OPTIMAL WAVELET DIFFERENCING METHOD FOR ROBUST MOTION DETECTION

Vladimir Crnojević, Borislav Antić and Dubravko Ćulibrk

Faculty of Technical Sciences, University of Novi Sad, Serbia

ABSTRACT

In real-world surveillance systems, where variation of light and camera parameters can sometimes severely impair the normal operation of background subtraction algorithms, better results are obtained with differencing schemes. We have earlier demonstrated that differencing of detail images produced by wavelet transformation can lead to more stable detection results. In this paper, we considerably extend that framework, by introducing the modified z-scores calculated from wavelet coefficient differences. Foreground pixels are detected as outliers in normal distribution by modified z-score test. The threshold value used in the outlier test is optimized by maximizing the precision and recall measures on several training frames. Finally, the elimination of ghosts from motion detection is done by double modified z-score testing, that is similar in idea to double frame differencing. The resulting motion detection method shows considerable resilience to changes in illumination and camera parameters and also produces a lower amount of detection errors than other motion detection methods.

Index Terms— Motion detection, Frame differencing, Wavelet transformation, Video surveillance

1. INTRODUCTION

Many important video analysis applications, such as visual motion tracking, human action or activity recognition and motion trajectory analysis, rely on the detection of moving objects as their initial step. Previous attempts to discover moving foreground objects in the scene were primarily targeted at building a comprehensive statistical model for the image background and its effective use through a set of techniques known as *background subtraction*. A good review of these techniques is provided by Piccardi [1].

Wren *et al.* [2] proposed to use a single Gaussian distribution for each pixel in the scene. After the initialization phase, during which the Gaussian models were built, Gaussian distributions are recursively updated using a simple adaptive filter. However, a single-mode distribution is incapable of modeling repetitive background motion (e.g. swaying trees). In their prominent paper, Stauffer and Grimson [3] modeled each pixel as a mixture of Gaussians (MoG) and used an online approximation to the *expectation-maximization* (EM) algorithm to update the model. The Gaussian distributions are then assessed to determine those which most likely belong to a background. However, backgrounds having fast variations are not accurately modeled with just a few Gaussians of a typical MoG model. Li *et al.* [4] proposed a Bayesian framework with *principal* features (BFPF) under which the background is represented by the most significant and frequent features at each pixel location. Their model incorporates spectral, spatial and temporal features. Static pixels are described with spectral and spatial features of the image, while dynamic pixels are described with temporal features. The potential problem with this method is that it can wrongly learn the features of foreground objects as the background if too many foreground objects are present in the scene. The codebook (CB) model for foreground-background segmentation, introduced by Kim *et al.* [5], samples pixel values at each location over long period of time without making any parametric assumption. The CB model can handle scenes containing moving backgrounds or illumination variations to some extent.

Background modeling is closely related to the problem of change detection. The goal of change detection is to identify the set of pixels with a significant difference between the last and previous images of a video sequence. Frame differencing, as a basic method for change detection, performs thresholding of the image differences between two consecutive video frames. If a sudden change in illumination or camera internal parameter occurs, frame differencing will produce a smaller amount of false positives than background subtraction methods. However, frame differencing is susceptible to *aperture* problem and *ghosting* [6]. The first problem refers to the low textured parts of moving objects, that are erroneously labeled as background. The second problem corresponds to false detections that occur when moving objects uncover some part of the background. In order to eliminate ghosts from the change detection results, Kameda and Minoh [7] proposed to use double frame difference (DFD). This method thresholds the differences between frames $t+1$ and t, and between frames t and $t-1$, and then combines the results using logical AND operation.

To get more consistent change detection masks than those obtained by simple and double frame differencing, we have previously proposed [8] to use differences of detail images

This research has been supported in part by EUREKA!4160 Project.

that were generated by an undecimated wavelet transformation. The elimination of ghosts that appear in thresholded difference images is achieved by performing a significance test on undecimated wavelet transformation coefficients. Only image locations with large wavelet coefficients that change a lot between two consecutive frames are labeled as foreground.

In this paper, we considerably extend the wavelet differencing framework introduced in [8]. Firstly, instead of simple wavelet differences, we calculate at each scale and location the modified z-score [9] from wavelet differences. The problem of finding the pixels that belong to moving objects is recast as that of detecting outliers among the normally distributed wavelet differences. Outlier detection is performed using modified z-score test, that compares the modified z-scores to a fixed threshold. Secondly, the fixed threshold value used in the statistical outlier test is established through an optimization procedure, that effectively minimizes the number of false positives and negatives for several manually segmented frames. Finally, the elimination of ghosts from motion detection masks is not accomplished by significance testing, but through a double modified z-score testing that is motivated by the idea of double frame differencing. This leads to a simpler optimization procedure, that needs to find only one threshold value instead of two used in the change detection and significance tests [8]. The resulting motion detection method shows considerable resilience to illumination and camera internal parameter changes (e.g. a change in automatic exposure and aperture). It also produces a lower amount of detection errors than other differencing and non-differencing techniques.

The rest of the paper is organized as follows. Section 2 explains in detail the proposed optimal wavelet based approach for motion detection. Experimental results are given in Section 3 and conclusion is given in Section 4.

2. OPTIMAL WAVELET BASED DETECTION OF MOTION

Multi-scale image representation plays a key role for understanding the saliency of visual information as perceived by humans and can be successfully used for object recognition. Recent visual neuroscience experiments suggest that robust object detection can be realized by sampling images at multiple scales so as to have access to context, shape and texture [10].

As in the human visual system, more robust motion detection and tracking can be achieved by using only the spatial gradient information derived from multi-resolution image representation. Our experiments show that undesirable changes of illumination or camera internal parameters have more influence on the low-pass information in images, than on the high-pass information. Therefore, low-pass image components are not used for the motion detection in this paper, but only the detail images generated by an undecimated wavelet transformation. The reason for choosing the undecimated wavelet transformation is that its detail images are already aligned because no decimation is performed. Detailed discussion of filter banks used for undecimated wavelet transformation is provided in [11].

At the beginning, the undecimated wavelet transformation of each frame of video sequence is calculated. Since scaling coefficients are not of interest in this paper, the wavelet transformation will provide only the set of wavelet coefficients $w_j^d(x,y,t)$, where $j = 1,...,J$ denotes a resolution level (scale), $d = LH, HL, HH$ denotes a subband orientation, and x,y and t denote spatial coordinates and frame number. The temporal change of wavelet coefficients is given as

$$\Delta_j^d(x,y,t) = w_j^d(x,y,t+1) - w_j^d(x,y,t). \quad (1)$$

If there is no motion in the image, the obtained wavelet differences at each resolution level and orientation are normally distributed with a zero mean and a small standard deviation. However, moving objects in the scene will cause some wavelet coefficients to change drastically, and the respective wavelet differences will become outliers to the normal distribution. Hence, the problem of finding pixels that correspond to motion is akin to the problem of detecting outliers in the set of wavelet differences.

Modified z-score test [9] is used to detect the outliers to the normal distributions by comparing modified z-score $Z_j^d(x,y,t)$ with a fixed threshold τ. The modified z-score is calculated by normalizing the wavelet coefficient difference with outlier resistent estimator of standard deviation. The median of absolute deviation about the median (MAD) is such an estimator [12], and here is calculated as $MAD(\Delta_j^d(x,y,t)) = median_{x,y}\{|\Delta_j^d(x,y,t)|\}$. Modified z-score is computed as

$$Z_j^d(x,y,t) = 0.6745 \cdot \frac{\Delta_j^d(x,y,t)}{MAD(\Delta_j^d(x,y,t))}. \quad (2)$$

Under the assumption that moving objects represent a smaller part of the frame, modified z-score test yields an outlier detection mask that identifies the pixels that correspond to visually distinctive parts of moving objects

$$O(x,y,t) = \begin{cases} 1 & \text{if } |Z_j^d(x,y,t)| > \tau \text{ for some } j,d \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

The elimination of ghosts from outlier detection mask is achieved by performing modified z-score test twice, and to mark as foreground only those locations that are both times characterized as outliers

$$FG(x,y,t) = \begin{cases} 1 & \text{if } O(x,y,t) = 1 \text{ and } O(x,y,t-1) = 1 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

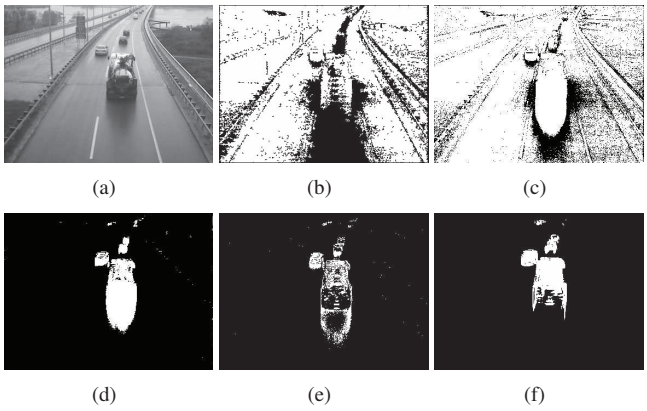

Fig. 1. Motion detection results for a frame of the *Bridge* video sequence. In order to regulate the amount of light after a large vehicle had entered the scene, the camera automatically increased the aperture size. The correction of aperture size caused a change of the values of many pixels in the image. (a) The original video frame. (b) Mixture of Gaussians model [3]. (c) Codebook model [5]. (d) Bayesian framework with principal features [4]. (e) Double frame differencing [7]. (f) Proposed algorithm.

The performance of the proposed wavelet based motion detection depends on the choice of fixed threshold value τ, used in the outlier detection test. As τ increases, the number of false positives (FP) falls, but the number of false negatives (FN) rises. If τ decreases, the opposite happens - FP rises, but FN falls. To find a value of threshold τ that minimizes both types of error, we performed the following optimization step on several ground truthed frames

$$\tau^{opt} = \arg\max_{\tau}\{w_p \cdot \frac{TP}{TP+FP} + w_r \cdot \frac{TP}{TP+FN}\}. \quad (5)$$

The first term is called *Precision* and it effectively minimizes the number of false positives. The second term is called *Recall* and it effectively minimizes the number of false negatives. Weights w_p and w_r can be adjusted to make criterion biased towards detecting accurate object outline (less false positives) or filling the internal holes in the detected moving objects (less false negatives). With $w_p = 2$ and $w_r = 1$, the obtained optimal threshold value is close to 10.

3. EXPERIMENTAL RESULTS

Experimental validation of the proposed method for motion detection and the comparison with other state-of-the-art methods is performed on two traffic video surveillance sequences. One sequence is recorded by a bridge traffic surveillance system, and exemplifies some challenging problems in real-world surveillance systems, such as sudden illumination change or camera internal parameter adjustment due to automatic control. Another video is the well-known *Karlsruhe Ettlinger-Tor* sequence, that is a short recording of the urban traffic. The proposed optimal wavelet differencing method for robust motion detection has been set to use three resolution levels of the undecimated Haar wavelet transformation.

Table 1. Quantitative analysis of the motion detection results given in Fig. 1, for a frame of the *Bridge* video sequence. Numbers in the table represent the percentages of true positives (TP), false positives (FP), false negatives (FN), precision (P), recall (R) and F-measure. Methods of comparison are Mixture of Gaussians (MoG), Codebook model (CB), Bayesian framework with principal features (BFPF), Double frame difference (DFD) and the proposed optimal wavelet based detection.

	TP	FP	FN	P	R	F
MoG [3]	3.1	69.0	1.3	4.3	70.6	8.1
CB [5]	4.3	85.5	0.1	4.8	**97.0**	9.1
BFPF [4]	4.2	4.6	0.2	47.9	95.6	63.8
DFD [7]	2.0	1.7	2.4	53.8	45.6	49.4
Proposed	3.4	1.5	1.0	**69.9**	78.0	**73.7**

Figure 1 shows a challenging situation for motion detection algorithms, when a large vehicle enters the scene and changes the amount of light in the image. In order to counteract the light change, the camera automatically adjusts the aperture size, but it causes a change of the values of many pixels in the image. Mixture of Gaussians (MoG) [3] and Codebook (CB) models [5] do not adapt quickly to the sudden change of pixel levels, and thus produce a lot of false positives. Bayesian framework with principal features (BFPF) [4] and Double frame differencing (DFD) [7] have more stable detection results, but they are still susceptible to shadows and light reflections. Table 1 provides a quantitative analysis of the motion detection results presented in Fig. 1. Percentages of the true positives (TP), false positives (FP) and false negatives (FN) are expressed relative to the total amount of pixels in the image. Beside the Precision (P) and Recall (R) measures that have been already defined in Section 2, numerical results are also provided for the F-measure, a metric for evaluating the integral performance of a detector. The F-measure is calculated as a harmonic mean of the Precision and Recall measures

$$F = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}. \quad (6)$$

Figure 2 and Table 2 provide the qualitative and quantitative analysis of motion detection results obtained for a frame of the *Karlsruhe Ettlinger-Tor* video sequence. As can be noted from Tables 1 and 2, in both experimental cases the proposed optimal wavelet differencing method outperformed other compared methods, yielding a higher F-measure and better motion detection. All video materials used in the paper can be found at `www.ursusgroup.com`.

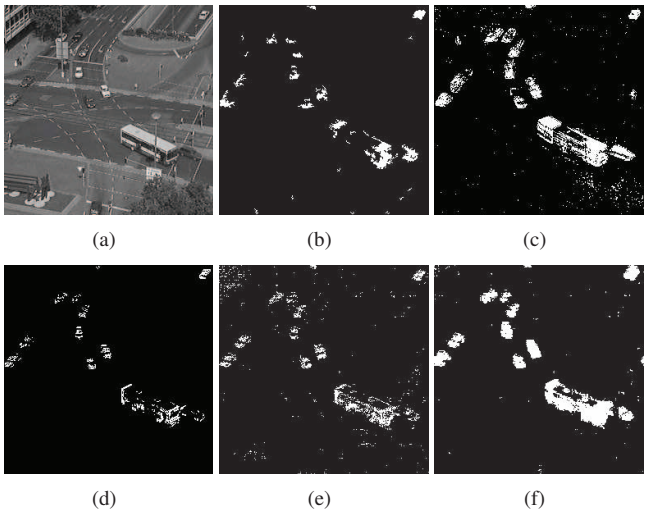

Fig. 2. Motion detection results for a frame of the *Karlsruhe Ettlinger-Tor* video sequence. (a) The original video frame. (b) Mixture of Gaussians model [3]. (c) Codebook model [5]. (d) Bayesian framework with principal features [4]. (e) Double frame differencing [7]. (f) Proposed algorithm.

Table 2. Quantitative analysis of the motion detection results given in Fig. 2, for a frame of the *Karlsruhe Ettlinger-Tor* video sequence. Please refer to Table 1 for an explanation of used symbols.

	TP	FP	FN	P	R	F
MoG [3]	1.7	0.9	3.8	64.3	30.5	41.4
CB [5]	4.3	13.6	1.1	24.3	**80.2**	37.3
Li [4]	2.9	2.0	2.5	58.7	53.2	55.8
DD [7]	2.1	1.0	3.4	**66.2**	37.9	48.2
Proposed	4.1	2.3	1.3	64.3	75.8	**69.6**

4. CONCLUSION

This paper proposes a novel algorithm for motion detection that is more resilient to changes in illumination and camera parameters than other motion detection methods. The wavelet differencing scheme, that operates on detail images of a wavelet transformation, has been considerably extended in this paper. The modified z-scores calculated from wavelet coefficient differences have been incorporated in order to classify pixels as foreground/background based on the modified z-score outlier test. The threshold value used in the outlier test is also optimized using several training frames with manually segmented foreground objects. At last, ghosts in motion detection results have been eliminated by double modified z-score testing. Experimental validation shows that the proposed method produces a lower amount of detection errors than other detection methods.

5. REFERENCES

[1] M. Piccardi, "Background subtraction techniques: a review," in *IEEE SMC 2004 Int. Conf. Syst., Man Cybern.*, Oct. 2004, vol. 4, pp. 3099–3104.

[2] C. Wren, A. Azabayejani, T. Darrell, and A. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 780–785, July 1997.

[3] C. Stauffer and E. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–757, Aug. 2000.

[4] L. Li, W. Huang, I.Y.H. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection," *IEEE Trans. Image Process.*, vol. 13, no. 11, pp. 1459–1472, Nov. 2004.

[5] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. S. Davis, "Real-time foreground-background segmentation using codebook model," *Real-Time Imaging*, vol. 11, no. 3, June 2005.

[6] D. Migliore, M. Matteucci, M. Naccari, and A. Bonarini, "A revaluation of frame difference in fast and robust motion detection," in *Proc. 4th ACM Int. Workshop on Video Surveillance and Sensor Networks (VSSN)*, Oct. 2006, pp. 215–218.

[7] Y. Kameda and M. Minoh, "A human motion estimation method using 3-successive video frames," in *Proc. Int. Conf. on Virtual Syst. and Multimedia (ICVSM)*, Sep. 1996, pp. 135–140.

[8] B. Antić, V. Crnojević, and D. Ćulibrk, "Efficient wavelet based detection of moving objects," in *Proc. 16th Int. Conf. Digital Signal Process. (DSP)*, July 2009, pp. 1–6.

[9] V. Barnett and T. Lewis, *Outliers in Statistical Data*, Wiley Series in Probability & Statistics. Wiley, April 1994.

[10] S. Bileschi, "Object detection at multiple scales improves accuracy," in *19th Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2008, pp. 1–5.

[11] F. Murtagh J.L. Starck, J. Fadili, "The undecimated wavelet decomposition and its reconstruction," *IEEE Trans. Image Process.*, vol. 16, pp. 297 – 309, 2007.

[12] A. Pizurica and W. Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising," *IEEE Transactions on Image Processing*, vol. 15, no. 3, pp. 654–665, March 2006.