A 23mW Face Recognition Accelerator in 40nm CMOS with Mostly-Read 5T Memory

Dongsuk Jeon[1,2], Qing Dong[1], Yejoong Kim[1], Xiaolong Wang[3], Shuai Chen[3], Hao Yu[3], David Blaauw[1], Dennis Sylvester[1]

[1]University of Michigan, MI; [2]Massachusetts Institute of Technology, MA; [3]Nanyang Technological University, Singapore

Abstract

This paper presents a face recognition accelerator for HD (1280×720) images. The proposed design detects faces from the input image using cascaded classifiers. A SVM (Support Vector Machine) performs face recognition based on features extracted by PCA (Principal Component Analysis). Algorithm optimizations including a hybrid search scheme that reduces the workload for face detection by 12×. A new mostly-read 5T memory reduces bitcell area by 7.2% compared to a conventional 6T bitcell while achieving significantly better read reliability and voltage scalability due to a decoupled read path. The resulting design consumes 23mW while processing both face detection and recognition in real time at 5.5 frames/s throughput.

Introduction

Reliably detecting and recognizing faces in an image is an active research topic in computer vision [1] for many application areas (Fig. 1). The Viola-Jones object detection algorithm is one of the fastest and most powerful approaches to face detection [2]. It consists of multiple weak classifier stages in which candidate images are either discarded or accepted. Candidates surviving through the last stage represent actual faces. SVM is widely used for classification in machine learning due to its sound theoretical background and consistently good performance across different application areas including face recognition. Although these algorithms provide good performance at reasonable hardware cost, they still require excessive amount of computations, especially for low-power mobile systems. In this work we propose a low-power face detection and recognition processor targeted at mobile applications.

Proposed Accelerator Architecture

The proposed accelerator first detects faces in an input image after which each detected face is recognized in a second stage (Fig. 2). Each face is normalized in size so that the recognition is performed consistently. In the recognition stage, well-known features called Eigenface [3] are extracted by applying PCA to the candidate images; the SVM takes these features as inputs to reduce dimensionality and computational cost while preserving recognition performance.

Fig. 3 shows the proposed accelerator architecture. As the image size grows the number of search windows per frame increases linearly, which translates to millions of windows for HD images. The detection step is typically a bottleneck and feature memory accesses dominate energy consumption. However, only a few of the windows contain actual faces and most candidate windows are discarded in early stages. The feature memory is therefore separated into two parts (Fig. 4); a small 1.5kB latch-based memory with lower read energy contains stages 1~5 (5% of total features) while a large SRAM memory stores all remaining stages. Testing on a custom database consisting of 180 HD images, 99.4% of rejected windows are discarded in the first 5 stages, reducing read energy by 20%; this contrasts with the more general approach of using a cache to reduce on-chip memory size [4].

To further reduce the number of feature memory accesses and energy consumption of the face detection block, we propose a hybrid search scheme (Fig. 5) that capitalizes on the innate tolerance of the Viola-Jones detector to deviations around the center of the face. Typically many windows are found that flag detection for a given face; these are then averaged to obtain an accurate center location. Instead, we first employ larger search steps to coarsely detect a face; faces are still correctly detected due to algorithmic tolerance and a low false negative rate. If the coarse window captures a face we expand the search around that point with smaller search steps so that the locations can be averaged for better localization. When the number of fine search windows flagged for positive detection exceeds a threshold value the point is passed to the recognition stage. This technique removes false positives that can occur in a single isolated location and also reduces the number of search windows by 12×. In simulation, it achieves 93% detection accuracy (F_1 score) when tested under the aforementioned custom database whereas the original algorithm has 76% accuracy due to high false positive rate. Recognition (classification) accuracy is 81% for 32-class (people) classification under the widely-used LFW face database [5].

Mostly-Read 5T Bitcell Memory

The proposed accelerator requires 492kB memory space to store coefficients such as support vectors and features. Once these memories are programmed at the beginning of operation, they need to be updated very infrequently (e.g., when new faces are added to database) and the system only reads out the stored values during operation. As the memory consumes the majority of the die area, a conventional 6T SRAM is a good candidate due to its small footprint. However, it suffers from limited voltage scalability and cannot provide aggressive low power operation. Other bitcell variants with isolated readout path (e.g., 8T [6]) have significantly better voltage scalability, but additional access transistors increase the memory area. To address these issues, we introduce a 5T SRAM bitcell with a decoupled read path (Fig. 6). The bitcell consists of an inverter pair and decoupled read transistor. The inverters use HVT devices for leakage reduction, while the SVT read transistor enables fast readout. The isolated read and write paths make readability insensitive to transistor size, allowing PDs and PUs to be minimum-sized. With an L-shape layout, the logic-rule 5T bitcell occupies 7.2% less area than a conventional logic-rule 6T bitcell.

The bitcell has separated cell VDD and VSS lines (Fig. 6). VDDL and VDDR are write bitlines shared per column, while VSSL and VSSR are write wordlines shared per row. In general, a value is written by raising VSSL (or VSSR) and pulling VDDR (or VDDL) down to write a 0 (or 1). However, since VSS lines are shared between two mirrored adjacent rows (Fig. 7), raising VSS lines will also flip the values in the neighboring row. Hence we first reset the entire memory array and then write values into consecutive rows from top to bottom (Fig. 8, left). During the reset process, all VSS lines are set to a low Vdd value VDDL and then are pulled down one after another from bottom to top, setting even and odd rows to 1 and 0, respectively. Write operation starts from the top by raising VSS1 and lowering VDDR0 to flip 1 to 0 (Fig. 8, right). Raising VSS1 may write 0s to the second row. However, this row is already reset to 0 and is unaffected. The same process continues for the entire array. Although complete reset is necessary before writing a new value, the accelerator requires extremely infrequent (or even one-time) bulk writing. This makes the 5T bitcell a good low power choice due to its large voltage scalability and HVT minimum-sized transistors; read energy is 38% less compared to a 6T design (Fig. 9 and Table 1). The design consumes 0.103pJ/bit/read at 100MHz and 0.6V (simulated).

Measurements

The proposed accelerator is fabricated in 40nm CMOS. Fig. 10 shows the die photo. The accelerator successfully detects and recognizes faces in test input images (example in Fig. 11). A 12kB 5T SRAM array has a measured V_{min} of 450mV. The system consumes 23mW from 600mV while processing HD images at 5.5 frames/s throughput with 4.5nJ/pixel efficiency, which enables continuous real-time face recognition even in mobile applications such as smartphones.

Acknowledgement

We thank TSMC University Shuttle Program for chip fabrication.

References

[1] S. Liao *et al*., CVPR, 2010. [2] P. Viola *et al*., CVPR, 2001.
[3] M. A. Turk *et al*., CVPR, 1991. [4] Y. Hanai *et al*., ISSCC, 2009.
[5] G. B. Huang *et al*., NIPS, 2012. [6] L. Chang *et al*., JSSC, 2009.

[7] N. Verma *et al*., ISSCC, 2007. [8] S. Nalam *et al*., CICC, 2009. [9] M.-P. Chen *et al*., VLSI, 2012.

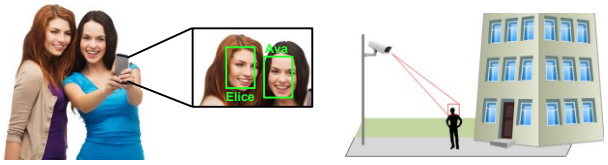

Figure 1. Applications of face detection and recognition.

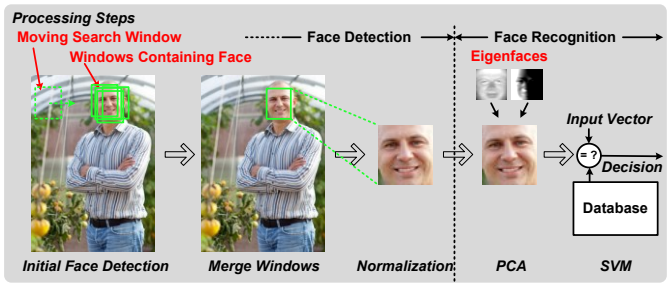

Figure 2. Processing steps of proposed accelerator.

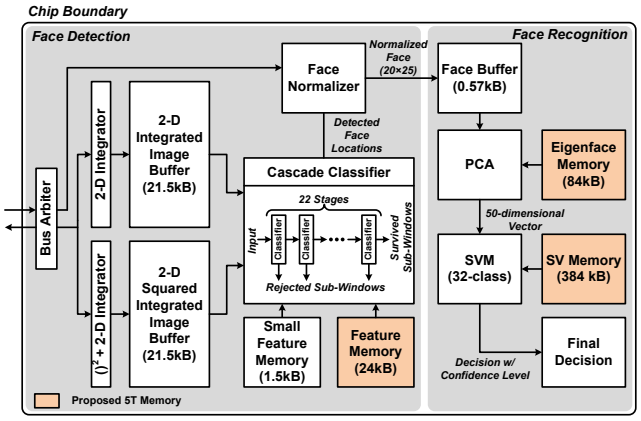

Figure 3. Proposed accelerator architecture.

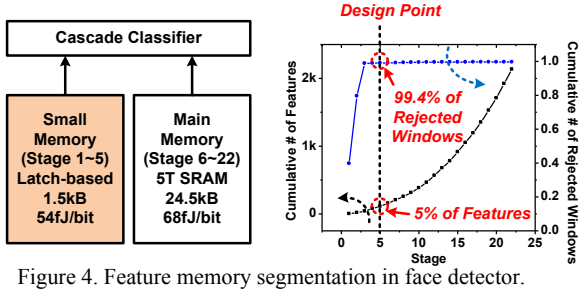

Figure 4. Feature memory segmentation in face detector.

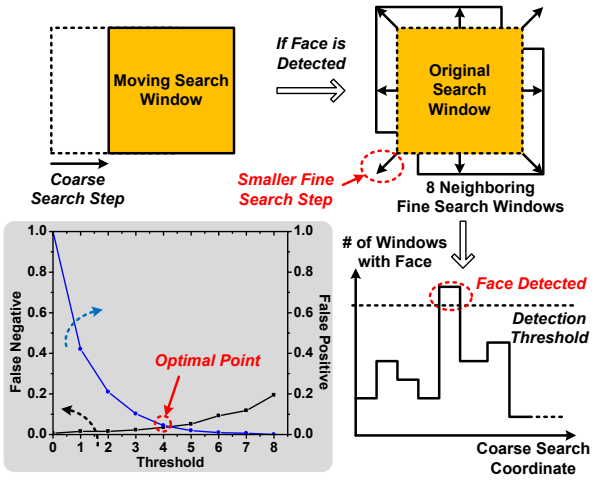

Figure 5. Hybrid search scheme for face detection.

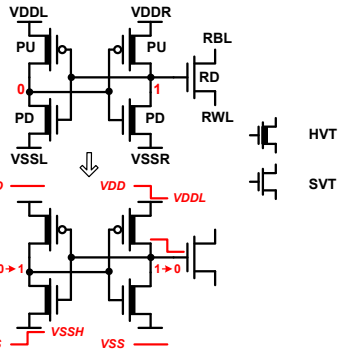

Figure 6. Proposed 5T bitcell and its basic write operation.

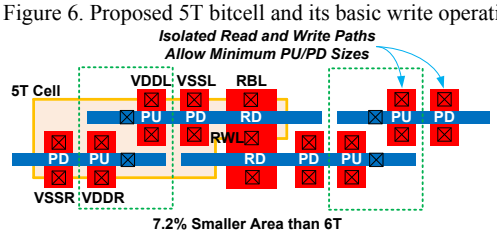

Figure 7. Layout of proposed 5T bitcell.

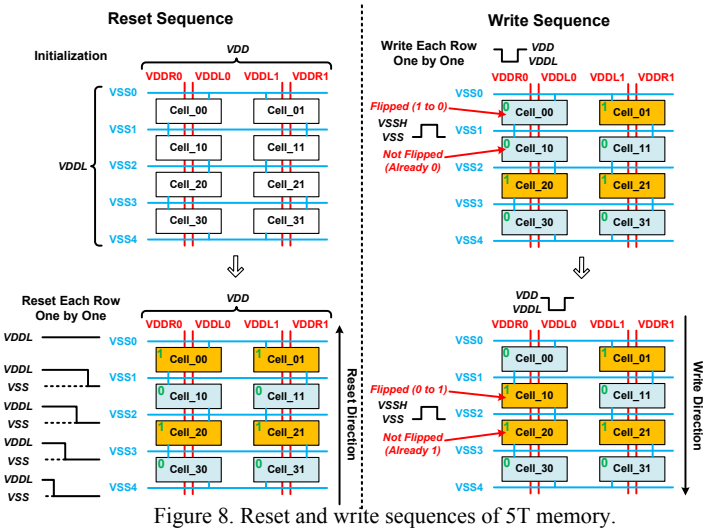

Figure 8. Reset and write sequences of 5T memory.

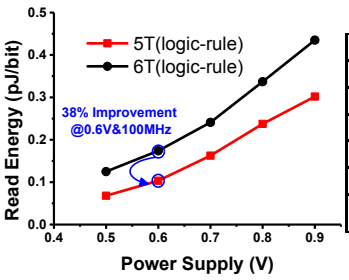

Figure 9. Simulated read energy.

Table 1. SRAM comparisons.

	[7]	[8]	[9]	This work
Process	65nm	45nm	65nm	40nm
Transistors	8T	5T	7T	5T
Voltage	0.35V	0.5V	0.26V	0.6V
Cell Size	1.3 x 6T	0.95 x 6T	1.15 x 6T	0.93 x 6T
Frequency	25kHz	250kHz	1.8MHz	100MHz
Read Energy (pJ/bit)	0.88 (0.35V)	8.8 (1V)	0.35 (0.26V)	0.103 (0.6V)

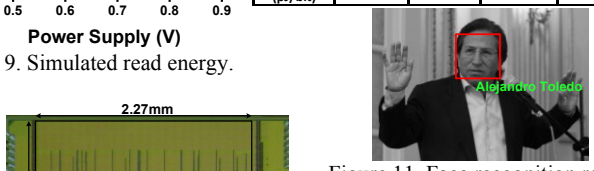

Figure 11. Face recognition result.

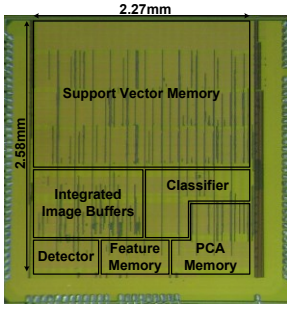

Figure 10. Die photo.

Table 2. Performance summary.

Technology	40 nm
Vdd	600 mV
Clock Freq.	100 MHz
Core Area	2.58×2.27 mm^2
Input Video	1280×720 (HD)
Power	23 mW
Efficiency	4.5 nJ/pixel