VIDEO OBJECT TRACKING USING REGION SPLIT AND MERGE AND A KALMAN FILTER TRACKING ALGORITHM

S.A.Vigus, D.R.Bull, C.N.Canagarajah

Department of Electrical and Electronic Engineering
University of Bristol
Woodland Road, Bristol, BS8 1UB, U.K.

ABSTRACT

This paper proposes a reliable method for tracking the trajectory of video objects using the vector Kalman predictor. Video objects, within the scope of this paper, are defined as groups of image pixels coherent spatially as well as in their values of luminance. The extent to which the quality of the unsupervised region split and merge segmentation affects the accuracy of the tracker is discussed, alongside the improvements made in the segmentation process as a result of the feedback from the tracking algorithm. The overall low complexity of the system, and the time savings made in using a spiral search algorithm, provide this method with prospects of being implemented in real-time.

1. INTRODUCTION

Motion tracking is the process whereby an object or region is tracked using information about its behaviour in terms of motion. There are many applications that already benefit from motion tracking including surveillance cameras, radar, and air traffic control systems. The simplest tracking algorithms involve block matching and motion estimation. These methods are easy to implement and produce reasonable results in a post-processing situation. However, when an object must be tracked for real-time applications, a faster, more efficient estimation method must be used.

This paper proposes a video tracking algorithm using the Kalman filter and the region split and merge method. The novelty of the proposal is in the combination of the segmentation and tracking techniques used, as well as the spiral search applied to subsequent frames in the sequence. Both the segmentation and tracking techniques employed in this paper keep complexity to a minimum, with a view to the real-time application of the complete system in the future. However the reduction of the complexity of a technique must not detriment the accuracy of the overall process.

The accuracy of a tracking algorithm can be thought of as the error between the estimation and measurement of the target location. This accuracy however, is itself dependent on the accuracy of the input measurement, which, in the case of video, is the result of the segmentation algorithm used. Two segmentation methods are tested in this paper, in order to investigate this dependence. The first is a very basic block luminance method, and the second the region split and merge method.

2. BASIC METHODOLOGIES

2.1 Segmentation

Segmentation methods can be broadly broken down into two categories, region-based[11] and edge-based processes[12]. Edge-based methods concentrate on the discontinuity in an image and region-based methods, the similarity. Edge-based methods are best employed where there are borders separating areas with significantly different properties, for example in images containing a majority of man-made objects, which are not as rich in visual information as natural scenes. For a natural environment it is usual to use region-based extraction techniques, which identify areas with similar properties in terms of texture and motion. In dynamic image sequences, motion can also act as a very useful cue.

One region-based technique is luminance thresholding. This method, in its most basic form, allocates any pixel with a value greater than a given threshold to one region, and those pixels with lower values to another region. When images contain many different regions, multiple thresholding is used. Threshold values are normally taken from a histogram of the image, but more complex segmentations can be carried out using tree-structured algorithms[11]. The correct choice of threshold value is imperative to obtain satisfactory segmentation.

0-7803-6725-1/01/$10.00 ©2001 IEEE

Too low a value would result in the object not being picked out from the image, and too high a value could result in every pixel belonging to the same region. Thus, no more information would be retrieved than before the segmentation was carried out.

Other region extraction techniques include relaxation, region growing[13], region merging[2], and split and merge. These are based on the systematic modification of an initial array of elementary regions, until a stable classification is reached. Similarity measures used in these techniques are usually based on region properties such as mean or variance. The main drawback with these procedures is that they yield unstable results, which are dependent on the order in which the regions have been analysed. Furthermore, regions which share similar properties, but which are unconnected, are labelled differently.

Spatio-temporal segmentation is an efficient means of describing dynamic scenes in terms of moving objects. There are many different spatio-temporal segmentation algorithms, including region merging. Moscheni and Bhattacharjee[2] discuss a robust region merging method, which exploits both the spatial and temporal information available in a scene. The results show that it can work well from various initial regions to form the same important final regions e.g. arm, racket and ball in the table tennis sequence.

2.2 Tracking

In order to enable motion tracking to be carried out in real-time, it is best to use a predictor operator to estimate where the object is most likely to be in the next time instant. This extra knowledge means that a spiral search can be conducted from the estimation point outwards, rather than over the whole frame. As the spiral search leads to a result more quickly, the object being tracked should not have moved as great a distance and so be easier to locate. Predictor operators generally fall into one of two categories, batch estimators or recursive estimators. The predictor discussed in this paper is the Kalman filter[7], which is a recursive predictor operator. Its recursive nature means that it is more applicable to real-time operations, as each observation is processed as it becomes available.

The Kalman filter predictor is a linear system in which the mean square error between the predicted output and the actual output is minimised. It is an algorithm that makes optimal use of imprecise data to continuously update the best estimate of the system's current state. There are two main sources of inaccuracy within a system, noise and error in readings. The Kalman filter predictor takes into account the inherent noise associated with any signal, in order to minimise prediction errors. The predictor is governed by the series of matrix equations (1-3) below[7].

Predictor Equation

$$\hat{\mathbf{x}}(k+1|k) = A\hat{\mathbf{x}}(k|k+1) + G(k)\left[\mathbf{y}(k) - C\hat{\mathbf{x}}(k|k-1)\right] \quad (1)$$

Predictor Gain

$$G(k) = AP(k|k-1)C^T\left[CP(k|k-1)C^T + R(k)\right]^{-1} \quad (2)$$

Prediction Mean Square Error

$$P(k+1|k) = [A - G(k)C]P(k|k-1)A^T + Q(k) \quad (3)$$

The simplest object to track is a ball, due to its uniform shape, regardless of the viewing camera's position. Ziliani and Moscheni[6] use a Kalman filter for recursive spatio-temporal segmentation and predicting the motion of the ball in the table-tennis sequence. Other researchers have successfully tracked footballs and ping-pong balls in carefully controlled environments, but there is still a lot of scope for improvement. The ideal tracking system should be able to automatically track any moving object. At present, trackers have only been implemented for given sets of motion circumstances, using sequence specific segmentation algorithms. As well as spatio-temporal segmentation, the Kalman filter has also been integrated with the wavelet transform[3] and various motion estimation techniques[4][5][8][9][10].

3. PROPOSED ALGORITHM

The proposed video tracking algorithm uses region split and merge to segment the first frame in the image sequence, thus locating the region of interest. To decide whether a block should be split or not, the region split and merge method uses a homogeneous criterion, in this case the standard deviation of luminance within a block (Eq. 4). If less than 80% of pixels within any one block comply with the homogeneous criterion then the block will be split.

$$\sigma = \left(\sum Lum_{mean}\right)^2 - \sum \left(Lum_{mean}\right)^2 \quad (4)$$

A successful region split and merge results in a number of regions, one of which is selected as the region of interest to be tracked. The coordinates of the located region provide the input to the Kalman filter, which having been initialized with the standard equation of motion (Eq.5) calculates the estimated coordinates of the region in the next frame. The segmentation of subsequent frames spirals outwards from the estimated location, until a suitably matching region is found.

$$s = ut + \tfrac{1}{2}at^2 \quad (5)$$

$$s_y = u\cos\theta t + \tfrac{1}{2}at^2 \quad (6)$$

$$s_x = u\sin\theta t \quad (7)$$

Adapting Eq.5 to include angles of departure results in a more general set of equations (Eqs.6-7), where s_x and s_y represent the displacements in the x and y direction respectively. These equations assume that the only acceleration applied to the object is that due to gravity, which only acts in the y direction. The angle in these equations can be adapted automatically to allow for changes in motion. For example in the table tennis sequence, the change from the up and down motion present at the beginning of the sequence, to a trajectory across the table.

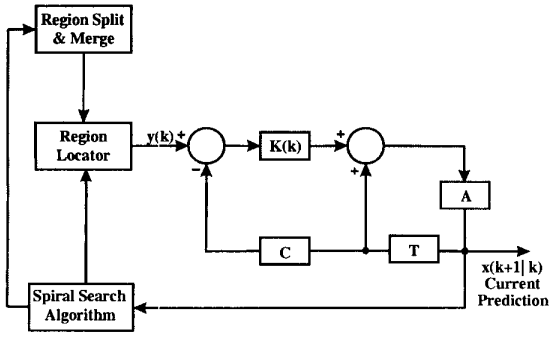

Figure 3.1: Segmentation and Kalman Filter Block Diagram

The block diagram in Figure 3.1 describes the way in which the region split and merge algorithm is linked with the Kalman filter to create the video tracker. The first frame is segmented using the region split and merge algorithm, the result of which is fed into the region locator module and on into the Kalman filter. The current prediction value, from the Kalman filter, is fed back into the spiral search algorithm, which if a match is found is linked directly to the region locator and back into the Kalman filter for the next frame. However, should a match not be easily found by the spiral search algorithm, the task is passed on to the original region split and merge process.

4. RESULTS

The results of the block luminance search and region split and merge methods used are shown in Figure 4.1 and Figure 4.2 respectively. The region split and merge routine segments down to a block size of 4*6 pixels, to maintain the correct aspect ratio, and is shown with a Laplacian function applied to highlight the region boundaries.

Method	Lum	Lum+Spiral	RSM	RSM+Spiral
Ave. Time/ Frame (s)	0.092	0.044	0.733	0.145

Table 4.1: Average segmentation time per frame

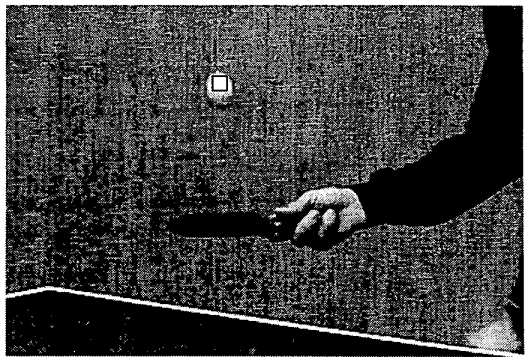

Figure 4.1: Initial segmentation of 'Table Tennis' #0

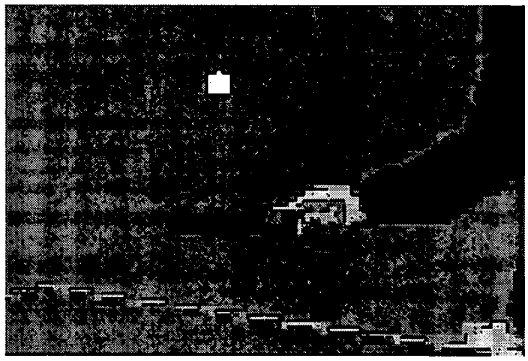

Figure 4.2: Region Split/Merge

Table 4.1 gives the average time taken to segment each frame using both the luminance search and the region split and merge methods. The time savings made in using the spiral search algorithm are also apparent. This is particularly true in the case of the region split and merge segmentation, where the average frame time is approximately $1/5^{th}$ that of segmenting every frame. The luminance and spiral search algorithm produces an average frame time of 0.044s. Given that in video each frame is shown for approximately 0.04 seconds, this segmentation is just 4 milliseconds away from real-time performance.

Figure 4.3 demonstrates the result of the basic Kalman filter tracking algorithm with region split and merge and the spiral segmentation. The black ball superimposed on the original images represents the estimated location of the ball received from the tracking algorithm. The tracking algorithm achieves just over 83% accuracy with the simple segmentation input. Given that each frame is only shown for $1/25^{th}$ of a second, with only 1 frame in every 6 displaying an incorrect track result, this is not greatly unacceptable to the human eye.

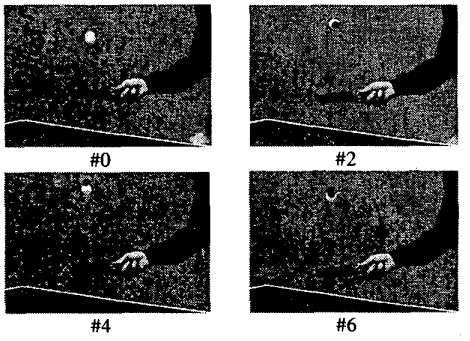

Figure 4.3: Results of Basic Kalman Filter Tracking

The graph in Figure 4.4 gives an idea of the difference in segmentation accuracy between the split and merge and the block luminance methods. It shows the error between the measured and estimated position of the ball in the y direction up to frame 30, where motion is mostly vertical. It can be seen that although both methods suffer errors from the abrupt changes in direction, the overall error in the region split and merge is not as great. The speed at which the filter regains the true track, particularly with region split and merge, is also evident in the graph.

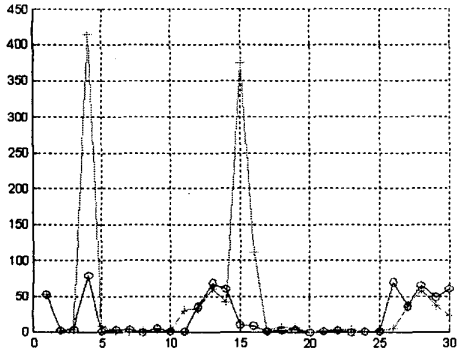

Figure 4.4: Error between estimated and measured y position. Where + represents the luminance and o represents the region split and merge method.

5. SUMMARY

This paper has presented a novel approach to video object tracking through the use of the region split and merge method and Kalman filtering. The spiral search algorithm applied to the estimated location, received from the tracker, resulted in an 80% increase in segmentation speed. The results show that real-time performance is obtainable with the luminance search algorithm, but at the sacrifice of the accuracy of the region split and merge segmentation.

6. REFERENCES

[1] Balasubramanian R., Goldgof D.B, and Kambhamettu C., *Tracking of Nonrigid Motion and 3D Structures from 2D Image Sequences Without Correspondences*, IEEE International Conference on Image Processing, Chicago, USA 1998.

[2] Moscheni F. and Bhattacharjee S., *Robust Region Merging for Spatio-Temporal Segmentation*, IEEE International Conference on Image Processing, Laussane, Switzerland 1996.

[3] Leduc J-P., Mujica F., Murenzi R. and Smith M., *Spatio-Temporal Wavelet Transforms for Motion Tracking*, IEEE International Conference on Image Processing, Santa Barbara, USA 1997.

[4] Jung S. and Wohn K., *Tracking and Motion Estimation of the Articulated Object: a Hierarchical Kalman Filter Approach*, Real-Time Imaging, Vol. 3, pages 415-432, 1997.

[5] Motsch J. and Nicolas H., *3D Motion Estimation of Video Objects Using A-Priori Data and 2D Apparent Motion*, IEEE International Conference on Image Processing, Chicago, USA 1998.

[6] Ziliani F. and Moscheni F., *Kalman Filtering Motion Prediction for Recursive Spatio-Temporal Segmentation and Object Tracking*, Workshop on Image Analysis for Multimedia Interactive Services, Belgium 1997.

[7] Bozic S.Z., *Digital and Kalman Filtering*, Second Edition, Edward Arnold, 1994.

[8] Calvagno G., Celeghin L., Rinaldo R., and Sbaiz L., *Statistical Based Motion Estimation for Video Coding*

[9] Efe M. and Atherton D.P., *Maneuvering Target Tracking with an Adaptive Kalman Filter*, Proceedings of the 37th IEEE Conference on Decision and Control, Florida 1998.

[10] Kim E.T., Han J.K. and Kim H.M., *A Kalman-Filtering Method for 3D Camera Motion Estimation From Image Sequences*, IEEE International Conference on Image Processing, Santa Barbara, USA 1997.

[11] Hsu F-J., Lee S-Y., Lin B-S., *Similarity retrieval by 2D C-Trees matching in image databases*, Journal of Visual Communication and Image Representation, *Vol. 9, No. 1, March '98, pp. 87-100.*

[12] Iannizzotto G. and Vita L., *Fast and accurate edge-based segmentation without contour smoothing in 2-d real images*, IEEE Transactions on Image Processing, July 2000.

[13] Papadimitriou Th., Diamantaras K.I., Strintzis M.G. and Roumeliotis M., *A Novel Rigid Object Segmentation Method Based on Multiresolution 3-D Motion and Luminance Analysis*, IEEE International Conference on Image Processing, Vancouver 2000.