EFFICIENT VIDEO CODING IN REGION PREDICTION IN ONLINE VIDEO SURVEILLANCE

Shivprasad P. Patil[1], Rajarshi Sanyal[2], and Prof. Ramjee Prasad[1]
[1] Center for TeleInFrastruktur, Aalborg University, Aalborg, Denmark.
[2] Belgacom International Carrier Services, Brussels, Belgium.

Abstract – *There is a recent trend to implement video coding techniques for video surveillance. Various coding techniques had been proposed to enhance estimation accuracy in progressive video coding. As the conventional coding approaches have constraints to optimize image processing for moving bodies whilst the background is typically static. The problem magnifies in case of rotating cameras. The challenge in this case is to segregate the dynamic entities while the background rotates at a fixed pace. This paper presents an approach for the improvement of error free coding in video surveillance by employing rotating cameras. A least mean estimator approach is used and the recurrent full search motion estimator logic is defined for the prediction of foreground moving elements from the video sequence, which comes from a rotating sensor. The benefits that we derive from this method are more accurate detection of the actual moving object, thereby, reducing data redundancy by eliminating the nonessential background information.*

Keywords: video surveillance, least mean estimator, motion estimator, rotating camera, FS-BMA.

1 Introduction

The development of digital video technology has made it possible to use digital video coding in various applications such as teleconferencing, digital broadcast codec, video telephony, and video surveillance etc. In purview of surveillance related applications, video coding finds its use in particularly traffic video surveillance, where there is an urge to optimize image processing and frame reduction given the channel capacity constraints of a typical city traffic surveillance network. With the state of the art method, the background is also considered as moving. There is no process to discriminate the background from a truly dynamic object (moving people, cars). We propose a new technique by, which we can segregate/discriminate the background from dynamic objects in the video frames. We treat dynamic objects as the intelligence and remove the non essential background information from the subsequent video frames. Thus, the background information is not transmitted repeatedly, but only the useful information pertaining to moving objects is conveyed. With this method we can leverage substantial benefits like reducing transmission overheads, and storage etc. of a video surveillance system by employing rotating cameras. To improve the true detection rate of moving objects and thereby, reducing the bandwidth of such a new application, in this work, a new coding technique has been proposed. Most video surveillance systems rely on the ability to detect moving objects in the video stream. Therefore, object detection remains an important information extraction step in a wide range of computer vision applications. Each image is segmented by automatic image analysis techniques. This should be done in a reliable and effective way in order to cope with unconstrained environments, non stationary background, and different object motion patterns. Furthermore, different types of objects are manually considered e.g., persons, vehicles, or groups of people. Many algorithms have been proposed for object detection in video surveillance applications. They rely on different assumptions e.g., statistical models of the background [1,2,3], minimization of Gaussian differences [4], minimum and maximum values [5], adaptivity [6, 7] or a combination of frame differences and statistical background models [8]. Two approaches have been recently considered to characterize the performance of video segmentation algorithms: pixel-based methods, and template based methods or object-based methods. In Pixel based methods we thrive to detect all the active pixels in a given image. Therefore, the problem of object detection is formulated as a set of independent pixel detection. The algorithms can therefore, be evaluated by standard measures used in the Communication theory e.g., misdetection rate, false alarm rate, and receiver operating characteristic (ROC) [9]. Several proposals have been made to improve the computation of the ROC in video segmentation problems e.g., using a perturbation detection rate analysis [10] or an equilibrium analysis [11]. The usefulness of pixel-based methods for surveillance applications is questionable since we are interested in the detection of object regions (in our case, a moving car is an object region), and not in independent pixel detection. The computation of the ROC can also be performed using rectangular regions selected by the user, with and without moving objects [12]. This improves the evaluation strategy since the statistics are based on templates instead of isolated pixels. As far as the object based segmentation concept is concerned, we do the evaluation of the object of interest. In this approach, most of the works aim to characterize the object on the basis of colour, shape, [13, 14, 15] or area based performance evaluation [16]. This approach is instrumental to measure the performance of image segmentation methods for video coding and synthesis, but it is not usually used in surveillance applications.

It is found that, in video surveillance applications, stationary cameras are often employed to form a network. Due to technological advancement and the cost factor, the rotating camera is finding its place in these applications. Employment of a rotating camera reduces the number of cameras to be installed, thereby, reducing installation and maintenance cost. Also, bandwidth requirement overheads are reduced.

It is imperative that video coding techniques are required for video surveillance especially, for city traffic surveillance where we have channel bandwidth restrictions and processing resource constraints. The video coding technique that we propose, aims to reduce the redundancy and hence, can be termed as channel coding. Motivation for the use of the video coding approach for segmentation purposes lies in more accurate detection of false motion arising due to the rotation of the camera than the other existing methods, including a statistical one. In the conventional approach of the coding technique, two successive frames are compared for estimation of motion. This is referred to as Full Search Block Matching (FSBMA). As our sensor is rotating, road side buildings also seem to be moving, which we call as false motion. The application of FSBMA does not deal with the rejection of false motion, thereby, reducing the accuracy of true motion detection. Hence, there is scope to improve the existing method. To achieve the objective of improvement in the existing coding algorithm, in this work the Recurrent Full Search Block Matching Algorithm(R-FSBMA) approach is proposed.

Basically, in video coding, for finding moving elements, two successive frames are compared, which is also called as the block matching algorithm (BMA). The pixels, which are not matching are taken as moving elements or motion vectors (MV). Obtained MV's are considered for further processing. However, in our case a stable background is also appearing as moving due to camera rotation. Also, we have to look into the movement pattern of moving objects, which can be linear or nonlinear. Hence, to find the actual moving object, conventional two successive frame comparisons are not effective. Hence, the direct application of BMA will not result in correct MV estimation. So, we have to go for comparing frames in a recursive manner, where in we have compared the current frame with its successive frame to detect MV. Here, we record the variations as linear and non- linear. As stable objects will have linear variation with frame, we can reject such coefficients, thereby, eliminating stable pixels falsely taken as moving.

The objective of our project is as follows:

1. Develop a new recurrent block matching approach for more effective and accurate detection of a moving object.
2. Apply an adaptive filtration method to attenuate the noise of the video sample generated from multi-segmented intersection.

The novelty of the proposed work is, rather than searching the successive video frames for motion, we go for searching the motion component in a set of frames, thereby, presenting a new recursive coding technique. Also, we are able to extract a moving object when a video file contains actual and false motion. It is worth to mention here, that our algorithm is not only supporting for a complex scene, where there are multiple segments at the intersection, but also for jitter in a rotating sensor.

The rest of the paper is outlined into six sections. In section 2, a conventional video coding system and its application to video surveillance is presented. In section 3, the error estimator logic is presented. Motion prediction technique is outlined in section 4. The proposed recurrent-FSBM algorithm is described in section 5. The observations obtained for the proposed work are presented in section 6. The conclusion of the developed work is outlined in section 7.

2 Video coding and application

Many applications have been proposed based on the assumption that an acceptable quality of video can be obtained for a bandwidth of about 1.5 Mbits/second (including audio).Computer vision systems often depend on the ability to distinguish or describe a moving object in an image space. An algorithm is designed to segment a moving foreground based on the block-matching motion method and recursive tracing of the resulting motion Vectors. The objective of this project is the creation of an algorithm that will separate moving foreground from a stationary background in a given video sequence. The separation of true motion associated with foreground, can further be utilized for sending the null values for stationary background, thereby reducing transmission bandwidth. The selection of a motion estimator model represents the first step in the problem. Gradient-based methods such as optical flow have shown high performance, but generally come with an increased computational overhead than block-based matching. The disadvantage of block methods is an expected loss of sharpness at the edge regions marking the boundary between foreground and background. Regardless of the motion estimator, careful attention must be paid to noise effects when estimating motion. Faulty motion vectors due to image noise can lead to visually unpleasant effects such as isolated background blocks in the resulting segmented image. Noise-reduction filters may be used to alleviate this problem. Another method is to examine the resulting mean-squared error of the known zero-motion vector regions. If any error exists, it must be due to the presence of noise in a particular image sequence. Accurate knowledge of all the motion vectors in a sequence theoretically, provides the means to segment the images into pixels associated with a moving object and pixels associated with a rigid background. The algorithm for tracing motion vectors throughout the sequence is highly recursive and can be computationally expensive, depending on the number of non-zero motion vectors present, which could be optimized in future works.

The moving frames are generally represented as a sequence of multiple frames. These frames are static in nature when isolated. All these frames together create a moving image as shown in figure 1. On a closer observation it can be seen that most of the moving frames have got correlated pixels among the successive frames. The transmission of these correlated pixels for low bit rate application is a very difficult task. To overcome this difficulty the moving image can be isolated from the stationary elements and can be transmitted isolately for more efficient low bit rate application.

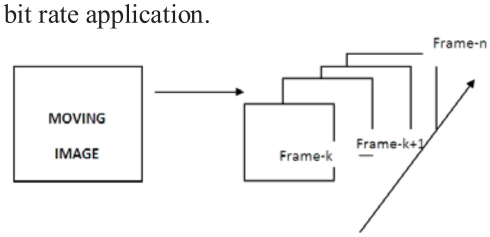

Figure 1: multi-frame representation of a video sample.

Generally, an image has two layers, namely-foreground and background. In case of a moving image, there are three possibilities:
1) Foreground and background moving
2) Foreground stationary and background moving
3) Background stationary and foreground moving

In the proposed work, the first case of both foreground and background moving is considered.
Normally, in video surveillance, vehicles and people are moving, whereas the other objects are stable. So, vehicle and people motion can be considered as true motion. But, in our case camera rotation gives a false motion to the hoardings, building etc. So, wherein conventional video coding is proposed for true motions only, a mixed model of true and false motion estimation has to be devised.
As, it is difficult to apply the conventional coding for video processing, in this paper a new coding is presented, which is briefed in the earlier section. Prior to the estimation approach, de-noising of a noisy video sample is required. In this work, an adaptive filtration based on the LMSE approach is used.

3 Denoising using LMS algorithm

The Least Mean Square (LMS) algorithm is an adaptive algorithm, which uses a gradient-based method of steepest decent. The LMS algorithm uses the estimates of the gradient vector from the available data. LMS incorporates an iterative procedure that makes successive corrections to the weight vector in the direction of the negative of the gradient vector, which eventually leads to the minimum mean square error [17]. Compared to other algorithms the LMS algorithm is relatively simple; it does not require the correlation function calculation nor does it require matrix inversions.
From the method of steepest descent, the weight vector equation is given by;

$$w(n+1)=w(n)+\mu[-\nabla (E\{e^2(n)\}] \quad (1)$$

Where μ is the step-size parameter and controls the convergence characteristics of the LMS algorithm; $e^2(n)$ is the mean square error between the output $y(n)$ and the desired output, which is given by,

$$e^2(n)=[d(n)-w(n)x^T(n)]^2 \quad (2)$$

The gradient vector in the above weight update equation can be computed as

$$\nabla (E\{e^2(n)\})= 2Rw(n) -2r \quad (3)$$

Where R is an autocorrelation of input signal $x(n)$ and r is a cross correlation between the desired response and input. In the method of steepest descent the biggest problem is the computation involved in finding the values r and R matrices in real time. The

LMS algorithm simplifies this problem by using instantaneous values;

$$R = x(n)x^T(n) \quad (4)$$

$$r = d(n)x(n) \quad (5)$$

Therefore, the weight update can be given by the following equation,

$$w(n+1) = w(n)+\mu x(n)[d(n)-x^T(n)w(n)]$$
$$= w(n)+\mu x(n)e(n) \quad (6)$$

The LMS algorithm is initiated with an arbitrary value w(0) for the weight vector at n=0.

The successive corrections of the weight vector eventually leads to the minimum value of the mean squared error.

Therefore, the LMS algorithm can be summarized in the following equations;

$$\left.\begin{aligned} y(n) &= w^T x(n) \\ e(n) &= d(n)-y(n) \\ w(n+1) &= w(n)+\mu x(n)e(n) \end{aligned}\right\} \quad (7)$$

This computed weight provides an optimal value for noise elimination. Over this de-noised video sample a new motion estimation approach is proposed. This approach is an extension to the FS-BMA approach.

4 Motion prediction

The motion estimation and compensation technique has been widely used in video compression due to its capability of reducing the temporal redundancies between frames. Most of the algorithms developed for motion estimation so far are block-based techniques, called the block-matching algorithm (BMA). In this technique, the current frame is divided into a fixed size of blocks, and then each block is compared with candidate blocks in a reference frame within the search area [18,19]. The widely used approach for the BMA is the full search BMA (FSBMA), which examines all the candidate blocks within the search area in the reference frame to obtain a motion vector (MV). The MV is a displacement between the block in the current frame and the best matched block in the reference frame in horizontal and vertical directions. The motion estimation algorithm is performed with a variable size of search area depending on block types varying from an 8x8 block to the complete frame. The video sequences for low bit-rate video coding applications such as videophone and video-conferencing have some restrictive motion characteristics. A block in a specific region in the previous frame can belong to the same region at that position in the current frame; in other words a block in the background region may lie in the background region in the current frame. The changing block shows the percentage of the difference from the background to the active region or vice versa. The other labels mean that the block types are the same in successive frames. In all video sequences, the percentage of background blocks in the successive frames is very high. The changing blocks occupy only 30% below, meaning that the motion field of each block is very high in the successive frames for the other blocks. Also, the pattern of distribution is very similar without regard to video sequences. It is shown that the temporal correlation between the successive frames is very high, that is, if a block in the previous frame belongs to background regions or active regions, the block, which is located in the same position in the current frame may be classified as a background block or active moving block, respectively, with a strong probability.

The basic idea of block matching is depicted in the figure 2, where the displacement for a block (LxL) in frame K (the present frame) is determined by considering a window of size [(L+2W) x (L+2W)] in frame K+1 (the search frame) for finding the location of the best-matching block of the same size. The search is usually limited to $(L+2W)^2$ region called the search window.

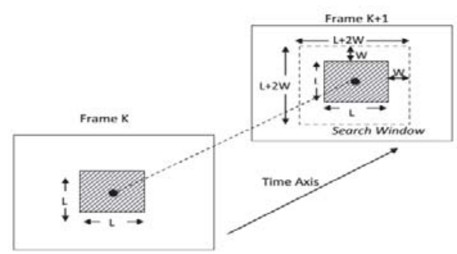

Figure 2: Matching approach.

Block matching algorithms differ in

- The matching Criteria
- The search strategy
- The determination of block size

Matching criteria:

The matching of the blocks can be quantified according to various criteria of, which the most popular and less expensive is mean absolute difference (MAD), given by equation (8). Another criteria mean square error (MSE), is given by equation (9).

$$MAD = \frac{1}{L^2}\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\left|P_{ij} - S_{ij}\right| \quad (8)$$

$$MSE = \frac{1}{L^2}\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}(P_{ij} - S_{ij})^2 \quad (9)$$

Where L is the side of the block, and Pij and Sij are the pixels being compared in the block from the present frame and the block from the search frame, respectively. Normally, MSE is not used, as it is difficult to realize the square operation in hardware.

The determination of block size:

The selection of an appropriate block is essential for any block-based motion estimation algorithm. There are conflicting requirements on the size of the search blocks. If the blocks are too small, a match may be established between blocks containing similar gray level patterns, which are unrelated in the sense of motion. On the other hand, if the blocks are too large, then the actual motion vectors may vary within a block, violating the assumption of a single motion vector per block. The block size for the proposed design is calculated by performing continuous testing, taking a different combination of frame sizes with different frame skips.

5 Recurrent estimation logic

Ideally, the tracer recognizes this and segments the region over all the frames, and not just the frames in which it moved. In general, this stage forms the computational bottleneck of the overall algorithm.

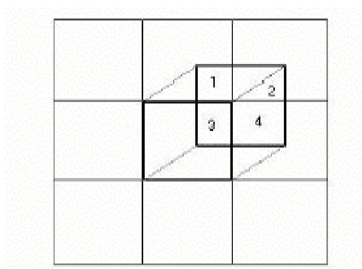

Figure 3: Recurrent searching of an overlapped pixel.

Tracing motion vectors lend itself naturally to a recursive solution. Each block with non-zero motion vectors in each frame represents a “seed” call to the tracing function. A moving block will, in general, translate into a region corresponding to four blocks. The tracing algorithm begins with a seed call. This seed block will move into as many as four other blocks, and each of these blocks is recursively called by the tracing function. The purpose of the tracing function is simply to identify the appropriate moving pixels based on the motion vectors and block regions, and then to seed further calls to it. Motion tracing has a straightforward solution only in one direction temporally. In other words, tracing must be done in both the forwards and reverse temporal directions for best segmentation results.

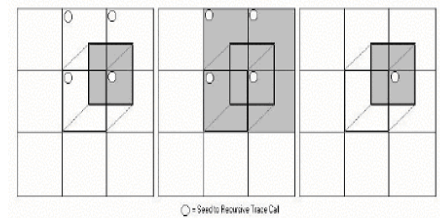

Figure 4: The process of searching in the frames using R-FSBMA.

For any moving block only the pixels corresponding to that moving block are associated with motion, but all four regions impinged by the block are seeded to the successive tracing call. This is the most accurate approach, but also the most computationally burdensome. The second approach is to seed all four blocks as well, but to treat all pixels within the four seeded blocks as having moved rather than just the actual moving pixels. This approximation greatly simplifies the tracing algorithm, and also increases the algorithm efficiency dramatically, since a block that is seeded to the tracing function need not be ever seeded again.

A final approach is to mark all moving pixels as in the general case, but to only seed the block corresponding to maximum overlap. If there are equal overlaps, then multiple blocks are seeded. Although this variation only approximates the tracing problem, it can be much faster since each trace call usually, only seeds one recursive call rather than four. In the most general case, the tracing algorithm runs slow. For improved speed, motion vectors are computed not between each frame, but between every n frames and tracing is done on this smaller set of motion vectors.

6 Simulation observation

To observe the developed work a video sequence is read, wherein a set of video frames is selected and the tracing algorithm is applied. The obtained results are as shown below:

The video file is captured at an elevated location at the center of a cross road, and the sensor is rotated for 360 degrees to capture the traffic images. The video sequence shows the vehicle movement and other static regions in the vicinity. The video sample is captured at 25fps, with a resolution of 272x 352.

Figure 5: Extracted Video frames from the video file.

A set of successive frames is extracted from the captured video sequence. Further, they are used for processing. The extracted frames are illustrated in figure 5.

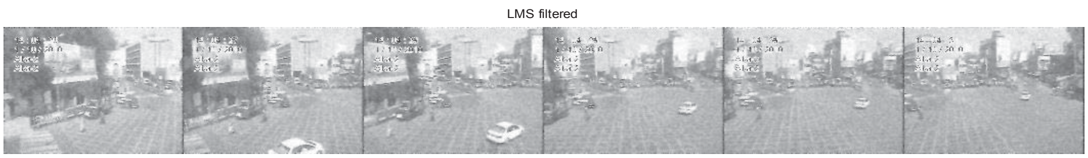

Figure 6: De-noised sample after LMS filtration.

It is required to eliminate the noises so as to achieve higher accuracy in the estimation of moving objects.
To achieve this, a conventional adaptive LMS filter is applied to denoise the affected sample. The obtained result for such filtration is given in figure 6. It is observed that a higher visual quality is achieved with this approach.

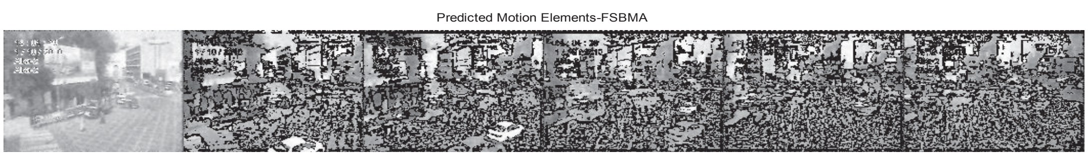

Figure 7: Predicted region by FSBMA approach.

Over the filtered sample, a full search block matching algorthm is applied to compute the moving element. It is observed that as the camera is in a rotating position, the background objects will also change their corresponding position for each frame. Hence, such components are also detected as moving elements in predicted video frames.

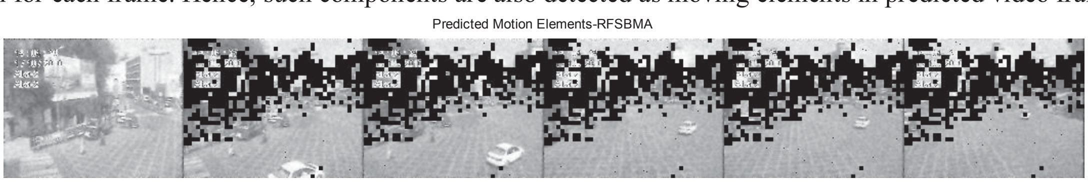

Figure 8: Predicted region after recurrent tracing.

In the case of the proposed Recurrent FSBMA approach, due to successive computation of Motion vector in both inter and intra frames, the elimination of a background element is possible. Hence, this approach detects the moving elements more accurately than the FSBMA approach, which is shown in figure 8.

7 Conclusion

A new coding approach for video surveillance is presented. The incorporation of new coding algorithms for denoising using the least mean error estimator results in higher estimation probability. This denoising approach is a dynamic model and hence, is suitable for all type of system interface. The proposal of recurrent motion estimation logic results in an improvement in the detection of a moving object in a video sequence, generated from a rotating camera. In the sequel, the authors would like to conclude that, the proposed work based on the recurrent block matching approach is found to be more effective and accurate in the field of video surveillance, by employing rotating sensors. It is worth to mention that, our proposal has applicability in the metropolitan surveillance network with wired or wireless rotating camera implementation, where bandwidth and processing resource optimization are the key challenges.

Acknowledgements

The datasets of the traffic at the intersection is kindly provided by Pune Municipal Corporation, India.

7 Reference

[1] Y. Ren, C. Chua, and Y. Ho, "Statistical background modeling for non-stationary camera," Pattern Recognition Letters, 24(1-3):183– 196, January 2003.
[2] C. Stauffer, W. Eric, and L. Grimson, "Learning patterns of activity using real-time tracking," IEEE Trans. Pattern Anal. MachineIntell., vol. 22, no. 8, pp. 747–757, August 2000.
[3] T. Bouwmans, F. Baf and B. Vachon. "Background modeling using mixture of gaussians for foreground detection - A survey." in Recent Patents on Computer Science 1, 3, pp 219-237,2008.
[4] N. Ohta, "A statistical approach to background suppression for surveillance systems," in Proceedings of IEEE Int. Conference onComputer Vision, pp. 481–486, 2001.
[5] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: Who? when? where? what? a real time system for detecting and tracking people,"in IEEE International Conference on Automatic Face and Gesture Recognition, pp. 222–227, April 1998.
[6] M. Seki, H. Fujiwara, and K. Sumi, "A robust background subtraction method for changing background," in Proceedings of IEEE Workshop on Applications of Computer Vision, pp. 207–213, 2000.
[7] D. Koller, J. Weber, T. Huang, J. Malik, G. Ogasawara, B. Rao, and S. Russel, "Towards robust automatic traffic scene analysis in real-time," in Proceedings of Int. Conference on Pattern Recognition, pp. 126–131,1994.
[8] R. Collins, A. Lipton, and T. Kanade, "A system for video surveillance and monitoring," in Proc. American Nuclear Society (ANS) Eighth Int. Topical Meeting on Robotic and Remote Systems, Pittsburgh, PA, pp. 25–29,April 1999.
[9] H. V. Trees, "Detection, Estimation, and Modulation Theory". John Wiley and Sons, 2001.
[10] T. H. Chalidabhongse, K. Kim, D. Harwood, and L. Davis, "A perturbation method for evaluating background subtraction algorithms," in Proc. Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS 2003), Nice, France, October 2003.
[11] X. Gao, T.E.Boult, F. Coetzee, and V. Ramesh, "Error analysis of background adaption," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 503–510, 2000.
[12] F. Oberti, A. Teschioni, and C. S. Regazzoni, "Roc curves for performance evaluation of video sequences processing systems for surveillance applications," in IEEE Int. Conf. on Image Processing, vol. 2, pp. 949–953, 1999.
[13] J. Black, T. Ellis, and P. Rosin, "A novel method for video tracking performance evaluation," in Joint IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS), Nice, France, pp. 125–132,2003.
[14] P. Correia and F. Pereira, "Objective evaluation of relative segmentation quality," in Int. Conference on Image Processing, pp. 308–311, 2000.
[15] C. E. Erdem, B. Sankur, and A. M.Tekalp, "Performance measures for video object segmentation and tracking," IEEE Trans. Image Processing, vol. 13, no. 7, pp. 937–951, 2004.
[16] V. Y. Mariano, J. Min, J.-H. Park, R. Kasturi, D. Mihalcik, H. Li, D. Doermann, and T. Drayer, "Performance evaluation of object detection algorithms," in Proceedings of 16th Int. Conf. on Pattern Recognition (ICPR02), vol. 3, pp. 965–969,2002.
[17]B.Widrow and S.D.Stearns, Adaptive Signal Processing, Prentice-Hall, Englewood Cliffs,NJ, 1985.
[18] Jianhua Lu , Liou, M.L. , "A simple and efficient search algorithm for block-matching motion estimation, " IEEE Trans. Circuits and Systems for Video Technology, vol. 7 , pp.429-433, Apr 1997.
[19]Chun-Ho Cheung, and Lai-Man Po, "A Novel Cross-Diamond Search Algorithm for Fast Block Motion Estimation", IEEE Trans. Circuits And Systems For Video Technology, vol 12., no. 12, pp. 1168-1177, December 2002.