVS automatic pixel matching, disparity calculating, range extraction program can discern the target plaque (but no features of the diver are matched) and the observer is able to find the coordinates of two holes on the top edge:
Right hole: (343,247) [v205] [dv196] X179 Y-224 Z5066 u(339,241)
Left hole: (320,250) [v202] [dv201] X13 Y-200 Z5015 u(316,244),
This indicates that the holes are about 5000mm distant and, leads to a size (distance between the holes) of 175.3mm. The actual distance between the features is 165 mm. Given that these images are shot at the lower resolution of 640x480 I consider this to be remarkably good accuracy of 1cm at 5meters

Auto-equalize:

Surprisingly this is not the best solution – the algorithm finds matches in the noise that have been caused by the extreme nature of the processed histogram. The next image shows the resulting disparity map for the extreme contrast enhancement method auto-equalize.

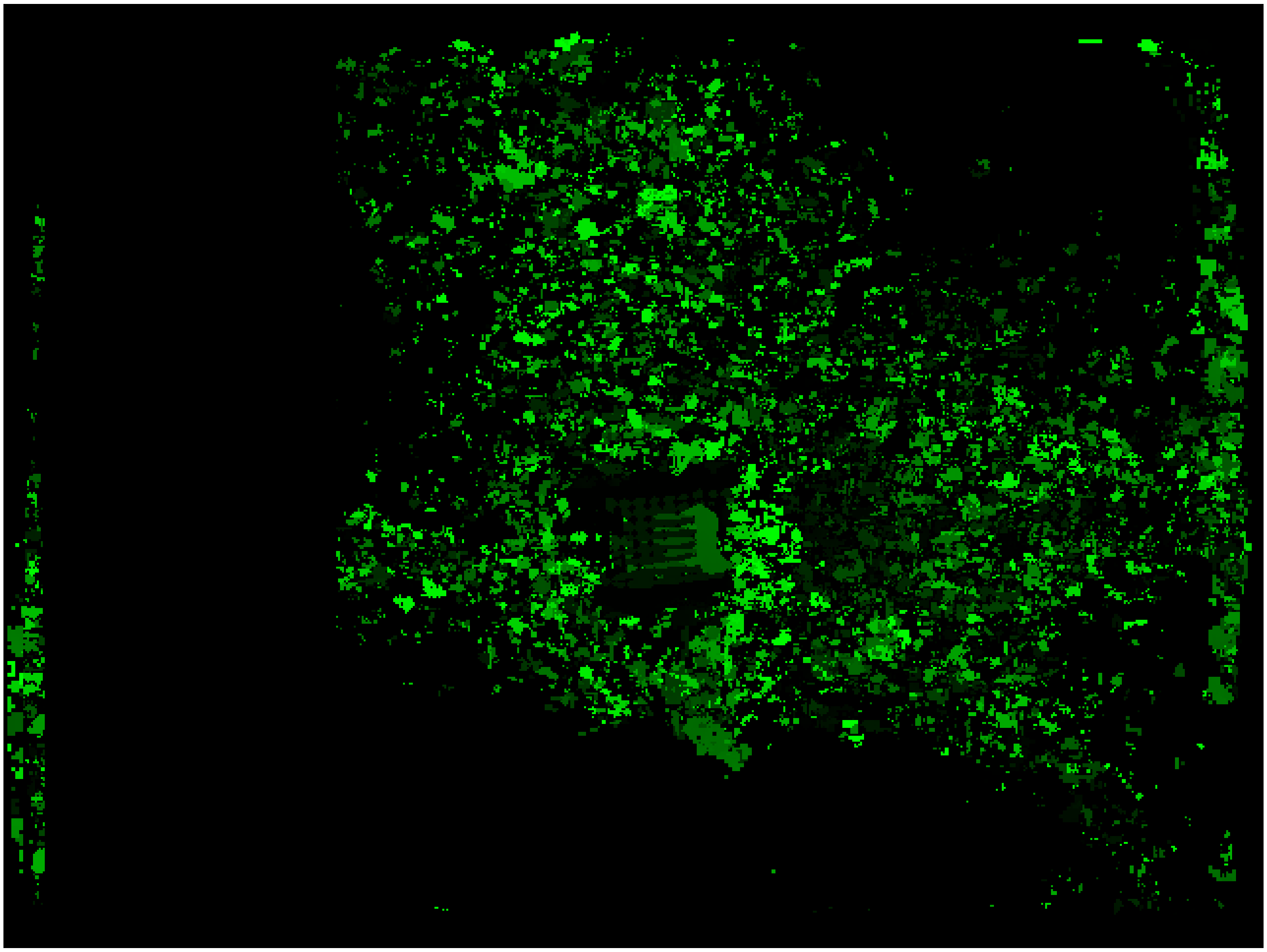


Figure **26** - Extreme contrast enhancement, auto-equalization method, results in many false matches with patterns generated by noise.

Auto-contrast stretch-a and Auto normalization:

Both of these contrast improving techniques allow for adequate range and size recovery. There is not a substantial difference between their performance on this measure and that of the auto-contrast stretch-b method. However, in a second test of the various automatic image enhancement techniques, processed images from a closer distance were submitted to the calibration routine to see if the routine could pick out the features on the calibration target. The calibration routine could only find all the features in the auto-contrast stretch-b pair of processed images. These two methods failed this test.

Example with fish:

Figure **27** - Left image of stereo pair, 1280x960 resolution. Note Black Grouper just below and to the right of center. Disparity with right image reveals Black Grouper is 2.8meters distant, and 55cm in length from tip of head to end of tail. Date:09Feb2005, Location: Benwood Wreck, Key Largo, Florida.

Figure **28** - Black Grouper - Left and Right images of stereo pair after rectification.

SVS software version 4.1f was used on the above stereo image pair to measure the x,y,z coordinates of two points on the Grouper. The coordinates of the Grouper's tip of the head (X=-190, Y=143, Z=2997) and tip of the tail (X=335, Y=118, Z=2832) in millimeters were used to calculate its length of 550.8mm, This is not inconsistent with a visual estimate of its length by the diver. Notice that the Grouper's orientation is not perpendicular to the line between it and the camera. The tail is closer to the camera than the head by 165mm so the effects of projection make it appear shortened. Calculations based on the coordinates show that the angle between the line from the camera to the grouper and the line from the tip of the head to the tip of the tail is 17.4 degrees.

Reliability:

As with any new prototype system we had some initial issues which affected the reliability of the system. Most were minor issues, but one issue in particular, loss of some critical part of the camera operating system and programs due to some form of hard drive corruption remained an intermittent problem until we eventually replaced the hard drive and upgraded to a newer version of the underlying Operating System. Most of the initial problems with the reliability of the system were traced to brief power interruptions which the hard disk and file system of the internal control and acquisition computer could not tolerate. We had initially chosen power cables and connectors that would be more convenient to use in the marine environment. Unfortunately these proved sensitive to the constant flexing at the connector when the camera was in use. We then modified the cables and connectors to a more rigid and conservative design which eliminated the brief power fluctuations. However occasional low battery power or operator error in removing power before completely shutting down the control computer would still result in major disk errors requiring long file system scans and/or re-imaging the onboard drive with the backup drive, neither of which were easily accomplished in the field. We had started with RedHat Linux 7.0, and then upgraded to RedHat Linux 9.0 as the Operating System for the control and acquisition computer. RedHat 9.0 introduced a journaling file system which allowed the system to recover gracefully from unexpected power interruptions. With this Operating System upgrade the camera system has been very reliable; with a total of over twenty mostly problem free dives. (There was a small problem on one dive while using a new experimental feature, but this was cured with a reboot and has subsequently been fixed. And on one recent dive in Feb2006 the data acquisition amount was so large, just over 2 hours of constant video resulting in over 60GB of data stored on the hard drive in the camera that the camera ran out of free disk space without the operator noticing the problem and it proceeded to get caught in an infinite reboot loop at the next data acquisition site.)

Color:

Our initial experiments with black and white (gray scale) images demonstrated that the method worked, but that B&W imagery was not sufficient for fish identification. Although the color sensor work was not part of original contract specification and we obtained external non NOAA/NMFS funding for the color sensor work we felt it would be appropriate to add some of our observations about using a color system. The biggest advantage of the color system for us was that it has a working auto-exposure system. The color sensors came with the ability to automatically regulate the gain and/or the exposure time based on the previously acquired image. This greatly reduced the number of poor exposures and the need for the diver to constantly monitor and adjust the exposure values. Furthermore, we feel the reliability and effectiveness of automatic target recognition, ranging and sizing algorithms will be greatly increased if color information is included in the methods.

Color complications in automatic pixel pattern matching: Although we consider color absolutely necessary for the purposes of this instrument, which necessarily includes fish identification, color imaging underwater introduces some complications. The effective sensitivity of the imaging array is reduced due to the color filters placed over each pixel. Compensating for this reduced sensitivity requires some combination of higher gain, longer exposure time or increased aperture. As discussed before, in a fixed focus system, the aperture cannot be increased without loss of depth of field, reducing the working focal range of the camera. Out of focus objects will present difficulties for the automatic pixel pattern matching algorithm. Increases in exposure times are also a problem for getting images of moving objects without blurring. Motion blurring will also prevent the automatic pixel pattern matching algorithm from succeeding. Increases in gain can be tolerated a bit more, but increased gain also introduces increased pixel level noise and the image acquires random brightness fluctuations that prevent automatic pixel pattern matching in stereo pairs.

Anyone who has taken a picture underwater is aware that the spectrum of natural light available is skewed towards the blue. Water preferentially absorbs more towards the red part of the spectrum than the blue. The natural “white” light of the solar source becomes progressively bluer as the light goes farther in the water because the red light is being progressively removed. This is well understood but it has some complications for color photography and for getting a correctly exposed image. If the imagers exposure system weights all pixels and colors (red, blue and green) equally in the exposure calculation and then determines the best exposure based on the average response of all the pixels, the exposure is going to show some problems in the image. The red pixels will be too dark, but their presence in the exposure calculation will cause the exposure algorithm to increase the gain or exposure time for all pixels. This will in turn result in the green and blue pixels becoming slightly overexposed. There are 2 solutions to this problem – using a reddish filter to attenuate some of the green and blue light, and using a imaging sensor where the gain can be controlled for the individual colors.

Goliath Grouper

Figure **29** - Stereo image pair of a Goliath Grouper nearly head on towards camera, 1280x960 resolution, 02Sep2005, Grecian Reef, Key Largo Florida

We were fortunate enough to encounter this Goliath Grouper on the first day that we had the color sensor equipped camera out for field testing. There were some problems with the automatic exposure algorithm resulting in image blur which have since been fixed. We were particularly fortunate that the camera was in movie mode, taking shots at about 5 frames a second, and as the diver swam up to this Goliath Grouper, it was initially facing them, and then it slowly turned and was nearly perpendicular to the camera allowing us to get stereo image pairs of both a nearly head on orientation and a nearly perpendicular orientation.

These images required some gentle image enhancement to bring out the detail without generating too much noise. After trying several automatic image enhancement algorithms available in image processing software we decided to develop our own. We have developed our own color balancing and contrast enhancing algorithm to maximize the image detail without substantially increasing noise. The idea of this image enhancement method is discussed in the section on image enhancement above, but instead of being simply applied to the black and white image values, the image is broken into its three color channels, and each color channel is optimized for contrast using the same technique as discussed for the black and white images.

Figure **30** - Left image of pair before contrast enhancement.

Figure **31** - Left image after color balancing and gentle contrast enhancement.

Using the SVS software to extract the coordinate data of the fish's tail and tip of the head to measure the length of the fish:

Nose: (610,500) [v36] [dv2016] X-254 Y-82 Z2386 u(621,507)
Tail (714,396) [v31] [dv1249] X-179 Y-272 Z3197 u(724,402)

And computing the length of the 3D difference vector leads to a measurement of 652.3mm. This measurement was difficult to obtain since in this orientation the location of the tail of the fish is difficult to visually pinpoint and the Automatic pixel pattern matching algorithm has difficulty in that region.

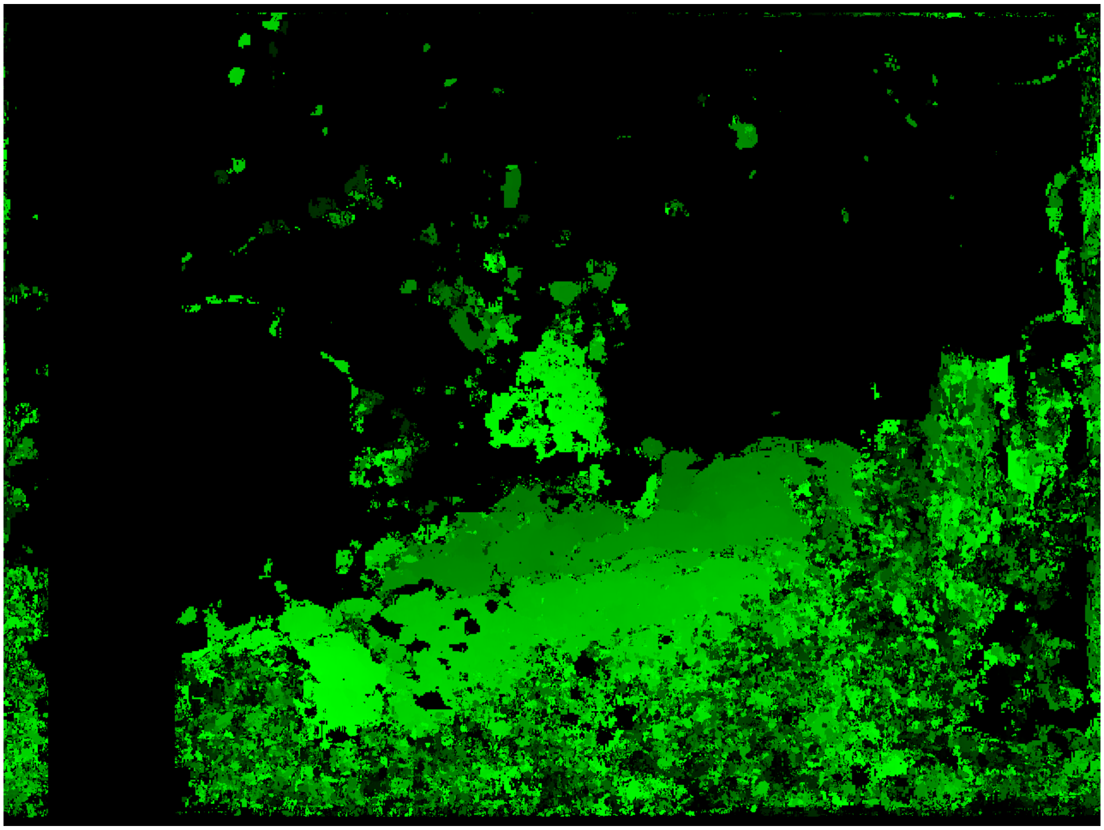


Figure **32** - Goliath Grouper disparity map, nearly head on. Notice foreground is inside the near point of the horopter.

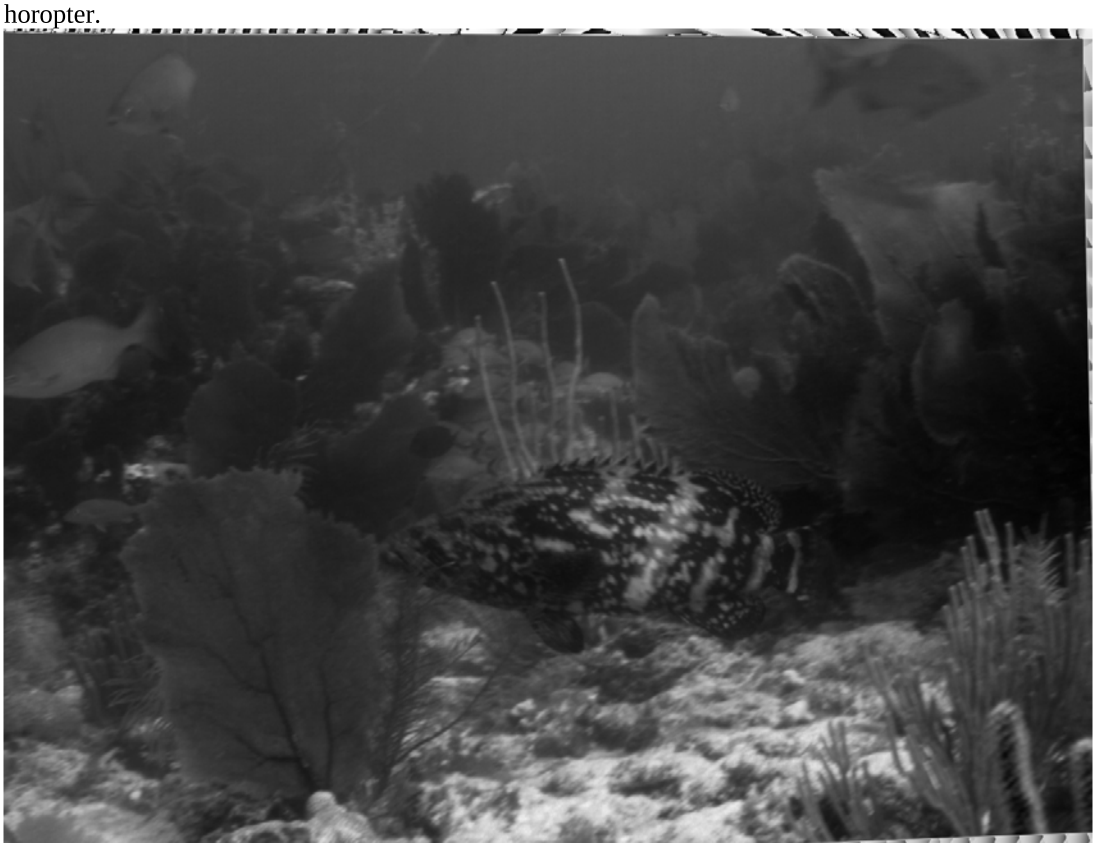


Figure **33** - Goliath Grouper seen nearly perpendicular to the camera view - left image.

Measurements from the associated stereo pair using SVS 4.1f reveal the coordinates of the:

Nose:(446,624) [v28] [dv1436] X-513 Y71 Z2754 u(462,630)
Tail: (944,584) [v16] [dv1372] X157 Y17 Z2823 u(959,593)

Computing the length of the 3D difference vector leads to a measurement of: 675.7mm.

These above two measurements are fairly consistent given the less than ideal orientation in the nearly head on sample. This essentially confirms the ability of the software to extract reliable length data over a large variety of orientations. Three other measurements at other orientations yielded 695.4mm, 715.4mm and 639mm.

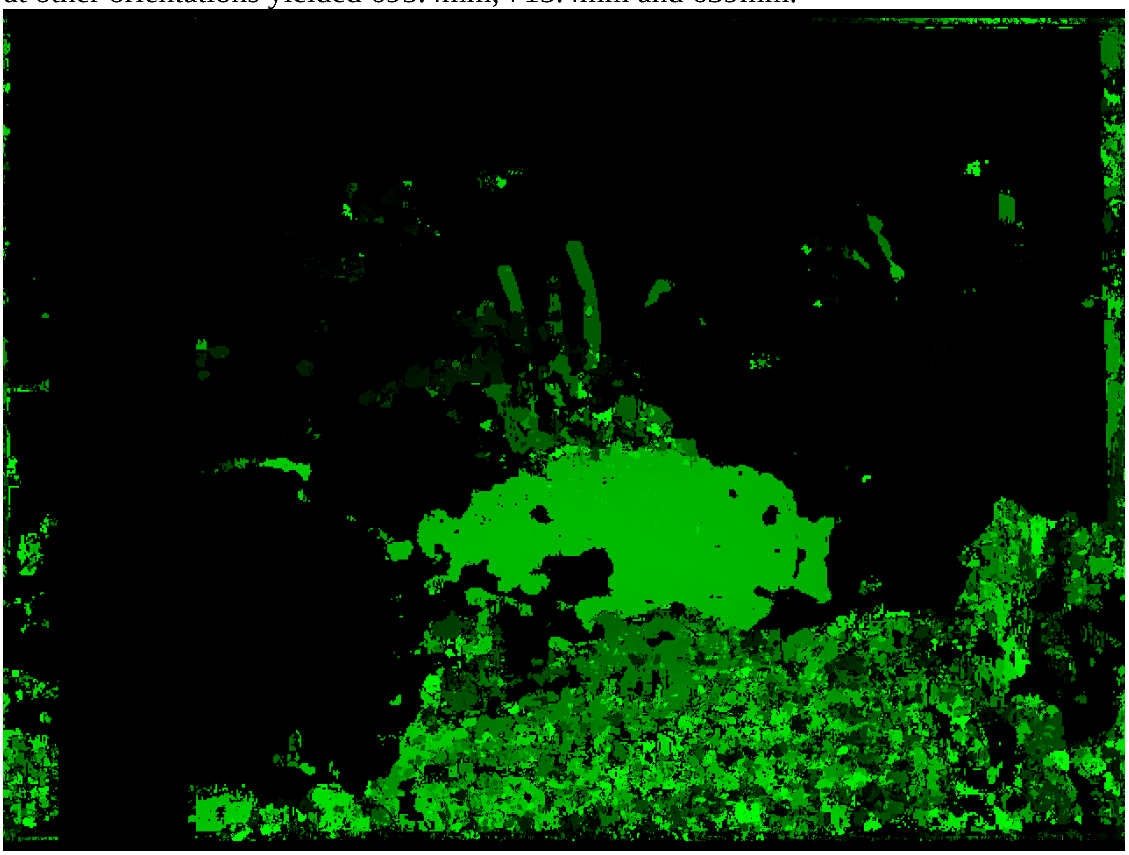


Figure **34** - Disparity map for Goliath Grouper perpendicular to camera view.

Miscellaneous issues: Image compression and associated artifacts.

We have been told that effective automatic object recognition requires a maximum of information in the images. Lossy compression (DV, JPEG, Gif, etc) all remove fine-scale information that is needed for image object recognition using texture.

Conclusions:

We initially chose a computer controlled camera because we wanted to guarantee synchronized simultaneous images pairs for the most effective stereo pair data analysis. This and our experience with digital imaging systems encouraged us to think of a completely digital design. We were convinced that the ability to download the already digitized images off the camera directly into a data analysis computer was a major advantage over the alternative method of using two separate cameras to record onto two video tapes, and then separately digitize the tapes and extract the data from post synchronized images. Not only did we felt that this digital environment would cut out the expense of separate video analog to digital conversion equipment, but it would also allow for more flexibility and control in the acquisition process.

Given our experience, the all digital method has many advantages, but there are some caveats. High resolution (1.2Mpixel) uncompressed images at 7 frames per second produces massive amounts of data. Stereo imaging generates two times that already large amount of data. Each five minute (300 seconds) segment of high resolution stereo video at 7 frames a second produces 300x7x2x1.2MB=5040MB or about 5 GB of data or about 1GB of data per minute. This amount of data becomes difficult and time consuming to process, transport, store, analyze, etc. With the color system we were only able to improve the frame rate and storage space requirements by modifying the standard software so that we would only store the raw images from the camera without any processing. When we circumvented the normal SVS color image acquisition and storage routines we were able to improve the maximum frame rate of the camera and reduce the storage needed on the camera This technique allowed us to significantly reduce the storage needed on the camera system by a factor of 4 (raw=2x1.2MB, processed=2x1.2 plus 2x3.6) greatly reducing the time that the computer is processing and writing frames onto the hard disk.

The initial B&W sensors were not sufficient for practical surveys, but the subsequent color imagers provided sufficient overall improvement to the system to provide some hope that this system can become a practical device for fish surveys.

Other Applications:

We have experimented extensively with using the camera for coral reef mapping and monitoring. We have mapped very large, up to 30m x10m areas using mosaicing and also done some 3D “swim arounds” of smaller corals to provide 3D models for accurate volume and feature measurements. We have taken some data as a test of the ability to use the data to measure rugosity of the benthos.

Future Work:

If asked what we would change about the hardware system we think that the fixed iris and fixed focus of the optical system introduces some difficult limitations under natural lighting. However, a variable focus would introduce a new variable in the calibration procedure. The software would need to be aware of the focus setting of the cameras and then only use the calibration settings for that particular focal setting. If one could acquire this kind of software and hardware, we believe that the issue with an automatic iris would not be a problem. It is unlikely to introduce major distortions as it changes for the lighting conditions. Another issue is that of the electronic shutter in the current imagers – we have found that a global shutter would be preferred since it would be less susceptible to image motion blur but at this time Videre Designs does not provide a high resolution (1MP) global shutter based system. Finally faster data transfer would be a major plus. Either providing an external USB 2.0 port or upgrading the network connection to 1Gbps would greatly reduce the data transfer time. An additional "nice to have" feature might be more compact batteries that are commercial airline friendly.

Reports and publications:

Voss, Kenneth J., Boynton, G Chris, Ault, Jerry, Bohnsack Jim, , Poster presentation: "An in water Digital Stereo Video system for Fish Sizing", CIMAS 25th Anniversary, RSMAS, University of Miami, Miami, Florida, 20Feb2003
Boynton, G. Chris, Voss, Kenneth J., "Development of a diver-held, stereo video camera", National Marine Fisheries Service Advanced Technology Workshop on Underwater Video Analysis, August 4-6, 2004, Alaska Fisheries Science Center, Seattle WA.

References:

Harvey, E. and M. Shortis, 1996, A system for stereo-video measurement of subtidal organisms, Mar. Tech. Soc. J. 10-22.

Van Rooij, J. and J. Videler, 1996, A simple field method for stereo-photographic length measurements of free swimming fish: merits and constraints, J. Exp. Mar. Biol. Ecol., 15, 237 – 249.

Woods, A., T. Docherty, and R. Koch, 1994, Experiences of using stereoscopic video with an underwater remotely operated vehicle, Underwater Intervention, San Diego, Ca, February.

Konolige, K. and Beymer, D, "SRI Small Vision System Users Manual Software version 4.1e", September 2005

Acknowledgments:

Funding was provided in part by the NOAA Fisheries Coral Reef Program (Grant #NA17RJ1226 SB33) and NOAA/NMFS Contract #GE133F04SE1220:SEFSC Underwater Stereo-Video Camera Research.

Thanks to Al Chapin for building the device and helping to work through some of the initial debugging, and being a tireless diving companion.
Thanks to Marco Monti for being a software engineer of incomparable value and expertise for developing and perfecting the camera system control program.
Thanks to Art Gleason for helping to provide field experiment logistics in the way of dive equipment, official paperwork, boat time, and dive time.

Appendix A: Camera Operation Manual with behavior, from power-on through shutdown:

Charge batteries

Check for O-ring seal integrity via positive pressure at Schrader valve.

Connect Power cable to top left connector, and snap cable strain relief in place.

Keep it as cool as possible while transporting – out of the sun. Also on a planing boat, place the camera towards the rear of the boat on some form of cushioning to reduce shock from a rough or "pounding" ride

When entering the water with the camera, we first hand the battery pack to the diver in the water and allow them to clip it onto D-rings on their BC, and then we hand the camera to the diver. Exiting the water is the reverse of the entry procedure. The length of the power cable is to allow this two part entry and exit procedure. .Also before descending the diver should look for any signs of a leak in the form of a persistent source of bubbles.

I power the camera on in the water, this is a little slower since there is the delay of having to watch the camera/computer boot up, but it reduces the possibility of it falling or having the power interrupted while on, and it allows it to cool off in the water before operation.

The following are the details of the power on behavior:

Figure **35** - At power on (switch on power at battery) the initial screen behavior is a banner display from the LCD screen indicating it is alive and well.

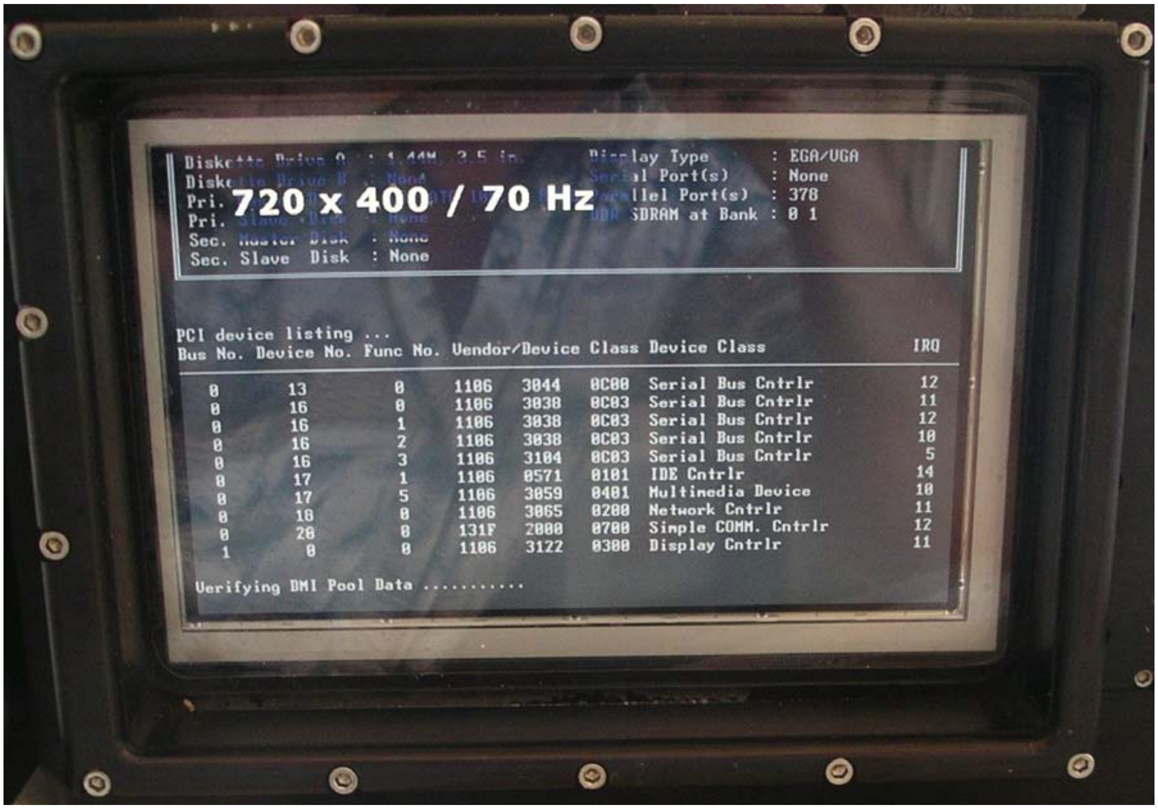


Figure **36** - The second screen is the computer BIOS screen indicating the status and activity of the computer and its peripherals

Figure **37** - The third screen is the Linux Boot manager, GRUB.

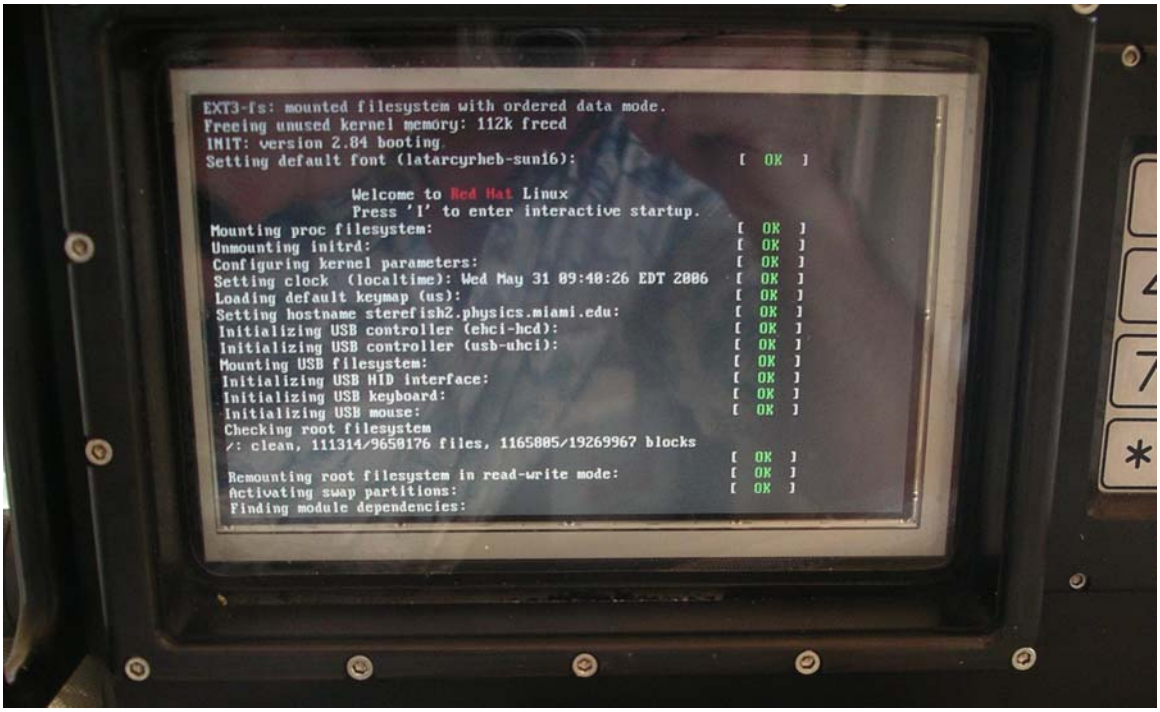


Figure **38** - The fourth screen is from the Linux boot and initialization process

Figure **39** - The fifth screen is an intermediate un-initialized graphics screen just before the Stereofish program takes control

Figure **40** - The initial fully operational screen –indicating Exposure setting (1-100), Gain setting (0-100), Step Size and buffer delay.

The initial fully operational screen –indicating Exposure setting (1-100), Gain setting (0-100), Step Size and buffer delay. The four top left quantities are all manually adjustable. The exposure and gain controls can also be set in four different modes, fully manual, manual exposure and automatic gain (default), manual gain and automatic exposure, and automatic gain and exposure. One can toggle through these modes by pressing the "Exposure mode" button (3) on the keypad. The yellow box adjacent to the Gain setting readout indicates that this setting is under automatic control.

Figure **41** - The Stereofish camera is fully operational at this stage and will accept a variety of imaging, adjustment or other special commands.

The Stereofish camera is fully operational at this stage and will accept a variety of imaging, adjustment or other special commands. The operator, by pressing the "selection" button (8) twice has selected the step size for adjustment (current selection is indicated by green lettering).

The following is a table of the commands listed on the back of the camera as a reminder to the operator:

Imaging:
- Single shot = 5
- Movie mode = 6
- Buffer mode = 4

Adjustments:
- Change Selection = 8
- Decrease value = 7
- Increase value = 9

Special commands:
- Toggle through (auto) exposure algorithms = 3
- Zoom mode = *
- Ramp exposure setting in a loop = 0
- Shutdown = #1
- Full Linux mode = #2

Our experience has shown that preferred setting is manual exposure with automatic gain. This is the default setting when the camera powers up.. This currently produces the highest quality automatic “exposure” images. The completely automatic exposure/gain setting provided in the camera software library has inappropriate behavior for imaging moving targets. The fully automatic algorithm, upon entering low light situations, favors increasing the exposure to a maximum of 100 before beginning to increase the gain. Long exposures often result in image blur which render the images useless for stereo pair information extraction. Therefore, we have set the initial exposure setting at 40, and the Gain is set on automatic. In the scene shown above, the light is more than sufficient for this exposure, the gain is fairly low, and the exposure setting should be manually reduced.

Figure **42** - The operator has lowered the Exposure value by first pressing the selection button (8) and selecting “Exposure” (indicated by green lettering) and then adjusting its value with a combinations of the 7 (decrease) and 9 (increase) keys.

The operator has lowered the Exposure value by first pressing the selection button (8) and selecting “Exposure” (indicated by green lettering) and then adjusting its value with a combinations of the 7 (decrease) and 9 (increase) keys**.**

Notice the presence of the two histograms for each sensor to aid in finding the optimum exposure. The presence of the yellow rectangles on the right side of the histogram section of the display indicates that greater than 5% of the total number of pixels are “saturated” or at the highest value of 255. Exposures with a large number of pixels in this range will appear washed out. The automatic exposure mechanism does not consider the number of saturated pixels in the image, only the average value of all of the pixels. The above scene has a large dynamic range with large numbers of dark pixels and large numbers of bright pixels.

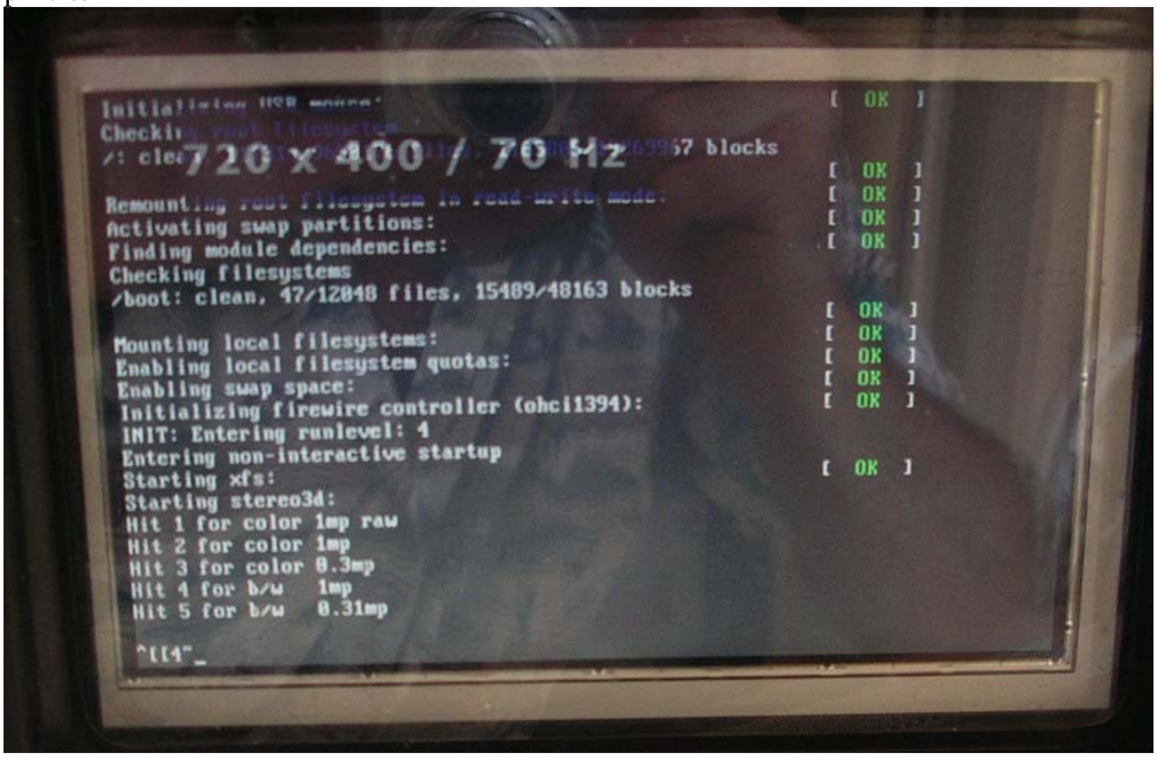


Figure **43** - Power down:using the two keystroke combination of #-1 to shutdown the camera, this is the final screen observed just before power off

Appendix B: Data acquisition dives

Field use - Data acquisition dives: To determine the utility of the camera towards its intended use we performed a number of field experiments. Initially these were simple test dives in a pool where we were able to iron out a number of reliability issues.
We then proceeded to take the camera out under realistic field conditions and uncover other operational issues (wet mate connectors, disk drive/OS software) issues, focus, aperture setting, exposure, and calibration issues.

Tue14Oct2003 – Initial open water dives, 3 separate dives, – Aperture settings are uneven and focus is off despite laboratory calibration. Disk drive problems end the dive

06Dec2004, 3 Dives, Key Largo
- 041206-120626 640x480 mode
  - Extremely turbid site. See movie 1020
  - Rampomatic failed exposure 0, gain 0,
  - Measurements, movie 322, 422, 546 (1200 shots)
  - Calibration sequence
- 041206-122823 640x480 mode
  - Turbid site, swim around, 6973 shots

041206-141851 640x480 mode
Turbid site
Some Rampomatic looks good
041206-144514 640x480 mode
Rampomatic sequences

25Jan2005 – Dive Key Largo Benwood, camera fails- disk errors

09Feb2005 – 2 Dives, Rampomatic, measurements, calibration movie
050209-120010 (12:00 – 6246 pairs)
050209-140409 (14:00 – 3160 +30pairs)

Modified software for 1MP mode

26Mar2005 3 dives Key Largo, Benwood– intermittent problems on first dive, second dive was picture perfect, black grouper aggregation (note that data is stored in 032505 folders) black grouper at 1mp is in 032505-140148/movie-2107->2233

09May2005 upgraded software libraries to SVS 3.2g
31May2005 – Dive for bottom mapping/survey data

18Jul2005 – Installed RedHatLinux 9, svs3.2h 80GB drive, modified folder format to save as was MMDDYY
26Jul2005 – New color sensors (MDCS2) installed. Modified software to acquire raw images for speed purposes.
02Sep2005 – Diving to test new color sensor – Goliath grouper, Exposure too slow, Aperture setting too small, auto exposure method is flawed.
08Nov2005 – 2 dives, shakedown for bottom survey, modified AE methods and aperture settings.
06Dec2005 – Dives with UM RSMAS Marine Geology, Reef/bottom surveys, Andros , Bahamas, AUTEC
07Dec2005 – Dives, Reef/bottom surveys, AUTEC, Andros, Bahamas.
09Dec2005 – Dives, bottom surveys, AUTEC Andros Bahamas