A Low-Power Low-Cost High-Speed 2D/3D Camera for Virtual Reality Headsets, Mobile Devices and Automobiles

Yibing M. Wang[1,3], Ilia Ovsiannikov[1,3], Jang-Woo You[3], Peter Deane[2], Dirk Smits[2], Yong-Hwa Park[3*], Maarten Niesten[2], Sungwoo Hwang [3], Chilhee Chung[3]

[1]Samsung Semiconductor, Inc., 2 N. Lake Ave, Suite 230, Pasadena, CA 91101
[2]Samsung Strategy and Innovation Center, 2440 Sand Hill Road, Suite 200, Menlo Park, CA 94025
[3]Samsung Advanced Institute of Technology, 130 Samsung-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do, Korea 443-803 * Now with KAIST
Email: michelle.w@samsung.com
Phone: 1-626-683-3465 ext 305

Abstract - In this paper, we present a 2D/3D camera capable of capturing RGBZ video simultaneously with low cost, low power, strong ambient light suppression, no depth ambiguity, zero motion blur and minimum multipath reflection errors. The camera consists of a traditional 2D color image sensor with pixels as small as 1um, and a miniature MEMS laser scanner. Timestamping and epipolar plane imaging techniques are applied to simplify sensor design, strongly reject scattering and multipath errors, and greatly reduce image processing complexity. A prototype has been built to demonstrate the performance of such a camera.

1. INTRODUCTION

Although 3D imaging technology has matured considerably, mobile devices with 3D cameras have not become popular on consumer market, let alone the monolithic RGBZ cameras. One of the reasons is that existing 3D solutions have many drawbacks in performance. For example, Stereo camera cannot handle objects without texture. Structured light camera and stereo camera, which are both triangulation based, has very high computational complexity. Time-of-flight (TOF) imaging with lock-in pixels consumes large power and uses special pixels with large size. TOF and structured light methods in particular suffer from strong ambient light[1,2]. Dynamic vision sensor (DVS) and laser scan was used for 3D in [3]. It suffers from high temporal noise, inaccurate timestamp order and low resolution due to large pixel size. Another reason is that 2D RGB and 3D sensor usually use heterogeneous designs, so it is very hard to merge them on a single die. There were two previous attempts to mix TOF pixel and 4T pixels in the same pixel array [4,5]. However, [4] does not allow simultaneous capturing of depth with color, but in sequential fields, and its pixel performance cannot be optimized for each mode. In [5], although RGBZ can be captured simultaneously, it is at the cost of losing 2D pixels and sacrificing 2D image quality.

To build a monolithic RGBZ camera with low computational complexity and strong ambient light rejection, we propose a new epipolar plane camera system (Fig.1) based on the existing 2D camera design. With an additional laser point scanner, depth information can be captured and calculated based on triangulation principle. By utilizing the epipolar plane imaging technique, (similar to the structured light camera in [6]), 3D operation can be realized by simply modifying the timing of any 2D CMOS image sensors with column-parallel single-slope ADCs using either up or down counters, with the same sensor and pixel architecture and minimum circuit modification. As a result, concurrent RGBZ image/video capturing can be easily realized by interleaving 2D and 3D timing with no compromise to color and depth resolution and performance. To reduce computational complexity associated with correspondence searching, epipolar plane imaging, laser event detection and timestamping techniques are used. Other advantage of combining these techniques are strong ambient light rejection, no depth ambiguity and no motion blur in 3D images.

Taking S5K4E6 with 5M pixels used in Galaxy S6 as an example, with a scan interval between adjacent points of 50ns, and a maximum line scan rate of 40,000 lines per second, we can achieve a VGA size depth frame rate of 60fps. The power consumption and the module size of the camera stay the same as the original 2D camera design. The average optical power can be below 1mW. With 1.34um pixel and 40 degrees field of view (FOV), the depth error is between 0.12mm at 25cm and 1.8mm at 1m with 5cm disparity between the camera and the laser source. This compact and low power design is suitable for many applications, e.g., VR headsets, mobile devices and automobiles.

For 3D only systems, APD or SPAD arrays can also be used instead of regular PPD arrays.

2. Principle of Operation and Sensor Design

We start from a regular 2D color CIS sensor design with column-parallel single-slope ADCs. In 3D mode, the 2D pixel array is broken into multiple 1D line sensors from operation point of view. One line sensor consists of one or multiple rows of pixels. The laser scanner scans the scene line by line. The laser can be any color or NIR depending on the selection of the filter and camera design. With precise alignment, only one line sensor that lies on the laser scanning plane, which is the epipolar plane, needs to be enabled to image one horizontal pass of laser scan beam. During each pass, the laser pulses to create distinct spots at a fixed time interval. Each pixel in the selected line generates an event as soon as it captures the image of a laser spot. This event is time-stamped by the corresponding time-to-digital converter (TDC) in column, which is re-purposed by the column ADC of 2D mode.

As a result, the computational intensive correspondence problem is solved by epipolar plane imaging. And timestamping ensures no ambiguity in the depth measurement by matching pixel event timestamps to laser directions. Pixels generating events almost instantaneously keeps ambient light strongly suppressed. In addition, the depth image generated is free of motion blur and shows clean edges. In addition, epipolar plane imaging greatly reduces errors from scatters and multipath reflection. Any photon arriving away from the enabled line is ignored.

To explain in detail how 3D capture with an existing 2D design is realized, a typical timing diagram (Fig. 2) for 2D mode using digital CDS and an up counter is taken as an example. Assuming readers of this paper are familiar with image sensor and single-slope ADC designs, no circuit details are explained. In the diagram, RSEL, RST, TX, ADC_RST and RAMP are row select, pixel reset, transfer, column ADC reset and ramp signals, respectively. PIX_CDS, Counter Clock and ADC are pixel output after CDS, clock input to column ADC, and ADC output, respectively. Only RST, TX and RAMP signals need to be changed to enable 3D operation. Fig. 3 shows a timing diagram for 3D linear mode operation. Since we are trying to capture the laser event right after it occurs, integrating the photoelectrons with a preset integration time as in 2D mode does not work anymore. The pixel outputs have to be monitored all the time. So the TX gates of all pixels are kept on to allow detected photoelectrons to transfer to floating diffusions right away. Since the laser events of one pass will happen only on a known epipolar line, only

RSEL of the row(s) falling onto the epipolar line needs to be enabled. The pixels on these row(s) are initialized with a reset. So do the column ADCs. The pixels not on the epipolar line are kept in reset to prevent blooming if any. The ramp signal is used as a threshold in this mode. It falls to a preset voltage after the reset period is over. Once a pixel output is pulled below the threshold due to a laser event, the corresponding column ADC starts counting. All column ADCs finish counting at the end of a predefined counter period. Thus, pixels being hit earlier have larger ADC counts than the ones later. The time difference between adjacent timestamps should match the time interval of the laser pulses if time-of-flight is ignored.

The linear mode timing works well at low ambient light level, but not sufficient at higher level. At high ambient light level, if an NIR laser source is used, using a band pass filter can cut down the ambient generated electrons by 50 to 100 times. However, if color laser is used, for example, green, at 100,000lux and 32us exposure time, the ambient electrons generated at one pixel can be over 60,000 in count, which cause the photodiode to saturate. To solve this problem, we implement a logarithmic mode operation as shown in Fig. 4. The only difference between the two modes is that RST is held high in logarithmic mode. By doing this, the reset transistor in the pixel works in subthreshold region, and the FD node has a logarithmic relationship with the incident photocurrent. Noticed that the pixel output dips only briefly after an event is detected, a latch may need to be added to the comparator output inside the ADC, depending on the original circuit design.

3. Prototype and Camera Performance

A prototype is setup to demonstrate the performance of this 2D/3D camera. A Celluon PicoPro projector is setup side-by-side with a 2D camera as shown in Fig. 5. Red laser source is chosen for scanning. A demodulation TOF camera using NIR light source is used for comparison. The prototype has a baseline of 11.5cm, pixel pitch of 7um and a focal length of 4.16mm. Fig. 6 shows the theoretical result of depth accuracy of this camera and the experiment result. They match very well. To demonstrate the superb multipath reflection rejection performance of epipolar plane imaging technique, a fog test is setup. Fig. 7 shows the 3D images taken by the prototype and the TOF camera under clear and foggy conditions. The prototype shows only a little change in depth measurement while the TOF image is totally smeared. This technique can be a very powerful tool for automotive cameras, including scanning LiDAR, under inclement weather conditions.

4. Conclusion

The proposed 2D/3D camera enabling concurrent RGBZ imaging has many good features, i.e. high depth accuracy, small size, low cost, low power, no motion blur, strong ambient light rejection and multipath reflection rejection. It can be a good candidate for many different applications, such as 2D and 3D imaging for mobile devices, SLAM for VR headset, 3D backup camera for automobiles and so many more.

References

[1] Gupta, M. et al. Structured light in sunlight. *Proc. IEEEE ICCV, 545-552.* Dec 2013.

[2] Lange, R. et al. Solid-state time-of-flight range camera. *IEEE J. Quantum Elec.* Mar 2001.

[3] Matsuda, N. et al. MC3D: motion contrast 3D scanning. *IEEE Conf. Computational Protography (ICCP).* Apr 2015.

[4] Kim, S.J. et al. A 640x480 image sensor with unified pixel architecture for 2D/3D imaging in 0.11um CMOS. *IEEE Symp. VLSI Circuits.* June 2011.

[5] Kim, W. et al. A 1.5Mpixel RGBZ image sensor for simultaneous color and range image capture. *IEEE Solid-State Circuits Conference (ISSCC).* Feb 2012.

[6] O'Toole et al. Homogeneous codes for energy-efficient illumination and imaging. *Siggraph.* August 2015.

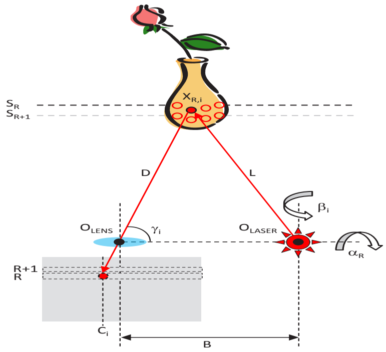

Fig.1 3D epipolar plane camera system with laser scan. When laser scans spot $X_{R,I}$, the sensor images that spot at known row R, column C_i. Its timestamp is recorded.

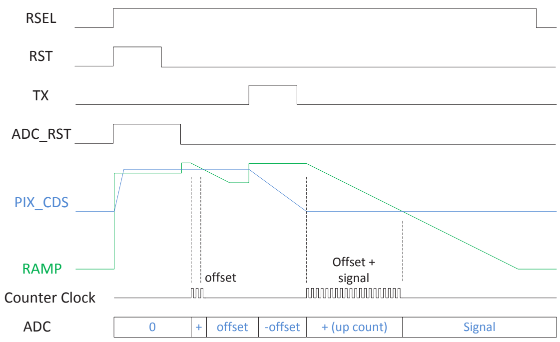

Fig. 2. Row readout timing for 2D mode.

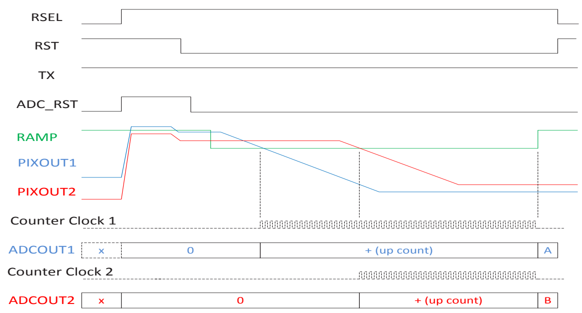

Fig. 3. Row timing for 3D linear mode.

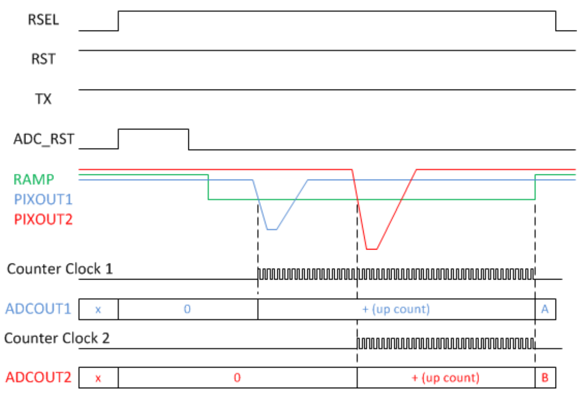

Fig. 4. Row timing for 3D logarithmic mode.

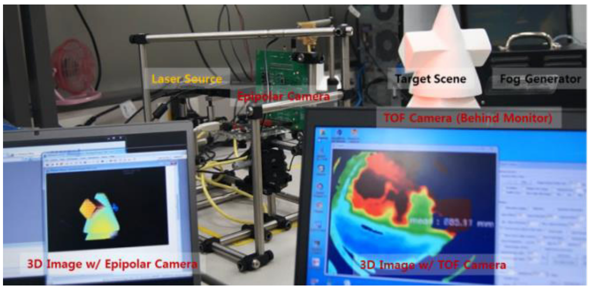

Fig. 5. Epipolar plane imaging test under foggy condition. The left camera uses red laser for scanning and epipolar plane imaging. The right camera uses diffused modulated NIR light and a TOF sensor with a fish-eye lens.

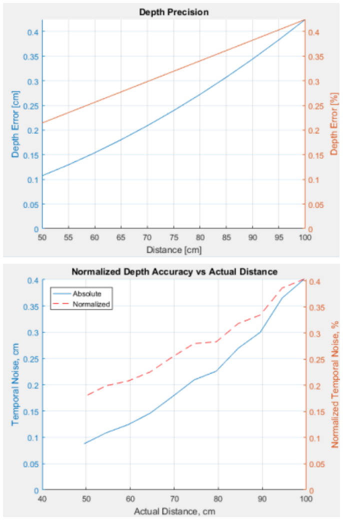

Fig. 6. Theoretical and measured depth precision.

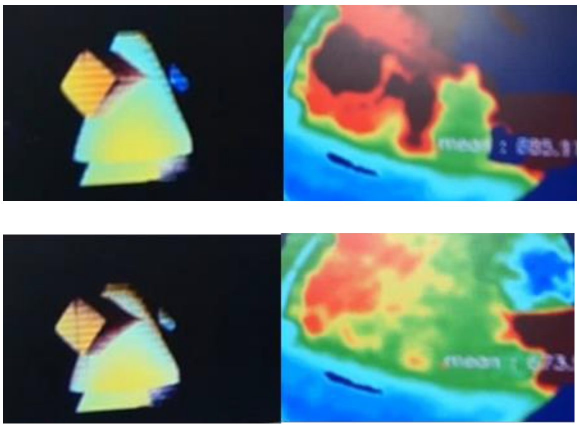

Fig. 7. 3D Images without and with fog. Top row: no fog. Bottom row: with fog. Left column: images from epipolar plane camera with red laser. Right column: images from non-epipolar TOF camera with NIR source. The bottom-left image is clear with a little depth error. The bottom-right image is entirely smeared.