

# Hyperbolic View Dependency for All-in-Focus Time of Flight Fields

Adam K. Taras, Donald G. Dansereau

Australian Centre for Field Robotics,  
School of Aerospace, Mechanical and Mechatronic Engineering,  
The University of Sydney, Australia  
{adam.taras, donald.dansereau}@sydney.edu.au

## Abstract

Time of flight (ToF) cameras have played a major role in enabling high-level decision making and scene representation in robotic systems. However conventional ToF camera performance degrades with long ranges and strongly absorptive or specular objects. Combining multiple ToF cameras in arrays has been proposed for mitigating these limitations, measuring a structure similar to a light field. In this work we expose a previously undescribed hyperbolic view dependency in ToF arrays and leverage it to construct an all-in-focus filter that improves robustness and accuracy of depth estimation. Our pipeline correctly deals with occlusion and saturation, and outperforms previous conventional and array-based approaches. This work has the potential to make robotic perception more accurate and robust, thus allowing robots to work in previously prohibitive conditions.

## 1 Introduction

Time of flight (ToF) and other depth-sensing cameras have found widespread use for 3D scene perception. Applications in domestic and industrial domains arise wherever grasping, localisation and mapping, and human-robot interaction are important. Warehouse robots, autonomous vehicles, drone-based search and rescue, and autonomous underwater survey and are all able to function more safely and consistently with effective 3D perception.

There are, however, limitations to all depth sensing approaches. Passive stereo and multi-view cameras struggle around textureless regions and poor illumination. Active speckle projection and ToF cameras suffer at long ranges, around strongly absorptive surfaces, and in the presence of interfering signals like sunlight, due to low signal-to-noise ratio (SNR). Paradoxically, they also suffer around strongly reflective specular sur-

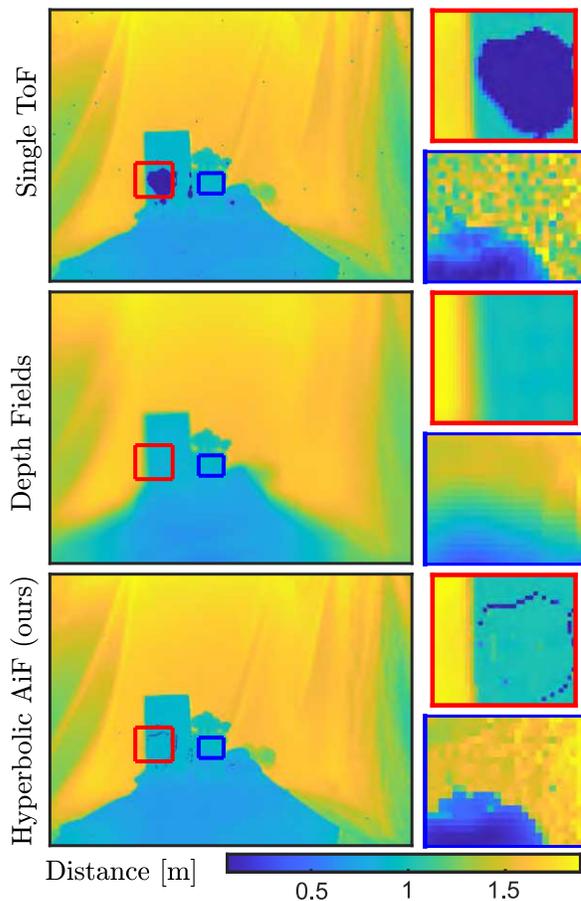


Figure 1: (top) ToF imaging yields noisy results around dark and/or distant regions (blue, contrast-stretched) and invalid estimates around specular reflection (red); (center) Previous work [Jayasuriya *et al.*, 2015] combines views from an array of ToF cameras to improve signal quality, but degrades accuracy and does not maintain sharp edges; (bottom) The proposed hyperbolic all-in-focus filter approach correctly combines views, preserves edges, and removes view-dependent artefacts, yielding more accurate and robust depth estimates to allow robots to better operate in challenging scenarios.

faces, where strong signal returns result in saturation and invalid depth estimates. The net effect is that even commonly occurring situations can yield inaccurate and incomplete depth maps in active depth sensing cameras.

Recent work has shown that combining ToF depth measurements over time can dramatically improve performance, especially in challenging conditions [Attal *et al.*, 2021]. Earlier work showed that combining ToF measurements from multiple spatial locations can yield a similarly significant improvement from an instantaneous measurement [Jayasuriya *et al.*, 2015]. By combining multiple views in a light field-like structure to focus on a single depth, SNR can be improved and the impact of specularities decreased. Performing this in an instantaneous measurement means it can drive decisions in time-critical robotics applications.

In this work we expose and overcome two key limitations of previous work combining multiple ToF cameras in an array. First, we show how to combine multiple views while maintaining sharp focus over the entire range of depths in the scene, rather than focusing on a single depth. Second, we show prior works miss an important view dependency in ToF field measurements, impacting accuracy. We describe this view dependency and propose a method that exploits it to yield more accurate and robust measurements via an all-in-focus filter.

The contributions of this work are:

1. We discover and describe a hyperbolic view dependency in the distance values of ToF fields,
2. We propose the first time of flight field all-in-focus filter, exploiting this view dependency, offering enhanced noise rejection and depth accuracy compared to previous methods, and
3. We augment the filter to correctly handle occlusion boundaries and reject saturation from specularities, further improving accuracy and robustness.

The dataset and code will be available at <https://roboticimaging.org/Projects/ToFF>.

To validate our approach we mount a single ToF camera on a robotic arm, imaging static scenes over a grid of camera poses to emulate a ToF camera array taking an instantaneous measurement. Figure 1 provides a qualitative view of the performance improvement our pipeline provides compared with conventional ToF capture and the previous ToF array approach. In Section 4 we further quantify the performance of our approach, showing more consistent and accurate depth estimates, and we evaluate how our method degrades with noise, showing substantially lower error than previous approaches. Finally, we show the proposed method better maintains edge information as well as offering more robustness to specularities compared to previous approaches.

The instantaneous capture and enhanced depth estimation offered by this work have the potential to improve robotic depth perception in a range of commonly occurring scenarios. From objects on reflective warehouse shelves to darkly coloured or strongly specular objects in dynamic scenes, this could allow robots to work more quickly and safely and in a broader range of conditions.

## 2 Related Work

The improvement of depth measurements from ToF cameras is an active research space. Recently, there has been a focus on learning-based approaches for use in challenging scenes. For example, DeepToF [Marco *et al.*, 2017] uses a ToF camera with no modifications feeding into an autoencoder to learn a basis for multipath interference and correct it. Su *et al.* [Su *et al.*, 2018] use a deep convolutional neural network from dual frequency measurements to allow for multipath removal, denoising and phase unwrapping. Instead, our work establishes a principled approach to gathering and combining multiple-aperture ToF fields to offer instantaneous capture while improving accuracy and robustness.

There are also several notable examples where augmenting the internal processing can allow the camera to see more clearly in challenging conditions. This includes coded illumination patterns for dealing with reflections [Heide *et al.*, 2013] or turbid media [Heide *et al.*, 2014], using different illumination locations or polarised light to extract surface normals and edges [Callenberg *et al.*, 2017] or constructing demodulation frequencies orthogonal to the illumination to extract information about the Doppler shift of points in the scene [Heide *et al.*, 2015]. Our proposed pipeline operates on the standard ToF camera, but we note that the concept of a camera array could also be combined with these forms of augmentation.

Combining multiple views to address shortfalls of vision systems is well explored for conventional cameras. Light fields use an array of cameras to capture richer scene content. Applications include denoising [Alain and Smolic, 2017], [Chen *et al.*, 2018] and occlusion removal [Vaish *et al.*, 2006], [Wang *et al.*, 2020]. Similarly, we leverage the richer information present in an array of sensors to improve ToF imaging, but also expose a key point of departure from conventional light field processing, namely the hyperbolic view dependency associated with ToF fields.

Depth fields (DF) [Jayasuriya *et al.*, 2015] first proposed using an array of ToF cameras for improved noise rejection. By leveraging conventional light field techniques, they demonstrated refocusing at a single depth, phase unwrapping and imaging through occluders. Whilst convincingly demonstrating the potential of using multiple views provided by ToF camera arrays, we

show how to keep the entire scene in focus, and that the view independence assumed by conventional light field techniques is not valid in ToF fields. We propose a pipeline that provides an all-in-focus filter by leveraging the hyperbolic view dependence (see [Subsection 3.1–3.2](#)).

Finally, TöRF [[Attal et al., 2021](#)] uses images from multiple ToF measurements. Whilst similar, our work solves a different problem with a well defined array of cameras, with a possibility of supporting instantaneous online sensing for robotic platforms. On the other hand, TöRF operates on multiple exposures over time in a more unstructured fashion.

### 3 Methodology

In this work we derive and exploit the appearance of a single point as imaged by an array of ToF cameras. Our experimental validation emulates a camera array using a robotic arm to moves both the light and aperture of a ToF camera. Our derivation thus describes this scenario, and requires minor modification for alternative configurations, e.g. employing a single light source in the center of an array. The hyperbolic view dependency that we uncover remains in both cases.

#### 3.1 Phase-Correct Depth Measurements

A camera’s position within a planar array can be described by its coordinates  $(s, t)$  relative to top corner of the array, as indicated in [Figure 2](#). Neglecting phase wrapping effects, the phase surface of a single point imaged by a camera in the array at  $(s, t)$  is given by

$$\phi(s, t) = k_\phi \sqrt{P_z^2 + (P_x - s)^2 + (P_y - t)^2}, \quad (1)$$

where  $\mathbf{P} = [P_x, P_y, P_z]^T$  is the world coordinate of the point being imaged and  $k_\phi = 4\pi f_m / \nu$  converts distance to phase expected for a ToF camera with modulation frequency  $f_m$  and propagation rate  $\nu$ . For simplicity, we scale our data using  $r(s, t) = \phi(s, t) / k_\phi$  so that all values represent distance rather than phase. The distance surface described by [Equation 1](#) is known as a hyperboloid of two sheets, and can be rearranged into canonical form.

Points measured by the camera array also follow the well known light field point-plane correspondence [[Dansereau et al., 2015](#)]. Discretising the time of flight field, we have cameras indices as  $(i, j)$  that sample  $(s, t)$ , and pixels within each image as  $(k, l)$  that sample  $(u, v)$ . Then, a point in the scene will be seen by pixels following the relationships

$$\frac{\Delta k}{\Delta i} = \frac{\Delta l}{\Delta j} = -\frac{\alpha d}{P_z(N-1)}, \quad (2)$$

where  $\alpha$  is the focal length of the camera,  $d$  is the physical dimension of the array in meters, and  $N \times N$  is the

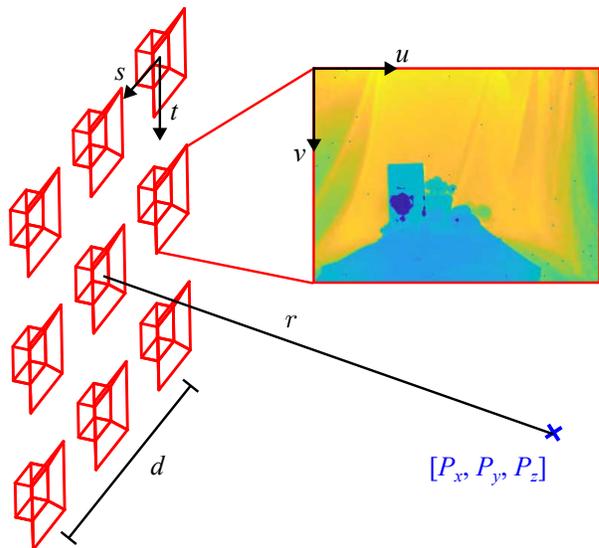


Figure 2: Schematic of a ToF camera array. Camera positions are defined by  $(s, t)$  and direction within images as  $(u, v)$ . In both cases, the top left corner is the origin. The camera array has baseline  $d$ . We find the distance to a point  $[P_x, P_y, P_z]$  imaged by a camera at  $(s, t)$  follows the hyperboloidal form  $r(s, t) = \phi(s, t) / k_\phi$  shown in [Equation 1](#).

number of cameras in the array. Note that a given pixel’s correspondence across the light field has only a single free parameter,  $P_z$ , while the observed value follows the hyperboloid of two sheets described by [Equation 1](#).

[Figure 3](#) shows this geometry for an unoccluded point. The figure depicts  $k, l$  in the center view, and examples of epipolar plane images, the consequence of mapping single rows or columns of pixels as a function of camera position in the array. As in conventional light field imaging, the epipolar images are dominated by straight lines. However, rather than taking a constant value as in a light field, [Equation 1](#) describes the dependence in distance seen along the straight lines in the epipolar images. We believe this is the first time this structure is described in the literature, and it captures both the light field imaging geometry of slope varying with depth, and the hyperboloidal dependence of observed depth on camera position.

The curvature of the hyperboloid at the vertex  $(s, t) = (P_x, P_y)$  is

$$\frac{\partial^2 r(s, t)}{\partial s^2} \Big|_{(P_x, P_y)} = \frac{\partial^2 r(s, t)}{\partial t^2} \Big|_{(P_x, P_y)} = \frac{1}{P_z}. \quad (3)$$

Thus, closer objects behave less linearly in the center of the image, however these are the objects that are the most important for robotic applications. The previous depth fields method [[Jayasuriya et al., 2015](#)] averages the

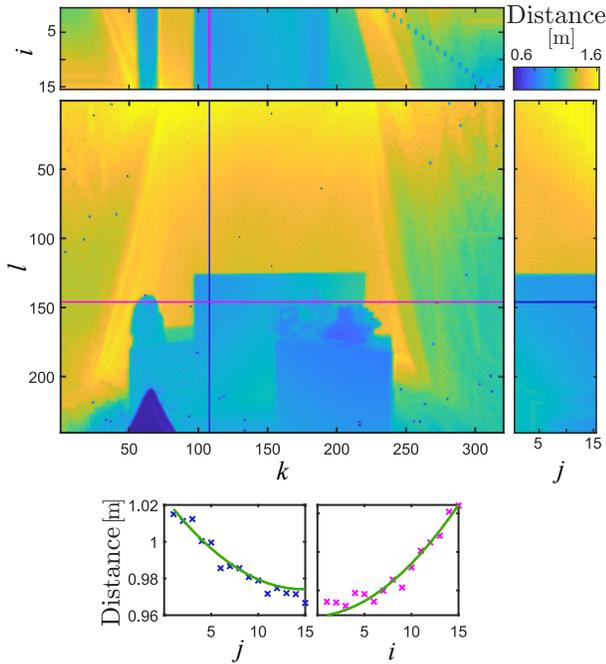


Figure 3: The hyperbolic view dependency in ToF fields: a row of pixels from a ToF image (center red) changes as a function of horizontal camera position  $i$  in an epipolar slice (top), and similarly a column of pixels (blue) evolves as a function of vertical camera position  $j$  (right). In light field imagery epipolar slices show lines of constant intensity, but inspecting them in the ToF field reveals variation. From first principles we derive this variation as a hyperboloid of two sheets, and show it to be a good fit to experimental measurements (bottom, green).

depths from all of the views, thus systematically overestimating depth (see [Subsection 4.2](#)).

### 3.2 All-in-Focus Depth Estimation

We leverage the observations from the above to create a pipeline that uses all measured views to generate a single, all-in-focus depth image, as seen from the pose of the central camera of the array. For each pixel in the center frame we optimise over  $P_z$ , testing at each putative depth the extent to which the observed data agrees with both the epipolar plane geometry and hyperbolic view dependency.

For each point in the image at each putative depth  $P_z$ , we apply [Equation 2](#) to find the corresponding samples in all views. We then perform a best-fit of the hyperboloid ([Equation 1](#)) to the measured distances at those samples. We measure the goodness of fit as the root mean squared error (RMSE) in [Equation 1](#). We use a simplex search method to find the  $P_z$  that minimises the RMSE. By repeating for each pixel, a full-frame all-in-focus depth image is produced.

Examples of the resulting best fit surfaces are shown in [Figure 6](#). There is strong agreement between our model and observations over a broad range of conditions. The following sections describes how we handle outliers in the fit caused by occlusions and saturation. In each case, the only change is to discard affected points from the goodness of fit calculation.

Our method also yields a goodness of fit for each pixel. We anticipate this measure of confidence in each estimate could be used to inform higher-level estimation and control algorithms on robotic platforms.

In this work we do not apply any form of regularisation, and expect doing so, e.g. through optimisation of total variation, could result in further improvement to the depth maps. Here we show results that treat each pixel independently to best understand the characteristics of the approach.

### 3.3 Handling Occlusion

Occlusion is a complex behaviour that breaks the assumptions of the previous derivation. The left column of [Figure 4](#) illustrates the behaviour of the pipeline if occlusions were not taken into account. Around edges, the depth appears blurry, as shown in the top row. By inspecting a surface near an edge, we observe that the set of points do not fit a single surface. Attempting to minimise the error over all points leads to a final estimate that is closer than the true depth. The edge is not always an instantaneous discontinuity, since a single pixel can measure a mixture of rays from different depths. The RMSE of the final surface fit indicates that this behaviour occurs at all edges and is most significant near the edge, where a correct depth fit is the most critical.

To remedy this, we use the fact that the occluder is always in front of the target world point. A computationally efficient way of filtering this is to apply a threshold i.e. if a point is some distance  $d_{occ}$  closer than point would be in the surface fit, it is classified as an occluder and not used. This parameter should be as close to zero as possible without clipping noisy data points. For our datasets,  $d_{occ} = 7$  cm is typical.

The result of using thresholding to handle occluders is shown in the right column of [Figure 4](#). The final depth image produced has much clearer edges. Inspecting the surface fit near an edge shows that the occluding points (blue) is correctly classified and removed, such that the fit still uses the unoccluded points and is able to produce a better depth estimate. The proportion of views used clearly shows that around edges we reject the occluding views, using fewer images but yielding more accurate and sharper results. We also test and demonstrate this approach on more difficult scenarios such as double oc-

cluders, holes and thin wires, some results are shown in Figure 6.

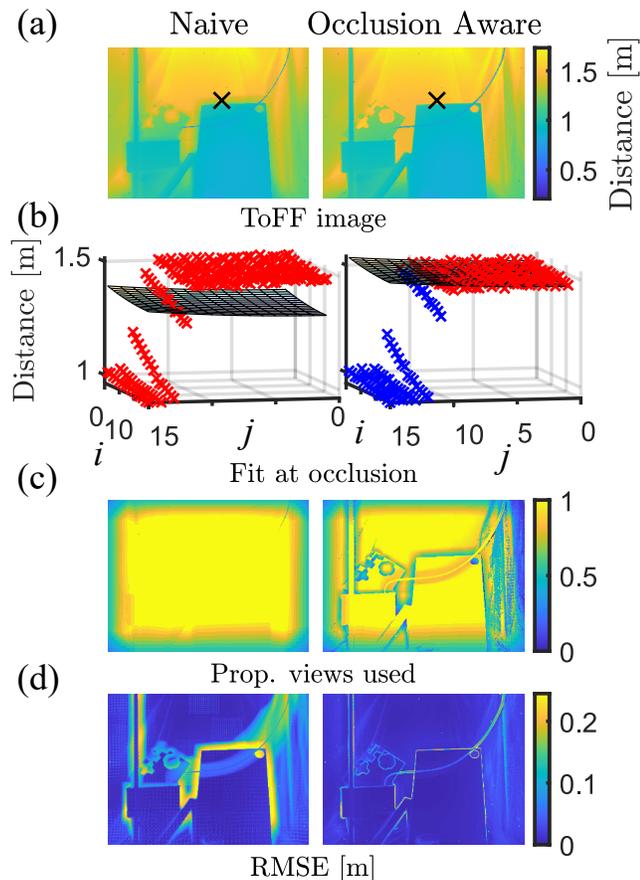


Figure 4: Occlusion-aware filtering maintains edge detail. (a) a point on an edge indicated by the black ‘x’ appears as (b) a discontinuity in  $(i, j)$  in the fitting process. The occlusion-aware fit (right) discards outliers (blue) yielding a better fit (gray) at the occlusion boundary. (c) The resulting estimate ignores some views, meaning a smaller proportion of total views get used, but (d) this results in sharper edge boundaries and lower error in the final depth map.

The consequence of occlusion-aware filtering is a more consistent accuracy across the image, including near edges. We do note some larger errors occur at the edge, rather than near it. This arises since the pixels at edges contain two different depths within the solid angle subtended at the sensor. A methods for handling this sub-pixel effect is left as future work.

### 3.4 Dealing with Specularity

As shown in Figure 1, specularity can cause the depth processing pipeline to fail by saturating the sensor. Similarly, there are cases where some pixels become unresponsive (i.e. dead pixels). Each of these are pixel-level

phenomena that vary across views in the ToF array. Where a particular pixel is saturated or unresponsive, other views may provide valid measurements.

We propose to extend our occlusion-handling approach to deal with saturated or dead pixels. When computing the error in the fitting process, we detect and reject pixels that are unphysical – too close as set by a threshold. If the center view is still valid, this approach closely resembles the occlusion removal approach. In the case where the pixel from the center view is invalid, we compute a coarse plane sweep over a candidate  $P_z$  range, initialise from the depth that gives the minimum fit error after removing invalid depths from the relevant views.

## 4 Results

To validate the ToF field approach we mount a Chronoptics Kea ToF camera on a UR5e robotic arm, emulating an array of ToF cameras. For all experiments we sample  $15 \times 15$  views in  $(s, t)$  with a baseline of  $d = 0.3$  m. The camera illuminates and demodulates the signal with  $f_m = 50$  MHz.

To provide an additional point of comparison, we also repeatedly image the scene from a stationary center view, capturing  $15^2$  images. This allows us to average multiple exposures in a stationary burst imaging approach, representing an upper bound in noise rejection but requiring a static scene and camera.

### 4.1 Qualitative Evaluation

We present a qualitative comparison between our proposed method and other candidates in Figure 1. DF [Jayasuriya *et al.*, 2015] is the previous state of the art method for processing time of flight fields. In the single image case (top), there is low SNR and regions of no information caused by saturation from highly reflective surfaces. The previous approach improves SNR but blurs edges outside of the focus depth. As we demonstrate, previous work also systematically overestimates the true depth since it assumes view independence. Our proposed method leverages the hyperbolic view dependency in the structure of the data. We filter view-dependent effects such as saturation from specularity producing high quality, low noise depth maps of complex scenes. Consequently, the output from our approach shows higher SNR, sharper edges, and more resilience to specularity than single-view or the previous filtering approach.

We also show a detailed comparison in Figure 5, imaging a mannequin head. A single ToF capture is noisy and obscures some features of the face such as the lips. A possible solution is averaging through a stationary burst, but this requires a static scenario that is not always realistic in mobile robotics applications. By capturing multiple views simultaneously, both the previous DF and proposed methods seek to improve SNR for instantaneous

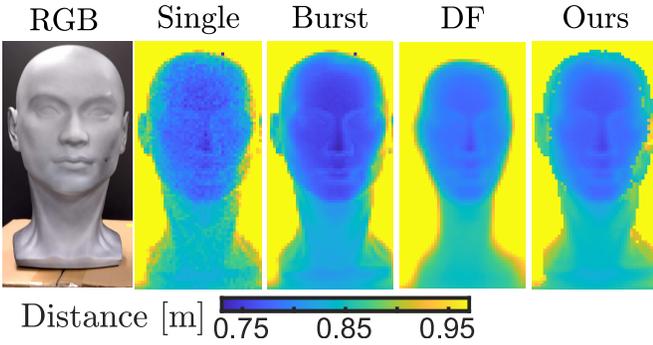


Figure 5: Qualitative performance comparison: (left to right) A mannequin head imaged with a single ToF camera shows significant noise; Combining a burst of frames in a static scene reduces noise but is not an instantaneous measurement; Previous depth fields approach improves SNR but loses information off the focal plane, apparent here as a loss of geometry around the neck and ears; The proposed approach improves SNR while maintaining sharp edges and features of the face.

capture. The DF approach is only sharply focused at a single depth, resulting in loss of edge information and detail at other depths. Our method maintains focus at all depths, improving SNR by combining multiple single images. The improvement is not as strong as that of the burst method, as evident in the spurious pixels around the face boundary. The increased depth around the left ear arises due to the large angle of incidence and is common between all methods. In particular, for DF the ear is out of focus and hence overestimated beyond the contrast stretch. A limitation that is also common is multipath interference, most noticeable near the bridge of the nose. This could be improved by combining our pipeline with previous approaches to handling multipath.

We also demonstrate the internal surface fits of our pipeline in a range of cases, shown in Figure 6. The data on unoccluded points strongly agrees with the model developed in Subsection 3.1, as seen in the top row. The distance spanned by the surface depends inversely on the depth of the point as in Equation 3, with Surface 1 spanning the largest range and Surface 3 spanning the least. We also demonstrate that our occlusion detection is valid, even when there are multiple occluders such that the proportion of points used is less than 0.5. The wire does get lost as it gets further away, since the thickness of the wire at that point is sub-pixel. As future work our pipeline could be augmented to handle sub-pixel effects to remedy this.

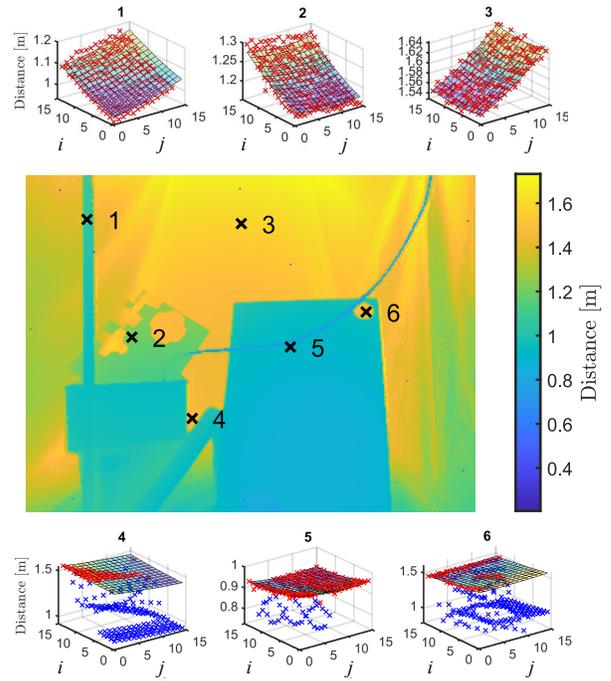


Figure 6: Illustration of surface fits in a variety of cases. Points which are closer (1) have a larger range of distances than points which are further (3), in accordance with Equation 3. Our pipeline fits the best surface while rejecting occluding views (blue). Our approach works around double occluders (4), thin occluders (5) and holes (6).

## 4.2 Quantitative Evaluation

We quantitatively compare methods by measuring the error in depth to a known planar surface. We crop the resulting depth image and evaluate depth RMSE from a planar fit to a gold-standard measurement taken as a long-exposure version of the stationary burst capture.

Figure 7 illustrates the errors of the final depth estimate by each pipeline. Table 1 additionally lists the performance of the methods in terms of completeness. For stationary burst and single capture, the RMSE provided in parentheses removes dead pixels to not confound results. For the DF method we focused at the plane depth, providing the best case result for this method. The single image has low SNR, whilst the stationary burst is the highest quality but requires multiple exposures for a static scene. Since the previous work assumes view-independence, it systematically overestimates distance. This difference can be critical in many robotics tasks such as object tracking or grasping. Our method shows lower noise than a single capture and has no systematic bias as in the DF method. We observe a slightly degraded accuracy of our method compared to the burst whilst maintaining the capture time of a single ToF im-

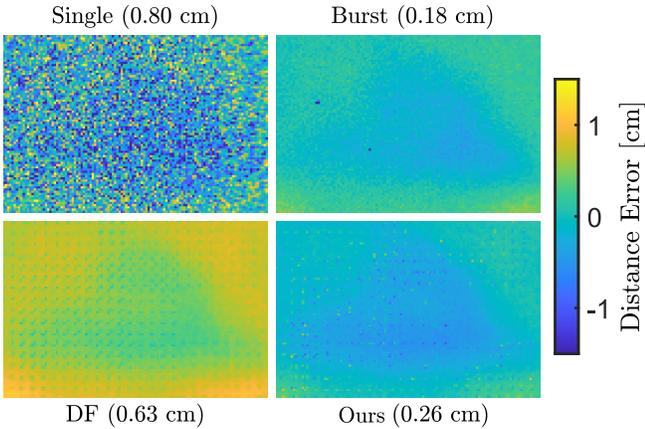


Figure 7: Performance comparison for a planar white screen at a distance of 1 m, showing per-pixel error in depth and RMSE in parentheses. Our method improves SNR almost as much as a burst without requiring a static scene. The previous work DF fails to account for view dependency so produces estimates that, even when in focus, systematically overestimate the distance. The numerical results are shown as accuracy in [Table 1](#).

age. The completeness measure counts the number of pixels that are not physical ( $< 60$  cm when the closest object in the centre frame is 70 cm). This is a view-dependent phenomenon and thus both the single and stationary burst have degraded performance. The depth fields has no such pixels, however this method still uses the invalid pixels in the average, underestimating the true depth. Our method deals with the dead pixels as occlusions, reducing the number of invalid pixels by an order of magnitude. Failure cases occur when the saturation causes invalid depth estimates greater than the saturation identification threshold, which can be improved if the minimum distance to the scene is known.

Table 1: Quantitative evaluation: accuracy measured as RMSE from a planar surface (dataset: `side_wall_low_light`) and completeness of the measurement (dataset: `fruit`, as in [Figure 1](#)) as the number of invalid points ( $r < 60$  cm). Our method provides more complete and accurate measurements while maintaining instantaneous capture.

Metric	Single	Burst <sup>a</sup>	DF <sup>b,c</sup>	Ours <sup>b</sup>
RMSE [cm]	0.80	0.18	0.63	0.26
Invalid Pts	660	681	0	82

<sup>a</sup> Capture time: 180.0 ms, <sup>b</sup> Capture time: 0.8 ms

<sup>c</sup> Depth fields [[Jayasuriya et al., 2015](#)]

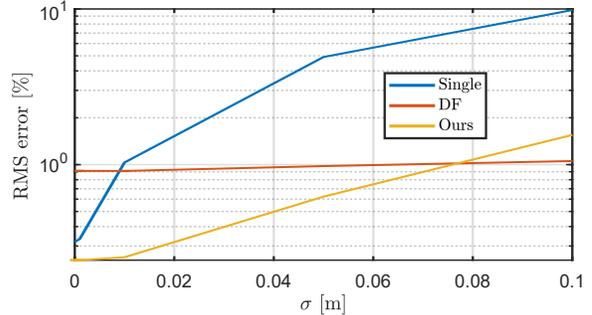


Figure 8: Performance of the model as synthetic noise is added to images of a planar sheet. This is the best scenario for the DF method since the whole scene is in focus. Our method correctly accounts for the view dependency, improving SNR and outperforming single capture and DF at typical noise levels. The depth fields pipeline yields biased results, limiting performance at low noise levels.

Finally, we investigate how robust the instantaneous capture methods are to noise. [Figure 8](#) shows the RMSE over the same planar surface after adding synthetic noise with standard deviation  $\sigma$ . As seen in the figure, the proposed approach is much more resilient to noise than single-image capture, and the DF approach shows poor performance at low noise levels due to the systematic error associated with view dependence. At high noise levels our method chiefly degrades due to the occlusion handling algorithm interpreting points as outliers. This reflects that the occlusion model should be adjusted to match the noise characteristics of the use case.

## 5 Conclusions

We have presented an all-in-focus filter for arrays of ToF cameras that exploits the previously undescribed hyperbolic view dependency present in ToF fields. Our pipeline combines information from across views to improve SNR and rejects view-dependent degradations such as specularity while maintaining sharp edges across all depths in the scene. By correctly accounting for the hyperbolic view dependency, we improved the accuracy of depth measurements compared with single image capture or previous works combining ToF views.

As future work we believe incorporating ToF fields with augmented internal processing e.g. [[Heide et al., 2013](#)], [[Callenberg et al., 2017](#)] would better deal with complex scenes including multipath rejection and imaging through translucent media. Furthermore, minor adjustments such as including the offset between light source and camera, explicitly dealing with dead pixels and including phase unwrapping could improve perfor-

mance in some applications. Finally, we believe information in the amplitude channel can be leveraged to improve occlusion detection and handling.

## Acknowledgments

The authors would like to acknowledge Jia Chen Lu for his suggestions on the implementation of this work and to Chronoptics for their support in using the Kea ToF camera.

## References

- [Alain and Smolic, 2017] Martin Alain and Aljosa Smolic. Light field denoising by sparse 5d transform domain collaborative filtering. In *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6. IEEE, 2017.
- [Attal *et al.*, 2021] Benjamin Attal, Eliot Laidlaw, Aaron Gokaslan, Changil Kim, Christian Richardt, James Tompkin, and Matthew O’Toole. TöRF: Time-of-flight radiance fields for dynamic scene view synthesis. *Advances in neural information processing systems*, 34, 2021.
- [Callenberg *et al.*, 2017] Clara Callenberg, Felix Heide, Gordon Wetzstein, and Matthias B Hullin. Snapshot difference imaging using correlation time-of-flight sensors. *ACM Transactions on Graphics (ToG)*, 36(6):1–11, 2017.
- [Chen *et al.*, 2018] Jie Chen, Junhui Hou, and Lap-Pui Chau. Light field denoising via anisotropic parallax analysis in a cnn framework. *IEEE Signal Processing Letters*, 25(9):1403–1407, 2018.
- [Dansereau *et al.*, 2015] Donald G. Dansereau, Oscar Pizarro, and Stefan B. Williams. Linear volumetric focus for light field cameras. *ACM Transactions on Graphics (TOG)*, 34(2):15, February 2015.
- [Heide *et al.*, 2013] Felix Heide, Matthias B Hullin, James Gregson, and Wolfgang Heidrich. Low-budget transient imaging using photonic mixer devices. *ACM Transactions on Graphics (ToG)*, 32(4):1–10, 2013.
- [Heide *et al.*, 2014] Felix Heide, Lei Xiao, Andreas Kolb, Matthias B Hullin, and Wolfgang Heidrich. Imaging in scattering media using correlation image sensors and sparse convolutional coding. *Optics express*, 22(21):26338–26350, 2014.
- [Heide *et al.*, 2015] Felix Heide, Wolfgang Heidrich, Matthias Hullin, and Gordon Wetzstein. Doppler time-of-flight imaging. *ACM Transactions on Graphics (ToG)*, 34(4):1–11, 2015.
- [Jayasuriya *et al.*, 2015] Suren Jayasuriya, Adithya Pediredla, Sriram Sivaramakrishnan, Alyosha Molnar, and Ashok Veeraraghavan. Depth Fields: Extending light field techniques to time-of-flight imaging. In *2015 International Conference on 3D Vision*, pages 1–9. IEEE, 2015.
- [Marco *et al.*, 2017] Julio Marco, Quercus Hernandez, Adolfo Munoz, Yue Dong, Adrian Jarabo, Min H Kim, Xin Tong, and Diego Gutierrez. Deeptof: off-the-shelf real-time correction of multipath interference in time-of-flight imaging. *ACM Transactions on Graphics (ToG)*, 36(6):1–12, 2017.
- [Su *et al.*, 2018] Shuochen Su, Felix Heide, Gordon Wetzstein, and Wolfgang Heidrich. Deep end-to-end time-of-flight imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6383–6392, 2018.
- [Vaish *et al.*, 2006] V. Vaish, M. Levoy, R. Szeliski, C.L. Zitnick, and S.B. Kang. Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures. In *Computer Vision and Pattern Recognition Conference (CVPR)*, volume 2, pages 2331–2338. IEEE, 2006.
- [Wang *et al.*, 2020] Yingqian Wang, Tianhao Wu, Jungang Yang, Longguang Wang, Wei An, and Yulan Guo. DeOccNet: Learning to see through foreground occlusions in light fields. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 118–127, 2020.