A 3.0_μW@5fps QQVGA self-controlled wake-up imager with on-chip motion detection, auto-exposure and object recognition

Arnaud Verdant, William Guicquero, Nicolas Royer, Guillaume Moritz, Sébastien Martin, Florent Lepin, Sylvain Choisnet, Fabrice Guellec, Benoît Deschamps, Sylvain Clerc, et al.

A 3.0µW@5fps QQVGA self-controlled wake-up imager with on-chip motion detection, auto-exposure and object recognition

Arnaud Verdant[1*], William Guicquero[1*], Nicolas Royer[1], Guillaume Moritz[1], Sébastien Martin[1], Florent Lepin[1], Sylvain Choisnet[1], Fabrice Guellec[1], Benoît Deschamps[2], Sylvain Clerc[2], Jérôme Chossat[2]
[1]CEA-Leti, Grenoble, France, [2]STMicroelectronics, Grenoble, France [*]These authors both contributed equally to the work.

Abstract

Analyzing image content usually comes at the expense of a power consumption incompatible with battery-powered systems. Aiming at proposing a solution to this problem, this paper presents an imager with full on-chip object recognition, consuming sub-10µW using standard 4T pixels in 90nm imaging CMOS technology, opening the path for both wake-up and high-quality imaging. It combines multi-modality event-of-interest detection with self-controlled capabilities, a key for low-power applications. It embeds a log-domain auto-exposure algorithm to increase on-chip automation. The power consumption figures range from 3.0 to 5.7µW at 5fps for a QQVGA resolution while enabling background subtraction and single-scale object recognition. This typically shows a measured 94% accuracy for a face detection use case.

Introduction

Preserving a well optimized 4T pixel while limiting ADC footprint and avoiding on-chip static power consumption were the three main design constraints for the wake-up imager. To achieve this while minimizing the dynamic power consumption, a scalable readout scheme has been developed, providing a thermometric-like bitstream format to lower complexity of digital pipeline operators. A 3-level awakening strategy implements 4b/6b motion detection and 6b object recognition. Fig. 1 reports its architecture including functional blocks. Fig. 2 describes *Fastscan* readout scheme. Fig. 3 depicts the 3 modes involving two interleaved exposure times.

Fastscan: from readout to A/D conversion

Instead of the conventional "reset-integrate-read-quantize" process to retrieve pixel intensity data, the *Fastscan* readout consists in multiple 1bit quantizations performed during the integration phase via successive entire focal plane scans, using rolling-based non-destructive pixels reads after the initial reset scan. *Fastscan* thus provides S bitplanes $X_b[p]$ ($p \in \{1,...,S\}$) such that $X_b[p] = (X_v[p] < V_{th})$ (vectorized: $\mathbf{x_b}[p]$), with V_{th} fixed and where $X_v[p]$ is the voltage-domain image at the output of the pixel source follower M_{sf}. $X_v[p]$ linearly evolves due to the photo-generated charge integration: $X_v[p] = V_{ref} - p\Psi T_s$; with T_s the inter-scan time, V_{ref} the reset voltage and Ψ the pixel photo-response in $[V \cdot s^{-1}]$. *Fastscan* data can be considered similar to Time-To-Saturation (TTS). Indeed, p such that $X_b[p] = 1$ implying $p > (V_{ref} - V_{th})/(\Psi T_s)$ leads to a multiplicative inverse representation of Ψ: $X_S = \Sigma_{p=1..S} X_b[p]$ (vectorized: $\mathbf{x_S}$). Fig. 2 details the *Fastscan* pixel initialization and readout. Using non-standard 4T readout has drawbacks in terms of pixel dispersions because of disabling Correlated Double Sampling to reduce Fixed Pattern Noise (FPN). To tackle this issue, an offset-calibration is performed during the initial reset scan using a pixel reset dependent on source-follower transistor's threshold voltage V_T. During the rolling reset of pixel rows, M_{sf} transistors are row-by-row configured in diode-connected mode with column readout buses biased to a reference V_{ref}. This way, the total M_{sf} gate&source capacitance made of pixel diffusion capacitance and power rail parasitic capacitances are discharged in 65ns, during the transition phase of M_{sf} from strong-to-weak inversion, until the gate's voltage equals $V_{ref} + V_T$. To maximize the pixel dynamic range, the transfer gate is permanently held ON during reset and integration. During each readout phase, the column capacitance is first set to ground. RS is then turned ON for 30ns to charge the column capacitance (250fF) without M_{sf} DC-current biasing, resulting in an incomplete settling of the column voltage V_{COL} from which the 1bit quantization is performed with an offset-compensated dynamic comparator. Combining low-level V_{COL} thresholding (50mV) with reduced voltage excursion on column capacitances, the associated dynamic power is restrained despite the multiple scans.

Digital Signal Processing and Sensor Scheduling

Fig. 1 depicts main digital components, which process successive bitplanes on-the-fly, without any full frame reconstruction. It enables column-based medians M extraction by detecting the scan index p for which the number of 1 is above half the size of the support (30/2=15). The column processors perform median-based background estimation and motion detection using a Δ modulation scheme. It also facilitates edge-detection filtering using basic XOR gates. Finally, a linear projection (K=1) on a 5×5 digitally binned image is performed on *Fastscan* data without any frame memory, thanks to in-line conditional accumulators (Condacc.) and a tree adder. To this end, 2 memory instances (3840bits) store the projection pattern (32×24) with 5b resolution. Fig. 3 depicts the sequence of SOC functions. Auto-exposure feedback control acts on a tunable ring-oscillator delivering the master clock that schedules *Fastscan* readout and processing. It also involves a frame interleaving to manage both processing modalities with two different equivalent exposure times (S times Ts), T_1 and T_2. T_2 is tuned to provide frames whose µ (the mean of M) is close to a programmable value (μ_t). T_2 frames are continuously adapted to the scene intensity, for mode3 (face recognition). On the contrary, T_1 which is dedicated to motion detection (mode1, mode2) does not change frame-to-frame. T_1 is periodically updated to T_2 at Timer2 interval to track slow scene variations. The ring-oscillator output frequency is tuned using a capacitor bank with a frequency step ratio of 1.25 (from 12kHz to 5MHz). The T_2 update undergoes the same ratios as $T_2 = T_m/1.25C$ with C the capacitor index (from 0 to 27) and T_m the longest period.

Test-chip measurements and sensor's characteristics

Fig. 4 reports test-chip acquisitions demonstrating FPN correction and image processing as performed on-chip. The SOC (Fig. 6) uses thick-oxide transistors to minimize leakages, with only 2 off-chip supply levels and no externally generated biases. For characterization purposes, all measurements are done under forced modes with or without auto-exposure, without considering IO ring power (2.0µW). The mode3 sensor's recognition capability has been qualified using a learned pattern (e.g., quantized SVM) applied to *Fastscan* edge-extracted images to discriminate face images from others. Compared to the characteristics of prior works (cf. Table I), the sensor offers a unique low power combination of key features, outperforming power consumption of [2], [3] and [5] in object recognition mode. It differs from [1] by preserving the pixel size and performance while implementing auto-exposure, from [2] and [3] which perform only acquisition or basic motion detection, and from [4] and [5] that embeds only intra-frame processing without any light conditions adaptation. Porting this design to a 3D-stacked technology will further decrease the power consumption and reduce the silicon footprint.

References

[1] J. Choi, et al., IEEE JSSC, pp. 289-300, Jan. 2014.
[2] J. Choi, et al., IEEE JSSC, pp. 130-140, Jan. 2016.
[3] O. Kumagai et al., ISSCC, pp. 86-88, Feb. 2018.
[4] K. Bong et al., IEEE JSSC, pp. 115-123, Jan. 2018.
[5] C. Young et al., ISSCC, pp. 98-100, Feb. 2019.
[6] Z. Liu et al., ICCV, pp. 3730-3738, Dec. 2015.
[7] A. Krizhevsky, https://www.cs.toronto.edu/kriz/cifar.

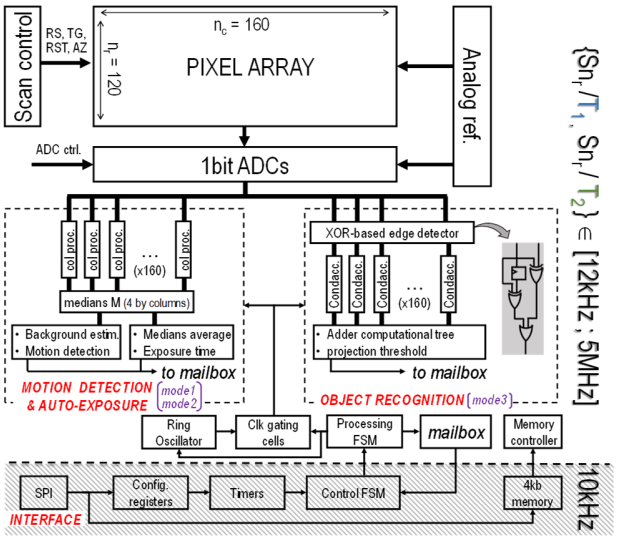

Fig. 1 Top-level wake-up CIS architecture.

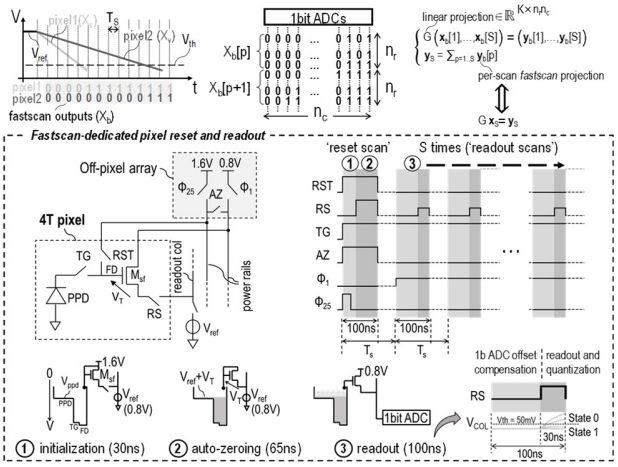

Fig. 2 *Fastscan* readout scheme concepts and related analog circuitry.

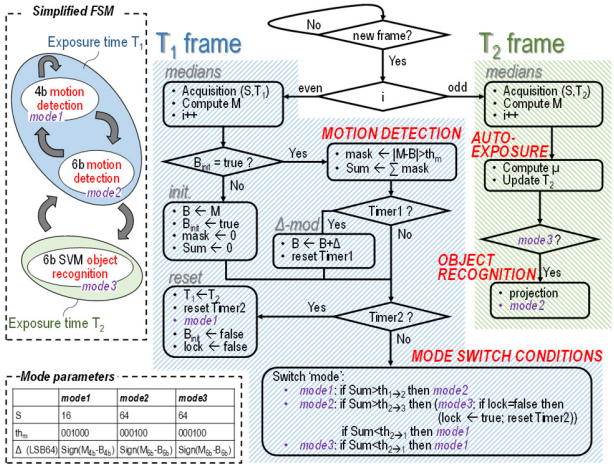

Fig. 3 Algorithm flow chart driven by the on-chip embedded FSM.

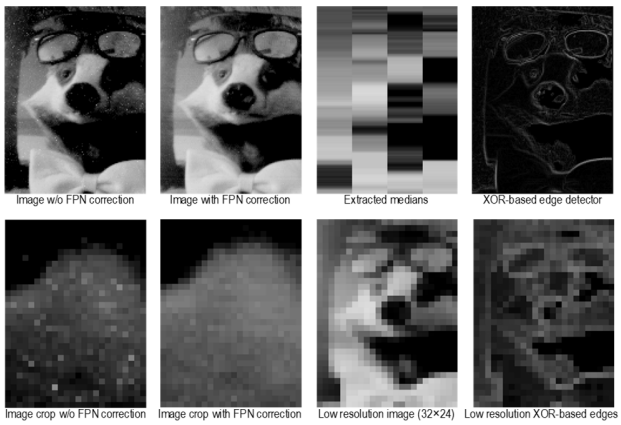

Fig. 4 Sensed images and related on-chip processing.

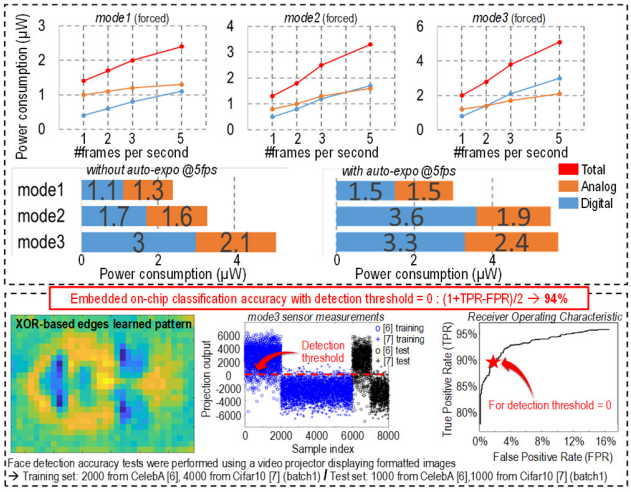

Fig. 5 Power measurements and inference performance.

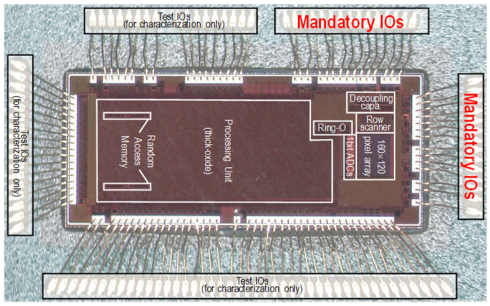

Fig. 6 Die micro-photograph.

TABLE I COMPARISON WITH PRIOR WORKS

	THIS WORK	Young et al. 2019 ISSCC [5]	Bong et al. 2018 JSSC [4]	Kumagai et al. 2018 ISSCC [3]	Choi et al. 2016 JSSC [2]	Choi et al. 2014 JSSC [1]
Technology	**90nm 1P4M (CIS)**	130nm 1P4M	65nm 1P8M	90nm 1AL4CU (CIS) 40nm 1AL6CU	110nm 1P4M	180nm 1P4M
Chip size (μm×μm)	**1830 × 4370**	4000 × 4000	3300 × 3380	4959 ×4401	5000 × 5000	2350 × 3180
Resolution	**QQVGA (160×120)**	QVGA (320×240)	QVGA (320×240)	3.9 M (2560 × 1536)	VGA (640×480)	(256×256)
Pixel type	**Standard 4T**	Standard 4T	Standard 4T	Shared cmd. 1.5T	Standard 4T	Custom 5.5T
Pixel size (μm×μm)	**3.2 × 3.2**	5.0 × 5.0	7.0 × 7.0	1.5 × 1.5	5.0 × 5.0	5.9 × 5.9
Master Clock frequency (Hz)	**10k ext. / 2k SPI 12k ~ 5M int.**	-	50M (analog) 5M ~ 100M (dig.)	6M ~ 27M	500k (320×240 pix)	25M
Frame rate (fps)	**5 (typ.)**	30 (typ.)	1 (typ.)	10 (sensing)	15 (typ.)	15 (typ.)
Dynamic Range (dB)	**36.0 (intra-frame)[1] 88.3 (inter-frame)[2]**	59.3	-	96	50	54.8
FPN	**<0.27%[1*]**	-	-	-	0.89%	0.05%
Supply voltage (V)	**1.6 (pixel, IOs) 0.8 (analog, digital)**	3.3 (analog)	2.5 (analog) 0.5 ~ 0.8 (digital)	1.8 (analog) 1.0 (digital)	0.9 (analog) 0.9 (digital)	1.3 (pixel) 0.8 (digital)
Power consumption (μW)		forced / auto-expo			45 (320×240 pix)	
Inter-frame proc.	*mode1*	**2.4 / 3.0**				
	mode2	**3.3 / 5.5**		1100		3.3
Intra-frame proc.	*mode3*	**5.1 / 5.7**	292	24 ~ 96		51
Energy/(pixel×frame) (pJ)		forced / auto-expo			34	
Inter-frame proc.	*mode1*	**25 / 31**				
	mode2	**34 / 57**		28		13.5
Intra-frame proc.	*mode3*	**53 / 59**	127	312 ~ 1250		52
ADC	• **1b comparator (*fastscan* ADC)**	• 2.75b Ratio-to-Digital Converter	-	• 10b SS ADC	• 12b SS ADC • 8b SAR ADC	• 8b SS ADC (with in-pixel amplifier)
Auto-exposure	**Yes**	No	No	Yes	No	No
Motion Detection	• **Background estimation with Δmodulation**	-	-	• Difference to static background	-	• Frame difference
Features extraction	• **Edge extraction**	• Log-gradients	• Haar-like	-	-	• HOG
Pattern recognition	• **5-bit linear SVM**	-	• Cascade	-	-	-

[1] in Time To Saturation (TTS) representation domain, i.e., on raw data in the *fastscan* format.
[2] taking into account auto-exposure (extrapolated from 1.25^{27}, i.e., 52.3dB).
[*] <0.175 LSB-6bit: no pixel FPN code dispersion observed under 6b quantization (i.e., 64 scans) more than 200 successive frames.