Hybrid wavelet-based Video tracking using Adaptive Contrast Change Detection in night-time Visual Surveillance Systems

Sherin M. Youssef, Meer A. Hamza, Arige F. Fayed

***Abstract*—A new method is proposed for detecting and tracking multiple moving objects on night-time lighting conditions. The method is based on integrating both the wavelet-based contrast change detector and locally adaptive thresholding scheme. For outdoor surveillance at night, the size of distant moving objects can be small and the motion between two adjacent images can be very small. The normalized cross-correlation is used to describe the correlation between frames. Normalized cross-correlation has the advantage of not being affected significantly by global changes in the lighting from one image to another. In the first stage, the contrast in local change over time is used to detect potential moving objects. Then motion prediction and spatial nearest neighbor data association are used to suppress false alarms. The experimental results prove the feasibility and usefulness of the proposed method. Experiments on real scenes show that the algorithm is effective for night-time object detection and tracking.**

***Index Terms*—Adaptive threshold, contrast change, tracking, moving objects.**

I. Introduction

Object detection and tracking at night remain very important problems for visual surveillance. The objects are often distant, small and their signatures have low contrast against the background. Traditional methods based on the analysis of the difference between successive frames and a background frame will do not work. Automated video surveillance is an important research area in the commercial sector [1, 2].

Surveillance cameras are already prevalent in commercial establishments, with camera output being recorded to tapes that are either rewritten periodically or stored in video archives. Keeping track of people, vehicles, and their interactions in an urban or battlefield environment is a difficult task [3, 4, 5].

Object detection and tracking at night remain very important problems for visual surveillance [2, 5]. The objects are often distant, small and their signatures have low contrast against the background. Traditional methods based on the analysis of the difference between successive frames and a background frame will do not work. Kaiqi3333qqq13e`6 Huang et.al [2] presented a novel real time object detection algorithm proposed for night-time visual surveillance. Experiments on real scenes show that the algorithm is effective for night-time object detection and tracking, the threshold is set by hand. Further enhancement on thresholding scheme was required.

Ying Wu and Ting Yu [5] presented a new approach based on a two-layer statistical field model that characterizes the prior of the complex shape variations as a Boltzmann distribution and embeds this prior and the complex image likelihood into a Markov field. This new approach has several advantages. It is intrinsically suitable for capturing local non-rigidity and is robust to partial occlusion. However, it is not robust for night-time surveillance.

In this paper, a novel object detection algorithm is proposed for night-time visual surveillance using adaptive threshold scheme. The algorithm is based on contrast analysis. In the first stage, the contrast in local change over time is used to detect potential moving objects. Then motion prediction and spatial nearest neighbor data association are used to suppress false alarms. A new change detector mechanism is used to detect the changes in a video sequence and divide the sequence into scenes to be encoded independently. Using the change detector algorithm (CD), it was efficient enough to detect abrupt cuts and help divide the video file into sequences. If the normalized difference is high then an abrupt cut is present. Fuzzy rules are used to video scenes into continuous, gradual change, or abrupt change cut type.

The paper is organized as follows. Section II introduces the discrete wavelet transform model. Section III presents the Object Detection and Tracking Algorithm for Night-time Virtual Surveillance using locally adaptive thresholding and contrast change detector. Experimental results and analysis are illustrated in section IV. Finally, conclusions are driven in section V.

II. Discrete wavelet transform

Wavelet transform provides a special basis that a signal can express features easily and efficiently. Localization in

Manuscript received February 28, 2010.

S. M. Youssef is with the Computer Engineering Department, College of Engineering & Technology, Arab Academy for Science and Technology, P.O.1029, Alexandria., EGYPT (phone:2-0109704237, 203-5622366; fax: 203-562-0245 (e-mail:sherin@aast.edu).

M. A. Hamza is with the College of Management and Technology, Arab Academy for Science and Technology, P.O. 1029, Alexandria., EGYPT (e-mail: meerh@aast.edu).

A. F. Fayed is with the Computer Engineering Department, College of Engineering & Technology, Arab Academy for Science and Technology, P.O. 1029, Alexandria., EGYPT (e-mail: arige_99@yahoo.com).

both frequency and time/spatial domains is the greatest advantage of discrete wavelet transform (DWT) over Fourier transform-based methods. The spatial localization indicates that after the wavelet transform takes place, the coefficients in a certain position at the wavelet sub-images correspond to the details of different frequencies in the corresponding spatial location [6, 7]. Two dimensional DWT can be used to decompose an image into four sub-images. Filters are applied in one dimension first, vertically or horizontally and then in the other dimension. Down-sampling is performed in two stages to reduce the overall number of computation. The wavelet transform connotes hierarchical features, and an image example can be decomposed into four sub-images. The four sub-images that the wavelet transform preserves not only the frequency features but also spatial features. An original image can be decomposed into four different bands (LL, HL, LH, HH) by using the discrete wavelet transform. These sub-bands contain different frequency characteristics after filtering. The high-pass filter extracts the high frequency part and the low-pass filter gives the low frequency information representing the most energy of an image [8, 9]. In this paper, only the low frequency part is used for processing due to the consideration of low computing cost and noise reduction issue. The proposed method decreases the computing cost by using three-level DWT.

III. MOTION DETECTION AND OBJECT TRACKING

A. Change Detection and Sequence Tester

A new change detector is used to detect the changes in a video sequence and divide the sequence into scenes to be encoded independently. A further enhancement is done using a sequence tester to test each scene for high activity. (Due to motion) and decide whether or not another reference frame (key frame) is needed. Simulation results prove the enhanced performance of the suggested. The mechanism proposed here, is based on segmenting the video into scenes using a change detector. For each scene, one or more key frames are chosen. The number and location of the key frame is chosen according to the length and characteristics of the scene. Each frame is then differenced from the nearest key frame.

A scene [10] is a sequence of shots that belong together semantically. There are two kinds of shot boundaries in videos: Abrupt shot changes called "cuts", and, Gradual transitions between two different shots. The proposed scene cut detector is simple and fast. The objective of the detector is to divide the whole video into independent sequences to be encoded separately. This is done in 2 steps. Firstly, the video as a whole is used to get the global cuts, thus dividing the video into sequences this is done by the change detector (CD). Secondly, each sequence is further investigated to ensure that there is no high level of changes between the frames that may cause artifacts in the decoded video. The results are efficient since it keeps complexity low and the presence of this module in the code improves efficiency.

Change Detector (CD)

The detector uses the normalized sum of absolute difference between each two successive frames on a pixel-by-pixel basis. If the normalized difference (normalized_diff) is high then an abrupt cut is present. It can be described in the following rules:

If (normalized_diff is low) then (cut_type is continuous)
If (normalized_diff is mid) then (cut_type is gradual change)
If (normalized_diff is high) then (cut_type is abrupt change)

Using the change detector algorithm (CD), it was efficient enough to detect abrupt cuts and help divide the video file into sequences. But in the case of long scenes where motion occurs, the high level of changes between frames must be considered. This is because the detector depends on differencing frames. For each scene a key frame is chosen and all the other frames are differenced from it. When long scenes with high level of changes between frames especially with respect to the key frame, the encoded frame carries a lot of information that may be lost due to lousy encoding and decoding so another key frame must be used.

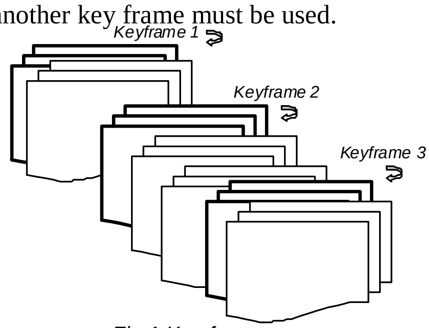

Fig.1 Key frames

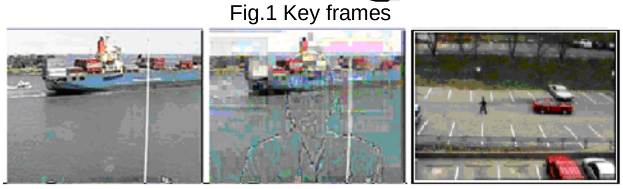
Fig.2 Change detector

For each scene a key frame is chosen and all the other frames are differenced from it. The change detector algorithm is used to determine the gradual change and abrupt change.

B. Object Detection and Tracking Algorithm for Night-time Virtual Surveillance

The object detection algorithm includes two steps. In the first step, the object is detected using local contrast computed over the entire image. In the second step the detected objects are tracked and falsely detected objects are removed using feedback from the tracking algorithm. Assume that I is the image frame, R is the inter-frame relation which describes the similarity between two frames, T1 is the contrast threshold and T is the contrast change threshold. Contrast-Change (CC) Image IC_{ji} is calculated between each two successive frames. This contrast-change image is deployed for object mask detection, which is used for object tracking. The details of the algorithm are given in the following section.

C. Content Detection based on Local Contrast Computation

To make the detection problem more precise, let {I1, I2 … IM} be an image sequence in which each image maps a pixel coordinate $x \in R^l$ to an intensity or color $l(x) \in R^k$. Typically, k = 1 (e.g. gray-scale images) or k = 3 (e.g. RGB color images). There are many methods to compute contrast [12, 13]. Typically, luminance contrast is defined as the relative difference between luminance of the object, Lo, and the surrounding background, LB, as C = (Lo − LB)/LB, which is called Weber contrast [11]. Michaelson defined contrast for elementary patterns as C = (L_{max} − L_{min}) / (L_{min} + L_{max}) [14]. We make use of a simple measure of contrast defined as the local standard deviation σ_L of the image intensities divided by the local mean intensity μ_L [15],

$$C_L = \frac{\sigma_L}{\mu_L} \tag{1}$$

The local mean intensity $\mu^{(p,q)}$ of a (2p + 1) x (2q + 1) block of pixels is

$$\mu_L^{(p,q)} = \frac{1}{(2p+1)(2q+1)} \sum_{i-p}^{i+p} \sum_{j-p}^{j+p} I(i,j) \tag{2}$$

The local standard deviation $\sigma^{(p,q)}$ of the block is

$$\sigma_L^{(p,q)} = \left(\frac{1}{(2p+1)(2q+1)} * \sum_{i-p}^{i+p} \sum_{j-p}^{j+p} \left[I(i,j) - \mu^{(p,q)}(i,j)\right]^2\right)^{1/2} \tag{3}$$

The local contrast is related to entropy in a statistical sense [16], but local contrast is simpler and faster to compute. It is clear that objects can be detected when the contrast exceeds a threshold. We can see that the mean is random (from very low to very high) and is thus of little use for detection. The standard deviation can describe the local contrast in some degree [17] but it is still not appropriate for reliable object detection. The reliable detection of objects can be achieved by thresholding the contrast. Some sub-images containing objects have very low contrast. These sub-images cannot be found by contrast detection alone. In addition, there are some sub-images which have a high contrast but which do not contain moving objects. The following formula is used to compute the mask for local contrast salience map M_{CM} [18]:

$$M_{CM}(x,y) = \begin{cases} 1 \; if \; C^{(p,q)}(x,y) \geq T \\ 0 \; if \; C^{(p,q)}(x,y) < T \end{cases} \tag{4}$$

$$I_{CM} = I(x,y) \cdot M_{CM} \tag{5}$$

where $C^{(p,q)}$(x, y) is the contrast image after local contrast computation, I (x, y) is the original image and I_{CM} shows the locations of the interested object. It is observed that the interested objects (people) and image structures (boundaries of illuminated areas) are detected by this step.

D. Moving Object Detection based on Changing Contrast

In the second step of the algorithm, changes in contrast saliency are used to filter the results of the first step and obtain the moving objects. This step is indicated by the following equations:

$$MD_{Cji}^{[p,q]} = T\left(\left|I_{Cmi}^{[p,q]} - I_{Cmj}^{[p,q]}\right|\right) \tag{6}$$

$$I_{Cji}^{[p,q]} = I_i \cdot MD_{Cij}^{[p,q]} \tag{7}$$

where $I^{[p,q]}_{Cmi}$ and $I^{[p,q]}_{Cmj}$ are contrast images obtained from the i^{th} image and the j^{th} image, [p, q] is the window for contrast computation in the first step, $MD^{[p,q]}_{Cji}$ is the mask for contrast difference image between two contrast salience images. T is the threshold to remove small changes caused by other factors such as noise or small variations in the light. The image $I^{[p,q]}_{Cij}$ takes large values in regions where the contrast is simultaneously high and changing.

Inter-frame Relation (R)

For outdoor surveillance at night, the size of distant moving objects can be as small as 2×3 pixels and the motion between two adjacent images can be very small. We choose the time interval between images by variance analysis as shown in formula 8. There are some methods such as mean absolute difference (MAD) or the sum of absolute differences (SAD) to decide on the time interval between frames. Here the normalized cross-correlation is used to describe the correlation between frames I_i and I_j. Normalized cross-correlation has the advantage of not being affected significantly by global changes in the lighting from one image to another. The correlation coefficient R is defined by

$$R = \frac{\sum_m \sum_n (I_{i_{mn}} - \bar{I_i})(I_{j_{mn}} - \bar{I_j})}{\sqrt{\left(\sum_m \sum_n (I_{i_{mn}} - \bar{I_i})^2\right)\left(\sum_m \sum_n (I_{j_{mn}} - \bar{I_j})^2\right)}} \tag{8}$$

where m, n parameterize the pixels in the image, I is the mean of I . Image I_i and I_j are decided when the differences between I_i and I_j are sufficiently large to make it worthwhile to look for moving objects.

$$if \; R(I_i, I_j) \geq T_R \; then \; j = j + 1 \; else \; i = j \tag{9}$$

The pair I_i and I_j is selected as soon as $R(I_i, I_j)$ satisfies (9).

E. Locally Adaptive Thresholding

In this class of algorithms, a threshold is calculated at each pixel, which depends on some local statistics like range, variance, or surface-fitting parameters of the pixel neighborhood. In what follows, the threshold T(i, j) is indicated as a function of the coordinates (i,j) at each pixel, or if this is not possible, the object/background decisions are indicated by the logical variable B(i, j).

Adaptive Threshold Change using Local Variance and Local Contrast Scheme

The method adapts the threshold according to the local mean m(i, j) and standard deviation s(i, j) and calculated a window size of b×b. comparing the gray value of the pixel with the average of the gray values in some neighborhood (15×15 window suggested. If the pixel is significantly darker than the average, it is denoted as foreground; otherwise, it is classified as background. In the local method, the threshold is set at the midrange value, which is the mean of the minimum I_{low}(i, j) and maximum I_{high}(i, j) gray values in a local window of suggested size w=31. However, if the contrast C(i, j)= I_{high}(i, j) - I_{low}(i, j) is below a certain threshold then that neighborhood is said to consist only of one class, print or background, depending on the value of T(i, j).

IV. Experimental Results

In this section, the results of proposed algorithm for nighttime object detection and tracking are assessed. All the night scene videos are captured by standard CCD cameras (Panasonic WV-W860A), with a frame size of 320 × 240 pixels. The effectiveness of the proposed method is verified by the detection and tracking results.

(a) (b)

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

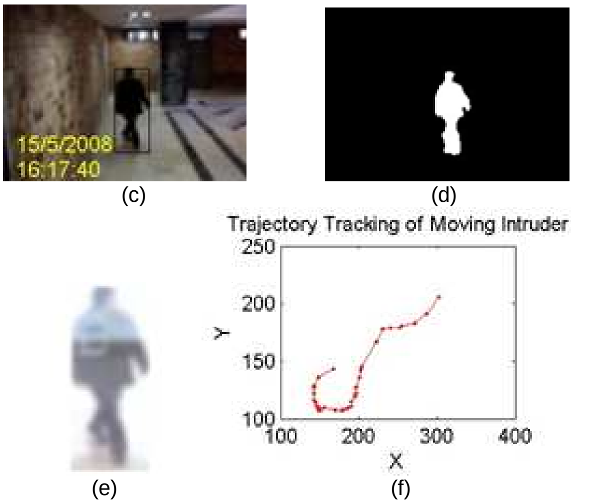

Fig.3 Tracking a single object in indoor environment with static background: (a) the scene with no moving object; and (b)–(d) an object enters the surveillance system and is tracked by system.

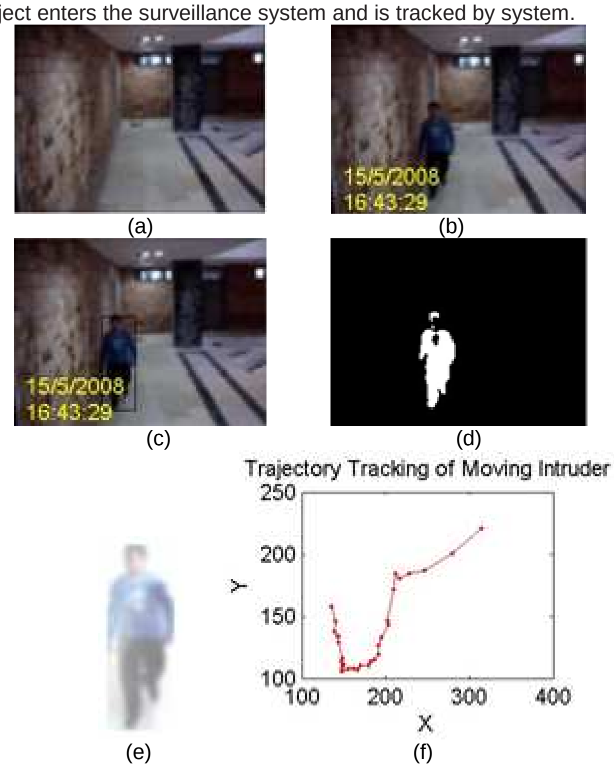

Fig.4 Tracking a single object in indoor environment with static background: (a) the scene with no moving object; and (b)–(d) an object enters the surveillance system and is tracked by the system.

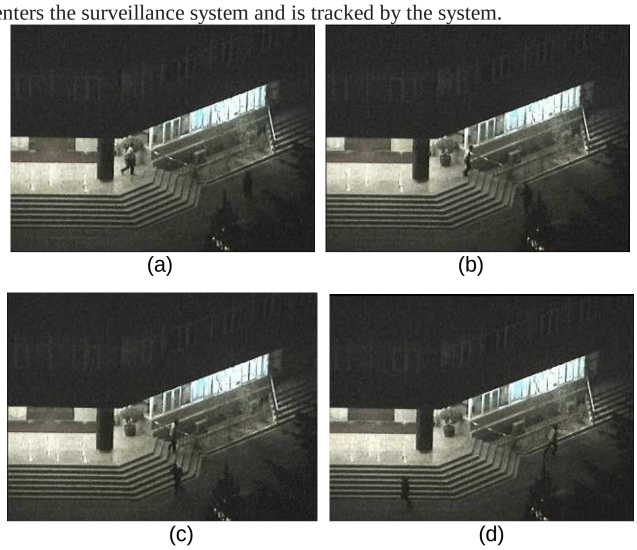

Fig.5 *"Two person at night"*visible objects found by local contrast computation

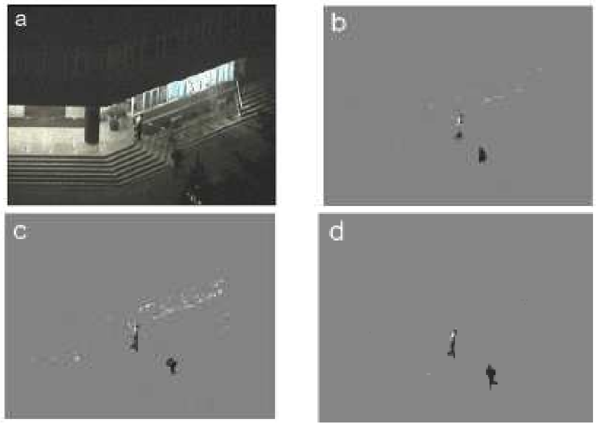

Fig.6 Comparisons of three object detection algorithms: (a) original image #110 in sequence "two persons at night", (b) result of MOG, (c) result of NPB, and (d) result of the proposed algorithm.

In order to test the proposed tracking method for multiple objects, we capture different video sequences indoors with multiple moving objects. The object detection algorithm is tested on the sequence "two person at night " from frame 1 to 120 (as shown in Fig. 5). Fig. 6 gives the detection results for frames 1–120. The moving objects and other visible content are detected in this step. It is clear that the moving objects are accurately located in all the frames.

Fig. 7 shows an example that the proposed method is effective to segment and track multiple people with occlusions. There are three almost complete occlusions happening in the video sequence. From Fig. 7, we can see that the three persons form a group and occlude one another. The proposed method effectively segments the three persons and correctly tracks all the three persons throughout the sequence.

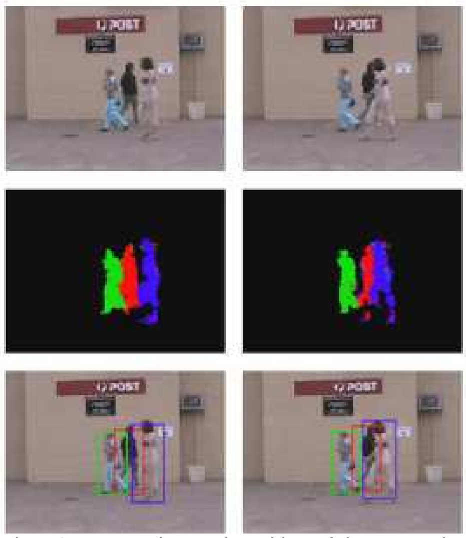

Fig.7 Segmentation and tracking of three people

A. *Comparison with other Methods for Object Detection*

The object detection algorithm is compared with two background-based object detection algorithms in Fig. 6. One of the algorithms uses an adaptive MOG model [19], the other uses nonparametric background subtraction (NPB) [20]. Fig. 6(a) is the original image. This image is low contrast and there are some distractors such as the light from the windows and the tree. Fig. 6(b) shows the detection results obtained from the MOG algorithm. The two objects are detected but not very well even though the parameters in the MOG algorithm are carefully chosen, especially the threshold T and standard deviation. Fig. 6(c) shows the detection result using

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

NPB. The results are better than those obtained from the MOG algorithm. Most of the pixels associated with the two objects are found. This algorithm is affected by high gray level gradients. For example, there are some false detections near the windows. Fig. 6(d) shows the detection results obtained using the proposed algorithm. It can be seen that the new algorithm performs better than MOG and NPB with more accurate detections and fewer false alarms.

The algorithm was tested in more cluttered night-time scenes. The camera is wide view, the objects have low contrast and a small apparent size. Most of the current algorithms, including NPB, fail to detect the objects.

Table 1 Comparison of the Normalized Jaccard coefficient between the new algorithm and the MOG-, NPB-based algorithms on different sequences of frames.

Video sample no.	MOG	NPB	The proposed method
"Two persons at night"	78.65%	85%	97.41%
"Traffic sequence at night"	55.71%	63.98%	96.87%
"Persons at a distance"	Less than 10%	Low than 10%	92.43%
"Zooming camera sequence"	Less than 10%	Low than 10%	95.23%

B. Evaluation of the Experiments

Comparisons are carried out between the proposed algorithm and both the MOG- and NBP-based algorithms. The performances of the three algorithms are measured using the Jaccard coefficient [21, 22], J defined by

$$J = TP/(TP + FP + FN)\% \qquad (10)$$

where TP (true positives) is the number of moving objects correctly detected, FP (false positives) is the number of false detections of moving objects and FN (false negatives) is the number of missed detections of moving objects. The ground truth is obtained manually in each frame of the sequence. Table 1 gives the values of normalized Jaccard (NJ) coefficient in Table 1, J is used to describe the detection accuracy of single frame and NJ gives the detection accuracy for the whole sequence. For the sequence “two persons at night”, the MOG and NBP algorithms both succeed to some extent. The NPB algorithm detects low contrast objects better than the MOG algorithm but it has too many false alarms. The proposed algorithm achieves the best (97.41%). For the sequence “traffic sequence at night”, the MOG and NBP algorithms do not work well because of the light reflections. The new algorithm gives satisfactory results (96.87%). For the two sequences “persons at a distance” and “zooming camera sequence”, it is very difficult to detect objects using the MOG and NBP algorithms because of the low contrast and the zooming camera. The NJ score is so low that it is ignored in Table 1. The proposed algorithm works relatively well on the two sequences (92.43% and 95.23%).

V. CONCLUSION

The paper introduced a new scheme for detecting and tracking of moving objects for night surveillance systems. The scheme is based on integrating both the wavelet-based contrast change detector and locally adaptive thresholding. Object detection is based on local contrast changes and detection results are improved by tracking the detected objects from one frame to the next. After processed by discrete wavelet transform, noise arises from fake motion will be decomposed into high frequency sub-band, so we can track moving objects in environment with varied background. Furthermore, the low-resolution level image bears analogy to the original image, therefore we can analyze the low frequency sub-image in the low-resolution level directly. This will highly reduce the computing cost. The proposed method has been tested and validated by a significant number of experiments. The proposed model has proved to be robust in various environments. Experiments demonstrate that our algorithm has the ability to detect and track objects robustly at night under conditions in which more conventional algorithms fail. There are several parameters and thresholds in the new algorithm. Parameters are adjusted adaptively, for example, the threshold to determine significant inter-frame differences, the size of rectangular region for contrast measure, the threshold on contrast measure and the threshold on the differences between contrast scores. The comparisons show that the proposed model achieves very promising results.

REFERENCES

[1] S. J. McKenna et al., *Tracking groups of people*, Computer Vision Image Understanding, 80 (2000) 42–56.
[2] W. Hu et al., “A survey on visual surveillance of object motion and behaviors,” IEEE Trans. Syst. Man Cybernetics–Part C 34 (3) (2004) 334–352.
[3] F. H. Cheng, Y. L. Chen, “Real time multiple objects tracking and identification based on discrete wavelet transform,” Pattern Recognition 39 (2006) 1126 – 1139.
[4] K. Huanga, L. Wanga, T. Tana, S. Maybankb, “A real-time object detecting and tracking system for outdoor night surveillance,” Pattern Recognition 41 (2008) 432 – 444.
[5] W. Wu, T. Yu, “A field model for human detection and tracking,” IEEE Transactions on pattern analysis and machine intelligence, Vol.28, No.5, May 2006.
[6] I. Daubechies, *Orthonormal bases of compactly supported wavelets*, Comm. Pure Appl. Math. (41) (1988) 909–996.
[7] M. T. Hanna, S. A. Mansoori, “The discrete time wavelet transform: its discrete time fourier transform and filter bank implementation,” IEEE Trans. Circuits Systems II: Analog Digital Signal Process. 48 (2) (2001) 180–183.
[8] A. S. Lewis, G. Knowles, “Image compression using the 2-D wavelet transforms,” IEEE Trans. Image Process. 1 (2) (1992) 244–250.
[9] F. Marino, “Two fast architectures for the direct 2-D discrete wavelet transform,” IEEE Trans. Signal Process. 49 (2001) 1248–1259.
[10] R. C. Gonzalez, R. E. Woods, *Digital Image Processing*, Addison Wesley, Reading, MA, 1992.
[11] C. Taskiran, J. Y. Chen, A. Albiol, L. Torres, C. A. Bouman, E. J. Delp, “VIBE: A Compressed Video Database Structured for Active Browsing and Search,” IEEE Trans. On Multimedia, Vol. 6, Issue 1, Feb. 2004, pp. 103-118.
[12] E. Peli, *Contrast sensitivity function and image discrimination*, J. Opt. Soc. Am. A 18 (2) (2001) 283–293.
[13] R. M. Haralick, L. G. Shapiro, *Computer and Robot Vision*, vol. 1, Addison-Wesley, Reading, MA, 1992 pp. 28–33.
[14] P. G. Barten, *Contrast sensitivity of the Human Eye and its Effects on Image Quality*, SPIE, Bellingham, Washington, 1999.
[15] E. Reinhard, P. Shirley, M. Ashikhmin, T. Troscianko, “Second order image statistics for computer graphics,” ACM Symposium on Applied Perception in Computer Graphics and Visualization, August 2004.
[16] Y. F. Ma, H. J. Zhang, “Detection motion object by spatial-temporal entropy,” ICME, 2001, pp. 381–384.

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

[17] K. Huang, Z. Y. Wu, Q. Wang, “Image enhancement based on the statistics of visual representation,” Image Vision Comput. 23 (2005) 51–57.

[18] A. Elgammal, L. S. Davis, “Probabilistic framework for segmenting people under occlusion,” Proceedings of the IEEE Eighth International Conference on Computer Vision, 2001, pp. 145–152.

[19] C. Stauffer, W.E.L. Grimson, “Adaptive background mixture models for real-time tracking,” CVPR, 1999, pp. 252–260.

[20] S. J. McKenna,Y. Raja, S. Gong, “Tracking color objects using adaptive mixture models,” Image Vision Comput. 17 (3) (1999) 225–231.

[21] L. Wang, W. Hu, T. Tan, “Recent developments in human motion analysis,” Pattern Recognition 36 (3) (2003) 585–601.

[22] P. Sneath, R. Sokal, *The Principle and Practice of Numerical Classification*, W.H. Freeman, New York, 1973.

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)