HIERARCHICAL MOTION ESTIMATION FOR EMBEDDED OBJECT TRACKING

Michael McErlean

Institute for System Level Integration
The Alba Campus, Livingston, EH54 7EG, Tel: - +44(0)845 2266508
Michael.McErlean@sli-institute.ac.uk,

***Abstract* – This paper presents an algorithm developed to provide automatic motion detection and object tracking embedded within intelligent CCTV systems. The algorithm development focuses on techniques which provide an efficient embedded systems implementation with the ability to target both FPGA and DSP devices. During algorithm development constraints on hardware implementation have been fully considered resulting in an algorithm which, when targeted at current FPGA devices, will take full advantage of the DSP resource commonly provided in such devices. The hierarchical structure of the proposed algorithm provides the system with a multi-level motion estimation process allowing low resolution estimation for motion detection and further higher resolution stages for motion estimation. An initial MATLAB prototype has demonstrated this algorithm capable of object motion estimation while compensating for camera motion, allowing a moving object to be tracked by a moving camera.**

***Keywords* – Wavelet Decomposition, Phase Correlation, Motion Detection, Motion Estimation, Object Tracking, FPGA, DSP, CCTV, Surveillance.**

I. INTRODUCTION

With the increasing numbers of CCTV cameras in everyday surveillance applications the burden on the surveillance operators is constantly increasing. The number of monitored video sources is growing beyond the effective capabilities of this human interface.

This paper describes a system which has been developed and prototyped in MATLAB, based on a proposed hierarchical block-based motion estimation algorithm. This system allows any standard Pan Tilt Zoom camera to independently and automatically detect and track a moving object within the field of view of the camera. This includes the capability to both pan and tilt the camera in order to maintain the tracked object in the camera view. The system comprises a hierarchical motion estimation algorithm integrated with a camera control loop allowing the system to behave as an automatic object tracking system. The novelty of this algorithm exists in the hierarchical motion estimation approach to object tracking and its specific development targeted at embedded hardware/software implementation. The motion estimation technique is based on an existing researched and qualified technique called Phase Correlation [1-9]. The proposed algorithm takes this base motion estimation technique and develops it to suit the application, making use of multiple image resolution levels as a means of handling variable object size and velocity. The adoption of a block based approach allows for ease of implementation in hardware and an effective use of the dedicated DSP hardware resource common in current FPGA devices.

This paper is organized as follows. Sections II and III describe the processes of Object Tracking and Motion Estimation, distinguishing the proposed application of both FPGA and DSP implementations. Section IV describes the hardware constraints relevant during algorithm development and the proposed solution to accommodate these constraints. Section V describes a MATLAB implementation of the proposed algorithm as a proof of concept. Sections VI and VII illustrate results from the prototype implementation and discuss conclusions from the proof of concept work carried out here.

II. OBJECT TRACKING

The ability to track an object in a scene requires a means of determining how much and in what direction the object has moved in a given time interval, commonly referred to as motion estimation. Object tracking subsequently involves the control loop around this estimation technique, updating the position of the tracking reference point before the next attempt at estimating the object's motion. Currently there are a variety of methods used to perform the task of motion estimation. Possibly the most common utilisation of motion estimation techniques is in predictive video compression algorithms. The majority of existing object tracking algorithms are software based, mainly stemming from the field of computer vision, and these commonly use a centralized processing resource to determine motion in captured video sequences. The distinguishing factor of this target application is the intention to distribute the tracking intelligence to the camera itself at the perimeter of the CCTV network, allowing it to handle object tracking and camera control in hardware as an autonomous system.

III. MOTION ESTIMATION

This distinguished hardware oriented object tracking approach initially focuses on examining methods of utilizing the information already provided by existing video compression algorithms as a means of feeding an object tracking algorithm. As a result this minimises any extra processing and hardware resource required for an embedded implementation. Current video compression algorithms such as MPEG4 and H.264 make use of motion estimation and motion compensation techniques to maximise compression along the time axis. These methods work on the basis that the overall content difference between consecutive video frames is usually small and a method of predicting the next frame in the sequence greatly reduces the amount of data which needs to be transmitted [2][8][10][11].

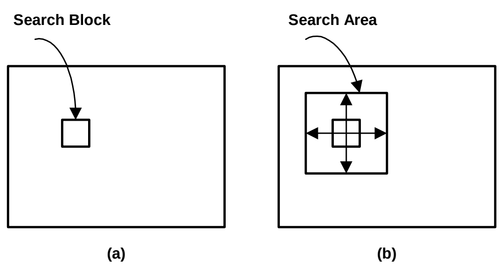

Fig.1. Block Matching technique (a) current block being matched, (b) constrained search area within which a best match is found.

Currently the most commonly used motion estimation method is Block Matching, Fig.1, where each N*N block of pixels is compared with its corresponding and neighbouring pixel elements in a previous reference frame to determine the block position which minimises the difference function. The result is a block of neighbouring pixels which can then be coded as a motion vector and a difference value, thereby reducing the amount of data to be transmitted but allowing the decoder the ability to reproduce the original image. With block matching algorithms the performance of the operation is dependent on how constrained the search routine is for the current frame; the time taken to perform the block matching search can be greatly reduced by limiting the amount of the image to be searched by the current block. An obvious drawback with limiting this search area is that this also limits the amount of object displacement which can be measured between consecutive frames. In terms of an object tracking application a more significant limitation of block matching is the fact that it provides a best match motion estimate within the given search area as opposed to a measure of the actual motion of a given object. This actual motion measurement is crucial for the ability to track an object using motion estimation.

An alternative technique to the Block Matching system which provides this actual motion measurement involves a frequency domain technique known as Phase Correlation [1]. This technique can provide a number of motion vector candidates which represent actual motion content between two images. With phase correlation, translational changes in the image in terms of motion can be measured as a phase shift in the frequency domain.

Let g_0 and g_1 be the two images which differ by a displacement of (x_0, y_0), the two images are related as follows:

$$g_1(x, y) = g_0(x - x_0, y - y_0) \tag{1}$$

Their Fourier transform is then related as follows:

$$G_1(\xi,\eta) = e^{-j2\pi(\xi x_0 + \eta y_0)} \cdot G_0(\xi,\eta) \tag{2}$$

The inverse Fourier transform of the phase correlation equation shown here:

$$\frac{G_0(\xi,\eta).G_1^*(\xi,\eta)}{|G_0(\xi,\eta).G_1^*(\xi,\eta)|} = e^{j2\pi(\xi x_0 + \eta y_0)} \tag{3}$$

produces a phase correlation surface (Fig.2). If the two images are identical but shifted the result will be an impulse at (x_0, y_0) which represents the translational displacement between the two images.

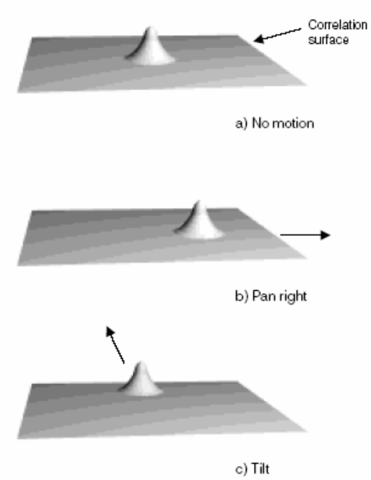

Fig.2. Phase Correlation measures translational shifts between images with individual peaks representing separate motion vectors.

In terms of an object tracking application, phase correlation provides a true match for an object between frames unlike the limited best match case of block matching. There are, however, limitations with the phase correlation method when used for object tracking. The size of the tracked object in terms of the overall area of the image has a direct impact on how well its motion can be detected. If the background area is greater than that of the tracked object then the background will have a dominating effect on the correlation process resulting in a suppressed correlation peak for the moving object and a dominating peak for the background. This would represent a real problem for the target application where the background may appear to be moving due to ego-motion introduced by camera motion during object tracking.

The use of shape adaptive phase correlation (SAPC) for motion estimation, where the algorithm is adapted to suit the size and expected velocity of the object being tracked, has previously been proposed [3]. This approach overcomes the identified problems associated with background dominance during phase correlation and as a result it can handle objects of any size with a greater motion between frames.

This paper proposes an alternative solution tailored towards hardware implementation. The following section describes the development of the proposed hierarchical block-based phase correlation technique (HBBPC). This technique overcomes the identified limitations of phase correlation, due to its dependence on object size, and also addresses hardware implementation concerns identified for the alternative SAPC technique.

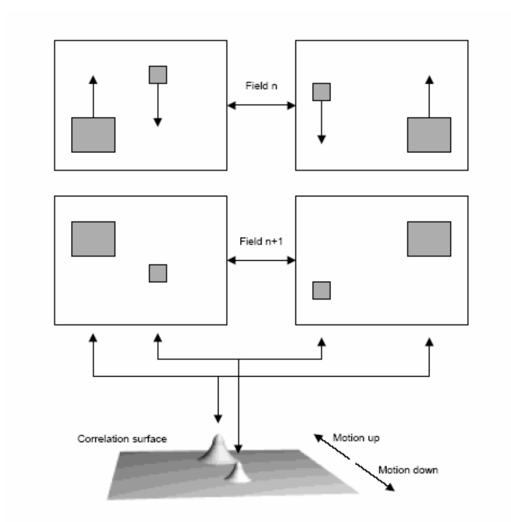

Fig. 3. Phase correlation measures a relative motion between images and is dependent on object size.

IV. HARDWARE IMPLEMENTATION CONSTRAINTS DURING ALGORITHM DEVELOPMENT

A number of hardware considerations have shaped the development of the tracking algorithm based on the phase correlation technique. Initial experimentation and prior literature indicate that the phase correlation technique provides a robust and very accurate means of determining the relative motion between two images. Initially developed for image registration it can handle changes in orientation and illumination and is capable of producing sub-pixel motion estimation measurements [1][2]. In terms of object tracking, the spatial presence of the object within the scene is a major factor in whether the dominating correlation peaks are representative of an object's motion or that of the background, Fig.3.

In the SAPC technique the foreground object is initially segmented from the background; the remaining background pixels are replaced with an average luminance value and the phase correlation is performed using the segmented object image. This technique requires the implementation of an adaptive phase correlation algorithm based on the size of the segmented object. Considering Fig.5's illustration of a possible hardware implementation of the phase correlation block diagram in Fig.4, the requirement for fixed size FFT resource and memory allocation make implementing a dynamic sized phase correlation difficult. The proposed HBBPC algorithm accommodates this hardware resource allocation consideration by using a fixed block size implementation of the phase correlation process thus providing a deterministic system where each block takes a preset time to process.

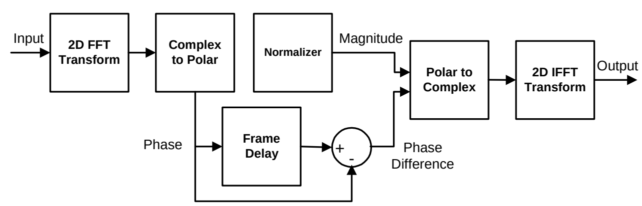

Fig.4. Phase Correlation Block Diagram.

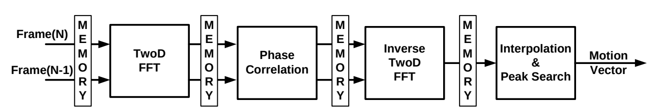

Fig.5. Phase Correlation Hardware Implementation.

As already established there are drawbacks with this block-based approach, smaller blocks require less hardware resource but limit the measurable displacement of the object between frames. However larger blocks risk the problem of the background dominating the correlation peak and giving motion vectors relating to the background motion rather than the foreground object motion, so tracking the object becomes impossible. Also with a typical 16*16 pixel block-based technique new information introduced at the block edges due to motion can introduce noise in the output of the phase correlation process. To handle this noise a raised cosine window is applied to the block, this window attenuates the signal frequencies at the perimeter of the block reducing the effects of new edge information. While this improves the results of the phase correlation it limits the effective area being tracked and the measurable velocity of the tracked object. Fig.6 illustrates how the use of a larger block size of 32*32 pixel elements gives a larger overlapping area for correlation and by applying the windowing function to this larger area it can be assumed that the attenuated frequencies have a negligible impact on the area inside the central 16*16 block. To compensate for the information suppressed due to attenuation at the edges the blocks are overlapped in each direction. This overlapping process quadruples the number of blocks to be processed but gives a better coverage while still providing good immunity to errors introduced by edge information.

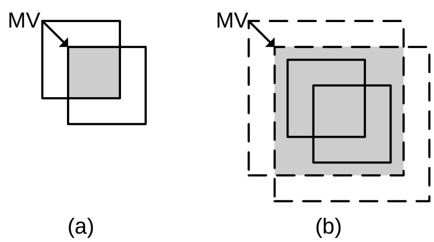

Fig.6. Correlation area using (a) 16*16 and (b) 32*32 blocks.

A means of providing a hierarchical range of process levels feeding into the block-based motion estimation algorithm is also proposed, allowing the motion estimation to be targeted at specific areas of the scene at the next level of the hierarchy. This system also provides an improved means of coping with larger displacements and can handle tracking of objects of different sizes. A quad tree reduction which systematically divides the segments into quadrants and performs phase correlation on each segment would provide this dynamic range [12]. However going back to hardware implementation considerations this technique would require the ability to process different sized blocks at each level of the quad tree hierarchy. Here the proposed solution to this, illustrated in Fig.7, utilises a wavelet decomposition of each image to provide multiple resolution levels feeding into a fixed block size phase correlation. This proposed technique provides a dynamic range for tracking variable sized objects and at different velocities but also allows the hardware implementation to use the same fixed size block based phase correlation at each level. Using this HBBPC algorithm different resolution levels can be exploited to approximate motion in given areas and adaptively concentrate higher resolution motion estimation in specific regions as required.

V. MATLAB PROTOTYPE

The proposed HBBPC algorithm has been implemented using MATLAB as a means of proving the performance of the tracking algorithm in a real environment. A control loop has been developed which determines where the tracked object is within the scene based on the position of a tracker box. From this coordinate position and the preset dimensions of the tracking box the control algorithm can determine in which direction to move the camera to maintain the object in the centre of the scene. The system processes a selected region of each video frame to determine the motion content at the multiple levels of the hierarchical algorithm. The best motion vector results are determined based on the confidence in the peak measured at the output of the phase correlation. This confidence measure is decided on by identifying the amplitude difference between the best peak and the second best peak. A large difference between these peaks indicates a higher confidence that the larger peak represents the best measure of motion in the block. For the three-level hierarchy a set of motion vectors are then produced and an average motion vector is determined which provides feedback to the tracker location function updating the tracker position for the next iteration. Once the tracker has been updated a separate function determines the current position of the tracker and decides in which direction the camera needs to move to maintain the object in its field of view.

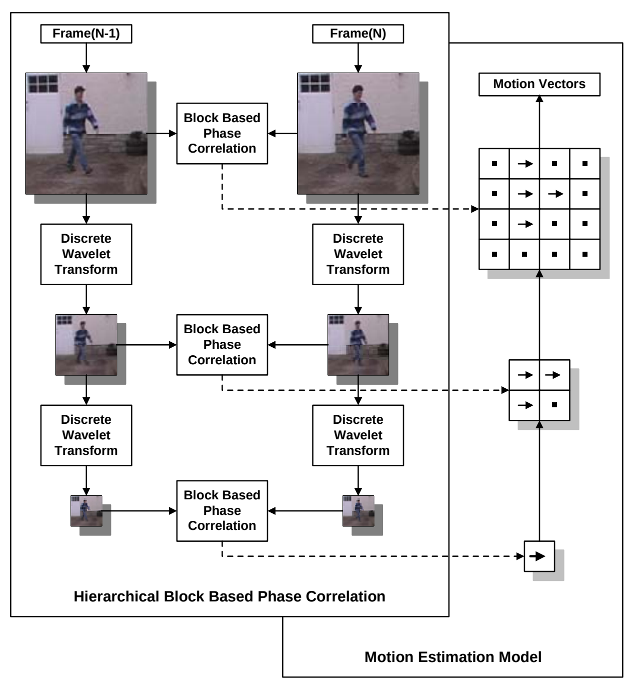

Fig.7. Hierarchical Block Based Phase Correlation.

VI. Results

The following image sequence is taken from the MATLAB prototype demonstrating the tracking of a person with a moving PTZ CCTV camera.

VII. Conclusions

A novel adaptive hierarchical algorithm has been developed specifically for embedded hardware implementation of a distributed object tracking algorithm which is both robust and flexible enough to handle real world applications. The algorithm takes an existing method for motion estimation and builds upon the obvious advantages it provides over other established techniques. The proposed hierarchical motion estimation technique overcomes the restrictions which have been identified due to the use of a block based approach for motion vector generation. The result is a dynamic model which can estimate motion in larger areas at lower resolutions and, based on this initial measurement, the effort at higher detail stages can be targeted to areas where motion is present. Using sub-pixel measurements in the phase correlation output of low resolution approximations it is possible to gain an accurate measurement of larger displacements which would have been beyond the limits of the basic block based approach. This multi-level approximation allows for tracking of varying sized objects and objects with a larger displacement between frames.

The operation of the proposed algorithm has been proven in software using MATLAB but limitations on processing performance mean object tracking is typically limited to a rate of one frame per second.

As a result of this initial MATLAB implementation further development will concentrate on an FPGA based pipelined hardware implementation of the algorithm with the goal of increasing the frame rate of the object tracking to full rate at 25fps.

ACKNOWLEDGMENTS

Figs [1 – 3] are used with permission from Snell & Wilcox, and sourced from “The Engineer’s Guide to Motion Compensation”.

This work is supported as part of EngD funding provided by the Engineering and Physical Sciences Research Council (EPSRC) and has been carried out in conjunction with:

Salent Technologies Ltd
The Rowan House
Wyndford Brae
Philpstoun
Nr. Linlithgow
West Lothian
EH49 6RN, U.K.
http://www.salent.co.uk

REFERENCES

[1] Kuglin C. and Hines D, "The phase correlation image alignment method", Proc. Of the IEEE Int. Con. On Cybernetics and Soc., 1975, pp.163-165.

[2] Thomas G, "Television motion measurement for DATV and other applications", 1987 BBC Research Department, Research Report 1987/11.

[3] Hill L. and Vlachos T, "Motion measurement using shape adaptive phase correlation", Electronics Letters., 2001, Vol. 37 No. 25.

[4] Hill L, Vlachos T, "On the estimation of global motion using phase correlation for broadcast applications", Image Processing and Its Applications, 1999. Seventh International Conference on (Conf. Publ. No. 465) , Volume: 2 , 13-15 July 1999 Pages:721 - 725 vol.2

[5] Erturk S, "Digital image stabilization with sub-image phase correlation based global motion estimation", Consumer Electronics, IEEE Transactions on, Volume: 49, Issue: 4, Nov. 2003 Pages:1320 – 1325

[6] Wu S.F, Fernando G.M.X, "Comparative study of two motion estimation techniques", Image Processing and its Applications, 1989, Third International Conference on , 18-20 Jul 1989 Pages:305 – 309

[7] Kumar E, Biswas M, Nguyen T.Q., **"**Global motion estimation in frequency and spatial domain"**,** Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on, Volume: 3 , 17-21 May 2004 Pages:iii - 333-6 vol.3

[8] Dabner S.C.; "Motion estimation for HDTV", Image Processing for HDTV, IEE Colloquium on , 26 Oct 1989 Pages:2/1 - 2/4

[9] Watkinson J, "The Engineer's Guide to Motion Compensation", Snell & Wilcox, 1994

[10] Dufaux F; Moscheni F; "Motion estimation techniques for digital TV: a review and a new contribution", Proceedings of the IEEE ,Volume: 83 ,Issue: 6 ,June 1995 Pages:858 – 876

[11] V. Bashkaran and K. Konstantinides, 'Image and video compression standards: algorithms and architectures', Kluwer Academic Publishers, 1995.

[12] Argyvriou, V.; Vlachos, T., "Motion estimation using quad-tree phase correlation," Image Processing, 2005. ICIP 2005. IEEE International Conference on , vol.1, no.pp. I- 1081-4, 11-14 Sept. 2005.