5.5 Dual-Tap Pipelined-Code-Memory Coded-Exposure-Pixel CMOS Image Sensor for Multi-Exposure Single-Frame Computational Imaging

Navid Sarhangnejad[1], Nikola Katic[2], Zhengfan Xia[1], Mian Wei[1], Nikita Gusev[1], Gairik Dutta[1], Rahul Gulve[1], Harel Haim[1], Manuel Moreno Garcia[3], David Stoppa[4], Kiriakos N. Kutulakos[1], Roman Genov[1]

[1]University of Toronto, Toronto, Canada
[2]Synopsys, Toronto, Canada
[3]Fondazione Bruno Kessler, Trento, Italy; [4]ams AG, Ruschlikon, Switzerland

Modern computational photography applications such as 3D sensing, gesture analysis, and robotic navigation drive the growing need for programmability, or coding, of the camera exposure at the individual-pixel level. Unlike conventional cameras, which record all light incident onto a pixel, the emerging class of coded-exposure-pixel (CEP) cameras can be programmed to selectively detect only *some* of that light [1] or, better, sort *all* of the light [2,3], depending on the pixel code. In conjunction with a concurrently coded illumination, this enables a wide range of new coded multi-exposure single-readout-frame imaging capabilities at video rates. This work demonstrates such an image sensor where multiple pixel-wise-coded exposures, or subframes, are accumulated during one video frame.

Over the past decade, CEP cameras were initially implemented using bulky, distortive, slow and expensive components like digital micromirror devices (DMD) together with off-the-shelf image sensors. In the last 2-3 years, the initial prototypes of fully integrated CEP image sensors have emerged, including a one-tap imager for compressive sensing [1]. 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 [2] and motion-deblurring [3] have recently been reported. By using two taps, all incident light is utilized regardless of applied codes. These imagers have large pixels (>12μm pitch) and low fill factors (<34%). None of them use a pinned photodiode-based (PPD) 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. Finally, the essential feature of global-shutter-style coded-subframe exposure for non-static codes is currently not available.

This work demonstrates a next-generation CEP image sensor that addresses all of the aforementioned limitations, and validates it in two novel single-frame computational imaging applications as illustrated in Fig. 5.5.1 (top). Figure 5.5.1 (top, left) depicts the single-frame structured light imaging application where: (1) 4 spatial sinusoidal patterns separated by a 90 degree phase shift are cyclically projected onto the scene; (2) synchronously with the projected patterns, 4 code matrices with time-multiplexed Bayer-like mosaic pattern, as shown, are cyclically submitted to the camera; and (3) the sorted photo-generated charges are accumulated across every 4 subframes and read out as two images per frame. Off-chip image processing reconstructs all lighting conditions at full spatial resolution. Finally, disparity and albedo maps are computed. Figure 5.5.1 (top, right) depicts the single-frame photometric stereo application. In this example, instead of sinusoidal patterns, 4 LED light sources are turned on, one at a time, during the 4 respective subframes within one frame. Following a procedure similar to that in structured light imaging, maps of surface normals and albedo are computed.

To enable such applications, a sensor with a general operating principle described in the flow chart in Fig. 5.5.1 (bottom, left) 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. To perform global coded-subframe exposure it requires an in-pixel dual digital code memory (Fig. 5.5.1 bottom, right). The code memory is pipelined: a code value is pre-loaded row-wise into each pixel during the previous subframe and is applied at the beginning of the new subframe. Photo-generated charge is collected based on the current code while the next code is being pre-loaded into the pixel.

Figure 5.5.2 (left) depicts the CMP architecture and the relevant timing diagrams. The CMP includes two D-latches, one controlled by ROW_LOAD to pre-load the code patterns row-by-row, and the other controlled by GLOB_LOAD to activate this pattern globally. Based on the code in each pixel, one of the two transfer gates, TG1 or TG2, connects the pinned-photodiode (PPD) to the corresponding floating diffusion node, FD1 or FD2, respectively. The pattern in each pixel is gated with EXPOSURE signal to stop any charge transfer during the read-out 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. Figure 5.5.2 (right) depicts the pixel layout floorplan and its cross-sectional electrostatic potential diagrams. A fill-factor of 45.3% is achieved with 27% of the area occupied by the latches and logic gates. The CMP operates at the maximum of 180 subframes/s with a high tap contrast of 99%.

Figure 5.5.3 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. The output is then serialized and buffered over 6 analog output channels.

Figure 5.5.4 depicts 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 3rd 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.

Figure 5.5.5 (top) depicts the albedo and disparity map reconstruction pipeline for the single-frame structured light 3D imaging setup where 4 subframes are used. As discussed above for Fig. 5.5.1 (top, left), at the end of a frame, each of the 2 tap images capture the visual scene sampled 4 times and coded by a four-pixel time-multiplexed Bayer-like mosaic pattern, corresponding to 4 90-degree-shifted sinusoidal illumination patterns. The Bayer-coded raw images from taps 1 and 2 (Fig. 5.5.5 top, left) are separated out into 4 images for each tap, each containing different structured light sine phase information (Fig. 5.5.5 top, middle). These images are then upsampled and processed to obtain 4 full-resolution images from which the output disparity and albedo maps are computed. Figure 5.5.5 (bottom) shows the results from a photometric stereo experiment. The top row depicts the surface normal maps of three different scenes captured at 20fps. The images are color-coded with the orientation of the surface normal at each pixel. The bottom row shows the albedo maps containing the intensity images of the same scenes.

Figure 5.5.6 compares the design with state-of-the-art multi-tap and other computational image sensors. Figure 5.5.7 depicts the micrograph of the chip.

Acknowledgement:
The authors gratefully acknowledge the support of the Natural Sciences and Engineering Research Council of Canada under the RGPIN, RTI and SGP programs, of DARPA under the REVEAL program, and of CMC Microsystems. We also wish to thank Hui Feng Ke, Hui Di Wang and Gilead Wolf Posluns for help on test system development.

References:
[1] J. Zhang, et al.,"Compact all-CMOS Spatiotemporal Compressive Sensing Video Camera with Pixel-wise Coded Exposure," *Optics Express*, vol. 24, no. 8, pp. 9013-9024, 2016.
[2] N. Sarhangnejad, et al.,"CMOS Image Sensor Architecture for Primal-Dual Coding," *Int. Image Sensor Workshop*, 2017.
[3] Y. Luo, et al.,"Exposure-Programmable CMOS Pixel With Selective Charge Storage and Code Memory for Computational Imaging," *TCAS-I*, vol. 65, no. 5, pp. 1555-1566, 2018.
[4] F. Mochizuki, et al.,"Single-Shot 200Mfps 5×3-Aperture Compressive CMOS Imager," *ISSCC Dig. Tech. Papers*, Feb. 2015.
[5] G. Wan, et al.,"CMOS Image Sensors With Multi-Bucket Pixels for Computational Photography," *JSSC*, vol. 47, no. 4, pp. 1031-1042, 2012.
[6] M. Seo, et al., "A 10.8ps-Time-Resolution 256×512 Image Sensor with 2-Tap True-CDS Lock-In Pixels for Fluorescence Lifetime Imaging," *ISSCC Dig. Tech. Papers*, 2015.

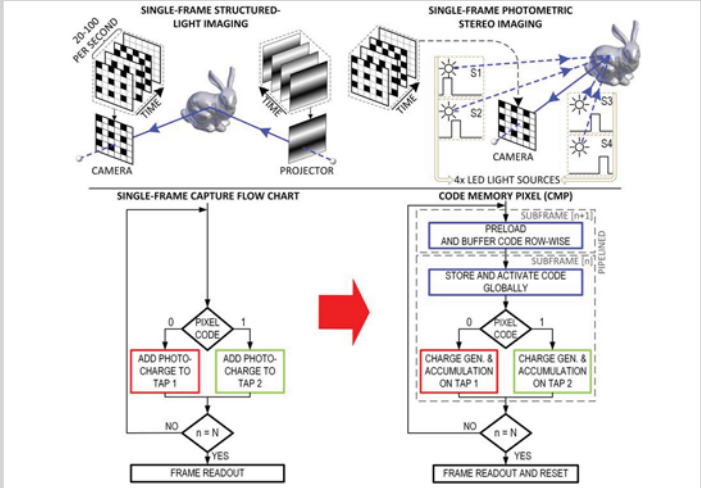

Figure 5.5.1: System configuration of single-frame structured light 3D imaging (top-left) and single-frame photometric stereo imaging (top-right). A generic flowchart of a dual-tap CEP sensor (bottom-left) and the flow chart for the CMP pixel architecture (bottom-right) are shown.

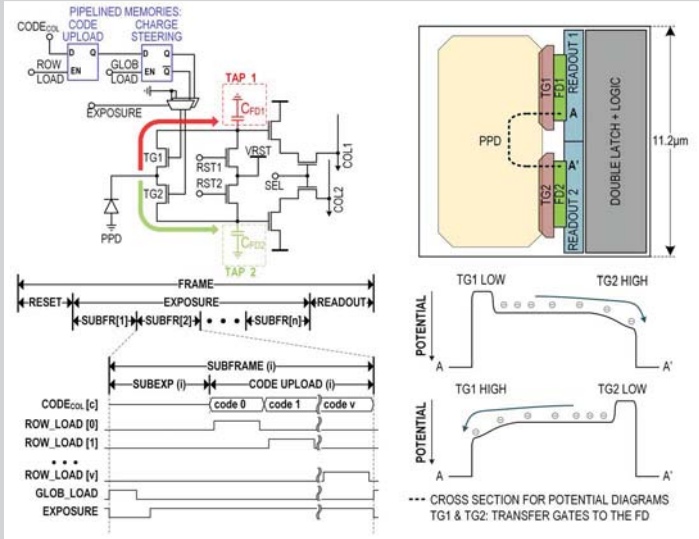

Figure 5.5.2: Schematic and layout of the code-memory pixel (CMP) with the associated timing and potential diagrams.

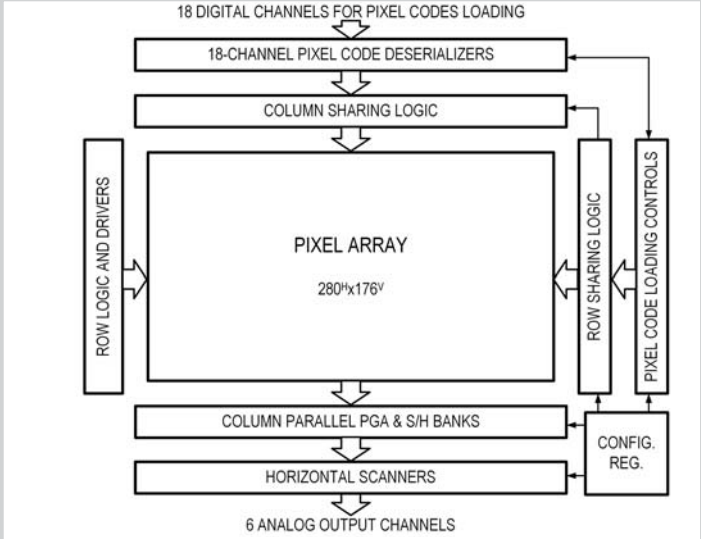

Figure 5.5.3: Image sensor top-level architecture with the code loading circuits and readout.

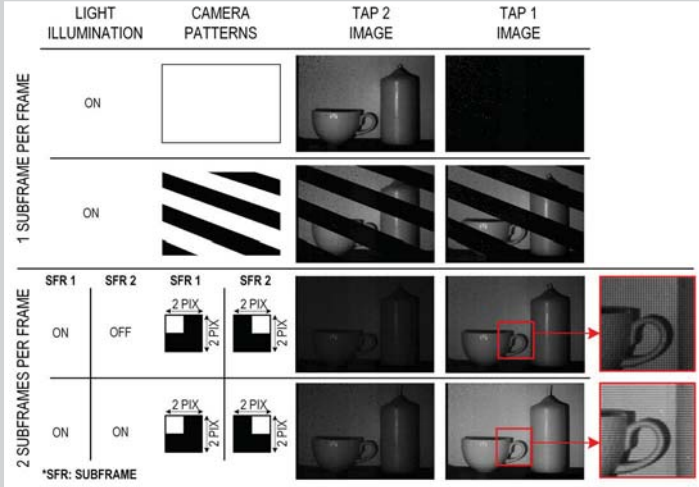

Figure 5.5.4: Raw output images of the sensor for two uniform-lighting cases: one (top) and two (bottom) subframes per frame. The insets on the right illustrate the difference between the two frames, with the same two sensor patterns but different illumination.

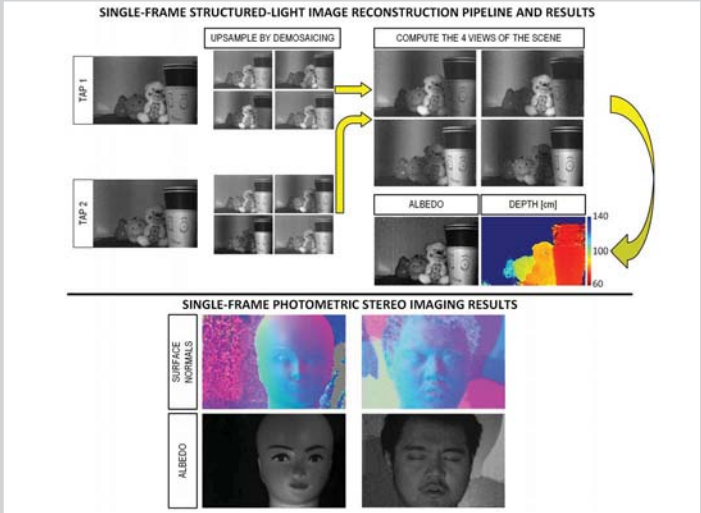

Figure 5.5.5: The albedo and disparity map reconstruction pipeline of single-frame structured light imaging (top); and examples of color-coded surface normal and albedo maps of the photometric stereo imaging (bottom).

		THIS WORK	[3]	[2]	[1]	[4]	[5]	[6]
	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
	PINNED PHOTODIODE	YES	NO (NW/P)	NO (PG)	YES	YES	YES	YES
	PIXEL PITCH [µm]	11.2	12.1x12.2	25	10	11.2X5.6	5	11.2X5.6
	FILL FACTOR [%]	45.3	33.2	20.5	52	–	42	–
	NUMBER OF TAPS	2	2	2	1 (PHOTONS ARE LOST)		2	2
	TAP CONTRAST	99 @ 180afps [1]	–	LOW [2]	N/A	N/A	–	94 [3]
ARCHITECTURE	PIXEL COUNT [Hx V]	244 x 162	10 x 10	80 x 60	127 x 90	64 x 108	640 x 576	256 x 512
	FRAME RATE [fps]	25	60	25	100	32	N/A	12
	POWER [mW]	34.4	1.23	7	1.3	1620	N/A	540
	POWER FoM [nJ] [4]	34	205	58	1.14	7324	N/A	343
SYSTEM	CODE MEMORY	IN-PIXEL (2 LATCHES)	IN-PIXEL (DRAM)	IN-PIXEL (2 LATCHES)	IN-PIXEL (SRAM)	OFF-PIXEL	N/A	N/A
	SUBFRAME RATE [afps [1]]	180	300000	–	100	N/A		
	PIXEL-CODE RATE [MHz]	7.1	30	–	0.11			
	ARBITRARY CODE / ROI	YES/YES	YES/–	YES/–	NO/–			
	SUBFRAME SHUTTER	GLOBAL	ROLLING	GLOBAL	ROLLING			
	APPLICATIONS	1.One-shot structured-light, 2.One-shot photometric stereo, 3.Other compressive sensing, etc.	Deblurring	Transport-Aware imaging	Spatiotemporal compressive sensing	Ultra high speed w/ compressive sampling	High-dynamic range	Fluorescence lifetime imaging

1: subframes per second
2: Limitations due to the type of the photo-detector (photogate-based)
3: Also known as Extinction Ratio
4: $FoM = Power/((Number\ of\ Pixels) \times (Frame\ rate))$
N/A: Not Applicable
–: Not Available

Figure 5.5.6: Comparison table.

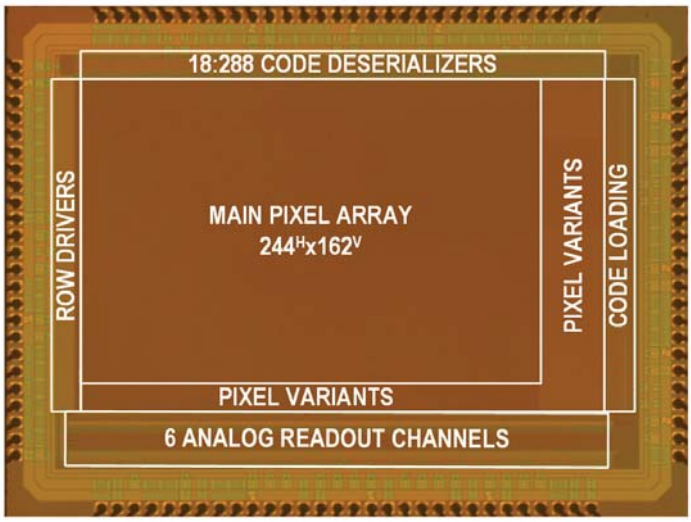

Figure 5.5.7: Chip micrograph (the dimensions are 4mm × 3mm).

978-1-5386-8531-0/19/$31.00