

A comparison of correlation-based plenoptic depth estimation techniques for digital image correlation

Bibek Sapkota^{1*}, Holger Mettelsiefen¹, Vrishank Raghav¹ and Brian S. Thurow¹

¹ Auburn University, Department of Aerospace Engineering, Auburn, AL, U.S.A

* bzs0084@auburn.edu

Abstract

This paper presents a comparison of several correlation-based methodologies for depth estimation using a single plenoptic camera. The plenoptic camera offers a distinct advantage by enabling the generation of many perspectives over a relatively small baseline. Unlike stereo reconstruction, which relies on a pair of images for depth estimation, these multiple perspectives are utilized collectively in two distinct approaches for depth estimation. The proposed methods are evaluated using synthetic and experimental data to assess their accuracy. Preliminary results indicate the robust performance of both methods, each exhibiting different strengths under varying conditions. Future work will assess how these methods perform in the context of a simultaneous DIC and 3D PIV measurement using a single plenoptic camera.

1 Introduction

The simultaneous measurement of fluid motion and surface deflection is critical for problems involving fluid-structure interaction where the flow physics and structural deformation are often coupled. In the case of three-dimensional flows and structural deformation, established measurement techniques typically require 4 cameras for tomographic particle image velocimetry (PIV) and 2 cameras for digital image correlation (DIC). Simultaneous measurements necessitate a complex 6-camera setup (Safi et al.) resulting in high hardware costs, complex alignment and operational challenges, as well as challenges related to depth of field. Moreover, in many experimental setups, optical access is not available for such multi-camera measurements. In such scenarios, a single camera with 3D measurement capability is sought. A promising solution is the use of a plenoptic camera that can potentially measure both 3D flow motion and surface deflection simultaneously using a single camera. This work represents the first step in this direction, focusing specifically on the development of a robust DIC technique that efficiently utilizes the unique information provided by plenoptic cameras. Future work will focus on the integration of PIV or PTV algorithms with the DIC algorithm.

The plenoptic camera has established itself as a valuable tool for optical 3D measurements, finding applications in flow measurements, temperature measurements, density gradient measurements, and surface reconstructions (Fahringer et al. (2022)) (Klemkowsky et al. (2017)). In the context of surface reconstructions, some progress has been made in utilizing plenoptic cameras for 3D digital image correlation (Zhao et al. (2019)). However, the simultaneous measurements for both surface and flow using a single plenoptic cameras remain a distinct challenge and are unexplored. This paper presents two novel 3D digital image correlation methodologies for depth estimation using a plenoptic camera to be used in simultaneous PIV and DIC measurement. The feasibility of the method is quantified with the synthetic result and experimentally verified with the reconstruction of flat plates at different depths.

1.1 Plenoptic Cameras

Plenoptic cameras possess a unique capability to encode 4D light field information of a scene into a single static image, which can be post-processed to extract 3D information. In a conventional camera, the entire aperture focuses light rays onto a point on the image sensor. In contrast, a plenoptic camera incorporates a

dense microlens array (MLA) positioned close to the sensor plane. The MLA serves to redirect light rays from different sections of the aperture to specific pixels on the sensor plane. While the MLA captures the positional information (s, t) of the light ray, individual pixels within the sub-image encode information from diverse angular origins (u, v). This unique arrangement enables a plenoptic image to capture 4D light-field information (u, v, s, t), including both positional and directional characteristics.

Sampling a pixel (u,v) from each sub-image (s,t) formulates an image with a unique angular view of the world. Thus a single plenoptic image is able to generate many (~ 100) perspective views over a small baseline. These perspectives also have a higher depth of field because they are sampled from small aperture sub-images. These features have been explored by Advanced Flow Diagnostics Laboratory (AFDL) at Auburn University to successfully develop single-camera-based 3D measurement techniques.

1.2 Stereovision and plenoptic DIC

Stereovision is a depth estimation technique using the apparent motion of the camera or disparity between two cameras. The disparity obtained between the two images is linearly proportional to depth. Since the plenoptic camera is able to generate perspective with a unique angular view, there exists a disparity between the position of an object in any two perspectives. This disparity can be directly related to depth using Equation 1 (Roberts and Thurow (2017)).

$$\frac{1}{S_0} = \frac{1}{F} - \frac{1}{S_{if}} \left(1 - \frac{d}{B}\right) \quad (1)$$

The relation is trivial and can be formulated from lens equations and similar triangle rules. The variables can be visualized from the figure 1 where S_0 is the object distance, F is the focal length of the main lens, S_{if} is the separation between the main lens and the MLA, d is the disparity between two perspectives and B is the baseline. The baseline is given by Equation 2 where f is the focal length of MLA and h is the absolute value of (u,v) on the sensor.

$$B = -\frac{S_{if} * h}{f} \quad (2)$$

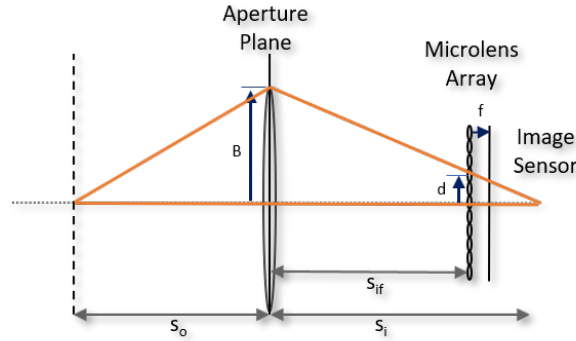


Figure 1: Optical arrangement showing disparity and aperture displacement

3D digital image correlation is a widely used technique that is highly effective in displacement and deformation tracking. It is composed of two steps, digital image correlation for disparity measurement and stereo vision for depth estimation. Cross-correlation is a mathematical operation to identify the similarity between two signals and is given by 3:

$$C_{x,y} = \frac{\sum_{i,j} (I_{a_{i,j}} - \mu_a) (I_{b_{i+x,j+y}} - \mu_b)}{\sqrt{\sum_{i,j} (I_{a_{i,j}} - \mu_a)^2 \sum_{i,j} (I_{b_{i,j}} - \mu_b)^2}} \quad (3)$$

The location of the peak value in the cross-correlation function gives the apparent disparity between the two images. In a plenoptic camera, the cross-correlation can be computed between any two arbitrary perspectives. The uniqueness of the plenoptic DIC is the availability of a large number of perspectives, meaning a large number of disparities can potentially be measured for a single object. In this work, this redundant information is used to formulate a methodology to be used in surface tracking with the additional information used to a) improve the accuracy of the measurement given a relatively narrow baseline and b) to improve the accuracy of the measurement in the presence of occlusion or noise, such as that produced by seed particles that may be present for simultaneous PIV/DIC measurements.

2 Objective

One of the major challenges associated in simultaneous Particle Image Velocimetry (PIV) and Digital Image Correlation (DIC) using a single plenoptic camera is acquiring accurate depth information from coupled particle and background images. Conventional methods struggle to handle the corruption of depth information caused by the presence of particles and random noise in the image. Particles located in front of the surface obstruct the view, while random noise contributes to poor correlation and low signal-to-noise. Consequently, the simultaneous process becomes particularly challenging when using a single camera.

Commonly used temporal image processing techniques, such as minimum noise filters, and spatial-based image processing techniques, like sliding mean filters, present challenges when applied to suppress image noise in speckle pattern images. The application of these filters compromises either the particle or the background image in simultaneous imaging, thereby affecting the quality of acquired background images and potentially corrupting the necessary depth information. As a result, traditional image processing methods are suboptimal for simultaneous PIV and DIC operations. Hence, a novel methodology, capable of withstanding noise, is required for depth estimation using speckle background displacement tracking with a correlation-based approach.

A plenoptic camera, with a large number of perspectives at a known baseline, offers unique information that can potentially mitigate these challenges. The effect of random noise can be mitigated by either averaging the signal to acquire a robust disparity or filtering out the outlier disparity to only sample potentially good disparity values for accurate depth estimation. In this study, both of these approaches are explored with the introduction of two novel methodologies for depth estimation using a plenoptic camera.

3 Methodology

In this work, two novel correlation-based techniques are presented to acquire robust depth information using a plenoptic camera. These methods are then compared to the traditional stereo triangulation method. The details of the implementation of these methods are described below.

3.1 Disparity bundling

The steps of disparity bundling based depth estimation methodology are shown in figure 2. The disparity bundling methodology is inspired by cross-correlation based depth estimation technique (Roberts and Thurow (2017)) and light-field ray bundling technique (Clifford et al. (2019)). In this method, first, a reference image is defined among all available perspectives, and all others are considered displaced images compared to the reference image. For this work, the perspective sampled from the center of the aperture (i.e. $u=0$ $v=0$) is chosen as a reference image for convenience.

Then, cross-correlation is performed between the reference image and all of the other perspective images. For correlation, both the reference image and displaced images are divided into sub-images. Multi-grid iterative cross-correlation is used to find the displacement of the sub-images. Normalized cross-correlation weighted with a Blackman window is used in the process. 2D-gaussian peak fitting is performed to identify the sub-pixel value of the correlation peak. Thus, the disparity between the reference perspective and all other perspective images is identified.

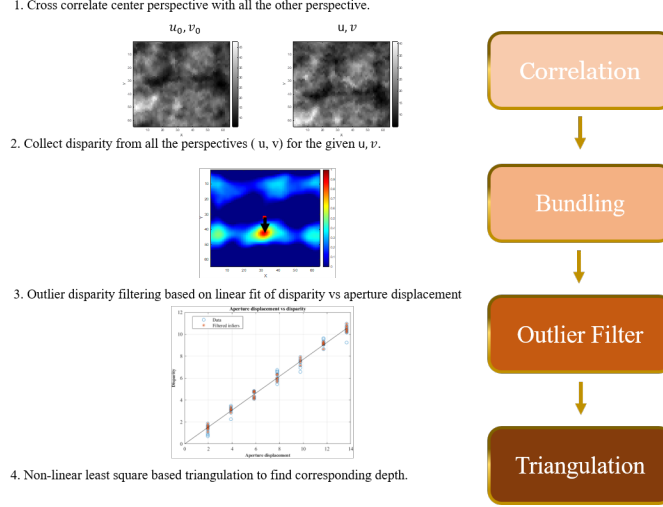


Figure 2: Steps of disparity bundling methodology

In the second step, the acquired disparity is bundled together for each sub-window. Disparity refers to the amount the sub-image has shifted from the reference image to the displaced image. So, when the disparity of the point (s,t) on reference image are collected together from all correlation maps, this gives the new position of this window in all the other perspectives. Thus, all the s,t are collected for all the u,v. The bundling step in this methodology is much simpler in comparison to the bundling based on the minimum distance between rays as mentioned in Clifford et al. (2019).

Third, the outlier disparities are filtered to obtain robust depth estimation. In the presence of inaccuracies such as noise or occlusion in the surface data, one or more disparities may be erroneous. Given the availability of a large number of disparity values, it is possible to exclude some from the depth analysis. To identify these outliers, prior knowledge of the linear relationship between disparity and aperture displacement is explored. Similar to (Roberts and Thurow (2017)), a line is fit through the acquired disparity and aperture displacement values. The Random Sample Consensus (RANSAC) algorithm is then applied to the data to determine the slope of the best-fit line. RANSAC is an iterative mathematical method widely used for outlier filtering and estimating mathematical models from noisy data. The fitted line is ensured to pass through the y-intercept at zero because the disparity at zero aperture displacement is known to be zero. Disparity values that deviate more than 0.2 pixels from the best-fit line are considered outliers. Thus, the outlier-filtered data is accepted, and the bundle is updated.

Finally, the depth location is estimated from the triangulation of image points. Direct light field calibration (DLFC) (Hall et al. (2018)) is a polynomial-based mapping of object space points (x, y, z) into image space $s = P(x, y, z, u, v)$ and $t = P(x, y, z, u, v)$. The updated bundle consists of a list of all image space points (s, t, u, v) , and the required variables are (x, y, z) . A non-linear least square optimization scheme using MATLAB's *lsqnonlin* function was implemented to solve for these variables. The obtained result is a point cloud in a 3D space.

3.2 Aperture Averaging

The aperture averaging method presents an alternative correlation-based approach for acquiring depth using a single plenoptic camera. This method involves averaging cross-correlation information from all available perspectives to obtain a robust disparity value, which is then triangulated to determine depth information. The utilization of all perspective information is made possible by the known baseline between the reference image and the displaced image. This method is inspired by the multi-frame pyramid correlation scheme (Sciacchitano et al. (2012)) but correlation planes are averaged based on spatial position. The method is explained in four steps, as depicted in Figure 3

In the first step, cross-correlation is performed to obtain a correlation map. Similar to the disparity

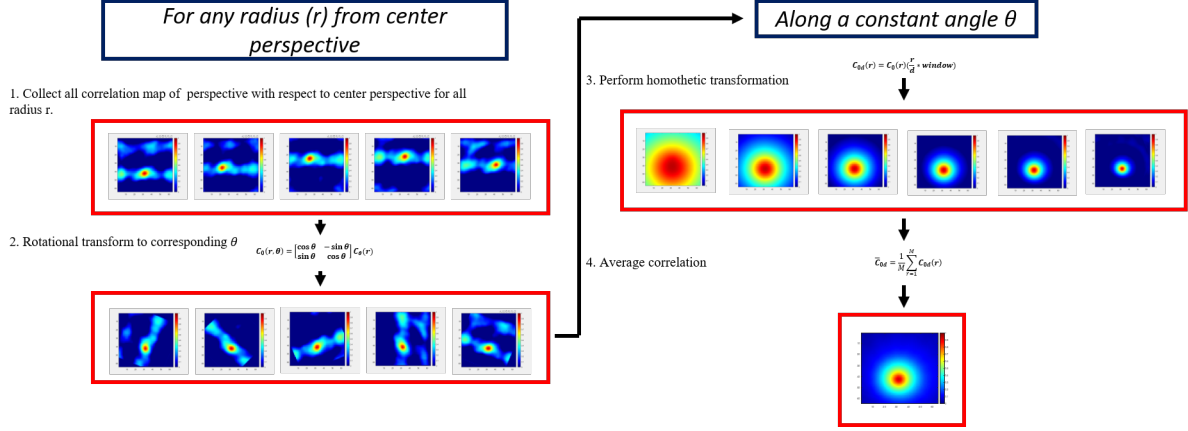


Figure 3: Steps of aperture average methodology

bundling method, the central perspective is considered as the reference perspective and correlated with all other perspectives to acquire correlation maps. Again, the multi-grid iterative cross-correlation method is utilized to generate the correlation map. Following the cross-correlation process, both the baseline (r), representing the distance from the center perspective, and the corresponding reference angle are recorded.

The second step of this method is to perform the rotational transformation of all the perspectives to a common line. Since the position of sampled aperture image is known, a rotational transformation of all the corresponding correlation maps brings the correlation peak to align to a single line. In fact, if the perspectives are at the same distance from the center, the correlation peak should coincide at the same point. For implementation, MATLAB's *imrotate* function is used with cubic interpolation and keeping the image size constant with rotation.

The third step of this method is to perform the homothetic transformation of all the rotated perspectives to the same size. In this step, all the correlation maps are transformed to match the disparity corresponding to the largest displacement. If d is the largest distance of baseline from the central perspective and r is the distance of any perspective after rotational under consideration, the homothetic transformation of the window to the corresponding maximum baseline window is given by equation 4.

$$C_{0d}(r) = C_0(r) \frac{r}{d} * window \quad (4)$$

The final step is to average the correlation maps obtained after homothetic transformation. The homothetic transformation is expected to align all the peaks to a common point. This allows a direct averaging of the correlation plane. The final correlation plane after averaging is expected to be clean and distinct peaks with high signal-to-noise. There is an option of performing a weighted average instead of direct averaging the plane from homothetic transformation. Further work is needed to observe different ways weights could be assigned and if a weighted average has any impact.

4 Synthetic Experiment

In order to quantify the performance of the methodology, a flat synthetic speckle plane was generated at different depths from -15 mm to 15 mm placed at the interval of 3 mm and the surface was reconstructed as a points cloud. To create a synthetic surface, randomly spaced round dots were projected on the image using ray tracing equations. The methodology for the creation of random dots is similar to the method of synthetic plenoptic particle field images as described by (Fahringer et al. (2015)) but for a surface with round dots. These random speckle dots is observed to provide a good correlation map for the measurement of displacement between perspectives. The synthetic image generated was for a modeled Imperx B6620 CCD camera with 6600*4400 pixels and a hexagonal micro-lens array of 472*361 resolution. The hexagonal microlens

array has a pitch of 0.077mm and a focal length of 0.308 mm. The simulation was performed with a main lens of a focal length of 85 mm and at a magnification of -0.5 mm.

4.1 Results of synthetic experiments

Three different methodologies, (i.e. normal stereo, disparity bundling, and aperture averaged) were compared with each other based on their performance to accurately resolve the surface at different resolutions and against different noise levels. For the comparison, the synthetic surface was placed at depths between -15 mm to 15 mm within the depth of field and with a uniform spacing of 3 mm. The corresponding uncertainty of each methodology was noted and compared.

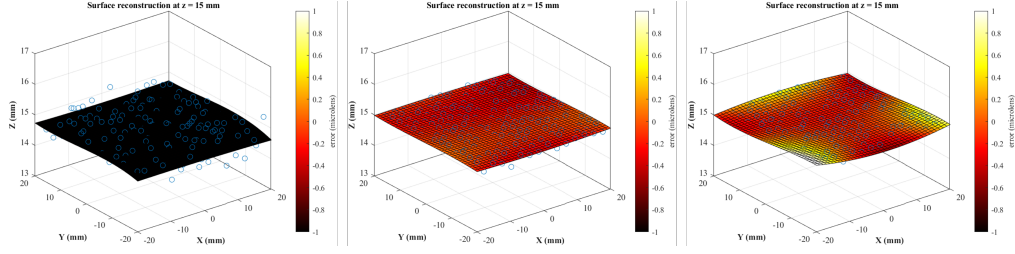


Figure 4: Synthetic surface reconstruction with error map at $z=15$ mm for stereo(left), disparity bundling (center), aperture averaged(right)

Figure 4 shows a synthetic reconstruction of a flat surface using all three methods at a depth of $z=15$ mm. The color map shows the local error (true depth - obtained depth) of reconstruction at a given depth normalized by the size of a microlens, which is equivalent to a pixel in conventional imaging. Both the shape of the fitted curve and the local error for stereo show the method underperforms comparatively. Aperture averaged method and disparity bundling shows similar performance and is able to capture the flat surface at the given plane. Analyzing the variations of an error on the surface, it is observed that the disparity bundling has a more uniform error meaning the uncertainty is lower. This is also evident in depth-wise uncertainty bar graph 5 at $z=15$ mm. Analyzing the depth-wise uncertainty at different depths, a similar trend holds for all the depth locations. The aperture averaging and disparity bundling show smaller uncertainty while the stereo methods lag in performance distinctively.

4.1.1 Performance at different resolution

To compare the performance of the algorithms at different spatial resolutions, the results were compared at 4 different final window sizes- 64×64 , 32×32 , 16×16 , 8×8 .

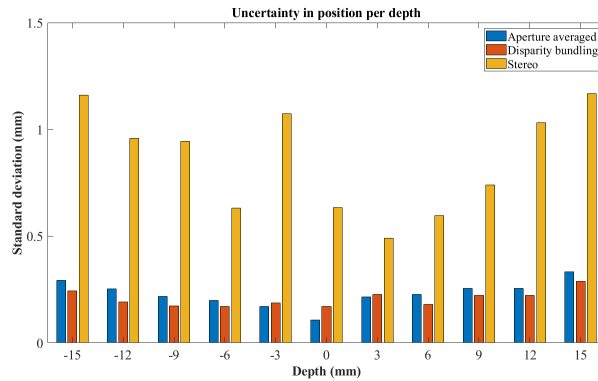


Figure 5: Depth-wise uncertainty comparison

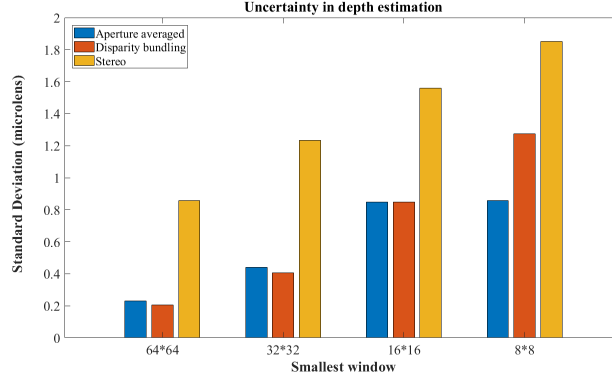


Figure 6: Uncertainty at different resolutions of correlation

Figure 6 compares the uncertainty of each algorithm at different resolutions. At larger window sizes, both the disparity bundling and correlation averaging methods exhibit similar performance, while distinctly outperforming the stereo method. These results align with our expectations because the stereo method relies solely on a single correlation and is therefore heavily reliant on the quality of that correlation. In contrast, the disparity bundling and correlation averaging methods leverage information from all available perspectives to resolve depth information.

As the window size decreases, the uncertainty in depth estimation increases across all cases. In comparison, the averaging method exhibits lower uncertainty at higher resolutions. Due to the averaging of a large number of correlation planes, this method generates a clean and distinct correlation map, even when the final window size is small. Hence, the algorithm's robustness, even at high resolutions, is recognized as one of the advantages of this method.

4.1.2 Performance at different noise level

Random Gaussian noise was introduced to the synthetic images, and their ability to resolve the same background was tested at various noise levels. The variance of the Gaussian noise level was kept fixed at 0.2, while the peak noise intensity was adjusted from 0 to 20 percent of the peak background intensity. The corresponding uncertainties for each method were compared at these noise levels

Figure 7 illustrates the performance of three different algorithms at various noise levels. Across all noise levels, the stereo algorithm consistently demonstrated lower performance compared to the other methods. This can be attributed to the fact that the stereo method relies on a single correlation value, making it more susceptible to increased uncertainty when noise is present. The performance of the disparity bundling and correlation methods remained comparable until a high noise level of 0.2 percent of the maximum intensity was introduced. At higher noise levels, the aperture averaging method exhibited a slight advantage in performance. While each individual correlation is prone to random noise, the averaging method mitigates the impact of randomness through the aggregation of the correlation map. As a result, the aperture averaging method holds a slight advantage specifically in the presence of high noise levels.

5 Moving plate experiment

In order to quantify the performance of the depth estimation methodologies with a single plenoptic camera, a bench-top experiment was set up with surfaces at the known distance from the camera. As analyzing depth accuracy is the major parameter of concern for surface reconstruction, the surface was moved along a known distance and its depth uncertainty evaluated was quantified.

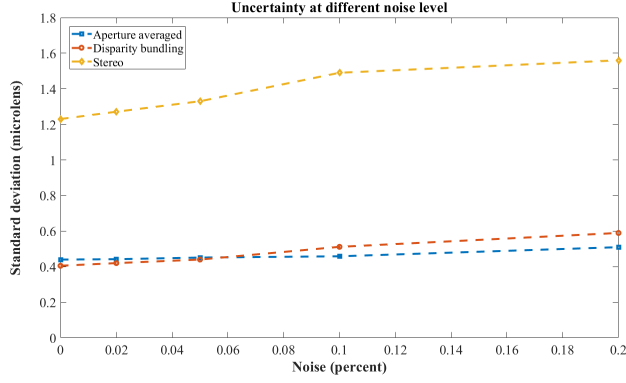


Figure 7: Change in uncertainty with noise

5.1 Experimental configuration

A simple bench-top experimental setup with a plenoptic camera and a surface to be tracked placed along the transitional stage is shown in figure 8. The movement of the flat plate placed on an electronic transitional stage was controlled with a computer. The flat plate consisted of an attached wavelet background which was used to track the plate displacement. The plate was moved every 3 mm along the depth and images were captured at each depth level. An Imprex B6620 CCD camera with a hexagonal micro-lens in front of the sensor was used for imaging. The light rays were focused with a 60 mm focal length main lens and at a magnification of -0.6.



Figure 8: Bench-top experimental setup

5.2 Results and discussion

Figure 9 shows a plot of the measured vs approximated depth position of the flat surface at different positions in the volume. A fitted line of approximate depth where the plane is expected to be and measured median depth is also plotted. The approximated depth is used here to take into account the experimental uncertainty of the position of the flat plate relative to the calibrated plane. The expected fit of approximated depth vs measured depth is supposed to be linear for a flat plate perpendicular to the camera. Here it is represented by a line of slope of 1.

The error in performance can be quantified in terms of the slope of the line observed. It is observed that the slope fitted with the disparity bundling method is 0.99 vs the slope with aperture averaged around 0.95. This suggests that the disparity bundling is closer to the expected data compared to the aperture averaging method. Looking at the z-position and line, clearly, the aperture averaging method is lagging to capture the depth value. So, the z-position captured is biased towards zero depth for all positions. This suggests the

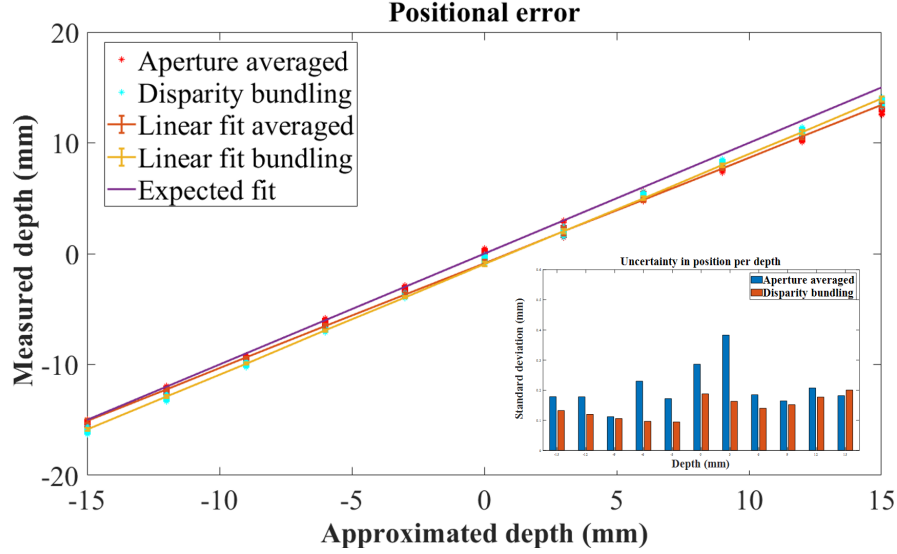


Figure 9: Measured vs approximated plane position

disparity value calculated by aperture methods was biased towards lower values.

While bias error of the aperture averaging method was observed in experimental data, it was not observed in synthetic data. This suggests a physical explanation is required to explain the discrepancy. In the aperture averaging process, it is implicitly assumed that perspective position (u,v) maps linearly from the image sensor plane to the aperture plane. While this holds true for synthetic cases, the perspective position mapping is non-linear. The polynomial calibration that maps the object space points (x,y,z) into image space $s = P(x,y,z,u,v)$ and $t = P(x,y,z,u,v)$ inherently take into the effect of this nonlinearity. However, the aperture averaging doesn't fully use the calibration before averaging process and thus the linear positional transformation could be the potential cause of the discrepancy. The bundling method directly uses the calibration and thus is not susceptible to these biases.

6 Conclusion and future work

Three different correlation-based approaches for depth measurement using a plenoptic camera are compared. Two robust methodologies that utilize the information from multiple perspectives at a known baseline are introduced in the process. The disparity bundling method bundles the position of each sub-images from all the perspectives followed by an advanced filtering technique to separate outliers from good disparity. The aperture averaging method averages out the correlation map to acquire robust disparity from all available perspectives. Both of the methods were compared to the stereo correlation method. It is observed that both methods outperformed the stereo methods in resolving depth. In most cases, the performance of both methods was similar for synthetic data. The aperture averaging method outperformed other methods at higher resolution, showing its advantage. However, the aperture averaging method showed a distinct biased result in experimental data. The potential cause of this behavior was explained as non-linear mapping of perspective position (u,v) in the experimental data. So, further calibration to correct this non-linearity is needed for the robust implementation of this method. On the other hand, the disparity bundling method showed robust performance in both synthetic and experimental data and is simpler in implementation.

The robust methodologies for surface reconstruction are expected to be used in a simultaneous PIV and DIC experiment. Since both PIV and DIC information needs to be extracted from a common image, the data is expected to be significantly corrupted for DIC making the surface reconstruction challenging with traditional methods. The correlation averaging method showed promising results in synthetic data at different noise levels and resolutions, its application to real data is challenging as the experimental results showed an overall lag in disparity estimation. The disparity bundling method showed good performance against most conditions but struggled when the resolution is very high. So, a novel method that takes the benefits of both

methods is desired. Possible alternatives like using the aperture averaged method to identify the peak position in the initial guess while using the bundling method for multigrid iterations will be explored as viable option that utilizes features of both methods.

The next step is to test this methodology with simultaneous PIV and DIC experimental data. An experiment in a water tunnel with simultaneous PIV and DIC data is planned with the flat plate as background moving at different depths. After the acquisition of simultaneous particle and background images, advanced image processing techniques to decouple occluding background from the particle field will be explored. The refinement of the robust method discussed earlier is expected to be helpful to reconstruct the surface data even in the presence of noise and particle occlusion.

Acknowledgements

This work was supported by the National Science Foundation under Grant no. 2145189, monitored by Dr. Ronald Joslin.

References

- Clifford C, Tan Z, Hall E, and Thurow B (2019) Particle matching and triangulation using light-field ray bundling. in *13th International Symposium on Particle Image Velocimetry*
- Fahringer TW, Danehy PM, Hutchins WD, and Thurow BS (2022) Design of a multispectral plenoptic camera and its application for pyrometry. *Applied Optics* 61:2459–2472
- Fahringer TW, Lynch KP, and Thurow BS (2015) Volumetric particle image velocimetry with a single plenoptic camera. *Measurement Science and Technology* 26:115201
- Hall EM, Fahringer TW, Guildenbecher DR, and Thurow BS (2018) Volumetric calibration of a plenoptic camera. *Applied optics* 57:914–923
- Klemkowsky JN, Fahringer TW, Clifford CJ, Bathel BF, and Thurow BS (2017) Plenoptic background oriented schlieren imaging. *Measurement Science and Technology* 28:095404
- Roberts WA and Thurow BS (2017) Correlation-based depth estimation with a plenoptic camera. *AIAA Journal* 55:435–445
- Safi H, Phillips N, Ventikos Y, and Bomphrey R () 3.8 implementing fluid-structure interaction computational and empirical techniques to assess hemodynamics of abdominal aortic aneurysms. *Artery Research* 20:55–56
- Sciacchitano A, Scarano F, and Wieneke B (2012) Multi-frame pyramid correlation for time-resolved piv. *Experiments in fluids* 53:1087–1105
- Zhao J, Liu Z, and Guo B (2019) Three-dimensional digital image correlation method based on a light field camera. *Optics and Lasers in Engineering* 116:19–25