IROS 2015 Workshop on Alternative Sensing for Robot Perception (WASRoP)

Closed-Form Change Detection from Moving Light Field Cameras

Donald G. Dansereau*, Stefan B. Williams†, and Peter I. Corke*
*ARC Centre of Excellence for Robotic Vision, Queensland University of Technology
Brisbane QLD 4000, Australia {donald.dansereau, peter.corke}@qut.edu.au www.roboticvision.org
†Australian Centre for Field Robotics; School of Aerospace, Mechanical and Mechatronic Engineering
University of Sydney, NSW, Australia s.williams@acfr.usyd.edu.au

***Abstract*—We present a novel approach to change detection from a moving camera, a task normally complicated by the nonuniform apparent motion of complex 3D environments. By combining methods for visual odometry and rendering from light fields, we derive what is, to our knowledge, the first closed-form approach to 3D change detection. No depth estimation is required – apparent motion is disregarded by exploiting the geometry implicitly encoded by the light field. The method has constant runtime and is suitable for real-time implementation on parallel hardware including FPGA and GPU. We show qualitative and quantitative results for Lytro-captured imagery, with our method significantly outperforming a naive 2D per-pixel method. We discuss a generalization of the derivation to other traditionally still-camera-only methods.**

I. Introduction

Having a static camera simplifies a wide range of computer vision problems: change / motion detection, object tracking, segmentation, isolation and removal, and a range of spatio-temporal filtering techniques [1]–[4]. If the camera is moving, however, nonuniform apparent motion complicates matters, necessitating costly structure-from motion approaches.

We show that light field cameras [5], [6] offer a simplification by allowing a virtual, stationary camera to be rendered from a dynamic light field sequence. This effectively reduces moving-camera problems to stationary-camera problems. The process begins by forming an estimate of camera motion, which is then used to render a novel view from a virtual, stationary camera.

This framework, depicted in Fig. 1, can be applied to generalize a wide range of classically still-camera methods, but we specifically demonstrate motion detection implemented as simple frame differencing. By employing closed-form solutions for each step of the framework, we arrive at the surprising result that change detection can be carried out with a single closed-form expression. To our knowledge this is the first closed-form solution to change detection from a moving camera in a 3D environment.

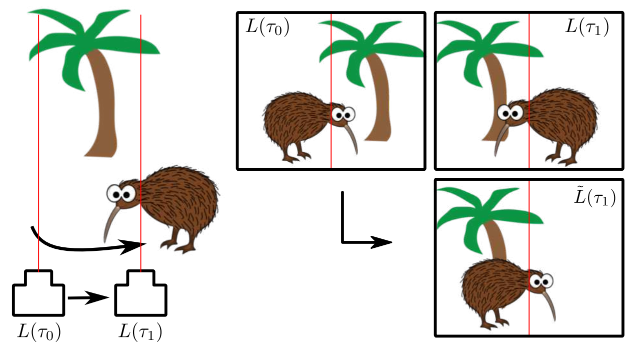

Fig. 1. Camera motion between times τ_0 and τ_1 (left) causes apparent motion in static scene elements like the tree (top insets), making them difficult to disambiguate from genuinely dynamic elements, like the Kiwi. We render a novel view $\tilde{L}(\tau_1)$ showing scene content from time τ_0 as seen from the point of view of the camera at τ_1 (bottom right). Static elements now appear static, opening a family of dynamic-camera problems to static-camera solutions. In the case of change detection, this process yields a closed-form solution.

II. Related Work

Change detection from dynamic platforms is nontrivial due to the apparent motion of the environment. This apparent motion is nonuniform in the case of non-planar 3D scene geometry, and so methods based on pixel-level statistics are insufficient.

For camera motion restricted to rotation, the lack of parallax has been exploited to simplify change detection [7]–[9]. 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 apply [10].

In the case of a freely moving camera and nontrivial scene geometry, more complex methods have been proposed, based on occlusion detection, or iterative camera motion and boundary estimation [11], [12]. Other approaches yield non-dense results or rely on feature extraction [13], [14].

Our method requires no feature tracking, no explicit 3D scene model is formed, no iterative optimization is required, and a dense result is produced directly. This is behaviourally and computationally simpler than existing methods, yields results in constant runtime, and is suitable for parallel implementations including FPGA and GPU.

This work builds on the concept of plenoptic flow [15], introducing a framework for simplifying moving-camera problems, deriving closed-form rendering from plenoptic flow, and providing a simple closed-form expression for change detection. A more detailed treatment can be found in [16].

III. Change Detection for Moving Cameras

A. Closed-Form Camera Motion Estimation

Plenoptic flow and its precursors were first introduced to estimate camera motion [15]–[17]. This operates much like motion estimation from 2D optical flow, but generalizing to six-degree-of-freedom motion.

The equation of plenoptic flow expresses the temporal light field derivative L_τ in terms of the spatial and angular derivatives L_s, L_t, L_u and L_v, and the camera's translation (q_x, q_y, q_z) and rotation (w_x, w_y, w_z). Partial derivatives are estimated using the first difference. The resulting linear system can be written compactly as

$$\boldsymbol{A}\boldsymbol{v} = L_\tau, \tag{1}$$

where $\boldsymbol{A}$ is constructed from the spatial and angular derivatives, and $\boldsymbol{v}$ is the change in pose, concatenating translation and rotation. A least-squares solution to this overdetermined system yields an estimate of the camera's motion $\tilde{\boldsymbol{v}}$ [15], [18]. In the following sections we will use this motion estimate to render a novel view which aligns two input light fields.

B. Closed-Form Rendering

Each of the columns of the matrix $\boldsymbol{A}$ can be interpreted as the change in the light field in response to one of six separate motion components. We will refer to these components as $L_x, L_y, L_z, L_{\omega x}, L_{\omega y}$ and $L_{\omega z}$, respectively. One of the immediate applications of this decomposition is that novel views can now be synthesized via the weighted addition of these six motion components to the original light field. Note that this approach breaks down for large camera motions and where non-Lambertian behaviour, e.g. specular reflection and occlusion, dominate.

C. Rendering a Stationary Virtual Camera

Given two frames, we begin by finding the least squares solution to the equation of plenoptic flow (1), yielding an estimate of the camera's motion $\tilde{\boldsymbol{v}}$. We then render a novel stationary view using the additive method described in the previous section. Again the equation of plenoptic flow gives us the tool to do this, by allowing us to derive the temporal derivative due to the estimated camera motion:

$$\tilde{L}_\tau = \boldsymbol{A}\tilde{\boldsymbol{v}}. \tag{2}$$

Rendering the light field measured at time τ_0 as though viewed from the camera's position at time τ_1 can be accomplished by adding

$$\tilde{L}(\tau_1) = L(\tau_0) + \tilde{L}_\tau. \tag{3}$$

D. Closed-Form Change Detection

Finally, we effect change detection through pixel differencing between the measured frame $L(\tau_1)$ and the estimated stationary frame $\tilde{L}(\tau_1)$. By substituting (3) and from the definition of the temporal derivative, we find

$$\boldsymbol{R} = L(\tau_1) - \tilde{L}(\tau_1) = L_\tau - \tilde{L}_\tau. \tag{4}$$

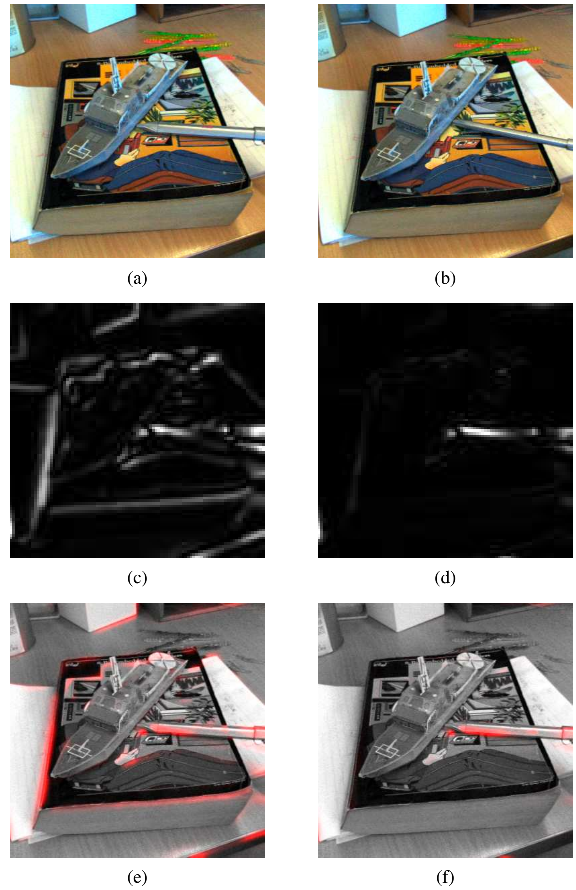

Fig. 2. Two frames (top) showing both apparent motion and a dynamic scene element. The temporal derivative (c) represents a naive pixel-differencing approach; the plenoptic residual (d) shows significantly less sensitivity to apparent motion while retaining dynamic elements. The first input frame is highlighted using each of these results (bottom). Notice that the pen rotated about its center, thus the pattern of decreasing velocity near its pivot.

In other words, the result of pixel differencing using this method simplifies to the residual error in the equation of plenoptic flow. This is a satisfying result, in that dynamic objects break the rules underlying plenoptic flow, appearing as areas of high error in the residual. It is also surprising, in that it shows that a conventionally complex task, change detection from a moving platform in a 3D environment, has a closed-form solution.

IV. Experiments

We applied the method of plenoptic residuals to pairs of images captured using a Lytro consumer-grade plenoptic camera. The camera was calibrated and imagery rectified using the MATLAB Light Field Toolbox [19]. We applied a numerically stable form of plenoptic flow, including the method for directly estimating derivatives from rectified light field imagery described in Section 5.3.1 of [16]. Finally, we computed the plenoptic residual (4) to build a map highlighting dynamic scene elements.

The top row of Figure 2 shows two input frames with a small inter-frame camera motion and a single dynamic scene

TABLE I
NAIVE PIXEL DIFFERENCE L_τ VS. PLENOPTIC RESIDUAL R

Scene	L_τ (dB)	R (dB)	Ratio (dB)
Jar	-31.81	-35.848	4.0386
Jar	-27.634	-31.029	3.3954
Jar	-36.452	-43.197	6.7448
Pen	-23.805	-28.842	5.037
Pen	-34.679	-39.917	5.2385
Toothpicks	-33.064	-33.55	0.48605
Toothpicks	-30.576	-32.087	1.5104
Toothpicks	-39.247	-42.276	3.0284
Mean	-29.684	-33.439	4.0905

element. The center row shows the magnitude of the difference between frames L_τ as computed after band-limiting for plenoptic flow (left), and the plenoptic residual $\boldsymbol{R}$ (right). The bottom row highlights dynamic scene elements in red using L_τ and $\boldsymbol{R}$. The results in Figures 2(c) and (e), representative of naive pixel differencing, show significant sensitivity to apparent motion. Though imperfect, the plenoptic residual results in Figures 2(d) and (f) show significant attenuation of apparent motion, while retaining those elements showing genuine motion within the scene.

Table I summarizes the signal energy resulting from naive pixel differencing and the method of plenoptic residuals, and their ratio. Values are shown for eight pairs of images from three test scenes. The tabulated values represent signal energy expressed in dB, for input light fields normalized to a peak value of one. The mean ratio of 4 dB establishes that the plenoptic residuals method is more than twice as selective as naive pixel differencing.

V. DISCUSSION AND FUTURE DIRECTIONS

We presented a general approach for converting moving-camera problems into stationary-camera problems. By effecting both camera motion estimation and rendering using closed-form plenoptic flow, we derived a closed-form method for change detection from moving platforms.

We showed the proposed method to outperform the naive 2D per-pixel method, which is sensitive to nonuniform apparent motion of the scene. The presented method copes with nonuniform apparent motion without requiring depth estimation or other complex scene modelling – apparent motion is disregarded by exploiting the geometric information implicitly encoded by the light field.

The proposed method is behaviourally and computationally simpler than previously explored feature-based and nonlinear methods. It operates in constant time independent of scene complexity, and is suitable for parallel implementations including GPU or FPGA.

As future work we envision demonstrating the technique on light field video sequences taken from dynamic platforms in complex environments. Establishing ground-truth data will allow a more formal quantitative analysis of the method's performance.

Finally, we believe our approach can be generalized to other forms of video processing which benefit from a static camera, e.g. noise reduction or object segmentation [2]. The potential for extension to temporally recursive filtering is especially interesting.

VI. ACKNOWLEDGMENTS

This work was supported by the Australian Centre for Field Robotics and the Australian Research Council Centre of Excellence for Robotic Vision (project number CE140100016).

REFERENCES

[1] M. Piccardi, "Background subtraction techniques: a review," in *IEEE Intl. Conference on Systems, Man and Cybernetics*, vol. 4, 2004, pp. 3099–3104.

[2] 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.

[3] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," in *Computer Vision and Pattern Recognition (CVPR)*, vol. 2. IEEE, 1999, pp. 246–252.

[4] S. Schauland, J. Velten, and A. Kummert, "Detection of moving objects in image sequences using 3D velocity filters," *Intl. Journal of Applied Mathematics and Computer Science*, vol. 18, no. 1, pp. 21–31, 2008.

[5] M. Levoy and P. Hanrahan, "Light field rendering," in *SIGGRAPH*. ACM, 1996, pp. 31–42.

[6] 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.

[7] E. Hayman and J. Eklundh, "Statistical background subtraction for a mobile observer," in *Intl. Conference on Computer Vision (ICCV)*, 2003, pp. 67–74.

[8] A. Mittal and D. Huttenlocher, "Scene modeling for wide area surveillance and image synthesis," in *Computer Vision and Pattern Recognition (CVPR)*, vol. 2. IEEE, 2000, pp. 160–167.

[9] 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.

[10] R. Pless, T. Brodsky, and Y. Aloimonos, "Detecting independent motion: the statistics of temporal continuity," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 22, no. 8, pp. 768–773, Aug. 2000.

[11] A. Ogale, C. Fermuller, and Y. Aloimonos, "Motion segmentation using occlusions," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 27, no. 6, pp. 988–992, Jun. 2005.

[12] 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," *IEEE Transactions on Image Processing (TIP)*, vol. 13, no. 11, pp. 1473–1490, Nov. 2004.

[13] R. C. Nelson, "Qualitative detection of motion by a moving observer," *Intl. Journal of Computer Vision (IJCV)*, vol. 7, pp. 33–46, 1991.

[14] S. Dey, V. Reilly, I. Saleemi, and M. Shah, "Detection of independently moving objects in non-planar scenes via multi-frame monocular epipolar constraint," in *European Conference on Computer Vision (ECCV)*, Oct. 2012.

[15] D. G. Dansereau, I. Mahon, O. Pizarro, and S. B. Williams, "Plenoptic flow: Closed-form visual odometry for light field cameras," in *Intelligent Robots and Systems (IROS)*. IEEE, Sept 2011, pp. 4455–4462.

[16] D. G. Dansereau, "Plenoptic signal processing for robust vision in field robotics," Ph.D. dissertation, Australian Centre for Field Robotics, School of Aerospace, Mechanical and Mechatronic Engineering, The University of Sydney, Jan. 2014.

[17] J. Neumann, C. Fermuller, Y. Aloimonos, and V. Brajovic, "Compound eye sensor for 3D ego motion estimation," in *Intelligent Robots and Systems (IROS)*, vol. 4. IEEE, 2005, pp. 3712–3717.

[18] F. Dong, S.-H. Ieng, X. Savatier, R. Etienne-Cummings, and R. Benosman, "Plenoptic cameras in real-time robotics," *The Intl. Journal of Robotics Research*, vol. 32, no. 2, pp. 206–217, 2013.

[19] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jun. 2013, pp. 1027–1034.