A Robust Background Subtraction Approach Based on Daubechies Complex Wavelet Transform

Anand Singh Jalal and Vrijendra Singh

Indian Institute of Information Technology, Allahabad, India
anandsinghjalal@gmail.com, vrij@iiita.ac.in

Abstract. This paper describes a simple and robust approach for background subtraction in Daubechies complex wavelet domain. A background subtraction approach exploiting noise resilience capability of wavelet domain combined with local spatial coherence and median filter in the training stage is proposed. The effectiveness of the proposed approach is demonstrated via qualitative and quantitative evaluation measures on both indoor and outdoor video sequences. The experimental results illustrate that the proposed approach outperforms state-of-the-art methods.

1 Introduction

Moving object detection from a video sequences is the primary step in many computer vision applications. These applications include object tracking, human computer interaction, vehicle traffic analysis and visual surveillance. Although a lot of studies have been conducted in recent years, the subject is still challenging. Some of the popular approaches proposed in the literature include background subtraction method, optical flow method and statistical learning method [1]. Algorithmic complexity is the major disadvantage of optical flow method. It requires higher time span than other methods. The requirement of training samples and higher computational complexity makes statistical learning methods infeasible for real time processing. The background subtraction approach is one of the very popular ways for extracting foreground objects from video sequences [1]. In this approach, the current frame and a reference frame is compared, to extract the moving object. However, accurate detection could be difficult due to a number of variations in environments such as illumination, shadow and camera jittering.

In recent years, many solutions have been proposed for background modelling and subtraction. Frame differencing [1] is a simple and easy way to extract moving object from a video sequence. In this approach image difference between consecutive frames is used and considerable differences in pixels value are considered as foreground region. However, Frame differencing methods suffer from fat boundary and thresholding problem. Wren et al. [2] proposed a method to model the background independently at each pixel location using a single Gaussian distribution. A recursive updating using a simple linear filter is used to estimate the Gaussian parameter. However it fails whenever some kind of variations occurs in the background. Stauffer and Grimson [3] proposed a method known as Gaussian Mixture Model (GMM), to

A. Abraham et al. (Eds.): ACC 2011, Part II, CCIS 191, pp. 516–524, 2011.

handle multi-modal distributions using a mixture of several Gaussians. The mean (μ), standard deviation (σ) and weight (w) is updated recursively to imitate the new observations for pixel value. The GMM is the most representative approach and has been widely used [4]. However, the major disadvantage of GMM is that it is computationally intensive and requires a tricky parameter optimization. Elgammal et. al. [5] exploited a nonparametric kernel density estimation to build a background PDF. The probability density estimation is performed using the recent historical samples without any assumption about background and foreground. The model is robust in nature and has good model accuracy as compare to Gaussian mixture model in the more complex scenes. However, the high computation cost limits its scope.

In recent years, wavelet domain is used for moving object segmentation [6] [7] [8]. In [6], the double-edge problem in the spatial domain is overcome, using a change detection method with different thresholds in four wavelet sub-bands. In [7], the authors proposed a method to extract moving object by using three consecutive frame differences in discrete wavelet transform (DWT) domain for frames at times (n-1), (n) and (n+1) and edge map of frame at time (n). In [8] a real-time multiple objects tracking algorithm is proposed and the fake background motion is suppressed by performing the background subtraction method in 2-level real discrete wavelet transform. The first frame of the video sequence is assumed to be a background image. Guan [9] proposed a method for foreground segmentation and shadow suppression using HSV color space in Multi-Scale Wavelet domain. An optimal threshold is automatically computed to extract moving objects from video sequences. The moving objects are extracted using the hue value. Even though, these wavelet based methods show promising results. However, they are not adaptive in nature and tested against simple scenarios. Also discrete real wavelet transform (DWT) suffers from shift-sensitivity.

In the proposed work, we have taken advantage of the Daubechies complex wavelet transform properties to develop a robust background subtraction approach to extract moving objects. The Daubechies complex wavelet transform is approximately shift-invariant and has better directionality information with respect to DWT. The motivation is, the noise resilience nature of wavelet domain, as the lower frequency sub-band of the wavelet transform has the capability of a low-pass filter. In this paper, we have discussed a simple and effective approach of background modelling and subtraction by exploiting the low frequency sub-band characteristics of the object image in complex wavelet domain. Besides, this we have also exploited the local spatial coherence of the foreground pixels, to make the proposed method more robust against camera jittering and illumination changes.

The paper is organized as follows. Section 2 presents an overview of Daubechies complex wavelet transform. Section 3 describes the proposed approach. Experimental results are discussed in section 4 and finally the conclusion is presented in section 5.

2 Daubechies Complex Wavelet Transform

The discrete complex wavelet transform (CxWT) like discrete real wavelet transform (DWT) provides an efficient framework for representation and storage of images at multiple levels [10]. The wavelet transform divide an image into four sub-images.

These sub-images are label as approximation coefficients (LL), horizontal coefficient (LH), vertical coefficient (HL) and diagonal coefficient (HH). The approximation coefficient appears just like the compressed (filtered) original, while other coefficients contain the detailed information. One of the features of discrete wavelet transform (DWT) is that the spatial information is retained even after decomposition of an image into four different frequency bands [10].

Real DWT is non-redundant and an efficient tool to analyze signals. However, it suffers from the problem of shift sensitivity [5]. Complex wavelet transform (CxWT) can reduce these short comings. We have used Daubechies CxWT, as it is approximately shift-invariant and less redundant as compare to other complex wavelets [11].

Any function $f(t)$ can be decomposed into complex scaling function $\varphi(t)$ and a mother wavelet $\psi(t)$ as:

$$f(t) = \sum_k c_k^{j_0} \varphi_{j_0,k}(t) + \sum_{j=j_0}^{j_{\max}-1} d_k^j \psi_{j,k}(t)$$

where, j_0 is a low resolution level, { $c_k^{j_0}$ } and { d_k^j } are known as approximation [$\varphi(t) = 2\sum_n a_n \varphi(2t-n)$] and detail coefficients [$\psi(t) = 2\sum_n (-1)^n \overline{a_{1-n}} \varphi(2t-n)$].

Where $\psi(t)$ and $\varphi(t)$ shares the same compact support [-N, N+1] and $a_n\ s$ are coefficients. The $a_n\ s$ can be real as well as complex valued and $\sum a_n = 1$.

The Daubechies wavelet bases $\{\psi_{j,k}(t)\}$ in one dimension are defined through the above scaling function and multiresolution analysis of $L^2(\Re)$. During the formation of solution if we relax the Daubechies condition for a_n [11], it leads to complex valued scaling function. We have used this symmetric Daubechies complex wavelet transform for tracking.

3 Proposed Approach

In background subtraction approach, we compare current frame with a reference frame known as background image. A significant difference indicates the presence of moving objects. However, if the reference is not modeled or updated adequately, this approach can be highly vulnerable to environment conditions like illumination and structural background changes. Also, with larger sizes of images, the above pixel by pixel operation may tend to slow down the overall computation. Hence, background subtraction in the wavelet domain at a higher level is employed in the proposed method. Performing the background subtraction in a wavelet domain provides a noise resilience capability to the system.

Since a background is defined as temporally stationary part of the video, so background scene represents stationary pixels in the video. Thus the background image can be computed exploiting the fact that moving objects reside in only some portions of image frames and disappear over time. In the proposed algorithm an initial background model is obtained through a median filter and then it is recursively updated in the wavelet domain by adjusting parameters. The formulation of background modelling and subtraction in wavelet domain is as follows:

Let $F_n(x, y)$ corresponds to the intensity value at each (x, y) pixel location in the n^{th} image frame in the spatial domain and $W_\varphi^L F_n(k,l)$ represents the wavelet coefficient at (k,l) position in approximation subband at L^{th} level in the n^{th} image frame. So the reference image frame in wavelet domain is defined as:

$$W_\varphi^L B(k,l) = median(W_\varphi^L F_1(k,l), W_\varphi^L F_2(k,l), ..., W_\varphi^L F_N(k,l)) \tag{1}$$

Where $W_\varphi^L B$ represents the wavelet coefficient at (k,l) position in the reference image.

Although the assumption is that the background is temporally stationary, we do allow certain amount of variation to make the algorithm noise resilience and robust to environmental changes.

The background subtraction is exploited to extract the moving object(s).

$$W_\varphi^L F_n^{diff}(k,l) = \left| W_\varphi^L F_n(k,l) - W_\varphi^L B_n(k,l) \right| \tag{2}$$

Where $W_\varphi^L F_n(k,l)$, $W_\varphi^L B_n(k,l)$, $W_\varphi^L F_n^{diff}(k,l)$ represent the wavelet coefficient at (k,l) position in approximation subband at L^{th} level for current frame, background frame and difference frame respectively. This background subtraction task is followed by thresholding to get the foreground object.

$$\begin{aligned} I_{obj+sha}(k,l) &= 1, \quad \text{if } W_\varphi^L F_{diff}(k,l) \ge Th(k,l) \\ &= 0, \quad \text{otherwise} \end{aligned} \tag{3}$$

Where $I_{obj+sha}$ represents a binary mask containing object and shadow region.

The following updating strategies are applied to the background image using the knowledge of the pixel's classification in the current frame.

$$\left.\begin{aligned} &W_\varphi^L B_{n+1}(k,l) = (1-\alpha) W_\varphi^L B_n(k,l) + \alpha W_\varphi^L F_n(k,l) \\ &\qquad\qquad \text{If } (k,l) \text{ is background} \\ &W_\varphi^L B_{n+1}(k,l) = W_\varphi^L B_n(k,l) \\ &\qquad\qquad \text{If } (k,l) \text{ is foreground} \end{aligned}\right\} \tag{4}$$

Where α is an adapting rate having value between 0 to 1; a smaller value of α tend to slow convergence, while a large value makes the modelling too sensitive.

A threshold describing a statistically significant change in the value of wavelet coefficient at each pixel position (k,l) is used for thresholding. An empirically determined value is used to initialize the threshold value. For each pixel the threshold is updated regularly using the following equations:

$$\left.\begin{array}{l} Th_{n+1}(k,l) = sqrt((1-\alpha)Th_n(k,l)^2 + \alpha(|W_\varphi^L F_n(k,l) - W_\varphi^L B_n(k,l)|))^2 \\ \qquad\qquad \text{If } (k,l) \text{ is background} \\ Th_{n+1}(k,l) = Th_n(k,l) \\ \qquad\qquad \text{If } (k,l) \text{ is foreground} \end{array}\right\} \tag{5}$$

Since foreground pixels have a propensity to appear in a sets of connected points as a blob. So we do not perform a pixel wise computation rather a weighted average of values $\left|W_\varphi^L F_n(k,l) - W_\varphi^L B_n(k,l)\right|$ is computed in a 3×3 neighborhood centered at (k,l). So the eq. 2 and eq. 3 can be modified and a pixel is classified as foreground if:

$$\left|W_\Phi^L F_n(k,l) - W_\Phi^L B_n(k,l)\right| * C > W_\Phi^L Th_n(k,l) * C \tag{6}$$

Where C is 3×3 Laplacian distribution mask used for convolution.

We can take mask bigger than 3×3 but that may cause degradation in detection close to the boundaries of the foreground objects. The reason of using Laplacian distribution mask is that in many practical applications, the Gaussian assumption may not hold completely, particularly indoor scenes and compressed video sequences [12].

4 Experimental Results and Discussion

To evaluate the performance of the proposed approach, we conducted experiments on several video sequences. Here, we are showing the results on two videos. The first video is recorded in our campus (outdoor environment), having the problem of camera jittering and noise in the background due to fluttering leaves and other local movements. The second video is Hall Monitoring video which is a commonly used benchmark test sequence especially for evaluating the effectiveness of background subtraction techniques. Due to the several lighting sources the Hall Monitoring video suffers from noise and variation in indoor illumination.

We have compared the performance of the proposed background subtraction approach with frame differencing [1] method and recently proposed improved Gaussian mixture model (IGMM) method [13]. Both qualitative and quantitative measures are used to compare the segmentations results. In order to provide a quantitative perspective, we used the false positive rate (FPR) and false negative rate (FNR) measures as given in [14]. These measures are defined as:

$$\text{FNR}=\frac{\text{the number of foreground pixels wrongly classified}}{\text{the number of foreground pixels in the ground truth}}$$

$$\text{FPR}=\frac{\text{the number of background pixels wrongly classified}}{\text{the number of background pixels in the ground truth}}$$

The ground truth is generated by manually labeling the corresponding frames. The white and black colors represent the foreground and background pixels respectively. To ensure a fair comparison, the experimental results shown in Figs. 1-2 are computed without any morphological processing.

Fig. 1 shows the experimental results of different methods on video 1. The results demonstrate the robustness of the proposed approach against camera jittering and small background movements such as fluttering leaves etc. The reason that our method is able to handle these variations so well is because we are also considering the local spatial coherence to classify pixel as foreground or background. From the detection results, we can observe that the person in the fourth frame is detected quite accurately by the proposed method. While the other two methods unable to detect them accurately due to small size and camera jittering. It has been observed that IGMM approach produces very poor results after frame #430, due to the problem of camera jittering.

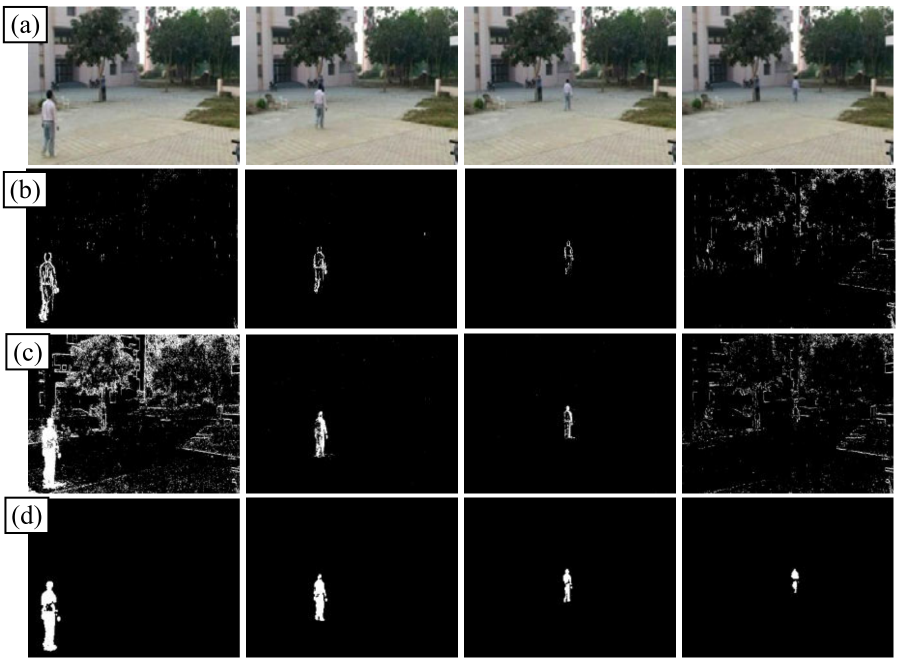

Fig. 1. Background subtraction results on Video 1 a) Original frames. b) Frame differencing results c) IGMM results d) Proposed method results.

In order to test the effectiveness of the proposed method in an indoor environment and lighting variations, we performed experiments on a well known indoor dataset i.e. 'Hall Monitoring'. Fig. 2 displays the results of the different methods on this video and shows the capability of the proposed method to handle the noise due to variations in the illumination. On the other hand for frame differencing and IGMM method, the noise is still considered as the foreground as shown in figs. 2(b-c).

From the experiments, we can observe that frame differencing method is very sensitive to variations due to illumination changes as shown in fig. 2(b). It is also observed that in frame differencing method, the extracted moving object is not filled in the overlapping area as shown in figs. 1(b) and 2(b). Yet another problem in frame differencing approach is selection of the threshold. Selecting a single threshold for all pixels makes it sensitive to different kinds of variations. Even though the IGMM method gives good results in general, but the method also shows poor performance in complex situations such as lighting variations, camera jittering as shown in fig. 1(c) and fig. 2(c).

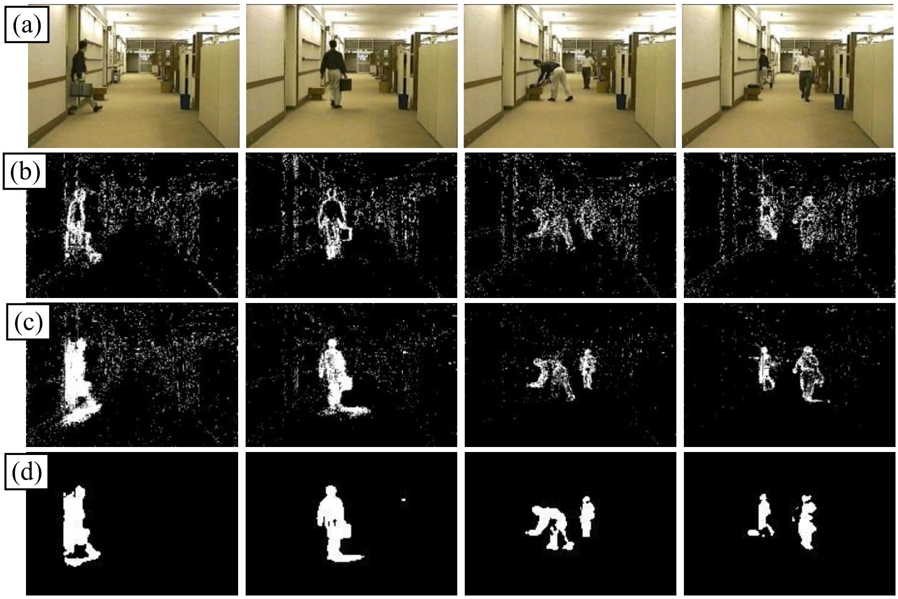

Fig. 2. Background subtraction results on Video 2 (Hall Monitoring) a) Original frames. b) Frame differencing results c) IGMM results d) Proposed method results

Table 1. Comparison of false positive rate and false negative rate

#Frame	Video 1						#Frame	Video 2					
	Proposed Method		IGMM		FD			Proposed Method		IGMM		FD	
	FP	FN	FP	FN	FP	FN		FP	FN	FP	FN	FP	FN
5	.0031	.1909	.1393	.1180	.0061	.5687	25	.0164	.1477	.0617	.0884	.0452	.4747
143	.0016	.1129	.0018	.3582	.0013	.7024	51	.0104	.0622	.0233	.1387	.0424	.6390
303	.0007	.2532	.0006	.5487	.0003	.7755	117	.0074	.2382	.0096	.5228	.0418	.6397
432	.0004	.3324	.0352	.8171	.0304	.9724	227	.0082	.3218	.0116	.4685	.0432	.5584

Table 1 illustrates the FPR and FNR values for different methods. It is evident from Table 1 that the proposed background subtraction method gives less false positives than the comparison methods. The values of false negative are also very less in most of the image sequences in our method. In our method, most of the false positive and false negatives occur on the boundary area of the moving object, because neighborhood pixels are considered in the classification. It should be noticed that the large false negative in case of FD is mostly due to misclassification in the inner areas of moving object. From the visual and numerical interpretation, as depicted in Figs. 1-2 and Table 1, it is clear that the proposed background subtraction method outperforms the other comparative methods, especially in conditions of camera jittering and illumination variations.

5 Conclusion

This paper addresses the problem of background subtraction in complex wavelet domain. In the proposed algorithm a temporal median filter is used to generate initial background model in a training stage and then foreground pixels are obtained by applying background subtraction scheme in the subsequent frames. The local spatial coherence of image pixels is exploited to makes the proposed approach more robust against illumination changes and camera jittering. We have demonstrated the robustness of the proposed approach in different video sequences with different kind of complexities.

References

1. Piccardi, M.: Background Subtraction Techniques: a Review. In: Proc. IEEE Int. Conf. Systems, Man, Cybernetics, pp. 3099–3104 (2004)
2. Wren, C., Azarbayejani, A., Darrell, T., Pentland, A.P.: Pfinder: Real Time Tracking of the Human Body. IEEE Trans. Pattern Analysis and Machine Intelligence 19(7), 780–785 (1997)
3. Stauffer, C., Grimson, W.E.L.: Adaptive Background Mixture Models for Real-Time Tracking. In: Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, pp. 246–252 (1999)
4. Parks, D.H., Fels, S.S.: Evaluation of Background Subtraction Algorithms with Post-processing. In: Proc. IEEE Int'l Conf. Advanced Video and Signal-based Surveillance, pp. 192–199 (2008)
5. Elgammal, A., Duraiswami, R., Harwood, D., Davis, L.S.: Background and Foreground Modeling using Non-parametric Kernel Density Estimation for Visual Surveillance. Proceedings of the IEEE, 1151–1163 (2002)
6. Huang, J., Hsieh, W.: Wavelet-based Moving Object Segmentation. IEE Electronic Letters 39(19), 1380–1382 (2003)
7. Huang, J., Hsieh, W.: Double Change Detection Method for Wavelet-based Moving-Object Segmentation. IEE Electronic Letters 40 (2004)
8. Cheng, F.H., Chen, Y.L.: Real Time Multiple Objects Tracking and Identification Based on Discrete Wavelet Transform. Pattern Recognition 39(6), 1126–1139 (2006)
9. Guan, Y.-P.: Spatio-temporal Motion-based Foreground Segmentation and Shadow Suppression. IET Computer Vision 4(1), 50–60 (2010)
10. Wang, Y., Doherty, J.F., Duck, R.E.V.: Moving Object Tracking in Video. In: Proceedings of 29th IEEE Int'l Conference on Applied Imagery Pattern Recognition Workshop, pp. 95–101 (2000)
11. Lina, J.-M.: Image Processing with Complex Daubechies Wavelets. Journal of Mathematical Imaging and Vision 7(3), 211–223 (1997)

12. Kim, H., Sakamoto, R., Kitahara, I., Toriyama, T., Kogure, K.: Robust Silhouette Extraction Technique using Background Subtraction with Multiple Thresholds. Optical Engineering 46(9) (2007)
13. Zivkovic, Z., Heijden, F.: Efficient Adaptive Density Estimation per Image Pixel for the Task of Background Subtraction. Pattern Recognition Letters 27(7), 773–780 (2006)
14. Heikkila, M., Pietikainen, M.: A Texture-Based Method for Modeling the Background and Detecting Moving Objects. IEEE Trans. Pattern Analysis Machine Intelligence 28(4), 657–662 (2006)