

Seabed Modeling and Distractor Extraction for Mobile AUVs Using Light Field Filtering

Donald G. Dansereau and Stefan B. Williams

Australian Centre for Field Robotics; School of Aerospace, Mechanical and Mechatronic Engineering
University of Sydney, NSW, Australia; Email: {d.dansereau, s.williams}@acfr.usyd.edu.au

Abstract—A method is presented for isolating moving distractors from a static background in imagery captured by a hovering or slowly moving Autonomous Underwater Vehicle (AUV). By reparameterizing a set of monocular images into a light field structure, it becomes possible to apply a linear fan filter and its inverse to extract the background and distractors, respectively. Results are shown for a hovering AUV imaging a region with non-trivial 3D structure and containing moving elements. The output is a distractor-free 3D light field model of the sea floor and a set of images of isolated distractors. We show that the technique is insensitive to parallax in the background elements, outperforming pixel differencing techniques.

I. INTRODUCTION

Autonomous Underwater Vehicles (AUVs) are playing an increasingly important role in exploring, surveying and monitoring benthic habitats. Recent advances in navigation allow dense, detailed surveys to be performed autonomously [1]–[3], while repeated surveys [4] open the possibility of directly observing habitat dynamics. The use of AUVs in long-term studies is particularly appealing as an alternative to dedicated, potentially intrusive and expensive underwater installations.

A common assumption in underwater surveys is that of a static environment – an assumption violated in many of the most interesting habitats: fish and drifting organisms, swaying vegetation, and dynamic lighting can all act as distractors. In long-term reef monitoring, it is desirable to establish patterns of change in the underlying reef structure while ignoring the movements of fish in the captured imagery. Conversely, one might wish to investigate the distractors themselves, isolating dynamic organisms from their underlying environment.

Distractor isolation from mobile platforms is non-trivial, due to the apparent motion of the environment in the captured imagery. This apparent motion is non-uniform in the case of non-planar 3D scene geometry, and so methods based on pixel-level statistics are insufficient for such applications. The key limitation of these techniques is in their direct use of 2D monocular imagery in what is fundamentally a higher-dimensional problem.

This paper proposes a method for collecting a set of monocular images into a light field structure [5], prior to distractor/background isolation. A light field represents the light permeating a scene and, unlike monocular images alone, implicitly represents the geometry of the scene as well as its textural content. This enables simple techniques to accomplish complex tasks – in particular we show that a linear filter

and its inverse can isolate static background elements and dynamic distractors. The output of the proposed process is a light field model of the sea floor, from which novel views may be rendered, as well as a set of images containing the isolated distractors.

Our method applies to monocular image sequences for which there exist sets of overlapping, approximately co-linear and equally spaced images. This would be the case, for example, for a station-keeping robot which slowly drifts about its ideal location for a large number of frames, or for an AUV moving at a constant velocity, with a low speed relative to the frame capture rate. We make use of an estimate of the camera's pose, obtained from the AUVs navigation system, to select images for inclusion in the light field, though this estimate need not be exact, and can be approximated using robust registration techniques [6]. We also assume that the camera is calibrated, though techniques exist for approximating a calibration directly from captured imagery [7].

The remainder of this paper is organized as follows: Section II provides background on foreground/background isolation and light field processing, Section III deals with constructing a light field from monocular image sequences, Section IV shows how a fan filter and its inverse can be used to isolate foreground and background elements, and Section V shows results for a nontrivial underwater sequence obtained from a hovering AUV. Finally, Section VI gives conclusions and future work.

II. BACKGROUND

The closely related tasks of foreground/background segmentation and change detection are fundamental problems of computer vision, with several successful approaches demonstrated under a variety of scene and camera constraints. For sequences with a static camera, the projection of the background onto the image plane is also static, and so it is possible to utilize simple pixel-based statistics to accomplish segmentation [8]–[10]. This is appealing for several reasons: it is computationally efficient, regardless of scene complexity, it is easily parallelized, and it does not rely on identifying and tracking features, which can be problematic in noisy or self-similar environments.

Extension to rotating cameras exploits the lack of parallax in the motion of the background [11]–[13], and so methods similar to the static-camera case may be utilized. Similarly,

approximately planar scenes with camera motion parallel to the plane – such as in aerial surveillance – present little or no parallax, and so similar techniques may be employed [14].

In the case of a freely moving camera and non-trivial scene geometry, background elements are not static and move with different projected velocities. Several approaches have been proposed for addressing this scenario, including the use of occlusion detection, and using concepts from optical flow to perform iterative camera motion and motion boundary estimation [15], [16]. Other interesting approaches exploit constraints on projected background motion in an orthographic camera, as in [17], which proposes tracking features across the image sequence, and modelling background motion as a sum of basis trajectories – dense per-pixel labelling is then performed in a final optimization step. In [18], motion between pairs of images is considered, for which background elements are shown to lie on a 1D locus. This constraint is exploited to detect foreground elements, though only low-density results are demonstrated. While promising, these techniques are all more conceptually and computationally complex than statistical per-pixel methods, and either generate sparse results or rely on iterative methods to converge on a solution.

It would be appealing, then, to accomplish foreground/background segmentation in a completely *linear* and *featureless* manner – one which does not rely on tracking features, iterative approaches, or optimization frameworks, but which operates simply on a per-pixel basis, as in the traditional static-camera segmentation techniques. In order to accomplish this, we propose collecting a number of input images into a light field structure, and filtering that structure.

Light fields first came about as an image-based approach to computer graphics [5], but have since gathered attention in image processing, allowing image-based, featureless techniques to accomplish complex tasks such as depth estimation and filtering [19]–[22]. Whereas a conventional photograph encodes variations in light as a function of direction for rays passing through a single position (aperture), a light field encodes variations in light as a function of both direction and position, and is typically measured using multiple cameras or lenses, or a single moving camera.

A common light field parameterization is the two-plane parameterization depicted by Fig. 1, which describes a light ray by its point of intersection with two reference planes: the s, t plane given by $z = 0$, and the u, v plane which is parallel to the s, t plane at some positive separation $z = d$. Coordinates in (s, t) can be thought of as defining the *position* of a ray, and in (u, v) as defining its *direction*, so a sample of the light field function $L(s, t, u, v)$ represents the value (monochrome or colour) of one specific ray within the scene.

Another intuitive way of visualizing the light field is to imagine the s, t plane as a grid of pinhole cameras facing the u, v plane. Fixing a value for (s, t) selects one specific pinhole camera, and (u, v) act as pixel coordinates for that pinhole camera. The only complication is that (u, v) coordinates are on a common plane for all cameras, and so are skewed as a function of (s, t) .

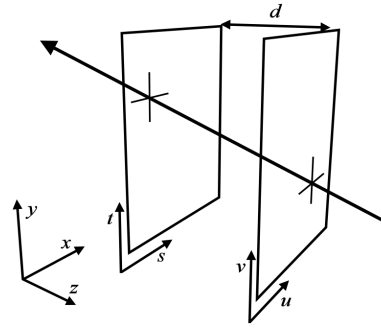


Fig. 1. Two-plane parameterization of light rays

As we shall see, a 3D subset of the light field, $L(s, u, v)$, suffices for the present task. This can be thought of as a horizontal row of images indexed by s , for a fixed value of t . Though this simplifies our task, it also prevents us from processing all images in a hovering AUV sequence – generalization to 4D is suggested as future work.

Our approach relies on constructing a light field from a moving camera. Several methods for rendering directly from arbitrarily posed cameras have been suggested [23]–[25], but only a few have focused on building the two-plane-parameterized light field which we will require for filtering [26], [27]. We will borrow heavily from these techniques in constructing our light field.

III. LIGHT FIELD CONSTRUCTION

Because the light field $L(s, u, v)$ can be thought of as a row of images along s , one way to construct it is to place a camera at each of the s sample locations, facing the u, v plane, and then reproject those images to common u, v coordinates. Assuming the camera is calibrated, we may even allow some relaxation in the orientations of the camera, so long as the images overlap sufficiently to be reprojected onto a common u, v plane. The critical point is that the camera’s aperture must coincide with the locations of the s samples. To this end, we require a set of co-linear, equally spaced camera poses.

A. Image Selection

We consider two scenarios: an AUV collecting a large number of images while hovering above an area of interest on the sea floor, and an AUV moving slowly over an area of interest at an approximately constant velocity, capturing images at a constant frame rate. Both scenarios can yield the co-linear and evenly spaced images we require. As we will be applying a linear fan filter to the light field, at least 10 such images are recommended [22] – more images will yield more selective results, meaning a higher sensitivity to slowly moving distractors, but requiring longer computation times.

In the case of a slowly moving AUV, the requisite images are straightforwardly obtained from a linear section of its path. In the case of a station-keeping robot a cloud of images is available, and we must select a subset of those images to construct a light field. Though several appropriate methods are available for this task, we have empirically determined

simulated annealing [28] to be simple and sufficient. The required number of images is prescribed as well as a minimum desired image separation, and position, orientation and image separation are allowed to converge, yielding a line of *ideal* grid locations s which are a best-fit to *actual* camera locations. During annealing, the vertical (z) distance between the actual and ideal locations is scaled to reflect its limited impact on the resulting images: movement in the vertical direction results in a maximum projected translation which is inversely proportional to the focal length of the camera.

B. Image Reparameterization

Having selected a set of camera poses close to an ideal line of samples, s , we must now reproject them onto a common u, v plane. We present a simplified approach then generalize, in all cases starting with rectification of the input images [29].

Starting with the assumptions of an approximately planar sea floor and a downward-facing camera moving parallel to it, we choose a u, v reference plane coincident with the sea floor and centered horizontally on the grid of camera positions, s . Because a downward facing camera has an image plane parallel to this u, v plane, reprojection of the images to common u, v coordinates can be carried out as a combination of 2D rotation and translation within the image plane. Rotation brings the horizontal pixel coordinate u into alignment with the line of positions, s , and translation aligns the images into a common u, v frame. To determine the appropriate transformations, the center image in s is first rotated to align its u with s based on the camera pose estimate. Next, the remaining images are brought into alignment with the center image based on statistically robust registration [6], ensuring immunity to distractors and small deviations from an ideally planar scene.

In the more general case of a non-planar scene which is not parallel to the direction of motion of the camera, and a non-downward-facing camera, we cannot always accomplish image registration using orthonormal transformations. Our approach in this case begins by reprojecting each image into a downward-facing pose, with aligned s and u axes, based on the AUV's pose estimates. We then perform robust registration of each image with the central image in the grid, using *projective* transformations. Applying these transformations directly would result in a non-parallel, non-orthogonal planar parameterization of the light field, and to correct this we extract and apply a set of best-fit *orthonormal* transformations from the projective transformations. This process guarantees a parallel and centered u, v reference plane, at a depth somewhere near the mean depth of the projected scene, and also corrects for inaccuracies in the estimated camera poses.

As a final step, the light field must be cropped in u, v such that only areas visible in all (or most) images along s remain. The reparameterization process can be summarized as:

- 1) Rectify and reproject images, rotating to align u with s
- 2) Find projective homographies with central image
- 3) Find and apply best-fit orthonormal transformations
- 4) Crop in u, v

IV. ISOLATING BACKGROUND AND DISTRACTORS

In [22] it was shown that an occlusion-free, Lambertian scene lying between two depths z_{min} and z_{max} will yield a light field whose frequency-domain region of support (ROS) is given by a dual-fan shape. For the Fourier transform of the light field, $L_{freq}(\Omega_s, \Omega_t, \Omega_u, \Omega_v)$, the ROS is given by

$$\begin{bmatrix} \Omega_s \\ \Omega_t \end{bmatrix} = \left(\frac{d}{p_z} - 1 \right) \begin{bmatrix} \Omega_u \\ \Omega_v \end{bmatrix}, \quad z_{min} \leq p_z \leq z_{max}, \quad (1)$$

where p_z is the depth of a point in the scene. This means that if the background is known to exist between two depths, we can guarantee that it exists within a fan-shaped ROS in Ω_s, Ω_u . This fan-shaped ROS arises because of smooth apparent motion of scene elements as the camera moves along s : elements are constrained to move in the same direction, and within a range of rates determined by the scene's geometry. This can be exploited to filter for a range of depths using a filter with a fan-shaped passband.

Because the images which make up our light field were taken at different times, moving elements within them will, in general, not conform to the smooth motion associated with background elements. In the case of a hovering AUV, in particular, images are typically distantly separated in time, so distractors display impulse-like behaviour along s : fish appear and disappear in adjacent images, introducing very high frequency components well outside the fan-shaped ROS of the background elements. For sequences with more slowly moving distractors, the frequency components may approach the fan shape, but are unlikely to exactly match those of a valid background element. Note that these observations are similar to those made in [16] and [18], but can be exploited much more simply because we have taken the additional step of reparameterizing the problem into a light field structure.

Note that occlusions and non-Lambertian (e.g. specular) scene elements may have frequency components outside the ideal fan-shaped ROS. Fortunately, most of the energy of interest in typical underwater scenes is both non-occluding and Lambertian, and indeed removal of specular reflections from these scenes may be desirable, depending on the application.

A. Fan Filter

In [22] filter banks are proposed for approximating a fan-shaped passband. Such a filter may be applied to the problem at hand by appropriately selecting filter coefficients such that the passband admits background elements within the depth range defined by z_{min} and z_{max} . The appropriate depth range will vary by the type of terrain in the scene, though a good starting point is to constrain the depth to lie between half and twice the nominal AUV altitude. A narrower depth range will be more selective to distractors, but will also attenuate any scene elements which violate the depth range.

For the purposes of this investigation, we implemented a simpler filter directly in the frequency domain. The light field is processed in slices in s, u , first taking a 2D discrete Fourier transform, then multiplying by a fan-shaped passband, and

finally taking the inverse discrete Fourier transform. The fan-shaped passband is designed directly in the frequency domain as the combination of an ideal fan passband and an edge-softening Gaussian, chosen to reduce ringing associated with discontinuities in the frequency domain. The variance of the Gaussian is increased along Ω_u , so as to maximize selectivity at low frequencies, and to further soften the filter at higher frequencies. The phase response is left at zero throughout. Though simplistic, this approach has yielded acceptable results. More complex schemes, such as those proposed in [30] and [31], may yield more exact results.

Because there are relatively few samples along s , it is necessary to introduce measures to deal with edge effects. In particular, the circular nature of the Fourier transform means images get blurred off one edge of the s dimension onto the opposite edge. To counteract this, zero-padding is introduced in the s dimension, though this introduces darkening near the edges. The darkening is partially dealt with by normalizing each slice along s to a consistent mean and standard deviation.

B. Inverse Fan Filter

As a trivial extension of the distractor-rejecting fan filter, a distractor-extracting filter can be formulated by utilizing the inverse fan filter. This is accomplished by applying one minus the magnitude response in the frequency domain.

V. RESULTS

The University of Sydney's Australian Centre for Field Robotics operates an ocean going AUV called Sirius capable of undertaking high resolution survey work. A sequence of 1853 geo-referenced images was collected during a station-keeping mission over the course of about half an hour, with position estimates obtained through simultaneous localization and mapping (SLAM) [1]. The navigation estimates of the camera's position are shown in a top-down view in Fig. 2 as grey dots. Over the sequence, the AUV drifted on the order of half a meter in each direction.

As described in Section III-A, simulated annealing was used to select 30 approximately co-linear camera positions, shown as dark circles in the figure. This yielded a best-fit grid of ideal positions, s , of length 0.2 m – the mean and worst-case error between the camera's estimated position and the ideal grid locations were 4 mm and 6.8 mm, respectively. Note that this error includes depth (not shown in the figure), and though the individual position errors seem high, the overall shape of the array is close enough to ideal for the filters to operate. Errors in position manifest themselves in the filtered scene model as a slight blurring of parallax motion, and in the isolated distractor output as ghosting of background elements.

The 30 selected images were reparameterized into a light field by registering the images using rotation and translation transformations, as described in Section III-B. The average altitude of the AUV over the 30 images was 1.98 m, and this determined the light field plane separation, d . The images, after cropping, contained 990 and 680 pixels in u and v , respectively. A fan filter was constructed with a passband

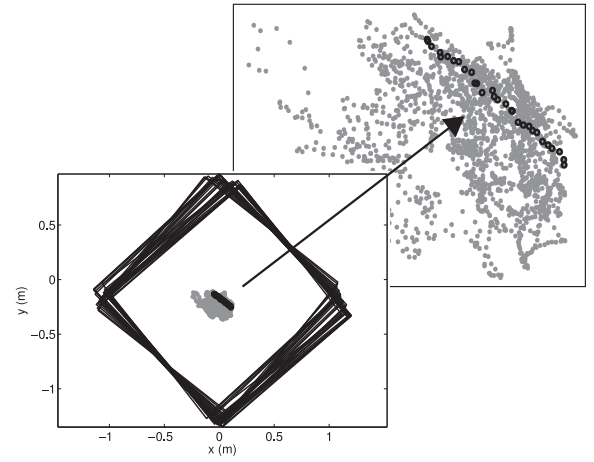


Fig. 2. A set of 30 camera poses (black circles) selected from a sequence of 1853 poses captured by a station-keeping AUV (grey dots); approximate image footprints are shown as black boxes

admitting elements between depths of 1.8 and 2.2 m – these were empirically chosen to closely conform to the geometry of the scene.

Fig. 3 depicts the results of the filtering operation: 3a shows a single input frame, taken near the center of s , and 3b shows a detailed zoom on a central region of the image. The inverse fan filter and fan filter outputs for this same central frame are shown in Figs. 3c and 3d, respectively. The inverse fan filter reveals two fish hiding in this frame – 30 such images were produced, one per s position, containing a total of 41 images of fish, though of course some are the same fish in different positions. The fan filter output is a distractor-free, 3D light field model of the background. A short video of these results is included as a video attachment to this paper.¹

By comparison, a simple pixel differencing scheme is unable to distinguish between changes caused by parallax and those caused by distractors. This is seen by comparing the inverse fan filter output with pixel differencing results for two of the images along s , shown in Figs. 3e and 3f, respectively.

Fig. 4 depicts the filtering results as slices in s and u : the horizontal axis represents samples along s – i.e. camera motion – while the vertical axis represents u – i.e. image pixels – for a particular v chosen to contain a distractor (seen near the top of the figure). The loss of contrast near the edges of the fan filter output is due to filtering edge effects, which increase in magnitude with the selectivity of the filter.

To investigate the effects of image separation on parallax components, we found the energy in the difference between the first image in s and each subsequent image, as shown in Fig. 5. As anticipated, parallax energy rises as image separation increases, and we believe that any 2D per-pixel method will suffer from this limitation. For comparison, the energy in the unnormalized fan filter output is shown – note that the increase in energy near the edge samples is due to filtering edge effects, not sensitivity to parallax.

¹Available at <http://www.youtube.com/watch?v=7IltUPFo3Ew>.

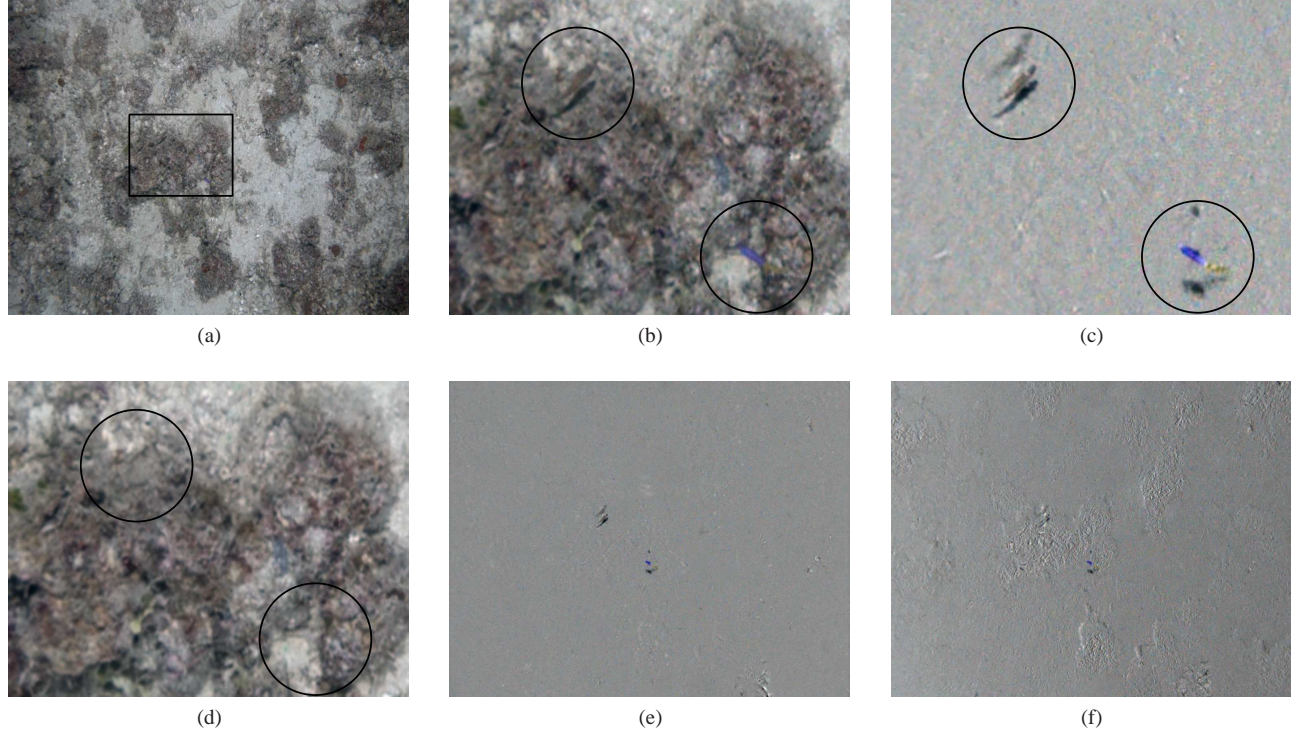


Fig. 3. a) A frame of the input light field taken near the center of s , and b) a zoomed in detail near the center of the image; c) isolated distractors, one of 30 such images produced by the inverse fan filter (note the shadows cast by the AUV's two strobes), and d) a frame of the distractor-free light field model produced by the fan filter; e) a full frame of the isolated distractors for comparison with f) simple pixel differencing, which shows higher sensitivity to parallax in the scene

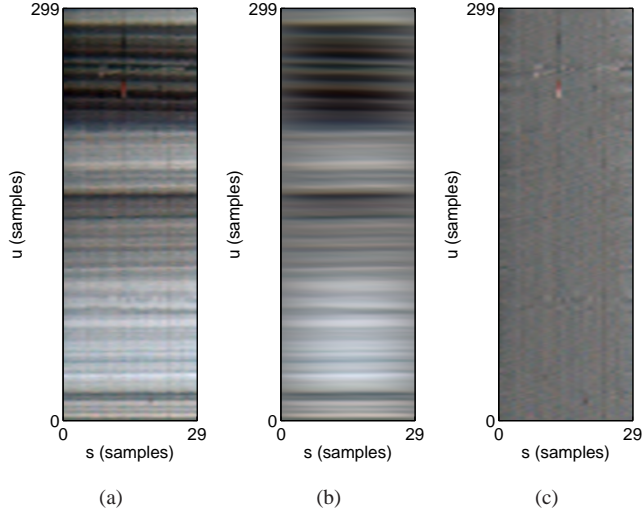


Fig. 4. A slice, in the s, u plane, of a) the input light field, b) the fan-filtered, distractor-free light field, and c) the inverse-fan-filtered extracted distractors; images have been cropped in u to show detail

VI. CONCLUSIONS AND FUTURE WORK

We have demonstrated a technique for building a 3D light field from monocular image sequences obtained from a station-

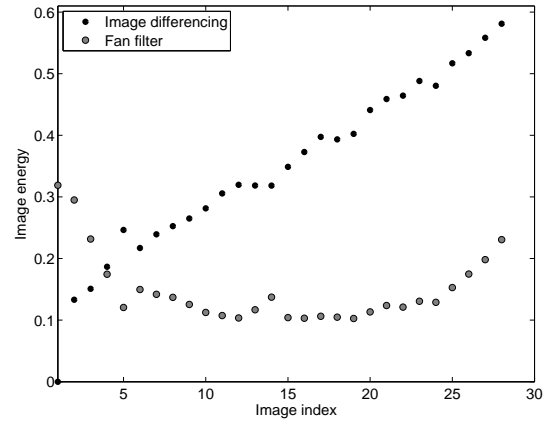


Fig. 5. Energy content from image differencing increases with image separation (black), driven by an increase in parallax; this effect is not seen in the inverse fan filter output (grey)

keeping or slowly moving AUV. By applying linear fan filters and inverse fan filters to these 3D light fields, we are able to construct a distractor-free light field model of the background and, conversely, images of the isolated distractors. Beyond the actual light field formation process, which relies on feature extraction for image registration, our technique is featureless – distractors and background are isolated using entirely linear, pixel-wise operations. We have shown our results to outper-

form 2D per-pixel methods, which are sensitive to parallax motion of background elements. The fan filters presented are conceptually and computationally simpler than the feature-based and nonlinear counterparts discussed in Section II. They operate in constant time independent of scene complexity, are suitable for parallelization, and are expected to show better noise performance than feature-based methods because they use all measured pixels in forming a solution.

As future work, generalization to a 4D light field would allow processing of all images from a hovering AUV. Applying the fan filter to sequences suffering from caustics and other illumination effects would be interesting, as would be demonstration of the technique for long-term distractor-free change detection. Finally, the presented techniques might find application outside those named – aerial surveillance or agricultural monitoring, for example.

ACKNOWLEDGMENT

This work is supported in part by the ARC Centre of Excellence programme, funded by the Australian Research Council (ARC) and the New South Wales State Government, the Australian Centre for Field Robotics, The University of Sydney, the Australian Government's International Postgraduate Research Scholarship (IPRS), and the Australian Institute of Marine Science (AIMS). The authors would also like to thank Lachlan Toohey for suggesting the inverse fan filter, and the ACFR underwater group for inspiration and guidance.

REFERENCES

- [1] S. B. Williams, O. Pizarro, I. Mahon, and M. Johnson-Roberson, "Simultaneous Localisation and Mapping and Dense Stereoscopic Seafloor Reconstruction Using an AUV," in *Experimental Robotics*, ser. Springer Tracts in Advanced Robotics, O. Khatib, V. Kumar, and G. Pappas, Eds. Springer Berlin / Heidelberg, 2009, vol. 54, pp. 407–416.
- [2] M. Grasmueck, G. P. Eberli, D. A. Viggiano, T. Correa, G. Rathwell, and J. Luo, "Autonomous underwater vehicle (AUV) mapping reveals coral mound distribution, morphology, and oceanography in deep water of the Straits of Florida," *Geophysical Research Letters*, vol. 33, p. L23616, Dec 2006.
- [3] C. Kunz, C. Murphy, R. Camilli, H. Singh, R. Eustice, C. Roman, M. Jakuba, C. Willis, T. Sato, K. Nakamura, R. Sohn, and J. Bailey, "Deep sea underwater robotic exploration in the ice-covered arctic ocean with AUVs," in *IROS*, 2008, pp. 3654–3660.
- [4] S. B. Williams, O. Pizarro, M. Jakuba, I. J. Mahon, S. D. Ling, and C. R. Johnson, "Repeated AUV Surveying of Urchin Barrens in North Eastern Tasmania," in *Proceedings of the 2010 IEEE international conference on Robotics and Automation*. IEEE, May 2010, pp. 293–299.
- [5] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 31–42.
- [6] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [7] R. Koch, M. Pollefeys, and L. Van Gool, "Robust calibration and 3D geometric modeling from large collections of uncalibrated images," in *Proc. DAGM*, vol. 99, 1999, pp. 413–420.
- [8] M. Piccardi, "Background subtraction techniques: a review," in *IEEE International Conference on Systems, Man and Cybernetics*, vol. 4, 2004, pp. 3099–3104.
- [9] S. Chien, S. Ma, and L. Chen, "Efficient moving object segmentation algorithm using background registration technique," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 7, pp. 577–586, 2002.
- [10] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2. IEEE, 1999, pp. 246–252.
- [11] E. Hayman and J. Eklundh, "Statistical background subtraction for a mobile observer," in *Ninth IEEE International Conference on Computer Vision*, 2003. *Proceedings*, 2003, pp. 67–74.
- [12] A. Mittal and D. Huttenlocher, "Scene modeling for wide area surveillance and image synthesis," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2. IEEE, 2000, pp. 160–167.
- [13] Y. Ren, C. Chua, and Y. Ho, "Statistical background modeling for non-stationary camera," *Pattern Recognition Letters*, vol. 24, no. 1-3, pp. 183–196, 2003.
- [14] R. Pless, T. Brodsky, and Y. Aloimonos, "Detecting independent motion: the statistics of temporal continuity," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 8, pp. 768–773, Aug. 2000.
- [15] A. Ogale, C. Fermuller, and Y. Aloimonos, "Motion segmentation using occlusions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 6, pp. 988–992, June 2005.
- [16] R. Feghali and A. Mitiche, "Spatiotemporal motion boundary detection and motion boundary velocity estimation for tracking moving objects with a moving camera: a level sets PDEs approach with concurrent camera motion compensation," *Image Processing, IEEE Transactions on*, vol. 13, no. 11, pp. 1473–1490, Nov. 2004.
- [17] Y. Sheikh, O. Javed, and T. Kanade, "Background subtraction for freely moving cameras," in *Computer Vision, 2009 IEEE 12th International Conference on*, Sept. 2009, pp. 1219–1225.
- [18] R. C. Nelson, "Qualitative detection of motion by a moving observer," *International Journal of Computer Vision*, vol. 7, pp. 33–46, 1991.
- [19] D. Dansereau and L. Bruton, "Gradient-based depth estimation from 4D light fields," in *Proceedings of the International Symposium on Circuits and Systems*, vol. 3. IEEE, May 2004, pp. 549–552.
- [20] W. Hong, "Light field applications to 3-dimensional surface imaging," Master's thesis, Massachusetts Institute of Technology, 2009.
- [21] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, vol. 2, 2005.
- [22] D. Dansereau and L. T. Bruton, "A 4-D Dual-Fan Filter Bank for Depth Filtering in Light Fields," *IEEE Transactions on Signal Processing*, vol. 55, no. 2, pp. 542–549, 2007.
- [23] M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch, "Visual modeling with a hand-held camera," *International Journal of Computer Vision*, vol. 59, no. 3, pp. 207–232, 2004.
- [24] B. Heigl, R. Koch, M. Pollefeys, J. Denzler, and L. Gool, "Plenoptic modeling and rendering from image sequences taken by hand-held camera," in *Mustererkennung 1999, 21. DAGM-Symposium*. Springer-Verlag, 1999, pp. 94–101.
- [25] R. Koch, M. Pollefeys, L. Van Gool, B. Heigl, and H. Niemann, "Calibration of hand-held camera sequences for plenoptic modeling," in *ICCV*, vol. 1. IEEE, 1999, pp. 585–591.
- [26] P. Sharma, A. Parashar, S. Banerjee, and P. Kalra, "An uncalibrated lightfield acquisition system," *Image and Vision Computing*, vol. 22, no. 14, pp. 1197–1202, 2004.
- [27] C. Vogelgsang, B. Heigl, G. Greiner, and H. Niemann, "Automatic image-based scene model acquisition and visualization," *Vision, modeling, and visualization 2000: proceedings: November 22-24, 2000, Saarbrücken, Germany*, pp. 188–198, 2000.
- [28] S. Kirkpatrick, C. Gelatt Jr, and M. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.
- [29] J. Heikkilä and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 1997, pp. 1106–1112.
- [30] A. Cetin, O. Gerek, and Y. Yardimci, "Equiripple FIR filter design by the FFT algorithm," *IEEE Signal Processing Magazine*, vol. 14, no. 2, pp. 60–64, 1997.
- [31] R. Ansari, "Efficient IIR and FIR fan filters," *IEEE Transactions on circuits and systems*, vol. 34, no. 8, pp. 941–945, 1987.