Dual-Tap Computational Photography Image Sensor With Per-Pixel Pipelined Digital Memory for Intra-Frame Coded Multi-Exposure

Navid Sarhangnejad, *Member, IEEE*, Nikola Katic, *Member, IEEE*, Zhengfan Xia, Mian Wei, Nikita Gusev, Gairik Dutta, Rahul Gulve, *Student Member, IEEE*, Peter Zhi Xuan Li, Hui Feng Ke, Harel Haim, Manuel Moreno-García, *Member, IEEE*, David Stoppa, *Senior Member, IEEE*, Kiriakos N. Kutulakos, *Member, IEEE*, and Roman Genov, *Senior Member, IEEE*

***Abstract*— A coded-exposure-pixel image sensor for computational imaging applications is presented. Each frame exposure time is divided into *N* subframes. Within each subframe, each pixel sorts photo-generated charge into two charge taps depending on that pixel's 1-bit binary code. *N* global updates of arbitrary pixel-wise codes are implemented in each frame to enable *N* short global pixel-specific subexposures within one frame. To make these subexposures global, two latches per pixel are utilized in a pipelined fashion. The code for the next subframe is loaded into latch 1 in a row parallel fashion, while the code for the current subframe is being applied by latch 2 globally for photo-generated charge sorting during the current subexposure. A $280^H \times 176^V$ image sensor prototype with 11.2-μm pixel pitch has been fabricated in a 0.11-μm CMOS image sensor (CIS) technology. The image sensor has been demonstrated in two computational photography applications, each using only a single frame of a video: 1) computing both albedo (a measure of reflectivity) and 3-D depth maps by means of structured-light imaging and 2) computing surface normals (3-D orientations) map by means of photometric stereo imaging. These demonstrations experimentally validate some of the unique capabilities of this computational image sensor, such as accurate 3-D visual scene reconstruction using only one camera, while maintaining its native specifications: the full spatial resolution and the maximum frame rate.**

***Index Terms*— 3-D imaging, 3-D scene reconstruction, active imaging, active pixel sensor (APS), albedo adaptive imaging, charge bucket, CMOS image sensors (CISs), coded-exposure imaging, coded-exposure-pixel (CEP), compressive sensing, computational cameras, computational photography, computer-vision, demodulation contrast, depth maps, disparity, dual-tap pixel, floating diffusion, global shutter, high frame rate, photo-generated charge, photometric stereo imaging, pinned photodiode (PPD), pixel tap, programmable exposure, programmable light projection, reflectivity, single-frame computational imaging, spatial light modulator, structured-light imaging, surface normals, and tap contrast.**

I. Introduction

In THE past decade, computational photography advances have enabled new imaging capabilities and improved imaging performance in three different ways: by repurposing the existing off-the-shelf camera modules electronically [1]–[4], by adopting unconventional optics together with off-the-shelf cameras [5]–[7], and by developing camera modules with custom-designed CMOS image sensors (CISs) with specific applications in mind [8]–[14].

Custom image sensors have demonstrated the possibilities for new ways of computational photography. For example, time-delay integration and spatiotemporal filtering by in-pixel ADC implementation are demonstrated in [8], dual and quad-bucket pixel architectures in [9] have enabled new approaches to several computational imaging techniques, including high-dynamic-range imaging, and compressive sensing for high-speed imaging at 200 Mfps by a multi-aperture image sensor is shown in [10]. Among these developments, a more recent focus has been on the programmability, or coding, of the camera exposure at the individual-pixel level, as shown in Fig. 1 in comparison to a conventional global-shutter camera. We refer to such cameras as coded-exposure-pixel (CEP) cameras. Unlike conventional global-shutter pixels

Manuscript received April 25, 2019; revised July 2, 2019; accepted July 20, 2019. Date of publication September 18, 2019; date of current version October 23, 2019. This article was approved by Associate Editor Vivek De. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under RGPIN, RTI, and SGP programs, in part by DARPA under REVEAL Program, and in part by CMC Microsystems. *(Corresponding author: Navid Sarhangnejad.)*

N. Sarhangnejad was with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada. He is now with Huawei Technologies, Markham, ON L3R 5B4, Canada (e-mail: sarhangn@ece.utoronto.ca).

N. Katic was with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada. He is now with Synopsys, Mississauga, ON L5B 1M2, Canada.

Z. Xia, N. Gusev, R. Gulve, H. F. Ke, and R. Genov are with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: roman@eecg.utoronto.ca).

M. Wei, H. Haim, and K. N. Kutulakos are with the Department of Computer Science, University of Toronto, Toronto, ON M5S 3H5, Canada.

G. Dutta was with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada. He is now with Rambus, Toronto, ON M5J 2M4, Canada.

P. Z. X. Li was with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada. He is now with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

M. Moreno-García is with Fondazione Bruno Kessler (FBK), I-38123 Trento, Italy.

D. Stoppa was with Fondazione Bruno Kessler (FBK), I-38123 Trento, Italy. He is now with AMS AG, 8803 Rüschlikon, Switzerland.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSSC.2019.2932623

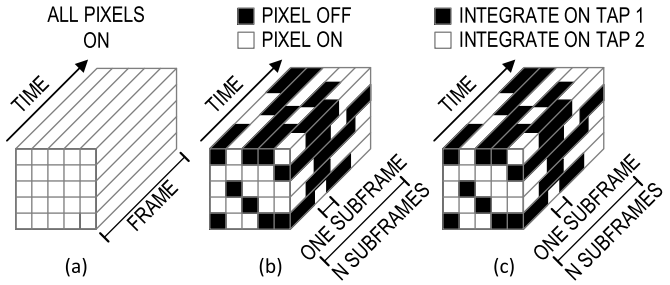


Fig. 1. Exposure principle of (a) conventional global-shutter image sensors where each pixel integrates all incident light, (b) single-tap CEP image sensors where incident light is selectively integrated based on a per-pixel 1-bit binary code, and (c) dual-tap CEP image sensors where incident light is sorted between two storage nodes known as taps.

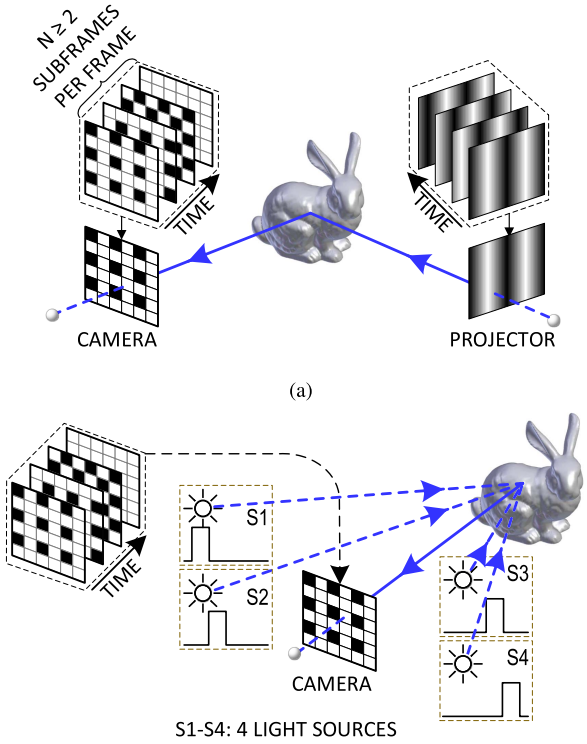


Fig. 2. Imaging principle for (a) single-frame structured-light imaging and (b) single-frame photometric stereo imaging [14].

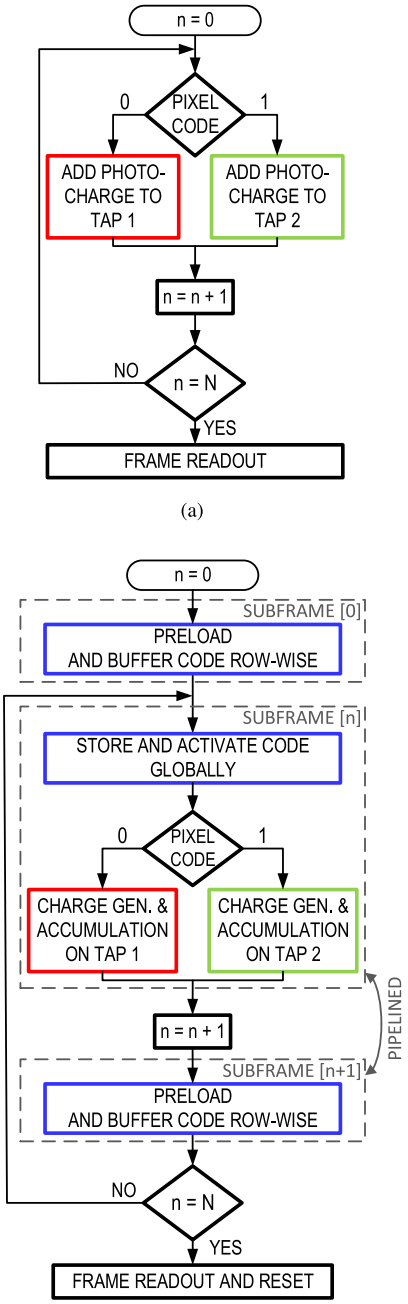


Fig. 3. Principle of operation of (a) generic dual-tap CEP and (b) presented dual-tap CMP [14].

which are all exposed for the same time interval (exposure time) and thus integrate all incident light in each video frame, the pixel in this emerging class of CEP cameras can be programmed to selectively sense only some of the light on a single storage node known as tap [11] [single-tap CEP imagers in Fig. 1(b)] or, better, sort all of the light between two taps [12], [13] [dual-tap CEP imagers in Fig. 1(c)], depending on the 1-bit binary pixel code. This is done N times per single frame, dividing the frame into N subexposures, which we refer to as subframes. In conjunction with a concurrently coded illumination, projected over the same N subframes, this enables a wide range of new coded multi-exposure single-readout-frame imaging capabilities at video rates. A visual scene can thus be interrogated N times within a single frame, with various combinations of projector pixel codes and camera pixel codes. This adds many degrees of freedom as to how a scene is illuminated and how the photo-generated charge is selectively integrated based on specific properties of photons arriving from the scene, for example based on the geometry of illumination or geometry of each photon's travel in space.

For applications such as 3-D sensing, gesture analysis, and robotic navigation, CEP cameras were initially implemented using bulky, distortive, slow, and expensive components [7] like digital micromirror devices (DMDs) together with off-the-shelf image sensors. In the last few years, the first prototypes of fully integrated CEP image sensors have emerged, including a one-tap imager for compressive sensing [11]. One-tap imagers lose light when the tap is off and cannot

sort photo-generated charge. Low-resolution (80×60 or less) two-tap light-sorting image sensors for primal–dual coded imaging [12] and motion-deblurring [13] have recently been reported. By using two taps and assuring one of the taps is always capturing light during frame exposure, all incident light is utilized regardless of applied codes. These imagers have large pixels ($> 12\ \mu$m) and low fill factors ($< 34\%$). None of them use a pinned photodiode (PPD)-based image sensor CMOS technology and thus suffer from low image quality and low tap contrast (defined as the transferred minus residual charge divided by their sum) and, as a result, are prone to degraded photo-generated charge sorting. Additionally, the code submission to the array in [11] and [13] is operated in a random access and row-wise manner, respectively. These architectures are not suitable where code update at each subframe has to be done globally. For example, active illumination in [7] is changing for the full field of view all at once and requires the camera to operate similarly, imposing a global code update at the start of the subframe in the CEP image sensor.

We present a next-generation dual-tap CEP image sensor, which is comprised of a $244^H \times 160^V$ main array of 11.2-μm pixels in a 0.11-μm CIS technology. The use of a CIS technology improves: 1) the dark current performance by isolating the Si–SiO_2 interface from the depletion region of the photodiode; 2) quantum efficiency by having deeper photodiode; and 3) tap-contrast due to complete charge transfer of the PPD device. The previously mentioned limitations of existing architectures are addressed by this sensor which features the combination of improved pixel performance and globally operated CEP functionality. The global operation is achieved by the pipelined dual-tap pixel architecture, which can apply any arbitrary codes to the full pixel array. Additionally, the design improves upon the state of the art by 1) implementing a two-tap pixel array at practical resolutions for demonstration; 2) a relatively small pixel pitch while maintaining a good fill-factor; and 3) achieving a high tap-contrast required for separation of charges from one subframe to another. In Section II, we begin with a brief overview of two multi-exposure single-frame computational photography applications and explain how these define the architecture of the pixel. In Section III, we describe the sensor architecture, which enables the operation of the presented CEP sensor. The system-level implementation is described in Section IV, and the experimental results and validation in the two computational photography applications are described in Sections V and VI, respectively.

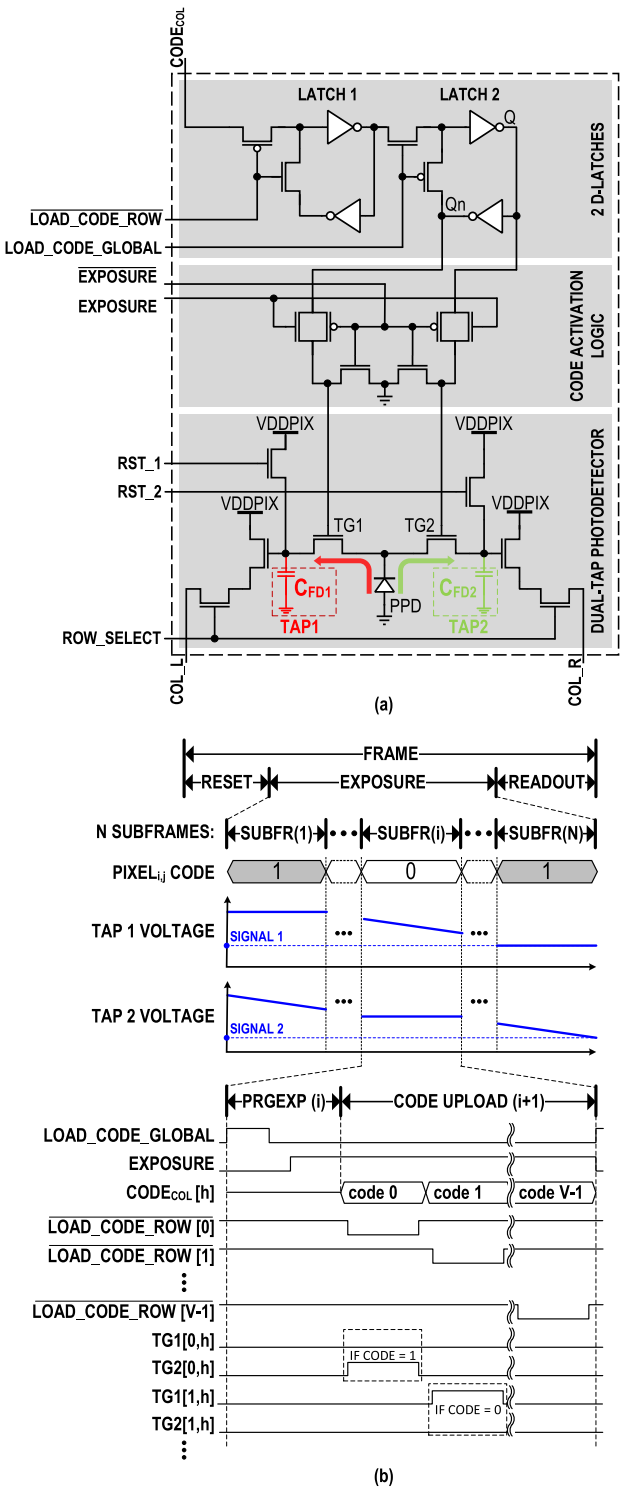


Fig. 4. (a) Pixel schematic of CMP pixel and (b) its simplified timing diagram. $H \times V$ are pixel array dimensions; h and v are the horizontal and vertical indices, respectively.

II. Coded-Exposure Pixel Architecture

A. Computational Photography Applications

The architecture of the CEP is designed to be generic, targeting many computational photography applications. Two such applications [15], as shown in Fig. 2, are introduced here with results demonstrated in Section VI. A novel aspect of the presented implementation of these imaging applications on our dual-tap CEP imager is that they are performed using only a single video frame at the sensor's native resolution.

1) Single-Frame Structured-Light Imaging: Fig. 2(a) shows the single-frame structured light imaging implementation. It is performed using the following steps.

1) Four spatial sinusoidal patterns separated by a 90° phase shift are cyclically projected onto the scene.

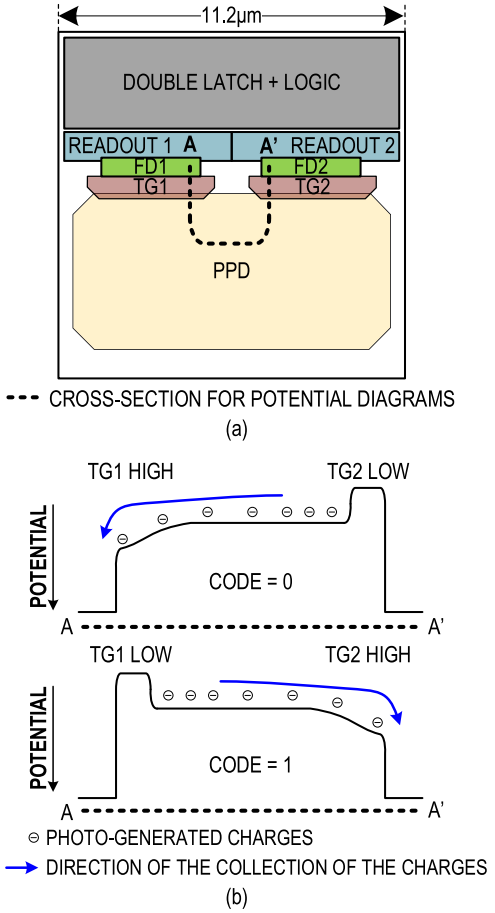


Fig. 5. (a) Floorplan of the pixel. (b) Electrostatic potential diagram for both code 0 and code 1 [14].

2) Synchronously with the projected patterns, four code matrices with time-multiplexed Bayer-like mosaic pattern, as shown, are cyclically submitted to the camera.
3) Next, the sorted photo-generated charges are accumulated across every four subframes and read out as two images per frame.
4) Off-chip image processing reconstructs all lighting conditions at full spatial resolution.
5) Finally, disparity or 3-D depth and albedo maps are computed. Disparity encodes the difference in horizontal coordinates of the point when viewed from both camera and projector viewpoints, which depends on the separation distance of the camera and projector. 3-D depth can be computed as inversely proportional to the disparity. Albedo is a measure of how much light is reflected off a surface without being absorbed.

2) Single-Frame Photometric-Stereo Imaging: Fig. 2(b) shows the single-frame photometric stereo implementation. It is implemented by the following steps.

1) Four LED light sources surrounding the camera are cyclically turned on and illuminate the scene, one at a time.
2) Synchronously with the LEDs, four code matrices with time-multiplexed Bayer-like mosaic pattern, as shown, are cyclically submitted to the camera.

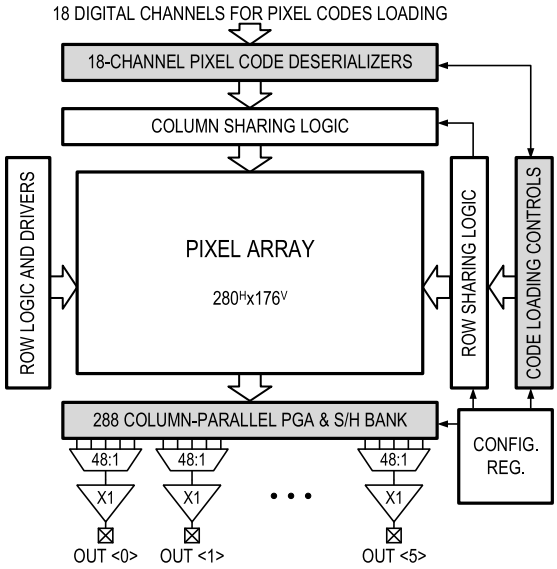


Fig. 6. VLSI architecture of the sensor. Blocks highlighted in gray are presented in more detail in Section III.

3) Next, the sorted photo-generated charges are accumulated across every four subframes and read out as two images per frame.
4) Off-chip image processing reconstructs all lighting conditions at full spatial resolution.
5) Finally, surface normals and albedo maps are computed. A surface normal is a vector perpendicular to the tangent plane to that surface at a given point. A map of normals is widely used in computer vision to encode 3-D information of a visual scene.

B. Pixel

Dual-tap and multi-tap pixels are commonly used for indirect time-of-flight (ToF) imaging applications to demodulate the received light and extract the phase information [16]–[18]. To enable CEP applications described in Section II-A, a dual-tap pixel with a general operating principle described in the flowchart in Fig. 3(a) is needed. The photo-generated charge is stored on taps 1 or 2 for codes 0 and 1, respectively, during each coded subframe, and the results are accumulated over N subframes within one video frame.

We introduce the code-memory pixel (CMP) architecture that operates following the aforementioned principle, as shown in Fig. 3(b). In order to perform global coded-subframe exposure, it requires an in-pixel dual digital code memory. The code memory is pipelined: a code value is preloaded row-wise into each pixel during the previous subframe and is applied at the beginning of the current subframe. Photo-generated charge is collected based on the current code while the next subframe's code is being preloaded into the pixel. The pipeline operation assures that the code update in the full array is done at once at the beginning of each subframe. This is suitable for applications with active illumination and does not require overhead time for code upload time.

Fig. 4(a) and (b) shows the CMP architecture and the relevant timing diagrams, respectively. The CMP includes two

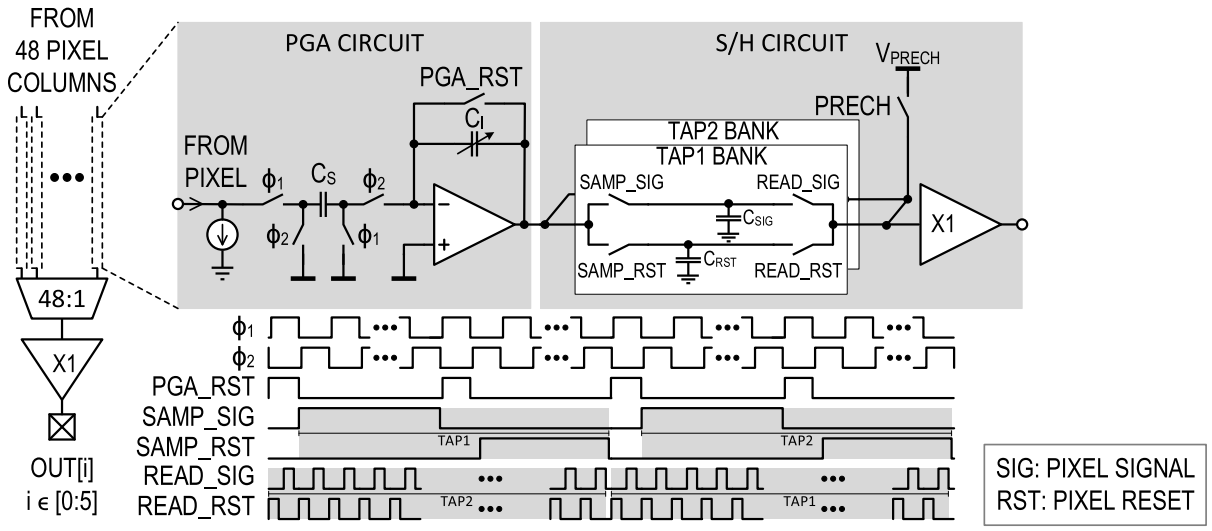


Fig. 7. Circuit and timing diagram of a 48-column slice of the column-parallel PGA and the S/H circuit in Fig. 6.

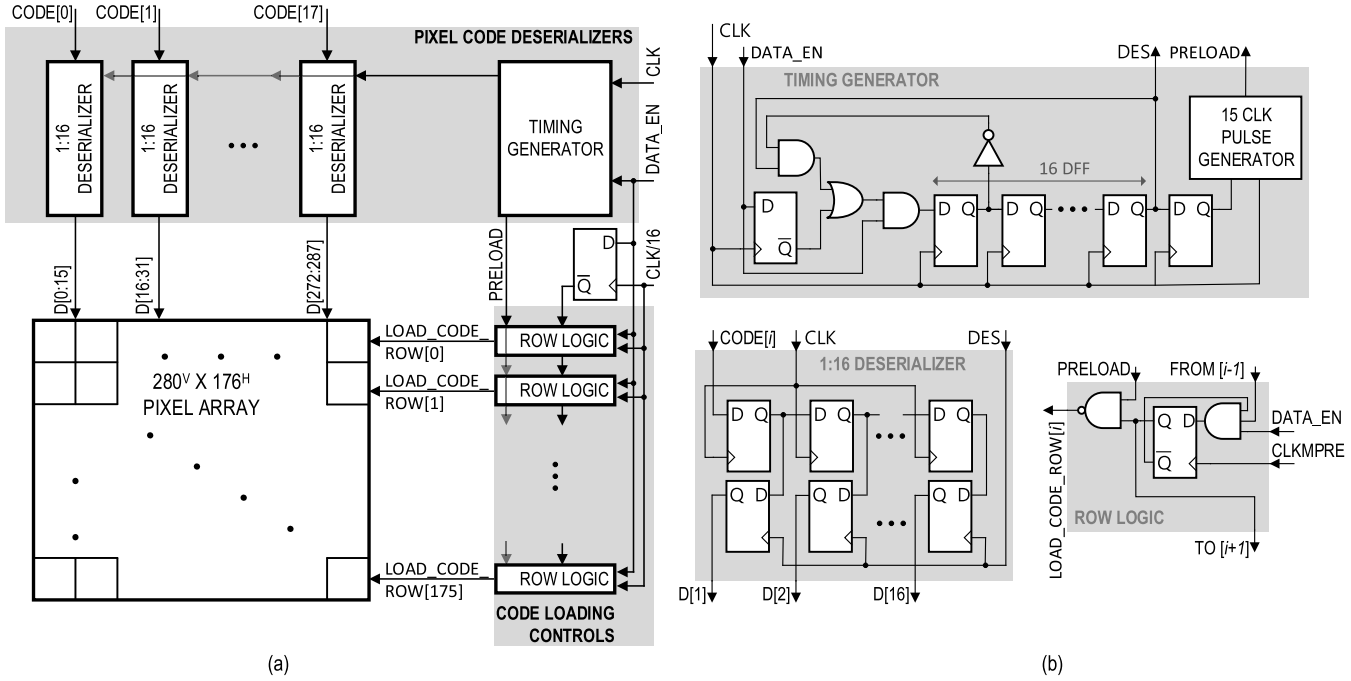


Fig. 8. (a) Block diagram of code deserializers and code loading circuit together with the pixel array. (b) Simplified circuitry of key internal blocks (other auxiliary circuitry for various configurations of the sensor are not shown here for simplicity).

D-latches: one controlled by LOAD_CODE_ROW—to pre-load the code patterns row-by-row—and the other controlled by LOAD_CODE_GLOBAL—to activate this pattern globally. Based on the code in each pixel, one of the two transfer gates, TG1 or TG2, connects the PPD to the corresponding tap, FD1 or FD2, respectively. The pattern in each pixel is gated with EXPOSURE signal to stop any charge transfer during the readout phase. EXPOSURE is also kept low during the global code updates to ensure that signal and supply glitches caused by digital switching in the array do not affect the tap contrast. The subframe phase, in addition to code upload period, includes a programmable period [PRGEXP in Fig. 4(b)] to increase the controllability of the exposure time. During the reset phase (not shown in detail), the EXPOSURE signal is toggled high and the reset signals of all pixels, RST_1 and RST_2, are asserted to reset both taps and the PPD through one of the taps. At the end of this phase, the EXPOSURE signal is first lowered and then the reset signals are set to the associated low voltage. During the readout phase (not shown in detail), the EXPOSURE is kept low while the pixels are accessed row-wise by ROW_SELECT for column-parallel readout. In Fig. 4(b), h and v refer to the column and row indices, respectively, for a generic pixel array size of $H \times V$.

Fig. 5 shows the pixel layout floor plan and its cross-sectional electrostatic potential diagrams. The PMOS transistors are placed at the maximum distance from the PPD devices to assure minimal interaction between the n-wells. A fill-factor of 45.3% is achieved with 27% of the area occupied by the latches and logic gates. As illustrated by the electrostatic potential in Fig. 5(b), the photo-generated charges

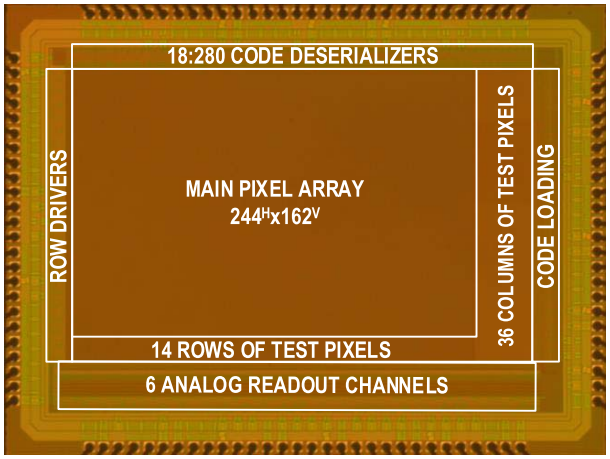


Fig. 9. Micrograph of a prototype fabricated in a CIS 110-nm process. The die size is 3 mm × 4 mm [14].

at any instance are sorted based on the transfer gate voltages determined by the 1-bit binary code stored in the pixel.

III. CEP Sensor Architecture

Fig. 6 depicts the 280 × 176-pixel sensor system architecture. Unlike conventional image sensors, during each subframe, the CEP sensor receives pixel codes to control all pixels' exposure individually. The sensor has 18 digital input channels for streaming the codes row by row, 18 code-deserializers with logic for arranging the codes, and vertically organized code loading control circuits to ship the codes to their respective rows. Additional column-sharing and row-sharing logic allows for faster streaming of repeated patterns to save power and time needed for shipping the same pattern from an off-chip DRAM. The column-parallel outputs are amplified by switched-capacitor programmable-gain amplifiers (PGAs), which can apply different gains to each of the two taps (with a gain range of 0.5× to 8×). The output is then serialized and buffered over six analog output channels with 48:1 multiplexing ratio.

The column-parallel PGAs and sample-and-hold (S/H) circuits are shown in Fig. 7. The PGA is a switched-capacitor integrator, where the integrating capacitor, C_I, is configurable to two different values for a gain of ×0.5 or ×1. By using an integrator, instead of a gain amplifier, multiple-sampling can be performed on the output of the pixels to amplify the signal level. This has the advantage of a programmable gain through timing adjustments (the number of integrations by ϕ_1 & ϕ_2 pulses, as shown in the timing diagram of Fig. 7) rather than serial peripheral interface (SPI) configurations for capacitor banks. The adjustable gain can be accommodated by first making $SAMP$_SIG high while ϕ_1 and ϕ_2 pulses sample the pixel signal data on C_S and integrate over C_I, respectively, and the output is sampled over the C_{SIG} capacitance of tap 1. Then, the pixel reset signal is sampled in a similar fashion over the C_{RST} capacitance. The sampled data of tap 1 are next multiplexed to the output by READ_{RST} and READ_{SIG} signals, while the tap 2 data are being processed by the PGA. The number of integrations of C_S charge over C_I determines the additional gain control from x1 up to x8.

There is one PGA per pixel column and the two taps are read out at different times. Two sampling capacitor banks

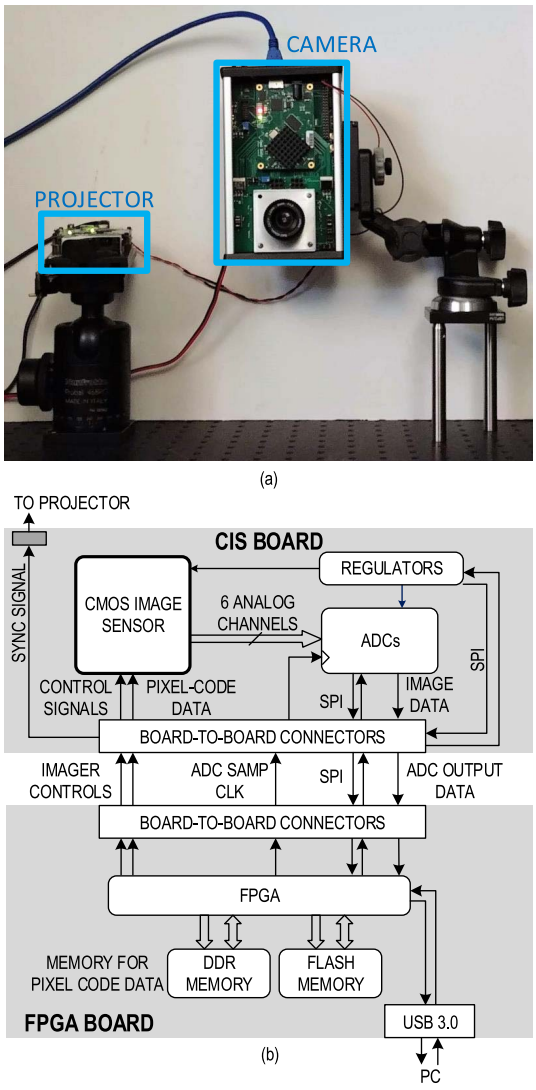
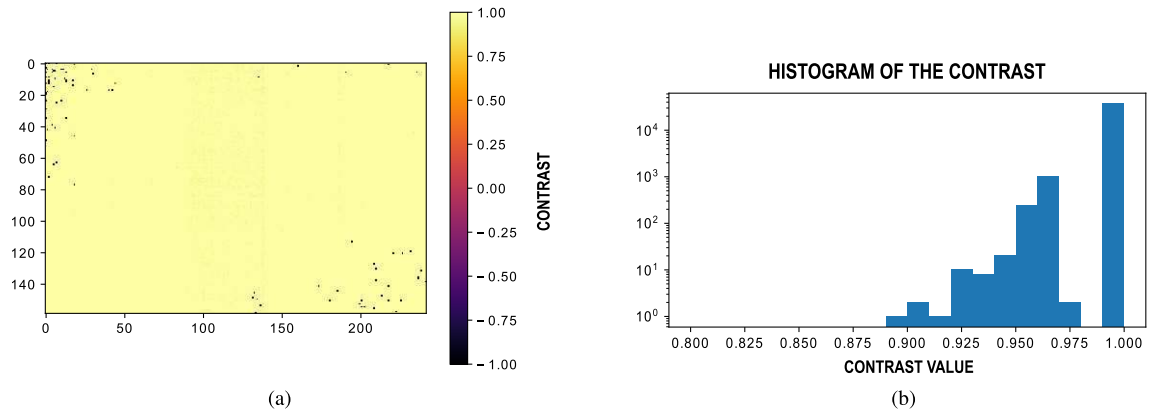


Fig. 10. (a) Imaging system experimental setup for single-frame structured-light imaging. (b) Camera module block diagram.

are implemented, one for each of the taps. When one tap is being sampled by the PGA, the previously sampled tap is buffered to the output I/O pads, as shown in the timing diagram of Fig. 7. Each bank has two capacitors: one for the reset level of the pixel (RST) and one for the signal level (SIG). Double sampling (DS) operation is performed by first reading the signal and then reading the reset level of the pixel. The subtraction of the two values is done on the off-chip ADC.

The simplified pixel code-deserializers and code-loading controls' circuits are shown in more detail in Fig. 8. The timing generator block ensures that the 18 deserializers register the received codes at every 16th clock cycle and sends it to the pixel array. Signal DATA_EN is asserted high with the arrival of the first code bit, and after 16 CLK cycles, sets DES for one CLK period. Signal DES registers the serialized data in the 1:16 deserializer block. One clock cycle later, when the pixel array columns settle to the correct code data, the PRELOAD signal enables the LOAD_CODE_ROW signal of the respective row in order to preload the codes to the first latches in the pixels of that row. The pulse generator block extends the PRELOAD signal width to 15 clock cycles to ensure that the latches have the maximum time for registering



(a) (b)

Fig. 11. (a) Experimentally measured tap contrast map of the sensor (the on-line version is in color). (b) Histogram of the tap contrast between 0.8 and 1 values (note that the vertical axis is shown on a logarithmic scale).

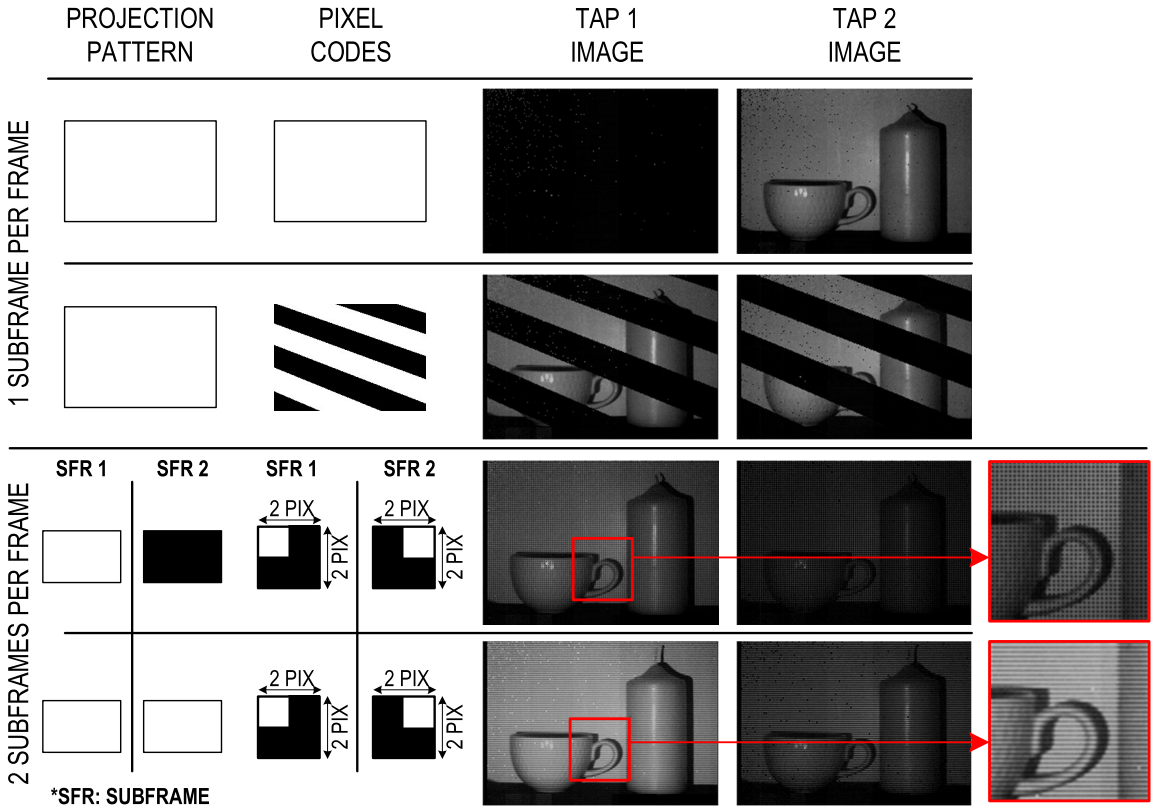


Fig. 12. Experimentally captured raw output images of the sensor for two uniform-lighting cases: for one (top) and two (bottom) subframes per frame [14].

the data. During these 16 CLK cycles, the code for the next row is deserilized. This is repeated row by row. After all the rows are preloaded with the code data, a global signal triggers the transfer of the codes to the second latch of pixels.

IV. System Implementation

Based on the pixel concept described in Section II and the sensor architecture in Section III, we have designed and fabricated an image sensor in a 110-nm CIS technology. 1.2-V supply is used for the digital peripheral circuits and 3.3 V for the analog readout block. The micrograph of the image sensor is shown in Fig. 9.

The camera module and the projector used for structured-light imaging (for the demonstration in Section VI-A) are shown in Fig. 10. The camera consists of two boards. One is a commercially available board carrying a DDR memory for the pixel code data and FPGA for controlling the system operation. The second board is a custom board including the designed image sensor and off-the-shelf ADC ICs for converting the analog image data of the sensor to the digital domain. At the start-up, the FPGA sets the supply and reference levels through SPI. Then, the PC uploads the pixel codes to the DDR through the FPGA and also uploads the projection patterns to the projector DDR. When the video streaming command is initiated by the PC, the FPGA ships the codes from the DDR to the sensor, controls the image sensor and ADC to receive the digitized data, and finally streams the data to the PC. The FPGA also controls the projector and makes sure that

TABLE I
COMPARISON TABLE

		THIS WORK	TCAS-I 2018 [13]	IISW 2017 [12]	OE 2016 [11]	ISSCC 2015 [10]	JSSC 2012 [9]	ISSCC 2015 [19]
	CODED-EXPOSURE	PER-PIXEL (*i.e.* PIXELWISE SPATIAL AND TEMPORAL CODING)				PER 1/15 OF ARRAY	PER FULL ARRAY (*i.e* TEMPORAL CODE ONLY)	
PIXEL	TECHNOLOGY [nm]	**110 CIS**	130 CMOS	350 CMOS	180 CIS	**110 CIS**	130 CIS	**110 CIS**
PIXEL	PINNED PHOTODIODE	**YES**	NO (NW/P)	NO (PG)	**YES**	**YES**	**YES**	**YES**
PIXEL	PIXEL PITCH [µm]	**11.2**	12.1x12.2	25	10	11.2X5.6	**5**	11.2X5.6
PIXEL	FILL FACTOR [%]	45.3	33.2	20.5	**52**	--	42	--
PIXEL	NUMBER OF TAPS	**2**	**2**	**2**	1	1	**2**	**2**
PIXEL	TAP CONTRAST	**0.99 @ 181sfps** [1]	--	LOW	N/A	N/A	--	0.94 [2]
ARCHITECTURE	PIXEL COUNT [HxV]	244 x 162	10 x 10	80 x 60	127 x 90	64 x 108	**640 x 576**	256 x 512
ARCHITECTURE	FRAME RATE [fps]	25	60	25	**100**	32	N/A	12
ARCHITECTURE	READ NOISE	**3.6mV 5.3DN (12-bit ADC)**	5.2mV	--	5.4DN	--	5.5e$^-$	1.75e$^-$
ARCHITECTURE	DYNAMIC RANGE [dB]	50.5	**52**	--	51.2	--	85 / 103 (HDR mode)	--
ARCHITECTURE	CONVERSION GAIN [µV/e$^-$]	33.5	**--**	--	--	--	51	85
ARCHITECTURE	POWER [mW]	34.4	**1.23**	7	1.3	1620	N/A	540
ARCHITECTURE	POWER FoM [nJ] [3]	34	205	58	**1.14** [4]	7324 [4]	N/A	343 [4]
SYSTEM	CODE MEMORY	IN-PIXEL (2 LATCHES)	IN-PIXEL (DRAM)	IN-PIXEL (2 LATCHES)	IN-PIXEL (SRAM)	OFF-PIXEL	N/A	N/A
SYSTEM	SUBFRAME RATE [sfps [1]]	181.8	**300000**	--	100	N/A		
SYSTEM	PIXEL-CODE RATE [MHz]	7.2	**30**	--	0.11			
SYSTEM	ARBITRARY CODE / ROI	**YES/YES**	YES/--	YES/--	NO/--			
SYSTEM	SUBFRAME SHUTTER	**GLOBAL**	ROLLING	**GLOBAL**	ROLLING			
SYSTEM	IMAGING APPLICATIONS	1.Single-frame structured-light, 2.Single-frame photometric stereo, 3.Other: compressive sensing, etc.	Deblurring	Transport-aware	Spatiotemporal compressive sensing	Ultra high speed w/ compressive sampling	High-dynamic-range	Fluorescence lifetime

1: subframe per second
2: also known as extinction ratio
3: $FoM = Power/\big((Number\ of\ pixels) \times (Frame\ rate)\big)$
4: FoM includes the on-chip ADC power

N/A: Not applicable
--: Not available
Bold font denotes the best performance

the projections of the patterns are synchronized with the codes sent to the sensor.

In the following, first the electrical testing and characterization of the sensor is described (Section V) and then the deployment of the camera in two single-shot computational imaging applications are shown (Section VI).

V. EXPERIMENTAL CHARACTERIZATION

The setup shown in Fig. 10(a) is used for the experimental characterization of the sensor. The most important specification of the camera is the tap contrast, which determines how much of the photo-generated charge in each pixel is directed to the tap corresponding to the code shipped to the pixel. As mentioned in Section I, tap contrast is defined as follows:

$$\chi = \frac{Q_1 - Q_2}{Q_1 + Q_2}$$

where Q_1 and Q_2 are the amounts of charge stored on taps 1 and 2 in the pixel, respectively. For measuring Q_1 and Q_2 in this experiment, alternating codes 1 and 0 are sent to all pixels of the sensor and the light is projected only for code value 0. This means that tap 1 should collect all the photo-generated charges and tap 2 should not collect any charges. In this scenario, the projection at codes 0 and 1 results to tap contrasts of 1 and -1 for an ideal pixel. The tap contrast map of the sensor is shown in Fig. 11(a). The measurement is done at four subframes per frame, at 25-frames/s frame rate. Considering the overhead times for the readout and data transfer to the PC, the effective subframe rate is about 181.8 Hz. The average tap contrast achieved is 0.99, and the histogram of the tap contrast between 0.8 and 1 is shown in Fig. 11(b). A small number of pixels in the two corners of Fig. 11(a), with tap contrast values of around -1, are the pixels that do not register the correct pixel code and collect the charges on the wrong tap (due to resolvable technology limitations).

Fig. 12 shows raw outputs of the sensor for four different camera configurations for the same scene under uniform lighting conditions. The first two rows show images where only one subframe is recorded per frame. The image is collected on taps 1 and 2 for black and white codes, respectively. In the third row, there are two subframes, with tiled 2×2-pixel codes, as shown, sent to the pixel array. The projector projects all-white during the first subframe and all-black during the second one. The magnified inset for tap 1 shows that the

image is black for 1 out of 4 pixels. The last row depicts the same conditions, but the projector projects all-white for both subframes. The magnified inset shows that the alternating rows are brighter because they have collected light for two subframes, while the darker rows have collected light for one subframe only.

A comparison to the state of the art is given in Table I, where CEP cameras are highlighted by light gray shading. The presented image sensor has the highest spatial resolution among the CEP architectures and additionally, it is the first dual-tap CEP using PPD pixels. A competitive pitch of 11.2 μm and a 45.3% fill factor are achieved with the proposed architecture. A very high tap contrast of 0.99 at 181.8 subframe per second rate is measured. The achieved power FoM is the second best to [11], which is implemented for low-power compressive sensing applications by incorporating pipelined readout and exposure with low-power successive approximation ADC (SAR) architectures in a single-tap sensor. The subframe rate and code rate at pixel level are the second best to [13] that has a significantly smaller pixel array size.

Comparison to non-CEP image sensors (not highlighted in gray shading) suitable for computational imaging is performed in the table as well. Ultra high-speed imaging is done by multi-aperture camera in [10], where temporal coding at the sub-array level is used. Dual-bucket pixel in [9] is designed for full-array level coded exposure applications, not for per-pixel-exposure coding, for example for high-dynamic-range imaging. Fluorescence lifetime imaging by a very high extinction ratio of 0.94 is proposed by dual-tap globally programmed pixels in [19].

VI. Experimental Demonstration

In this section, two computer vision applications as briefly explained in Section II-A, and in more detail in [15], are demonstrated using the proposed CEP image sensor.

A. Single-Frame Structured Light Imaging Results

Structured-light imaging is used for 3-D imaging by projecting patterns to the scene and analyzing them from a different view point than that of the projection. The depth and surface information can be calculated by finding the correspondence between the sent patterns and the received ones. This technique conventionally requires multiple frames to reconstruct the disparity (the inverse depth) and albedo map of the scene. We show that by using dual-tap CEP camera, one can obtain the disparity (and consequently depth) and albedo maps at full pixel array resolution in a single frame, as described in Section II-A1.

Fig. 13(b) shows the albedo and depth map reconstruction pipeline for the single-frame structured light 3-D imaging setup, where four subframes are used. As discussed before for Fig. 2(a), at the end of a frame, each of the two tap images captures the visual scene sampled four times and coded by a four-pixel time-multiplexed Bayer-like mosaic pattern, corresponding to four 90° shifted sinusoidal illumination patterns. The Bayer-coded raw images from taps 1 and 2 are separated out into four images for each tap, each containing different structured light sine phase information. These images are then upsampled and processed to obtain four full-resolution images from which the resulting depth (or, alternatively, disparity) and albedo maps are computed.

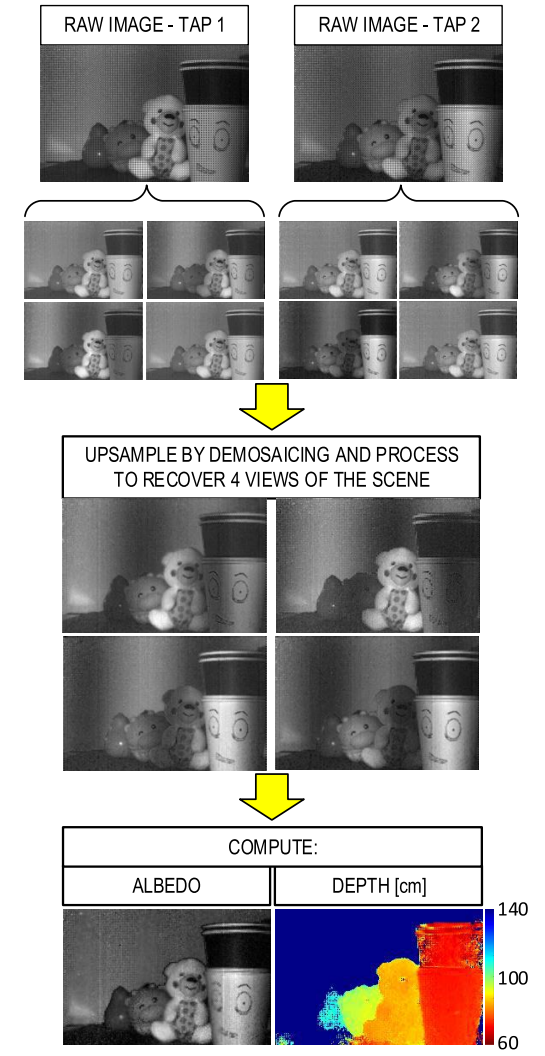


Fig. 13. Experimentally measured single-frame structured-light imaging results. (a) Albedo and depth map reconstruction pipeline [14]. (b) Albedo and disparity (i.e., inverse depth) maps from other scenes for both static (left) and dynamic (right) scenes, all generated at 20 frames/s.

B. Single-Frame Photometric Stereo Imaging Results

Photometric stereo is an imaging technique for determining the scene surface 3-D orientation at each image point [20]. This technique conventionally requires capturing multiple

Fig. 14. Experimentally measured single-frame photometric stereo imaging results: surface normals maps and albedo maps generated at 20 frames/s, for both static (left) and dynamic (right) scenes [14].

images or deploying an image sensor with multiple taps (usually three or more taps [21]). We demonstrated that four images with different directions of illumination light can be acquired by a dual-tap CEP sensor within a single frame as described in Section II-A2.

Fig. 14 shows the results from a single-frame photometric stereo experiment. Fig. 14 (top) depicts the surface normal maps of two different scenes captured at 20 frames/s. The images are color-coded with the orientation of the surface normal at each pixel. Fig. 14 (bottom) shows the albedo maps containing information on the reflectivity of the same scenes.

VII. Discussion

The proposed dual-tap-pixel camera has been demonstrated in two computational photography applications, and the depth and surface normals maps have been successfully obtained using only a single frame. Alternative solutions for such applications using conventional image sensors require multiple frame acquisitions [22] or, in some cases, more than two taps per pixel [21]. Multiple-frame readout suffers from motion blur, consumes more power, and typically suffers from higher read noise. The noise of multiple readouts is higher if it is dominated by the readout circuitry noise (pixel reset noise, source follower noise, column-parallel circuitry noise, and so on) rather than the photon shot noise. Existing multi-tap pixels with more than two taps do not offer the versatility, flexibility, and universality of a CEP camera, which are some of the advantages of the presented design.

VIII. Conclusion

A method of camera re-programmability at the pixel level, multiple times during a single frame exposure, is presented. This enables many new imaging applications in computational photography including applications that previously required bulky and distortive optics used together with off-the-shelf cameras. The presented CEP camera has been demonstrated in two such applications, but using only a single image sensor. These 3-D imaging techniques are: structured-light imaging and photometric stereo imaging, both implemented within just a single frame. Compared to an equivalent implementation using a high-frame-rate camera operating at the same frame rate as the subframe rate in CEP cameras, the presented camera does not suffer from added read noise in each subframe, offers real-time operation without constraints of prohibitively high output data rate, and does not require excessive output buffer memory. Much lower power and complexity are also important advantages.

Author Contributions: K. N. Kutulakos and R. Genov conceived, devised, and oversaw this project and provided close direction for the entire team. N. Sarhangnejad and N. Katic designed the image sensor integrated circuit including the pixels, and P. Z. X. Li helped with peripheral blocks. M. Moreno-García and D. Stoppa contributed to the pixel design. N. Sarhangnejad and Z. Xia developed the camera module and its interface with help of H. F. Ke in the FPGA test setup design. M. Wei and H. Haim deployed the camera in single-frame computational imaging applications. N. Gusev, G. Dutta, and R. Gulve contributed to the PCB design as well as to test the camera. All the authors conceived various elements of the experiments, analyzed the results, and reviewed the manuscript. N. Sarhangnejad, N. Katic, K. N. Kutulakos, and R. Genov wrote and edited the manuscript.

References

[1] F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich, "Low-budget transient imaging using photonic mixer devices," *ACM Trans. Graph.*, vol. 32, no. 4, p. 45, Jul. 2013.

[2] A. Kadambi, A. Bhandari, R. Whyte, A. Dorrington, and R. Raskar, "Demultiplexing illumination via low cost sensing and nanosecond coding," in *Proc. IEEE Int. Conf. Comput. Photographs (ICCP)*, Jan. 2014, pp. 1–10.

[3] M. O'Toole, S. Achar, S. G. Narasimhan, and K. N. Kutulakos, "Homogeneous codes for energy-efficient illumination and imaging," *ACM Trans. Graph.*, vol. 34, no. 4, p. 35, 2015. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2809654.2766897

[4] C. Callenberg, F. Heide, G. Wetzstein, and M. B. Hullin, "Snapshot difference imaging using correlation time-of-flight sensors," *ACM Trans. Graph.*, vol. 36, no. 6, p. 220, Nov. 2017.

[5] Y. Hitomi, J. Gu, M. Gupta, T. Mitsunaga, and S. K. Nayar, "Video from a single coded exposure photograph using a learned over-complete dictionary," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 287–294.

[6] D. Liu, J. Gu, Y. Hitomi, M. Gupta, T. Mitsunaga, and S. K. Nayar, "Efficient space-time sampling with pixel-wise coded exposure for high-speed imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 248–260, Feb. 2014.

[7] M. O'Toole, R. Raskar, and K. N. Kutulakos, "Primal-dual coding to probe light transport," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 39, Jul. 2012. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2185520.2185535

[8] B. Tyrrell *et al.*, "Time delay integration and in-pixel spatiotemporal filtering using a nanoscale digital CMOS focal plane readout," *IEEE Trans. Electron Devices*, vol. 56, no. 11, pp. 2516–2523, Nov. 2009.

[9] G. Wan, X. Li, G. Agranov, M. Levoy, and M. Horowitz, "CMOS image sensors with multi-bucket pixels for computational photography," *IEEE J. Solid-State Circuits*, vol. 47, no. 4, pp. 1031–1042, Apr. 2012.

[10] F. Mochizuki *et al.*, "6.4 Single-shot 200 Mfps 5×3-aperture compressive CMOS imager," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, 2015, pp. 1–3.

[11] J. Zhang, T. Xiong, T. Tran, S. Chin, and R. Etienne-Cummings, "Compact all-CMOS spatiotemporal compressive sensing video camera with pixel-wise coded exposure," *Opt. Express*, vol. 24, no. 8, pp. 9013–9024, 2016.

[12] N. Sarhangnejad, H. Lee, N. Katic, M. O'Toole, K. Kutulakos, and R. Genov, "CMOS image sensor architecture for primal-dual coding," in *Proc. Int. Image Sensor Workshop*, 2017, pp. 356–359.

[13] Y. Luo, D. Ho, and S. Mirabbasi, "Exposure-programmable CMOS pixel with selective charge storage and code memory for computational imaging," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 65, no. 5, pp. 1555–1566, May 2018.

[14] N. Sarhangnejad *et al.*, "Dual-tap pipelined-code-memory coded-exposure-pixel CMOS image sensor for multi-exposure single-frame computational imaging," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2019, pp. 102–104.

[15] M. Wei *et al.*, "Coded two-bucket cameras for computer vision," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 55–73.

[16] B. Buttgen, F. Lustenberger, and P. Seitz, "Demodulation pixel based on static drift fields," *IEEE Trans. Electron Devices*, vol. 53, no. 11, pp. 2741–2747, Nov. 2006.
[17] D. Stoppa, N. Massari, L. Pancheri, M. Malfatti, M. Perenzoni, and L. Gonzo, "An 80 × 60 range image sensor based on 10 μm 50 MHz lock-in pixels in 0.18 μm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2010, pp. 406–407.
[18] B. Rodrigues *et al.*, "Indirect ToF pixel integrating fast buried-channel transfer gates and gradual epitaxy and enabling CDS," in *Proc. IISW*, 2017, pp. 266–269.
[19] M.-W. Seo *et al.*, "11.2 A 10.8 ps-time-resolution 256×512 image sensor with 2-tap true-CDS lock-in pixels for fluorescence lifetime imaging," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2015, pp. 1–3.
[20] R. J. Woodham, "Photometric method for determining surface orientation from multiple images," *Opt. Eng.*, vol. 19, no. 1, 1980, Art. no. 191139.
[21] T. Yoda, H. Nagahara, R.-I. Taniguchi, K. Kagawa, K. Yasutomi, and S. Kawahito, "The dynamic photometric stereo method using a multi-tap CMOS image sensor," *Sensors*, vol. 18, no. 3, p. 786, 2018.
[22] Y. Quéau, R. Mecca, J.-D. Durou, and X. Descombes, "Photometric stereo with only two images: A theoretical study and numerical resolution," *Image Vis. Comput.*, vol. 57, pp. 175–191, Jan. 2017.

Navid Sarhangnejad (S'16–M'18) received the B.Sc. degree in electrical and computer engineering from the University of Tehran, Tehran, Iran, in 2008, and the M.S. degree in electrical and computer engineering from the Delft University of Technology, Delft, The Netherlands, in 2010. He is currently pursuing the Ph.D. degree in electrical and computer engineering with the University of Toronto, Toronto, ON, Canada.

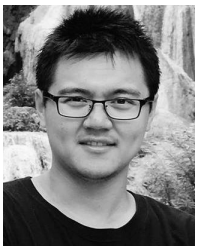
From 2011 to 2014, he was an Analog and Mixed-Signal Design Engineer with CMOSIS, Antwerp, Belgium. During this time, he worked on readout and peripheral circuits for CMOS image sensors. In summer 2016, he was a Visiting Researcher with the Integrated Radiation and Image Sensors (IRIS) Research Unit, Fondazione Bruno Kessler (FBK), Trento, Italy, where he was involved in pixel simulation and design. Recently, he joined Huawei Technologies, Toronto, ON, Canada, as an Analog Design Engineer. His research interests include analog and mixed-signal integrated circuits, over-sampled ADCs, CMOS image sensors, and high-speed electronics.

Nikola Katic (S'10–M'14) received the B.Sc. degree in electrical engineering from the School of Electrical Engineering, University of Belgrade, Belgrade, Serbia, in 2008, and the M.Sc. and Ph.D. degrees in electrical and electronic engineering from the Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland, in 2010 and 2014, respectively.

From 2014 to 2016, he was a Senior Research and a Development Engineer with DRAM Design Team, Samsung Electronics, where he was involved in the first-generation of HBM2 DRAM memory chip. He joined the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, in 2016, where he was a Postdoctoral Fellow until 2017. Since 2017, he has been a Senior Research and Development Engineer with Synopsys Inc., Mississauga, ON, Canada, where he was involved in the high-speed IP circuit design. His research interests include CMOS image sensor design, high-speed analog and mixed-signal integrated circuit design, and signal and image processing.

Zhengfan Xia received the B.E. degree in electronic and information engineering from the China University of Geosciences, Beijing, China, in 2008, and the M.S. and Ph.D. degrees in information sciences from Tohoku University, Sendai, Japan, in 2011 and 2014, respectively.

From 2014 to 2017, he was a Research Engineer with the Toshiba Research and Development Center, Kawasaki, Japan, where he was involved in hardware and embedded system security. In 2017, he joined the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, where he is currently a Postdoctoral Fellow. His research interests include computational camera system design, asynchronous circuits, computer architecture, and security.



Mian Wei received the B.Sc. degree in computer science and mathematics and the M.S. degree in computer science from the University of Toronto, Toronto, ON, Canada, in 2015 and 2017, respectively, where he is currently pursuing the Ph.D. degree in computer science.

Nikita Gusev received the four-year B.A.Sc. degree in electrical and computer engineering with a minor in robotics and mechatronics from the University of Toronto, Toronto, ON, Canada, in 2017, where he is currently pursuing the M.A.Sc. degree with the Intelligent Sensory Microsystems Laboratory.

His research interests are ASIC design and system level design of computational camera systems.

Gairik Dutta received the B.Tech. degree in instrumentation engineering from IIT Kharagpur, Kharagpur, India, in 2016, and the M.A.Sc. degree in electrical and computer engineering from the University of Toronto, Toronto, ON, Canada, in 2019.

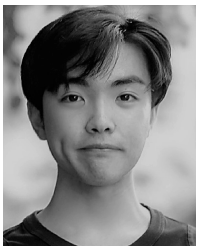
He is currently an Analog Design Engineer with Rambus, Toronto. His research interests include analog and mixed-signal integrated circuits, data converters, and high-speed digital communication circuits.

Rahul Gulve (S'18) received the B.Tech. and M.Tech. degrees in electrical engineering from IIT Madras, Chennai, India, in 2017. He is currently pursuing the Ph.D. degree in electrical and computer engineering with the University of Toronto, Toronto, ON, Canada.

His research interests include design and development of mixed signal systems and pixel architecture in transport aware 3-D image sensors and cameras for computational photography.

Peter Zhi Xuan Li received the B.A.Sc. degree in engineering science from the University of Toronto, Toronto, ON, Canada, in 2018. He is currently pursuing the Ph.D. degree in electrical engineering and computer science with the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA.

From 2013 to 2018, he was a Research Assistant with the University of Toronto, where he was involved in low-power analog IC design for a variety of biomedical applications. From 2016 to 2017, he was with the High-Speed Converters Group, Analog Devices, Toronto, as an Integrated Circuit Engineer. He is currently a Graduate Researcher with MIT and focuses on the co-design of algorithms and specialized hardware for localization, mapping, and path-planning on energy-efficient miniature robots.



Hui Feng Ke is currently pursuing the five-year B.Sc. degree in electrical and computer engineering with the University of Toronto, Toronto, ON, Canada.

From 2018 to 2019, he was a Design and Verification Intern with AMD, Markham, ON, Canada. His research interests include CMOS image sensors, image processing, and digital system design.

Harel Haim received the B.Sc. degree in electrical engineering from the Ben-Gurion University of the Negev, Beersheba, Israel, in 2010, and the M.Sc. and Ph.D. degrees within the framework of the Direct Program for Outstanding Ph.D. Candidates from the School of Electrical Engineering, Tel-Aviv University, Tel Aviv-Yafo, Israel, in 2013 and 2017, respectively, under the supervision of Prof. E. Marom and Prof. A. Bronstein.

He is currently a Postdoctoral Fellow with the Department of Computer Science, University of Toronto, Toronto, ON, Canada, with Prof. K. Kutulakos. His research interests are computational photography, time-of-flight and occluded imaging, compressive imaging, computer vision, and machine learning.

Dr. Haim was a recipient of the Weinstein Research Institute Award for published paper in 2013 and 2016 and the Electro-Optic Fund Award for outstanding research work in 2015. From 2015 to 2017, he served as a Principal Investigator in a government-funded project ("KAMIN" program), in pursuit of creating a small-scale computational camera for extended depth-of-field and depth estimation applications.

Manuel Moreno-García (M'16) received the five-year B.Sc. degree from the Universidad de Sevilla, Spain, in 2009, the M.Sc. degree in nuclear physics and microelectronics from the University of Münster, Germany, and the Universidad de Seville, and 2011, and the Ph.D. degree from the Instituto de Microelectrónica de Sevilla, Seville, Spain, in 2015.

Since 2015, he has been with the Integrated Radiation and Image Sensors (IRIS) Research Unit, Fondazione Bruno Kessler (FBK), Trento, Italy. His research interests are PPD and SPAD-based sensors, the simulation, design, and characterization of semiconductor photon detectors for ToF imaging, and integrated circuits design.

Dr. Moreno-García was a recipient of the First Prize in the National Award for Excellence in Academic Performance during his B.Sc. degree.

David Stoppa (M'97–SM'12) received the Laurea degree in electronics engineering from the Politecnico di Milano, Italy, in 1998, and the Ph.D. degree in microelectronics from the University of Trento, Italy, in 2002.

From 2002 to 2012, he was teaching courses on analog electronics and microelectronics with the Telecommunications Engineering Faculty, University of Trento. He was a Group Leader of the Smart Optical Sensors and Interfaces Group from 2010 to 2013. From 2014 to 2017, he was the Head of the Integrated Radiation and Image Sensors Research Unit, Fondazione Bruno Kessler, Trento, Italy, where he has been a Research Scientist since 2002. In 2017, he joined AMS, Rüschlikon, Switzerland, where he is currently in charge of the research and development of next-generation range sensors. He has authored or coauthored over 120 papers in international journals and presentations at international conferences, and holds several patents in the field of image sensors. His research interests are mainly in the field of CMOS integrated circuits' design, image sensors, and biosensors.

Dr. Stoppa was a recipient of the 2006 European Solid-State Circuits Conference Best Paper Award. Since 2011, he has been serving as a Program Committee Member for the International Solid-State Circuits Conference and the SPIE Videometrics, Range Imaging, and Applications Conference, and was a Technical Committee Member of the International Image Sensors Workshop in 2009, 2013, 2015, and 2017. He was a Guest Editor of the IEEE JOURNAL OF SOLID-STATE CIRCUITS special issues on ISSCC'14 in 2015. He has been serving as an Associate Editor since 2017.

Kiriakos N. Kutulakos (M'90) received the B.A. degree in computer science from the University of Crete, Greece, in 1988, and the M.S. and Ph.D. degrees in computer science from the University of Wisconsin Madison, Madison, WI, USA, in 1990 and 1994, respectively.

Following his dissertation work, he joined the University of Rochester, Rochester, NY, USA, where he was an NSF Postdoctoral Fellow and later an Assistant Professor until 2001. He is currently a Professor of computer science with the University of Toronto, Toronto, ON, Canada.

Dr. Kutulakos was a recipient of the Best Student Paper Award at the IEEE Conference on Computer Vision and Pattern Recognition in 1994 and 2017, the David Marr Prize in 1999, the David Marr Prize Honorable Mention in 2005, the Best Paper Honorable Mention at the European Conference on Computer Vision in 2006, the CAREER Award from the U.S. National Science Foundation (NSF), the Premiers Research Excellence Award from the Government of Ontario, and the Alfred P. Sloan Research Fellowship. He was the Program Co-Chair of CVPR 2003, ICCP 2010, and ICCV 2013. He was an Associate Editor of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE from 2005 to 2010.

Roman Genov (S'96–M'02–SM'11) received the B.S. degree in electrical engineering from the Rochester Institute of Technology, Rochester, NY, USA, in 1996, and the M.S.E. and Ph.D. degrees in electrical and computer engineering from Johns Hopkins University, Baltimore, MD, USA, in 1998 and 2003, respectively.

He is currently a Professor with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, where he is also a member of the Electronics Group and the Biomedical Engineering Group and the Director of the Intelligent Sensory Microsystems Laboratory. His research interests are primarily in analog integrated circuits and systems for energy-constrained biological, medical, and consumer sensory applications.

Dr. Genov was a member of the IEEE International Solid-State Circuits Conference International Program Committee. He is also a member of the European Solid-State Circuits Conference International Program Committee. He was a co-recipient of the Jack Kilby Award for Outstanding Student Paper at the IEEE International Solid-State Circuits Conference, the NSERC Discovery Accelerator Award, the Best Paper Award from the IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS, the Best Paper Award at the IEEE Biomedical Circuits and Systems Conference, the Best Paper Award in Biomedical Systems and the Best Student Paper Award at the IEEE International Symposium on Circuits and Systems, the Brian L. Barge Award for Excellence in Microsystems Integration, the MEMSCAP Microsystems Design Award, the DALSA Corporation Award for Excellence in Microsystems Innovation, and the Canadian Institutes of Health Research Next Generation Award. He was the Technical Program Co-Chair at the IEEE Biomedical Circuits and Systems Conference. He was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-II: EXPRESS BRIEFS and the IEEE SIGNAL PROCESSING LETTERS. He is currently an Associate Editor of the IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS and its Guest Editor of Special Issues on the IEEE International Solid-State Circuits Conference.