Memristive Threshold Logic Face Recognition

Akshay Kumar Maan¹, Dinesh S. Kumar², and Alex Pappachen James³

¹ Griffith University, Brisbane, Queensland, Australia
² Enview R&D labs, Kerala, India
³ Nazarbayev University, Astana, Kazakhastan

Abstract

This paper presents a face recognition method implemented using reconfigurable network of memristive threshold logic cells that can be practically realised in a secondary plane to the pixel arrays. Among the most distinguishing features of the presented system are a) an early detection and storage of only the relevant information directly from the sensors, b) a parallel, scalable information storage and detection architecture in hardware, as opposed to an algorithmic approach, and c) a fast and robust face recognition system. The threshold logic cell is inspired from a simplistic cortical neuron model that has multiple inputs with corresponding input memristors and one binary output. These cells when used with a set of input memristors are able to detect significant pixel variations in the incoming video frame and memorize the output template depending on the logic of selection of the resistor values. The implemented face recognition circuit shows small chip area, low power dissipation and ability to scale the networks with increase in image resolutions.

Keywords: Threshold Logic, Resistance Networks, Memristors, Object Detection

1 Introduction

The earliest attempts of pattern matching hardware [2], [1] focused on developing neuromorphic circuits to learning the fixed pattern images [4], [6]. The neuromorphic vision circuits such as retina pixel array [7], [11] have the limitation of low spatial resolution and difficulty of scaling to higher resolution. In contrast, the ability of the human visual system to use large number of pixels and with higher efficiency [10], [8] makes it significantly better than the existing neuromorphic solutions.

The need for higher imaging resolution increases with increase in natural variability such as that reflected through illumination changes [9], and age related changes. In real-time implementations, a finer space time resolution, i.e. high speed and high resolution would enable good quality features, however, it is practically difficult to accommodate such requirements in existing
systems due to high bandwidth, processing complexity of algorithms and memory requirements posed by recognition algorithms. In this paper we propose a hardware solution to the problem of the face recognition by proposing the idea of reconfigurable resistive threshold logic networks for memorising the image features and providing a matching process of face-image pixels from the sensor array. The proposed system is suitable to be implemented in subsequent planes to the pixel array, and pixel parallel analog to digital converters.

2 Proposed Method

Template Formation Block 1 and block 2 in Fig. 1a are used in the training stage in the proposed face recognition system. Block 1 takes digital data as input and does a bit-by-bit average of a fixed number of frames in time and applies the average to an inverter. The incoming picture frames come directly from sensor data and spatially aligned with respect to stable features in the images. These stable features are, for example, the location of eyes, nose, and mouth in facial images. Block 1 performs this average for each bit per pixel per a fixed number of frames in time. For an \( n \) input averaging circuit with input voltages \( V = \{V_1, V_2, \ldots, V_n\} \), the output voltage \( V_{\text{block}1} \) is given by:

\[
V_{\text{block}1} = \begin{cases} 
1 & \text{if } (1/n) \sum_{i=1}^{n} V_i \leq V_{th} \\
0 & \text{otherwise}
\end{cases}
\]  

(1)

where \( V_{th} \) is the threshold of the inverter. As an illustration, a batch of three frames arrive at the input in Fig. 1. For each bit of each pixel in the incoming frame, we have the circuit of Fig. 1. The first bits (MSB, in this case) of the first pixels of the first three frames are input to an averaging circuit and inverter (a logic NOT gate). The averaging circuit takes \( n \) (\( n = 3 \) in Fig. 1) equal resistors and connects one end of each to a common point, then apply input voltages (one to each of the resistors’ free ends), so the voltage seen at the common point is the mathematical average of the three. For example, if the three bits of the corresponding pixel in the three incoming frames are [1, 0, 1], the output of the averaging circuit will be 0.67 V which in turn applied to the NOT gate gives an output of 0V. Similarly, inputs of [1, 1, 0]V and [0, 0, 0]V average respectively to 0.67V and 0V and are inverted to 0V and 1V at the output of block 1. This block of the circuit is a pre-processing circuit employed to detect significant variations in the pixel intensities of the frames that may result due to, for instance, changing lighting conditions, facial expressions, and occlusions. Any of these changes within the frames will be detected as a state change from either 0V to 1V or 1V to 0V at the output of block 1. This change will be memorised by training the resistance values \( R_H \) and \( R_L \) in block 2 by the following logic:
Memristive Threshold Logic Face Recognition

Maan, Kumar and James

\[ R_{\text{block}2} = \begin{cases} R_H & \text{if } V_{\text{block}1} = 1V \\ R_L & \text{if } V_{\text{block}1} = 0V \end{cases} \] (2)

where \( R_H \) is a resistance of high value (\( \sim 10^9 \Omega \)) and \( R_L \) is a resistance of low value (\( \sim 10^6 \Omega \)).

Meanwhile, the input digital stream continues to be received at the input of block 1, while the input to block 2 \( V_{\text{block}2} \) is kept at logic 1 irrespective of the input changes. The purpose of this arrangement is to preserve the stable features of an image even under changing lighting conditions and occlusions. This input of logic 1 is applied to a voltage divider circuit followed by an inverter of block 2. The voltage divider circuit is composed of \( R_H \) or \( R_L \) depending on the output voltage of block 1 (\( V_{\text{block}1} \)) that controls the setting of the resistors in block 2, and ground resistance \( R_o (10^7 \Omega) \) so that the output \( V_{\text{block}2} \) of the block 2 is given by:

\[ V_{\text{block}2} = \begin{cases} 1V & \text{if } \frac{V}{R_H/L} \frac{1}{R_H/L + R_o} \leq V_{th} \\ 0V & \text{otherwise} \end{cases} \] (3)

**Face Recognition** Consider the circuit in Fig. 1b for an illustration of the working of a threshold logic cell. The voltage inputs for the cognitive cell are \( V_1, V_2 \) and \( V_3 \) and for which the corresponding resistance parameters are \( R_1, R_2 \) and \( R_3 \) respectively. A threshold logic cell usually processes a small region in an image, and groups several such cognitive cells in parallel to process the entire image. The input resistance parameters are set to either \( R_H \) or \( R_L \) depending upon the input values. The resistor parameters in a cognitive cell are set only when the template image is applied. Generally, for an \( n \)-input cognitive cell, the resistance values \( R = \{ R_1, R_2, R_3, ... R_N \} \) for a corresponding set of input values \( V = \{ V_1, V_2, V_3, ..., V_N \} \) are set based on:

\[ R_i = \begin{cases} R_H & \text{if } V_i > V_t = \frac{1}{n} \sum_{i=1}^{n} V_i \\ R_L & \text{otherwise} \end{cases} \] (4)

where \( V_t \) is the threshold value provided by the average of the applied inputs. The output voltage \( V_o \) of the potential divider of a cell with \( n \) inputs is given by \( V_o = \frac{\sum_{i=1}^{n} V_i/R_i}{1/R_o + \sum_{i=1}^{n} V_i/R_i} \).

Thus the output \( V_{OUT} \) of the inverter is given as:

\[ V_{OUT} = \begin{cases} 1V & \text{if } V_o > V_{th} \\ 0V & \text{otherwise} \end{cases} \] (5)

where \( V_{th} \) is the threshold of the inverter. Eq. 5 is a simplified form of a more realistic logistic function, and it is implied by default that in practical realization logical functions are used. At the output of these cells another averaging circuit performs an average of all the threshold logic cell outputs. Hence, all the cells are essentially turned on by the template image.

In order to extract the facial features from the images we develop a edge feature extraction [5] method with memristive circuits. Fig 2 shows the overall process of edge detection with the proposed threshold logic networks. The grayscale normalised input image is applied to the black to white and white to black networks to find the intermediate output voltage. The inverter threshold voltage \( V_{th} \) is chosen as 0.3V. The input image is applied to the network by taking 2 x 2 pixels blocks a time. Here, the input resistance value of the threshold logic network cell is set at high memristance \( R_H \), when the input voltage is a logic low state, i.e. \( V_i < V_{th} \). The input resistance, is set at low memristance \( R_L \), when the input voltage is a logic high state i.e. \( V_i > V_{th} \). This is the white to black network cell. The input image is inverted and applied to another network which is black to white network. In this network the input resistance value
Figure 2: Figure shows how $2 \times 2$ pixel blocks of the input image are applied to the black to white network and white to black network, whose outputs are XOR to form output pixels detecting the edges in the input image.

of the threshold logic network cell is set at high memristance $R_H$, when the input voltage is a logic high state, i.e. $V_i > V_{th}$ and the input resistance is set at low memristance $R_L$ when the input image is a logic low state, i.e. when $V_i < V_{th}$. This is the black to white network. The XOR output voltage then forms the final image containing the edges of the original image.

**Experimental Setup** The simulations of area and power requirements of the proposed cognitive cell were performed with a SPICE model of memristor [3] using feature size of 0.25μm TSMC process, BSIM models and HP memristor model. For studying the system performance the template formation and recognition phases of the system were tested on three face-recognition databases.

**Results** Fig 3a shows the effect on the recognition accuracy of the three databases with the resolution of the train and test images. Fig. 3b shows this useful thresholding result for the hardware edge detection in the recognition accuracy achieved. We find that the accuracy increased up to an optimal value of the block size (6) after which it began to decrease. This has important implication in the density and resulting area and power dissipation of the cognitive memory network.

Table 1a, shows the performance analysis of the circuit for template formation. In this circuit, each pixel has been divided into 8 bits. Each bit has its own template formation circuit as shown in Fig. 1a. For this single bit circuit the area, power dissipation and leakage power have been found to be resp. $9.41 \mu m^2$, $2.03 nW$ and $4.60 pW$. This circuit has a delay of 2.55 ns. Since each and every bit of all the pixels has its own circuit, the area, power dissipation, leakage power and delay for images of any dimension will scale linearly. This has been verified by simulations as shown in Fig 4a.

---

1YALE Face Database, ORL and faces94 using MATLAB. The images in the databases are fed one after the other to the cognitive network to simulate video frames from a camera. YALE database contains 15 classes each containing 11 images of human faces in different poses and lighting conditions. ORL database has 40 different classes each containing 10 images. For faces94 database we tested on 20 classes each containing 20 images. Each database was equally and randomly divided into train and test images.

2In our experiments an image resized to $80 \times 80$ pixel resolution gives the best results for ORL and faces94 databases, whereas YALE peaks at $60 \times 60$. We found that a block size of $6 \times 6$ gave the best accuracy on the YALE and ORL databases.
Figure 3: (a) Variation of recognition accuracy with increase in pixel resolution for train and test images for the three databases - AR, ORL and faces94. (b) Variation of recognition accuracy with edge detection cell size resolution for train and test images for databases - YALE, ORL.

Table 1: (a) Performance analysis of the circuit for template formation. (b) Performance analysis of the Face recognition circuit

(a) Dimension | Area ($m^2$) | Power dissipation (W) | Leakage Power (W)
---|---|---|---
3 x 3 | 677.37μ | 145.87n | 333.45p
10 x 10 | 7500μ | 1.62μ | 3.71n
100 x 100 | 0.75m | 162.08μ | 370.50n
1000 x 1000 | 75.26m | 16.21m | 37.05μ

(b) Dimension | Area ($m^2$) | Power dissipation (W) | Leakage Power (W)
---|---|---|---
3 x 3 | 18.80μ | 15.17n | 9.4 p
10 x 10 | 117.52μ | 75.51n | 58.75 p
100 x 100 | 11751μ | 7.55 μ | 5.88 n
1000 x 1000 | 1.175 m | 0.76 m | 0.59 μ

Table 1b gives the performance analysis of face recognition network. The basic cell is a 4 input cognitive cell. The area, power dissipation and leakage power of this basic cell are 4.7 μm$^2$, 3.88 nW and 2.35 pW respectively. As in template formation stage these parameters

Figure 4: (a) Variation of power dissipation, leakage power and area with the number of pixels processed in the template formation stage. (b) Variation of power dissipation, leakage power and area with the number of pixels processed in the face recognition stage.
scale almost linearly with number of the pixels processed since each bit of each pixel has an independent cognitive memory circuit. The linear variation is shown in Fig 4b.

3 Conclusion

The presented approach addresses the problem of face recognition in real-time at high frame-rates. It provides a hardware-only method of image template formation and recognition, whose recognition performance takes advantage of but is not adversely affected by high resolution images or high frame rates. The method includes a repeated estimation of object properties based on intensity changes from one frame to another frame. The cognitive cells that form the primitive level components of a cognitive network can have multiple numbers of digital inputs but can have only one digital output. Such a cell is used to form a network that in turn is able to detect the pixel intensity changes in the images across time and across space. Image frames can be memorized in the network by adjusting the cognitive cell parameters according to the input. Face recognition is then implemented by quantifying the change in the current image frame with respect to a memorized frame at the output of the network.

References


