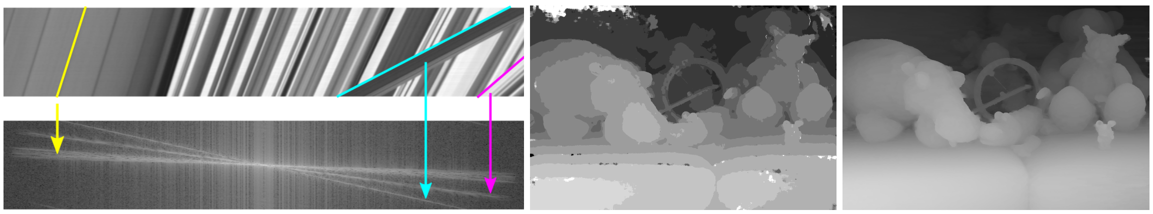
Nonuniform Depth Distribution Selection with Discrete Fourier Transform

Lode Jorissen[1] Patrik Goorts[1] Gauthier Lafruit[2] Philippe Bekaert[1]
[1]Hasselt University - tUL - Expertise Centre for Digital Media, Belgium {firstname.lastname@uhasselt.be}
[2]Université Libre de Bruxelles - LISA department, Belgium {gauthier.lafruit@ulb.ac.be}



(a) (b) (c)

Figure 1: *(a) The frequency domain contains lines representing the dominant depths. (b) Depth map generated using a uniform depth distribution. (c) Depth map generated using the nonuniform depth distribution obtained using our method.*

Keywords: Depth Distribution, Fourier Transform, Light Fields, Epipolar Plane Images

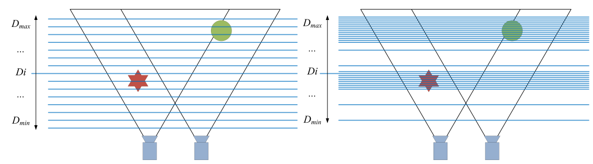
Concepts: •**Computing methodologies** → *Epipolar geometry; Matching;*

1 Introduction

In recent years there is a growing interest in the generation of virtual views from a limited set of input cameras. This is especially useful for applications such as Free Viewpoint Navigation and light field displays [Tanimoto 2015]. The latter often requires tens to hundreds of input views, while it is often not feasible to record with as many cameras. View interpolation algorithms often traverse a set of depths to find correspondences between the input images [Stankiewicz et al. 2013; Goorts et al. 2013]. Most algorithms choose a uniform set of depths to traverse (as shown in Figure 2(a)), but this often leads to an excessive amount of unnecessary calculations in regions where no objects are located. It also results in an increased amount of mismatches, and thus, inaccuracies in the generated views. These problems also occur when a too large depth range is selected. Hence, typically a depth range that encloses the scene tightly is manually selected to mitigate these errors. A depth distribution that organizes the depth layers around the objects in the scene, as shown in Figure 2(b), would reduce these errors and decrease the number of computations by reducing the number of depths to search through. [Goorts et al. 2013] determine a nonuniform global depth distribution by reusing the generated depth information from the previous time stamp. This makes the algorithm dependent on previous results.

We propose a depth distribution selection method that uses the Discrete Fourier Transform (DFT) of light field structures that arise in linear multi-view camera setups. This allows us to generate a depth distribution per scanline, without injecting possibly unstable data from previous time stamps.

SIGGRAPH '16, July 24-28, 2016, Anaheim, CA,
ISBN: 978-1-4503-4371-8/16/07
DOI: http://dx.doi.org/10.1145/2945078.2945133



(a) (b)

Figure 2: *(a) Uniform depth distribution. (b) Nonuniform depth distribution. From Goorts et al. [Goorts et al. 2013].*

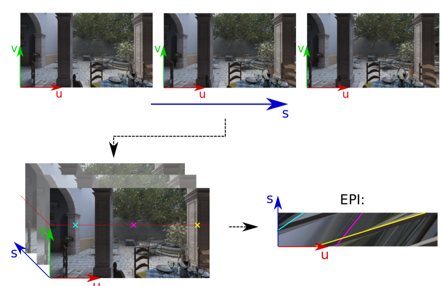


Figure 3: *Epipolar Plane Image Generation.*

2 DFT Based Depth Distribution

Our algorithm assumes that the input images are captured with a linear multi-view camera setup. An EPI v is generated by stacking the rectified input images behind each other in a 3D-cube, and taking the (u, s)-slice corresponding to scanlines v. This process is shown in Figure 3 [Kim et al. 2013]. Each EPI contains line-structures, called EPI-lines, that correspond to 3D-points in the scene. The slope of each EPI-line corresponds to the depth of the point in the scene: points farther away from the cameras have a more vertical slope than points closer to the camera setup. The frequency representation of an image, obtained with the Discrete Fourier Transform (DFT), represents the decomposition of that image into a set of 2D-waves that have a particular frequency, magnitude, phase-shift and

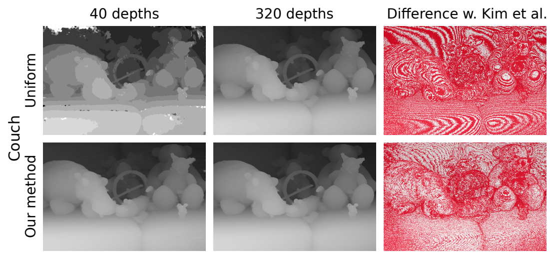


Figure 4: *Results for the 'Couch' dataset. Compared to Kim et al. [Kim et al. 2013], our nonuniform method has* 50.07% *good pixels while the uniform distribution only has* 34.95% *good pixels in the case of 320 depth layers (right column). White pixels indicate correct pixels.*

orientation. In the special case of EPIs, only the waves that are 'perpendicular' to the line structures will have high magnitudes. This results in a set of 'lines' running through the DC-component of the frequency representation (see Figure 1(a)). We consider the total magnitude along these lines to be an indication of the importance of their corresponding depths. The depth distribution for an EPI is calculated by converting the EPI to grayscale and taking the magnitude spectrum from this image its Fast Fourier Transform. Next, we loop over a set of lines running through the center of the magnitude spectrum. Each line corresponds to a disparity d from a uniformly spread set of disparities. A score s_d is calculated by taking the sum of the the magnitudes along the line, but leaving out magnitudes that are likely to be the result of undersampling in the case of a small set of input images. Next, a Cumulative Distribution Function (CDF) is determined for the resulting scores. If the depth distribution needs to contain a set of N depth layers, N evenly spread points on the y-axis of the CDF are selected and their corresponding disparities (the x-value) are considered to be the optimal depth distribution. In the end, every scanline has N depth layers that can be traversed by a depth estimation or view synthesis algorithm.

3 Results

We validate our method by comparing the depth maps obtained with a uniform distribution and our nonuniform depth distribution. For every dataset we used 10 input images. In this section we will show the results for two datasets: 'Couch' and 'Statue'. Both datasets are provided by Kim et al. [Kim et al. 2013]. The left column in Figure 4 shows the results for a multi-view depth estimation algorithm using only 40 depth layers for both a uniform and a nonuniform depth distribution. The depth layers were spread over a disparity range of 0 to 45 pixels, while the actual disparity range of the 'Couch'-scene goes from 15 to 30 pixels. The different depth layers are clearly visible in the depth map obtained from a uniform distribution. This is not the case for our method, thanks to a more tight fitting of the depth layers around the objects. Compared to the depth map of Kim et al., our method results in a much higher PSNR value (36.3dB) than the uniform distribution (21.5dB).

The middle column of Figure 4 shows the results using 320 depth layers. The result of the uniform distribution is now very similar to the result of our method using only 40 depths per scanline. This uniform distribution, however, required a calculation time of 4613ms compared to 690ms with our approach. This uniform distribution result is, at first sight, also very similar to the result of the nonuniform distribution with 320 depth layers. But, when we compare these depth maps to those generated by Kim et al. we can still clearly see the depth layers in the uniform case, while this is not so much the case for our method (Figure 4, right column).

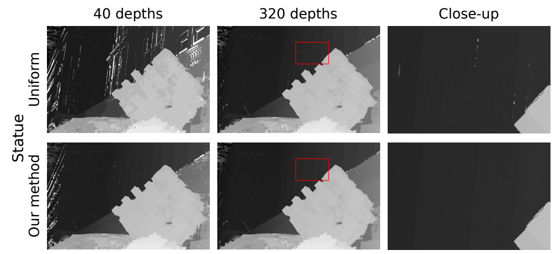


Figure 5: *Results for the 'Statue' dataset. The right column shows a close-up for the 320 depth layers case: less noise can be observed while using our method case.*

Compared to Kim et al., our nonuniform method has 50.07% good pixels while the uniform distribution only has 34.95% good pixels in the case of 320 depth layers. As shown in Figure 5 our nonuniform depth distribution tends to result in less noise, compared to the uniform distribution, even when a larger set of depth layers is used.

4 Conclusion

In this abstract we presented a method to select a more optimal nonuniform depth distribution, using the Discrete Fourier Transform, that can be used to reduce computational time and artifacts in depth estimation and virtual view synthesis applications. Contrary to the related work, our method determines a distribution per scanline instead of a global distribution for the whole scene. It also does not require possibly unstable data from previous time stamps.

5 Future Work

One of the next steps is the reduction of the computational time of the distribution selection to allow a real-time or near real-time calculation of accurate depth maps and view interpolations. We would also like to generalize our findings to arbitrary camera arrangements, as it would allow more freedom in camera placement.

6 Acknowledgements

Part of this research is funded by the European Commission in the scope of the FP7 DREAMSPACE project.

References

GOORTS, P., MAESEN, S., DUMONT, M., ROGMANS, S., AND BEKAERT, P. 2013. Optimization of Free Viewpoint Interpolation by Applying Adaptive Depth Plane Distributions in Plane Sweeping. In *Proc. of the Tenth Intl. Conf. on Signal Processing and Multimedia Applications (SIGMAP 2013)*, INSTICC.

KIM, C., ZIMMER, H., PRITCH, Y., SORKINE-HORNUNG, A., AND GROSS, M. H. 2013. Scene Reconstruction from High Spatio-Angular Resolution Light Fields. *ACM Trans. Graph. 32*, 4, 73–1.

STANKIEWICZ, O., WEGNER, K., TANIMOTO, M., AND DOMASKI, M., 2013. Enhanced Depth Estimation Reference Software (DERS) for Free-viewpoint Television, ISO/IEC JTC1/SC29/WG11/M31518, Oct.

TANIMOTO, M. 2015. FTV Standardization for Super Multiview and Free Navigation in MPEG. In *SPIE Sensing Technology+ Applications*, International Society for Optics and Photonics, 94950T–94950T.