Random Access Sensors and their Implications for Optical Data Processing

Adrian Trenaman, David Barry, David Vernon[1].
Department of Computer Science, Maynooth College, Co. Kildare, Ireland

Telephone: 708 3847
Fax. 708 3848

trenaman@cs.may.ie, drbarry@cs.may.ie, dvernon@cs.may.ie

ABSTRACT

We present an investigation into the usefulness of a recent development in the field of optical sensors: a random access digital optical sensor. This new sensor exploits CMOS, instead of the traditional CCD technology, to reduce access individual pixel from an array of light-sensitive photosites in approx. 0.2μ s. This new development has important implications for real-time optical data processing which will be investigated in this paper. In particular, since image content is locally-specific, image interpretation (or analysis) is inherently random access; that is, relevant information is spatially localised and, more often that not, much of the data contained within an image is extraneous. For this reason, it is often not necessary to process an entire image, but instead, only meaningful regions should be considered. This fact is sometimes overlooked when designing optical data processing algorithms. With the advent of new random access sensors, the possibility of alternative analysis strategies arises. This paper first presents a specification of a new random access algorithms, which incorporate a pre-attentive and an attentive stage. Finally, we present results which indicate the time savings that can be achieved using a random access sensor coupled with a local contour-based segmentation and analysis random access algorithm.

Introduction

This paper concerns itself with a recent development in the field of optical scanners: A random access image sensor [1]. This sensor is an array of pixels where the intensity level at a pixel can be retrieved via X-Y addressing in an manner similar to random access memory. We present a model based on the pre-attentive/attentive model of human vision which in our opinion best utilises the sensor for a class for computer vision problems - namely tracking.

With the advent of random access sensors, the need has arisen to develop algorithms which best utilise this device, and further more to develop a frame work for producing such algorithms. One such approach, which has been borrowed from the human vision system, is the pre-attentive/attentive model of vision[2-4]. Regions of interest are located at the pre-attentive stage, and these regions are examined in more detail at the attentive stage. Due to the nature of the random access sensor, it is best utilised if only a small sub-set of the pixels on the sensor array are accessed while analysing any one image.

Firstly, a presentation of the pre-attentive/attentive model of human vision will be made. How "regions of interest" are determined is explained. A brief review of the psychology [5] behind this model is also presented. We are keen to point out that the human vision system is but one of many different vision systems to be found in nature, and not to be considered as nature's optimal solution. However, it does serve as a good example of a vision system capable of processing huge amounts of information in real time. We indicate the implications that attentive mechanisms (i.e. the mechanism that locates "region of interest") have on machine vision systems, particularly with respect to random access machine vision techniques. A frame work for the use of this model in conjunction with the random access sensor is then put forward. The paper then concludes by illustrating an example, that has been implemented, where the pre-attentive and attentive model of vision has been used in conjunction with the random access sensor to good effect.

The Psychology of Perception

One of the pioneers of investigations into the problem of processing vast quantities of visual information, Yarbus, demonstrated *scan-paths* [2] in Human vision. The sequence of fixation points of the observer's eye were analysed (see Figure 1). The location of these fixation points follow a path inspecting the salient features of the visual field. For example, when an observer is inspecting a face, the eyes and the mouth are critical points, and along the pathway these points are constantly revisited. This lead to research into attentional mechanisms, which will be addressed later.

---

[1] This work is funded by ESPRIT Project 20557 RAMAP

guided by pre-attentive mechanisms, working on global image information. In doing so, the pre-attentive mechanism effectively reduces the volume of data to be processed. This reduction of data-set is crucial with respect to high speed random access vision, where the aim is to achieve accurate results while minimising pixel accesses. By reducing the number of pixels required to analyse the image, we reduce acquisition time, and thus the image acquisition bottleneck present in some current systems. This section presents a worked example.

An example: Tracking a blob.

As an example to illustrate the advantages of using a random access camera coupled with the pre-attentive and attentive model of vision, a blob tracking algorithm was implemented. The implementation, and results are presented here.

Problem outline

The task of this vision system is quite straight forward: the system is required to locate the position of a circular icon in successive images. In essence it tracks the movement of the icon (or blob as it would appear) given a sequence of images. The iconic mark is a white circular mark that is placed against a black background. Although the problem does appear trivial it is a valid industrial machine vision problem, and the real goal is not to just solve the problem, but to solve it using the minimum of time. Given that CCIR standards (a 50Hz interleaved video signal) dictate that a vision system using the standard can only acquire 25 frame per second, then any image sequence processing rates that are greater that this must be consider a significant advance. This is especially true when tracking fast moving objects, such as electronic component with the use of SMD machines.

The Algorithm

The algorithm used to inspect each image is again quite straight forward. The pre-attentive and attentive model well defines the goals of each of the two stages of the image analysis process. The pre-attentive stage must locate regions of interest., while the attentive stage must inspect these regions of interest in more detail. Further more the attentive stage must decide if a region of interest represents the iconic mark, and if it does it must locate its position on the sensor:

```
FOR each image in the sequence DO
/*Pre-attentive stage. Goal: Located regions on the sensor that are likely to represent the blob. */
/*Approach: */
        IF first image
                THEN perform a binary search a long the Y-axis (fast) of the sensor until a 1-d edge above a certain threshold is located. Pass the co-ordinate of this region (pixel) of interest to the attentive stage.
                ELSE start the search for a 1-d edge along the Y-axis in both directions from the position of the blob computed from the previous image. Pass the co-ordinate of this region (pixel) of interest to the attentive stage.
                        IF this fails
                                THEN perform the same operation along the X-axis. Pass the co-ordinate of this region (pixel) of interest to the attentive stage.
                                        IF this in turn fails
                                                THEN return to the original binary search.
/*end pre-attentive stage.*/
/*Attentive stage. Goal:  Determine if the region of interest located at the pre-attentive stage is the blob, if it is extract the location of the blob.*/
/*Approach: */
        Follow the contour of the edge (see below) located at the pre-attentive stage to generate a 1-d representation of the blob.   /**This 1-d representation is called a boundary chain code (BCC). See [9]
        From the BCC compute the first and second moment of the blob. /** See [9]
        From these two moments compute the area and centroid of the blob. /** See [9]
        IF the area of the blob is  above a pre-selected threshold
                THEN select object as correct, and return the centroid to the pre-attentive stage
                ELSE  return to the pre-attentive stage to select a further region of interest.
/*end attentive stage.*/
```

Contour following

The edge magnitude and direction of the first pixel that is believed to be on the boundary of the blob are computed using a Sobel edge detector. From this pixel three candidate pixels are chosen in a direction that is normal to the first pixels edge direction, see Figure 4. The pixel then which is judged to be most likely the next pixel on the boundary is the chosen from the three candidate pixels. It is possible that at this stages none of these pixels are deemed to be on the boundary of the blob, in which case the search is widened. Contour is followed in the direction normal to edge choosing local maxima (see Figure 3) until the starting pixels revisited. A full description of this technique can be found in [15].

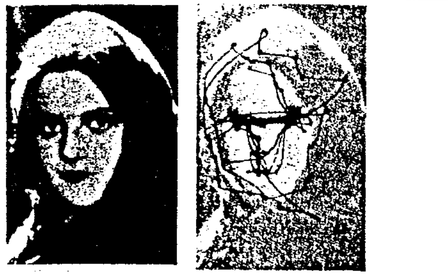

Figure 1: The figure on the right indicates the scan path that an observers eye follows when presented with the picture on the left.

Scene understanding, leading from object recognition, is a very complex task. The human vision system is, to the best of our knowledge, the most highly developed system for performing this task. Visual information passes from the retina to the central cortical neurons. Along this journey the information passes through many transformations. From a functional viewpoint, early visual information processing is characterised by two distinct modes [4,9]:

- Pre-attentive mode - The information is processed in a spatially parallel manner, without the need for attentional resources. This parallel execution is characterised by: i) complete independence of the number of elements; ii) an almost instantaneous execution; iii) the lack of sophisticated scrutiny; iv) a large visual field.
- Attentive mode - for more complex analysis the allocation of attentional resources to specific locations or objects is needed. A sequence of regions of interest is scrutinised by a small aperture of visual attention. The object can be consciously identified only within this attentional aperture.

The Pre-attentive Mode

The phenomenon by which objects that share a single feature not shared by the rest of the objects within that scene are detected is called *pop-out* - the common objects literally pop out of the picture. Treisman and Gormican [10] put forward the theory that features (such as colour and orientation) each form a dimension of mutually exclusive values.
Figure 2 illustrates this. In the far right picture, the T's and L's are both black and grey. Shape and grey level produce pop-out, but their conjunctions does not. Furthermore Treisman [11], showed that the absence of a feature is not always detected at the pre-attentive stage.

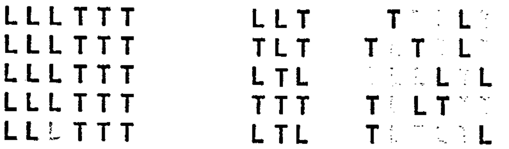

Figure 2: Three cases of regions composed of three different textures are shown. The partition of the couples of textures is quite evident in the left and middle region and not detectable in the right region: shape and grey level produce pop-out, but their conjunction does not.

The pre-attentive mode achieves its instantaneous analysis of the image by means of parallel processing. Every pixel of the image is processed at once. Features such as edges and colour are instantly noticed, and grouped, although it would appear that conjunctions of features are not processed at this stage. The processing allows the eye to direct it attention to area that may be of interest.

Spatio-Temporal Stimuli in the Pre-Attentive Model

One form of visual stimuli that merits a mention in its own right is that of optical flow stimuli. Optical flow information is the rate of change of points in the visual field, with respect to time. Many, perhaps all, vision systems found in nature use optical flow information. Insects use optical flow information on a very primitive level to guide their flight paths. In the human vision system the optical flow information is gathered at the pre-attentive stage. Optical flow information is used to guide to the focus of interest of the eye. This is can be done on a voluntary level (e.g. when tracking an object), or on an involuntary level (e.g. when an unexpected object suddenly darts across the eyes filed of view.) One of the simplest use of optical flow information is to infer distance due to motion parallax. For example, when we are driving, optical flow information is used to steer the car in a straight line.

The Attentive Mode

The focus of attention of the eye is directed to an area of interest by the pre-attentive stage. This is a pre-conscious process: a stimuli response [3]. Human vision also works with feed-back loops, that is, *voluntary attention* also occurs. For example, when searching for red X's amongst a collection of red and green X's and O's, the person must consciously execute a search of all the red objects. These red objects may be highlighted in the pre-attentive stage, for they differ from the green objects in one feature: colour. The search of the red objects then continues on a conscious level: this is a voluntary attentive stage.
Two kinds of attention are usually considered:

- Selective attention - This is the model-driven case of the search for a known target and regards how appropriate an instance selected from the scene is. By model-driven we mean the case where the search target is already known.
- Spatial attention - This is the mechanism which orients attention to a particular location in the scene. This process may be voluntary or involuntary (data-driven involving an out of place target, i.e. a distraction in the scene). Involuntary spatial attention is in fact a pre-attentive process.

Implications to Random Access Vision

Random access sensors provide speedy direct access to individual pixels in the scene, and suggest the need for algorithms working on a small subset of the image, the regions of interest. What are we to learn from this enquiry into the human vision system? We have seen how attentive mechanisms, working with high resolution on local areas of interest, are

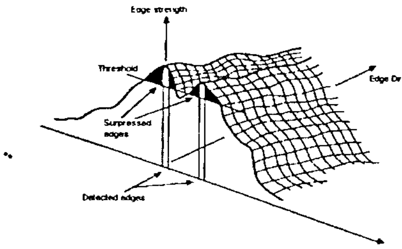

Figure 3: Contour Following

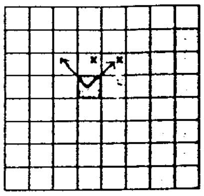

Figure 4: Edge Normal

Results

In order to test the performance of the blob tracker in conjunction with the random access sensor a virtual random access sensor was constructed. This model includes a cache and a sensor monitor. The Parameters can be changed: Sensor size, addressing speeds and cache size. This is used to test algorithms in software.

All tests were done in software, and it should be made clear that certain assumption about the characteristics of the sensors response to the test data have been made. First, images were captured using a standard CCIR camera. So the images represent a static moment in time. Second, the logarithmic response of the random access sensor was not mimicked by the CCIR camera, and it is yet unclear what the implications of this are. Further more, the times calculations are based on the makers figures, and no tests have as yet been done to see if the sensors access speeds lives up to expectations. The next step will be to perform test on a physical, rather that virtual random access sensor. Table 1 show the results from the 3 images tested, which can been seen in figures 10-13. Figure 5 illustrate the algorithm in action. The horizontal lines indicate the binary search path, the contour followed by the blob is also included.

Image No.	Quick Accesses	Slow Accesses	Total Accesses	Time
Image 1	3192	781	3973	2.2004ms
Image 2	2814	1977	4791	6.7682ms
Image 3	6082	3652	9734	8.5204 ms

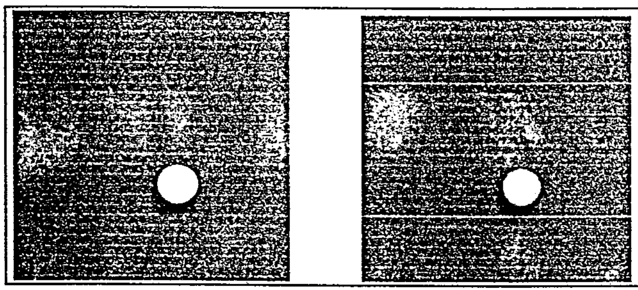

Figure 5: Left: image of blob, Right: Image with scan path superimposed

Conclusion

This paper presented what we believe to be a very useful model for implementing successful solutions for machine vision problems that wish to avail of high speed random access sensor technology. Compared with today's technologies random access sensors do represent significant speed saving, if used sensibly. The main advantage of the sensor is that the configuration of the photosite array can be very diverse. It is clear however, that sensors which can image a scene, and transfer that scene to computer memory in ultra fast speed are not that far away, and perhaps random access technology may soon become redundant. For the present this is not the case, and the work that we present here is clearly a way forward in reducing image analysis speeds. In any case the advent of faster hardware, of both sensors and computers, will not illuminate the need for random access vision. Machine vision has always been a very computationally expensive area of computing, and the model that we presented here does allow for very significant reductions computational cost.

Bibliography

[1] IMEC's Fuga 18 sensor litriture.

[2] Yarbus, L., *Eye movement and vision*, New York: Plenum Press, 1967

[3] Yantis, S., Jonides, J. *Abrupt visual onsets and selective attention: voluntary versus automatic allocation.* Journal of Experimental Psychology: Human Perception and Performance, Vol. 16, pp. 121-134, 1990.

[4] Neisser, U., *Cognitive psychology*. New York: Apple-Century-Crofts, 1967.

[5] Schneider, W., Pimm-Smith M., Worde, M., *Neurobiology of attention and automativeity,* Current Opinion in Neurobiology, Vol. $, pp. !77-182, 1994.

[6] Yamamoto, H., Yeshurun, Y., and Levine M.D. *An Active Foveated Vision System: Attentional Mechanisms and Scan Paths Convergence Measures.* Computer Vision and Image Understanding Vol. 63, No.1 pp 55-65.

[7] OE Reports No. 143/ Nov. 1995 pp1,13

[8] Swain, M.J. and Stricker, M.A. *Promising Direction in Active Vision.* International Journal of Computer Vision. pp 109-126. 1993.

[9] Wilf, J.M. *Chain-Code.* Robotics Age, Vol. 3, No. 2.