Hybrid Background Subtraction in Video using Bi-level CodeBook Model

Soumya Varma

Federal Institute of Science and Technology,

Mahatma Gandhi University, Kerala

India

summu121@gmail.com

Sreeraj M

Federal Institute of Science and Technology,

Mahatma Gandhi University, Kerala

India

sreerajtkzy@gmail.com

***Abstract*—Detection of Objects in Video is a highly demanding area of research. The Background Subtraction Algorithms can yield better results in Foreground Object Detection. This work presents a Hybrid CodeBook based Background Subtraction to extract the foreground ROI from the background. Codebooks are used to store compressed information by demanding lesser memory usage and high speedy processing. This Hybrid method which uses Block-Based and Pixel-Based Codebooks provide efficient detection results; the high speed processing capability of block based background subtraction as well as high Precision Rate of pixel based background subtraction are exploited to yield an efficient Background Subtraction System. The Block stage produces a coarse foreground area, which is then refined by the Pixel stage. The system's performance is evaluated with different block sizes and with different block descriptors like 2D-DCT, FFT etc. The Experimental analysis based on statistical measurements yields precision, recall, similarity and F measure of the hybrid system as 88.74%, 91.09%, 81.66% and 89.90% respectively, and thus proves the efficiency of the novel system.**

Keywords— Background Subtraction,Codebook Model,Foreground Detection DCT, FFT

I. INTRODUCTION

Object Detection in Video has enormous applications in the field of Target Recognition, Security Surveillance, Intelligent Monitoring, Pedestrian Detection, Object Tracking etc. The identification of an object in a video could be made very easily, if the redundant video background is eliminated. In Computer Vision related applications, Background Subtraction or Filtering is a wide research area which focuses on subtracting the unessential "background area" from the "foreground region". Any background subtraction methodology must not demand much processing time and memory usage. Codebooks are used to represent compressed form of information without demanding much processing time and memory usage.

The remaining of the work is organized as follows. Section II describes the Literature Review; Section III elucidates the System Architecture. In Section IV, the details of the Implementation are explained. Results and Evaluations are described in Section V, and Conclusion and Future Scope of the work are discussed in Section VI.

II. LITERATURE REVIEW

The process of Background Filtering is carried out by analyzing the video frame-by-frame, maintaining a temporal continuity between the consecutive frames of a video. The approaches for Background Subtraction on the basis of processing each frame could be broadly classified as pixel-based and block-based methods. In the former approach, the pixels of a frame that provide detailed information is taken care of and decision is made on each pixel-whether it belongs to foreground or background. But in the latter approach, each frame segmented to different blocks of fixed size, is taken into account and a decision is adopted for a block so that it is either classified as foreground or background.

A. Pixel based methods

1) Frame Differencing

The simplest method of Background Subtraction is the Frame Differencing strategy [1], in which the pixel characteristics of a frame are subtracted from its previous frame. In this approach, a pixel is classified as foreground if the absolute difference between the pixel values in two successive frames is greater than a predefined threshold. The value for threshold must be carefully chosen to achieve accurate results. However, this method offers only coarse form of foreground which is least precise.

For any two frames, $frame_i$ and $frame_{i+1}$, pixel P at position (x, y), is classified as foreground or background as follows

$$P_{(x,y)} = \begin{cases} P_{fg} \text{ if } |frame_i - frame_{i+1}| > Th \\ P_{bg} \text{ otherwise} \end{cases} \quad (1)$$

where Th is the predefined threshold for classification. Here the value that the parameter Th is highly sensitive in obtaining the accurate output.

2) Background Modeling

A background subtraction strategy compares an observed image with a background image. This process,

978-1-4799-2259-14/$31.00©2014

divides the observed image into two disjoint sets of pixels that together comprises the entire image:the foreground pixel set and the background pixel set.

In Background Modeling approach, the background region is estimated from the sequence of frames by constructing a reference background image. Most of the background subtraction strategies make use of probability distribution to accurately estimate the reference model. Stauffer and Grimson proposed the Mixture Of Gaussians (MOG) [1,2] in which background modeling is accomplished by using multiple Gaussian distributions to represent each pixel. It has the advantage of effectively modeling the background even for stationary objects and better adaptability to background model. However, It is not capable of handling shadows, dynamic backgrounds etc. An easy and simple approach is adopted in [3] in which the foreground objects are detected considering the correlations between RGB channels. Here, a thresholding approach is based on color distortion and brightness distortion is being used. But the method is dependent on the background still frame to be loaded with each video, and is highly sensitive to thresholds. A typical Background Subtraction procedure is depicted in the following figure, Fig.1.

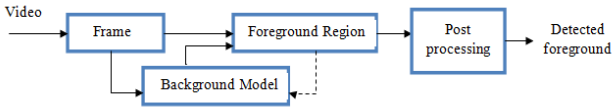

Fig.1.A typical Background Subtraction Procedure

3) Background Modeling using heterogeneous features

In [4], Han et.al use a set of visual features to effectively model background. In their work, a multiple feature based pixel-wise background modeling and subtraction technique is introduced and classification is performed by combining generative (using Kernel Density Approximation) and discriminative (using Support Vector Machine) techniques. The set of heterogeneous features are color, gradient and 6 Haar-like features. The background is modeled using visual features and 1D KDA is the density function used to perform this. After modeling background, distinction for background and foreground pixels are made. The final classification of the pixels is based on the output of SVM.

4) Background Modeling and Subtraction using Probability based Framework

In [5], Chiu et.al views the background modeling and subtraction system as comprising of two modules namely: Initial background extraction and Object segmentation. The work is on color image frames. The major advantage of this probability based framework is that, it does not demand for large memory requirements as well as processing time. Also the threshold parameters used in the algorithm is automatically adjusted based on lighting variations, hence it is easy to update the initially extracted background.

5) Background Modeling and Subtraction using Codebooks

K.Kim et al [6] introduced a codebook based foreground-background segmentation algorithm that could be used in real-time with minimal memory consumption. Codebooks [6,7] are used to represent background model in a compressed form, and is efficient in terms of speed and memory use compared to state-of-the-art background subtraction techniques.

In [6], a pixel based codebook construction algorithm is introduced that works using a clustering algorithm like Linear Vector Quantization (LVQ). In order to segment a pixel as foreground or background two other relevant metrics like color and brightness difference are used. However, the method cannot handle dynamic background and sudden illumination changes.

In [7], foreground is detected by a hierarchical scheme combining the pixel based and block based codebooks. The block based stage makes use of Block Truncation Coding (BTC), a lossy compression scheme and use some mean intensity values to represent each non-overlapping blocks of an image in the video frame sequence. Pixel based stage makes use of the codebook model as described in [6].

B. Block based methods

Majority of the works in background subtraction is carried at pixel level, and a few deals with region based or patch based. [8] deals with block based foreground detection with a cascaded classifier sequence to make decision for the blocks. The cascade comprises three classifiers , 1) location specific Gaussian model probability measurement, 2) similarity measurement using cosine distance metric, 3) temporal correlation checker. A pixel is probabilistically classified either as foreground or background based on the count of the blocks that contain this specific pixel that are classified as foreground or background.

- *Descriptor Generators for block-based methods*

In block based method of background subtraction, the features that represent a block can be generated using the block descriptor. The standard descriptors are

1) Local Binary Pattern (LBP)

LBP [9] is a powerful kind of texture descriptor that is gray-scale invariant. The LBP histogram computed over a circular region around a pixel is used as the feature vector. Large scale texture primitives could be captured, if the neighborhood is extended.

2) Block Truncation Coding (BTC)

It's a lossy image compression algorithm which uses a fixed length compression method that uses a Q level quantizer to quantize a given region of the image.

3) Discrete Cosine Transform (DCT)

DCT [8] is a widely used signal/image compression technique that transforms an image from spatial domain to frequency domain. It helps to separate a frame into regions of varying importance.

Computing the 2D DCT

- apply 1D DCT Column-wise (Vertical)
- apply 1D DCT Row-wise to resultant column-wise DCT (Horizontal)

- or alternatively Horizontal to Vertical

4) Normalized Vector Distance (NVD)

NVD [10] provides a normalized measurement of the inter vector distance between a multi-valued vector pair.

III. SYSTEM ARCHITECTURE

The generic architecture of a system for Background Subtraction is depicted in the following figure, Fig.2.

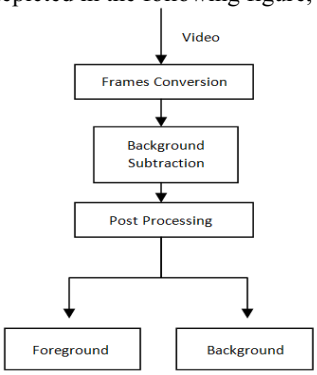

Fig.2.Generic System Architecture

A. Video to frame Segmentation

The video to frame segmentation module takes a video as input and produces a number of frames determined by the 'frames per second' property and 'duration' of the video. Mathematically, each frame of the video is represented as in (2).

$$X_n: n > 1 \ \& \ n = fps \times t, \ \& \ t > 1 \qquad (2)$$

where

- X_n is the nth frame
- n is the number of frames
- fps represents frames per second
- t is the time duration

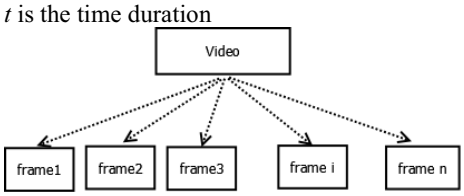

Fig.3.Video Segmentation Module

B. Background Subtraction

In this work, a two level Codebook Model is employed to achieve coarse-to-fine foreground detection efficiently. Block based processing has greater processing speed, though it has lower precision. Pixel based methods provide detailed information regarding the characteristics of the pixels and produce better True Positive Rate, however they take long processing time. The work focus on a coarse-to-fine foreground detection using a combination of Block based and Pixel based Stages. This hybrid method makes use of two codebooks namely Block-Based CodeBook and Pixel Based CodeBook. The Block based codebooks store the features of each blocks of a frame, whereas the Pixel based codebooks store pixel features in each frame. These codebooks are constructed during the training phase and are used in the subsequent Detection phase. Figure 4 illustrates the architecture of Hybrid Background Subtraction Module.

The Background Subtraction procedure consists of two phases.

1) Training Phase

It is during the training phase that, both the Block based and Pixel based Codebooks are constructed. The first T frames of the video are used for training and the rest frames are used in Detection phase. During Codebook training, all T frames are considered at once. Block features, extracted using Block Descriptors, for the same block of entire T frames are used to generate a Block codebook for that block position. Likewise, for each block position, Block codebooks are constructed. Similarly, Pixel Intensity characters of a pixel at a particular position of entire T frames are used to construct Pixel codebook for that pixel position. Each pixel position has a Pixel codebook containing varying number of codewords. The constructed Codebooks are provided to the subsequent Detection phase for the purpose of foreground detection.

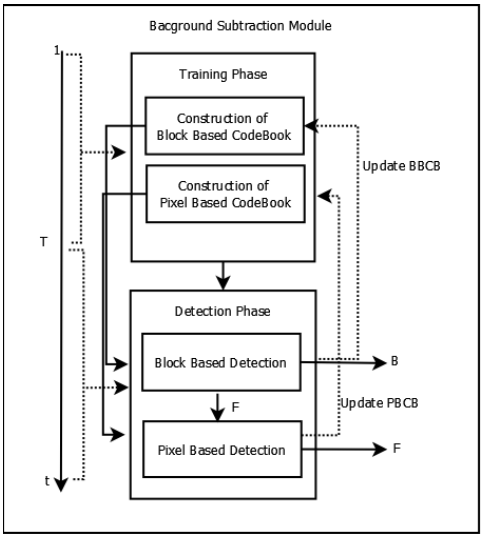

Fig.4.Background Subtraction Module

1. Block Based CodeBook Construction

The video frame is divided into non-overlapping blocks of size m x m. Feature Descriptor of each block is computed using 2D-Discrete Cosine Transform (2D-DCT).For each block, 4 strong DCT coefficients for each color channel (R,G,B) is extracted and thus a total of 12D-Descriptor represents a block. This feature vector of the block act as a codeword to the Codebook of the block position. Each block position has a Codebook consisting of C codewords from blocks residing in the same position of several frames. C varies for different codebooks.

The image sequence with M x N could be divided to blocks of size m x m. Each block of *Frame f(i)* is accessed by indexing it as *Block, B(p,q)*.The descriptor for each block is named as *Feature Vector*, *fv*.

The procedure of block based codebook training is as follows:

- Access block *B(p,q)* with m x m size of frame *f(i)*.

- Compute 2D Discrete Cosine Transform (DCT) of the accessed block, and retain 4 strong DCT coefficients per color channel.
- This *4* x *3*, *12D* descriptor of the block *B(p,q)* is the feature vector or codeword *fv(p,q)* which is to be added to the CodeBook *CB(p,q).*
- Each codewords are entered to the CodeBook *CB(p,q)* only when there is no match with existing codewords in the *CB(p,q),* after checking with the codebook entries.
- Thus for each block position *B(p,q)* ,a Codebook comprising of *C* codewords are stored.
- Totally *(M/m)* x *(N/m)* block based CBs are present for each frame *i* where *1<=i<=T.*

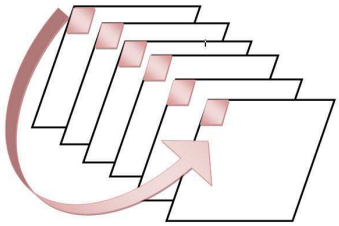

Fig.5.Block Based CodeBook

2. Pixel Based CodeBook Construction

During Pixel Based CodeBook construction, as in BlockBased CB construction, all *i* frames are considered at once where *1<=i<=T.* Here the intensity value of each pixel in the three color channels is accounted as a pixel vector *pv(x,y)*. So, for each pixel position in a frame a 3D feature vector is created which is the codeword to be entered to Pixel codebook at position *(x,y)* where *1<=x<=M, 1<=y<=N.*

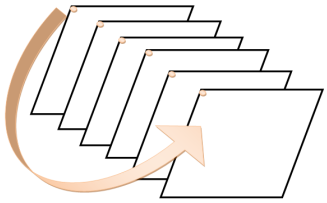

Fig.6.Pixel Based CodeBook

2) Detection Phase

Using the training frame sequence, Codebooks are already constructed. Now, using the Test sequence, from *T+1* to *t* foreground could be detected first by using Block-based detection method. Subsequently, this coarse foreground region is provided for the Pixel based detection where a fine detection is made. Each incoming block is checked against each codeword entry of the constructed codebook at the exact block position, to get a match. If a match is obtained, then the input block belongs to background region, otherwise it is a foreground region. The blocks that are marked as foreground by the block based stage is used by subsequent pixel based stage to make further refined foreground region. The results of this detection can be used to update both the pixel based and block based codebooks. Ultimately, the background and foreground could be separated in the test sequence.

The procedure is as follows:

- Each frame from *T+1* to *t* is processed one at a time to detect the foreground objects in that frame.
- Detection phase first uses Block Based CBs for Block Based CodeBook detection, and then Pixel Based Codebooks for Pixel Based Detection.
- Each block of the input frame is extracted and DCT descriptor is computed. This is then checked against all existing codewords in the CodeBook at position *B(p,q).*
- Match occurs if the block at hand belongs to background region. Update the CodeBook.
- If no match is encountered, then the block is likely to be a foreground region.
- Pixel Based Detection is then carried out to make fine refinements to the obtained coarse foreground.
- Thus a refined foreground is obtained using the proposed Hybrid CodeBook Model.

Match Function

$$match(fv_{input}, fv_{codeword}) = \begin{cases} 1 & if\ ((dist \times dist')/dim(dist)) < \lambda^2 \\ 0 & otherwise \end{cases} \quad (3)$$

$$where \qquad dist = (fv_{input} - fv_{codeword}) \quad (4)$$

In Detection phase Euclidean distance is used to find a match.

Codeword Updation

Block based and pixel based codeword updation are as follows

$$codeword_i = (1-\alpha) \times codeword_i + \alpha \times fv(p, q) \quad (5)$$

$$codeword_i = (1-\alpha) \times codeword_i + \alpha \times pv(x, y) \quad (6)$$

where α is the learning rate coefficient and α=0.05

C. Post Processing

The background subtracted image may consist of noise, incompleteness in shapes etc. Post processing operations are a solution to such problems. Figure 5 shows the schematic representation of the Post Processing Module. This module comprises Median Filter and Morphological Filter.

1) Median Filter

Median filtering is a nonlinear signal enhancement method used to remove salt and pepper noise and to smooth the signals by preserving edges. Window size 3 x 3 is used.

2) Morphological Filtering

Morphological filtering helps in maintaining the shape of the foreground object, by adding pixels or removing pixels in the detected foreground object. Morphological Close which is dilation followed by an erosion- using the same structuring element for both operations is used.

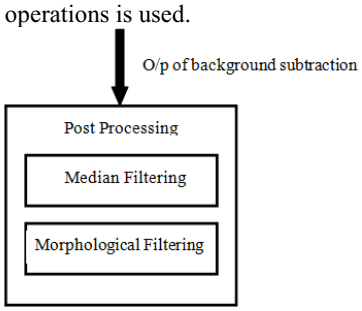

Fig.5.Post processing Module

IV. EXPERIMENTAL EVALUATION

The evaluation of the system is carried out using standard datasets like WAVING TREES[13], Datasets from I2R Database like CAMPUS[13], WATERSURFACE[13], FOUNTAIN[13] etc. Training of the system is carried out using the first half frames of these datasets, and Tested using the rest frames.

A. Statistical measurements for System Performance

The proposed hybrid system has been evaluated with Pixel based Codebook method, Block based Codebook methods and with Mixture of Gaussians. In order to evaluate the results, various statistical metrics like Precision, Recall, F-measure, and Similarity are used, which are as follows.

$$Precision=tp/(tp+fp) \quad (7)$$

$$Recall=tp/(tp+fn) \quad (8)$$

$$F\text{-}measure=2*((precision*recall)/(precision+recall)) \quad (9)$$

$$Similarity=tp/(tp+fp+fn) \quad (10)$$

The experimental results elucidates that the hybrid approach for background subtraction are superior either to stand alone block based or pixel based methods (Table I). The threshold value for match function is set to λ =4 for 5x5 and 4x4 block sizes, throughout the work. For other block sizes, λ =6.

B. Block Based Descriptor Evaluation

The descriptors DCT and FFT are being evaluated for their averaged performance on test sequence of Waving Trees. There is no specific distinction in their performance for this test sequence. However, DCT proves to be superior to FFT in terms of precision, similarity and F measure. This evaluation is depicted in table II.

TABLE II Averaged Performance Comparison of DCT & FFT on Test Sequence of Waving Trees

	DCT	**FFT**
Precision	0.8874	0.8794
Recall	0.9109	0.9150
Similarity	0.8166	0.8130
F-measure	0.8990	0.8968

C. Evaluation of Processing speeds of DCT and FFT

Block descriptors DCT and FFT were evaluated. FFT brings tremendous processing speed in comparison with DCT

The table III depicts the time (in seconds) required by FFT and DCT descriptor to train a sequence containing just 20 frames. The experiment shows that processing speed of FFT is much more compared to DCT. Thus FFT produce output faster than DCT.

TABLE III Processing time Comparison of Block Descriptors

	FFT	**DCT**
Hybrid(4x4)	19.7699	188.6593
Hybrid(5x5)	13.2607	121.6438
Hybrid(8x8)	5.4065	48.0231
Hybrid(10x10)	3.5781	31.0414

In figure 6, first row depicts the sequence in RGB color channel, and Row II indicates the respective binary sequences.(1)refers to Waving Trees(frame287) with its Ground truth.(2) refers to the result obtained by Block Based method (size:8x8) and its binary (3) indicates Pixel Based result (4) shows the results with the hybrid approach(8x8).

The performance of various block sizes for the hybrid method is evaluated; fig 7 depicts its results.

In figure 7, each row represents the results for a specific block size. (a) refers to block size (8x8), (b) (10x10),(c) (4x4),(d) (5x5).In each row, the first image depicts result of hybrid approach, second shows the same after morphological close operation, third depicts result after applying median filter of size (3x3).

TABLE I Performance Comparison using WAVING TREES Test Sequence

	Precision	**Recall**	**Similarity**	**F score**
Block-Based(4x4)	85.21	94.07	80.87	89.42
Block-Based(5x5)	83.22	96.82	81.01	89.51
Block-Based(8x8)	81.24	95.43	78.19	87.76
Block-Based(10x10)	80.99	96.34	78.57	88.00
Pixel Based	87.94	91.50	81.30	89.68
Hybrid(4x4)	88.58	90.89	81.35	89.72
Hybrid(5x5)	88.80	91.41	81.96	90.09
Hybrid(8x8)	88.60	90.92	81.40	89.74
Hybrid(10x10)	88.98	91.17	81.93	90.06
MOG	63.64	80.00	54.91	70.89

D. Output Observations

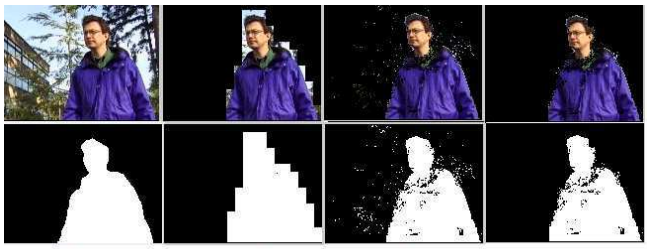

Fig.6. Results of background removed foreground for Waving Trees Test Sequence

The figure 8 shows the non-post processed test sequence image for various BGS methods. The results potrays the efficiency of the hybrid method over the Gaussian modeling techniques like Mixture of Gaussians. The proposed method is effective under dynamic environments and for outdoor sequences.

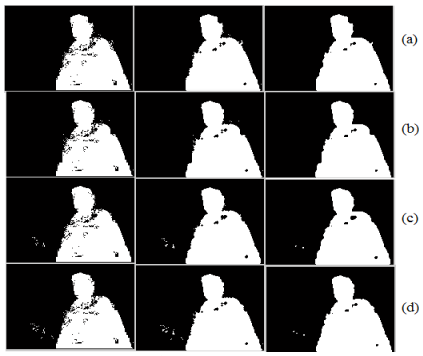

Fig.7. Performance Evaluation for various block sizes

Fig.8. The Background Subtraction Output

The Background Subtraction output for 1.Mixture of Gaussians, 2.Hybrid method of block size (8x8), 3.(10x10), 4.(4x4), 5.(5x5) is described in the figure 8.

Figure 9 depicts the output of the Codebook Background Subtraction method for Fountain[13] test sequence without using any post processing operations. Figure 10 is the output for Campus[13] test sequence which is highly dynamic.

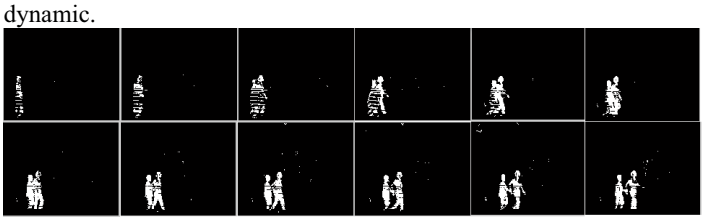
Fig.9. The Background Subtraction output of Fountain test sequence

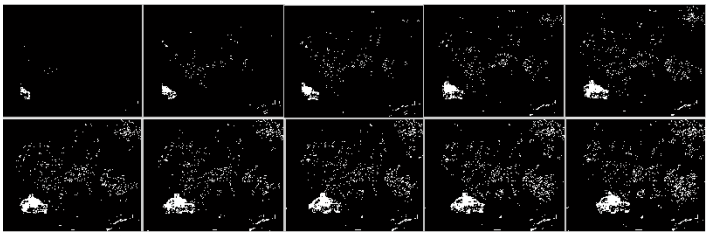
Fig.10. The Background Subtraction output of Campus test sequence

V. CONCLUSION AND FUTURE WORKS

The proposed hybrid system for Background Subtraction is evaluated to be very efficient in processing and in accuracy. Much of the background could be removed in Block Stage itself, the Pixel Stage is also incorporated so that better precision and rich contextual information could be achieved. Codebooks reveal to be useful in applications that demand speedy processing of data. The method is efficient in a different perspective; it could process the images in RGB color channels which is computationally thrice complex than a gray-scale image. The system is also efficient in terms of the feature descriptor used for representing a block. As obtained in the experimental evaluations, FFT descriptor processes blocks much faster in comparison with DCT; still DCT is slightly superior to FFT in terms of precision, similarity and F-measure. Another advantage of these descriptors is that, only with 12D features, better performance is obtained. Selection of Block size varies for different datasets. Performance of the system with Hybrid (8x8) and Hybrid(10x10) for WAVINGTREES reveals that Hybrid(10x10) produces slightly better results than Hybrid(8x8).

A. Future Scope

This work need to be extended to include the following

- More comparisons based on various other Block Descriptors.
- The new algorithm of MIT called "Sparse FFT" would be introduced as a block descriptor that is not yet used as feature descriptor.
- Efficient Conical color model that can effectively segregate true foreground from shadow and highlight.

REFERENCES

[1] T. Bouwmans, F. El Baf, B. Vachon," Background Modeling using Mixture of Gaussians for Foreground Detection- A Survey", Recent Patents on Computer Science 1, 3 ,2008 219-237

[2] C Stauffer, W Grimson, "Adaptive background mixture models for real - time tracking". Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999, 2(6) : 248 – 252

[3] M. Sivabalakrishnan and D. Manjula. "An efficient foreground detection algorithm for visual surveillance system." International Journal of Computer Science and Network Security, 9(5):221–227, May 2009

[4] B Han,L.S Davis,"Density based multifeature background subtraction with SVM",IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL. 34, NO. 5, MAY 2012

[5] C Chiu, M Ku,"A Robust object segmentation system using a probability based background extraction algorithm", IEEE Transactions on Circuits & Systems for Video Technology, Vol. 20, NO. 4, April 2010

[6] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model", Real-Time Imaging, vol. 11, no. 3, pp. 172–185, Jun. 2005

[7] Jing-Ming Guo, Yun-Fu Liu, , Chih-Hsien Hsia, Min-Hsiung Shih, and Chih-Sheng Hsu," Hierarchical Method for Foreground Detection Using Codebook Model", IEEE Transactions on Circuits & Systems for Video Technology, Vol. 21, NO. 6, June 2011

[8] Vikas Reddy, Conrad Sanderson, Brian C. Lovell, "Improved Foreground Detection via Block-Based Classifier Cascade With Probabilistic Decision Integration", IEEE Transactions on Circuits & Systems for Video Technology, Vol. 23-1, Jan 2013

[9] Marko Heikkil, Matti Pietik inen," A Texture-Based Method for Modeling the Background and Detecting Moving Objects", IEEE Transactions on Pattern Analysis &Machine Intelligence,Vol.28-4, Apr 2006

[10] Takashi Matsuyama, Toshikazu Wada, Hitoshi Habe, and Kazuya Tanahashi, "Background Subtraction under Varying Illumination", Systems and Computers in Japan, Vol. 37, No. 4, 2006

[11] O. Barnich and M. Van Droogenbroeck. "ViBe: a powerful random technique to estimate the background in video sequences". In Int. Conf. on

Acoustics, Speech and Signal Processing (ICASSP), pages 945–948, April 2009
[12] M.Piccardi, "Background Subtraction Techniques- a review", UIT, Sydney
[13] Statistical Modeling of Complex Background for Foreground Object Detection [Online]. Available:http://perception.i2r.astar.edu.sg/bk−model /bk−index.html