https://doi.org/10.2352/EI.2023.35.17.3DIA-107
© 2023, Society for Imaging Science and Technology

# A Comparative Evaluation of 3D Geometries of Scenes Estimated using Factor Graph Based Disparity Estimation Algorithms

***Hanieh Shabanian; Computer Science Department, School of Computing and Analytics, Northern Kentucky University, Highland Heights, Kentucky, USA***
***Madhusudhanan Balasubramanian; Electrical and Computer Engineering Department, University of Memphis, Memphis, Tennessee, USA***

## Abstract

*Passive stereo vision systems are useful for estimating 3D geometries from digital images similar to the human biological system. In general, two cameras are situated at a known distance from each other and simultaneously capture images of the same scene from different views. This paper presents a comparative evaluation of 3D geometries of scenes estimated by three disparity estimation algorithms, namely the hybrid stereo matching algorithm (HCS), factor graph-based stereo matching algorithm (FGS), and a multi-resolution FGS algorithm (MR-FGS). Comparative studies were conducted using our stereo imaging system as well as hand-held, consumer-market digital cameras and camera phones of a variety of makes/models. Based on our experimental results, the factor graph algorithm (FGS) and multi-resolution factor graph algorithm (MR-FGS) result in a higher level of 3D reconstruction accuracy than the HCS algorithm. When compared with the FGS algorithm, MR-FGS provides a significant improvement in the disparity contrast along the depth boundaries and minimal depth discontinuities.*

## Introduction

The scene being imaged may consist of several elements or objects. Because photographs and images are only 2D projections of 3D scenes, they cannot be used to estimate the scene 3D geometry. In order to reconstruct the 3D geometry of a scene non-invasively, two or more distinct scene-to-camera orientations are required. An image can be used to estimate any point's 3D coordinate by triangulating the light rays that emerge from the scene onto two or more cameras by modeling a point within scene space as a virtual ray of light from the scene onto the sensor. There are many computer vision applications that benefit from non-invasive 3D reconstructions from multiple lower-dimensional observations, such as robotic navigation [1], satellite image based reconstructions [2], reconstructions using aerial imaging [3], estimating depth distribution of water bodies [4].

Based on a taxonomy of methods [5], 3D acquisition methods are broadly categorized as active and passive methods. The active reconstruction methods use controlled illuminations such as structured illumination [6], photometric stereo [7], and shape from shading techniques [8]. Unlike the active reconstruction algorithms, which require spatial and temporal modulation of the scene illumination, passive methods do not require any specialized scene illumination requirement. Active reconstruction techniques, in general, are more accurate but also more expensive because of the unique lighting needs. With recent advancements in the 3D reconstruction process, such as disparity estimation [9], passive methods can be employed for a wide range of computer vision problems, such as for autonomous navigation.

According to the quantity of images or views with distinct scene-camera orientations needed, passive algorithms are further divided into single and multi-viewpoint algorithms. The *passive stereo vision* needs two images of a scene, each of which has a distinct camera orientation and has significant overlap between the images. 3D coordinates of scene points are estimated based on the disparity of pixel coordinates of scene points mapped onto stereo cameras. As a result, accurate stereo disparity estimates are crucial for reconstructing the 3D geometry of a scene with high accuracy.

Prior to this study, we developed three disparity estimation algorithms with higher accuracy, namely the hybrid stereo matching algorithm (HCS) [10], factor graph-based stereo matching algorithm (FGS) [9], and multi-resolution FGS algorithm (MR-FGS) [11]. This paper aims to demonstrate passive 3D reconstruction using these new disparity estimation methods and assess the quality of their 3D reconstructions.

The remainder of this paper will describe the passive stereo vision system, briefly describe the previously developed disparity estimation algorithms, present 3D reconstruction results, and conclude with our findings.

## Passive Stereo Vision System

Uchida et al. [12] developed one of the early facial recognition systems using a passive stereo vision model. The stereo imaging system consisted of two stereo cameras (a total of 4 CCD cameras), each with a narrow baseline to minimize occluded pixels within each stereo camera. A wider baseline between these two stereo cameras was used to achieve a more accurate 3D reconstruction though such configuration leads to pixel occlusions. The final 3D information was obtained using integrated 3D data from left and right parallel cameras using a phase-based image matching technique.

In stereo vision systems with two cameras, intensity differences and illumination variations are generally due to smaller differences in focal lengths and zoom levels between the cameras. Thus, the accuracy of the estimated disparities or stereo correspondences are affected by differences in the pixel intensity differences and illumination differences. Lee et al. [13] introduced a new stereo vision setup, including a camera and a biprism to create two images with different viewpoints of a scene and reconstruct a few lines of the scene. Recently, Liang et al. [14] used this acquisition technology to reconstruct 3D information of the textureless weld pool surface during many penetration phases.

Passive stereo vision systems are useful for estimating 3D geometries from digital images similar to the human biological system. In general, two cameras are situated at a known distance from each other and simultaneously capture images of the same scene from different views. While several stereo cameras are commercially available such as Bumblebee XB3 and Bumblebee 2, Kinect 3D, MEGA-DCS, Ensenso N10, Capella, and Scorpion 3D Stinger, it is possible to build a passive stereo arrangement using only two webcam cameras.

The general 3D reconstruction pipeline involves identifying the optical characteristics of the cameras in the stereo setup using a camera calibration procedure (focal length, sensor distortion characterized as intrinsic parameters), geometrical relationship between the cameras (relative orientation and offset), correspondences between the stereo images (disparity estimation) and estimating the 3D coordinates of the entire scene imaged by the stereo camera arrangement. A geometric model of image formation based on a pinhole camera design is commonly used to relate the 2D pixel coordinates in the acquired images with the 3D object coordinates in the scene. Thus, the model allows to characterize any distortions in the geometrical mapping such as scaling and skewness introduced by the camera optics and its imaging sensor.

In a stereo arrangement, the relative geometries of two pinhole models (one for each camera) are represented using *extrinsic characteristics* of each camera, namely the relative position or offset and orientation with respect to a chosen *frame of reference* (reference coordinate system) in the scene. Specifically, the *essential* matrix of a camera is useful for geometrically relating the corresponding pixels in a stereo setup using a normalized pinhole model (where the effects of the camera intrinsic parameters are removed from the corresponding pixel coordinates). More specifically, with one of the cameras as a reference, the essential matrix is useful for estimating the relative orientation and offset characteristics of the second camera with respect to the reference camera. In an uncalibrated stereo setup, a *fundamental matrix* relates the corresponding pixel coordinates using a non-normalized pinhole model. The essential matrix and relative camera orientations are then estimated from the fundamental matrix. Finally, the 3D scene coordinates are estimated using the pinhole model of image formation using the estimated point correspondences and the camera parameters (intrinsic and extrinsic) [15].

In general, the reconstruction process described earlier provides a sparse geometric description of the scene and may not be sufficient for achieving a good visual representation of the scene. Therefore, disparity maps estimated from *rectified* stereo images are used for 3D reconstruction. In brief, rectification simplifies the stereo correspondence problem by applying a single linear transformation to the stereo image to make the image planes of both the cameras parallel to each other [16]. After the rectification process, the horizontal displacement between corresponding points in the first and second images is obtained using stereo matching. The horizontal separation of an object seen by first and second images (similar to the left and right eyes) creates stereo disparity [17]. Based on the triangulation process in epipolar geometry and the fact that the depth information is inversely proportional to the disparity between two corresponding points, the depth information is perceived by knowing the stereo disparity and distance between two cameras.

## Disparity Map Estimation Algorithms

Estimates of disparities among multiple views of a scene are useful for building a dense 3D architecture of a scene. Stereo matching algorithms can estimate a disparity map comprised of dense pixel-level correspondences between two views in a stereo pair. In the following subsections, we briefly present our previously developed disparity estimation techniques and subsequently evaluate these techniques for building 3D geometries of scenes using stereo cameras.

### *1. A hybrid of cross-correlation between stereo-image pairs and scene segmentation (HCS)*

In this study, we developed a new stereo-matching algorithm using guided image filtering (GIF)-based cost aggregation [10]. The main contribution of this framework is a combination of prior scene segmentation and cross-correlation between stereo images to produce an initial disparity map [10]. In brief, a new hybrid stereo-matching algorithm (HCS) was proposed that incorporates a priori distribution of disparity estimates and a normalized cross-correlation cost measure. The initial disparity map was efficiently calculated by guiding the search locations in the Middlebury benchmark stereo images. The cost volume was aggregated using edge-preserving guided filtering to improve accuracy. The proposed method was compared with recent state-of-the-art approaches, resulting in lower disparity errors in 8 out of the 15 tested stereo pairs among all evaluated non-learning methods. Although a relatively higher disparity error was observed in stereo images with large disparity, the overall performance of the proposed approach was better, as indicated by the lower average disparity error.

### *2. A novel factor graph-based optimization technique for stereo correspondence estimation (FGS)*

The FGS disparity estimation algorithm utilized spatial dependencies between image intensity and disparity estimates to produce more accurate disparity estimates [9]. In brief, image segments were formed by grouping scene elements from a reference image in each stereo pair and *a priori* disparity distributions were established using a select set of disparities within each segment. The factor graph model associated each pixel within a segment with its corresponding *a priori* disparity distribution. The spatial dependencies among disparity estimates were taken into account to estimate *a posteriori* disparity for each pixel using a larger and spatially variable neighborhood system. This resulted in improved disparity estimates in areas with smoother texture and reduced occlusion, as well as higher disparity contrast along depth boundaries. The final disparity map was produced through post-processing of the generated posterior disparity map.

The proposed factor graph methodology can be utilized to arrive at the maximum *a posteriori* estimates from models or optimization issues that have a complex relationship between hidden variables. When factor graph-based inference problems require the computation of message convergence, using *a priori* distributions with a shorter support and spatial dependencies can be helpful. A new factor-graph-based disparity estimation algorithm was rigorously evaluated using Middlebury benchmark stereo datasets. Factor-graph algorithms provide higher accuracy disparity estimates compared to recent non-learning and learning-based disparity estimation algorithms using the Middlebury evaluation dataset version 3.05.

### *3. Multi-Resolution Factor Graph Based Stereo Correspondence Algorithm (MR-FGS)*

The MR-FGS algorithm is a probabilistic graphical model that uses spatial and multi-resolution dependencies among random variables of the multi-resolution factor graph (MR-FGS) [11]. With a potential for a variety of applications, previously we demonstrated that the MR-FGS model improved accuracy and contrast along depth boundaries over the earlier FGS model. A comparison was made between the proposed multi-resolution probabilistic factor graph model and the FGS results using the Middlebury benchmark stereo datasets. We found that the multi-resolution factor graph algorithm improved depth boundary contrast and provided more accurate disparity estimates. In comparison to the FGS algorithm, the MR-FGS algorithm eliminates the need for computationally intensive left-right consistency checks for the disparity estimates.

## 3D reconstruction experimental Results

In this section, we demonstrate the 3D reconstruction of stereo image pairs utilizing the proposed novel disparity map methods after conducting sparse and dense 3D reconstruction methodologies. Using a custom-built stereo imaging system, we captured stereo images and generated sparse and dense 3D reconstructions. In addition, we used stereo images captured using hand-held, consumer-market digital cameras, and camera phones of a variety of makes/models available from [18] to demonstrate the 3D reconstruction procedure.

### *1. Results of Real-World Stereo Images Captured From Our Imaging System*

We constructed a simple imaging system consisting of two webcams (USB2.0 PC Camera, focal length 3.85mm 10x) placed six inches apart. The focus of each camera was verified and adjusted as a pair. The illustration in Figure 1 depicts our simple stereo camera setup.

Figure 1: A basic stereo imaging setup was created. The prototype is comprised of two USB webcam cameras with a 3.85mm 10x focal length, and were positioned six inches apart from each other.

Stereo images captured using our stereo imaging system are shown in Figure 2, image (a) captured by the left camera and image (b) captured by the right camera simultaneously. After using the proposed disparity estimation algorithms, (c) initial disparity map was obtained using HCS algorithm [10] without cost aggregation step, (b) HCS after cost aggregation [10], (d) FGS algorithm [9], and (f) MR-FGS algorithm [11].

Figures 3 (a) to (d) display illustrations of dense 3D reconstructions obtained from the initial disparity maps and the disparity maps of HCS [10], FGS [9], and MR-FGS algorithm [11].

### *2. Results of Real-World Stereo Images Captured From Other Imaging Systems*

Aside from the stereo images captured with our simple stereo imaging setup, we also used consumer-market digital cameras and camera phones of a variety of brands and models to produce real-world stereo images [18].

Figures 4 and 5 show disparity estimations and sparse/dense 3D reconstruction results of the *hand print* stereo images. Figures 6 and 7 show disparity estimation and 3D reconstruction results of the church stereo images.

As illustrated by the 3D reconstructions, the accuracy of the 3D point-cloud estimated from the stereo image pairs was significantly higher for the FGS and MR-FGS algorithms when compared to the HCS algorithm. In particular, there were fewer depth discontinuities and reduced depth jumps along object boundaries when reconstructions were based on MR-FGS disparity maps.

## Conclusions

We compared 3D scene geometries estimated using three recent disparity estimation algorithms, namely the Hybrid Stereo Matching Algorithm (HCS) [10], the Factor Graph-based Stereo Matching Algorithm (FGS) [9], and the Multi-Resolution Factor Graph-based Stereo Matching Algorithm (MR-FGS) [11]. Comparisons were made using images captured with our stereo setup as well as using various consumer-grade cameras and camera phones, as documented in [18]. Our experiment results show that both the FGS [9] and MR-FGS [11] algorithms produce more accurate 3D reconstructions compared to the HCS [10] algorithm. The MR-FGS algorithm [11] showed significant improvement in the quality of 3D reconstruction as evidenced by enhanced disparity contrast along depth boundaries and reduced depth discontinuities (spurious depth jumps) when compared to the FGS algorithm [9].

## References


[1] Y. Sumi, Y. Kawai, T. Yoshimi, and F. Tomita, "3d object recognition in cluttered environments by segment-based stereo vision," *International Journal of Computer Vision*, vol. 46, no. 1, pp. 5–23, 2002.

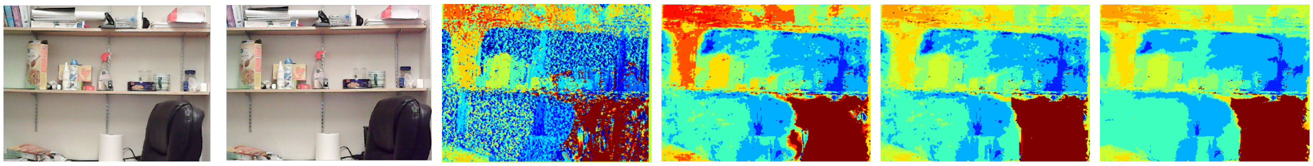
[2] T. Moons, D. Frère, J. Vandekerckhove, and L. Van Gool, "Automatic modelling and 3d reconstruction of urban house roofs from high resolution aerial imagery," in *European Conference on Computer Vision*. Springer, 1998, pp. 410–425.

[3] J. Byrne, E. O'Keeffe, D. Lennon, and D. F. Laefer, "3d reconstructions using unstabilized video footage from an unmanned aerial vehicle," *Journal of Imaging*, vol. 3, no. 2, p. 15, 2017.

[4] M. Marghany, M. Hashim, and A. P. Cracknell, "3-d reconstruction of coastal bathymetry from airsar/polsar data," *Chinese Journal of Oceanology and Limnology*, vol. 27, no. 1, p. 117, 2009.

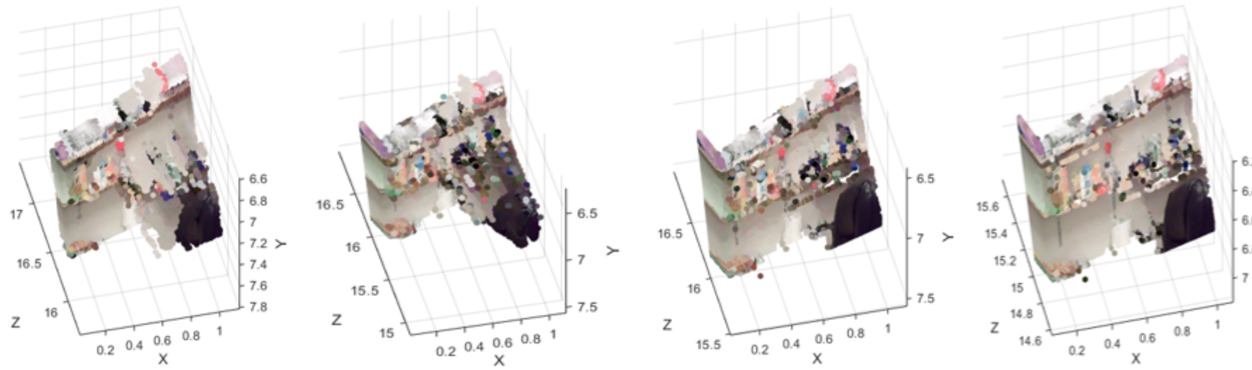
[5] T. Moons, L. Van Gool, and M. Vergauwen, *3D reconstruction from multiple images: principles*. Now Publishers Inc, 2009.

[6] K. Zhang, M. Yan, T. Huang, J. Zheng, and Z. Li, "3d reconstruction of complex spatial weld seam for autonomous welding



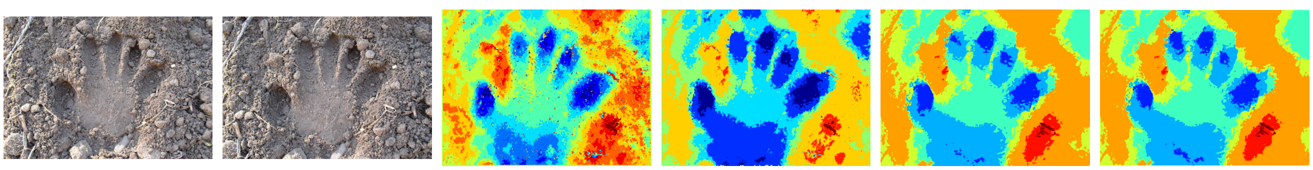
(a) (b) (c) (d) (e) (f)

Figure 2: Disparity between the stereo images of an office view. (a) left image (b) right image, (c) initial disparity map, computed using the proposed approach in [10] without cost-aggregation and post-processing steps (d) HCS algorithm [10] , (e) FGS algorithm [9], and (f) MR-FGS algorithm [11].



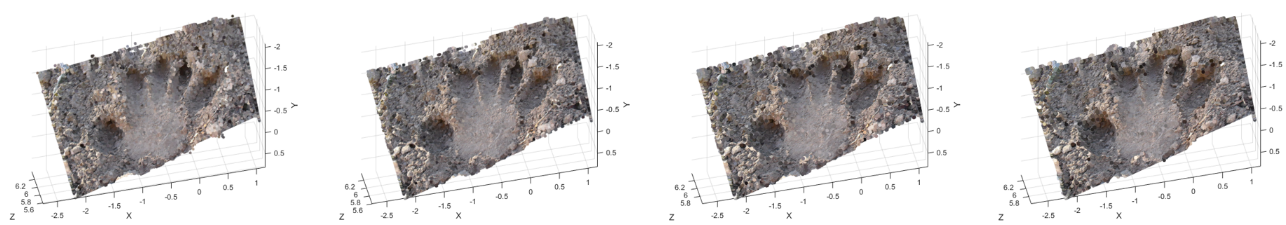

(a) (b) (c) (d)

Figure 3: 3D view of an *office scene* reconstructed based on stereo disparity maps estimated using (a) an initial disparity map (b) HCS disparity map, (c) FGS disparity map, and (d) the MR-FGS disparity map.



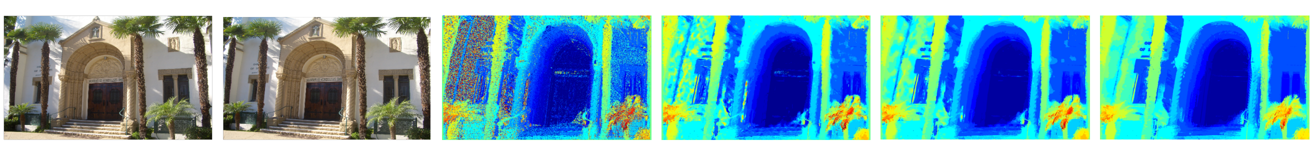
(a) (b) (c) (d) (e) (f)

Figure 4: Disparity results of the hand print stereo images. (a) left image (b) right image, (c) initial disparity map computed using the proposed approach in [10] without cost-aggregation and post-processing steps (d) HCS algorithm [10] , (e) FGS algorithm [9], and (f) the MR-FGS algorithm [11].



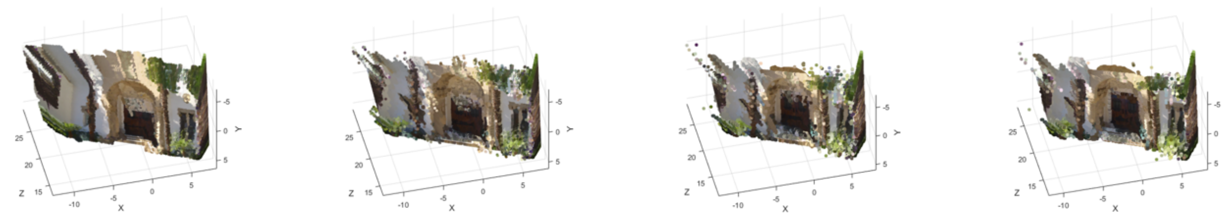
(a) (b) (c) (d)

Figure 5: 3D views of the *hand print* reconstructed based on stereo disparity maps estimated using (a) an initial disparity map (b) HCS disparity map, (c) FGS disparity map, and (d) MR-FGS disparity map.



(a) (b) (c) (d) (e) (f)

Figure 6: Disparity estimates of the *church* stereo images (a) left image and (b) right image; (c) an initial disparity map, computed using the proposed approach in [10] without cost-aggregation and post-processing steps considering occluded regions (d) HCS algorithm [10] , (e) FGS algorithm [9], and (f) the MR-FGS algorithm [11]



(a) (b) (c) (d)

Figure 7: 3D reconstruction of the *church* stereo images based on the disparity maps estimated using (a) an initial disparity map (b) HCS disparity map, (c) FGS disparity map, and (d) the MR-FGS disparity map [11].


by laser structured light scanning," *Journal of Manufacturing Processes*, vol. 39, pp. 200–207, 2019.

[7] G. Vogiatzis and C. Hernández, "Practical 3d reconstruction based on photometric stereo," in *Computer vision*. Springer, 2010, pp. 313–345.

[8] C. Q. Forster and C. L. Tozzi, "Towards 3d reconstruction of endoscope images using shape from shading," in *Proceedings 13th Brazilian symposium on computer graphics and image processing (Cat. No. PR00878)*. IEEE, 2000, pp. 90–96.

[9] H. Shabanian and M. Balasubramanian, "A novel factor graph-based optimization technique for stereo correspondence estimation," *Scientific Reports*, vol. 12, no. 1, p. 15613, 2022.

[10] ——, "A new hybrid stereo disparity estimation algorithm with guided image filtering-based cost aggregation," 2021.

[11] ——, "Multi-resolution factor graph based stereo correspondence algorithm," *arXiv preprint arXiv:2202.01309*, 2022.

[12] N. Uchida, T. Shibahara, T. Aoki, H. Nakajima, and K. Kobayashi, "3d face recognition using passive stereo vision," in *IEEE International Conference on Image Processing 2005*, vol. 2. IEEE, 2005, pp. II–950.

[13] D. Lee and I. Kweon, "A novel stereo camera system by a biprism," *IEEE Transactions on Robotics and Automation*, vol. 16, no. 5, pp. 528–541, 2000.

[14] Z. Liang, H. Chang, Q. Wang, D. Wang, and Y. Zhang, "3d reconstruction of weld pool surface in pulsed gmaw by passive biprism stereo vision," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 3091–3097, 2019.

[15] S. J. Prince, *Computer vision: models, learning, and inference*. Cambridge University Press, 2012.

[16] R. I. Hartley, "Theory and practice of projective rectification," *International Journal of Computer Vision*, vol. 35, no. 2, pp. 115–127, 1999.

[17] H. Shabanian, "Stereo correspondence using probabilistic graphical models for 3d reconstruction," Ph.D. dissertation, The University of Memphis, 2021.

[18] Y.-F. Wang, "A comparison study of five 3d modeling systems based on the sfm principles," Technical Report 2011-01. Visualize Inc., Goleta, USA, Tech. Rep., 2011.


## Author Biography

*Hanieh Shabanian, Ph.D., is an Assistant Professor in Computer Science at the School of Computing and Analytics at Northern Kentucky University. Dr. Shabanian earned her Ph.D. and Ms. in Computer Engineering with a prior undergraduate degree in Software Engineering. Her research interest and expertise are at the intersection of computer vision, machine learning, and image processing.*

*Madhusudhanan Balasubramanian, Ph.D. is the Director and Principal Investigator of the Computational Ocularscience laboratory. Currently, Dr. Balasubramanian serves as Associate Professor in the Department of Electrical and Computer Engineering at the University of Memphis. His current research interests include computer vision, optical imaging and functional assessment of ocular changes in ocular hypertension and glaucoma, understanding the role of retinal structure and physiology in motion detection, and biophysical modeling of ocular therapies.*