A passive infrared gesture recognition system

T. David Binnie, A. F. Armitage
Edinburgh Napier University
Edinburgh. UK.
td.binnie@napier.ac.uk

Piotr Wojtczuk
Pyreos Ltd
Edinburgh. UK
piotr.wojtczuk@pyreos.com

Abstract **A device for a low-to-intermediate level of gesture recognition which uses a passive thermal-infrared (PIR) sensor array is described. The detection system which discriminates between a small number of simple dynamic gestures, such as 'hand swiping' in different directions and at varying velocities. The technology is low powered, in terms of energy consumption and computational power. The sensor enables a gesture sensing input, for mobile device application areas, where hands-free operation can be useful and where low-power consumption is essential.**

Keywords—gesture detection; infrared sensor array; low power sensor;

I. Introduction

In human interaction, gesturing is a natural and constituent part of communication. Hand gestures are second in importance only to facial expression. In human-machine interfaces, along with voice recognition, gesture based interaction has becoming increasingly important. Such interaction covers a broad range of modalities from the simple switching on of a hand drier by proximity sensing, to high precision three dimensional tracking for a full-body gaming experience. For many applications the technology that is used is determined by the level of information that is required from the gesture.[1] This can range from a simple replacement for the function of a non-contact on/off switch, to a method of selecting from drop-down menus, or onto a highly sophisticated system that recognizes one of many possible human gestures. In this paper the device is briefly described but the primary focus is on an evaluation of the practical detection of human hand gestures rather than the physical specification and test bench performance of the sensor.

II. Gesture Characterisation

Human and motions can be classified into static and dynamic gestures. Byung-Woo Min et al.[2] Dynamic gestures are generally more intuitive for the user, as the direction of motion or its shape reflect the resulting interpretation and can mimic the required response. For example, an up or down motion can represent increase or decrease of room lighting brightness. These gestures are commonly universal and as such do not rely on cultural or knowledge of sign-language. Static gestures are fixed hand or arm postures that correspond to a sign. Generally these are more difficult tack for machine based detection, as each object of interest has to be identified by its relations with other objects, e.g. an angle of an arm to body has to be evaluated by identifying two objects. Additionally, tolerances have to be taken into consideration as each user may have a different perception of their own body position.

This study deals with simple dynamic gestures e.g. simple left-right and up-down swiping; static gestures are not considered.

To define the detectable hand motion(s) which will act as the input to the device, a participant study was undertaken. An evaluation of hand trajectories were produced using images from a high frame rate camera with skin colour hand feature extraction and frame-by-frame location tracking. The results provided the practical limits of velocities of human hand motion. The analysis included path direction. The techniques used here are implementations of standard image processing techniques and statistical analysis of resulting traces.

A camera mounted five metres vertically above the hand motion captures the horizontal plane, which corresponds to the plane of gesture. This is the primary camera from which the profile of the hand motion can be extracted. A second horizontally mounted camera captures any component of movement perpendicular to the gesture plane. The height measurement is also used to correct any parallax error in the video stream of the gesture. Markers of known dimensions provide the means spatial calibration within the image frame.

The image analysis process to extract hand positions within each video frame was based on that described by O'Hagan et. al.[3] The process consists of: Background Subtraction; Filtering – HSL, range to match skin tone - Binarisation; Centroiding ('blob' detection). In practice, up to half of the gestures presented by participants were rejected because they could not be clearly tracked in the video analysis. This mostly happened because the hand motion was too close to a frame edge, or because of an interruption to tracking.

The directional deviations from the main axes, particularly in case of the fast gestures is explained by the participants being mainly right-handed. Slower hand movements are better controlled and more linear, while fast movements displayed a pivoting motion around the elbow joint. The arching motion, observed in medium speed hand motion is shown for left-right and up-down swipes in Figs. 1a and 1b.

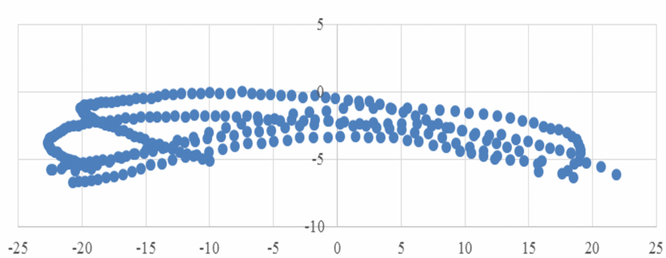

Fig. 1a. X-Y trace of repeated Left-Right Hand Motion extracted from Video Footage. Axes dimensions are centimetres.

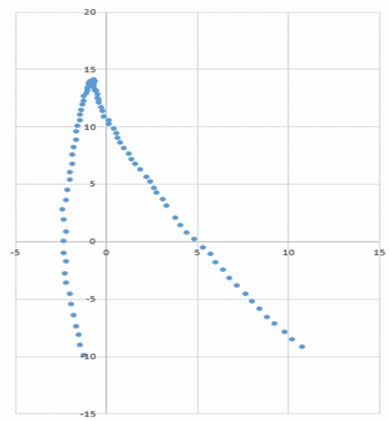

Fig. 1b. X-Y trace of Up-Down Hand Motion extracted from Video Footage. Axes dimensions are centimetres.

III. INFRARED SENSOR STRUCTURE

For the Gesture Capture Device a custom-packaged pyroelectric infrared sensor was designed and then fabricated through Pyreos Ltd.[4] The sensor die is a silicon MEMS structure with the elements laid out on a thin film membrane. Several types of arrays were made available by Pyreos for the purpose of this study. Results are presented for a 2x2 array which can be used in different modes. The infrared radiation is detected by a coated PZT layer which has a high pyroelectric sensitivity and a relatively high frequency response.

CMOS wafer processing techniques were used for the fabrication process, coupled with an additional MEMS etching process. The wafer is overlaid with layers of metal and pyroelectric material. After the sensor stack is deposited on the wafer, the backside is partially etched. This creates a suspended membrane under the sensor elements, with a frame around each die. The physical characteristics of the membrane can be tuned to affect the thermal performance of the elements. PZT is self-polarised to form a pyroelectric while being grown on a silicon substrate heated to 450 °C. Pyroelectric coefficients of $2x10^{-4}$ C/m^2K are achieved, with a dielectric constant of 300 and dielectric loss of 0.01. The process for obtaining polarised PZT during pyroelectric deposition was developed by Bruchhaus et al [5] and patented at Siemens A.G. before Pyreos Ltd obtained the intellectual property.

The fabrication process results in an infrared array detector with a relatively high frequency response (τ) of 10-15 ms, as opposed to the 150 ms found in standard PZT motion detectors. The response time is an important device criteria as the detector has to be capable tracking of fast hand movements of up to 10 m/s.

IV. SENSOR STRUCTURE

The array used in the study was a 2x2 pyroelectric sensor array mounted in a package with a custom designed readout ASIC. The die size was 2 mm x 2 mm. The individual 200 µm x 200 µm square elements were symmetrically placed around the centre of the die. Results were assessed with the four sensing elements connected in several different configurations. The cross-coupled differential mode provided the best response for the gesture detection application. The package was fitted with a lid which included a Germanium window The Germanium window was used to block visible and near-IR spectral electromagnetic radiation, while allowing transmission of the radiation in the 8–14 µm wavelength range. A 700 µm diameter pinhole in a metal film directly below the germanium window was used to form a two-dimensional spatially related signal from a field of view of 50 degrees. [6] For the intended application area, the inclusion of a germanium lens would improve performance. The limited light gathering properties of the pinhole limit the sensor response.

V. DETECTION ALGORITHM

The pyroelectric signal of a warm hand moving in front of the detector typically consists of a positive peak as the temperature changes from the cool background to the warm target. This is followed by a trough as the warm target moves out of the field of view, and the cool background is now imaged. Fig. 2 shows this signal from a single swipe from each of the four elements. The direction of motion is obtained by finding the sequence of the peaks, denoting the sequence in which the target moved in front of the elements. The peak detection is shown for the 2x2 sensor connected as two differential pairs. In Fig. 3, a target has moved first in one direction then the other. The motion is parallel with the axis on one pair of elements: these (lower trace) produce a much higher signal than the pair that is nearly at right angles to the direction of motion (upper trace). A statistical analysis of these signals produce the direction of the hand motion. The time difference between the peaks and troughs give the speed of the hand. Software filters were applied to reduce offset and noise.

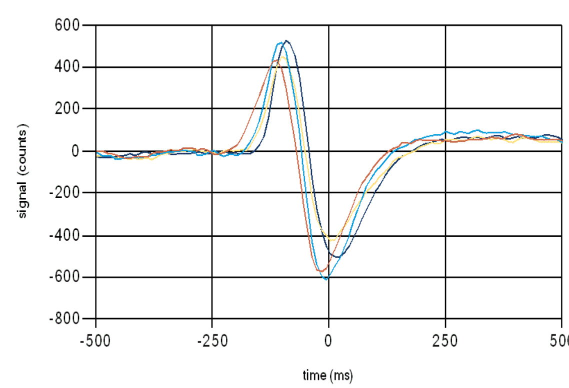

Fig. 2. The signal responses against time to a single hand swipe from each of the four sensing elements of the 2x2 sensor array. Amplitude scaling is arbitrary; time scale is milliseconds.

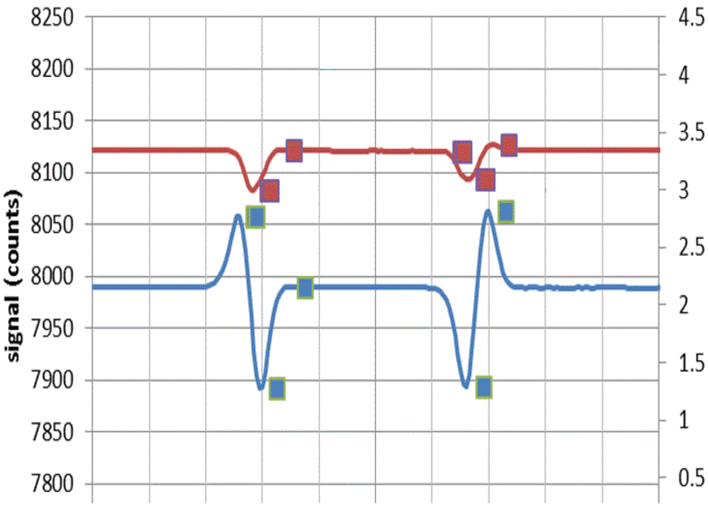

Fig. 3. Example signal response to a single hand swipe from each of the two cross-coupled differential pairs in a 2x2 sensor array. The squares indicate the detected signal peaks and troughs from each of the pixel pairs. The horizontal axis scaling is 200 samples per tick and the vertical axis scaling is the signal amplitude in arbitrary units.

VI. Sensor Response

For the 2x2 array, both noise and responsivity were uniform; there is little practical difference in where the elements are placed in these arrays. This contrasts with, for example, a 16 element (4x4) array where there are central, edge and corner elements, and each element will have a slightly different thermal environment. The frequency and sensitivity of each of the sensing elements is highly dependent on the thermal environment. The frequency response for a 2x2 array with a range element sizes and spacings was explored. The elements on the smallest gap layout have the poorest low frequency performance, as in that region the thermal crosstalk is most prevalent.

VII. Gesture Detection

The study was performed with the two video cameras synchronised to record gestures above the infrared sensor on a workbench with. In total there were 2479 valid gesture samples. The participants were asked to produce a swiping gesture at 'normal' 'slow' and 'fast' speeds within a height range of 10-20 cm above the horizontal sensor plane. Although these are subjective terms, the objective was to determine the range of gestures that would be encountered in practice. 898 samples were performed at moderate speeds. The middle values of 103 to 157 cm/s were used for the optimisation of the algorithms. The scenario is intended be typical of gestures that might be made over a mobile device. The height at which the gesture was made was measured but deliberately not controlled.

A natural tendency for faster gestures to be made higher above the detector than moderate or slow gestures was observed. Generally, the left-right gesture is bio-mechanically easier than the up-down gesture; the left-right gestures were generally made closer to the detector than the up-down gestures. 227 gestures were used to optimise the detection algorithms. This involved determining the settings for the peak threshold, minimum detection interval and a number of other parameters

Left-right gestures were detected more reliably than the up-down gestures. This was attributed to the relative ease of making and repeating a linear left-right gesture. A typical up-down gesture was more curved, and still has strong components in the left-right direction, making it harder to identify. Further work is needed to compensate for this effect. It is also likely that he difference in this gesture as performed by left and right handed people will need to be accounted for. It was difficult for the participants to produce gestures at precise speeds, but because they were asked to perform gestures at normal, as well as fast and slow speeds, it was possible to develop an understanding of how this affects the accuracy of gesture determination. The discrimination accuracy drops at higher speeds as gestures variability in motion path increases. When participants were asked to perform gestures at a normal speed, the measured 'normal' range from 103 to 157 cm/s falls within the range of speeds where the performance is best. In this range the accuracy of detection of left-right and right-left swipes was 76%. 20% of the swipes produced an indeterminate result and 4% produced an error.

VIII. Conclusion

A simple gesture detector based on a 2x2 passive infrared sensor array which discriminates a range of human hand motions has been demonstrated. Although it was found that reliability of the gesture detection depends on hand speed, these initial findings revealed that the detector was viable in non-critical situations. It was further found that accurate detection rates were very dependent on the human characteristic; some operators could reliably produce accuracy measurements above 95%. This implies that with the inclusion of an adaptive algorithm, far greater detection accuracies could be achieved.

IX. Acknowledgment

The support of Pyreos Ltd in fabricating and packaging the sensor samples is gratefully acknowledged. This work was supported by the UK Engineering and Physical Sciences Research Council and Pyreos Ltd under EPSRC CASE Award EP/I50137.

X. References

[1] S. Mitra and T. Acharya, "Gesture recognition: A survey," IEEE Trans. Syst., Man, Cybern, vol. 37, pp. 311-324, 2007.

[2] B.-W. Min, H.-S. Yoon, J. Soh, Y.-M. Yang and T. Ejima, "Hand gesture recognition using hidden Markov models," in Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on, 1997.

[3] R. O'Hagan, A. Zelinsky and S. Rougeaux, "Visual gesture interfaces for virtual environments," Interacting with Computers, vol. 14(3), pp. 231-250, 2002. *(references)*

[4] Pyreos Ltd. Scottish Microelectronic Centre. Edinburgh EH9 3FF

[5] R. Bruchhaus, D. Pitzer, R. Primig, W. Wersing and Y. Xu, "Deposition of self-polarized pzt films by planar multi-target sputtering," Integrated Ferroelectrics, vol. 14, no. 1-4, p. 141–149, 19. 1997.

[6] P. Wojtczuk, T D Binnie and C. Geibler."Sensor System for detecting movement of an Infrared Light Source" Patent No. WO 2014029698 A1. 2014.