A Conical Panoramic Stereo Imaging System for 3-D Scene Reconstruction

Pezhman Firoozfam, Shahriar Negahdaripour, and Caroline Barufaldi
Underwater Vision & Imaging Lab
Electrical and Computer Engineering Department
University of Miami
Coral Gables, FL 33124-0640

***ABSTRACT*: The concept of a "conical panorama" was introduced in an earlier paper, involving the view of an oblique down-look camera rotating about an axis, either directly through or at some distance from the center of projection. This provides the framework 1) to analyze multi-camera realizations developed for various underwater visual tasks, and 2) to develop methods for the computation of visual motion that are more robust relative to processing of images covering a finite field of view. This paper presented a stereo conical imaging system, comprising 12 cameras with relatively large overlap between pairs of neighboring cameras. This panoramic stereovision system is targeted for 3-D visualization and improved scene reconstruction by utilizing correspondences in more than 2 views of the scene at modest distances. The calibration of the system is addressed, and experimental results demonstrate the potential in the use of this system for 3-D mapping.**

I. INTRODUCTION

Despite extensive research on the development of autonomous underwater vehicles, a variety of operations involving unmanned submersible, including scientific explorations, will human assistance and supervision in the near future. Both the operator and the scientists rely heavily on the images or video of the remote scene. Undoubtedly, visualization tools that provide 3-D virtual presence, as realistic as they can be made, will be highly desirable; that is, to be able to see the remote 3-D environment surrounding the vehicle,rather than the most current frame of a video. In this regard and (or) for building a global picture or map of the environment, high-resolution images covering a large field of view (FOV) are very informative [24, 25, 27]. This is primarily why the construction of photo-mosaics from video or snap shots, particularly online during the mission, has now become the most widely recognized application in underwater video survey (e.g., [4, 8, 15, 17, 18, 23]).

Capturing a large amount of information rapidly in a single image has motivated fast growing popularity of cameras with wider fields of view [12]. Unlike a photo-mosaic that is generally generated by moving the camera viewpoint, the goal is to generate a composite image, preferably a panorama, from a fixed camera position. To capture a much wider field of view, an omnidirectional sensor uses refracting and reflecting elements in arranged configurations [24], a panning camera[16], or combination of several cameras [14, 19]. Some work has addressed improving the quality of panoramic images through optics and hardware design [6]. Others have focussed on designing suitable mirrors for capturing omni-images with the desired viewing directions and angles [6, 20], as well as the construction of stereo omni-images [16]. Clearly, the geometry, modeling and calibration of omnidirectional cameras are important issues in utilizing these systems for a variety of applications [5].

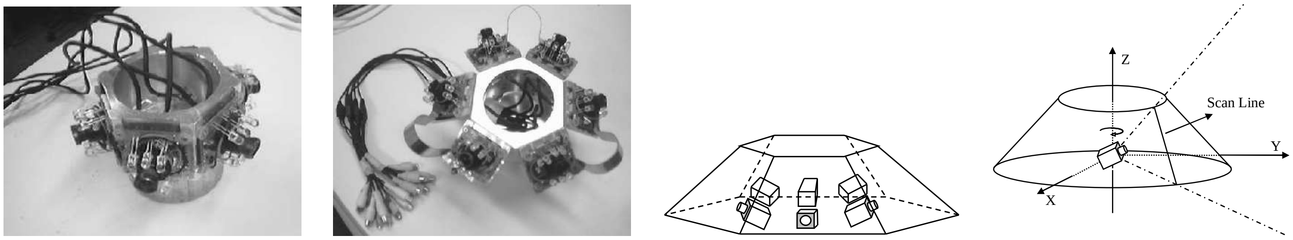
In earlier papers [14, 3], we described the design, construction, calibration and various underwater applications of two multi-cam panoramic cameras. Each comprised 6 cameras for the construction of forward-look and down-look panoramic views; see fig. 1. These systems were configured for minimum overlap between neighboring cameras, primarily serving the purpose of registering images in constructing a *panoramic view*. Such panoramic imaging systems encode similar visual cues as in a monocular vision system with limited field of view. Target applications comprise use of visual motion cues for obstacle avoidance and target tracking, accurate motion estimation for positioning, and 2-D mosaicing. In particular, it has been shown both analytically and through experiments with synthetic and real images that motion estimation can be done more robustly and accurately with sequences of panoramic views, in comparison to data collected over smaller viewing angles [3].

In this paper, we describe a new stereo conical system comprising 12 cameras with substantially larger overlap between pairs of neighboring cameras. In particular, 3 or more of the cameras can view the same object (region of the scene), positioned at some modest distance away. This panoramic stereovision system is targeted for 3-D scene visualization and improved reconstruction by utilizing more than 2 views of the scene at each position. We will present the calibration of the system, as well as some experimental results that address 3-D scene reconstruction.

II. PRELIMINARIES

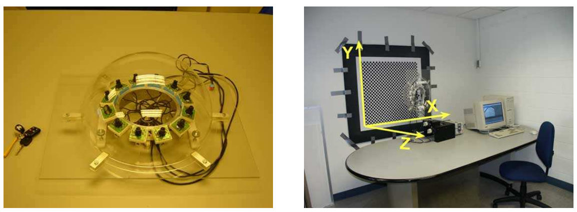
In our earlier paper [3], describing a conical panoramic imaging system with 6 cameras in a down-look configuration (see fig. 1), we provided details on the following key issues: 1) Development of a mathematical model that describes the geometry of a conical view 2) Realization with a multi-camera system; 3) Mapping to and from the conical view and the images from the multi-cam realization; 4) Calibration of the multi-cam system, enabling the alignment of 6 camera images for the reconstruction of a seamless panoramic image. Doubling the number of cameras, a new 3-D imaging system has been developed that provides substantial overlap between pairs of neighboring cameras; see fig. 2a.With two or more cameras imaging the same part of the scene at modest distances, the system is targeted for 3-D reconstruction with improved performance over traditional binocular stereo systems. Just as in the earlier six-camera realizations for con-

0-933957-31-9



(a) (b) (c) (d)

Figure 1: Realization of panoramic imaging with 6 cameras in out-look (a) and down-look (b) configurations. Down-look multi-cam provides a conical view that can be analyzed by the model of a rotating single camera (c) and (d).



(a) (b)

Figure 2: Stereo multi-cam conical camera constructed at the Underwater Vision and Imaging Lab of the University of Miami b) setup for conical camera calibration.

ical views, the key issue in optimizing performance is the extrinsic calibration process; that is, to determine with high precision the position and orientation of each of the 12 cameras relative to a global coordinate system. In binocular systems, this is traditionally done by matching a number of points on a calibration target in the left and right images, often in several poses of the stereo system for better accuracy. To do the same with our multi-cam conical stereo imaging system, the optimization has to be carried out simultaneously over all views, utilizing a very large grid comprising thousands of feature points. Fig. 2b depicts the imaging system positioned in front of the grid patter during calibration, which is described in the next section. It is noted that the system is also to be calibrated to determine the extrinsic parameters [9]– namely, the focal length, aspect ratio, image center position, and the coefficients need to rectify the lens distortion. The distortion problem is particularly significant in our conical system, which was built with 12 inexpensive security cameras each with a 72 degrees field of view.

III. SYSTEM CALIBRATION

In practice, it is difficult to align the cameras in a multi-cam system precisely. In other words, the projection of point $\mathbf{P}$ onto the image plane of each camera is not known precisely. The projection matrices can be found using *bundle adjustment* techniques. Consider a set of 3D points X_j is viewed by a set of cameras. Given the observed image points x_{ij} (the projection of the$j$-th 3D point in the i-th camera image), bundle adjustment is to optimize the estimate of the projection matrices of the cameras $\tilde{\mathcal{P}}_i$ and/or the 3D points $\tilde{X}_j$ such that the distance between x_{ij} and $\tilde{\mathcal{P}}_i \tilde{X}_j$ is minimized. Fig. 3 shows this schematically. Representing it in mathematical language, we solve

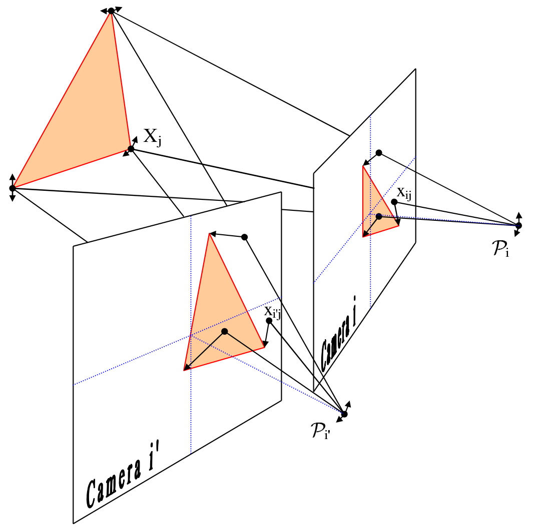


Figure 3: Bundle Adjustment

$$\arg\min_{\tilde{\mathcal{P}}_i, \tilde{X}_j} \sum_{ij} \left\| x_{ij} - \tilde{\mathcal{P}}_i \tilde{X}_j \right\| \qquad (1)$$

Bundle adjustment consists of minimizing the back projection error by varying simultaneously the viewing geometry (camera positions and orientations) and/or the 3D point estimates [21]. This can be accomplished efficiently by sparse *Levenberg-Marquardt* optimization techniques.

We propose a flexible technique to readily calibrate the multicam system, which is a generalization of the camera calibration method by Zhang [26]. In this work, a planar pattern is observed from different orientations to estimate the intrinsic and extrinsic parameters of stereo cameras. We extend this work by considering several camera in the calibration process. This method still requires the cameras to observe a planar pattern at a few different orientations. The motions need not be known and are found from the optimization process. Let $P_{j,0}$ be a point space of coordinate vector$[X_{j,0}, Y_{j,0}, Z_{j,0}]^\top$ in the grid reference frame (see fig. 2-b), and$P_{i,j,0} = [X_{i,j,0}, Y_{i,j,0}, Z_{i,j,0}]^\top$ be the coordinate vector in the reference frame of camera i. Then $P_{j,0}$ and$P_{i,j,0}$ are related through the following rigid motion constraint:

$$P_{i,j,0} = R_i \; P_{j,0} + T_i \qquad (2)$$

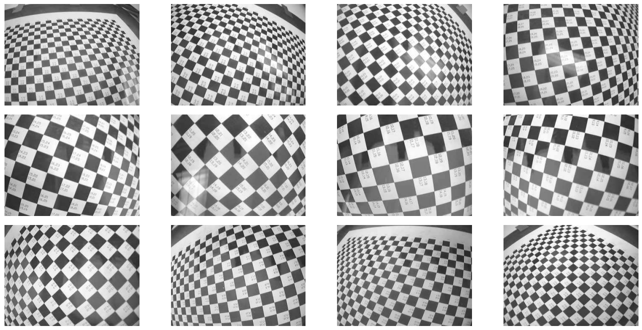


Figure 4: A sample set of pattern images used for camera calibration

where R_i and T_i are the rotation and translation of the pattern respect to the coordinate system of camera i. After changing the orientation of the pattern, the position of the point $P_{j,0}$ is related to its new position trough the following rigid motion equation:

$$P_{j,0} = R_k \; P_{j,k} + T_k \tag{3}$$

where R_k and T_k are the rotation and translation of the pattern with respect to its original position. Computation of the calibration parameters is done through global optimization based on minimizing error between the projection of point $P_{j,k}$ onto camera i from its observed position $p_{i,j,k}$ in the image:

$$(R_{i=0...N-1}, T_{i=0...N-1}, R_{k=1...K-1}, T_{k=1...K-1}) = \text{arg min} \left\{ \sum_{k=0}^{K-1} \sum_{i=0}^{N-1} \sum_{j=1}^{M-1} \left| \mathbf{p_{i,j,k}} - \mathbf{Proj_i} \left(\mathbf{M_{ik}} \mathbf{P_{j,k}} \right) \right|^2 \right\}$$

$$\text{where } \mathbf{M_{ik}} = \begin{bmatrix} R_i & T_i \\ 0_{3\times1} & 1 \end{bmatrix} \begin{bmatrix} R_k & T_k \\ 0_{3\times1} & 1 \end{bmatrix} \tag{4}$$

Note that $R_{k=0} = I$ and $T_{k=0} = [\begin{array}{ccc} 0 & 0 & 0 \end{array}]^\top$. $\mathbf{Proj_i}$ stands for camera projection. This optimization process has been implemented by bundle adjustment, utilizing thousands of points on a 40×40 checker-board grid. The actual calibration experiment is described next, in details.

III. EXPERIMENTS

A. Calibration

A total of 10 sets of conical views, each consist of 12 camera images for the camera calibration, were used. Fig. 4 shows a sample set. The total number of observed points in the images are more than 18,000. The solution of the optimization for finding the extrinsic parameters, applied to equation 4, corresponds to a mean error of less than 0.5 pixel over these points.

Table 1 shows the estimated extrinsic matrices for each camera in stereo conical view structure, expressing the rotation R_i and displacement T_i of camera i with respect to the reference frame in the form of a 4x4 homogeneous transformation:

$$H_i = \begin{bmatrix} R_i & T_i \\ 0_{1\times3} & 1 \end{bmatrix} \tag{5}$$

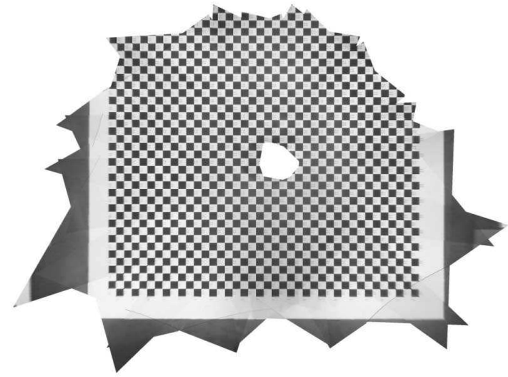


Figure 5: Grid-board mosaic created by arranging the conical views in a common reference frame.

These results have been summarized in table 2, in terms of the misalignment error of each camera from the designed pose and position. These confirm reasonable accuracy in achieving the target structural design and viewing geometry. Two other relevant parameters, determined through calibration, are the distance $d = 131.44[mm]$ from the central axis to the focal point of each camera and the down-look angle $\theta = 75.3[^o]$. Fig. 5 shows the mosaic of the target grid board that is generated by putting the conical views into a common reference frame.

B. 3-D Reconstruction:

An experiment with a data set is presented, consisting of 3 stereo conical views. From a distance of about 50 [cm] from a scene, a $4' \times 4'$ area was imaged. Fig. 6 top-left shows the top view of the scene taken by a still color camera. Circles on this images shows the approximate position of, and area covered by each conical view. The two pie sections in view 2 within dotted and dashed-dotted lines are roughly the regions seen in two neighboring views, depicting the size of the overlap at a $40[cm]$ scene distance. Also, The distance between each of the 3 camera positions is roughly $40[cm]$. In the same figure on the top-right side, we have given the 12 images at each of the three views. In the bottom, we have shown magnification of selected frames (middle rows of each of 3 views). Figure 7 shows the reconstructed 3-D scene as a mesh plot. The processing involves stereo rectification of images from various views at 640x480 resolution, and the 1-D disparity computations from a maximum-liklihood stereo algorithm [2]. The estimated motion between pair of views, based on the method given in [3], allows the mosaicing of the 3-D views.

VI. CONCLUSION

A conical view, referring to a panorama constructed from a rotating camera in an oblique down-look configuration, can be realized by utilizing multiple cameras in the appropriate configuration. A system with 12 cameras has been described for viewing regions of the scene in two or more cameras based on large overlap in neighboring cameras. This stereovision capability enables 3-D scene reconstruction, which is of tremendous value as a visualization tool for 3-D virtual presence, as well as for quantitative measurements in underwater exploration. The mathematic models, enabling the system calibration and the construction of a conical view, were

$$M_{ext}^{0} = \begin{bmatrix} -0.0044 & 1.0000 & 0.0042 & -1.4450 \\ -0.9644 & -0.0054 & 0.2643 & 127.0217 \\ 0.2643 & -0.0028 & 0.9644 & -34.4647 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

$$M_{ext}^{1} = \begin{bmatrix} 0.4977 & 0.8673 & 0.0065 & -0.7740 \\ -0.8432 & 0.4821 & 0.2377 & 127.3303 \\ 0.2031 & -0.1238 & 0.9713 & -32.1809 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

$$M_{ext}^{2} = \begin{bmatrix} 0.8643 & 0.5009 & 0.0450 & -0.5982 \\ -0.4945 & 0.8300 & 0.2582 & 127.6311 \\ 0.0920 & -0.2454 & 0.9650 & -32.7947 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

$$M_{ext}^{3} = \begin{bmatrix} 0.9989 & 0.0334 & 0.0313 & 3.5597 \\ -0.0407 & 0.9614 & 0.2721 & 127.5189 \\ -0.0210 & -0.2731 & 0.9618 & -34.7610 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

$$M_{ext}^{4} = \begin{bmatrix} 0.8668 & -0.4974 & 0.0356 & 0.5120 \\ 0.4724 & 0.8420 & 0.2605 & 127.1339 \\ -0.1596 & -0.2090 & 0.9648 & -33.1324 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

$$M_{ext}^{5} = \begin{bmatrix} 0.5091 & -0.8607 & 0.0067 & 0.2870 \\ 0.8314 & 0.4937 & 0.2548 & 127.6305 \\ -0.2226 & -0.1241 & 0.9670 & -32.7178 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

$$M_{ext}^{6} = \begin{bmatrix} 0.0056 & -0.9995 & 0.0326 & 0.8709 \\ 0.9651 & 0.0140 & 0.2615 & 127.1648 \\ -0.2619 & 0.0300 & 0.9646 & -33.3782 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

$$M_{ext}^{7} = \begin{bmatrix} -0.4866 & -0.8734 & 0.0187 & 1.8168 \\ 0.8460 & -0.4657 & 0.2596 & 127.1858 \\ -0.2181 & 0.1421 & 0.9655 & -31.9602 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

$$M_{ext}^{8} = \begin{bmatrix} -0.8808 & -0.4730 & 0.0212 & -3.9735 \\ 0.4620 & -0.8489 & 0.2568 & 126.4759 \\ -0.1034 & 0.2360 & 0.9662 & -33.6249 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

$$M_{ext}^{9} = \begin{bmatrix} -1.0000 & -0.0011 & -0.0092 & -1.9969 \\ -0.0011 & -0.9721 & 0.2344 & 125.8363 \\ -0.0092 & 0.2344 & 0.9721 & -31.4822 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

$$M_{ext}^{10} = \begin{bmatrix} -0.8485 & 0.5290 & 0.0146 & -5.7699 \\ -0.5097 & -0.8242 & 0.2469 & 125.6056 \\ 0.1426 & 0.2020 & 0.9689 & -33.4626 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

$$M_{ext}^{11} = \begin{bmatrix} -0.5089 & 0.8607 & 0.0161 & 1.1763 \\ -0.8323 & -0.4968 & 0.2460 & 127.9090 \\ 0.2198 & 0.1118 & 0.9691 & -33.5790 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

Table 1: The extrinsic matrices of cameras of stereo panoramic system determined from calibration process.

cam	rotation error [radian]			displacement error [mm]		
#	ω_x	ω_y	ω_z	δX	δY	δZ
0	0.001	-0.010	0.025	-1.88	-1.59	-0.23
1	0.012	0.017	0.026	-1.71	0.05	1.93
2	-0.014	0.033	0.027	-1.43	1.46	1.37
3	-0.007	0.024	-0.004	3.65	2.17	-0.91
4	0.022	0.025	0.027	1.16	0.78	0.22
5	0.013	-0.006	0.019	1.68	0.48	0.88
6	0.038	-0.004	0.024	1.61	-1.02	0.33
7	0.020	-0.019	0.014	0.79	-2.66	2.50
8	0.013	-0.025	0.060	3.86	1.04	-0.06
9	-0.018	0.005	0.028	1.73	0.38	1.80
10	-0.011	-0.010	0.063	4.74	-2.83	0.05
11	-0.013	-0.000	0.018	-2.93	0.05	0.66

Table 2: Misalignment of each camera in the multi-cam conical imaging system. Rotation error is the angle between the poses for the designed system and the actual assembly. Displacement error is the displacement of each camera from its designed position.

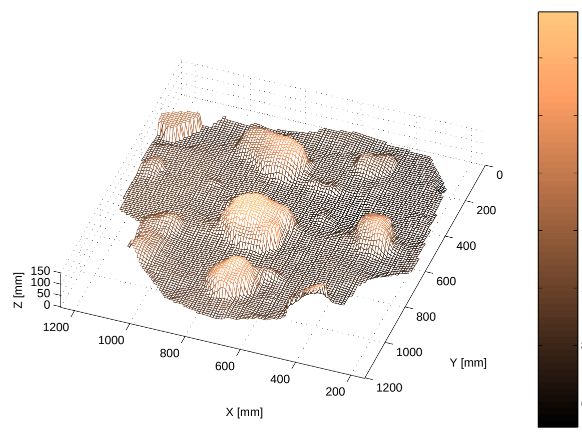


Figure 7: 3D mesh representation of reconstructed scene

given in an earlier paper which described a similar system (with 6 cameras) with very limited overlap. Utilizing these models and a generalization of well-known external calibration method, the geometry of each camera in the coordinate system of the conical view has been computed. This is a critical step for the construction of conical views, and 3-D reconstruction from disparity cues. The 3-D reconstruction capability as well as the mosaicing of multiple 3-D views were demonstrated in an example with images acquired at three nearby positions. This is part of an ongoing project on the development of high-resolution large-area 3-D mapping capability for benthic studies.

Acknowledgement The first author has been supported under grant No. N000140310074, from the Office of Naval Research (ONR), involving the demonstration of automated vision-based ship hull inspection. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the ONR.

References

[1] G. Adiv, "Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field," *IEEE PAMI*, Vol 11(5), May, 1986.

[2] I. J. Cox, S. L. Hingorani, S. B. Rao and B. M. Maggs, "A Maximum LikelihoodStereo Algorithm," *Computer Vision and Image Understanding*, Vol. 63, no. 3, May 1996.

[3] P. Firoozfam and S. Negahdaripour, "Application of Panoramic Images of the Seafloor for Improved Motion Estimation and Photo-mosaicking,"*MTS/IEEE Oceans*, Biloxi, 2002.

[4] S. Fleischer and H.W. Wang and S.M. Rock and M.J. Lee, "Videomosaicking along arbitrary vehicle paths," *Proc. Symp. AUV Technology*, Monterey,CA, June 1996.

[5] J.M. Gluckman, and S.K. Nayar, "Planarcatadioptric stereo: geometry and calibration," *Proc.Conference Computer Vision and Pattern Recognition*, Fort Collins,June 1999.

[6] J.M. Gluckman, and S.K. Nayar, "Rectified Catadioptric StereoSensors," *Proc. CVPR'00*, Hilton Head Island, South Carolina, June 2000.

[7] N. Gracias, and J. Santos-Victor, "Underwater video mosaics asvisual navigation maps," *Computer Vision Image Understanding*, Vol 79, July, 2000.

[8] N. Gracias, and J. Santos-Victor, "Underwater mosaicing andtrajectory reconstruction using global alignment," *Proc.Oceans'01*, Honolulu, HI, November, 2001.

[9] J. Heikkila, and O. Silven, "A four-step camera calibrationprocedure with implicit image correction," *Proc. CVPR*, Puerto Rico, June, 1997.

[10] J.F. Lots, D.M. Lane, E. Trucco, "Application of 2.5D visual servoing to underwater vehicle station keeping," *Proc.Oceans'00*, Rhode Island Convention Center, Sept, 2000.

[11] H.C. Longuet-Higgins, and K. Prazdny, "The Interpretation of a Moving Retinal Image," *Proc. the Royal Society of London*, Vol B-208, 1980.

[12] V. Nalwa. "A true omnidirectional viewer,"Technical report, Bell Laboratories, Holmdel, NJ, Feb 1996.

[13] S. Negahdaripour, C.H. Yu, "Robust Recovery of Motion: Effects ofSurface Orientation and Field of View," *Proc. CVPR*, Univ.of Michigan, Ann Harbor, June, 1988.

[14] S. Negahdaripour, H. Zhang,P. Firoozfam, and J. Oles, *Utilizing Panoramic Views forVisually Guided Tasks in Underwater Robotics Applications, Proc. Oceans*, Honolulu, USA, 2001.

[15] S. Negahdaripour and X. Xu, "Mosaic-based positioning and improved motion estimationmethods for automatic navigation of submersible vehicles," *IEEE Trans. on OceanicEngineering*, Vol. 27, No. 1, 2002.

[16] S. Peleg, M. Ben-Ezra, and Y. Pritch, "OmniStereo: PanoramicStereo Imaging," *IEEE Trans. PAMI*, March 2001.

[17] C. Roman and H. Singh, "Estimation of error in large area underwater photomosaics using vehicle navigation data", *Proc. Oceans*, Honolulu, HI, 2001.

[18] Y. Rzhanov and L.M. Linnett and R. Forbes, "Underwater video mosaicking from seabed mapping", *Proc. Int. Conf. Image Processing*, Vancouver, CA, 2000.

[19] R. Swaminathan, and S.K. Nayar, "Polycameras: Camera cluster forwide angle imaging," CUCS-012-99, Dept. CS, Columbia Univ., 1999.

[20] R. Swaminathan, M.D. Grossberg and S.K. Nayar,"Caustics of Catadioptric Cameras," *Proc. International Conference Computer Vision*, Vancouver, Canada, July 2001.

[21] B. Triggs, P.F. McLauchlan, R.I. Hartley, A.W. Fitzgibbon,"Bundle Adjustment – A Modern Synthesis," *Workshop on Vision Algorithms*, 1999.

[22] R.Y. Tsai, T.S. Huang, "Estimating Three-Dimensional MotionParameters of a Rigid Planar Patch, II: Singular Value Decomposition," *IEEE Transactions on Acoustics, Speech, andSignal Processing*, Vol. ASSP-30, No. 4, August 1982.

[23] X. Xu and S. Negahdaripour, "Vision-based motion sensing for underwater navigation andmosaicing of ocean floor images", *Proc. Oceans*, Halifax, October 1997.

[24] Y. Yagi, and M. Yachida, "Real-Time Generation ofEnvironmental Map and Obstacle Avoidance Using OmnidirectionalImage Sensor with Conic Mirror," *Proc. CVPR*, Honolulu, HI,1991.

[25] Z. Zhang, Weiss, R., and Riseman, E.M.,"Feature Matching in 360^o Waveforms for Robot Navigation,"*Proc. CVPR*, Honolulu, HI, 1991.

[26] Z. Zhang, "Flexible Camera Calibration By Viewing a Plane From Unknown Orientations," *ICCV99*.

[27] J.Y. Zheng, Tsuji, S., "Panoramic representation for routerecognition by a mobile robot," *International JournalComputer Vision*, Vol 9(1), 1992.

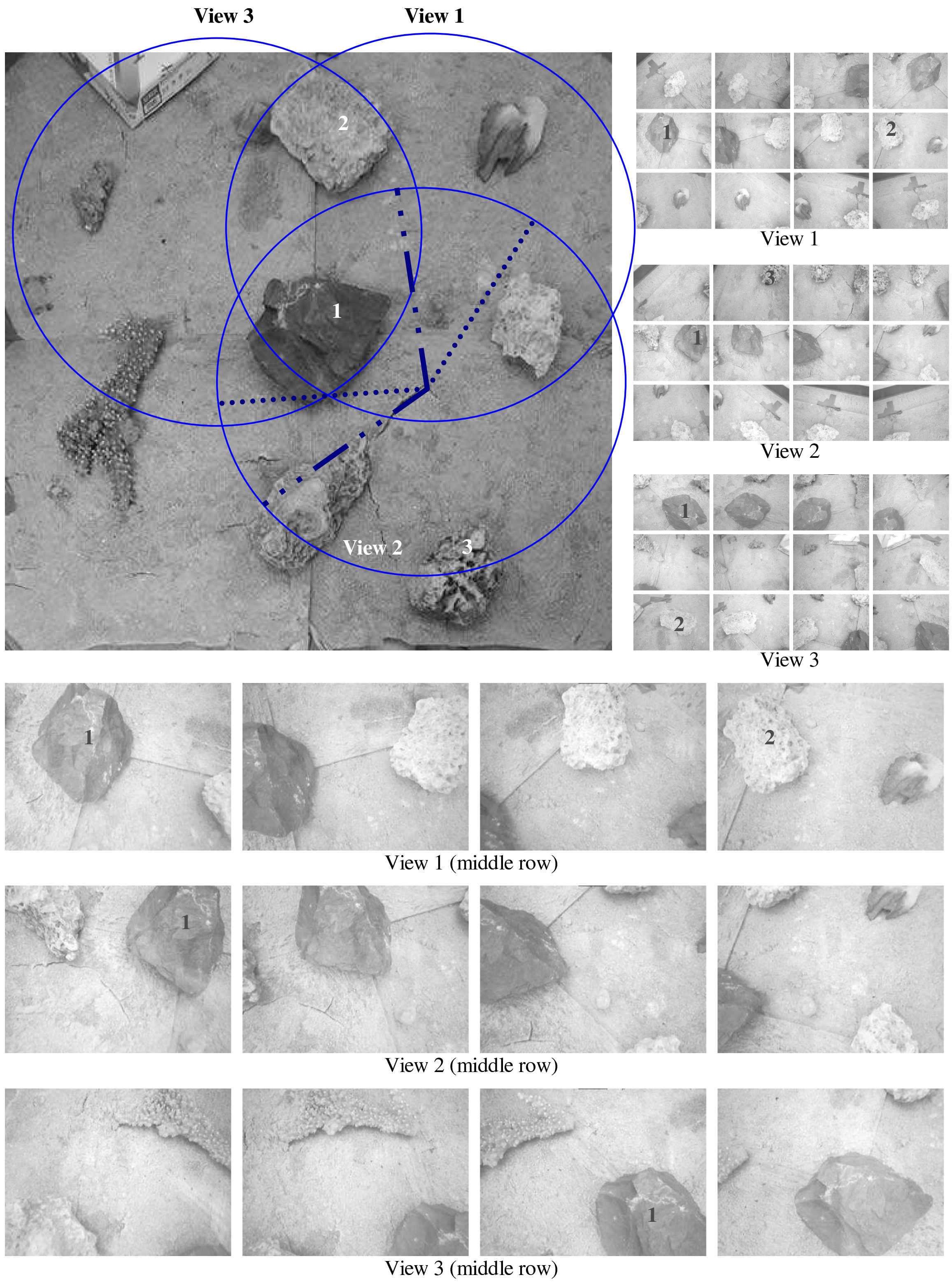


Figure 6: Stereo conical imaging: Top-left is the top view of the scene, where each circle shows rough coverage area of a single camera. Top right shows all of the images in the three conical views. Bottom is the blow-up of the middle-row frames of each conical view.