International Journal of Computer Applications Technology and Research
Volume 3– Issue 4, 221 - 228, 2014, ISSN: 2319–8656

Algorithmic Analysis to Video Object Tracking and Background Segmentation and Wavelet Transform as New Approach to Video Object Tracking

Ashmita Rathi
BAMU University
Aditya Engineering
College, Beed, India

S.H.Kadam
BAMU University
Aditya Engineering
College, Beed, India

Syed A.H
BAMU University
Aditya Engineering
College, Beed, India

Abstract: Video object tracking and segmentation are the fundamental building blocks for smart surveillance system. Various algorithms like partial least square analysis, Markov model, Temporal differencing, background subtraction algorithm, adaptive background updating have been proposed but each were having drawbacks like object tracking problem, multibackground congestion, illumination changes, occlusion etc. The background segmentation worked on to principled object tracking by using two models Gaussian mixture model and level centre model. Wavelet transforms have been one of the important signal processing developments, especially for the applications such as time-frequency analysis, data compression, segmentation and vision. The key idea of the wavelet transform approach is to represents any arbitrary function f (t) as a superposition of a set of such wavelets or basis functions. Results show that algorithm performs well to remove occlusion and multibackground congestion as well as algorithm worked with removal of noise in the signals.

Keywords: Temporal differencing, occlusion, gaussian mixture, level set, wavelets transform

1. INTRODUCTION

Video object tracking and segmentation had the roots in solving problem like traffic surveillance system, suspicious person monitoring. Several algorithms have been proposed for video object tracking and segmentation. Partial least square analysis worked on principle of object tracking posed as a binary classification problem in which the correlation of object appearance and class labels from foreground and background is modeled by partial least squares (PLS) analysis, for generating a low-dimensional discriminative feature subspace. The algorithm had high performance rate and lower tracking algorithm. Markov random field model worked only on background subtraction having drawback like occlusion. Temporal differencing [1] is very adaptive to dynamic environments, as only the current frames are used for analysis, but generally does a poor job of extracting all the relevant object pixels corresponding to object motion. Back-ground subtraction [2][3], provides the most complete object data but is extremely sensitive to dynamic scene changes due to lighting and extraneous events [4]. More recent adaptive back grounding methods can cope much better with environment dynamism. However, they cannot handle multi-modal backgrounds and have problems in scenes with many moving objects.

Background Segmentation is a more advanced adaptive background modeling method. Here, each pixel is modeled using a mixture of Gaussians and is updated by an on-line approximation. The adaptive background model based segmentation method would alone suffice for applications where a rough estimate of the moving foreground, is in the form of irregular space blobs, is sufficient. Here the exact shape of the moving object need not be determined and only some post processing of the segmentation output using appropriate filters would give the desired blobs of interest. Recently, the level set method has become popular for object shape extraction and tracking purposes. The central idea is to evolve a higher dimensional function whose zero-level set always corresponds to the position of the propagating contour. There are several advantages of this level set formulation. Topological changes such as splitting and merging of contours are naturally handled. The final extracted contour is independent of the curve initialization, unlike other active contour models like the snakes, where the final object contour is very much determined by the contour initialization.

Wavelet transform is theory of wavelets having roots in quantum mechanics and the theory of functions though a unifying framework is a recent occurrence. Wavelet analysis is performed using a prototype function called a wavelet. Wavelets are functions defined over a finite interval and having an average value of zero. The basic idea of the wavelet transform is to represent any arbitrary function f (t) as a superposition of a set of such wavelets or basis functions. These basis functions or baby wavelets are obtained from a single prototype wavelet called the mother wavelet, by dilations or contractions (scaling) and translations (shifts). Efficient implementation of the wavelet transforms has been derived based on the Fast Fourier transform and short-length fast-running FIR algorithms in order to reduce the computational complexity per computed coefficient. All wavelet packet transforms are calculated in a similar way. Therefore we shall concentrate initially on the Haar wavelet packet transform, which is the easiest to describe and implement. The Haar wavelet packet transform is usually referred to as the Walsh transform.

2. LITERATURE SURVEY

This topic of paper shows the analysis and from analysis guides about the drawback of various algorithms that have been proposed for video object tracking and segmentation for smart surveillance system and the advantage of background segmentation and wavelet transform on it. In this topic each algorithm is studied in finer detail.

2.1 Analysis Of Partial Least Square Algorithm

Partial least squares analysis (PLS) [5], is a statistical method for modeling relations between sets of variables via some latent quantities.PLS starts at the point where maximum likelihood covariance-based system reach their limit. PLS works well with object tracking by using methods that find new spaces where most variations of the samples can be preserved, and the learned latent variables from two blocks are more correlated than those in the original spaces PLS therefore combines information about the variances of both the predictors and the responses, while also considering the correlations among them [10].

The major limitations of PLS algorithm are high risk of overlooking real correlations and sensitivity to relative scaling of descriptor variables [22].

2.2 Hidden Markov Model Explanation

The hidden markov model is stochastic model based on markov property which studies the detail of each hidden states. The video object tracking process can be modeled with the hidden Markov model (HMM) shown in Fig. 1. In this model, system state, *Xt*, is the object state to be estimated at time *t*

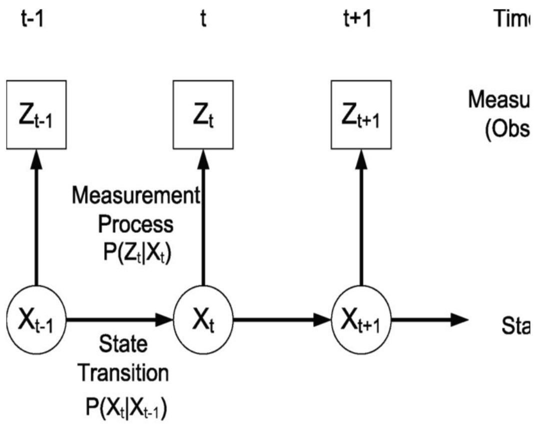

Figure 1: Illustration of system model. The video object tracking problem can be modeled with a hidden Markov model (HMM).

(e.g., object position, object velocity, objects size, etc). Measurement, *Zt,* is the observation from the current frame. Video object tracking is an inverse problem; that is, object states are inferred from the measurements (features extracted from video sequences) such as color, texture, and gradient [8]. The model in Fig. 1 can be specified further using the following equations:

$Xt = f\,(Xt{-}1) + nx,$ -------------------------------- (1)

$Zt = g\,(Xt) + nz,$ ----------------------------------- (2)

Where $f\,(\cdot)$ is the state transition function, $g(\cdot)$ is the Measurement function, and *nx* and *nz* are noise. HMM requires the prior knowledge for building the architecture. The major disadvantage of the model is speed Almost everything one does in an HMM involves: “enumerating all possible paths through the model”. This model is still slow in comparison to other methods.

2.3 Analysis Of CAMSHIFT

Continuously Adaptive Mean Shift algorithm (CAMSHIFT) is a popular algorithm for visual tracking [7], providing speed and robustness with minimal training and computational cost. CamShift is an adaptation of the Mean Shift algorithm for object tracking that is intended as a step towards head and face tracking for a perceptual user interface using minimum CPU cycles and thereby a single color hue. The CamShift algorithm can be summarized in the following steps:

1. Set the region of interest (ROI) of the probability distribution image to the entire image.

2. Select an initial location of the Mean Shift search window. The selected location is the target distribution to be tracked.

3. Calculate a color probability distribution of the region centered at the Mean Shift search window.

4. Iterate Mean Shift algorithm to find the centroid of the probability image. Store the zeroth moment (distribution area) and centroid location.

5. For the following frame, center the search window at the mean location found in Step 4 and set the window size to a function of the zeroth moment. Go to Step 3.

The algorithm may fail to track multi-hued objects or objects where hue alone cannot allow the object to be distinguished from the background and other objects. It can fail rapidly when the camera moves since it relies on static models of both background and the tracked object [09]. Furthermore, it is unable to track objects passing in front of backgrounds with which it shares significant colors [10].

3. PROPOSED ALGORITHM

The above topic depicts the various problems faced by the algorithm that are used for video object tracking and segmentation. More advanced adaptive background modeling methods have been proposed for video object tracking and segmentation. Here, each pixel is modeled using a mixture of Gaussians and is updated by an on-line approximation. Filters would give the desired blobs of interest. Recently, the level set method has become popular for object shape extraction and tracking purposes. Both the algorithm explained deals with background segmentation and is explained in minute detail below:

3.1 Background Segmentation

3.1.1 Gaussian mixture model

Background subtraction involves calculating a reference background image, subtracting each new frame from this image and thresholding the result that results in a binary segmentation image that highlights the regions of non-stationary objects. Non-adaptive background models, computed over a training sequence of sufficient length involving only the background, could only be used in case of short-time surveillance applications, where we assume that the background model, both in terms of pixel intensity distribution and background composition, have not changed significantly to undo the background subtraction philosophy. But the above assumptions cannot be guaranteed to hold good for long-time surveillance applications, where not only scene composition but also the intensity distributions of the background can change over time. When the background scene involves large or sudden changes, a single Gaussian model is not adequate, and a multi-model distribution is needed to fully describe the scene dynamics. GMM allows the representation of a background scene with multi-modal distributions [11]. Here, multi-model distribution means multiple- Gaussian distribution, which essentially means a "multiple-surface" background representation of a pixel. In GMM, each pixel is modeled parametrically by a mixture of K Gaussians as the intensity of each pixel evolves over time (temporally). The model is parameterized by a mean, a covariance matrix, a priori probability for each of the K components. Similar to the single Gaussian method, they can be implemented using a running average method. The parameters are updated for each frame and hence the method does not require a large buffer of video frames with high memory requirement. GMM is a more general approach when compared to the single Gaussian model. For every pixel, each of the K Gaussian distributions corresponds to the probability of observing a particular intensity. The probabilities are then evaluated to determine which are most likely to be resulted from the background scene, and which are most likely to be resulted from the foreground objects. The probability of observing the pixel intensity in the current frame is:

$$P(I_t) = \sum_{i=1}^{K} \omega_{i,t}\eta(I_t, \mu_{i,t}, \Sigma_{i,t})$$

………………………… (3)

where K is number of distributions, it is chosen to be usually 3 . . . 5, ωi,t is an estimate a priori probability (what portion of the data is accounted for by this Gaussian) of the ith Gaussian at time t, η(It, μi,t,_i,t) is the ith Gaussian model, with mean μ and covariance matrix of the pixel intensities. For computational simplicity, the covariance matrix is assumed to be diagonal so that the inverse can be determined easily. The distribution of recently observed values of each pixel in the scene is characterized by a mixture of Gaussians. A new pixel in the current frame is representing by one of the K components of the mixture model. To update the background model, each new pixel in the current frame, is checked against the existing K Gaussian distributions, until a "match" is found. A match is defined as a pixel value within 2.5 standard deviations of a distribution (out of the K components), and this matched component can be a background component or a foreground component which will be verified later.

3.1.2 Foreground Segmentation

As the background model has a stochastic nature, we expect the entire segmentation problem to be cast in a similar stochastic framework. Following steps are required to deal with random variables which follow some particular distribution.

1. To calculate incoming pixel intensity using Gaussian random variable, whose mean is the observed intensity and variance is some small prefixed value, i.e. if I(x) denotes the incoming pixel intensity is at location x, then the observed intensity is modeled as random variable X, such that X N (I(x)). This random model is in fact intuitive considering random factors such as camera feed perturbations and source intensity fluctuations.

2. To solve the above problem the observed pixel intensity as a Gaussian signal is modeled, highly spiked about its mean, rather than treating it as a deterministic signal.
3. Then usual background subtraction procedures need to be recast into this stochastic framework. This leads to the deployment of divergence measures between the two probability distributions: the background distribution and the incoming pixel distribution. The Jeffrey's divergence measure is used which is similar to the KL measure but having the additional property of symmetry about its arguments.
4. Jeffrey's information measure between two distributions having density functions f and g is defined as:

$$J(f,g)=\int[f(x)-g(x)]log(\frac{f(x)}{g(x)})dx$$

------------------------------ (4)

As obvious from the above definition, J(f,g) satisfies the following desirable properties:

1. J(f,g)≥0
2. J(f,g)==0 if and only if f and g are identical functions
3. J(f,g)==J(g,f)

3.1.3 Background Model Initialization

In order to avoid the assumption of a sequence starting in the absence of foreground objects, simple temporal frame differencing is used for the initial phase of background model initialization until the background pixels are "stable" [12] The temporal frame difference FDt(x, y) at time t is defined as:

$$FD_t(x,y) = |I_t(x,y) - I_{t-1}(x,y)|$$

…………….(5)

where It(x, y) is the intensity of pixel (x, y) in the frame at time t. The foreground binary mask FGt(x, y) is determined by comparing FDt(x, y) to a threshold T1 which is empirically determined and set to be 20, a pixel is considered as having significant motion, and labeled as a foreground pixel, if the difference is greater than a threshold,

$$FG_t(x,y) = \begin{cases} 1 & \text{if } FD_t(x,y) > T1 \\ 0 & \text{otherwise.} \end{cases}$$

…--………(6)

A pixel is considered as a "stable" background pixel if there is no significant motion detected (i.e. FDt(x, y) < T1) for a certain number of frames (denoted by Tfr). Consider a frame count Cfr that is incremented by 1 each time FDt(x, y) < T1, when this frame counts Cfr > Tfr (the consecutive background frame count Tfr is empirically determined and set to be 100), we can use this pixel in the current frame to construct the background model:

$$BM_t(x,y) = \begin{cases} I_t(x,y) & \text{if } C_{fr} > T_{fr}, \\ 0 & \text{otherwise.} \end{cases}$$

……………(7)

This background model initialization method assumes every pixel of the background will be uncovered at some time.

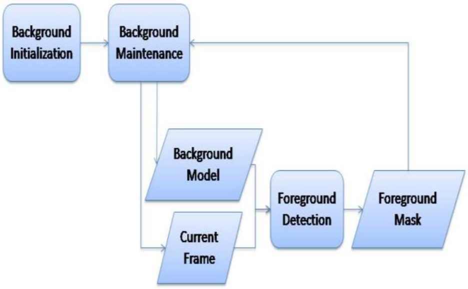

Figure 2: Background Subtraction

4. Wavelet Transform Method

A wavelet series is a representation of a square-integral (real- or complex-valued) function by orthonormal series generated by a wavelet[13]. Nowadays, wavelet transformation is one of the most popular candidates of the time-frequency-transformation.The Haar wavelet packet transform is usually referred to as the Walsh transform. Figure 3 represents the wavelet transform processes in terms of time and frequency.

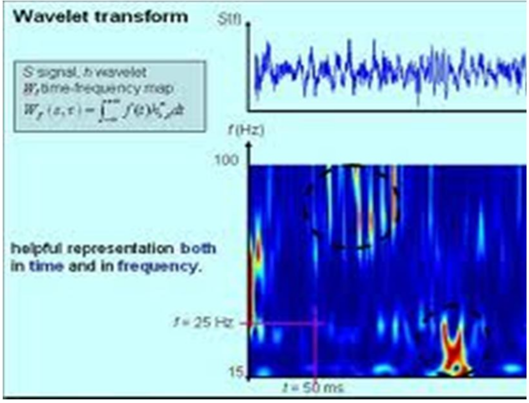

Figure: 3 Wavelet Transform Processes in terms of time and frequency

4.1 Haar Transform Method

The Haar transform is the simplest of the wavelet transforms. This transform cross-multiplies a function against the Haar wavelet with various shifts and stretches, like the Fourier transform cross-multiplies a function against a sine wave with two phases and many stretches.]

The attracting features of the Haar transform, includes fast for implementation and able to analyze the local feature, make it a potential candidate in modern electrical and computer engineering applications, such as signal and image compression

The Haar transform is found effective as it provides a simple approach for analyzing the local aspects of a signal. The Haar transform is derived from the Haar matrix which can be used for removing occlusion. An example of a 4x4 Haar transformation matrix is shown below.

$$H_4 = \frac{1}{\sqrt{4}}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{bmatrix}$$

The Haar transform can be thought of as a sampling processes in which rows of the transformation matrix act as samples of finer and finer resolution [14].

Property of haar transform

1. No need for multiplications. It requires only additions and there are many elements with zero value in the Haar matrix, so the computation time is short. It is faster than Walsh transform, whose matrix is composed of +1 and -1.
2. Input and output length are the same. However, the length should be a power of 2, i.e. $N = 2^k$.
3. It can be used to analyse the localized feature of signals. Due to the orthogonal property of Haar function, the frequency components of input signal can be analyzed.

Working of Haar transform

1. Form the orthogonal series of the periodic mean-square spectrum estimate.
2. Construct the ECG waveform. The objective of ECG signal processing is manifold and comprises the improvement of measurement accuracy and reproducibility (when compared with manual measurements) and the extraction of information not readily available from the signal through visual assessment. The ECG waveform will help in removing the noise in signals.
3. Then the Gaussian filter analysis is performed (explained in section 3.1.1)
4. The magnitude response is calculated giving the desired output. (Error free video).
5. Near-optimal performance is obtained at substantially reduced complexity, due to

the availability of fast computational schemes.

The figure represents the result for the images considering the time, frequency and amplitude.

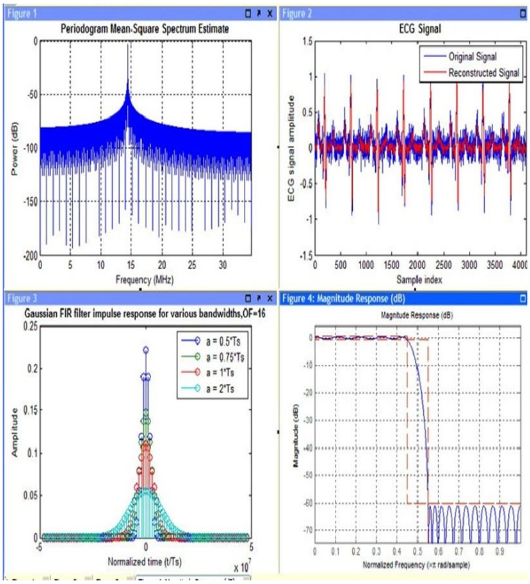

Figure 4: Haar Wavelet Transform Processes

5. CONCLUSION:

The paper proposed guides about the previous algorithm that were proposed for video object tracking and segmentation for smart surveillance system. Paper also explains the drawback that each algorithm were having (occlusion, illumination changes, least accurate results etc.).

In the next section of paper background segmentation and wavelet transform a more advanced algorithm are discussed The background segmentation algorithm extracts the background and foreground data using Gaussian mixture and level-set algorithm.

The wavelet transform helps in compressing the data collected for segmentation and vision analysis

Both the algorithm work well in solving the problems like occlusion, sudden background change, detecting stationary background, solving illumination change problem.

6. ACKNOWLEDGEMENT

Firstly, I express my key thanks to the authors whose paper I have studied and analyzed as without their contribution the paper would have not been made. I offer my sincere gratitude to my co-authors who helped in making the paper.

7. REFERENCES

[1] Galic S, Loncaric S, Ericsson Nikola Tesla “Spatio-temporal image segmentation using optical flow and clustering algorithm”, Image and Signal Processing and Analysis, Pula, Croatia, June 14-15, 2000,

[2] A. Elgammal, D. Harwood, and L. Davis, “Non-parametric model for background subtraction,” in *Proc. Eur. Conf. Comput. Vision*, May 2000,pp. 51–767.

[3] L. Cheng, M. Gong, D. Schuurmans, and T. Caelli, “Real-time discriminative background subtraction,” *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1401–1414, May 2011.

[4] Y. Benezeth; B. Emile; H. Laurent; C. Rosenberger (December 2008). "Review and evaluation of commonly-implemented background subtraction algorithms".19th International Conference on Pattern Recognition. pp. 1–4.

[5] R. Rosipal and L. Trejo, “Kernel partial least squares regression in reproducing kernel Hilbert space,” *J. Mach. Learn. Res.*, vol. 2, pp. 97–123, Mar. 2002.

[6] H. Saigo, N. Kramer, and K. Tsuda, “Partial least squares regression for graph mining,” in *Proc. ACM SIGKDD Int.*

Knowl. Discovery Data Mining Conf., 2008, pp. 578–586.

[7] O. Williams, A. Blake, and R. Cipolla, “Sparse bayesian learning for efficient visual tracking,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1292–1304, Aug. 2005.

[8] P. Kohli and P. Torr, “Dynamic graph cuts for efficient inference in Markov random fields,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29,no. 12, pp. 2079–2088, Dec. 2007.

[9] T. Tanaka, A. Shimada, D. Arita, and R.-I. Taniguchi, “Object detection under varying illumination based on adaptive background modeling considering spatial locality,” in *Lecture Notes in Computer Science*, vol.5414. 2009,

[10] S.-Y. Chien, Y.-W. Huang, B.-Y. Hsieh, S.-Y.Ma, andL.-G. Chen, “Fast video segmentation algorithm with shadow cancellation, global motion compensation, and adaptive threshold techniques,” *IEEE Trans. Multimedia*, vol. 6, no. 1, pp. 732–748, Oct. 2004.

[11] T. Bouwmans; F. El Baf; B. Vachon (November 2008). "Background Modeling using Mixture of Gaussians for Foreground Detection -A Survey". *Recent Patents on Computer Science* **1**: 219–237.

[12] D. Gutchess, M. Trajkovic, E. Cohen-Solal, D. Lyons and A.K. Jain. A Background Model Initialization Algorithm for Video Surveillance. In *Int. Conf. Computer Vision*, pages 733-740, 2001.

[13] H. Donelan and T.O’Farrell, “Method for generating sets of orthogonal sequences,” *Electron. Lett.*, vol. 35, no. 18, pp. 1537–1538, Sep. 1999.

[14] R. Poluri and A. N. Akansu, “New orthogonal binary user codes for multiuser spread spectrum communications,” in *Proc. EUSIPCO*, Antalya, Turkey, Sep. 2005, vol. 1, pp. 2–4.

[15] Jacques: Continuous wavelet transform, viewed 6 February 2010 "Novel method for stride length estimation with body area network accelerometers", *IEEE BioWireless 2011*, pp. 79-82

[16] Fino, B.J.;Algazi,V.R(1976). "Unified Matrix Treatment the Fast Walsh–Hadamard Transform".*IEEE* Transactions Computers **25** (11):11421146. Doi:10.1109 /TC.1976.1674569

[17] Sethian, James A. (1999). Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science. Cambridge University Press. ISBN 0-521 -64557-3.

[18] Richard .D.Cramer “PLS: Its strengthens & Limitations” perspective in drug discovery and design (1993) 267-278