VLSI Implementation of a Skin Detector Based on a Neural Network

Farid Boussaid [(1)], Abdesselam Bouzerdoum [(2)], and Douglas Chai [(3)]

[(1)] University of Western Australia, School of Electrical and Computer Engineering
[(2)] University of Wollongong, School of Electrical, Computer and Telecommunications Engineering
[(3)] Edith Cowan University, School of Engineering and Mathematics
AUSTRALIA

Abstract – **This paper describes the VLSI implementation of a skin detector based on a neural network. The proposed skin detector uses a multilayer perception with three inputs, one hidden layer, one output neuron and a saturating linear activation function to simplify the hardware implementation. The skin detector achieves a classification accuracy of 88.76%. To reduce mismatch associated errors, a single skin detection processing unit is used to classify all pixels of the input RGB image. The current-mode fully analog skin detection processing circuitry only performs computations during the read-out phase, enabling real-time processing. Fully programmable, the proposed skin detection processing circuitry allows for the external control of all classifier parameters to compensate for mismatch and changing lighting conditions.**

I. INTRODUCTION

Skin detection enables a wide range of machine vision tasks such as the detection, tracking and recognition of face and gesture, which are required for human-machine interaction [1]. Skin detection is typically achieved using color information [2][3][4], which is a feature that can be computationally inexpensive and thus well suited for real-time applications. A number of algorithms have been proposed to achieve color-based skin detection, including statistical methods, neural networks or template matching [2][5][6]. Most of the research conducted in the past has aimed at achieving robust skin detection irrespective of image capture conditions (e.g. lighting conditions) or human physical characteristics. The actual integration of skin detection processing on a single silicon chip has received little attention. In [7], Perez and Koch proposed a CMOS analog implementation of an RGB (Red-Green-Blue) to HSI (Hue-Saturation-Intensity) color space conversion, to provide for reduced sensitivity to illumination conditions. In [8], Etienne-Cummings *et al.* implemented the principle of HSI-based pixel segmentation. The fabricated CMOS image sensor uses template matching to achieve pattern or skin recognition, and relies on memory elements to store the different templates. As a result, the proposed VLSI implementation leads to a substantial increase

in silicon area and does not allow for real-time processing. In this paper, we present a skin detector that is well suited for VLSI implementation. Based on multilayer feedforward neural network architecture, the proposed skin detector enables real-time processing and removes the need for on-chip memory elements. The paper is organized as follows. In the next section, the proposed skin detector is presented and evaluated. Section 3 describes its VLSI implementation for real-time skin detection processing. Finally, concluding remarks are given in Section 4.

II. SKIN DETECTION USING NEURAL NETWORKS

Neural networks have the ability to learn complex data structures from a set of example patterns [9]. They have the advantage of working fast (after the training phase) even with large amount of data. The results presented in this paper are based on a multilayer feedforward network architecture, known as the multilayer perception (MLP). The MLP is a powerful tool that has been used extensively for classification, nonlinear regression, speech recognition, hand-written character recognition and many other applications [9]. The elementary processing unit in a MLP is called a neuron or perceptron. It consists of a set of input synapses, through which the input signals are received, a summing unit and a nonlinear activation transfer function. Each neuron performs a nonlinear transformation of its input vector; the input-output relationship is given by

$$\varphi(X) = f(\sum_{j=1}^{P} w_j x_j + \theta) = f(W^T X + \theta), \quad (1)$$

where W is the synaptic weight vector, X is the input vector, θ is a constant called the bias, f is the activation function, superscript T is the transpose operator, and $\varphi(X)$ is the neuron output signal.

An MLP architecture consists of a layer of input units, followed by one or more layers of processing units, called hidden layers, and one output layer. Information propagates, in a feedforward manner, from the input to the output layer; the output signals represent the desired information. The input layer serves only as a relay of information and no information processing occurs at this layer. Before a network can operate to perform the desired task, it must be trained. The training process changes the parameters of the network in such a way that the error between the network outputs and the target values (desired outputs) is minimized.

In this paper, we propose a method to detect skin color that is suitable for VLSI implementation. The skin detector uses an MLP with three inputs, one hidden layer and one output neuron (see Figure 1). To simplify the VLSI implementation, we use a saturating linear activation function. Each pixel is represented by its RGB (red, green and blue) color components. These three color components are used as inputs by the neural network. The output of each hidden neuron is given by (1), and the network output is given by

$$y = \sum_{j=1}^{Q} c_j \varphi_j(X) + \beta \qquad (2)$$

where $\varphi_j(X)$ is the output of the j-th hidden neuron, and c_j is the synaptic weight of the output neuron.
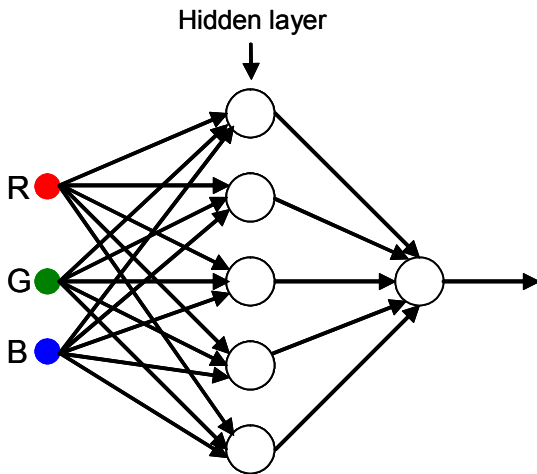

Figure 1: Neural network architecture for skin detection.

To estimate the neural network parameters (i.e. synaptic weights and biases), a training set containing 3,135 skin and non-skin pixels was extracted from set of images. The network was trained using the Levenberg-Marquardt backpropagation algorithm [1]. The generalization ability of the trained network is tested using a set containing 1.5673 million skin and non-skin pixels. The training and test sets were extracted from images containing skin colors of people from different races and under different lighting conditions. Figure 2 shows the ROC (receiver operating characteristic) curve on the test set. Overall, the neural network achieves a classification accuracy

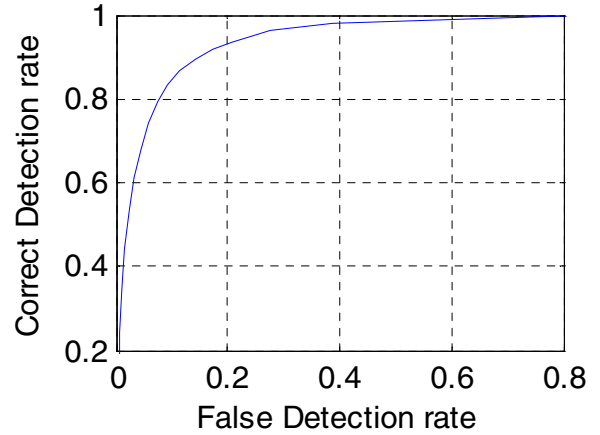of 88.76% on the test set. Figure 3 presents an image and the detected skin regions.

Figure 2: Test set ROC curve of the trained neural network: the inputs are the RGB components.

Figure 3: Original image (top) and skin segmented image (bottom).

III. VLSI IMPLEMENTATION

Figure 4 depicts the proposed VLSI architecture for a CMOS image sensor integrating skin detection processing. The image sensor uses currents as pixel output signals to take full advantage of current-mode processing and enable real-time processing [10][11]. In the adopted current-mode approach, sums are computed by simply wiring the appropriate signals,

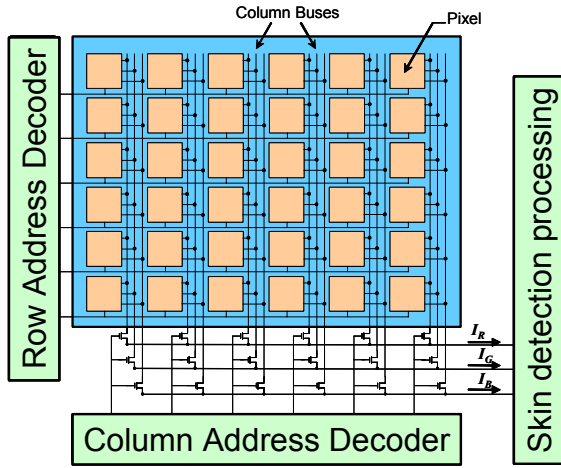and differences by means of simple current mirrors (see Figure 5).
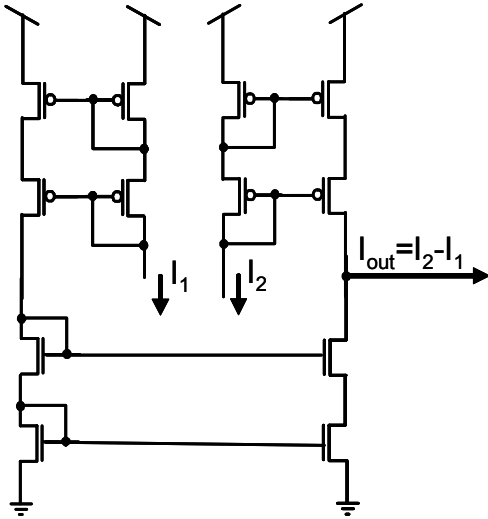

Figure 4: Image sensor architecture.

Figure 5: Cascade current-mode subtractor.

To make the three primary colors R, G and B available from each pixel at any given time, three vertically integrated photodiodes are used as an RGB in-pixel color detector. Here, color separation is achieved using the strong wavelength dependence of the absorption coefficient in silicon. This wavelength dependence causes a very shallow absorption of blue light and enables red light to penetrate deeply in silicon [12]. As a result of the selected color capture mode, pixels within a column will thus share three common column buses, corresponding each to a primary color (see Figure 4). Pixels are selected individually for output current read-out using conventional row/column counter address decoders. Each time a pixel is selected for read-out, three output currents I_R, I_G and I_B are simultaneously handed out to the skin detection processing circuitry, which classifies the pixel as skin or non-skin pixel. In the proposed approach, on read-out skin detection processing is thus achieved enabling "real-time"

processing. The hardware realization of the proposed skin detector can be described as follows. The bias associated to each neuron is implemented by means of an externally controlled current source. A saturating linear activation function $f(x)$ is used to simplify the VLSI implementation. The saturating linear activation function is defined as follows: (a) $f(x) = 0$ for $x < 0$; (b) $f(x) = 1$ for $x > 1$ and (c) $f(x) = x$ for $0 < x < 1$. Such an activation function can easily be implemented using current comparators as shown in Figure 6, which determine whether the current is positive or negative (here zero is represented by a small current value).
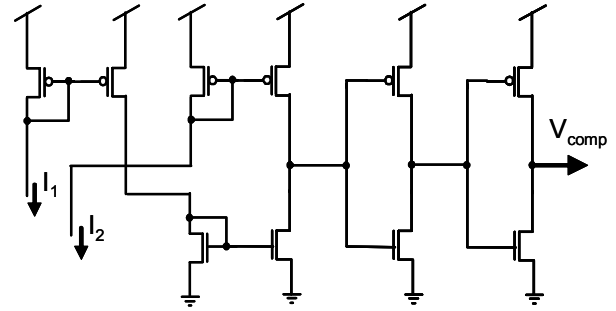

Figure 6: Current-mode comparator.

Each synaptic weight is implemented by means of a tunable active current mirror. A tunable active current mirror, based on a customized version of the circuit topology [13], is also used to clamp each of the three output buses. As a result, photocurrents as small as leakage currents can be readout with acceptable delays. In Figure 7, C_{bus} represents the large capacitance of an output bus capacitance whereas I_{in} refers to the total current flowing into the output bus. For $V_{G1} = V_{G2}$, the output node, labeled N, is clamped to V_{clamp} and $I_{in} \approx I_{out}$ provided that $M1$ and $M2$ are matched. To make the active current mirror tunable, choose $V_{G1} \neq V_{G2}$ and size transistors $M1$ and $M2$ so that they operate in the subthreshold region for the entire range of variation of the output bus photocurrent. Under these assumptions, $M1$ and $M2$ drain currents can be expressed [13] as:

$$I_{in} = I_{o1} e^{(V_{G1} - V_s)/nU_t}$$

$$I_{out} = I_{o2} e^{(V_{G2} - V_s)/nU_t}$$

where $U_t = kT / q$; V_s is the voltage at the source of both transistors $M1$ and $M2$. If both transistors are properly matched, then $I_{o1} \approx I_{o2}$ and the active current mirror gain G_i will be given by:

$$G_i = \frac{I_{out}}{I_{in}} = e^{(V_{G2} - V_{G1})/nU_t}$$

The output current I_{out} can thus be controlled exponentially by tuning $V_{G2} - V_{G1}$. Therefore, the current can be amplified or attenuated over an ultra-wide dynamic range. If properly compensated, the active input current mirror, shown in Figure

6, remains stable [13] even for an arbitrary small input photocurrent I_{in}. However, as photocurrent levels approach the pA range, the current mirror becomes increasingly slower. For C_{bus} = 1pF and I_{out} = 10pA, careful design led to a time constant of 100ns. As seen, a tunable active current mirror provides a mean for setting the synaptic weights and in turn for compensating for mismatch associated errors.
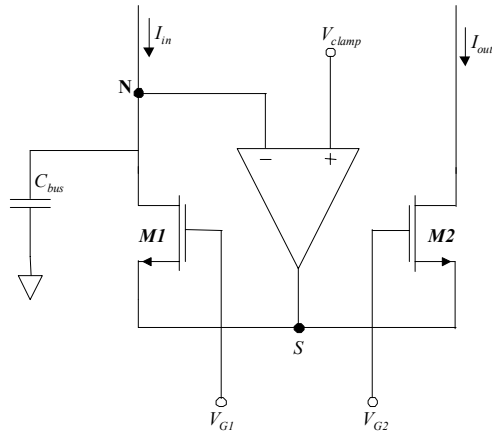

Figure 7: Active input current mirror topology.

IV. CONCLUSION

In this paper, we propose a method to detect skin color that is suitable for VLSI implementation. The skin detector uses an MLP with three inputs, one hidden layer and one output neuron. To simplify the VLSI implementation, we use a saturating linear activation function. Each pixel is represented by its RGB color components, which are used as inputs by the neural network. A current-mode fully programmable VLSI implementation is proposed to achieve skin detection processing on-read-out. The proposed skin detector offers a good trade-off between skin detection performance and implementation complexity.

V. ACKNOWLEDGMENTS

VI. REFERENCES

[1] R. Kjeldsen and J. Kender, "Finding skin in color images," *Proc. Conf. Automatic Face and Gesture Recognition*, pp. 379-384, 1996.

[2] J. Yang, W. Lu and A. Waibel, "Skin-color modeling and adaptation," *Proc. ACCV'98*, vol. II, pp. 687-694, 1998.

[3] D. Chai and K. N. Ngan, "Face segmentation using skin color map in videophone applications," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 9, no. 4, pp. 551-564, 1999.

[4] S. L. Phung, A. Bouzerdoum and D. Chai, "Skin segmentation using color pixel classification: analysis and comparison," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, pp. 148-154, Jan. 2005.

[5] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 274-280, 1999.

[6] D. Chai, S. L. Phung and A. Bouzerdoum, "Face localization based on color and shape information in a neural network approach," *International Conference on Information, Communications and Signal Processing*, Singapore, Oct. 2001.

[7] F. Perez and C. Koch, "Towards color image segmentation in analog VLSI: algorithms and hardware," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 17-42, 1994.

[8] R. Etienne-Cummings, P. Pouliquen and A. Lewis, "Color segmentation, histogramming and pattern matching chip", *Proc. IEEE ISCAS 2002*, Phoenix, Arizona, USA, pp.320-323, May 2002.

[9] J. Zurada, *Introduction to Artificial Neural Systems*, PWS publishing company, 1992.

[10] F. Ismail and T. Fiez, *Analog VLSI Signal and Information Processing*, McGraw Hill, 1994.

[11] E. Vittoz, "Analog VLSI signal processing: why, where and how?" *Analog Integrated Circuits and Signal Processing*, pp. 27-44, 1994.

[12] F. Boussaïd, D. Chai and A. Bouzerdoum, "On-chip skin color detection using a triple-well CMOS process," *Proceedings of SPIE: Microelectronics: Design, Technology, and Packaging*, vol. 5274, pp. 206-214, 2004.

[13] T. Serrano-Gotarredona, B. Linares-Barranco and A. G. Andreou, "Very wide range tunable CMOS/bipolar current mirrors with voltage clamped input," *IEEE Trans. on Circuits and Systems-I,* vol.46, no.11, pp.1398-1407, 1999.