Işık Alanı Kamerası ve Normal Kamera İçeren Hibrid Stereo Görüntüleme
Hybrid Stereo Imaging Including a Light Field and a Regular Camera

M. Zeshan Alam and Bahadir K. Gunturk
Dept. of Electrical and Electronics Engineering, Istanbul Medipol University, Istanbul, Turkey
mzalam@st.medipol.edu.tr, bkgunturk@medipol.edu.tr

***Özetçe* —Işık alanı görüntüleme, ışığın hem yön hem de yer dağılımını kaydeder; böylelikle, kayıt sonrası sayısal odaklama, perspektif kaydırma, derinlik kestirme gibi yeni kabiliyetler getirir. Mikro lens dizisi (MLD) tabanlı ışık alanı kameraları maliyet ve etkinlik açısından uygun bir çözüm sunar. MLD tabanlı ışık alanı kameralarındaki en önemli iki sorun düşük görüntü çözünürlüğü ve derinlik kestiriminin erişim ve hassasiyetini sınırlayan dar görüntü-arası mesafedir. Bu makalede, ışık alanı kamerasından ve normal bir kameradan oluşan hibrid bir stereo görüntüleme sistemi sunuyoruz. Bu hibrid sistem MLD tabanlı ışık alanı kameralarının hem görüntü çözünürlüğü hem de dar kamera-arası mesafe sorunlarına çözüm getirirken ışık alanı kaydetmenin getirdiği kabiliyetleri de korumaktadır.**

Anahtar Kelimeler—Işık alanı kamerası, hibrid stereo görüntüleme.

***Abstract*—Light field imaging involves capturing both angular and spatial distribution of light; it enables new capabilities, such as post-capture digital refocusing, perspective shift, and depth estimation. Micro-lens array (MLA) based light field cameras provide a cost-effective approach to light field imaging. The two major issues with MLA based light field cameras are low spatial resolution and narrow baseline, which limits the depth range and accuracy. In this paper, we present a hybrid stereo imaging system that includes a light field camera and a regular camera. The hybrid system addresses both the spatial resolution and narrow baseline issues of the MLA based light field cameras while preserving all the light field capabilities.**

Keywords—light field imaging, hybrid stereo imaging.

I. INTRODUCTION

Light field (LF) imaging brings in new capabilities and applications to the field of computer vision and image processing [1] [2] [3]. Unlike regular cameras, light field cameras capture the directional light information, and this enables new capabilities, including post-capture adjustment of camera features (such as focal depth and aperture size), post-capture change of camera position and angle, and depth estimation. The depth information can be utilized in low-level and high-level computer vision applications, including segmentation, restoration, object recognition, and tracking.

Light field imaging systems can be implemented in a variety of ways, including camera arrays [4], micro-lens arrays [5], coded masks [6] and lens arrays [7]. Among these different implementations, micro-lens array (MLA) based light field cameras offer a cost-effective approach, resulting so far in two commercial light field cameras, namely, Lytro [8] and Raytrix [9].

Other than the camera arrays, which are in general expensive and bulky, light field cameras suffer from the fundamental resolution trade-off between spatial and angular resolution. For example, the first-generation Lytro camera has a spatial resolution of 380x380=0.14 megapixels. The recently announced second-generation Lytro camera has a spatial resolution of 4 megapixels; this is a decent number, however it should be noted that the sensor used in the camera is about 40 megapixels, and this large resolution capacity translates to only 4 megapixels due to the angular-resolution trade-off.

The second issue associated with MLA based light field cameras is narrow baseline. The distance between sub-aperture images decoded from light field data is very small, significantly limiting the depth range and accuracy [10].

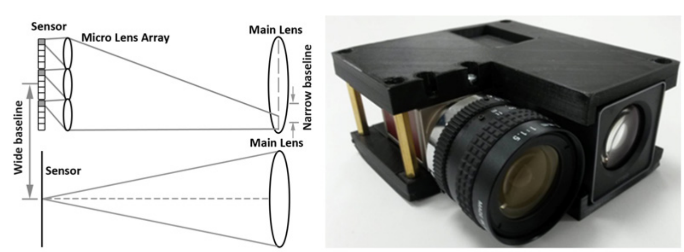

Figure 1: Hybrid imaging system.

To address both the resolution and baseline issues, we propose a hybrid stereo imaging system, consisting of a light field camera and a regular camera. The imaging system is shown in Figure 1. The system has two main advantages. First, the high-resolution image captured by the regular sensor is fused with the low-resolution sub-aperture images to enhance the resolution of each sub-aperture image. That is, a high

This work is supported by TUBITAK Grant 114E095. Dr. Gunturk is supported by TUBITAK Grant 114C098.

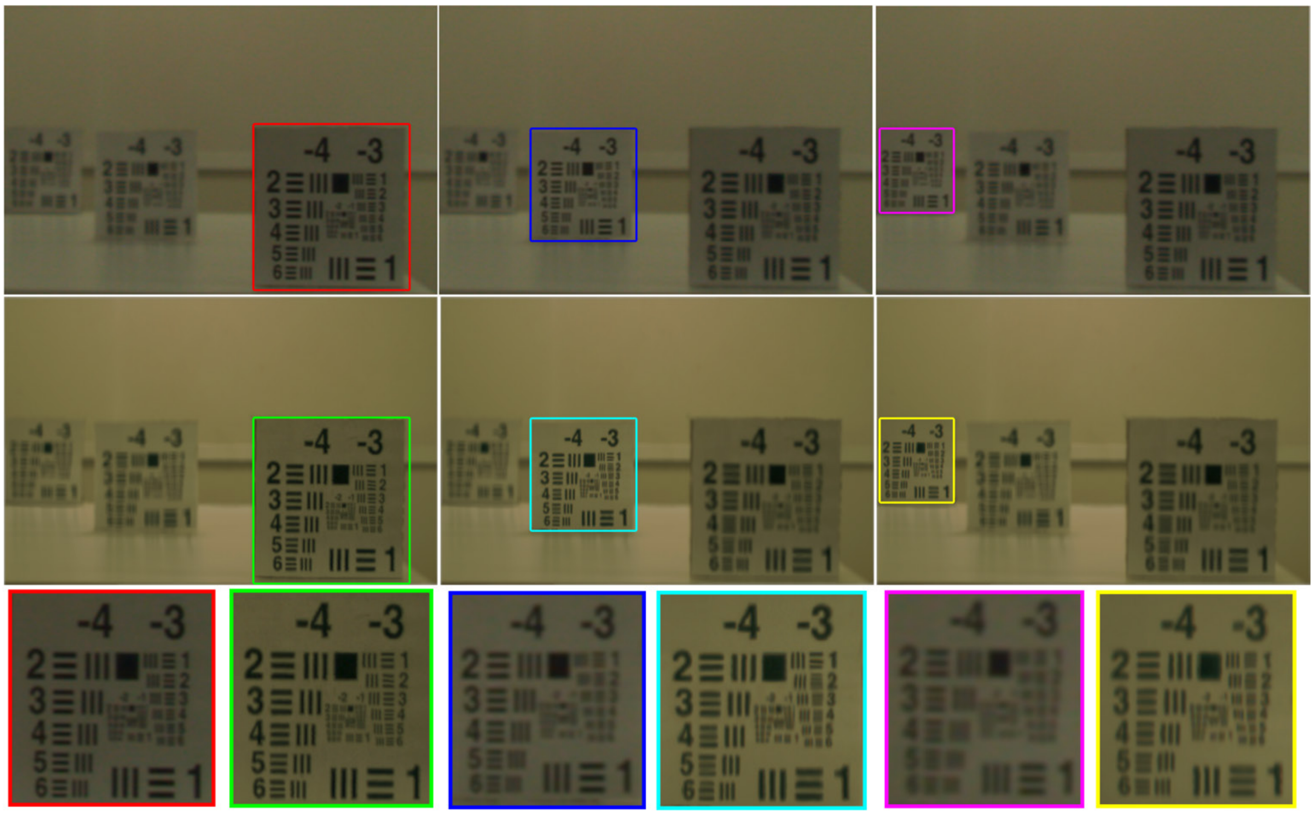

Figure 2: Post-capture digital refocusing using shift-and-sum technique. (Top row) Lytro light field images refocused at close, middle and far depths. (Middle row) Resolution enhanced light field images refocused at close, middle and far depths. (Bottom row) Zoomed-in focused regions before and after resolution enhancement. (The shift amounts to focus at a region are determined using the Fourier based motion estimation method given in [11].)

spatial resolution light field is obtained. Second, the separation between the light field camera and the regular camera produces a large baseline, improving the depth range and accuracy. Fixing the cameras as a stereo system also enables offline calibration and rectification, which takes off from the computational cost in online application.

II. Related work

There are various methods, in the light field literature, proposed to address the low spatial resolution and narrow baseline issues. Regarding the spatial resolution issue, the application of super-resolution restoration to the sub-aperture images of light field data has been investigated and demonstrated in improving the quality of light field images [12] [13]. Regarding the narrow baseline issue, sub-pixel accurate disparity estimation algorithms specific to the light field data have been proposed [14]. It is also proposed to use defocus and shading cues to improve the disparity estimation accuracy [15].

An alternative approach, proposed in [16], is to merge light field images with a high resolution image taken by a regular camera. The method uses dictionary learning and patch matching to enhance the resolution of Lytro images utilizing the high resolution image captured by a DSLR camera. Our proposed method is based on the same hybrid imaging idea. While the camera positions are arbitrary in [16], in our method, we set up a fixed stereo system, which enables calibration of the cameras and reduces to the computational cost of the merging process. Our system effectively increases the baseline, improving both the accuracy and range of the depth map. Related to our work, we can also mention multi-baseline stereo systems, which consist of several regular cameras [17]. Such multi-baseline systems have been shown to have uniform depth accuracy over a wide range, as opposed to traditional stereo vision systems, which have depth dependent disparity accuracy.

III. Proposed Method

To enhance spatial resolution of each light field sub-aperture image, we utilize the high resolution image captured by the regular camera. This requires registration of light field and regular camera images. For computational efficiency, we perform offline calibration between the regular camera and the middle sub-aperture of the Lytro camera. For a new image set, the pre-computed calibration parameters are used to rectify the regular camera image. It is of course necessary to estimate correspondence for a new image set; however, the rectification step reduces the search space thus speeding up the process.

We further increase speed of the registration process by using the fact that light field sub-aperture images are captured on a regular grid. This is done by estimating the optical flow between only three sub-aperture images and deducing the rest from these three optical flow fields: Essentially, we estimate the optical flow between the rectified regular image and the middle sub-aperture LF image, the leftmost sub-aperture LF image and the topmost sub-aperture LF image. The flow vectors are then interpolated and extrapolated according to their positions with respect to the middle, leftmost, and topmost sub-aperture images. In case of the Lytro camera, there are 11x11=121 sub-aperture images; with this process, we cut down images, for which the correspondence should be estimated, from 121 to only 3. We tested this method with real data, and did not notice any significant performance degradation.

Once a light field sub-aperture image and the regular camera image are registered, we can use an image fusion algorithm to transfer the resolution of the regular camera image to the light field sub-aperture image. The problem of image fusion for resolution enhancement has been well studied in the literature, with applications in satellite imaging as pan-sharpening [18] and in digital camera pipelines as demosaicking [19].

IV. Experimental Results

Our stereo system consists of a Lytro camera, with spatial resolution 380x380 for each sub-aperture image, and an AVT Mako CMOS camera, with spatial resolution of 1200x780. The light field data is decoded by [20] to obtain 11x11 sub-aperture images. Each sub-aperture image is initially interpolated to match the size of the regular camera. Calibration between between the middle sub-aperture image and the regular camera image is done by Matlab stereo calibration toolbox. The regular image is first photometrically transformed to match the Lytro images by performing histogram based intensity mapping of each channel [21]. The optical flow vectors between the middle, leftmost, and topmost sub-aperture images are estimated using the algorithm given in [22]. These vectors are then interpolated/extrapolated to obtain the optical flow vectors for each sub-aperture image. Warped images and the residuals given in figures 3 and 4 show that the registration process works effectively.

One of the key features of light field imaging is post-capture digital refocusing through a simple shift-and-sum procedure [2]. In Figure 2, we show refocusing at different distances with the Lytro light field images and the resolution-enhanced sub-aperture images obtained through photometric registration and warping. It is seen that we can obtain sharper refocusing compared to the original Lytro images.

Another feature of light field imaging is the ability of estimating depth. However, it is known that the depth accuracy and range is low in MLA based systems due to narrow baseline. With our hybrid stereo imaging system, this issue is addressed. In Figure 5, we see target objects placed with 10cm separation, starting from 60cm away from the imaging system. On the left, the leftmost and rightmost sub-aperture images of the Lytro camera are shown on top of each other. On the right, the middle sub-aperture image and the regular camera image are shown on top of each other. It is clear that the disparity is very small and diminishes quickly on the left, making it very difficult to have an accurate depth estimation. On the right, the hybrid system has larger disparity and extended range due to its wide baseline.

Figure 6 shows the disparity in pixels at the centers of the targets given in Figure 5. For the original Lytro light field, the

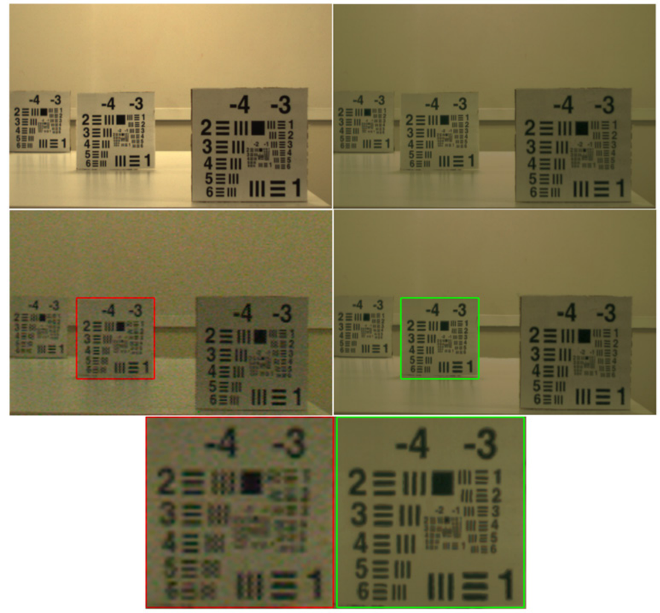

Figure 3: (Top left) Regular camera image. (Top right) Regular camera image after photometric registration. (Middle left) One of the bicubicly resized Lytro sub-aperture image. (Middle right) Regular camera image after warping on the Lytro sub-aperture image. (Bottom left) Zoomed-in center region from the Lytro sub-aperture image. (Bottom Right) Zoomed-in center region from the warped regular camera image.

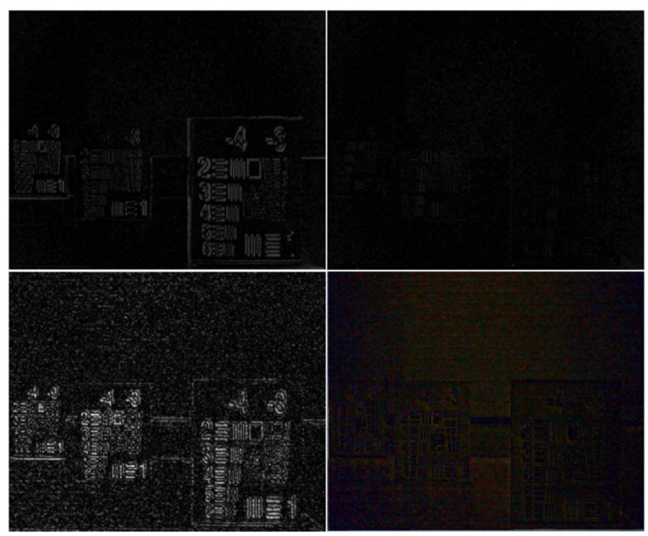

Figure 4: (Top left and right) Difference between the Lytro middle sub-aperture image and the rectified regular camera image before warping and after warping. (Bottom left and right) Difference between a Lytro sub-aperture image, for which the motion vectors are interpolated, and the Lytro middle sub-aperture before warping and after warping.

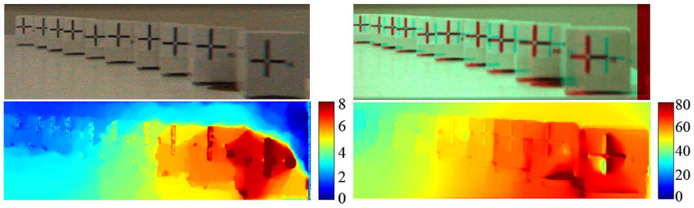

Figure 5: (Left) Disparity between extreme Lytro sub-aperture images; (Right) Disparity between Lytro middle sub-aperture image and the regular camera image.

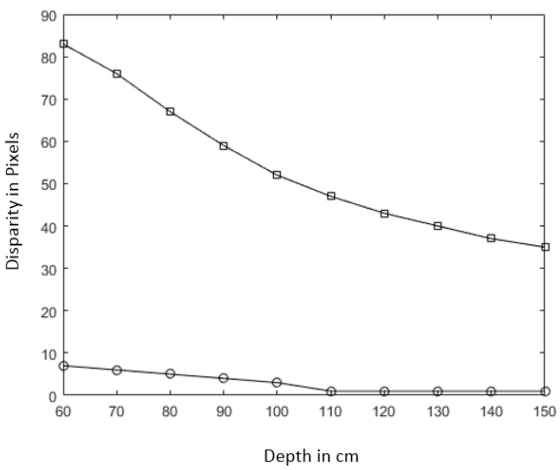

Figure 6: Disparities at the centers of the target objects given in Figure 5. (Circle) Disparities for the original Lytro leftmost and rightmost sub-aperture images. (Square) Disparities for the hybrid system between the Lytro middle sub-aperture image and the regular camera image.

disparity becomes indistinguishable after about 100cm, limiting the range of the depth estimation. For the hybrid system, however, the disparity difference can be easily distinguished, demonstrating the extended depth range.

One advantage that light field camera brings to the hybrid stereo imaging system is the ability to focus at very close objects. Hence the hybrid stereo imaging extends the depth range not only towards the far objects but towards the closer object as well.

V. Conclusions

In this paper, we present a hybrid stereo vision system that includes a light field camera and a regular camera. The system, while keeping the capabilities of light field imaging, improves spatial resolution as well as depth estimation range and accuracy. Because the system allows pre-calibration of images, the registration of low-resolution light field sub-aperture images and high-resolution regular camera images is simplified: Limiting the search range for image registration along the epipolar lines provides computational advantage and potential improvement in accuracy.

References

[1] E. H. Adelson and J. R. Bergen. The plenoptic function and the elements of early vision. In *Computation Models of Visual Processing*. Editors: Landy, M., Movshon, J. A. Cambridge: MIT Press, 1991.

[2] M. Levoy and P. Hanrahan. Light field rendering. In *ACM conference on Computer Graphics and Interactive Techniques*, pages 31–42, 1996.

[3] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *ACM conference on Computer Graphics and Interactive Techniques*, pages 43–54, 1996.

[4] 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. In *ACM Transactions on Graphics*, pages 765–776, 2005.

[5] R. Ng. Fourier slice photography. In *ACM Transactions on Graphics*, pages 735–744, 2005.

[6] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. In *ACM Transactions on Graphics*, pages 1–12, 2007.

[7] T. Georgeiv, K. C. Zheng, B. Curless, D. Salesin, S. Nayar, and C. Intwala. Spatio-angular resolution tradeoffs in integral photography. In *Eurographics Conference on Rendering Techniques*, pages 263–272. Eurographics Association, 2006.

[8] Lytro, inc. https://www.lytro.com/.

[9] Raytrix, gmbh. https://www.raytrix.de/.

[10] Z. Yu, X. Guo, H. Ling, A. Lumsdaine, and J. Yu. Line assisted light field triangulation and stereo matching. In *IEEE International Conference on Computer Vision*, pages 2792–2799, 2013.

[11] M. G. Sicairos, S. T. Thurman, and J. R. Fienup. Efficient subpixel image registration algorithms. In *Optical Letter*, pages 156–158. OSA, 2008.

[12] T. E. Bishop, S. Zanetti, and P. Favaro. Light field superresolution. In *IEEE International Conference on Computational Photography*, pages 1–9, 2009.

[13] D. Cho, M. Lee, S. Kim, and Y. Tai. Modeling the calibration pipeline of the lytro camera for high quality light-field image reconstruction. In *IEEE International Conference on Computer Vision*, pages 3280–3287, 2013.

[14] T. Georgiev and A. Lumsdaine. Reducing plenoptic camera artifacts. In *Computer Graphics Forum*, pages 1955–1968. Blackwell Publishing Ltd, 2010.

[15] M.W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi. Depth from combining defocus and correspondence using light-field cameras. In *IEEE International Conference on Computer Vision*, pages 673–680, 2013.

[16] V. Boominathan, K. Mitra, and A. Veeraraghavan. Improving resolution and depth-of-field of light field cameras using a hybrid imaging system. In *IEEE International Conference on Computational Photography*, pages 1–10, 2014.

[17] M. Okutomi and T. Kanade. A multiple-baseline stereo. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 63–69, 1991.

[18] H. B. Mitchel. Image fusion: theories, techniques and applications. 2013.

[19] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau. Color plane interpolation using alternating projections. In *Image Processing, IEEE Transactions on*, pages 997–1013, 2003.

[20] D. G. Dansereau, O. Pizarro, and S. B. Williams. Decoding, calibration and rectification for lenselet-based plenoptic cameras. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1027–1034, 2013.

[21] I. Hossain and B. K. Gunturk. Joint photometric registration and optical flow estimation. In *IEEE International Conference on Image Processing*, pages 1201–1204, 2010.

[22] C. Liu. Beyond pixels: exploring new representations and applications for motion analysis. MIT, 2009.