HYBRID SALIENT MOTION DETECTION USING TEMPORAL DIFFERENCING AND KALMAN FILTER TRACKING WITH NON-STATIONARY CAMERA

Xuesong Le[1], Ruben Gonzalez[2],

School of ICT, Griffith University, Australia, School of ICT Griffith University, Australia

Abstract

Uncertain motion of typical surveillance targets, e.g. slow moving or stopped, abrupt acceleration, and uniform motion makes a single salient motion detection algorithm unsuitable for accurate segmentation. It becomes even more challenging in case of the camera is non-stationary. In this paper, first, a simple local adaptive temporal differencing method is proposed to detect moving objects boundaries and partial interiors. To improve the accuracy of detection, a Bottom-up Variable Block Size block matching method is employed to identify the existence of possible moving object blocks and then an adaptive Kalman filter is applied to distinguish salient motions from other distracting motions. At last, the motion data from two algorithms are successfully fused to determine whether a region has been changed or not. Experimental results comparing the proposed and other competing methods are evaluated objectively and show that the proposed method achieves promising motion results for a variety of real environments

Introduction

Detecting salient motion in image sequence is one of main tasks in some promising applications, such as video surveillance, traffic monitoring, etc. However, it is still in its early developmental stage and needs to improve its robustness depending on the specific scene conditions. Some of the most challenging problems are those in which motion is being exhibited not just by the objects of interest, but also by other factors such as varying illumination, dynamic backgrounds, crowded scenes and occlusions. In case of the camera is non-stationary, moving object detection problem becomes even more challenging, since the background is not static and background subtraction methods cannot be employed anymore. In this work, we examine the feasibility of using the temporal differencing in conjunction with Kalman filtering based BM algorithm for motion detection with a non-stationary camera.

Methodology

Figure 1 illustrates the overall methodology. After camera motion compensation, two motion detection based techniques are proposed to accurately segment moving objects out. First, a simple local adaptive temporal differencing method is proposed to detect moving objects boundaries and partial interiors. To improve the accuracy of detection, a Bottom-up Variable Block Size block matching method is employed to identify the existence of possible moving object blocks and then an adaptive Kalman filter is applied to distinguish salient motions from other distracting motions. At last, the motion data from two algorithms are successfully fused to determine whether a region has been changed or not.

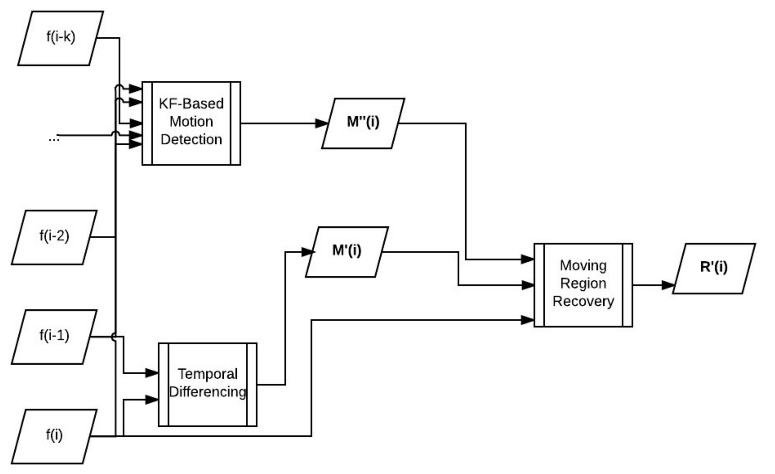

Fig. 1 Overall methodology

TEMPORAL DIFFERENCING

As shown in fig.2, the motion detection algorithm using temporal differencing starts with frame subtraction between inverse transformed current frame F^(i) and its previous frame f(i-1) Next, a simple local adaptive thresholding method is applied to remove camera noise. The observation under the null hypothesis is modelled as a Gaussian random variable with zero mean and variance. The unknown parameters are approximated using simple robust statistics method. Then a bi-level difference magnitude thresholding is applied to incorporates spatial context into the thresholding decision, and effectively enables small isolated regions to be eliminated without fragmenting larger regions.

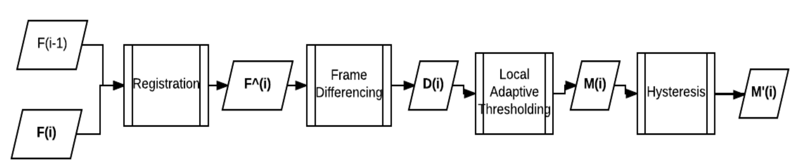

Fig. 2 Motion Detection Using Temporal Differencing.

KALMAN FILTER BASED Motion Detection

Temporal differencing has the following disadvantages: it tends to cause small holes; it cannot detect the complete shape of a moving objects; and in the case of a non-stationary camera, any error caused by inter-frame registration in the background can be easily classified as the foreground by temporal difference. To solve this problem, a block-based KF motion tracking solution Figure 3 is proposed to add robustness to the moving object detection by improving the reliability of correctly detected objects while reducing the noise presence at the same time. A Bottom-up Variable Block Size (BVBS) based block matching method is employed to identify the existence of possible moving object blocks first. Then an adaptive Kalman filters is proposed to track motion trajectory of the target so that one can verify the correctness and refine the object location estimates, and even reinitialize tracking in the case of target loss.

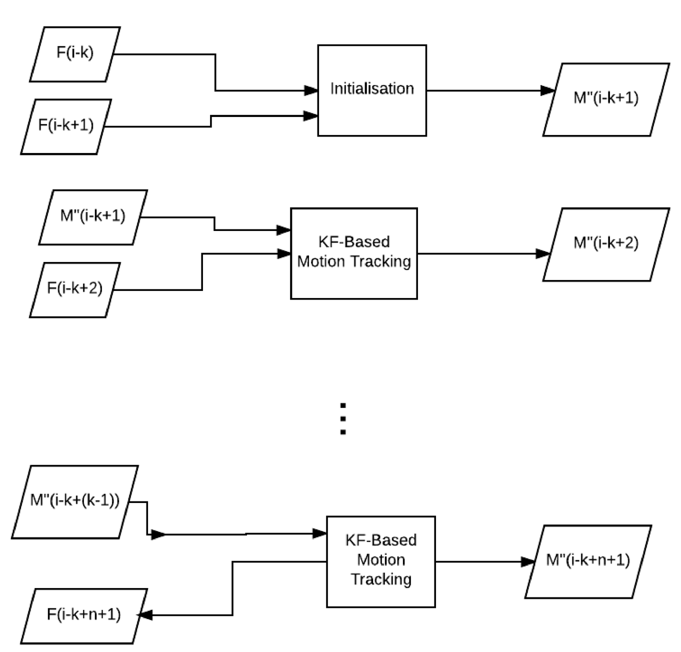

Fig. 3 Motion Detection Using Kalman-Filter based Block Tracking

Fusion of Motion Detection Algorithms for Region Recovery

Referring to the block diagram Figure 4, the motion results from temporal differencing and motion tracking, and segmented image regions are re-inspected to group moving regions from motion data. The moving region recovery process begins by the processing of each individual frame. To represent an image in a more compact and perceptually meaningful way, an efficient colour image segmentation algorithm is proposed to group pixels in the image into coherent atomic regions at first. Then the motion activity within each segmented region can be calculated based on the two motion detection algorithms respectively. At last, the motion information from two motion detection algorithms are successfully fused together and used as the main criterion to identify the moving regions.

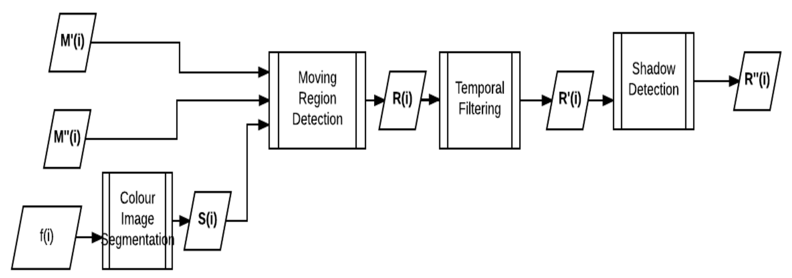

Fig. 4 Fusion of Motion Detections

Experimental Data Set

Since there are no existing video datasets available specifically designed for video object segmentation under non-stationary camera with affine camera motions, 5 video sequences under camera translation and 4 video sequences under rotation in Figure 5 are used in our experiments. The image size for each video sequence is 896×504 pixels. The results of algorithms are compared to ground truth images which are obtained from manual segmentations done by human users. In our experiments, each type of video sequences has been further tested under varying scaling, rotation, and translation camera transformations.

Fig. 5 Testing Videos with camera translation in (a)-(e) and with camera rotation in (f)-(i) (a) Pedestrian 1_T (b) Pedestrian 2_T (c) Traffic_T (d) Carpark_T (e) Pedestrian 3_T (f) Pedestrian 4_R (g) Traffic_R (h) Pedestrian 5_R (i) Hopspital_R

Comparing Methods in Experiment

The following methods are compared with proposed method: global temporal differencing method proposed by Rosin [1], local temporal differencing method proposed by Aach in [2], difference accumulation method proposed by Leng [3], and hybrid method of temporal differencing and optical flow method proposed by Tian [4]. To evaluate the similarity between the segmentation and the ground truth, F_1 score, as a trade-off between precision and recall, is used. where TP is the number of the true positives pixels which are correctly classified foreground pixels. FP is the number of background pixels, wrongly classified as foreground pixels. FN is the number of foreground pixels, wrongly classified as background pixels. 10 pairs of consecutive frame in each sample video are tested with those algorithms and F_1 score from each pair against a ground truth is calculated.

$$precision = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \quad recall = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}},$$

$$F_1 = 2\frac{\mathrm{precision} * \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}},$$

Fig. 1 Overall methodology

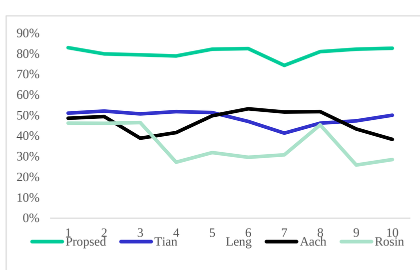

Figure 6 F1 score from Pedestrian 2_T

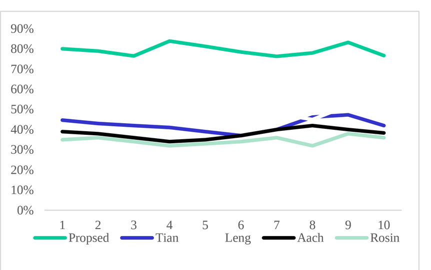

Figure 7 F1 score from Traffic_T

Experiment results measured with F_1 score from moving vehicles and pedestrian video sequences are shown in figure 6 and 7. It shows that proposed method achieved highest F_1 score in both test cases, close to 80%. Rosin's method has the worst performance overall due to its global thresholding. Aach's local method works slight better in comparison to Rosin's method. However, both methods have poor performance in detecting complete shapes of moving objects as shown in Figure 8 and 9. Also as shown in figure 8, both methods have poor performance in object localization along the boundaries of pedestrian due to registration noise. Leng's method detects slow motion of pedestrian by accumulating past N frame differences but such accumulation also expands the actual changed region when the object is moving fast, such as in figure 9(d). Tian's method has better detection rate in terms of F measure but still cannot track the motion accurately if objects stop, are occluded, or move fast as shown in figure 8(e) and figure 9(e).

Motion Segmentation Results

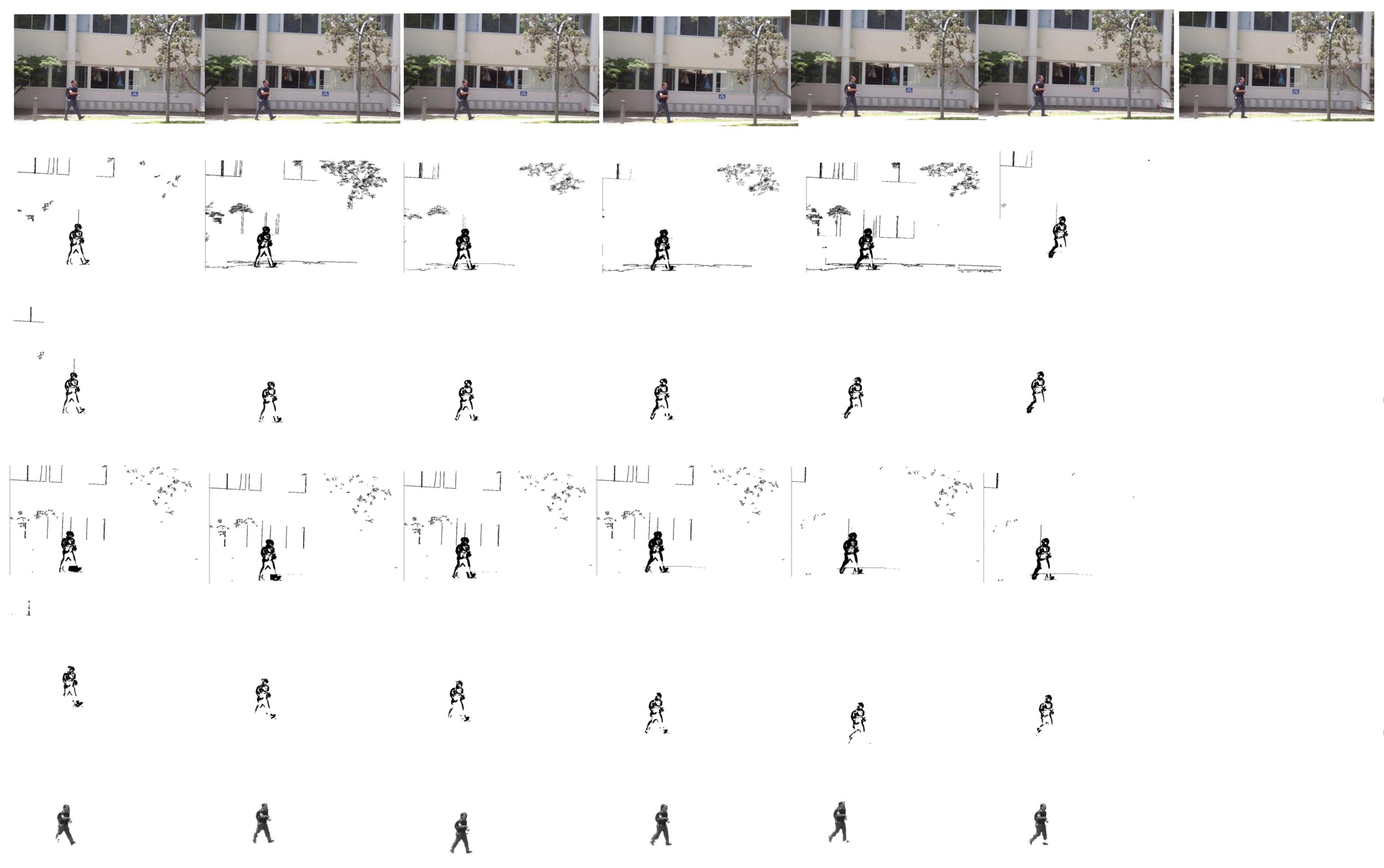

Figure 8: comparison of motion segmentation results from Pedestrian 2_T
(a) original frames 11-17 (b) Rosin's method (c) Aach's method (d) Leng's method (e) Tian's method (f) Proposed approach.

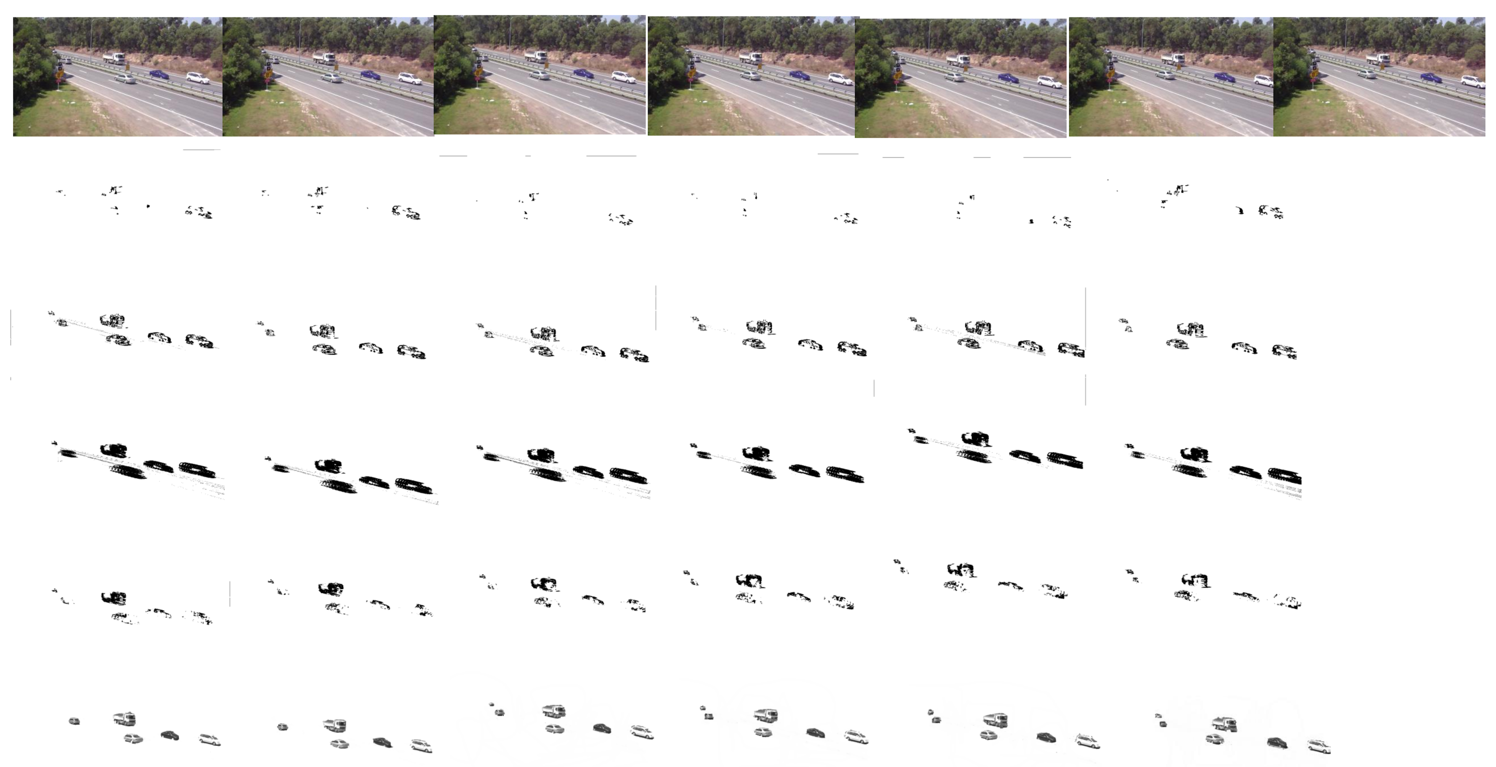

Figure 9: comparison of motion segmentation results from Traffic_R
(a) original frames 14-17 (b) Rosin's method (c) Aach's method (d) Leng's method (e) Tian's method (f) Proposed approach.

Conclusion

The proposed algorithm separates the background interference and foreground information effectively with non-stationary camera and detects the local moving object accurately. It addresses the issues of uncertainty of motion, robustness to noise presence. The effectiveness of the proposed algorithm to robust detect salient motion is demonstrated for a variety of real environments.

References

[1] P. Rosin, "Thresholding for change detection," in Sixth International Conference on. IEEE omputer Vision, 1998.
[2] T. Aach,, A. Kaup and R. Mester, "Statistical model-based change detection in moving video., 31(2), pp.165-180.," Signal processing, 1993.
[3] L. Bing and Q. Dai, "Video object segmentation based on accumulative frame difference," in Picture Coding Symposium , 2007.
[4] Y. Tian and A. Hampapur , "Robust salient motion detection with complex background for real-time video surveillance," in Seventh IEEE Workshops In Application of Computer Vision, 2005.