Fast Video Processing Using a Spiral Coordinate System and an Eye Tremor Sampling Scheme

J. Fegan,[1] S.A., Coleman,[1] D. Kerr,[1] B.W., Scotney[2]

[1]*School of Computing and Intelligent Systems,*
[2]*School of Computing and Information Engineering,*
Ulster University, Northern Ireland

Abstract

In the advent of autonomous machines, the need for real time video processing is becoming an increasingly important issue. Although technological advances have brought us closer to achieving this goal, they are often based on expensive and uniquely designed hardware solutions. It can be argued that as the complexity of image processing increases, it becomes more desirable to focus on portability and cost effective processing strategies. In this paper, we present a biologically inspired processing strategy that can be integrated with common, cost-effective image hardware. The results demonstrate that this approach can achieve a six-fold speedup, against a traditional image processing strategy, without any hardware modifications and a ten-fold speedup on adapted hardware. Alongside this, we present a novel type of processing that is used to detect video features in a space-time continuum. The results of this also demonstrate real-time processing potential and appear promising for motion focused tasks such as robot navigation.

Keywords: Fast Video Processing, Spiral Coordinate System, Eye Tremor, Edge Detection, Space-time Processing

1 Introduction

Efficient video processing is essential in many important machine vision tasks where computer hardware is expected to operate on a stream of consecutive image frames under strict time constraints. The prevalent way to accomplish these tasks, where runtime performance is important, often relies on long-standing principles that do not reflect our current understanding of biological vision. For example, in Traditional Image Processing (TIP) a digital image is sampled on a rectangular lattice and stored as a matrix of picture elements (pixels) according to a two-dimensional (2D) Cartesian or raster coordinate system. By contrast, the Human Visual System (HVS) senses stimuli on a hexagonal lattice of light sensitive cells [1]. This observation has inspired a one-dimensional (1D) spiral coordinate system that is effective for storing images sampled on a hexagonal lattice [2]. Unfortunately, the benefits of this scheme, including fast image processing performance, are currently undermined by a lack of hardware that can capture hexagonal images, and the subsequent computational cost needed to map an image to a hexagonal pixel structure [3]. To circumvent these issues, the spiral coordinate system was adapted for traditional, rectangular hardware [3].

Building on the work in [4], this paper presents the application of a spiral coordinate system and a biologically inspired sampling procedure to conduct fast video processing. Here the implementation has been optimised to ensure high speed performance whilst maintaining accuracy. Furthermore, we extend the implementation for temporal processing based on a sparse ‘space-time neighbourhood’ operation which demonstrates promising initial results for future work. An overview of the spiral framework is presented in Section 2, with spatial processing being presented in Section 3. Section 4 introduces the concept of a ‘space-time neighbourhood’ and provides a set of preliminary results with the work being concluded in Section 5.

2 Image Representation

The traditional way to store an image using a 2D Cartesian coordinate system is intuitive but the pixels must be stored as a sequence of rows or columns. Consequently, the pixels are not kept in proximity with all of their nearest neighbours and this means that the pixels cannot be processed with their nearest neighbours in a linear sequence. By comparison, a spiral coordinate system allows some pixels to be stored beside their nearest neighbours in a contiguous vector and this formation can be exploited to improve the runtime performance of image processing algorithms.

2.1 Square Spiral Coordinate System

In the square spiral (Squiral) coordinate system a single coordinate is used to locate a point in 2D space. In this system, the origin (numbered 0) is at the centre of a region being sampled. The eight points surrounding the origin are numbered in an outward spiral, thus each number represents a cardinal or intermediate direction, for example:

{C, W, NW, N, NE, E, SE, S, SW} = {0, 1, 2, 3, 4, 5, 6, 7, 8}

In a similar way, a base 9 number is assigned to each point, such that a point's distance from the origin is determined by the position of the digits in its coordinate. For example, a pixel at coordinate 315 is nine (3^2) pixels north, three (3^1) pixels west, and (3^0) one pixel east from the origin. Further information on the Squiral Coordinate system can be found in [3], [5] and [6]. In accordance with this system, each pixel at a coordinate 0 mod 9 is stored beside its eight nearest neighbours in a contiguous vector. This simplifies and facilitates fast spatial processing because these pixels and their eight nearest neighbours can be traversed sequentially. However, it is difficult to process pixels with neighbours contained in multiple spiral regions because these pixels are not stored contiguously, for example pixels 1 and 15 in Figure 1. To overcome this difficulty an approach based on the simulation of involuntary eye movements called tremors was proposed in [7] where a series of images which incorporate small pixel shifts are used to facilitate processing across multiple spiral regions. We extend that approach here by adapting it for video sequences.

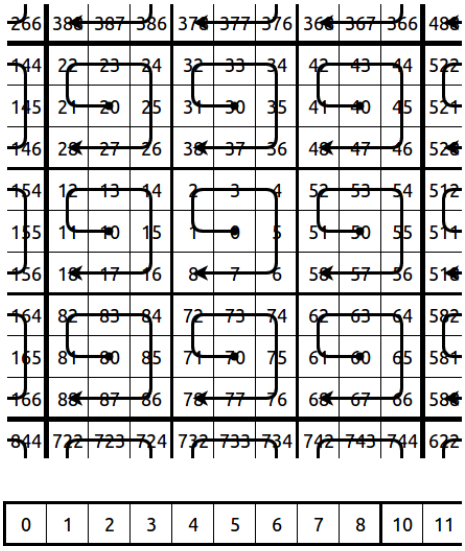

Figure 1: Squiral Coordinate System

2.2 Eye Tremor Frame Sampling

In video processing, the biological behaviour of eye tremor can be simulated by shifting the origin of the Squiral coordinate system, by one pixel, on each new frame. For example, Figure 2 shows a static 5x5 region where nine 3x3 frames are captured using the Squiral coordinate system. In the first frame (F_0) the origin of the Squiral coordinate system is located at the centre of the sampling lattice, and thus each pixel sampled at a coordinate 0 mod 9 is stored adjacent beside its eight nearest neighbours in a contiguous vector. In the next frame (F_1) the origin of the Squiral coordinate system is shifted left by one pixel. By doing this, the points that were previously coordinated 1 mod 9 assume the coordinates 0 mod 9 and are sampled in sequence with their surrounding neighbours. This process is repeated until a set of nine 'eye tremor' images are captured, one for each set of mod 9 pixels.

3 Spatial Processing

Spatial processing, commonly referred to as neighbourhood operations, describes an image operation where an output is computed by considering the properties of a pixel in relation to those of its surrounding neighbours. In TIP, all the pixels in an image are processed by considering the local neighbourhood area and a complete feature map output is obtained. This is not representative of the HVS, which sparsely interprets the visual stimuli it receives. In this section, a similar approach facilitated by the Squiral coordinate system and eye tremor sampling scheme is discussed.

3.1 Methodology

In this approach, only the central pixel in each spiral region is processed using its contiguous neighbours. As a result, the output pixels sparsely occupy one-ninth of a complete feature map. Therefore, the eye tremor-sampling scheme is used to 'focus' on a different mod 9 pixel in each spiral region allowing them to be sparsely processed in the same way. The outputs can be combined to produce a full-sized feature map. For example, Figure 2 demonstrates how a 3x3 image region can be sparsely processed across nine eye tremor frames (F_0 – F_8). In this situation, it takes nine initial frames to achieve a complete representation of the 2D scene. Thereafter each subsequent frame can be sparsely processed and combined with the output of the previous eight frames to construct a single, approximate feature map.

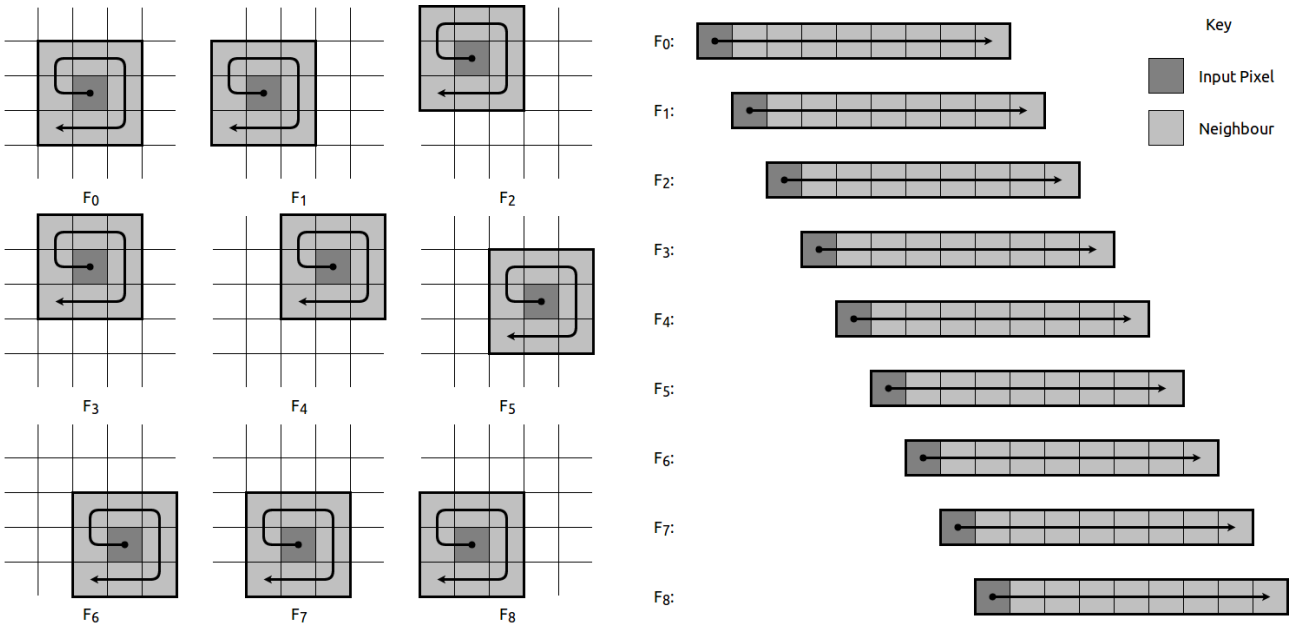

Figure 2: Spatial Processing Using Eye Tremor

3.2 Visual Evaluation

An implementation of the Sobel operator was tested on several videos and used to obtain features maps for a TIP approach and Squiral, eye tremor image processing approach. In this paper, ten frames (234 – 243) are shown from a video [8] which depicts a woman raising her right arm. The video was chosen for inclusion in this paper because it contains static and dynamic features that clearly illustrate the similarities and differences of the two processing approaches. The unaltered frames of the video are presented in Figure 3.1: the feature maps obtained using the TIP approach are shown in Figure 3.2; and the feature maps obtained using the Squiral, eye tremor approach are shown in Figure 3.3.

Figure 3.1: Arm Gesture

Figure 3.2: Traditional Feature Maps

Figure 3.3: Eye Tremor Feature Maps

In the eye tremor feature maps, the more relaxed features of the woman such as the face and torso closely resemble their counterparts in the traditional feature maps. However, there are some discrepancies around the lower arm and hand where there is fast movement. This is consistent with the feature maps from other test videos and indicates that the strength of a detected feature is affected by the rate of its spatial change. Based on this observation it is thought that a sparse eye tremor processing strategy will detect features more clearly if a higher framerate is used or if the spatial changes within the video sequence are small between frames. Regardless of these anomalies, the features detected in all of the test videos appear complete enough to support machine vision tasks.

3.3 Runtime Evaluation

The system used to measure the runtime performance had an Intel Core i7-4790 CPU @ 3.60GHz x 8, 16GB RAM and Ubuntu Linux 16.04 LTS 64-bit. The time taken to apply the Sobel operator to each and every frame of the video [8] at different resolutions are shown in Table 1. The times are measured in frames per second (fps) and correspond with a traditional and eye tremor implementation. The times given for the eye tremor implementation also list the time taken to map the 2D Cartesian frames to 1D Squiral frames; the time taken to map the 1D Squiral feature maps to 2D Cartesian feature maps (for display purposes); and the total time taken for all three actions. The timings show that a Squiral coordinate system and eye tremor sampling scheme can increase runtime performance significantly compared to a traditional implementation. At larger scales, such as layer 6, the speedup is almost 6 times faster than its traditional counterpart. If the overhead needed to map a traditional image to and from a Squiral image is removed the speedup is almost ten times faster.

Image Size		Traditional	Eye Tremor			
Layer	Pixels	Total	2D -> 1D	Edge Detection	1D -> 2D	Total
1	3x3	401,929fps	577,367fps	2,283,110fps	5,025,130fps	422,119fps
2	9x9	168,265fps	494,805fps	1,148,110fps	2,881,840fps	308,737fps
3	27x27	27,462fps	246,305fps	260,824fps	1,058,200fps	113,135fps
4	81x81	3,285fps	48,377fps	32,621fps	197,472fps	17,733fps
5	243x243	371fps	6,772fps	3,605fps	22,281fps	2,128fps
6	729x729	42fps	775fps	406fps	2,918fps	244fps

Table 1: Spatial Processing Runtimes

4 Space-time Processing

Previous research on the Squiral coordinate system combined with eye tremor sampling was primarily conducted on individual, static images, and it was only considered in the spatial domain. In video processing the influence of time is also considered. In most instances, temporal image processing examines how a pixel at a given location changes over time. In other words, a pixel is compared with a pixel at the same location in a previous or succeeding frame. In this section, we discuss a novel processing approach that incorporates spatial and temporal characteristics by using the vertical eye tremor processing strategy discussed in [6].

4.1 Methodology

In this space-time processing approach, the pixels in one frame are spatially processed using their neighbours in succeeding (future) frames. This idea is illustrated in Figure 4 where a pixel in the frame F_0 is processed using its eight spatial neighbours in the succeeding frames ($F_1 – F_8$). In practice, the Squiral, eye tremor frames are stacked to form a matrix and the pixels in the top row (F_0) are processed using the pixels that are parallel in the other rows. In this situation, a parallel pixel represents a spatial neighbour at a different point in time, a ‘space-time

neighbourhood' (Figure 4). In the implemented approach, a set of nine frames is needed before a single frame can be processed. Thereafter, each time a new frame is loaded, one-ninth of the top frame is processed. In other words, a set of mod 9 pixels are processed every time a new frame is loaded. The new frames are mapped to a second matrix. A limitation of this approach is that only one-ninth of the video frames can be processed. An alternative approach is to drop the oldest frame and append a new frame to the matrix. However, this presents a problem because only the first frame in a set of nine is vertically adjacent with its spatial neighbours. For example, the central pixel in F_1 does not have an immediate neighbour at the centre of $F_4 - F_6$. A possible solution to this problem is to use expensive base 9 computation or a lookup table similar to the one in [5] to locate a pixel's neighbours. This is considered a subject for further work.

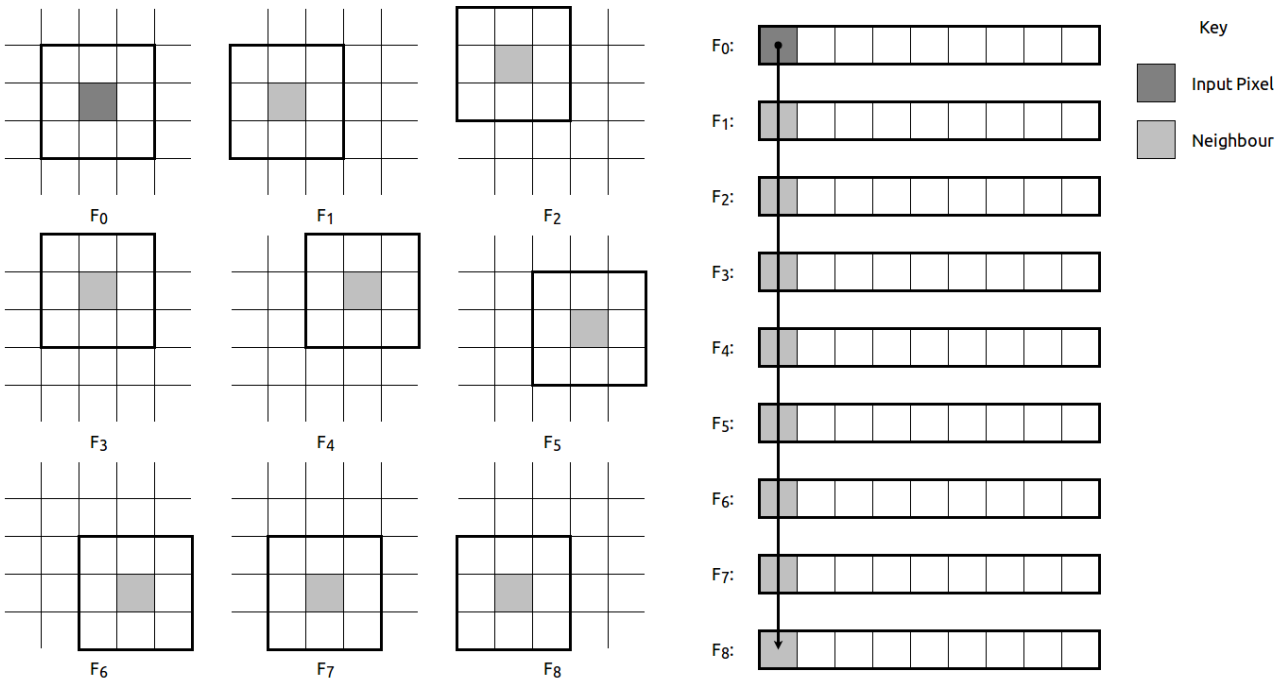

Figure 4: Space-time Processing Using Eye Tremor

4.2 Visual Evaluation

A set of features maps obtained from 'space-time neighbourhood' processing is shown in Figure 5. The Sobel operator was used with the system setup described in Section 3. In this example, it is visually apparent that the outline of the woman's right arm is the strongest detected feature. Therefore, it is hypothesized that feature detection in a space-time continuum will place more emphasize on features that change the most in space over time. Incidentally, in the complete video it is noticeable that some features such as the woman's shadow are identified more clearly in space-time. The usefulness of these unique space-time features is yet to be determined and will be considered in future research. Overall, it is thought that 'space-time neighbourhood' processing could be very useful in tasks where motion and run-time are key considerations.

Figure 5: Space-time Feature Maps

4.3 Runtime Evaluation

The runtime results for space-time image processing are shown in Table 2. The results indicate that this approach is slower than the eye tremor spatial processing results in Section 3, but they are still significantly faster than the TIP results at larger resolutions. This was expected, because a pixel's temporal neighbours are not contiguous with it in the conceptualised space-time matrix. It is envisioned that the performance of space-time processing could be further improved by sparsely storing pixels in each frame beside their space-time neighbours in a contiguous sequence.

Image Size		Runtimes			
Layer	Pixels	Cartesian -> Squiral	Edge Detection	Squiral -> Cartesian	Total
1	3x3	524,109fps	1,404,490fps	4,464,290fps	351,617fps
2	9x9	425,532fps	579,374fps	2,680,970fps	224,770fps
3	27x27	230,840fps	115,808fps	977,517fps	71,479fps
4	81x81	46,531fps	22,592fps	192,604fps	14,070fps
5	243x243	6,347fps	2,769fps	23,361fps	1,781fps
6	729x729	744fps	322fps	3,314fps	210fps

Table 2: Space-time Processing Runtimes

5 Conclusion

The effectiveness of a spiral coordinate system and eye tremor sampling scheme was evaluated by processing video footage at different resolutions. The results have shown that a spiral, eye tremor approach can be used to extract image features significantly faster than a traditional approach using a Cartesian coordinate system. In addition, it has been shown that a spiral, eye tremor scheme can facilitate processing in a space-time continuum and this approach could be more meaningful for motion focussed tasks. Future work will consider how multi-stage operators such as corner detectors can be applied to a spiral vector. Furthermore, we will continue to explore the characteristics of space-time processing and develop unique space-time operators.

References

[1] A. Róka, Á. Csapó, B. Reskó and P. Baranyi, “Edge Detection Model Based on Involuntary Eye Movements of the Eye Retina System,” *Acta Polytechnica Hungarica,* vol. 4, no. 1, pp. 31 - 46, 2007.

[2] L. Middleton and J. Sivawamy, Hexagonal Image Processing: A Practical Approach, Springer, 2005.

[3] M. Jing, B. Scotney, S. Coleman and M. McGinnity, “A Novel Spiral Addressing Scheme for Rectangular Images,” in *International Conference on Machine Vision Applications*, Tokyo, 2015.

[4] J. Ming, S. Coleman, B. Scotney and M. M., “Biologically Motivated Spiral Architecture for Fast Video Processing,” in *IEEE International Conference on Image Processing*, Quebec, 2015.

[5] J. Fegan, S. Coleman, D. Kerr and B. Scotney, “Fast Corner Detection Using a Squiral Architecture,” in *Irish Machine Vision and Image Processing*, Galway, 2016.

[6] J. Fegan, S. Coleman, D. Kerr and B. Scotney, “An Implementation Framework for Fast Image Processing,” in *International Conference on Robotics and Vision*, Wuhan, 2017.

[7] S. Coleman, B. Scotney and B. Gardiner, “Biologically Motivated Feature Extraction,” in *International Conference on Image Analysis and Processing*, 2011.

[8] T. U. o. Tokyo, “Services for High-speed Image Processing - Videos (SHIP-v),” Ishikawa Watanabe Laboratory, [Online]. Available: www.k2.t.u-tokyo.ac.jp/ship-v.