Air Gesture Control Using 5-Pixel Light Sensor

Zoran Zivkovic
NXP Semiconductors, The Netherlands

***Abstract*--A minimal low-cost solution for touch-less gesture control is presented. An integrated sensor containing 5 photo diodes is introduced. As commonly used for infra-red based proximity detection the system operated using modulated near infra red light. However, the 5 sensors setup enables effective way for detecting also position of the close-by object and direction of the object movements. A basic but rich set of sweeping and selection gestures can be recognized using the system as demonstrated here. To minimize the costs, the system does not include any additional optical elements but an innovative approach is used where the directionality of the movements is estimated by using shadows of the existing enclosure of the chip.**

I. INTRODUCTION

Touch-free air-gesture controlled applications are becoming increasingly interesting for the consumers. Instead of swiping your fingers over the display and making it dirty, you can simply swipe the air, leaving your display clean.

Recently, the low cost infra red (IR) light depth cameras were introduced [1][2] offering rich full body touch-free interaction possibilities. On the other hand, for many applications, e.g. tablets and mobile phones, only a basic set of swipe and touch gestures would be enough for sufficiently rich gesture based control. An effective low cost system is presented here containing only 5 light sensors. The system can be regarded as a 5-pixel camera. Modulated IR LED is used to generate illumination and detect only close-by objects, similar to the common proximity detection components present in many mobile phones. Various solutions using multiple LEDs and separate detectors were previously presented. To the best of our knowledge this is the first system with a single LED and an integrated sensor containing multiple photo diodes.

II. AIR GESTURE AS DIRECTIONAL PROXIMITY DETECTION

The main blocks of the system are illustrated in Figure 1. Invisible IR light is detected by 5 integrated light sensors. Movement in front of the device will introduce different patterns on the different sensors. Different movements will introduce different motion patterns. A detector is used to recognize the different movements. The detector can be constructed using various machine learning techniques to learn different set of gestures. A set of gestures and experiments is presented later.

In order to be able to distinguish different movements, some "directional response" is required. The presented integrated sensors are very small and close to each other. The dimensions and positions of the sensors are depicted in Figure 2 (top). As result of the close placement, the light sensors would have very similar response and there would be little information about the movement direction in the sensor readings. Figure 2 (middle) shows the simulated spatial response of the central and one side sensor. The responses are indeed very similar. However, there is still a small difference on the side. This difference comes from the walls of the chip enclosure that are reasonably close to the sensors. The enclosure will cast different shadows on different sensors. These small effects can still be used for distinguishing different movements as it will be demonstrated. Using the existing shadows leads to a simple and highly effective solution for the air-gesture detection.

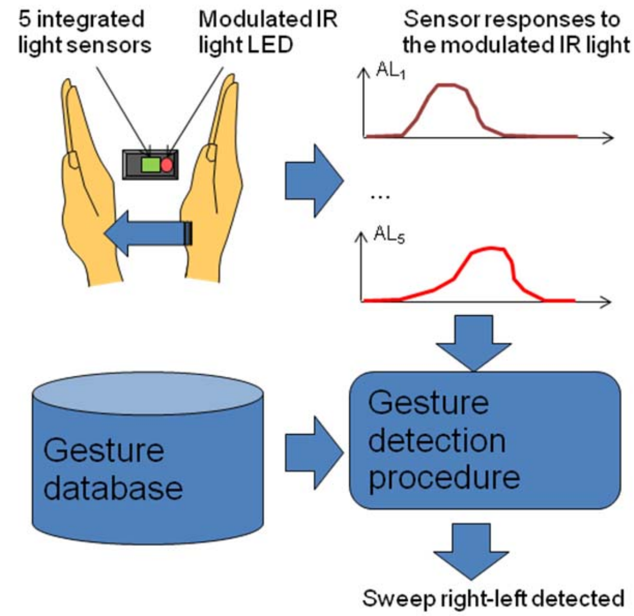

Fig. 1. Interactive air-gesture system using modulated IR light.

A more costly option is to add an optical element, e.g. a lens. This would ensure that different parts of the space project to different sensors differently, e.g. if the hand is on the left side only one of the sensors would have large response and a different one if the hand is on the right side. Figure 2 (bottom) illustrates the spatial response of the central and one side sensor when a pinhole optical element is installed in front of the sensors.

III. GESTURE DETECTION EXPERIMENTS

In order to detect different gestures, a detector is learned on a set of examples. Five basic gestures were considered for this experiment: push, sweep left, sweep right, sweep up, sweep down. For each gesture 600 examples were recorded. The movements were performed with different speeds, holding the hand in different configurations, using only fingers, etc. The recorded examples contain a diverse set of how different people would perform each of the gestures. The gestures were recorded for two cases, the sensor without any optical

978-1-4799-1291-9/14/$31.00 ©2014 IEEE

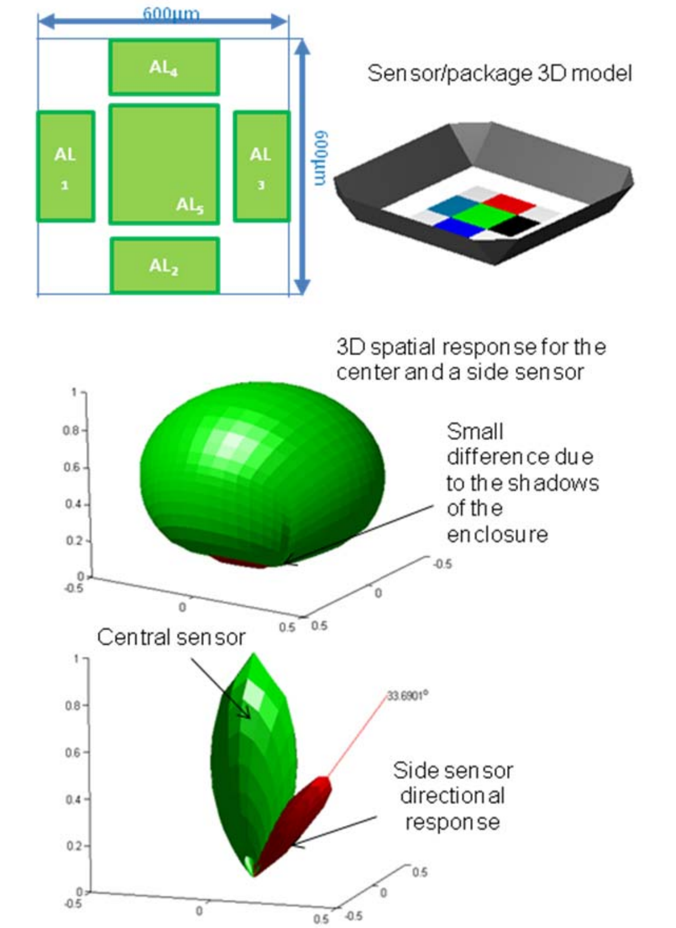

Fig. 2. Integrated light sensors lay out (top), 3D model including enclosure and spatial response of different sensors.

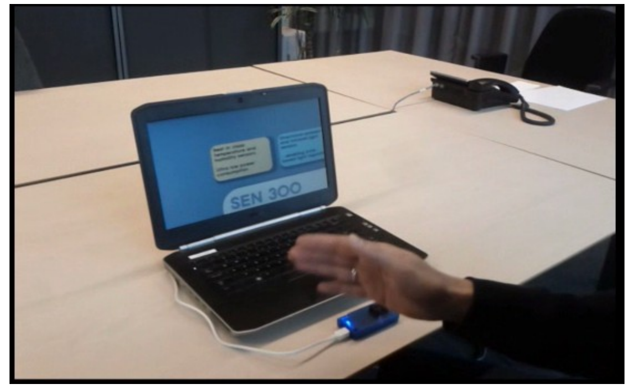

Fig. 3. A demo system. The sensor is in the demo box (blue box) ant it is connected to the PC. The detection is performed in the demo box and used to control PC applications.

Pinhole optic element			
	Push	Left	Right
Push	299	0	1
Left	0	300	0
Right	1	0	299

No optics			
	Push	Left	Right
Push	297	2	1
Left	1	299	0
Right	1	0	299

Pinhole optics (full gesture set)					
	Push	Left	Right	Top	Bottom
Push	298		1	1	
Left		300			
Right	1		298		1
Top	2	1		297	
Bottom	1		2		297

No optics (full gesture set)					
	Push	Left	Right	Top	Bottom
Push	290	1	1	5	3
Left	2	288	1	4	5
Right	5	2	280	6	7
Top	10	5	7	275	3
Bottom	9	11	13	5	262

Fig. 4. Recognition results

elements relying on the shadow effects of the chip enclosure and the sensor with a pinhole optical element.

The recorded data was randomly split in training and test sets of equal size. The training data is used to train a classifier. For classifier we used here the Ada-Boost [3]. The classifier is applied on the independent testing data set for performance evaluation.

A simpler set of gestures is selected first, containing just push, sweep left and right. The confusion matrices for this experiment are in Fig 4. For each gesture, number of correctly classified examples is presented together with numbers of classifications as other gestures. Very good accuracy is demonstrated. Without optical element the performance degrades slightly, but still very robust detection is achievable. For the full gesture set, very accurate detection is achieved with the optical element. Without the optics the number of misclassified gestures is rising. However, the results are encouraging. A reasonably robust detection is achievable without any optical elements.

IV. CONCLUSIONS

Effective system for air-gesture control is described. Five integrated light sensors are used for directional proximity detection. The direction of the movement is used to recognize various gestures. The experimental results demonstrate very robust detection. Adding optical element improves recognition since the signals depend more on the position of the object/hand. However, even without any optics the recognition can be reasonably good. The directional information in that case comes from already existing small shadow effects of the chip enclosure. Having no additional optical elements simplifies the design and in practice reduces the costs of the system significantly.

In the end it is also worth mentioning that in common case of display gesture control, the display itself can be used as the source of modulated light removing the need for the additional LED usually used in proximity detection, and even further reducing the system costs [4][5].

REFERENCES

[1] Microsoft Xbox Kinect, www.xbox.com/kinect
[2] Leap Motion, www.leapmotion.com
[3] Y. Freund, R.E. Schapire, “A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting”, Computational learning theory, 1995.
[4] A. D. Wilson, “Touch Light: An imaging touch screen and display for gesture-based interaction”, Proceedings of the 6th ACM International Conference on Multimodal interfaces, pp. 69 – 76, 2004.
[5] Z. Zivkovic, H. Groot de Hulze, “Mid-Air Interactive Display Using Modulated Display Light”, IEEE ICCE, 2013.