

# A 130 nm 165 nJ/frame Compressed-Domain Smashed-Filter-Based Mixed-Signal Classifier for “In-Sensor” Analytics in Smart Cameras

Anvesha Amaravati, *Student Member, IEEE*, Shaojie Xu, *Student Member, IEEE*, Justin Romberg, *Senior Member, IEEE*, and Arijit Raychowdhury, *Senior Member, IEEE*

**Abstract**—To avoid data deluge at the back end, it is imperative that advanced sensor nodes perform “in-sensor” processing to extract relevant features from data. We propose a compressed-domain smashed-filter-based object recognition system and measurements from a 130 nm mixed-signal test chip to demonstrate reconfigurable and ultra-low power operation. We measure greater than 90% accuracy in object localization with 165 nJ energy per frame on image data sets.

**Index Terms**—Gesture recognition, smashed-filter.

## I. INTRODUCTION

AS THE Internet of Smart Things continues to have tremendous societal impacts, human-machine interfaces are evolving in the modalities, accuracies and improved energy efficiencies. “Always-on” sensors, where these user interfaces need to be perpetually on pose significant power management challenges. To address this issue in camera based sensors, we present an “always-on” camera front-end which allows a sensor node to continuously acquire videos and monitor for a trigger that will wake up the back-end for video transmission when necessary; thus enabling exciting new usage models. To enable ultra-low power embedded classification, we take advantage of recent advances in compressed domain (CD) data processing which allows trigger detection with significantly lower power and computational requirements. There has been extensive work on compressive sampling high speed ADC architectures reported in [1]–[3]. In contrast to existing algorithms which work directly in the pixel domain [4], we extract image features in the compressive domain, which are linear random combinations of pixel values, followed by smashed filter based classification. Conventional compressed domain sensor front-ends include Nyquist ADCs followed by all-digital compression Fig. 1 [5]. This is typically followed by digital classifiers. For example, back-end support

Manuscript received October 22, 2016; revised February 18, 2017 and March 15, 2017; accepted March 18, 2017. Date of publication April 7, 2017; date of current version March 8, 2018. This brief was recommended by Associate Editor H. Schmid. (*Corresponding author: Anvesha Amaravati.*)

The authors are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: aamaravati3@gatech.edu; kyle.xu@gatech.edu; jrom@ece.gatech.edu; arijit.raychowdhury@ece.gatech.edu).

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

Digital Object Identifier 10.1109/TCSII.2017.2690861

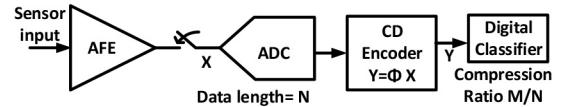


Fig. 1. Traditional CD classifier [5].

vector machines (SVMs) have been demonstrated in [5] for biomedical applications. These all-digital techniques do not take advantage of the energy-efficiency of mixed signal (MS) design. Embedding signal conversion along with classification has recently gained attention with hardware demonstrations in [6] and [7]. Embedded signal conversion and computation helps to reduce the power significantly since the classification is performed along with A to D conversion. The circuits reported in [6] and [7] operate on static images and applications related to image processing. However, a significant portion of the embedded applications on video data-streams, like gesture recognition, localization of moving objects etc. require the system to operate on the difference of frames. Also, the A to D converters reported in [6] and [7] operate in Nyquist speed therefore power efficiency is drastically reduced. The work reported in [7] reduces energy per compute however it uses the convolutional neural network (CNN) classifier which is computationally expensive from in-sensor classifiers. Research reported in [6] and [7] are targeted towards object classification.

Most of the video processing/object localization tasks operate on the difference of image frames. Object localization algorithms require parameter such as motion centers to be extracted. This is a highly computationally expensive process [8], [9]. Since difference image also performs background subtraction, we propose a mixed signal CD image processing pipeline which includes (1) switched capacitor based frame subtractor, followed by (2) in-situ random compressive measurements and (3) smashed filter based classification and Winner-Take All (WTA) back end to enable energy-efficient detection of key image features for always on sensor nodes. The design features a reconfigurable, gain-programmable compressive Digital to Analog Converter (DAC) followed by a Matrix-Multiplying SAR ADC MMADC) which performs smashed filter classifications directly in the CD. The proposed hardware, fabricated

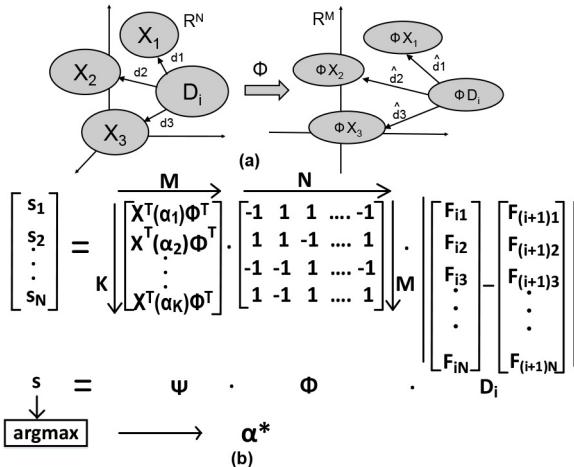


Fig. 2. (a) Inherent structure of a manifold is preserved in the CD. (b) Mathematical representation of the smash-filter operation.

in 130nm CMOS provides reconfiguration between a high accuracy (HA) and a low power (LP) mode with a graceful control of energy-accuracy trade-off. Section II describes mathematical background of smashed filter based object localization technique. Section III describes circuit implementation details. Section IV describes experimental results and Section V concludes this brief.

## II. SMASHED-FILTER-BASED CLASSIFICATION

The proposed design leverages recent advances in compressive data processing, where it has been shown that a manifold can retain its inherent structure in the CD with very high probability [10]. In particular, given a set of high dimensional vectors \$\{D\_i, X\_1, X\_2, X\_3 \in \mathbb{R}^N\}\$ lying on a manifold, the order of the distances between different vectors can be maintained after random linear projections into a space of much lower dimensionality (Fig. 2 (a)). This structure preserving property allows us to directly apply nearest-neighbor classifiers in the CD, reducing the overall computation (and hence, power consumption) without loss of classification accuracy. We perform motion center estimation/ moving-object localization in the CD, by judiciously embedding computation in the mixed-signal and digital domains.

Difference images are capable of capturing the motions. Denote \$F\_i\$ as the \$i\$th vectorized full resolution image output from the camera containing \$N\$ pixels in total. Vectorized difference image \$D\_i\$ of two consecutive frames is calculated as \$D\_i = |F\_{i+1} - F\_i|\$. To transfer \$D\_i\$ into the CD, we construct a random matrix \$\Phi\$ of size \$M \times N\$. Each entry of \$\Phi\$ is uniformly chosen from \$\{+1, -1\}\$ at random. The projection of the vectorized difference image in the compressed domain is calculated as:

$$\hat{D}_i = \Phi D_i \quad (1)$$

In the uncompressed domain, a difference image only has large non-zero values at the place where motion is present. We therefore construct a dictionary of templates, with each template being the same size as the difference image and having uniform non-zero values within a small rectangular region

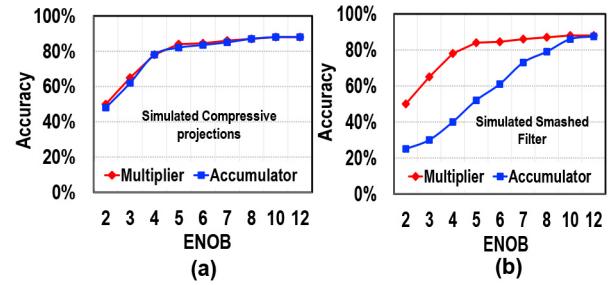


Fig. 3. Accuracy versus ENOB trade off for (a) compressive sensing and (b) smashed filters.

and zeros elsewhere. We vectorize these templates the same way we do to each frame and denote the set of \$K\$ vectorized templates as \$X(\alpha)\$, where \$\alpha\$ represents the coordinates of the center of the small non-zero rectangle. In this way, our templates resemble the difference images that have motion present at different locations. Motion is localized by solving

$$\alpha^* = \arg \min_{\alpha} \|D_i - X(\alpha)\|_2 \quad (2)$$

The collection of templates forms a manifold in \$\mathbb{R}^N\$ with intrinsic parameter \$\alpha\$. Using the result from [11], we can directly localize motion in the compressed domain. That is:

$$\hat{\alpha}^* = \arg \min_{\alpha} \|\hat{D}_i - \Phi X(\alpha)\|_2 \quad (3)$$

\$\hat{\alpha}^\* \approx \alpha^\*\$ with high probability for some \$M \ll N\$. By normalizing all the templates to have the same energy, we can further write motion center estimation as solving:

$$\hat{\alpha}^* = \arg \max_{\alpha} X^T(\alpha)\Phi^T \hat{D}_i = \arg \max_{\alpha} \Psi_{\alpha} \cdot \Phi D_i \quad (4)$$

The above process is known as smashed filtering [11] which is akin to matched filtering in the CD. It can be decomposed into multiplication with a compression matrix \$\Phi\$, a smashed filter matrix \$\Psi\$, followed by WTA (Fig. 2 (b)).

*Design Partitioning Between Digital and MS:* The proposed algorithm includes multiplication of the input with random number sequences and followed by smashed filter coefficients, followed by accumulation (Fig. 2). We analyze the tradeoff between the classification accuracy vs Effective Number of Bits (ENOB) used to represent the multiplier and accumulator outputs. Fig. 3 (a) illustrates that for compressive sensing using random binary matrix both multiplier and accumulator can be within 5-6 bits. Fig. 3 (b) illustrates that 80% (90%) accuracy is achieved with 5b (7b) multiplier ENOB but the accumulator needs to be at least 10b wide. We compare the power consumed in a MS vs. an all-digital implementation [12] and note: (1) for ENOB of 5-7, MS multipliers allow \$> 6X\$ reduction in power (lower process \$C\_{MIN}\$ lowers power consumption but increases linearity requirements at higher ENOB) and (2) high dynamic range requirement in accumulators make digital implementations \$> 12X\$ energy efficient at ENOB of 10. This allows us to make judicious design partition where the MMADCs enable MS multiplication followed by a digital accumulation and WTA. It should be noted that passive charge sharing multipliers having capacitors lower than 5fF (ENOB 4-5 bits) have low energy/operation compared to traditional

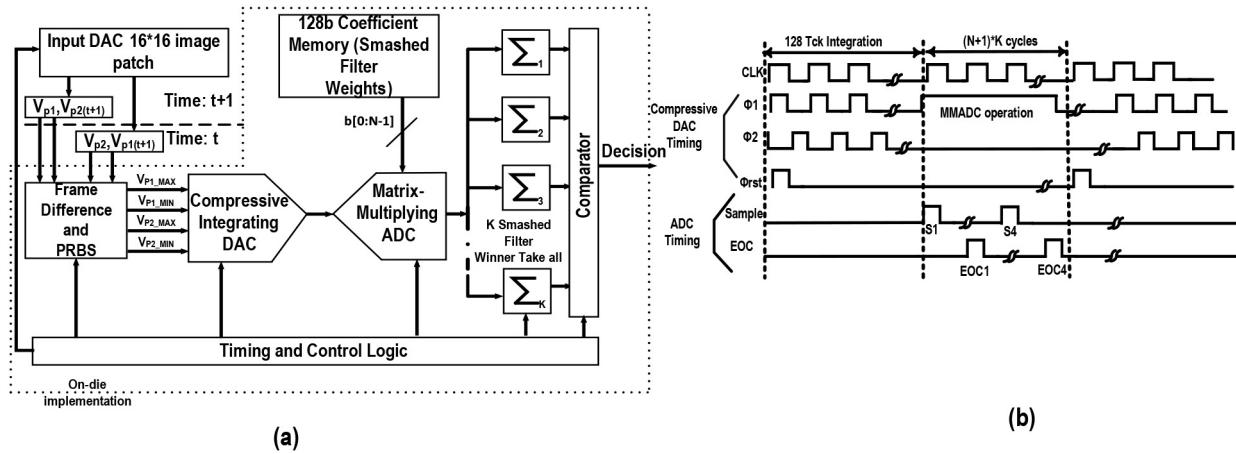


Fig. 4. (a) Proposed system architecture. (b) Timing diagram for the proposed system.

digital multipliers [12]. However charge sharing accumulators require amplifier with high gain, higher supply to enhance the dynamic range to more than 5-6 bits, which makes them infeasible for the proposed application (required ENOB is 10).

We propose compressed domain mixed mode classification system. The proposed reconfigurable system is shown in Fig. 4. We propose a reconfigurable system for “always-on” and high performance/power mode. “Always-on” mode is suitable for low power classification. If an object of interest is present the system can go into high performance mode which acquires the full signal/image from the sensor. During always ON mode we use compressive sampling to reduce the power in the signal acquisition and fixed point inner product circuit enabling low power computations. Mixed signal computation using switched capacitor is high energy efficiency compared to digital computation when the resolution is less than 6 bits. This is consistent with previous findings [12].

### III. HARDWARE IMPLEMENTATION

Fig. 4 illustrates the top level system architecture and the timing diagram of control signals as implemented in a 130nm test-chip. Frame-data is serialized and applied through an FPGA interface. The current implementation allows four parallel inputs (two from frame  $t$  and two from frame  $t + 1$ ). This represents a slice which can be parallelized in a video application processor. Key design blocks are described below.

**Multi-Input Compressive Sampling and Integrating DAC:** We propose a novel multi-input integrating DAC for in-situ compressive sensing. In the first stage: (1) a comparator identifies  $\max(V_{P(t)}, V_{P(t+1)})$ , (2) a PRBS generator combines a random sequence from +1, -1 with the comparator output to create  $V_{P1MAX}$  and  $V_{P1MIN}$  (Fig. 5). The truth table encoding the PRBS, absolute input difference is shown in Fig. 5. Mixed signal implementation reduces encoding power compared to a purely digital encoding and it requires less hardware resources for comparison and finding absolute value. The output of PRBS and absolute value circuit is given by:

$$V_{P1MAX} = \text{Phi} * \max(V_{P(t)}, V_{P(t+1)}) \quad (5)$$

and

$$V_{P1MIN} = \text{Phi} * \min(V_{P(t)}, V_{P(t+1)}) \quad (6)$$

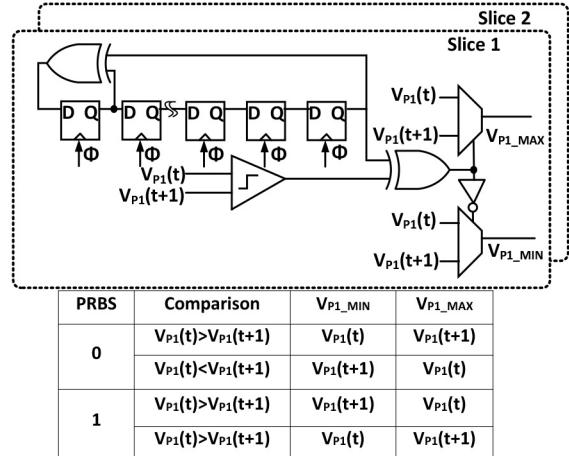


Fig. 5. Mixed signal PRBS and absolute value mixing circuit.

This acts as input to an integrating DAC (Fig. 6). We propose multi-input integrating DAC as opposed to single input integrating DAC reported in [13]. The test chip consists of 4 input integrating DAC. Multi-input integrating DAC helps to increase the throughput of the system by a factor of  $K$ , where,  $K$  is the number of inputs to integrating DAC. Capacitance  $C$  is the gain setting capacitor used to control the dynamic range at the output of the integrator. A reconfigurable OTA draws 14A (80A) of current for 5b (7b) operation. For an image patch of  $16 * 16$ , compressive sampling and integration is performed over 128 cycles and the output is given by:

$$V_{out}^{DAC} = \sum_{i=1}^{128} \frac{C_i}{C} * \Phi * abs(V_{Pi}(t) - V_{Pi}(t+1)) \quad (7)$$

The matrix multiplication is a true MS implementation where digital PRBS signals are combined with the absolute difference value of analog sensor inputs.

**Fixed Point Matrix Multiplying ADC:** Smashed filter operation is performed in a MM-ADC which operates on  $V_{OUT}^{DAC}$ . Smashed filter co-efficient ( $b_i$ , 5b or 7b) select  $V_{OUT}^{DAC}$  or 0 and the sampled voltage across the capacitive DAC of the SAR ADC is given by:

$$V_{out,samp} = \frac{1}{32} \sum_{i=0}^{4096} \left( 2^i b_i V_{OUT}^{DAC} \right) \quad (8)$$

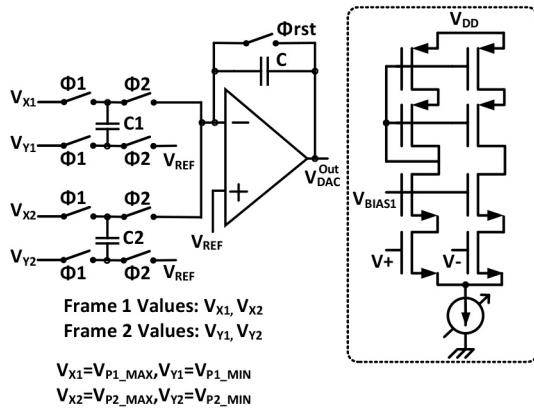


Fig. 6. Compressive integrating DAC and reconfigurable OTA.

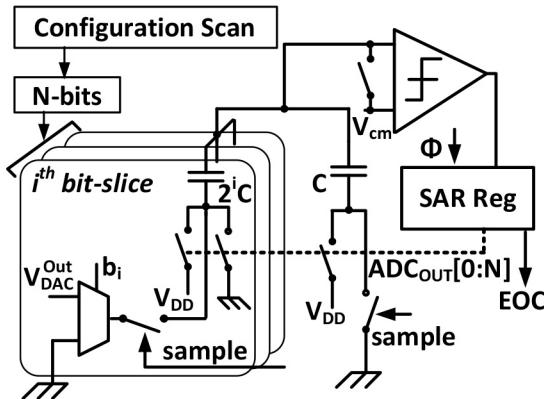


Fig. 7. Proposed matrix multiplying ADC.

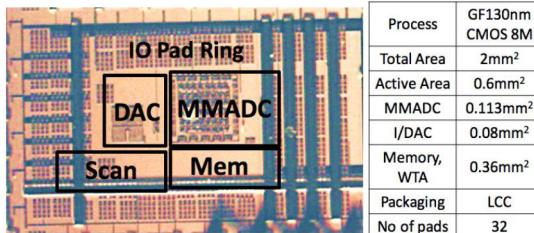


Fig. 8. Die picture of the chip.

Initial investigations reveal that the application requires 5-7b ENOB and justifies the use of a fixed-point SAR ADC architecture (Fig. 7). Smashed filter weights are stored in an 128b on-die memory array. The MMADC output is accumulated over K accumulators and a digital WTA identifies match/no-match with template patters. The entire design supports full-scan and internal nodes are exposed to high-speed pads for test and debug. Die photo and chip characteristics are shown in Fig. 8.

#### IV. EXPERIMENTAL RESULTS

The input waveform is applied to the compressive integrating DAC with the other input biased at the common mode voltage to measure the linearity of the front-end. Fig. 9 shows the FFT output of the ADC in the high accuracy 7b mode

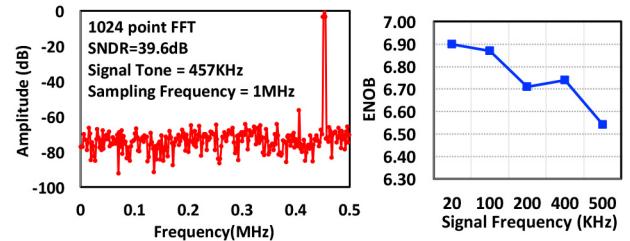


Fig. 9. Measured ENOB and SNR of the proposed front end.

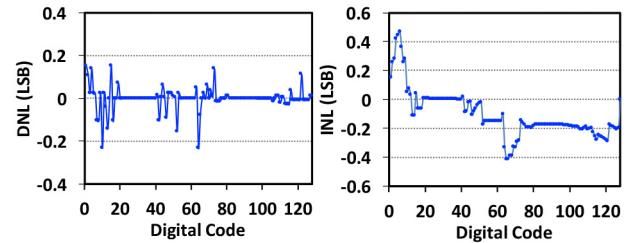


Fig. 10. Measured INL and DNL of the proposed front end.

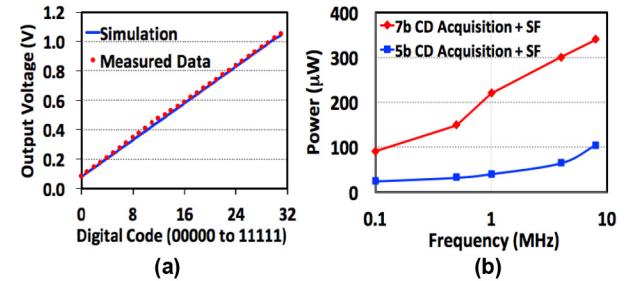


Fig. 11. (a) Measured linearity of proposed front end. (b) Power of the front end at two frequencies.

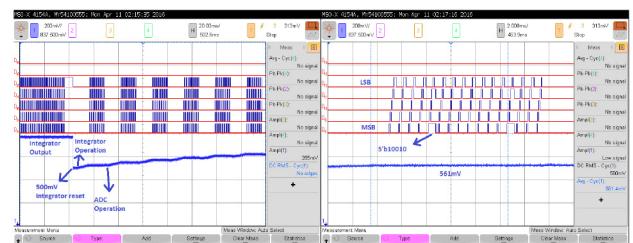


Fig. 12. Oscilloscope capture of Integrator and ADC outputs.

and the ENOB as a function of the signal frequency at a sampling rate of 1MHz. SNDR of 39.6dB is noted. Fig. 10 shows the measured linearity performance of the proposed front-end (SAR 7b mode). Both INL and DNL are less than 0.5LSB.

Fig. 11 (a) shows the linearity of the 5-bit MMADC where the output is converted to the ideal analog equivalent. Between the codes 8 to 15 the measured gain non-linearity is less than 0.1LSB. Fig. 11 (b) shows the power dissipated by the front-end during the two modes of operation. At 8MHz a total dissipated power of 104W and 340W are measured.

Fig. 12 shows the oscilloscope capture of the integrator and ADC outputs showing different phases of operation. Here the ADC is configured to operated in the 5 bit mode and number of localization centers is 16. We observe that while integrator

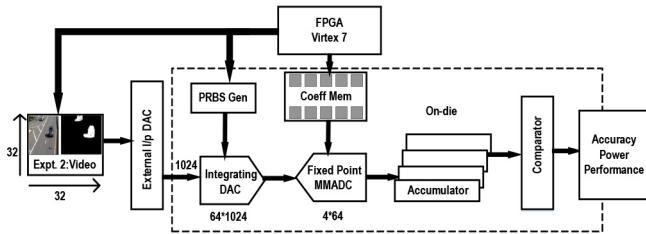


Fig. 13. Setup for system testing.

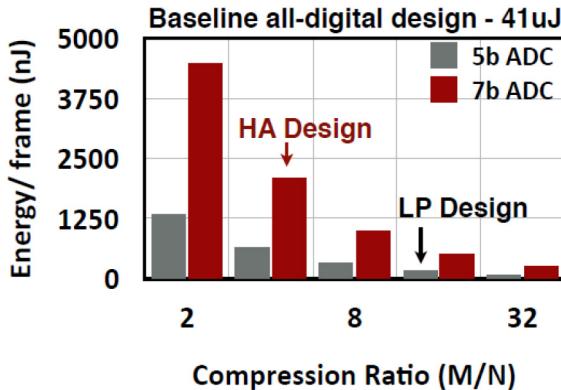


Fig. 14. Measured energy of the proposed system.

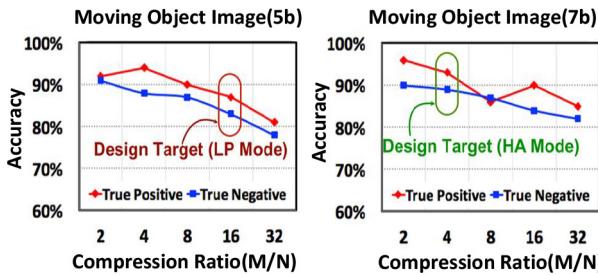


Fig. 15. Recognition accuracy for object classification and object location in a frame difference image.

does random compression, ADC outputs are zero. When ADC does fixed point matrix multiplication the integrator output stays constant. A typical ADC output of 5b10010 corresponding to 562.5mV is shown (561mV is the integrator output showing error within 0.5LSB).

System level measurements are carried out with image and video databases [8] (Fig. 13) to understand the power and accuracy trade-offs. In HA mode (target accuracy > 90%), a compression ratio of 4X with 7b MMADC is chosen and it is reconfigured to a compression ratio of 16X with 5b MMADC in the LP mode (target accuracy > 80%).

Measured energy-efficiency showed in Fig. 14 has 2108nJ/frame (HA mode) and 165nJ/frame (LP) mode illustrating a 248X improvement in energy efficiency compared to a baseline all-digital design using matched filters in the pixel domain. We estimate that 16X energy reduction comes from compressive sensing. Another 4X energy reduction is achieved by multi-input compressive DAC. 3.8X energy efficiency is achieved by using 5 bits of resolution in MMADC as opposed to a 12b digital data-path. Fig. 15 illustrates the true positive

TABLE I  
COMPARISON WITH REPORTED WORKS ON OBJECT DETECTION

	ISSCC 2015 [6]	ISSCC 2016 [7]	TCAS-I [5]	TCAS-II [1]	This work
System	Object/ ECG detection	Object detection	EEG Classification	CS ADC	Object/ object position detection
Detection Algorithm	Adaboost	CNN*	Compressed SVM	CS Recovery	Smashed Filter
Classifier System	ADC/WTA	DAC/ADC	Digital CORDIC	L1 norm recovery	Integrator/ ADC/WTA
Energy/ classification**	1.38μJ	0.8μJ	3.28mJ (cr=1) *** 0.13mJ (cr=16)	-	1.3μJ (cr=1) 0.16μJ (cr=16)
Technology	130nm	40nm	130nm	-	130nm
A/D Rate	20KS/s ENOB 7Bits	39MS/s 5-7bits	-	4GS/s 6bits	1MS/s 5-7 bits
Resolution	Analog/4b*/8b	Analog/3b/6b	-	Analog/6b	Analog/5b/7
TPR/TNR	0.83/0.89	0.86 (overall)	0.96/0.91	-	<b>0.87/0.83</b>

\* No ReLU(rectified linear), FC fully connected

\*\* 4 classes considered (32\*32 image)

\*\*\*No ADC

and true negative rates greater than 80% (90%) in 5b (7b) mode. The proposed design shows competitive figure of merit when compared to previously published results (Table. I).

## V. CONCLUSION

This brief presents a camera front-end with mixed-signal CD smashed-filter classification demonstrating an energy-efficiency of 165nJ/frame, thus making it suitable for low power mixed-signal camera front-end.

## REFERENCES

- [1] B. Hu, F. Ren, Z.-Z. Chen, X. Jiang, and M.-C. F. Chang, "An 8-bit compressive sensing ADC with 4GS/s equivalent speed utilizing self-timed pipeline SAR-binary-search," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 63, no. 10, pp. 934–938, Oct. 2016.
- [2] M. Trakimas, R. D. Angelo, S. Aeron, T. Hancock, and S. Sonkusale, "A compressed sensing analog-to-information converter with edge-triggered SAR ADC core," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 60, no. 5, pp. 1135–1148, May 2013.
- [3] W. Guo, Y. Kim, A. Sanyal, A. Tewfik, and N. Sun, "A single SAR ADC converting multi-channel sparse signals," in *Proc. IEEE ISCAS*, vol. 60. Beijing, China, 2013, pp. 2235–2238.
- [4] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: A survey," *Artif. Intell. Rev.*, vol. 43, no. 1, pp. 1–54, 2015.
- [5] M. Shoaib, K. H. Lee, N. K. Jha, and N. Verma, "A 0.16–107μW energy-scalable processor for directly analyzing compressively-sensed EEG," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 4, pp. 1105–1118, Apr. 2014.
- [6] J. Zhang, Z. Wang, and N. Verma, "A matrix-multiplying ADC implementing a machine-learning classifier directly with data conversion," in *Proc. IEEE ISSCC*, San Francisco, CA, USA, 2015, pp. 1–3.
- [7] E. H. Lee and S. S. Wong, "A 2.5GHz 7.7TOPS/W switched-capacitor matrix multiplier with co-designed local memory in 40 nm," in *Proc. IEEE ISSCC*, San Francisco, CA, USA, 2016, pp. 418–419.
- [8] A. Anvesha, S. Xu, N. Cao, J. Romberg, and A. Raychowdhury, "A light-powered, 'always-on', smart camera with compressed domain gesture detection," in *Proc. IEEE ISLPED*, San Francisco, CA, USA, 2016, pp. 118–123.
- [9] S. Xu, A. Amaravati, J. Romberg, and A. Raychowdhury, "Appearance-based gesture recognition in the compressed domain," in *Proc. ICASSP*, 2017.
- [10] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Found. Comput. Math.*, vol. 9, no. 1, pp. 51–77, 2009.
- [11] M. A. Davenport *et al.*, "The smashed filter for compressive classification and target recognition," in *Proc. Electron. Imag. ISOP*, San Francisco, CA, USA, 2007, Art. no. 64980H.
- [12] D. Bankman and B. Murmann, "Passive charge redistribution digital-to-analog multiplier," *IEEE EL*, vol. 51, no. 5, pp. 386–388, 2015.
- [13] D. Gangopadhyay *et al.*, "Compressed sensing analog front-end for bio-sensor applications," *IEEE JSSC*, vol. 49, no. 2, pp. 426–438, Feb. 2014.