# StereoMeasure and StereoFeatures: Measuring Fish and Reconstructing Scenes Using Underwater Stereo Photographs

Yuri Rzhanov[1] and George Cutter[2]

[1]Center for Coastal and Ocean Mapping, University of New Hampshire, Durham, New Hampshire
[2]Southwest Fisheries Science Center, NMFS, NOAA
La Jolla, California
yuri.rzhanov@unh.edu

## Abstract

Photogrammetric techniques can and should be used for reconstruction of underwater scenes, in particular for estimation of fish sizes and species' recognition. However, specifics of the medium render most of the algorithms developed for in-air conditions inapplicable without major alterations. The authors describe results of their experimentation with collected imagery and work in progress.

## Introduction

Remote sensing and non-lethal approaches for assessments of underwater habitat and biological resources increasingly has relied on underwater images. NMFS routinely deploys cameras on ROVs, AUVs, and other platforms for underwater surveys of fish and habitats. Some of these images are visually inspected and annotated. Relatively few are being processed and used for semi-quantitative analysis of fish abundance, estimation of coverage by certain species, size measurements, etc., particularly when stereo imaging is considered (e.g., Rooper, et al. 2010). The sheer amount of collected data makes obvious the necessity of development of automated processing techniques allowing for substantial reduction of involvement of human operators. In this paper the authors report the ongoing work on automatic reconstruction of 3-dimensional scenes including fish targets ("3D reconstruction") from underwater stereo imagery, which is an initial step for fish detection in complex environments and subsequent species recognition.

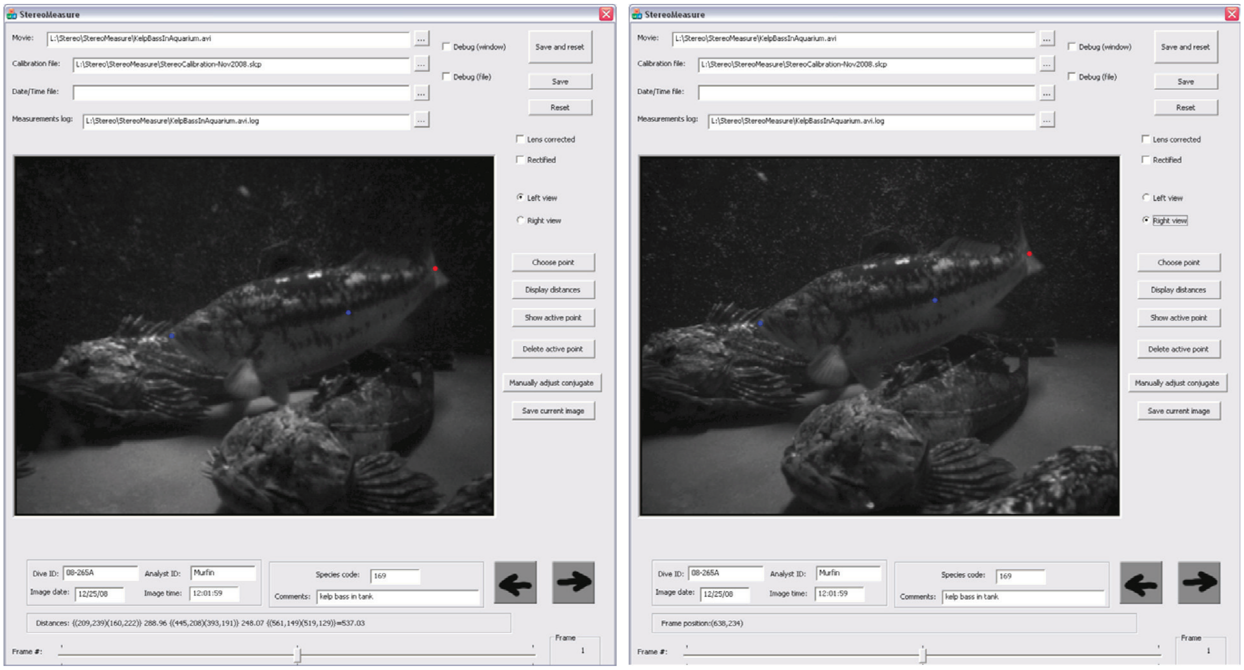
## Methods

Multi-view imaging proved to be a reliable tool for reconstruction of 3-dimensional scenes: geo-registered urban scenes, reverse CAD engineering, etc. (Frahm et al. 2010). Since the foundational work on multiple view geometry in the 1970s, the associated mathematics have been well understood and formulated.

Almost all of the significant results in this area have been achieved working with the imagery acquired in the air under ambient lighting. It appears that direct application of the same techniques to underwater imagery does not meet performance expectations. The reasons for this include 1) distance- and wavelength-dependent absorption of light by the medium; 2) particles suspended even in a very clear water increase noise in acquired imagery; and 3) the effects of artificial illumination that is required underwater beyond several meters deep.

These reasons also complicate construction of mosaics from underwater images, compared to those from images taken in-air. Many 3-D reconstruction techniques utilize a brightness constancy constraint. This assumption rarely holds even for in-air imagery – it expects only Lambertian light scattering and careful photometric calibration of all cameras (Pons et al. 2007). Although these conditions are not encountered in the real world, this assumption is still being used due to its advantages – it allows working with a single-pixel resolution and to use computationally efficient global optimization algorithms (e.g., Kolmogorov and Zabih 2002). All other algorithms require rectification of image pairs prior to searching for conjugate points (points in different images corresponding to the same feature in the scene). Due to the noisiness common in underwater images, the algorithm of choice must be noise-tolerant (Leclercq and Morris 2003). It also must be local, that is, window-based, as line-based algorithms (Birchfield and Tomasi 1996) are known to suffer from streaky artifacts. The most robust algorithm is one that utilizes normalized cross-correlation (NCC) as a measure of matching quality, and it was chosen for our implementation of "StereoMeasure" software.

Careful calibration of each camera and the stereo rig as a whole allowed for correction of lens distortions and rectification of each stereo pair of images. Rectification performs such transformations on the images that any point feature (visible in both frames) appears in these images on the same row. In other words, epipolar lines become horizontal, and vertical disparity becomes zero. (The experiments have shown later that features located near the frame edges still often have nonzero vertical disparity, which indicates lack of accuracy in the calibration procedure, or inability



Movie: L:\Stereo\StereoMeasure\KelpBassInAquarium.avi
Calibration file: L:\Stereo\StereoMeasure\StereoCalibration-Nov2008.slcp
Date/Time file:
Measurements log: L:\Stereo\StereoMeasure\KelpBassInAquarium.avi.log
☐ Debug (window)
☐ Debug (file)
☐ Lens corrected
☐ Rectified
● Left view
○ Right view
Dive ID: 08-265A
Analyst ID: Murfin
Image date: 12/25/08
Image time: 12:01:59
Species code: 169
Comments: kelp bass in tank
Distances: {(209,239)(160,222)} 288.96 {(445,208)(393,191)} 248.07 {(561,149)(519,129)}=537.03
Frame #:
Frame 1

Movie: L:\Stereo\StereoMeasure\KelpBassInAquarium.avi
Calibration file: L:\Stereo\StereoMeasure\StereoCalibration-Nov2008.slcp
Date/Time file:
Measurements log: L:\Stereo\StereoMeasure\KelpBassInAquarium.avi.log
☐ Debug (window)
☐ Debug (file)
☐ Lens corrected
☐ Rectified
○ Left view
● Right view
Dive ID: 08-265A
Analyst ID: Murfin
Image date: 12/25/08
Image time: 12:01:59
Species code: 169
Comments: kelp bass in tank
Frame position:(638,234)
Frame #:
Frame 1

Figure 1. *StereoMeasure* user interface showing three measurement points and their conjugate points used to measure a fish in images from the left and right camera.

of the used calibration model to capture the complexity of the optical system.) Measurement of distances between points in 3-dimensional space was performed by the user choosing (clicking in the GUI) these points in one image from a stereo pair. Points were assumed to belong to some opaque surface (e.g., fish skin, rock in the background, etc.). "StereoMeasure" software automatically converts integer pixel locations to a corresponding location in a rectified image (represented by floating point coordinates), searches for a conjugate point in the complimentary rectified image (utilizing epipolar constraint and NCC score), and converts its location back to the space of the original complimentary frame. All points – manually chosen (MC) and automatically found (AF) – are displayed in the GUI for visual verification (Fig. 1). Occasionally, automatic procedure finds conjugate points incorrectly. The primary reasons for this include: 1) choice of initial point in a textureless area; and 2) choice of the point in the area with the repeatable pattern (like fish scales). In these rare cases (less than 2% during our experimentation), the user can open a dialog box with an upsampled version in the vicinity of either an MC or AF point and shift it in any direction. Upon closing the dialog box, all the calculations are refreshed. Changing the location of a MC point would find another AF point (which is then reviewed for accuracy); changing an AF point overrides the automatically found result and accepts manually corrected result as final. Once MC and AF pairs are finalized, "StereoMeasure" performs triangulation in 3D space and calculates distances between sequential triangulated points. The results can be saved in an ASCII file along with metadata provided by the user (e.g., operator's name, number and date of the mission, etc.) More than two points can be identified by the user to enable measurement of length along a curved surface such as the side of a bent fish.

## Discussion

"StereoMeasure" proved to be useful and reliable tool for underwater distance measurements, and the extension of the work is underway as the "StereoFeatures" project. The current task is to build a dense disparity map for the image pair which allows 3D reconstruction of a scene—estimation of the spatial locations of all targets and seabed or background elements. The scene resolution is limited only by the resolution of input images. Rectified images then undergo extraction and pair-wise matching of salient point features. We have experimented with SIFT (Lowe 2004) and SURF (Bay 2008) keypoints and descriptors, and the results proved to be very similar.

On smooth surfaces, disparity also changes smoothly, so starting with the extracted "seed" points (which are all assumed to be correctly matched), matching is continued in all directions. The search for a conjugate point is conducted within a square window of predetermined size, and the NCC score is aggregated over a square window of a different size. Obviously, the procedure stops working near occlusions (NCC score decreases), textureless areas (spatial variability of NCC score decreases), and around pixels which, despite being conjugate, demonstrate highly

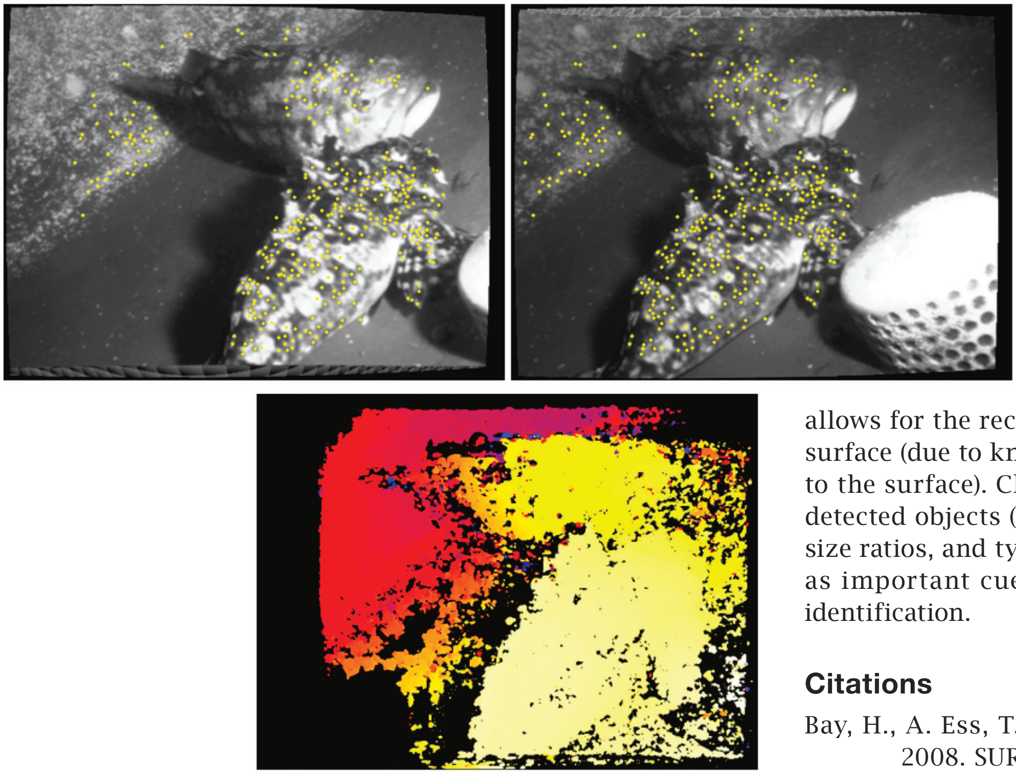


Figure 2. Matched SIFT keypoints for fish and part of the background in left and right stereo camera images acquired from a tank, and (bottom) a rendition of the reconstructed 3-D scene, where colors (yellow to red = near to far) represent distances from the origin, referenced to one of the cameras.

different brightnesses (due to specular reflection, or other reasons). These considerations allow for the formulation of stopping criteria: propagation in a certain direction stops if its NCC score or "textural richness" drops below some threshold. The latter has been defined as the variance of deviations of brightness values from planar surface optimally fit to the current window.

During the first stage, the NCC score is calculated using only integer pixel values. All matches with a score exceeding the threshold are ranked according to this score, and the *n* top-scoring matches are saved as potential candidates (*n* usually ranges from 5-10). Once the entire image is processed, all potential matches are considered in an attempt to minimize the disparity difference for neighboring pixels. The selection process works on "belief propagation" principles. Chosen matches are improved by calculation of subpixel locations of conjugate points. Finally, successful matches are triangulated, creating a cloud of points in the system of coordinates where the left camera is at the origin (Fig. 2). Points calculated from the neighboring matches are immediately linked together creating triangular facets. Those with gaps remain unlinked.

Further processing steps will include detection of 3D objects that potentially represent bodies of fish, separation of these forms from their background, and automated estimation of size and shape. Measurement of water properties (specifically, the dependence of absorption coefficient on light wavelength) allows for the reconstruction of true color of its surface (due to known distance from the camera to the surface). Characteristic measurements of detected objects (sizes in all directions, various size ratios, and typical color patterns) may serve as important cues in species recognition and identification.

## Citations

Bay, H., A. Ess, T. Tuytelaars, and L. Van Gool. 2008. SURF: Speeded-up robust features. Computer Vision and Image Understanding (CVIU). 110:346-359.

Birchfield, S., and C. Tomasi. 1996. Depth discontinuities be pixel-to-pixel stereo. Technical report CS-TR-96-1573. Department of Computer Science, Stanford University.

Frahm, J.-M., M. Pollefeys, S. Lazebnik, B. Clipp, D. Gallup, R. Raguram, and C. Wu. 2010. Fast robust reconstruction of large scale environments. Conference on Information Sciences and Systems (CISS). Princeton, NJ.

Kolmogorov, V., and R. Zabih. 2002. Multi-camera scene reconstruction via graph cuts. Proc. ECCV. 3:82-96.

Leclercq, P., and J. Morris. 2003. Robustness to noise of stereo matching. Proceedings of ICIAP'03. Mantova, Italy.

Lowe, D.G. 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision. 60:91-110.

Pons, J.-P., R. Keriven, and O. Faureras. 2007. Multi-view stereo reconstruction and scene flow estimation wit a global image-based matching score. Int. J. Comp. Vision 72:179–193.

Rooper, C.N., G.R. Hoff, and A. De Robertis. 2010. Assessing habitat utilization and rockfish (*Sebastes* spp.) biomass on an isolated rocky ridge using acoustics and stereo image analysis. Can. J. Fish. Aquat. Sci. 67:1658-1670.

NOAA Technical Memorandum NMFS-F/SPO-121

# Report of the National Marine Fisheries Service Automated Image Processing Workshop

**September 4-7, 2010**
**Seattle, Washington**

by
Kresimir Williams, Chris Rooper, and John Harms (editors)

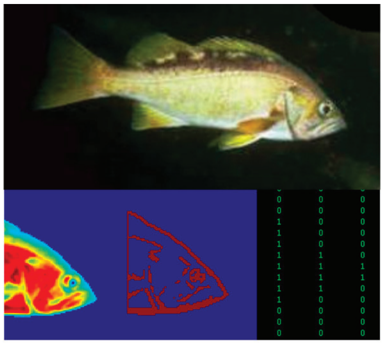
U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Marine Fisheries Service

March 2012

NOAA Technical Memorandum NMFS-F/SPO-121

# Report of the National Marine Fisheries Service Automated Image Processing Workshop

**September 4-7, 2010**
**Seattle, Washington**



Kresimir Williams[1], Chris Rooper[1], and John Harms[2] (editors)

[1]Resource Assessment and Conservation Engineering Division
Alaska Fisheries Science Center
7600 Sand Point Way N.E.
Seattle, WA 98115
*www.afsc.noaa.gov*

[2]Fishery Resource Analysis and Monitoring Division
Northwest Fisheries Science Center
2725 Montlake Boulevard East
Seattle, WA 98112
*www.nwfsc.noaa.gov*

U.S. DEPARTMENT OF COMMERCE
John Bryson, Secretary
**National Oceanic and Atmospheric Administration**
Jane Lubchenco, Under Secretary and Administrator
**National Marine Fisheries Service**
Samuel Rauch III, Acting Assistant Administrator for Fisheries

## NOAA Technical Memorandum NMFS

This document should be cited as follows:

Kresimir Williams, Chris Rooper, and John Harms (editors). 2012. Report of the National Marine Fisheries Service Automated Image Processing Workshop. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-F/SPO-121, 48 p.

**A copy of this document may be obtained from:**

Alaska Fisheries Science Center
NMFS, NOAA
7600 Sand Point Way N.E.
Seattle, WA 98115

*online at: http://spo.nmfs.noaa.gov/tm/*