Self-timed vertacolor dichromatic vision sensor for low power pattern detection

R. Berner, P. Lichtsteiner, T. Delbruck
Inst. of Neuroinformatics, UZH-ETH Zurich
dollbrain.ini.uzh.ch

***Abstract*—This paper proposes a simple focal plane pattern detector architecture using a novel pixel sensor based on the dichromatic vertacolor structure. Additionally, the sensor transfers dichromatic intensity values using a self-timed time-to-first-spike scheme, which provides high dynamic range imaging. The intensity information is transmitted using the address event representation protocol. The spectral information is sampled automatically at each intensity reading in a ratioed way that maintains high dynamic range. A test chip consisting of 20 pixels has been fabricated in 1.5 um 2P 2M CMOS and characterized. The combined pattern detector/ imager core consumes 45 uA at 5 V supply voltage.**

I. Introduction

Custom digital and mixed signal face detection circuits have been developed which allow for skin color [1] or face detection [2]. These systems are highly efficient, but because they partition their operation into image sensor followed by analog or digital processor, still require system level power consumption of at least 100 mW, making it impossible to run them continuously under battery power. It may be desirable to burn full power only when the presence of a human desiring interaction is detected.

The high power consumption of traditional vision systems is partly due to their repetitive processing of highly redundant input. Avoiding the readout and post processing of every pixel by doing necessary processing directly on chip at the pixel-level can reduce the power consumption significantly. This approach is feasible for basic pattern recognition especially if the detection of a pattern can wake up more power-hungry but more reliable post processing when it is likely to be needed.

We propose an architecture that combines focal plane face pattern detection with dichromatic imaging capability that can be used for more sophisticated post processing after wake up. The face detection is based on the fact that skin has high reflectance in the near IR and that the intensity distribution coming from a face has prominent dips at the eyes.

To extract intensity and chromatic information in a pixel, we exploit the fact that photon absorption length in silicon is strongly wavelength dependent. This allows us to build a dichromatic photo-sensor in standard CMOS technology. Using wavelength-dependent absorption length has been proposed in 1987 [3], but employing it for color imaging requires special process steps to achieve sufficient image quality [4].

We present here the architecture, implementation and measurement results of a dichromatic time-to-first-spike imager and a simple pattern detector test chip. The pattern detection is neither size nor shift invariant but the presented architecture is built so that it can be extended to achieve shift invariance by parallel implementation of pattern detector units. The proposed pattern detector combines basic face features and will in the future be extended to a more realistic face detector.

II. Pattern Detector Architecture

Our face pattern detector detects a face blob that is redder than its surrounding and the presence of “eyes” that are dark pixels around the upper middle of the face.

Angelopoulou showed that skin strongly reflects light in the near IR independent of skin color (race) [5]. Therefore the dichromatic pixel proposed here, which has the ability to discriminate bluish and IR spectral characteristics, should serve as a decent indicator of skin color.

Viola and Jones showed that the eyes are the most prominent features of a face in a monochromatic image [6]. More specifically, their face detection algorithm uses a cascade of simple rectangle filters which are selected and trained by an AdaBoost learning algorithm. The first two (and therefore most important) filters illustrate the fact that the eyes are shadowed and appear darker than the cheeks and the nose due to their position in the skull. This means that the average of the pixels representing the eyes is darker than the average of pixels representing the cheeks or the nose.

Fig. 1 shows a schematic overview of the face pattern detector for one possible face position. The detector assumes the presence of a face if eye (E) pixel 2 is darker than pixel 3 and 7, eye (E) pixel 4 is darker than pixel 3 and 9 and the average color of face pixels (R pixels) is redder than the surrounding (B pixels).

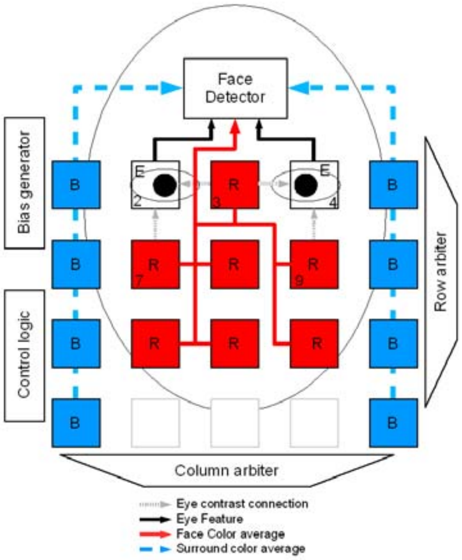

Fig. 1 Architecture of the face detector.

III. IMPLEMENTATION

This section describes the vertacolor structure, the pixel architecture, the face detector and the imaging capability.

A. Vertacolor structure

The pixel sensor uses the simplest vertacolor structure that can be built in every standard CMOS process. The vertacolor structure consists of two stacked photodiodes formed by active-well and well-substrate diodes. The spectral response of the U diode current peaks in the green and the L diode current peaks in the near IR.

B. Pixel Architecture

The pixel sensor is based on the same structure as the dichromatic spectral measurement circuit proposed by Fasnacht and Delbruck [7] but utilizes a different sampling method to achieve a time encoding of the intensity and a voltage encoding of the ratio of U and L photocurrents.

The pixel sensor structure consists of the two vertacolor photodiodes L and U and two switches driven by control signal φ (Fig. 2). The intensity and chromatic information is sampled in two phases. In the first phase both photodiode capacitances are charged to reference voltages $V_A=V_{refL}$ and $V_C=V_{refH}$. In phase 2 the switches are opened and the photocurrents I_{phL} and I_{phU} discharge the parasitic photodiode capacitances C_L and C_U:

$$\Delta V_C = -\frac{I_{phL}}{C_L}\Delta t$$

$$\Delta V_A = \Delta V_C + \frac{I_{phU}}{C_U}\Delta t$$

Phase 2 ends by sampling V_A when V_C reaches an adjustable V_{Clo}. By measuring V_A and the time Δt to reach V_{Clo}, we can calculate $I_{phL}\cdot K_L$ and $I_{phU}\cdot K_U$, where K denotes a constant factor predominantly determined by the photodiode capacitance. However, for the face detection it is enough to determine if the face area has more spectral power in the red than the surrounding area and for that it is enough to measure V_A, because:

$$\Delta V_A = \Delta V_C \left(1 - \frac{I_{phU}}{I_{phL}}\frac{C_L}{C_U}\right) \text{where } \Delta V_C = V_{Clo} - V_{refH}$$

and therefore the higher the voltage V_A at the sampling time, the bluer the scene, independent of intensity.

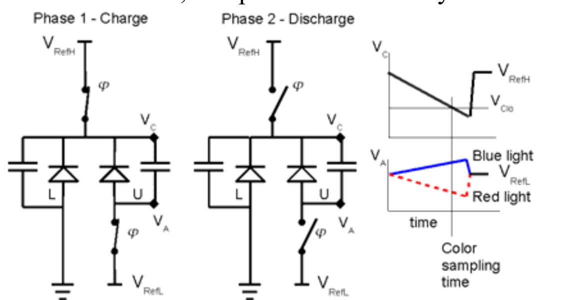

Fig. 2 The two phases of the pixel operation. Phase 1 recharges nodes V_C and V_A to reference values. In phase 2 both photodiodes are discharged by the photocurrents. V_A depends on the ratio of the photocurrents and therefore on the spectral content of the light.

The pixel sensor is embedded in a communication and control structure (Fig. 3). The sampling phases are scheduled by a state machine. After all the pixels complete phase 1 (determined by a global wired OR signal *BUSY*), phase 2 is started simultaneously in all pixels. Pixels then wait for the internal signal V_c to go below threshold V_{Clo} to conclude phase 2. Upon conclusion of phase two, the state machine signals via Address-Event-Representation (AER) communication to the periphery, and the pixel enters phase 1 again to prepare for the next integration cycle. The pixel circuit (Fig. 3) includes a sample and hold circuit for V_A and AER interfacing circuits.

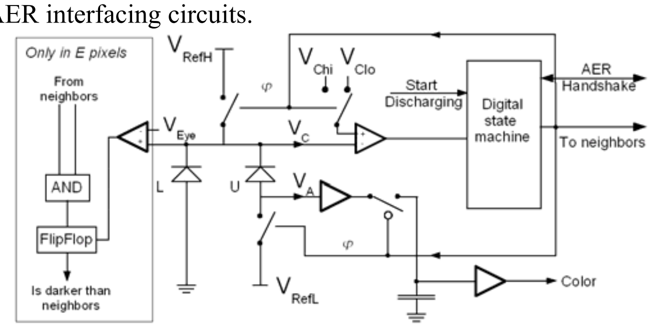

Fig. 3 Pixel architecture, including control state machine, sample and hold for the color value and the circuits for the computation of the eye feature (section C). The pixel consists of two photodiodes, two comparators, two source-followers, a capacitor and some digital control logic.

C. The Pattern Detector

The pattern detector unit is built by combining detection of three features, face/surround color contrast, left eye and right eye luminance contrast. Color contrast is computed by capacitive averaging of color voltages—one averaging circuit each for surround and face color. The average face color is compared to the average surround color by a simple opamp comparator and the comparator decision is latched on falling edge of *BUSY*.

When V_C of an E-pixel (Fig. 1) reaches an adjustable threshold $V_{eye} > V_{Clo}$ it samples the state (still integrating or finished integrating) of neighboring pixels. If the neighbors have already finished discharging, they are brighter and the eye is flagged as detected. V_{eye} sets the necessary contrast for the presence of an eye.

The three face features (color, left eye, right eye) are the input to an AND gate whose binary output indicates presence of the pattern.

D. Time-to-first-spike imager

The sensor's imaging capability is provided by extending the control state machine and adding AER interfacing circuits [8] to implement an enhanced time-to-first-spike encoding [9]. At the start of integration, a spike with a special address (*frame start address*) is emitted. As soon as each pixel reaches V_{Clo}, it emits a spike by putting its address on the bus. At this moment the color voltage is output as an analog signal.

IV. Measurement Results

A test chip with 20 pixels and one pattern detector unit has been fabricated in a 1.5um 2-metal 2-poly process. Fig. 4 shows a die micrograph.

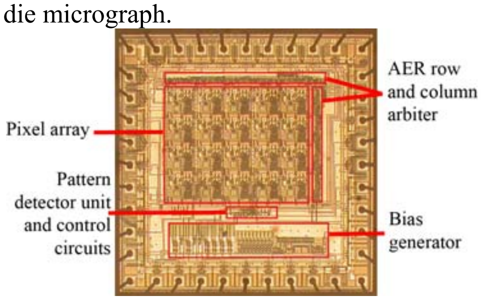

Fig. 4 Die micrograph. Die size is 2.2 by 2.2 mm.

The chip includes a bias generator for fixed bias currents [10]. For testing and characterization, the chip is placed on a PCB with a SiLabs C8051F320 USB1.1 transceiver/microcontroller and an Analog Devices AD5391 16-channel DAC for generating reference voltages and the possibility to override internal bias voltages. The system is interfaced to the jAER software [11].

Fig. 5 shows the average integration time of all the pixels over more than 3 decades of irradiance. The measurements were conducted with an incandescent light source and Kodak Wratten neutral density filters (due to the IR component in the spectrum of the light source we measured and corrected the attenuation factor of the filters). The plot shows that the integration time is nearly inverse proportional to irradiance over more than 3 decades. At very high irradiance, integration time saturates due to finite circuit speed, at very low irradiance integration time saturates due to dark current. The plot also shows that the color value is varying about 5%. The shift towards red with decreasing intensity can be attributed to the fact that the neutral density filters are less effective in the near IR.

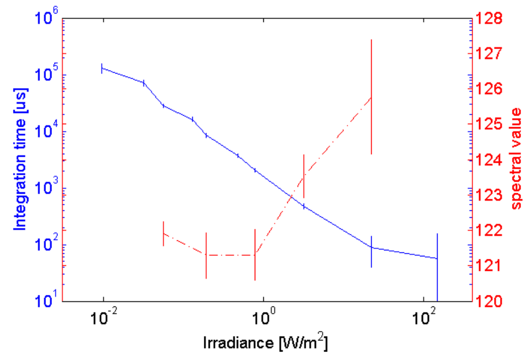

Fig. 5 Integration time (blue, solid) and 8 bit color value (dashed, red) *vs.* irradiance.

Fig. 6 displays the 8 bit encoding of monochromatic light for all 20 pixels. The encoding of the color saturates below 500nm and above 750nm. The blue limit could not be fully explored because we used an incandescent light source. However between 500nm and 750nm we get a good resolution encouraging further exploration.

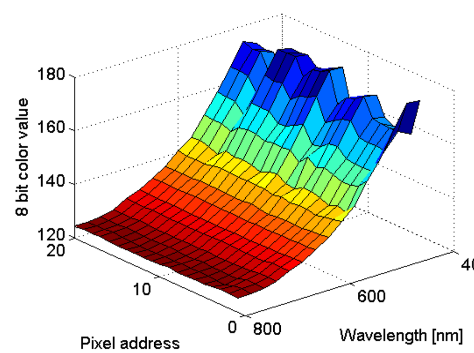

Fig. 6 Spectral response of all 20 pixels vs wavelength.

Fig. 7 shows the integration time versus wavelength for all pixels normalized by irradiance. The data shows a strong fixed pattern which was observed in all dies. Fig. 8 shows measurements of the pattern detector circuits. The left plot shows the output of the capacitive averaging circuits when the sensor is aimed at a computer monitor displaying a stimulus similar to the one on the right of Fig. 8, where either the hue of the "face" is held constant and the hue of the surround is stepped from 0 to 1 or vice versa. It can be seen that the average surround color voltage is higher for a given hue, which means that the detector is biased towards detecting a red blob even in uniform color stimulus. This is unwanted and has to be corrected for future implementations.

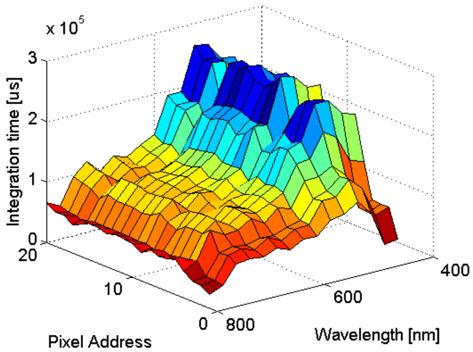

Fig. 7 Integration time for all 20 pixels vs wavelength. The integration time has been normalized by the irradiance measured with a Tektronix J17 photometer.

The right plot of Fig. 8 shows the maximum V_{eye} threshold voltage, at which the sensor still can detect the eye feature for a given luminance ratio between E pixel and its neighbors. The measurements demonstrate that the pattern detector circuit works. Its performance can be improved by spending more area on the comparator and adding a threshold for the color feature.

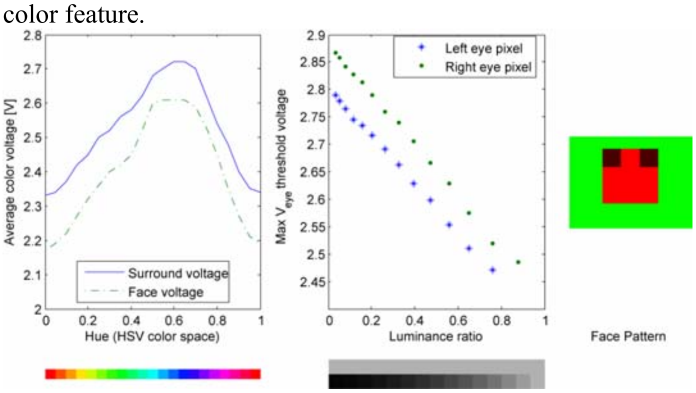

Fig. 8 Pattern detector circuit response. Left plot shows the output of the capacitive averaging circuits. Right plot shows the maximum V_{eye} threshold voltage at which the sensor still can detect an eye for the given luminance ratio between eye pixel and surround. Reference voltages are V_{Chi}=2.985 and V_{Clo}=2.463 V. Right pattern is the stimulus "face" pattern.

V. Conclusion And Outlook

This paper describes a novel approach for color vision, based on the wavelength separation capabilities of vanilla silicon. Measurements demonstrate the feasibility of the pattern detection approach. Future work will focus on improving pixel speed as well as improving and extending the pattern detector circuit.

1. DollBrain1 vision sensor specifications

Functionality	Asynchronous time-to-first-spike imager, with analog dichromatic spectral value output
Pixel size μm (lambda) *Fill factor (%)*	244x256 (305x320) 4.06% (PD area 2540μm^2)
Fabrication process	2M 2P 1.5um
Pixel complexity	99 transistors (11 analog), 3 capacitors
Array size	4x5
Interface *Dynamic Range*	5-bit word-parallel AER 70dB
Power (@5V)	Analog: 30μA Digital @300 Hz frame rate: 14μA Pads: 550μA (80% analog output pads for characterization)

Acknowledgements

This project was supported by the Swiss National Science Fund grant 200021-112354 / 1; the project website is *dollbrain.ini.uzh.ch*. We acknowledge Pit Gebber's prior work in developing parts of the USB interface.

References

[1] R. Etienne-Cummings, P. Pouliquen, and M. A. Lewis, "A vision chip for color segmentation and pattern matching," *Eurasip Journal on Applied Signal Processing*, vol. 2003, pp. 703-712, 2003.

[2] Y. Hori and T. Kuroda, "A 0.79-mm(2) 29-mW real-time face detection core," *IEEE Journal of Solid-State Circuits*, vol. 42, pp. 790-797, 2007.

[3] R. F. Wolffenbuttel, "Color filters integrated with the detector in silicon," *IEEE Electron Device Letters*, vol. EDL-8, pp. 13-15, 1987.

[4] G. Gilder, *The Silicon Eye: How a Silicon Valley Company Aims to Make All Current Computers, Cameras, and Cell Phones Obsolete*: W. W. Norton & Company, 2005.

[5] E. Angelopoulou, "Understanding the color of human skin," 2001 SPIE Conference on Human Vision and Electronic Imaging, pp. 243-251, 2001.

[6] P. Viola and M. Jones, "Robust real time face detection," Eighth IEEE Conference on Computer Vision, pp. 747-747, 2001.

[7] D. Fasnacht and T. Delbruck, "Dichromatic spectral measurement circuit in vanilla CMOS," IEEE International Symposium on Circuits and Systems (ISCAS 2007), New Orleans, pp. 3091-3094, 2007.

[8] K. A. Boahen, "A burst-mode word-serial address-event link-I transmitter design," *IEEE Transactions on Circuits and Systems I-Regular Papers*, vol. 51, pp. 1269-1280, 2004.

[9] X. Qi, X. Guo, and J. Harris, "A Time-to-first-spike CMOS imager," 2004 International Circuits and Systems Conference (ISCAS2004), Vancouver, Canada, pp. 824-827, 2004.

[10] T. Delbruck and A. van Schaik, "Bias current generators with wide dynamic range," *Analog Integrated Circuits and Signal Processing*, vol. 43, pp. 247-268, 2005.

[11] Available: http://jaer.wiki.sourceforge.net