Exploiting Parallax in Panoramic Capture to Construct Light Fields

Donald G. Dansereau[1*], David Wood[2], Sebastian Montabone[2], Stefan B. Williams[1]
[1]Australian Centre for Field Robotics; Aerospace, Mechanical and Mechatronic Engineering
University of Sydney, NSW, Australia {d.dansereau; s.williams}@acfr.usyd.edu.au
[2]Ocular Robotics, Sydney, Australia {d.wood; s.montabone}@ocularrobotics.com

Abstract

We show that the parallax motion resulting from non-nodal rotation in panorama capture can be exploited for light field construction from commodity hardware. Automated panoramic image capture typically seeks to rotate a camera exactly about its nodal point, for which no parallax motion is observed. This can be difficult or impossible to achieve due to limitations of the mounting or optical systems, and consequently a wide range of captured panoramas suffer from parallax between images. We show that by capturing such imagery over a regular grid of camera poses, then appropriately transforming the captured imagery to a common parameterisation, a light field can be constructed. The resulting four-dimensional image encodes scene geometry as well as texture, allowing an increasingly rich range of light field processing techniques to be applied. Employing an Ocular Robotics REV25 camera pointing system, we demonstrate light field capture, refocusing and low-light image enhancement.

1 Introduction

Light field (also "plenoptic") imaging and signal processing are emerging as powerful alternatives to conventional imaging. By capturing a rich four-dimensional representation of the light permeating a scene, light field imagery encodes three-dimensional scene geometry as well as texture. Recent advances in light field processing have spanned enhanced low-light imaging [Dansereau *et al.*, 2013], refocussing [Fiss *et al.*, 2014], labeling [Wanner *et al.*, 2013], video stabilisation [Smith *et al.*, 2010] and depth estimation [Bishop and Favaro, 2011].

Light field capture is typically carried out using either multiple-camera arrays [Wilburn *et al.*, 2005] or through the installation of a lenslet array in the optical path of a conventional camera [Ng *et al.*, 2005]. Gantry-based

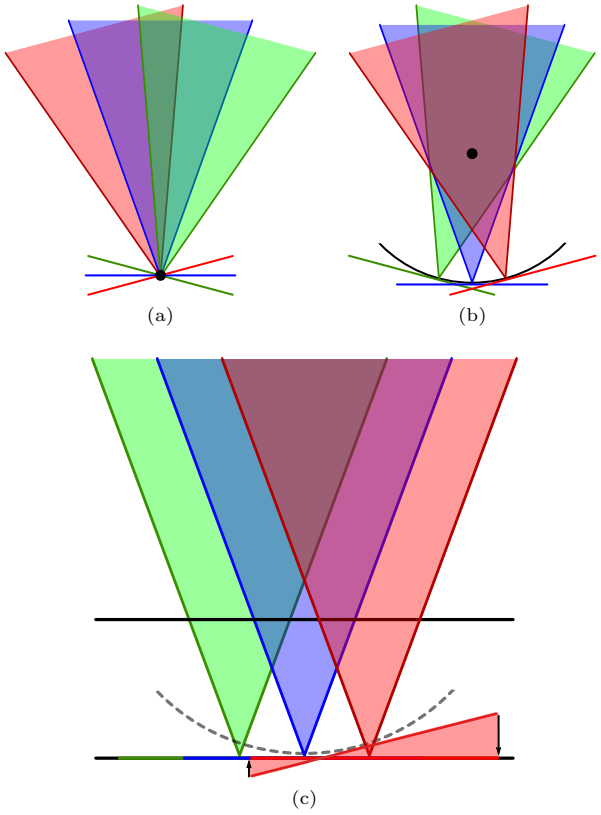


Figure 1: (a) Ideal panoramic rotation is about the nodal point of the camera, resulting in zero translation between images; (b) Even a slight optical misalignment can result in rotation off the nodal point, resulting in translation between frames of the panorama; (c) If images are collected over a regular grid of poses, the resulting frames can be transformed into a light field by reprojecting images into a common two-plane parameterisation – a co-planarity approximation simplifies this process.

systems are also employed, in which a single camera is translated over a regular grid of poses [Dansereau and Bruton, 2003]. All these methods capture the same *form* of information, encoding light rays in terms of both position and direction, in a 2D array of 2D images. Crucial to the measurement of a light field is the *spatial baseline*, the distance between camera poses, or virtual apertures in the case of lenslet-based cameras. This baseline yields parallax motion in the captured imagery which is ultimately the source of depth information in the light field.

Panoramic imaging systems are typically designed to avoid parallax. As such, these devices seek to rotate an optical system about its nodal point – the point for which there is no parallax motion, as depicted in Fig. 1(a). Misalignment in such systems, depicted in Fig. 1(b), yields parallax which is seen as undesirable. In this work we show that the parallax encountered in such systems can be exploited to build light field images, opening the collected imagery to an increasingly wide range of plenoptic processing algorithms. Though focused on panoramic imaging systems, this idea applies equally well in any multiple-image collection scheme for which parallax is conventionally considered a detriment.

The following section provides background on light field capture and parameterisation as well as the panoramic image capture devices we employ for light field construction. Section 3 describes the reparameterisation process employed to construct the light fields, and Section 4 shows results typical of the method. Finally, Section 5 gives conclusions and indicates directions for future work.

2 Background

2.1 Light Field Parameterisation

A planar grid of cameras is a light field camera. It captures a 2D array of 2D images, which can be parameterised according to the two-plane parameterisation depicted in Fig. 2. Note that each pixel captured by such an array corresponds to a unique ray in the scene, the position and direction of which is entirely described by its points of intersection with the two reference planes.

An s, t plane, by convention coincident with the planar grid of cameras as shown in Fig. 2, describes the position of the camera that captures each image. A u, v plane at distance D describes the orientations of the rays captured in each image. The continuous-domain light field signal $L(s, t, u, v)$ describes all light rays passing through the s, t and u, v planes. Although other parameterisations exist, most recent plenoptic signal processing work deals with two-plane-parameterised light fields.

2.2 Light Field Capture

Two straightforward methods for light field capture are the camera array and the lenslet-based camera. The first

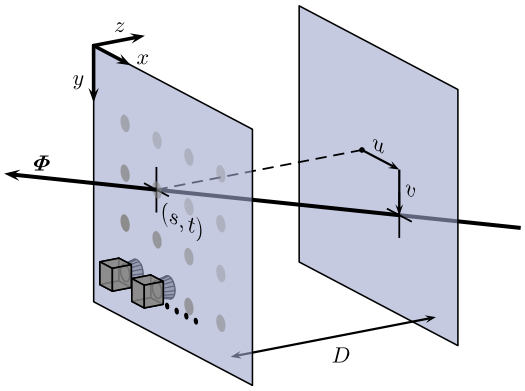


Figure 2: Two-plane parameterisations of light rays: The points of intersection of a ray with two parallel planes completely describes its position and orientation in space.

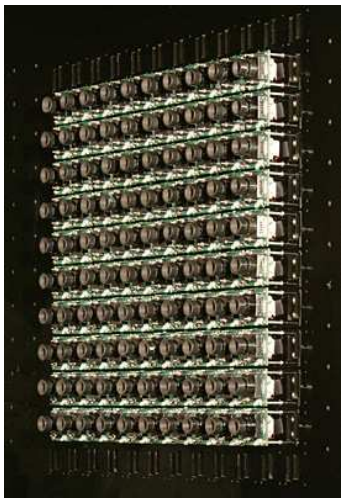


Figure 3: The 100-camera Stanford array[Wilburn *et al.*, 2005].

of these arranges discrete sensors and apertures into a grid, as in the Stanford multi-camera array shown in Fig. 3 [Wilburn *et al.*, 2005]. These systems have the ability to capture extremely high-pixel-count light fields, even from low-resolution base cameras – for example an array of 100 VGA-resolution (640×480 pixel) cameras generates a 30 MPix light field. These systems capture a large amount of information, but require significant outlay for both the camera array and the capture hardware capable of handling the large amounts of data produced.

The lenslet-based approach, following the work of [Adelson and Wang, 2002; Ng *et al.*, 2005; Lumsdaine and Georgiev, 2009] and commercialised by companies such as Lytro and Raytrix GMBH, installs an array of microlenses in the optical path of a conventional, single-sensor camera. Though more compact and cost-effective,

these systems feature shorter baselines and therefore weaker depth resolutions than array-based cameras.

Closely related to the camera array is the gantry-based approach, in which a single camera is translated over a regular grid of poses [Dansereau and Bruton, 2003]. This reduces costs in both camera and acquisition hardware, but presents the disadvantage of capture time, typically requiring on the order of minutes or tens of minutes to capture a light field.

Another approach that has not been well-explored is to use a high-resolution camera system coupled with a high-precision pointing system. By rotating the camera about a point other than the optical centre of the imaging system, parallax information is gathered and forms the basis for light field construction.

This approach has the advantage of yielding high-resolution, multi-megapixel light fields from a single sensor. It also results in an extremely flexible system, with field of view and spatial resolution being on-the-fly reconfigurable to suit the scene and application. There are however three distinct disadvantages to this approach: Firstly the accuracy of camera pointing using traditional pan-tilt or gimbal systems is generally not comparable to the pixel resolution of the image sets being collected. Image arrays captured from these traditional systems therefore require substantial post-processing for image alignment, at an associated computational cost. Poor repeatability of the pointing system means realignment must be performed for each captured light field, and the resulting irregular array spacing can significantly complicate processing and impact performance.

The second issue with the camera pointing approach is the limited baselines possible. Camera pointing devices are typically designed for panoramic capture, generating arrays of images with minimal parallax. An example of such a system is the Gigapan EPIC Pro, shown in Fig. 4. An obvious option for light field capture is to intentionally displace the camera from the centre of rotation of the device, increasing the effective baseline of the capture. This may result in a mechanically imbalanced system, yielding reduced lifetime, increased sensitivity to shock and vibration, and longer capture times due to the increased moment of inertia of the imaging system.

The third issue with camera pointing systems is the time required to capture the image sets. Systems such as the Gigapan EPIC Pro require from ten to twenty minutes to grab a typical panorama of 21 by 11 images. This level of time commitment precludes the use of such devices in many real-world applications, particularly where dynamic scene elements and variations in illumination can corrupt the light field.

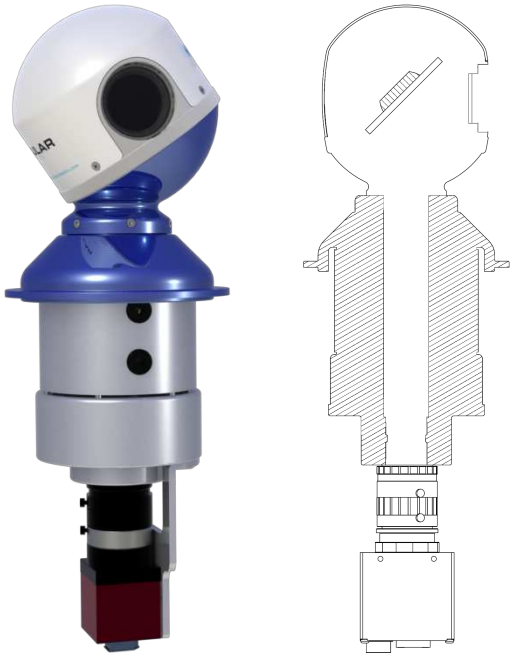
Figure 4: The Gigapan EPIC Pro panorama capture stage for digital SLR cameras (`www.gigapan.com`).

2.3 Steerable Mirror Systems

Although the technique developed in this work apply to panoramic image capture systems in general, we focus in particular on imagery captured using a steerable mirror-based camera pointing device. These devices, by virtue of moving a lightweight mirror rather than the entire camera, can operate at much higher speeds, partially addressing the temporal limitations described above. Furthermore, the stable mechanics of these systems mean significant baselines can be obtained without increased mechanical stress. Finally, their high degree of repeatability simplifies image alignment, requiring only a one-time calibration, rather then per-acquisition recalibration.

For the purposes of this work we employed an Ocular Robotics REV25 camera pointing system, depicted in Fig. 5. This high-performance camera pointing system steers a 25 mm aperture using a "RobotEye" pointing device, effectively steering the field of view of a camera mounted on the back end of the device. The camera we employed was an AVT Prosilica 2450GC-C colour camera, the basic parameters of which are summarised in Table 1. This camera has a pixel count of 5 MPix over a field of view of approximately 30 deg.

The RobotEye pointing system is a high-performance alternative to traditional pan-tilt and gimbal-based optical pointing systems. The performance statistics for the REV25 relevant to this work are shown in Table 2. The primary advantages are in the reporting accuracy of the eye and the array capture speed. In fact, the light field capture rate is practically limited by the integration time of the imaging sensor rather than the movement speed of the pointing system. Even allowing for long exposure



(a) (b)

Figure 5: The Ocular Robotics REV25 system, (a) out of its enclosure and (b) in cutaway, showing the camera, optical path and steerable mirror (`www.ocularrobotics.com`)

Table 1: Relevant parameters of the GC2450-C camera.

Sensor Resolution	2448×2050
Sensor Size	Type 2/3 (2/3 inch)
Cell Size	3.45 µm
Interface	Gigabit Ethernet

times due to poor indoor lighting, arrays on the order of hundreds of images can be captured in tens of seconds.

Poses are reported with an accuracy of 0.01°, simplifying calibration by providing near-optimal initial pose estimates for each of the captured images. Repeatability is sufficiently high that calibration need only be carried out once – poses need only be estimated once for a given grid configuration, greatly simplifying the light field construction process.

Finally, the REV25 system provides an extremely stable offset between the centre of rotation and the optical centre of the camera system. The optical centre effectively takes on a constant distance behind the mirror independent of pointing direction. REV25's with different optical configurations are available to suit various applications. This allows the offset between mirror and optical centre to be specified without impacting the dynamics or stability of the pointing system.

Table 2: Relevant parameters of the REV25 mirror-based camera pointing system.

Peak Aperture Speed	10,000 °/s
Peak Aperture Acceleration	100,000 °/s^2
Pointing Accuracy	±0.05°
Reporting Accuracy	±0.01°
Capture Speed: Arbitrary	10 locations/s max
Capture Speed: Grid	25 locations/s max

3 Light Field Construction

The REV25 system is extremely flexible, and can be programmed to point the camera over arbitrary collections of poses. As a starting point we explore light field construction from images captured over a regular grid of azimuths and elevations. Given each azimuth Az and elevation El, the optics of the system effectively yield images with the following transformations:

$$\phi = \text{-}Az + C, \tag{1}$$

$$\theta = \text{-}El, \tag{2}$$

$$\psi = \text{-}Az, \tag{3}$$

taking $[\phi, \theta, \psi]$ as roll, pitch and yaw, respectively, and C as a constant rotational offset determined by the camera's mounted rotation.

Some of the complexity in what follows is due to the unknown rotation C, which must be compensated for, and our use of a regular sampling of azimuth and elevation, which yields aperture positions that deviate from an ideal grid. We discuss as future work the possibility of optimising the camera's poses to simplify light field construction.

3.1 Planar Approximation

The effective aperture positions lie on a sphere, as depicted in Fig. 1(b), and so a direct approach to light field construction is to begin with a spherical light field parameterization, from which a two-plane light field can be interpolated. However, aperture positions typically occupy narrow subsets of the sampling sphere and are therefore approximately co-planar, as depicted in Fig. 1(c). As such, a simplification may be employed in which the transformation to a planar light field is well described by 2D per-image affine transformations.

This simplification effectively reduces the light field construction process to one of registering each image in the array with the central image. There is a close parallel to earlier work dealing with co-linear sets of images [Dansereau and Williams, 2011], except here the image poses are more tightly controlled, and they occupy a full 2D grid rather than a single co-linear set of poses. As in earlier work, the key issues to address are

establishing the appropriate image transformations, and dealing with misalignment and skew between the resulting light field axes.

3.2 Calibration

Under the planar approximation, each image in the captured grid is individually transformed into registration with the central image. This process is best carried out on a dataset of a distant scene, in order to minimise the impact of parallax on the registration. The high predictability of the REV25 allows a closed-form initial estimate of the per-image registrations to be derived. Conventional image registration techniques are then employed to refine the transformations [Lowe, 2004]. The repeatability of the system means the estimated poses are stable between acquisitions, and calibration need only be carried out once.

3.3 Alignment

Once the images are co-registered, the complication arises that the u, v images are rotated relative to the s, t grid axes, due to the unknown camera rotation C and azimuth-dependent roll term (1). A further complication stems from our use of a regular grid in azimuth and elevation, rather than in pose, resulting in a skewed pose grid. These issues are both illustrated in the Results section in Fig. 6.

To address these two forms of misalignment, both a rotational offset and a degree of grid skew are estimated from the calibrated per-image affine transformations. From these, a rotation is applied to every image in u, v, correcting for the rotational offset and bringing the u, v axes in line with the s, t axes – this is a close parallel to the procedure introduced in [Dansereau and Williams, 2011].

Skew is similarly addressed through 2D per-image transformations, either in s, t or in u, v. For the purposes of this work we compose the skew compensation with the rotation correction in u, v, which is itself composed with the affine transformation bringing the images into a co-registered light field. The result is that only a single per-subimage affine transformation is required which composes registration, rotation and skew correction.

In summary, light field construction consists of two stages: 1) a one-time calibration yielding a single 2D affine transformation per u, v sub-image; and 2) a run-time construction phase which applies these transformations.

4 Results

We employed the REV25 to measure a 7×7 grid of 5 MPix images over the period of about 20 seconds. Because the scene was indoor and poorly lit, the limiting factor on the capture rate was the exposure time of the camera. Starting from initial estimates derived from the optical pointing equations (1–3), we found the optimal affine transformation required to bring each image into registration with the central image in the array. The projected 2D pose estimates corresponding to the estimated affine transformations are visualised in Fig. 6(a).

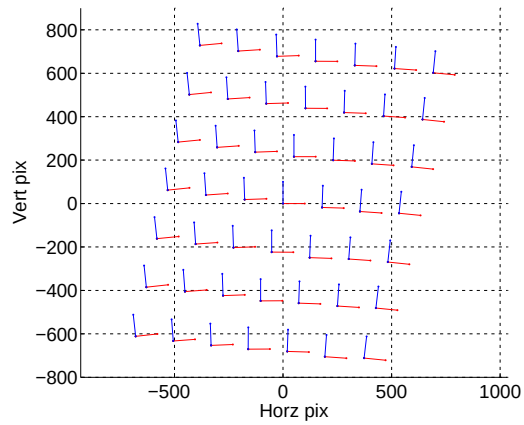
To bring the image axes into alignment with the sampling grid, we estimated the s, t grid alignment from the mean orientation of the best-fit lines for each grid row and column. We then composed the inverse rotation into each affine image transformation, yielding an axis-aligned array of transformations – the resulting projected poses are visualised in Fig. 6(b). Per-subimage rotations include correction for the azimuth-dependent roll term (1).

Note the skew present in this new grid of poses: the horizontal axes are all horizontal, but the vertical axis slopes towards the right. Estimating the skew from the average orientation of the best-fit column and row lines, we composed the inverse skew into the affine transformations, yielding the final set of transformations visualised in Fig. 6(c). Note that the final grid of transformations is a close match to an ideal grid, but it is not perfect due to the small error introduced by the small-angle approximation.

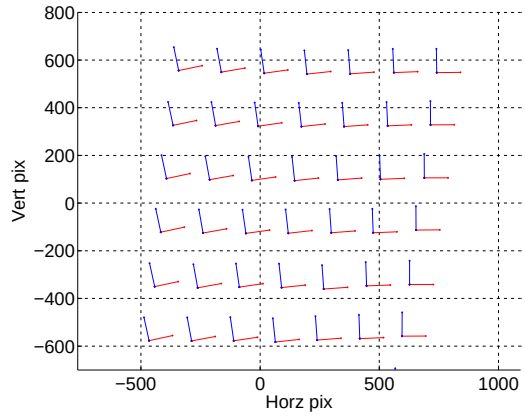
With calibration complete, light field construction is carried out by measuring a set of 7×7 frames and applying the calibrated 2D transformations. Each light field is $7 \times 7 \times 2448 \times 2050$, about 240 MPix, though only the overlapping portions of the image are useful. There are tradeoffs in selecting the grid spacing, field of view, image overlap, and spatial sampling resolution. In the light fields we captured the image overlap was about 80%, and the useful image region was around $7 \times 7 \times 2048 \times 1650$, or about 165 MPix.

Fig. 7(a) shows a slice of the light field in t, v – this is a slice through the printed resolution test pattern in the background, and ribbon cable in the foreground, as seen in Fig. 7(b) – and it is clear from the inset that the difference in slopes between foreground and background elements is on the order of tens of pixels over the seven spatial light field samples. The ribbon cable in the foreground spans about 20 pixels in v, while the background elements span close to zero pixels. This confirms that the light field has captured parallax motion and therefore significant depth information, opening the light field to a wide range of processing techniques including depth estimation and filtering [Bishop and Favaro, 2011; Dansereau *et al.*, 2013].

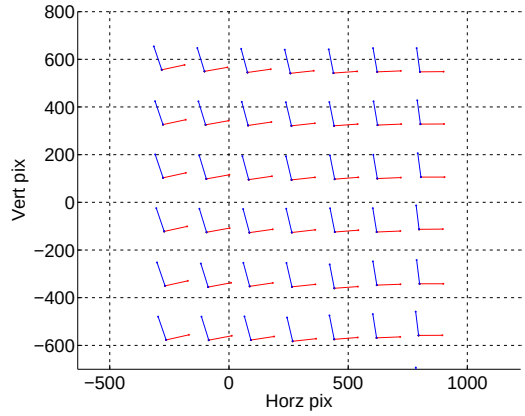
Fig. 7(b) depicts a u, v slice of one of the captured light fields, while (c) and (d) show the output of a simple refocus algorithm based on [Ng *et al.*, 2005], focused on foreground and background scene elements, respectively.



(a) Measured



(b) With rotation compensation



(c) With rotation and skew compensation

Figure 6: A visualisation of the poses of the u, v sub-images based on estimated per-image transformations relative to the central image of the array: (a) The measured sub-images are misaligned with the pose grid; (b) A per-subimage rotation brings them into alignment with the pose grid, but the pose grid is skewed; and (c) Adding a per-subimage skew yields an approximately orthogonal grid of aligned images.

The insets clearly show a significant amount of defocus blur for out-of-focus elements, while desired scene elements stay sharply focused.

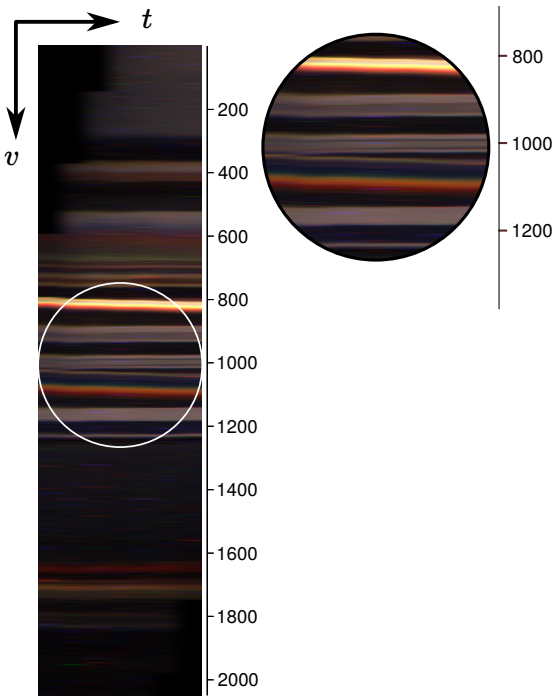
One complication associated with the application of skew correction in u, v rather than s, t, is that, although the light field has aligned and orthogonal u, v and s, t axes, the output slices in u, v are skewed. As such, each of the outputs in u, v shown here have had a single, 2D inverse skew correction applied.

As a demonstration of light field filtering, we selected a low-light region from the light field depicted in Fig. 7, and adjusted its contrast to enhance the dark regions of the image. A slice of the resulting input light field is shown in Fig. 8(a), and the filtered light field is shown in (b). The filter in this case is the same refocussing filter depicted in Fig. 7 – as in optical focus, virtual refocus has the advantage of rejecting noise. Notice how the fingerprints on the computer case, clearly visible in the filtered image, are nearly entirely obscured by noise in the input image. With the addition of synthetic Gaussian noise, the input and filtered light fields in Figs. 8(c) and (d) depict a more extreme low-light scenario, again showing significant gains through filtering. For applications requiring a wider depth of field, volumetric focus might be applied to reduce noise without impacting depth of field [Dansereau *et al.*, 2013].

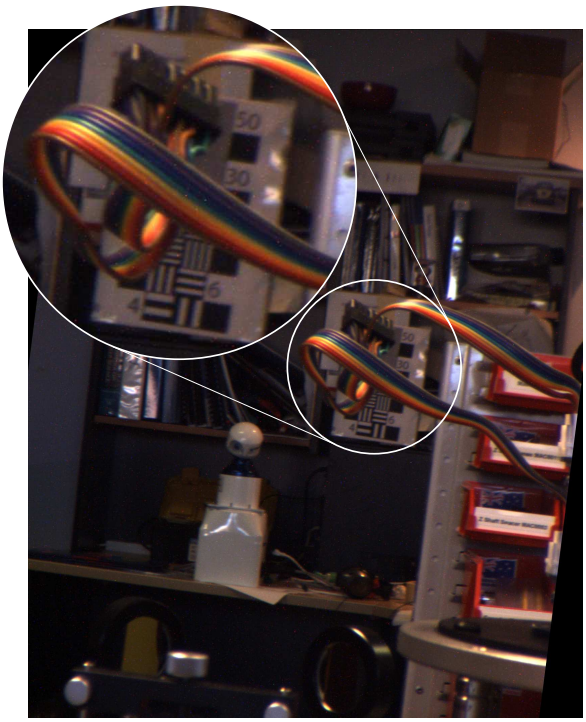
5 Conclusions and Future Work

We have demonstrated that panoramic image capture suffering from parallax presents an opportunity for light field construction. By building light fields from panoramic sequences, we have shown that, rather than a detriment, parallax motion can represent an advantage. We demonstrated post-capture refocus and noise reduction from our measured light fields. By employing a high-performance mirror-steering camera pointing device, we were able to capture hundred-megapixel light fields in tens of seconds. Due to the high repeatability of this device, calibration needed only be carried out once, with subsequent light field construction reduced to a set of per-image affine transformations. Though the transformations took about two minutes to execute in Matlab, affine transformations can be straightforwardly and significantly accelerated on GPU, and might be skipped altogether by employing algorithms that deal directly with raw sub-images and transformation estimates.

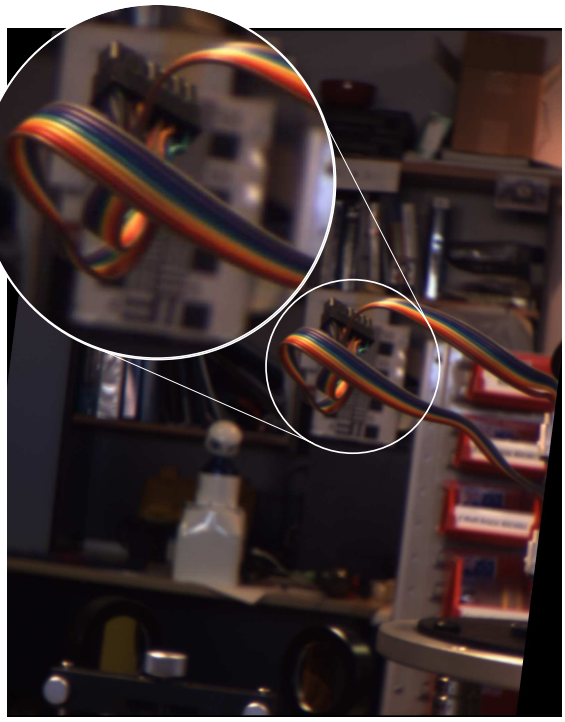
We have not addressed lens distortion or other optical aberrations introduced by the optical pointing system. Near the edges of the refocused images, and in particular when interacting with the light fields live, it is evident that significant radial distortion is present in each of the images making up the light field. One question for future investigation is in the potential interplay between the optical pointing system and lens distortion.



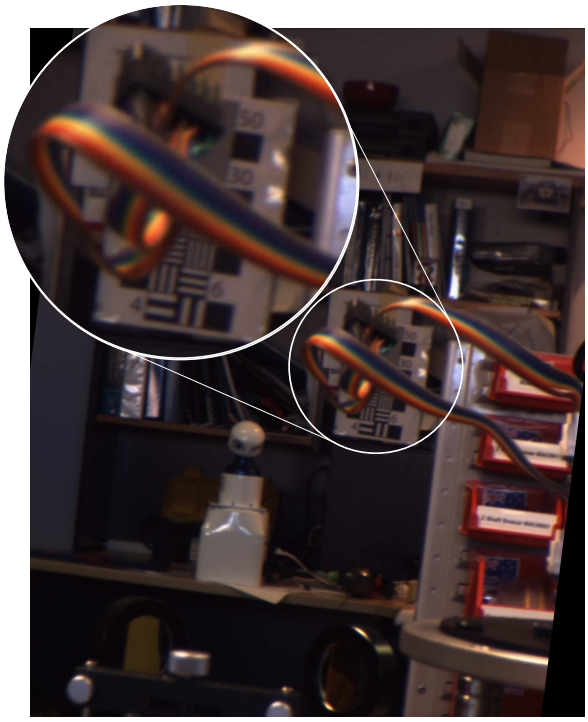
(a) Slice in t, v



(b) Input

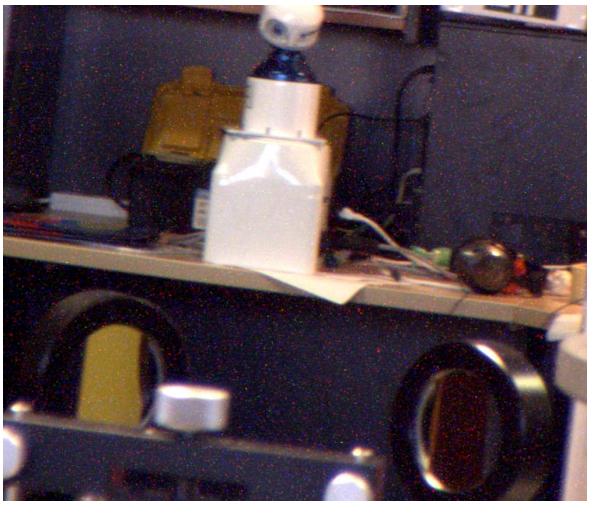


(c) Focused on foreground

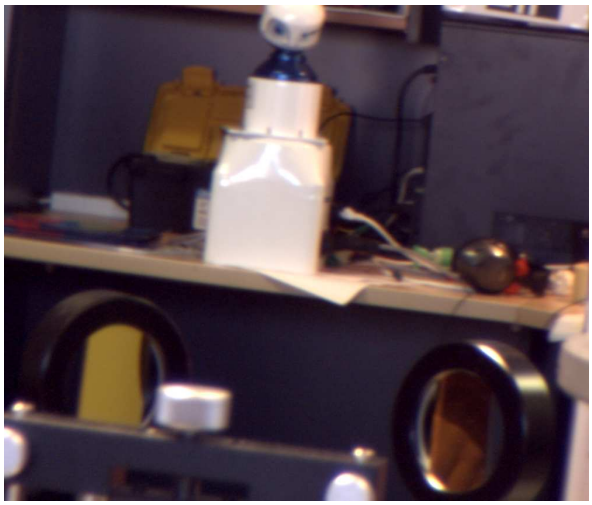


(d) Focused on background

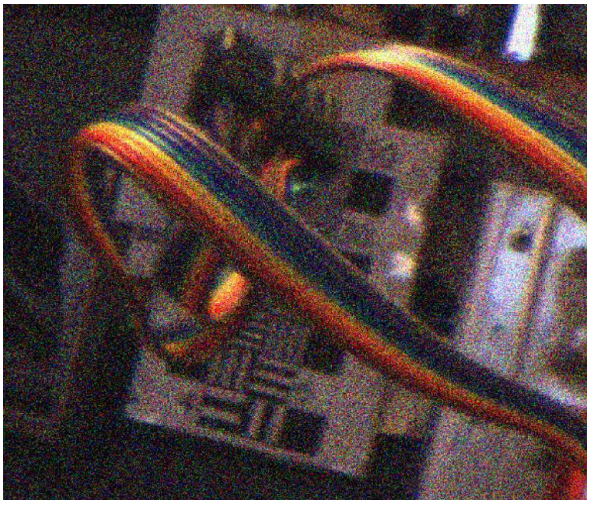
Figure 7: (a) A slice in t, v of the input light field, showing the multiple-pixel difference in slope between foreground and background elements; (b) A slice in u, v of the input light field and (c,d) light fields focused on foreground and background elements, respectively, showing significant depth selectivity.



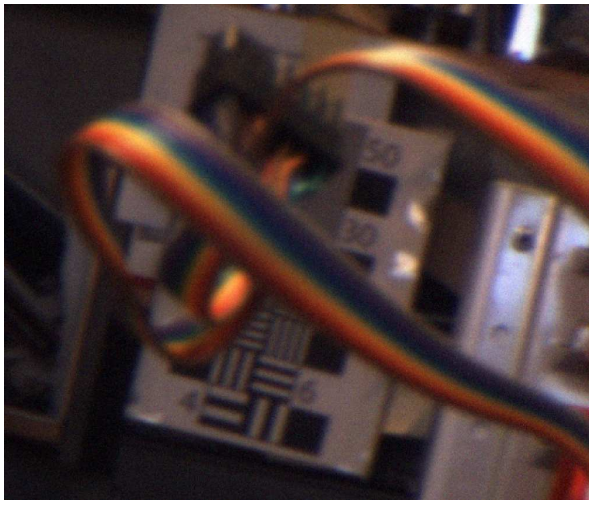
(a) Input



(b) Filtered



(c) Input with added noise



(d) Filtered

Figure 8: (a) Contrast-adjusted input and (b) focused light fields showing how virtual focus, like optical focus, improves signal-to-noise ratio; Notice the fingerprints on the computer case, top-right, hidden by the noise in the input but appearing clearly in the filtered image; (c) A different part of the scene with synthetic Gaussian noise added to emulate a more extreme low-light imaging scenario, and (d) the filtered image; Note the improvement in legibility of the numbers on the test pattern.

The calibration and rectification required to reverse this distortion is unclear.

When constructing the light field, we corrected for rotation between the camera and grid, as well as skew in the grid, by transforming the individual images in u, v. It should be possible, instead, to perform a one-time optimisation of the grid itself, the idea being to measure a grid of images that is already rotation- and skew-adjusted to yield an orthogonal and aligned light field. The repeatability of the pointing system should allow these elements of the calibration to be carried out entirely through manipulation of the capture grid.

One of the strengths of the REV25 camera steering device is its flexibility. It would be interesting to dynamically adjust the parameters of the sampling grid – say changing the number of samples or the grid spacing – to suit the dynamic needs of a real-world application such as low-light imaging or 3D target tracking.

Finally, we have demonstrated light field construction from a specific device, the REV25, but the underlying concept of this work applies wherever parallax is introduced in a multiple-image acquisition system. It should therefore be possible to generalise the technique described here over many similar devices.

Acknowledgments

This work is supported in part by the Australian Centre for Field Robotics, the New South Wales State Government, The University of Sydney, and the Australian Government's International Postgraduate Research Scholarship and Endeavour Research Fellowship.

References

[Adelson and Wang, 2002] E. H. Adelson and J. Y. A. Wang. Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 14(2):99–106, 2002.

[Bishop and Favaro, 2011] Tom Bishop and Paolo Favaro. Full-resolution depth map estimation from an aliased plenoptic light field. In *Asian Conference on Computer Vision (ACCV)*, pages 186–200. Springer, 2011.

[Dansereau and Bruton, 2003] D. G. Dansereau and L. T. Bruton. A 4D frequency-planar IIR filter and its application to light field processing. In *Intl. Symposium on Circuits and Systems (ISCAS)*, volume 4, pages 476–479. IEEE, May 2003.

[Dansereau and Williams, 2011] D. G. Dansereau and Stefan B. Williams. Seabed modeling and distractor extraction for mobile AUVs using light field filtering. In *Robotics and Automation (ICRA)*, pages 1634–1639. IEEE, May 2011.

[Dansereau *et al.*, 2013] D. G. Dansereau, Daniel L. Bongiorno, Oscar Pizarro, and Stefan B. Williams. Light field image denoising using a linear 4D frequency-hyperfan all-in-focus filter. In *Proceedings SPIE Computational Imaging XI*, page 86570P, Feb. 2013.

[Fiss *et al.*, 2014] Juliet Fiss, Brian Curless, and Richard Szeliski. Refocusing plenoptic images using depth-adaptive splatting. In *Computational Photography (ICCP)*, pages 1–9. IEEE, 2014.

[Lowe, 2004] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Intl. Journal of Computer Vision (IJCV)*, 60(2):91–110, 2004.

[Lumsdaine and Georgiev, 2009] A. Lumsdaine and T. Georgiev. The focused plenoptic camera. In *Computational Photography (ICCP)*, pages 1–8. IEEE, 2009.

[Ng *et al.*, 2005] 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*, 2, 2005.

[Smith *et al.*, 2010] B.M. Smith, L. Zhang, H. Jin, and A. Agarwala. Light field video stabilization. In *Intl. Conference on Computer Vision (ICCV)*, pages 341–348. IEEE, 2010.

[Wanner *et al.*, 2013] S. Wanner, C. Straehle, and B. Goldluecke. Globally consistent multi-label assignment on the ray space of 4D light fields. In *Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jun. 2013.

[Wilburn *et al.*, 2005] B. Wilburn, N. Joshi, V. Vaish, E.V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy. High performance imaging using large camera arrays. *ACM Transactions on Graphics (TOG)*, 24(3):765–776, 2005.