

Effective Detection of Moving Objects Using Complex Background Subtraction

¹R.Kalai selvi, ²D. Durai kumar

¹M.Tech Student, ²Associate Professor & Head

Department of Information Technology,

Ganadipathy Tulsi's Jain Engineering College, Vellore, Tamilnadu, India

Abstract - The detection of moving object is based on background subtraction under complex wavelet transform (CWT) domain for video surveillance system. Initially it starts with background initialization by choosing start frame or taking initial few frames with approximate median method. Then the CWT is applied to both low and high frequency sub bands. In order to remove some unwanted pixels, morphological operation is performed for object edge smoothness. The proposed approach has some advantages of background noise insensitiveness and invariant to varying illumination or lighting conditions. After the object detection, performance of method is measured through metrics such as sensitivity, accuracy, correlation. This object detection also helps to track detected object using connected component analysis.

Index Terms - Video surveillance, CWT, Approximate median, Sub band differencing, Morphological operation

I. INTRODUCTION

There are instantaneous needs for programmed video supervision systems in viable, law enforcement and military applications for providing security. Mounting video cameras is cost-effective, but finding obtainable human property to observe the output is expensive. Although observation cameras are previously ubiquitous in banks, stores, and parking lots, video data currently is used only "Subsequent to the fact" as a forensic tool, thus losing its most important benefit as an active, real-time intermediate [4]. The abundant military applications include patrolling national borders, measuring the stream of refugees in concerned areas, monitoring peace treaties, and on condition that secure perimeters around bases and embassies. Real-time moving object detection and tracking from stationary and moving camera platforms, recognition of generic object classes (e.g. human, sedan, truck) and specific object types (e.g. campus police car, FedEx van), object facade evaluation with admiration to a geospatial situate sculpt, vigorous camera organize and multi-camera supportive tracking, human gait analysis, recognition of simple multi-agent behavior, concurrent data propagation, data logging and self-motivated scene hallucination.

Surveillance is the monitoring of behavior. Systems surveillance is the progression of monitoring the activities of inhabitants, substance or processes within systems for conventionality to predictable or preferred norms in trusted systems for protection or societal organize for all forms of surveillance or monitoring, not just illustration observation. Surveillance in numerous contemporary cities and buildings frequently uses closed-circuit television cameras. Although surveillance can be a constructive apparatus for law enforcement and security companies, various inhabitants encompass concerns about the loss of solitude.

Background subtraction (BS) is an essential precursor in most machine vision applications. In most video analytics systems, BS is performed for moving foreground detection. Although most security cameras are static in the sense that these are attached to fixed poles with panning and zooming disabled, the scene background observed by such a camera is still vulnerable to casual jerks of the pole, changes due to lighting conditions, and background with intrinsic motion. Thus, BS for even static cameras is performed by maintaining a statistical model of the dynamic background and then comparing its difference from each incoming video frame [5]. The underlying background model is kept updated to reflect illumination variation or any structural change in the background over time. Existing BS techniques are reliable and produce acceptable detection results either with scenario specific parameter tuning or when scene dynamics remains stable. However, due to over-reliance on statistical observations, these techniques show unpredictable performance in dynamic unconstrained scenarios where the characteristics of the operating environment are either unknown or change abruptly.

II. SYSTEM MODEL

The frames are separated into number of still images from an input video. These images are decomposed into different sub bands with respect to high and low frequency. CWT is applied on different sub bands based on background initialization. By subtracting background frame and foreground frame the foreground is detected based on sub band differencing. With help of binary map and inverse CWT and algorithm is used to enhancing the object detection process. Then the morphological erosion and dilation operations are used to attenuating the color variations generated by background motions while still highlighting the moving objects. And finally using morphological process to smoothing the foreground region and determine the performance measures.

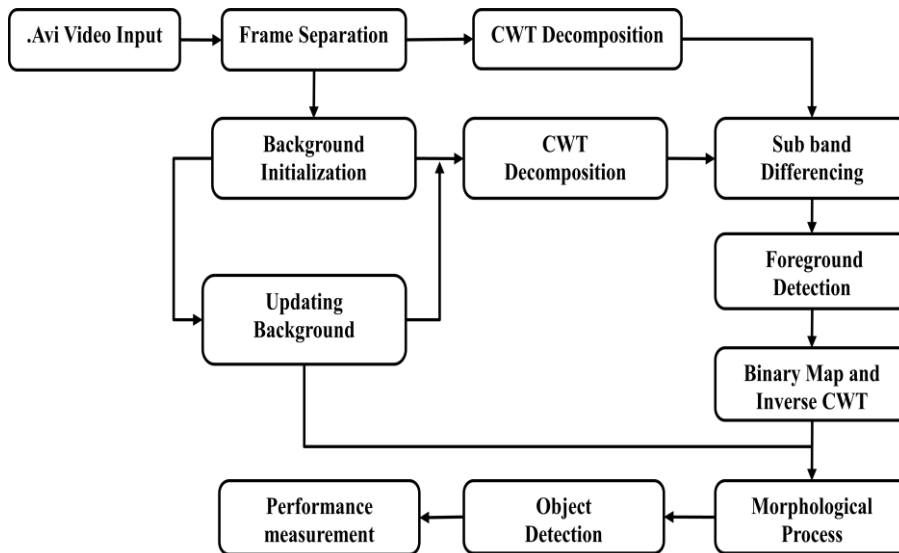


Fig 1: Architecture for object detection

III. RELATED WORK

1. FRAME SEPARATION

An Input Video (.avi files) is converted into still images for processing it and to detect the moving objects. These sequences of frames are gathered from video files by finding the information about it through 'aviinfo' command. These frames are converted into images with help of the command 'frame2im'. Create the name to each images and this process will be continued for all the video frames.

2. GAUSSIAN SMOOTHING PROCESS

A Gaussian smoothing is the result of blurring an image by a Gaussian function. It is a extensively used consequence in graphics software, normally to condense image noise and decrease feature. Gaussian smoothing is furthermore used as a pre-processing point in computer vision algorithms in order to improve representation structures at different scales. Mathematically, applying a Gaussian blur to an image is the equivalent as convolving the image with a Gaussian function. By contrast, convolving by a circle (i.e., a circular box blur) would more accurately reproduce the bookeed effect. Since the Fourier transform of a Gaussian, by applying a Gaussian blur is used to reducing the image's high-frequency components. The idea of Gaussian smoothing is to use this 2-D allocation as a 'point-spread' function is achieved by convolution. Since the image is stored as an assortment of discrete pixels are need to generate a separate approximation to the Gaussian function before perform the convolution.

$$\text{Gauss Coeff} = (1/\sqrt{2\pi \cdot \text{sig}^2}) \cdot (\exp(-x^2 - y^2 / 2 \cdot \text{sig}^2)) \tag{1}$$

Where, x, y, sig - input coordinates corresponds to the target and standard deviation.

3. COMPLEX WAVELET TRANSFORM (CWT)

This CWT is used to separate the approximation and detailed coefficients from the time domain images. It decomposes the image into different sub band images, namely, LL, LH, HL, and HH for embedding the messages in the pixel coefficients of sub bands. Complex scheme is a technique to convert DWT coefficients to Integer coefficients without losing informational sub bands contains the significant part of the spatial domain image. High-frequency sub band contains the edge information and noise component of input image.

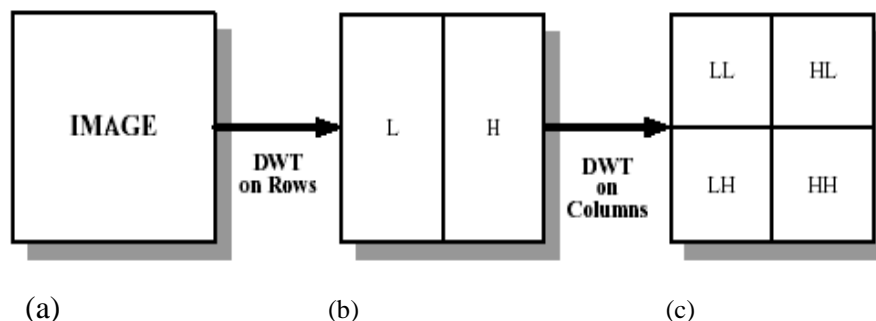


Fig 2: Block Diagram of CWT

(a) Original Image (b) Output image after the 1-D applied on Row input

(c) Output image after the second 1-D applied on row input

The coefficients are selected as reserved space foe hiding the text records. The covert text data is implanted into the wavelet coefficients of high frequency sub bands because it is non sensitive to human visual system. The basis set of wavelets is generated from the mother or basic wavelet is defined as:

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) ; a, b \in \mathbb{R} \text{ and } a > 0 \tag{2}$$

The variable ‘a’ (inverse of frequency) reflects the scale (width) of a particular basis function such that its large value gives low frequencies and small value gives high frequencies. The variable ‘b’ specifies its translation along x-axis in time. The term $1/\sqrt{a}$ is used for normalization. In order to obtain an efficient wavelet computation, it is significant to reduce as many preventable computations as potential. A suspicious assessment of promote and reverse transforms shows that about half the operations either lead to data which are destroyed or are null operations (as in multiplication by 0). The one-dimensional wavelet transform is computed by separately applying two analysis filters at alternating even and odd locations. The converse progression first doubles the duration of all signals by inserting zeros in every supplementary point, then applies the suitable combination filter to each signal and adds the filtered signals to get the final reverse transform. Inverse Integer wavelet transform is formed by reverse lifting proposal procedure is parallel to the forward lifting scheme.

Algorithm 1: Decomposition of images into sub bands

- 1: Inputs-R (Resolution level), N*M (Matrix) to H(high pass filter) and L (Low pas filter),i(iterations),n.
- 2: for (i=0;i>=count; i++)
- 3: convert frames into still images
- end for
- 4: filtering process
- 5:for(j=0; j>=count; j++)
- 6: column wise dispensation to obtain H and L
 $H = (Co-Ce)$ and $L = (Ce+ [H/2])$
- 7: row wise processing to get LL, LH, HL and HH
- 8: divide odd and even rows of H and L,
 $LH = L_{odd}-L_{even}$, $LL = L_{even} + [LH / 2]$
 $HH = H_{odd} - H_{even}$, $HL = H_{even} + [HH / 2]$
- end for
- 9:obtain image information visible at resolution level R.
- 10: outputs-sub bands-LL, LH, HL, HH

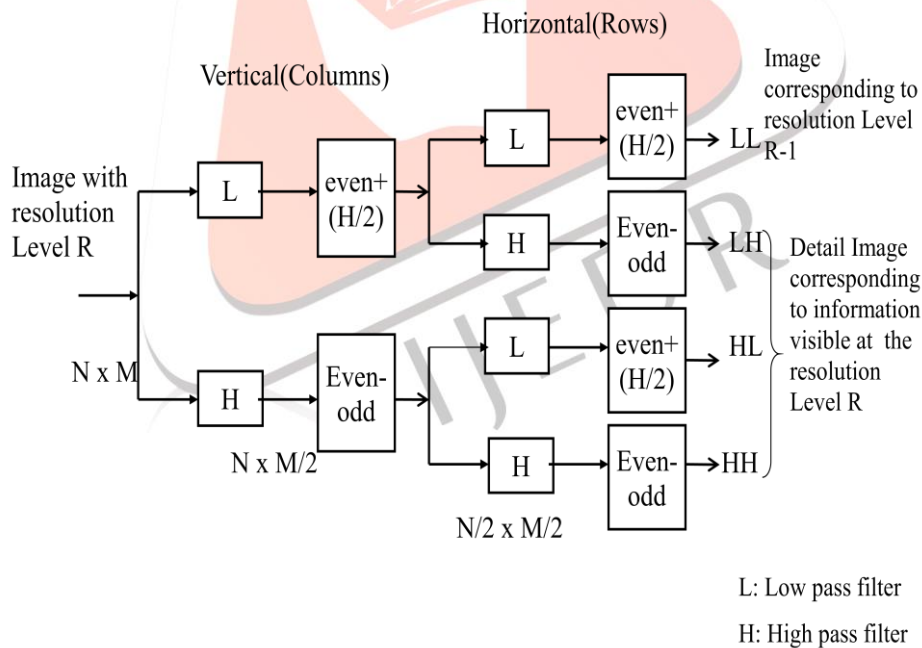


Fig 3: CWT Decomposition

4. BACKGROUND SUBTRACTION

Background subtraction is the first step in the process of segmenting and tracking inhabitants. Discriminate among foreground and background in a very dynamic and unconstrained outdoor environment over several hours is a challenging task. Background subtraction is the first step in the process of segmenting and detecting objects.

A. Background Model

The principle of the background model is to quantify the variation at each pixel into a set of vectors. Each vector is called a codeword [7], and a set of codeword’s forms a codebook. The quantization into codeword’s is not based on any parametric

assumptions. If a background pixel has very little variation then only a single codeword is used to model the background at that pixel [1].

B. Subtraction and Update of the Background Model

After initialization, temporally subsequent samples are fed to the network. Each incoming pixel P_t of the sequence Frame I_t is compared to the current pixel model C to determine if there exists a weight vector that best matches it. If a paramount similar weight vector C_m is found, it means that P_t belongs to the background and it is used as the pixel encoding approximation, and the best similar weight vector, simultaneously with its region, is durable [4]. To determine which weight vector gives the excellent match, a number of metrics for detecting changes in color imagery, such as those reported in and in references therein, could be adopted. Experiments are show the way to provide work for the Euclidean distance of vectors in the HSV color hex cone [7] that gives the distance between two pixel p_i and p_j .

$$p_i = (h_i, s_i, v_i) \quad i = (h_i, s_i, v_i) \quad \text{and} \quad p_j = (h_j, s_j, v_j) \quad \text{as} \quad d(p_i, p_j) = \left\| (v_i s_i \cos(h_i), v_i s_i \sin(h_i), v_i) - (v_j s_j \cos(h_j), v_j s_j \sin(h_j), v_j) \right\| \quad (3)$$

C. Foreground/Background Classification

When the background model has been initialized through the training period the classification of pixels as either foreground or background is a relatively simple task. If a pixel at image coordinate (x,y) lies within the subspace of RGB space [4] described by any of the inward pixel position (x,y) , then the pixel is classified as background, otherwise it is foreground. Whether or not a pixel lies within a the subspace then it is determined in the same way as in the training where the difference in chromaticity must be smaller than the subspace radius, and the intensity must be within the intensity limits low and high frequencies. This utilization of the background model gives the foreground/background. In the classification the ability to handle shadows [7] and highlights since the model accounts for the intensity variation. The background model's way of dealing with chromaticity differences also gives the foreground/background classification a high detection sensitivity to foreground objects that are similar in color to the background. In the classification process the values of min and max can be different from the values used during training. It is reasonable to only allow small variation in the chromaticity's during the relatively short training period, whereas the variation over longer periods of time would be greater due to large changes in illumination [3] conditions. The preprocessing, mentions that smoothing of the images can be used to reduce camera noise and remove transient environmental noise such as rain. Many algorithms use a Gaussian blur first to average out fluctuating pixel values to alleviate big differences. Alternatively, when temporal data can be exploited in a video, if a pixel's value is constantly changing over time then it can be assumed it is part of a non-static background object. The background model can deal with events such as objects changing positions by implementing an effective update rule to change the model over time. Background modeling is an area of research itself, however one example of an update process is to track object locations. If an object moves and then remains constantly in the same position over a length of time it can then be considered to be a part of the background. Illumination changes [8] can be handled by exploiting illumination invariance within the color space used. Post-processing can be used for data validation to eliminate false positive matches. This can be in the form of rejection of isolated foreground pixels as they can be assumed to be noise or thresholding on foreground region size. As the subtraction usually only looks at a single pixel, this stage can also examine the value of the neighborhood pixels [6].

5. SUB-BAND DIFFERENCING

The sub band filtering process is done between background frame and foreground frame. This process is done by subtracting current frame sub band and previous frame sub band for detecting object from background. Then background is updated by comparing the process frame and background frame. This process continued for all consecutive frames. The first k video frames are used to train the background model to achieve a model that represents the variation in the background during this period.

The frames (from $k + 1$ and onwards) are each processed by the Sub-Band Differencing module to produce a mask that describes the foreground regions identified by comparing the incoming frame with the background model. Information from frames $k + 1$ and onwards are used to update the background model either by the continuous revise method, the encrusted updating, or equally. The facade obtained from the Sub-Band Differencing is processed further in the post dispensation component which minimizes the consequence of noise in the facade.

6. MORPHOLOGICAL PROCESS

Morphological image processing is a collection of non-linear operations related to the shape or morphology of features in an image. Morphological operations rely simply on the relative ordering of pixel values, not on their numerical values, and consequently are particularly appropriate to the processing of binary images. Morphological operations can also be applied to grey scale images such that their light transfer functions are unknown and therefore their absolute pixel values are of no or minor interest. A morphological operation on a binary image creates a new binary image in which the pixel has a non-zero value only if the test is successful at that location in the input image. Morphological techniques probe an image with a small shape or template called a structuring element. The structuring element is placed at all possible locations in the image and it is compared with the corresponding neighborhood of pixels. The structuring element is a small binary image, i.e. a small matrix of pixels, each with a value of zero or one. The matrix dimensions indicate the range of the structuring element. The sample of ones and zeros specifies the shape of the structuring element. An foundation of the structuring element is frequently one of its pixels, although generally the origin can be outside the structuring element. A common practice is to have odd dimensions of the structuring matrix and the origin defined as the centre of the matrix. Structuring elements cooperate in morphological image processing the same role as convolution kernels in linear image filtering. Erosion and dilation process are used to enhance (smoothing) the object region by

removing the unwanted pixels from outside region of foreground object. Erosion among small (e.g. 2×2 - 5×5) square structuring elements shrinks an image by stripping away a layer of pixels from both the inner and outer boundaries of regions. The holes and gaps between different regions become larger, and small details are eliminated.

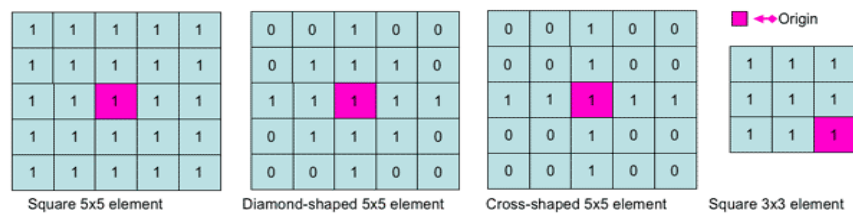


Fig 4: Examples of simple structuring elements

Dilation, in general, causes objects to dilate or grow in size and erosion causes objects to reduce in size. The quantity and the approach that they produce or shrink depend upon the choice of the structuring element. Dilating or eroding not as well as specifying the structural element makes no more sense than trying to low pass filter an image without specifying the filter. After this process, the pixels are applied for connected component analysis and then analysis the object region for counting the objects.

A. Connected Component Analysis

The output of the change detection module is the binary image that contains only two labels, i.e., '0' and '255', representing as 'background' and 'foreground' pixels respectively, with some noise. The goal of the connected component analysis is to detect the large sized connected foreground region or object. This is one of the important operations in motion recognition [9]. The pixels that are cooperatively associated can be clustered into changing or moving objects by analyzing their connectivity. In binary image analysis, the object is extracted using the connected component cataloging procedure, which consist of transmission distinctive label to each maximally connected foreground province of pixels. One of the imperative classification [7] approaches is "classical sequential labeling algorithm". It is based on two raster scan of binary image. The original scan performs the provisional labeling to each foreground region pixels by checking their connectivity of the scanned image. When a foreground pixel with two or more than two foreground neighboring pixels carrying the same label is found, the labels associated with those pixels are registered as being equivalent. That means these regions are from the same object. The handling of equivalent labels and merging thereafter is the most complex task. The first scan gives temporary labels to the foreground pixels according to their connectivity. The connectivity check can be done with the help of either a 4-connectivity or 8-connectivity approach. 8-connectivity approach is used. Here, the idea is to label the whole blob at a time to avoid the label redundancies. The labeling operation scans the image moving along the row until it comes to the point P, for which $S = \{255\}$.

7. PERFORMANCE MEASURE

The velocity of object is evaluated based on distance travelled by an object and frame rate

$$\text{Velocity} = \text{Distance travelled} / \text{Frame rate}$$

Sensitivity is measures the proportion of actual positives which are correctly identified

$$\text{Sensitivity} = \text{Tp} / (\text{Tp} + \text{Fn}) \quad (4)$$

Where, Tp = True Positive: Object pixels correctly classified as object, Fn = False negative: Object pixels incorrectly classified as background.

Correlation coefficient is used to find the similarity between two different images with their intensities. It will be described by,

$$\text{Cor_coef} = [\text{sum}(\text{sum}(u1 * u2))] / [\text{sqrt}(\text{sum}(\text{sum}(u1 * u1)) * \text{sum}(\text{sum}(u2 * u2)))] \quad (5)$$

Where, $u1 = F1 - \text{mean of } F1$, $u2 = F2 - \text{mean of } F2$, $F1$ - Obtained result and $F2$ - Ground truth

IV. CONCLUSION

The efficient detection of object is based on complex background subtraction using frame difference with thresholding, complex wavelet transformation and mathematical morphology. In the present work, a robust and efficiently computed method for segmentation of moving objects using approximate median filter based method in complex wavelet transform domain have been proposed with effective way of handling dynamic backgrounds. The performance of the proposed method has been evaluated and compared with other standard methods in consideration in terms of various performance metrics. It is concluded that the proposed method is performing better in comparison to other methods as well as capable of alleviating the problems associated with other spatial domain methods.

REFERENCES

- [1] Baochang Zhang, Yongsheng Gao, Sanqiang Zhao, and Bineng Zhong, "Kernel Similarity Modeling of Texture Pattern Flow for Motion Detection in Complex Background," IEEE Trans. Circuits and Systems for Video Technology, vol. 21, no. 1, pp. 29-38, January 2011.
- [2] L. Li, W. Huang, I.Y.-H. Gu, and Q. Tian, "Statistical Modeling of Complex Backgrounds for Foreground Object Detection," IEEE Trans. Image Processing, vol. 13, no. 11, pp. 1459-1472, Nov. 2004.
- [3] L. Li and M. Leung, "Integrating intensity and texture differences for robust change detection," IEEE Trans. Image Processing, vol. 11, pp. 105-112, Feb. 2002.

- [4] L. Maddalena and A. Petrosino, "A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications," *IEEE Trans. Image Processing*, vol. 17, no. 7, pp. 1168-1177, July 2008.
- [5] Mahfuzul Haque, Member, IEEE, and Manzur Murshed, "Perception-Inspired Background Subtraction," *IEEE Trans. Circuits and Systems For Video Technology*, vol. 23, no. 12, pp. 2127- 2140, December 2013.
- [6] Mike McHugh, Member, Janusz Konrad, Fellow, Venkatesh Saligrama, and Pierre-Marc Jodoin, Member, "Foreground-Adaptive Background Subtraction," *IEEE Signal Processing Letters*, vol. 16, no. 5, pp. 390-393, May 2009.
- [7] R. Cucchiara, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1-6, Oct. 2003.
- [8] O. Javed, K. Shafique, and M. Shah, "A hierarchical approach to robust background subtraction using color and gradient information," in *Proc. IEEE Workshop Motion Video Computing*, Dec. 2002, pp. 22-27.
- [9] Y.-L. Tian and A. Hampapur, "Robust salient motion detection with complex background for real-time video surveillance," in *Proc. IEEE Comput. Soc. Workshop Motion Video Comput.*, vol. 2. 2005, pp. 30-35.

