Simulating multi-camera imaging systems for depth estimation, enhanced photography and video effects

Ginni Grover, Ram Nalla and Ramkumar Narayanswamy
Intel Corporation, 2200 Mission College Boulevard, Santa Clara, CA 95054
ginni.grover@intel.com

Abstract: Multi-camera systems have numerous applications in computational photography such as for depth estimation to enable many photography effects. We present a method for simulating multi-camera systems and results comparing different algorithms and camera geometries.
OCIS codes: (150.6910) Three-dimensional sensing; (150.0155) Machine vision optics

1. Introduction

Simulation of digital cameras has recently been proposed [1,2] and is also available as commercial software packages such as Imaging Systems Evaluation toolbox (ISET) [3] and simulation capability provided by FiveFocal [4]. These systems accurately model scene radiometry, optics and sensors of a traditional camera and allow simulating scenes with different camera parameters and scene properties. Such simulations of digital cameras have previously allowed scientists and engineers to quickly analyze, understand and designs various camera components [2].

In this paper, we report a full system simulation method based on ISET and photorealistic computer generated content for simulating hybrid multi-camera systems. Using the simulated images, we present two cases to evaluate depth estimation algorithms and multi-camera geometries. These examples are representative of the analysis which can be conducted with simulation based evaluation.

2. Multi-camera image simulation

Image simulation package ISET divides camera simulation in different components; namely scene, optics, sensor and image processing, allowing control of parameters for each of those. The input parameters are either manufacture specified or measured. In this work, we use ISET as the basis of our image simulation for multi-camera system. As an input to ISET, we computer generate scene images for each of the cameras using standard rendering packages such as VRay and 3DStudioMax. These images contain linear, high dynamic range scene irradiance data.

The next steps of the simulator are briefly described here. Please refer to Ref. [1] for full details of ISET simulator. As a next step, we simulate the effect of optics on the scene i.e. calculate irradiance right at the sensor plane albeit not ‘captured’ by the sensor. The optics simulation allows simulating either ideal lens blur or blurring the scene with aberrations as in real lenses. We have modified the optics module to speed up application of the shift-variant blur by using windowed Fourier transform technique. In practice, we notice that this improves simulation speed by 40x over brute force method of applying PSF at each pixel. In the next step, photon capture and electronic effects of the sensor are applied which include but are not limited to pixelation, quantum efficiency, color filter arrays, noise, etc. The sensor module delivers the RAW image which is passed to the imaging processing module to convert to RGB image.

An example of simulated images using different camera viewpoints in the same scene is shown in Fig. 1. It can be seen from the images that the cameras are shifted vertically, leading to vertical disparity between the captured images.

3. Depth estimation with simulated images

We present two cases which simulate images for multi-camera systems and analyze the depth measurement performance of these systems. The simulated images are appropriately rectified and passed to disparity estimation algorithm [5].

3.1 Comparing error in depth measurement for different baselines in a 3-camera system

As is well known in theory, error in depth estimation depends on object depth, baseline and focal length of the cameras, and error in disparity estimation. Based on the disparity estimation algorithm used and scene properties such as texture, the error in disparity estimation varies. We compare error in measured depth for a 3-camera system, with all the cameras aligned on a line. The cameras simulated are generic mobile phone cameras with 2MP sensor resolution. A special scene with equally spaced boards along the depth is simulated. ‘Dead-leaves’ charts are

Figure 1: Simulating multi-camera systems. (a) and (b) are two different camera positions capturing the same scene. The cameras are shifted vertically leading to vertical disparity between different objects, based on their depth with respect to the cameras. Few regions in the image such as the fork on a napkin which is closer to the camera has so much shift that it is not visible in (a). Farther objects have smaller shift between the two images.

displayed on these boards since they are scale invariant [6]. The boards are spaced a meter apart, starting from depth positions of 1 m to 10 m. Figure 2 shows the results of depth estimation with this setup. It can be seen that, as expected from theory, the error increases as the depth of the object increases. Also, the error reduces if the maximum baseline between the cameras increases. The theoretical curves are also shown in the plot and the simulation follows the limits closely. Going into further details, it can be seen that in the case with 52 mm baseline, the placement of the cameras changes the error. It means that with the particular disparity algorithm used here, two cameras spaced closer are better for depth estimation than equally spaced three cameras.

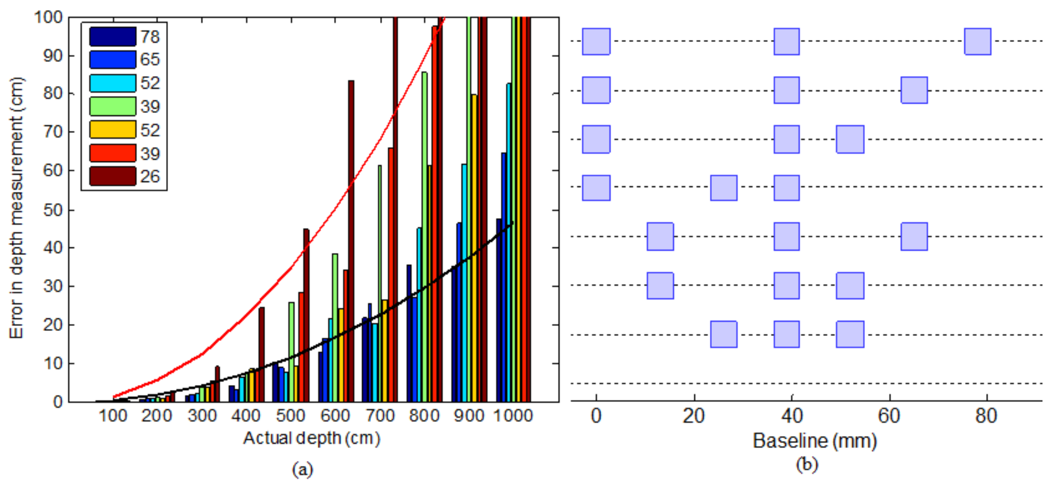

Figure 2: (a) Plot of error in measured depth vs depth for different baselines in a 3-camera array (all 1080p mobile cameras). The legend color indicates camera systems with different (longest) baselines. (b) shows the camera geometry arranged in the same order as the legend in (a), going from wider baseline to shorter baseline. The black and red curves in (a) are theoretical curves with 0.5 pix disparity error for the longest and shortest baselines, respectively.

3.2 Analyzing the effect of sensor resolution on depth resolution

In the next case, we compare stereo camera pairs with different sensor resolutions while keeping all the other optical and sensor parameters the same. Figure 3 shows the simulation results and depth maps of two systems with a 720p and a 1080p mobile phone camera with the optical specs of a generic 1080p camera and 37.5 mm baseline. From the results we note that that higher resolution camera not only provides higher resolution depth map but provides higher depth resolution and depth range. The measured discrete depth levels in Fig 3(d) and (e) show that the higher resolution 1080p camera has 8 discrete depth levels whereas the 720p camera has 7 depth levels. This consequence of higher resolution in the depth map can improve accuracy in applications such as depth based segmentation.

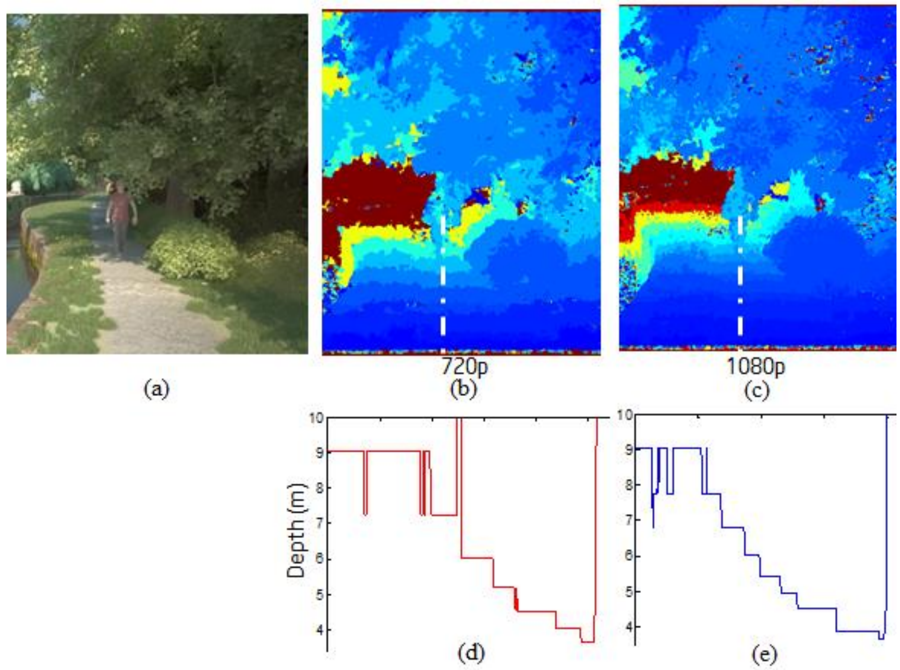

Figure 3: Depth estimation for stereo pairs with simulated images. (a) A photorealistic scene simulated for a stereo pair with a baseline of 37.5 mm. Two cases with both the cameras as generic mobile phone cameras with 720p and 1080p resolution, respectively, are simulated and analyzed, keeping all other camera parameters the same. (b) and (c) show the depth maps for both the cases. Colormap show increasing depth from blue to red. (d) and (e) are slices along the dashed white vertical lines in (b) and (c) in the depth maps. (b)-(e) show that using a higher resolution camera leads to finer depth resolution and edges.

4. Conclusion

In this paper, we present a method to simulate multi-camera systems using photorealistic rendering and image simulation tool ISET. This unique approach to use computer generated content and creating multi-camera systems has many advantages of speed, ease of use than hardware prototyping. It allows testing camera geometries, camera parameters and processing algorithms and doing iterative optimization to obtain optimal parameters for required applications. Here we have shown two examples out of many of the use cases this method is applicable to and the studies we have conducted.

5. References

1. J. E. Farrell, P. B. Catrysse, and B. A. Wandell, "Digital camera simulation," Appl. Opt. **51**, A80–A90 (2012).
2. J. Chen, K. Venkataraman, D. Bakin, B. Rodricks, R. Gravelle, P. Rao, and Y. Ni, "Digital Camera Imaging System Simulation," IEEE Transactions on Electron Devices **56**, 2496–2505 (2009).
3. Imaging Systems Evaluation Toolbox (ISET), http://imageval.com/
4. Five Focal, http://www.fivefocal.com/engineering-services/modeling-and-simulation/
5. D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," International Journal of Computer Vision **7**, 7–42 (2002).
6. F. Cao, F. Guichard, and H. Hornung, "Dead leaves model for measuring texture quality on a digital camera" in http://spie.org/Publications/Proceedings/Paper/10.1117/12.838902 edited by F. Imai, N. Sampat, and F. Xiao, (2010), p. 75370E–75370E–8.