

DESIGN AND CALIBRATION OF AN UNDERWATER STEREO-VIDEO SYSTEM FOR THE MONITORING OF MARINE FAUNA POPULATIONS

Mark R. Shortis

Department of Geomatics
University of Melbourne
Parkville 3052, AUSTRALIA
M.Shortis@unimelb.edu.au

Euan S. Harvey

Department of Marine Science
University of Otago
Dunedin, NEW ZEALAND
Euan.Harvey@stonebow.otago.ac.nz

Commission V, Working Group IC V/III

KEY WORDS : Video camcorders, underwater, calibration, stereo matching, marine fauna, environmental monitoring

ABSTRACT

Assessment of age and size structure of marine populations is often used to detect and determine the effect of natural and anthropogenic factors, such as commercial fishing, upon marine communities. A primary tool in the characterisation of population structure is the distribution of the lengths or biomass of a large sample of individual specimens of a particular species. Rather than use relatively unreliable visual estimates by divers, an underwater stereo-video system has been developed to improve the accuracy of the measurement of lengths of highly indicative species such as reef fish. In common with any system used for accurate measurements, the design and calibration of the underwater stereo-video system are of paramount importance to realise the maximum possible accuracy from the system. Aspects of the design of the system, the calibration procedure and algorithm, the determination of the relative orientation of the two cameras, stereo-measurement and stereo-matching, and the tracking of individual specimens are discussed. Also addressed is the stability of the calibrations and relative orientation of the cameras during dives to capture video sequences of marine life.

1. INTRODUCTION

A main theme of global ecosystem monitoring programs is how the marine ecosystems of our planet will be affected by environmental impacts and how, in turn, this will effect global climate change. To this end, major research is required on the response of the marine fauna and flora to changes in physical and biological factors. With many reef fish occupying positions at or near the top of food webs, their abundance and population structure are highly dependent on the availability of food and the state of their environment. Accurate information on the size structure of a fish population, when linked with knowledge of the biology of the species, can allow analysis regarding fishing intensity, environmental impacts and rates of recovery (McCormick and Choat, 1987). Species of reef fishes are therefore useful indicators of the status of near shore temperate and tropical ecosystems. Visual census techniques (Thresher and Gunn, 1986) utilise SCUBA divers to count reef fish abundance and in some cases estimate the lengths of reef fish to determine the size frequency or mean length of a population. These visual census techniques are used in marine reserves and sanctuaries around the world to monitor whether changes in the abundance and size frequency of species populations are occurring (Francour, 1994). More recently they have been used in fisheries management as a tool for assessing the standing stock or biomass of individual species based on the relationship between the estimated length and the weight of individual fish of a certain species (Russ and Alcala, 1996).

However, making accurate and precise visual estimates of the length of objects underwater is extremely difficult and requires the observers to be well trained and experienced (English *et al.*, 1994). The estimation of the length of an object underwater is complicated by the refractive effects of water, which increases the apparent size of objects and causes objects to appear to be closer to the observer than the actual range. Further, researchers using SCUBA are not efficient workers when performance underwater is compared to similar activities in the air (Hollien and Rothman, 1975). In addition, the sampling bias and errors resulting from the detrimental physiological effects related to SCUBA diving can be significant (Baddeley, 1965)

and where data from visual size or length estimates have been published, few authors include the precision or accuracy of the data. Problems with long-term studies occur when different observers may be involved in making estimates of size or length of marine organisms at different spatial and temporal scales. Even though calibration procedures are used by some researchers (Bell *et al.*, 1985) inter-observer variability has the potential to cause major biases. If the data collected is to be used to compare the size estimates recorded for different times, places or species then it is important that the level of precision and accuracy is known to enable rigorous analysis of the comparisons. Due to observer error and biases it is probable that many studies lack the statistical power to detect small changes in the length of the organisms being studied (English *et al.*, 1994; Fairweather, 1991). To overcome the problem of subjectivity in visual estimates and enhance accuracy and precision, an impersonal system of measurement is preferable. Clearly, any impersonal system of measurement must be technology based, but within the limits imposed by the underwater environment and finite resources of research organisations.

Many marine scientists and biologists have experimented with conventional and video imagery. For example, Klimley and Brown (1983) describe the use of stereophotography for estimating the size and dispersion of free swimming sharks. The system was viable underwater, convenient to use for measurement and could be developed or purchased at a reasonable cost. As a consequence, stereo-video cameras were quickly adopted for a wide range of applications in the marine environment (Hamner *et al.*, 1987; Vrana and Schwartz, 1989). In recent times there have been rapid technological improvements in video cameras which has improved the utility and accuracy of such systems.

Metric photogrammetry has been used specifically for various types of biological recording and analysis, generally using stereo photographs. Film-based stereophotogrammetry has been used to make many types of biological measurements, such as the demography of underwater plants (Kaczynski and Szmaja, 1988). Conventional film and video systems have been

used for non-biological underwater measurements, such as oil rig inspections, in both single camera and stereo-configurations (Turner, 1992). More recently, a single video camera has been used to monitor the shape of deployed fishing nets (Schewe *et al.*, 1996).

This paper details the continuing development of a stereo-video system specifically designed for the monitoring of populations of marine fauna such as reef fish (Harvey and Shortis, 1996). The long term aim of the research and development is a reliable system which can accurately and precisely determine the size and shape of objects underwater. Whilst the current version of the system utilises camcorders and analog video tape, the algorithms and techniques developed are independent of the basic hardware and therefore can take advantage of advances in video technology, such as the new generation of digital video cameras. Current development of the system is concentrated on enabling a very high level of automation of the calibration and measurement processes to facilitate wider use of the system by marine scientists.

2. SYSTEM DESIGN

The design of the stereo-video system was influenced by many factors, but the principal design aim was to make precise size estimates of large mobile reef fish at distances of 3 to 10 metres. A typical range between divers and fish in clear water is 5 metres, and common transect widths used in population sampling are 2.5 or 5 metres. One of the critical factors affecting the precision of measurement is the base separation of the cameras. The separation is dictated by the size of the camera base bar and influences the ability of the diver to manipulate the cameras underwater. A larger frame improves the overall measurement precision due to the increase in the base separation of the cameras, but is more difficult to manoeuvre underwater. The compromise adopted to reconcile these two conflicting requirements was a base separation of 1.4 metres, realising a base to distance ratio of 3.6 at the typical range of 5 metres.

The stereo-video system uses two Sony Hi8 video camcorders mounted in water proof housings which are fixed to a neutral buoyancy base bar and frame. The CCD arrays used in the video cameras are 1/3" format, colour sensors with an actual resolution of 795 by 596 pixels and a sensor element spacing of 6.3 microns. The minimum focal length of the camcorders in air is approximately 3.75mm, corresponding to 5.5mm in water. As the image format is approximately 5mm by 3.8mm, the maximum diagonal field of view is of the order of 60°.

The focal lengths are typically set at 7.5mm in air (10mm in water) to obtain fields of view of 2.5 metres at a range of 5 metres. Each camera is inwardly converged at 8.5° to gain an optimised field of view (figure 1). Like the base separation, these design issues are adopted as a compromise between competing considerations. For example, shorter focal lengths increase the field of view, but decrease the measurement precision. Similarly, a more acute convergence would improve the measurement precision, but will decrease the useable field of view and increase the apparent perspective distortion.

The quality of images recorded on analog video tape is notoriously poor (Shortis *et al.*, 1993), so a quite conservative *a priori* image measurement precision of 0.5 pixels, which has been borne out by experience, is appropriate. Using this value and the geometry of the stereo-video system, the expected object space precisions are 3mm and 22mm in lateral position and depth away from the cameras respectively, at a 95% confidence level. The expected depth precision is significantly poorer than the position precision due to the non-optimal base to depth ratio. Clearly, the orientation of the fish with respect

to the cameras will have a strong influence on the precision of the length.

A calibration check plate (figures 1 and 3) is mounted at 2.5 metres from the centre of the camera base bar and can be used to verify the stability of the camera relative orientation. The check plate is positioned so that it can be seen in the images recorded by both cameras whilst not unduly intruding in the images. Periodic measurement of the points on the check plate can be made to detect any variability during a dive. A light emitting diode is mounted above the check plate. The diode can be switched on and off manually by the operator at appropriate times to indicate when useful measurements can be made. It also serves as a means of synchronising the left and right images, as the camcorder frame rates tend to drift with time. Synchronisation avoids motion parallax from movement of the cameras or the object of interest, which would decrease the accuracy of measurements due to the introduced systematic error.

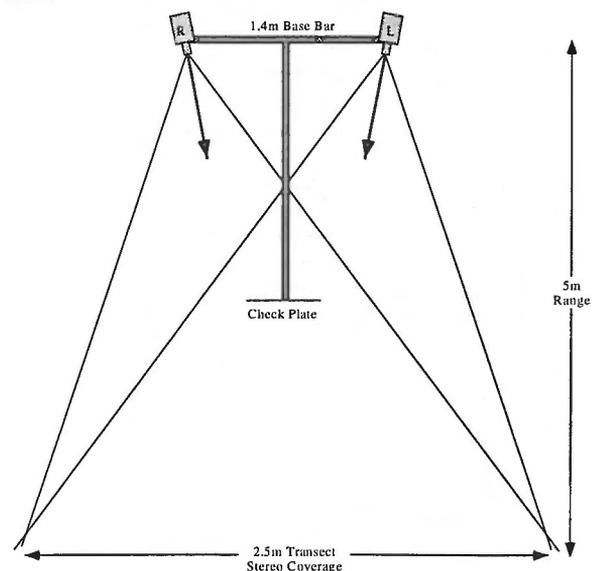


Figure 1. Geometry of the underwater stereo-video system.

The paired video images are converted from an analogue video signal to digital image files using an IBM PC compatible frame grabber. The files are grabbed and stored with a slightly reduced geometric resolution, due to a limitation of the frame grabber, and with a radiometric resolution of 256 colours. Once in digital form the images can be readily adjusted for brightness and/or contrast if necessary, and image locations can be measured within the frames. Grey scale images are often grabbed (using the S-VHS luminance signal only) to simplify subsequent processing during calibration, whilst colour images are generally used for size measurements to enhance the interpretation of the images.

3. SYSTEM CALIBRATION

Calibration of the system is necessary for two reasons. First, the interior orientation of the cameras must be defined to determine the internal geometric characteristics of the cameras, using physical parameters for principal distance, principal point location, radial and decentering lens distortions, plus affinity and orthogonality terms to compensate for bias in the spacing of the pixels on the CCD sensor and effects of the analog tape recording (Shortis *et al.*, 1993). Second, the relative orientation of the two cameras with respect to one another must be determined. The relative orientation effectively defines the separation of the perspective centres of the two lenses, the

pointing angles of the two optical axes of the cameras and the roll rotations of the two CCD sensors.

The camera calibration model does not contain explicit terms for the refractive effects of the perspex camera ports and the refractive interfaces, as analysis of the effects of the refractive surfaces in the optical path in an ideal camera housing shows that images are displaced radially from the principal point (Li *et al.*, 1996). Whilst the assumptions that the optical components of the housing are symmetric around the optical axis of the camera and refractive surfaces are in general perpendicular to the optical axis are unlikely to be perfectly fulfilled in practice, it is clear that the principal component of the refractive effect is radial. As a consequence, the approach which has been widely adopted has been to allow the refractive effects of the optical components and refractive interfaces to be absorbed by the conventional, physical camera calibration parameters. The principal component is implicitly taken up by the standard, odd-ordered polynomial model for radial distortion, whilst any residual effects from asymmetric components of the housing are partly or wholly absorbed into other parameters of the camera calibration, such as decentring lens distortion or the affinity term. No assumptions need to be made concerning the refractive indices of the air, glass or water media, and modeling of the optical components of the underwater housing is unnecessary. This approach has been used successfully by previous systems (Turner, 1992; Schewe *et al.*, 1996), whereas a rigorous approach to optical ray tracing requires a two phase calibration approach and assumed values for the refractive indices of the media (Li *et al.*, 1997). The refractive index of water is known to change with depth, temperature and salinity (Newton, 1989) and the shape of the camera housings and port may change with depth due to the changing pressure. A procedure which incorporates implicit calibration of the complete system under prevailing conditions is likely to be more accurate and reliable.

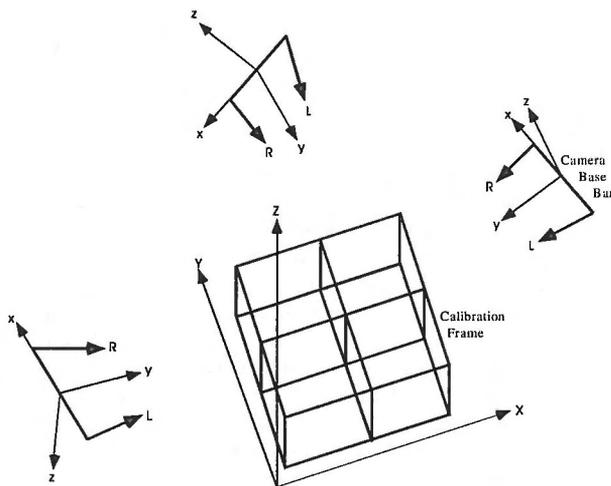


Figure 2. Calibration of the cameras and camera base bar using a calibration frame.

The camera and relative orientation calibration is accomplished using a purpose built frame (figures 2 and 3). The frame is an open cuboid constructed from light weight aluminium. It is black in colour and has 56 white, circular targets riveted to the surface. The targets provide high contrast, unambiguous points which allow a simultaneous, self-calibration of both cameras. Video footage is captured during four rotations of the frame held at an oblique angle by a diver (figure 3), the first rotation

with the camera base in the "normal" horizontal position, and the following rotations with the camera base vertical and in other orientations. The rotation of the frame is more efficient than the diver manoeuvring the cameras around the frame, and emulates a set convergent of camera stations surrounding the frame (figure 2). The rotation of the camera base is necessary to de-couple external and internal parameters of the self-calibration, such as the principal point position and camera station locations. Note that accurate information for the positions of the targets on the frame is not required, as coordinates of the targets are derived as part of the self-calibration procedure. Hence it is immaterial if the frame distorts or is dis-assembled between calibrations, however the frame must retain its structural integrity during a calibration sequence.

From the video footage, 16 synchronised pairs of frames are gathered and the locations of the target images measured. The bundle solution provides estimates and precisions of the camera calibration parameters and the locations and orientations of the cameras at each synchronised pair of exposures. Whilst the camera calibration data are used directly in the subsequent calculations, the location and orientation data must first be transformed. The data for the 16 pairs of synchronised exposures are initially in the frame of reference of the calibration frame. Each pair is transformed into a local frame of reference for the camera base (figure 2). The local frame of reference is adopted as the centre of the base between the camera perspective centres, with the axes aligned with the base direction and the mean optical axis pointing direction. The final parameters for the relative orientation are computed as the average of the values for the 16 pairs. Whilst the computation of the relative orientation is currently a post-process after the bundle solution, a future development of the system will be the incorporation of a stereo-pair constraint solution such as that developed by King (1995).

4. MEASUREMENT SYSTEM

Once the relative orientation is established, measurements within the common field of view of the cameras can be made by locating objects of interest in the left and right stereo images. Again, the images must be synchronised to avoid systematic errors caused by the false shift of objects in one frame relative to the other.

The computer interface for calibration or stereo measurement is shown in figure 3. The left and right images are shown along with a variable zoom window showing the current measurement point. Measurements in the two selected fields of view are made by simple mouse clicks which instigate a centroid computation or an operator defined position within the zoom window. In order to minimise gross errors, operator measurements are aided by displaying epipolar lines in all other images. In calibration mode the system operates as an image comparator which compiles image observations for the self-calibrating photogrammetric solution. The two camera calibrations are simultaneously computed as a multi-camera block-invariant solution.

In image sequence mode, the two pairs of image space coordinates are converted into three object space coordinates using a straightforward intersection computation based on the camera calibration and relative orientation data. Also computed is the RMS image error which is used as an estimator of the quality of the measurement. The operator can step through a declared image sequence in order to, for example, make multiple length measurements of a single fish or to be certain of the identification and measurement of individual fish through the recognition of distinctive patterns of motion.

It is well known that, from the geometry of stereo-photogrammetry, the precision of the computed intersection degrades with distance and the square of the distance for the lateral and depth directions respectively. Experience with the system has shown that, as expected, the RMS image residual values used as an image quality estimator does deteriorate markedly with distance. However the principal use of the quality estimator is to detect mistakes in the image measurements. Measurements made with the stereo-video system are dependent on the clear definition of the objects to be measured. The discrete sampling of the CCD sensors combined with noise artefacts from the video tape recording and frame grabber tends to smear edges and blur detail, which can lead to mis-identification of left and right images of objects to be measured. The quality estimator is tested against a preset criterion and non-identical points always produce very poor quality estimators. Image measurements which fail the test can be immediately re-observed to correct the error.



Figure 3. The image measurement computer interface.

5. PRECISION AND ACCURACY

The precision and accuracy of the system were tested using a simple procedure used for calibrating diver estimates of the lengths of reef fish. PVC sticks or plywood silhouettes of fish are placed in the water and the length estimated (Bell *et al.*, 1985; English *et al.*, 1994). The precision and accuracy of the stereo-video estimates can then be assessed by the difference between the real size and the measured length. To determine the accuracy and precision of the stereo-video system, 16 plastic silhouettes of fish, ranging in size from 100 to 490 mm in length were placed at distances between 2.5 and 6.6 metres from a rope in a salt water pool. The stereo-video system was moved along the rope recording each of the sixteen silhouettes. This process was repeated four times. Ten measurements of the recorded silhouette were made from each set of paired images of the 16 silhouettes.

In all, 64 images of silhouettes were recorded and a total of 640 measurements made. However, during the processing it was observed that, as expected, the quality of the results altered as the angle of orientation of the silhouette changed from perpendicular to parallel to the direction of view. The precision and accuracy of the measurements was clearly degraded if the orientation of the silhouettes was greater than 50° to the direction of view or where the images did not show the head or the tail. Eight silhouette images were subsequently removed on the basis that an operator or an automated measurement process always have the discretion to reject unfavourable orientations of fish within a sequence of images.

The remaining data showed a mean precision of ± 4 mm across all the silhouettes (Harvey and Shortis, 1996). The range of precisions for the silhouettes was ± 2 to ± 13 mm, and showed only a weak correlation with the length and orientation of the silhouettes. The actual precision values are compatible with the precisions predicted by the design of the system. The mean accuracy across all silhouettes was -12.7 mm, with a range of 3.7mm to -24.3 mm, and again showed no strong correlations with length or orientation. This result implies a consistent under-estimation of length which was unexpected. Although this under-estimation can be readily compensated assuming it is a consistent systematic error, the source of the under-estimation is under investigation. Despite this error, the stereo-video system has better accuracy and precision than visual estimates by both novice and experienced divers. Whilst the accuracy of the underwater stereo-video system was 2-4% of the length of the silhouette, the mean errors in the visual estimates made by the divers ranged between 5% to a maximum of 12% of the length (Harvey *et al.*, 1998). The precision of the underwater stereo-video system was typically less than 3% of the silhouette length, compared to diver precisions ranging between 10% and 30% of the length (Harvey *et al.*, 1998).

6. CALIBRATION STABILITY

There are many issues to be considered when predicting or testing the stability of the calibration of cameras used for quantitative measurements. Perhaps the first consideration should be that no camera will be perfectly stable because all cameras are handled in some way during routine operations. The influence of handling is greater for cameras which are not designed for quantitative measurement, such as video camcorders. Like 35mm SLR based digital still cameras, the flexing of the body of the camera and possible movement of the focal plane CCD sensor contribute to the variation, both instigated by handling during photography (Shortis and Beyer, 1997). The stereo-video system introduces the additional stability issue of the relative orientation. Once more, use of the system could be expected to vary the relative orientation due to handling, pressure changes and movement stresses on the camera frame and housings, as well as the possibility of the cameras moving within the housings. However, substantive changes to the camera calibration are unlikely whilst the cameras are sealed in the waterproof housings and the frame remains rigid. Therefore, small variations in the camera calibration and relative orientation might be expected to result from routine use, such as a single dive where the cameras are being manipulated in the underwater environment. Larger changes would be likely to occur between dives, as the water proof housings must be opened and the cameras are handled more to retrieve video tapes. More fundamental changes such as camera disassembly (Shortis and Beyer, 1997) or refocussing (Shortis *et al.*, 1996) would of course result in dramatic changes in the camera calibrations. Similarly, a disassembly of the frame or camera housings and mountings would have a similar affect on the relative orientation.

To test the calibration stability of the underwater stereo-video system within and between dives, calibrations in the controlled environment of a swimming pool have been contrasted against calibrations in open water during routine use of the system (Harvey and Shortis, 1998). In the pool test, nine calibrations were recorded sequentially over a 45 minute period. No changes were made to the stereo-video system during the calibrations, which were recorded at a depth of 5 metres with an estimated visibility of 25 metres. Fifteen calibrations were made under field conditions in open water during routine field work monitoring reef fish populations in Milford Sound, on the west coast of the south island of New Zealand. Calibrations in

the field were undertaken at a depth of eight metres and required only a few minutes to complete. Visibility was estimated to range between 6 and 15 metres, which in all cases was adequate for the typical 5 metre range to the calibration frame.

The stability of the calibration and the relative orientation is indicated in figure 4, which shows the statistical significance of changes in selected parameters. Significance is defined here as

$$\text{Significance} = \frac{|\text{parameter}_n - \text{parameter}_{n-1}|}{\sqrt{\text{variance}_n + \text{variance}_{n-1}}}$$

where the variance is the square of the standard deviation.

The significance value effectively gives a weighted, dimensionless estimator of the change in the parameter which can be tested against a Student-T distribution. For a 95% confidence level and 15 degrees of freedom the critical value is 2.1, however comparative values are generally sufficient for the purpose of this investigation.

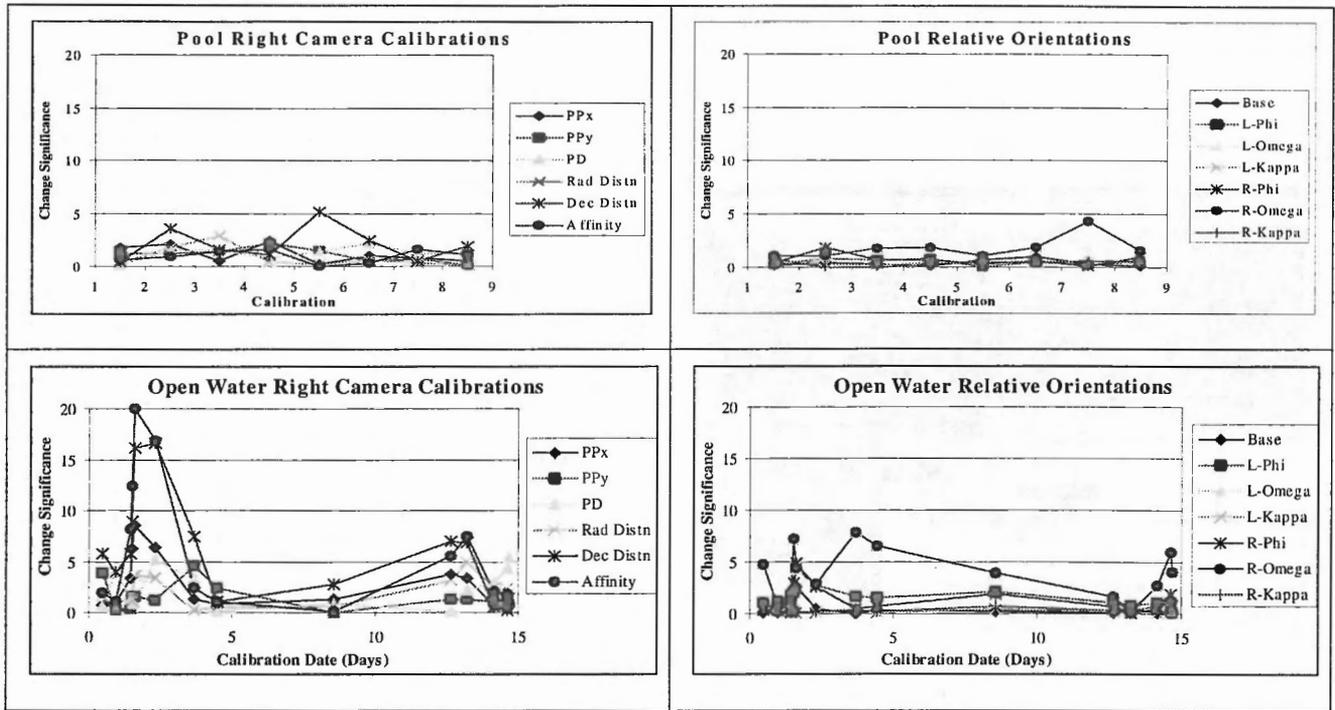


Figure 4. Change significance for camera calibration and relative orientation parameters.

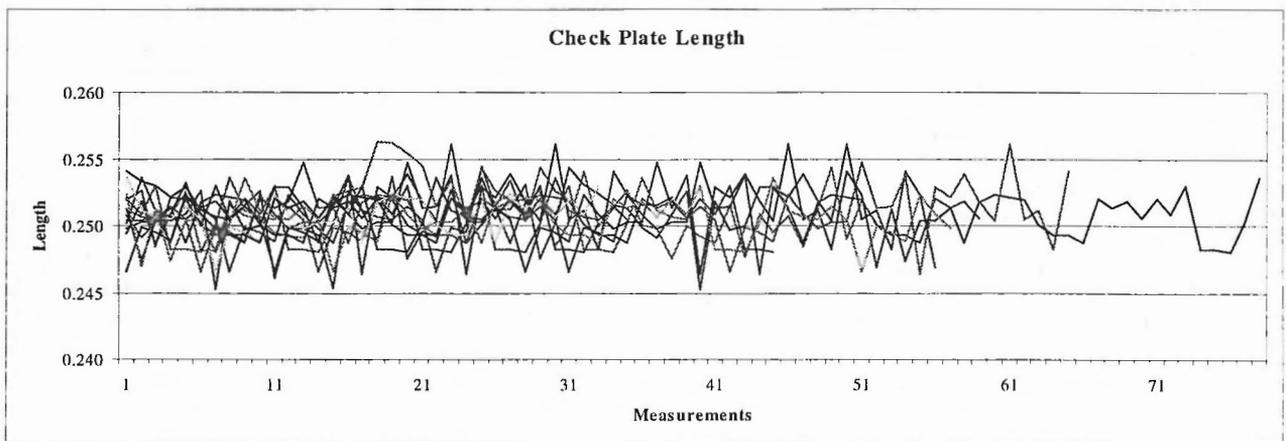


Figure 5. Calibration check plate length measurements for open water dives.

The graphs clearly show that the pool calibrations are quite stable, with some exceptions such as the omega rotations in the relative orientation and the decentring distortion in the camera calibrations. However, the significant changes in these

parameters may be effected by correlations, so there is some doubt as to whether these values represent real variation or are an artefact of the calibration technique. The exactly parallel changes in the left and right omega values are a result of the mathematical transformation employed for the relative

orientation, which seeks a mean rotation of the camera optical axes with respect to the camera base. In contrast, the variations in the parameters for the open water calibrations are extreme in some cases, such as the changes which occurred between two sites on the second day where the disassembly of the stereo-video system, or some other factor such as a refocusing of the lenses, has clearly had a dramatic effect. Although the influence of correlations cannot be ruled out completely, significance values which are ten times the critical value must be indicative of extraordinary change which is not statistically acceptable. Whilst there are clear differences in stability between the pool and open water calibrations, the relative orientation stability of the system in routine use in open water is nevertheless evident.

Whilst the variation in the parameters indicates the fundamental stability of the system, more important is the influence of the variations on measured lengths obtained from the captured images. To test the consistency of length measurements, the length of the calibration check plate was repeatedly measured during the analysis of a number of the open water dives (Harvey and Shortis, 1998). Raw results of the check plate measurements are shown in figure 5 and it is apparent that there are no substantial trends in the length measurements. For individual dives, the mean length measured over the full duration of each dive varied from 0.2504 to 0.2514 metres. The precisions of the length measurements were 2mm or less in all cases and, based on a trend analysis, the maximum apparent change in the length of the check plate for any dive was 1.5mm. The precision of the length measurement is acceptable in terms of the design prediction, and the maximum length change again indicates a high level of stability for the underwater stereo-video system. Considering that the check plate is at half of the typical range for fish length measurements, the maximum error in any such length measurement caused by variation in the system is likely to be no worse than a few millimetres.

7. AUTOMATION

Whilst developments in optical, acoustic, photo-optical systems and under way sampling techniques are providing strong gains in data acquisition within the marine environment, there are few equivalent productivity gains in data analysis. Net, bottle, visual or photographic sampling of the marine environment, followed by manual taxonomic classification and estimation of abundance or size frequency of marine flora and fauna, persist as the most important techniques for marine biological research. The main reason for the slow development in automated measurement of indicative species such as reef fish is that the basic data used in ecological and biological science are living organisms. These present extremely complex signatures to any data collection system and automated high volume data analysis systems must be developed for handling these and similar biological specimens. The effective use of advanced systems of observation and monitoring for the marine environment, such as the underwater stereo-video system, depends upon the elimination of the analysis bottleneck.

Therefore, additional research and development is required to automate both the calibration and measurement processes associated with the underwater stereo-video system. Coded targets or other identifiable markings on the control frame is one solution which is under investigation for automated processing of calibrations. Photogrammetric systems designed for industrial inspection rely on small, uniquely identifiable, high contrast targets (Schneider and Sinnreich, 1992), but experience with image systems in the underwater environment suggests that this approach is not feasible, as the required size of the targets would make the control frame impractical. Some other unique signature is needed, such as colour coding the targets, bar code type markings or shape coding of auxiliary targets. A critical factor is the robustness of the targeting

schemes when used in conjunction with image scanning and segmentation procedures, and the subsequent robust, closed form orientation solutions such as the widely used algorithm of Zeng and Wang (1992). The advantage of a colour coding system is that it can be the basis of a radiometric calibration process. An unambiguous scheme of colour coded targets, perhaps with auxiliary coding, would implicitly provide the orientation of the control frame. Simultaneously, the variation of colour and intensity would provide information on colour calibration of the cameras and the relationship between image intensity and target distance. Although the dispersion of the visible electromagnetic spectrum within water is well known in general, local variations due to changing water quality require compensation to maintain colour fidelity for the captured images. A higher quality level for the underwater images would certainly improve automated discrimination by segmentation and classification techniques.

As part of the automation of the measurement process, image quality must also be automatically enhanced where possible. The multiple phenomena of refraction, dispersion, absorption and backscatter combine to degrade underwater video images in terms of clarity, contrast, brightness and colour rendition. Poor visibility conditions exacerbate the image quality loss. One potential solution for image quality enhancement is the combination of image segmentation and frame averaging for situations where the camera and object are motionless, or nearly so. Frame averaging has been shown to effectively eliminate noise and improve the resolution of images (Fryer and Macintosh, 1997), and therefore has potential as an image quality enhancement tool. In the simplest case, short image sequences of bottom dwellers could be directly averaged. More complex cases of moving objects, such as reef fish, within a cluttered scene can be segmented using a combination of image subtraction, edge detection using operators such as Canny and Sobel, and "optical flow" type algorithms to discriminate between moving and non-moving portions of the images (Barron *et al.*, 1992). The integration and further development of these three techniques is required to resolve single fish and overlapping fish which both may be composed of several segments. Such segments must be clustered or segregated based on information from all techniques across a number of images in a sequence.

8. IMAGE SEQUENCE ANALYSIS

Location of objects within image sequences is fundamentally a geometric problem. At present manual observations made by an operator are employed to take limited measurements in each image of the stereo pair. These image measurements are then converted into object space coordinates or length measurements by a mathematical transformation based on a stereo intersection. This process can be greatly enhanced by the automatic detection and stereo-matching of reef fish. Due to the large number of image sequences that must be analysed to obtain statistically significant data sets across a range of sites for any particular monitoring program, an efficient, robust image matching algorithm is required if biologically meaningful samples are to be routinely obtained.

Template matching is a fundamental tool for image location within the scenes captured by video and digital image systems, and is applied to small areas or patches within the images. Automated image matching systems conform to the following sequence of processes: segmenting corresponding image pairs; extracting features; matching those features; then refining the correspondence polynomial and extracting points. A point feature matching procedure, based on the work of Foerstner (1986), within the segment boundary, combined with some initial approximation of the viewing geometry, can provide adequate initial 3D interest points associated with each fish.

The image matching process will span both stereo images from the two cameras and sequences of stereo images. This logically leads beyond the concept of geometrically constrained multi-photo matching (Gruen and Baltsavias, 1988) and multi-epoch stereo matching (Tao, 1996) which has been applied to many tasks in videometry where more than two images of the object are available. Stereo-image sequences of moving fish will necessitate the integration of geometrically constrained image matching with not only a shape model (Li and Gruen, 1997), but also a shape change detector to refine the descriptions of real objects to produce more accurate and reliable estimates of shape and position. The shape change detector, such as those developed for medical applications of videometry (Pilgrim, 1995), must be incorporated as the object reconstruction task will be complicated by the dynamic nature of the target objects.

Sequences of images of straight body movement will allow the accumulation of image matching information to improve the overall accuracy of the object reconstruction, and therefore the accuracy of such information as shape and size. Basic models of the (expected) object shape can be catalogued into a shape library using CAD model construction using appropriate descriptors, such as spline curves and patches used in engineering, or 3D shape fitting templates (Delingette *et al.*, 1992) used for natural surfaces. The characteristics of the sinuous body motion derived from image sequences may contribute to the identification of particular species. The reliability of this process will be absolutely dependent on the robustness of the shape change detection, which will be based on analysis of the precisions of the image matching and derived object space coordinates.

The tracking problem in this circumstance is complicated by the fact that both the objects to be tracked and the cameras are moving, individual objects may be partly or totally obscured, and some objects will enter and leave the field of view. The investigations into tracking will centre on predictions of motion based primarily on a history of the trajectory (Robson and Shortis, 1997) and, if needed, the body shape and size of individuals. The latter becomes important as a discriminator in the case of occlusion or near collision of objects being tracked. In many cases range differences will be sufficient to maintain a tracking lock on individuals, however the additional information can be used as a verification.

The initial state in an image sequence can be established by applying the segmentation processes to one camera image from the stereopair, conducting an epipolar line search for the conjugate image of the object in the other camera image, and then applying geometrically constrained matching. The next image pair in the sequence can then be scanned for objects based on the assumption that little movement has taken place, therefore constraining the initial search windows. Objects which are successfully detected and matched can then form the basis for a transformation of the frame of reference due to camera movement. However in many cases the determination of the trajectory of the cameras will not be uniquely defined by this process. The absolute orientation of the camera pair is not required, only a relative measure within any sequence is needed so as to maintain the integrity of the trajectories of individual objects. Static background objects can be employed, when available, to contribute to the relative pose estimation of the stereopair. However it will be necessary for the developed strategies to be adaptable to the changing environment, variation in the background and continual change in the number of tracked objects in order to maintain a high level of automation.

Simultaneously with the camera trajectory determination, the trajectories of individual objects will be established and subsequently predicted. This procedure will require iteration

and filtering across several frames at the start of a sequence, considering the complex nature of a field of moving objects imaged by a moving pair of cameras. Algorithms and methodologies to recognise unpredictable, rapid changes in trajectory must be developed. These algorithms will be essential when tracking many, densely packed individuals in a school, where consistent tracking of identified individuals is paramount to avoid re-enumeration. In terms of the ultimate practicality of the technique, the speed of processing will be important for long image sequences. Once the trajectories of the objects and the camera pair are established, the process can then proceed automatically for a specified sequence of frames.

9. CONCLUSIONS

Biologists making visual estimates of size or length should aim for the maximum achievable accuracy and precision realistically available to them. As the accuracy and precision of size or length estimates improves, so does the ability to detect real or relative changes in variables involving biological length or size. Stereo-video offers an alternative to standard visual census techniques where data collection emphasis is on accurate and precise estimates of size or length. Further, the measuring system can be applied to tasks other than recording lengths, especially those measurements which are too complex or too inefficient to be carried out through visual estimation by divers. For example, biomass estimation for environment monitoring requires the accurate three dimensional mapping of reef communities in order to assess changes in volume. Periodic determinations of reef biomass may be feasible using a combination of image matching and shape fitting templates. The stereo-video system is also well suited to population studies based on correlation from indirect information, such as the external tubes or sand funnels of bottom dwelling marine animals. Visual estimation is not sufficiently accurate to detect very small dimensional changes, and the fragility of these structures demands a non-contact measurement method.

The system described in this paper is a prototype. However, the initial testing has clearly demonstrated the potential for efficient and appropriate under-water measurement. The system can be scaled down or up for various applications with corresponding changes in accuracy and precision. The technique is robust and insensitive to user experience, therefore removing biases resulting from inter-observer variability. Underwater stereo-video measurement has many advantages over visual census techniques. It is anticipated that stereo-video systems will gain widespread acceptance in many applications in the future, especially as systems are developed to incorporate automated and accurate calibration and measurement routines. The increased power of statistical testing enabled by large quantities of accurate data generated from automated analysis will revolutionise environmental monitoring programmes dependant on the visual classification, enumeration and/or manual measurement of specimens.

REFERENCES

- Baddeley, A. D., 1965. The influence of depth on manual dexterity of free divers : a comparison between open sea and pressure chamber testing. *J. Appl. Psy.*, 50: 81-85.
- Barron, L. J., Fleet, D. J., Beauchemin, S. S. and Burkitt, T. A., 1992. Performance of optical flow techniques. *IEEE Trans. on CV&IP*, 3 : 236-242.
- Bell, J. D., G. J. S. Craik, D. A. Pollard and B. C. Russell, 1985. Estimating length frequency distributions of large reef fish underwater. *Coral Reefs*, 4: 41-44.
- Delingette, H., Hebert, M., and Ikeuchi, K., 1992. Shape representation and image segmentation using deformable surfaces. *Image and Vision Computing*, 10(3) : 132-144.

- English, S., C. Wilkinson and V. Baker, 1994. Survey Manual for Tropical Marine Resources. ASEAN-Australia Marine Science Project : Living Coastal Resources Project. 368pp.
- Fairweather, P. G., 1991. Statistical Power and Design Requirements for Environmental Monitoring. *Aust. J. Mar. Freshwater Res.*, 42: 555-567.
- Foerstner, W. 1986. A feature based correspondence algorithm for image matching. *Int. Arch. of Photogrammetry and Remote Sensing*, 26(3): pp150-166.
- Francour, P., 1994. Pluriannual analysis of the reserve effect on ichthyofauna in the Scandola natural reserve (Corsica, Northwestern Mediterranean). *Ocean. Acta.*, 17(3): 309-317.
- Fryer, J. G. and Macintosh, K., 1997. Integration of low resolution digital images. *Videometrics V*, SPIE Vol. 3174, pp 136-144.
- Gruen, A. and Baltsavias E., 1988. Geometrically constrained multi-photo matching. *Photogrammetric Engineering and Remote Sensing*, 54(5) : 633-641.
- Hamner, W., E. Kristof and A. Chandler, 1987. 3-D video and computer analysis of fish schooling. *Oceans '87*, IEEE Publishing Services, New York.
- Harvey, E. S., Fletcher, D. and Shortis, M. R., 1998. A comparison of the precision and accuracy of visual estimates of reef fish length made by divers and a stereo-video system. Submitted to *Marine Ecology Progress Series*.
- Harvey, E. S. and Shortis, M. R., 1996. A system for stereo-video measurement of sub-tidal organisms. *Marine Technology Society Journal*, 29(4) : 10-22.
- Harvey, E. S. and Shortis, M. R., 1998. Calibration stability of an underwater stereo-video system: Implications for measurement accuracy and precision. Submitted to *Marine Technology Society Journal*.
- Hollien, H. and H. B. Rothman, 1975. The effectiveness of diver's work with and without the aid of communication systems. *Marine Technology Society Journal*, 9(8): 3-10.
- Kaczynski, R. and J. Szmaja, 1988. The application of underwater stereophotogrammetry to ecological studies of macrophytes. *Int. Arch. of Photogrammetry and Remote Sensing*, 27(B5): 306-309.
- King, B. R., 1995. Bundle adjustment of constrained stereo pairs - mathematical models. *Geomatics Research Australia*, 63 : 67-92.
- Klimley, A. P. and S. T. Brown, 1983. Stereophotography for the field biologist: measurement of lengths and three dimensional positions of free swimming sharks. *Env. Bio. Fishes*, 12(1): 23-32.
- Li, H. and Gruen, A, 1997. LSB-Snakes for industrial measurement applications. *Optical 3D Measurement Techniques IV*, Zurich, Wichmann. pp169-178.
- Li, R., Li, H., Zou, W, Smith, R. G. and Curran, T. A., 1997. Quantitative photogrammetric analysis of digital underwater video imagery. *IEEE Journal of Oceanic Engineering*, 22(2) : 364-375.
- Li, R., Tao, C., Zou, W, Smith, R. G. and Curran, T. A., 1996. An underwater digital photogrammetric system for fishery geomatics. *Int. Arch. of Photogrammetry and Remote Sensing*, 31(B5) : 319-323.
- McCormick, M. I. and J. H. Choat, 1987. Estimating total abundance of a large temperate-reef fish using visual strip-transects. *Marine Biology*, 96(4): 469-478.
- Newton, I., 1989. Underwater Photogrammetry. *Non-Topographic Photogrammetry*, Second Edition, Karara, H. M. (Ed.), ASPRS Science and Engineering Series, Falls Church, Virginia. 445 pages.
- Pilgrim, L., 1996. Robust estimation applied to surface matching. *ISPRS Journal of Photogrammetry and Remote Sensing*, 51(5) : 243-257.
- Robson, S. and Shortis, M. R., 1997. Surface characterisation by tracking discrete targets. *Fourth Conference on Optical 3-D Measurement Techniques*, A. Gruen and H. Kahmen (Editors), Wichmann, pp 179-186.
- Russ, G. R., and A. C. Alcala, 1996. Do marine reserves export adult fish biomass? Evidence from Apo Island, central Philippines. *Marine Ecology Progress Series*, 132: 1-9.
- Schewe, H., Moncreiff, E. and Gruendig, L., 1996. Improvement of fish farm pen design using computational structural modelling and large-scale underwater photogrammetry. *Int. Arch. of Photogrammetry and Remote Sensing*, 31(B5) : 524-529.
- Schneider, C.-T. and Sinnreich, K., (1992) Optical 3-D measurement systems for quality control in industry. *Int. Arch. of Photogrammetry and Remote Sensing*, 29(B5) : 56-59.
- Shortis, M. R. and Beyer, H. A, 1997. Calibration stability of the Kodak DCS420 and 460 cameras. *Videometrics V*, SPIE Vol. 3174, pp. 94-105.
- Shortis, M. R., Robson, S. and Short, T., 1996. Multiple focus calibration of a still video camera. *Int. Arch. of Photogrammetry and Remote Sensing*, 31(5) : 534-539.
- Shortis, M. R., Snow, W. L., Childers, B. A. and Goad, W. K., 1993. The influence of storage media on the accuracy and repeatability of photogrammetric measurements using CCD cameras. *Videometrics II*, SPIE Vol. 2067, pp 80-92.
- Tao, C., 1996. An integrated approach to road centre-line reconstruction using stereo image sequences from a mobile mapping system. *Int. Archives of Photogrammetry and Remote Sensing*, 31(B3) : 857-862.
- Thresher, R. E. and J. S. Gunn, 1986. Comparative analysis of visual census techniques for highly mobile, reef associated piscivores (Carangidae). *Env. Bio. Fish.*, 17: 93-116.
- Turner, J., 1992. The development of an operational digital photogrammetric system for the North Sea oil and gas industry. *Videometrics II*, SPIE Vol. 2067, pp 136-144.
- Vrana, K. J. and J. Schwartz, 1989. Instrumented sled, ROV join to provide enhanced images of Edmund Fitzgerald. *Sea Technology*, 30(12): 17-21.
- Zeng, Z. and Wang, X. (1992) A general solution of a closed-form space resection. *Photogrammetric Engineering and Remote Sensing*, 58(3): 327-338.