International Journal of Innovative
Computing, Information and Control
Volume 8, Number 7(A), July 2012

ICIC International ©2012 ISSN 1349-4198
pp. 4451–4468

REAL-TIME MULTIPLE MOVING OBJECTS DETECTION AND TRACKING WITH DIRECT LL-MASK BAND SCHEME

Chih-Hsien Hsia[1] and Jen-Shiun Chiang[2]

[1]Department of Electrical Engineering
National Taiwan University of Science and Technology
No. 43, Section 4, Keelung Road, Taipei 106, Taiwan
chhsia@ee.tku.edu.tw

[2]Department of Electrical Engineering
Tamkang University
No. 151, Yingzhuan Road, Tamsui District, New Taipei City 25137, Taiwan
chiang@ee.tku.edu.tw

Received March 2011; revised July 2011

Abstract. *This paper presents a new approach, direct LL-mask band scheme (DLLBS), for the detection and tracking of moving objects using a low resolution image. Moving object detection is an important basic task for intelligent video surveillance systems, because it provides a focus of attention for post-processing. However, the successful detection of moving objects in a real environment is a difficult task, due to noise cause by fake motion, such as the motion of leaves in trees. Many methods have been developed in constrained environments, for the detection and tracking of moving objects. The DLLBS method can effectively reduce this noise, with low computing cost, in both indoor and outdoor environments. For circumstances where occlusions occur, we propose a new approach, characteristic point recognition (CPR). Together with DLLBS and CPR, the problems associated with occlusions are alleviated. The experimental results indicate that the proposed DLLBS method (for* 320×240 *and* 640×480 *image frames) can provide higher precision in the detection and tracking of moving objects, and multiple moving objects where occlusion is a problem, for real-time intelligent video surveillance applications.*
Keywords: Direct LL-mask band scheme, Moving object detection and tracking, Low resolution image, Occlusion, Characteristic point recognition

1. **Introduction.** In recent years, video surveillance systems for the purpose of security have been developed rapidly. More and more researches try to develop intelligent video surveillance systems to replace the traditional passive video surveillance systems [1,2]. The intelligent video surveillance system can detect moving objects in the initial stage and subsequently process the functions such as object classification, object tracking, and object behaviors description. Detecting moving object is a very important aspect of computer vision and has a very wide range of surveillance applications. The accurate location of the moving object does not only provide a focus of attention for post-processing but also can reduce the redundant computation for the incorrect motion of the moving object. The successful moving object detection in a real surrounding environment is a difficult task, since there are many kinds of problems such as illumination changes, fake motion [3], night detection [4], and Gaussian noise in the background [5] that may lead to detect incorrect motion of the moving object. There are three typical approaches for motion detection [1,2,6]: background subtraction, temporal differencing, and optical flow. The background subtraction method detects moving regions between the current frame and the reference background frame. It provides the most complete motion mask data, but is susceptible

to dynamic scene changes due to lighting and extraneous events. Therefore, it has to update the reference background frame frequently. The temporal differencing approach extracts the moving region by using consecutive frames of the image sequences. It is suitable for dynamic environment, but often extracts incomplete relevant motion object pixels. The optical flow method uses characteristics of flow vectors of moving objects over time to detect moving regions. However, most optical flow methods are with higher complex computation. Generally, the above three moving object detection methods are all sensitive to illumination changes, noises, and fake motion such as moving leaves of trees.

In order to solve the mentioned problems, several approaches for object detecting and tracking were proposed [1-4,6-9,13-20,22,23]. Video tracking systems have to deal with variously shaped and sized input objects, which often result in a massive computing cost of the input of images. Cheng et al. [3] used discrete wavelet transform (DWT) to detect and track moving objects. The 2-D DWT can be used to decompose an image into four-subband images (LL, LH, HL and HH). It only processes the part of LL-band image due to the consideration of low computing cost and noise reduction issues. Although this method provides low computing cost (low resolution) for post-processing and noise reduction based on the conventional DWT, the LL-band image produced by the original image size via two dimensions (row and column) calculation may cause high computing cost in the pre-processing. Especially they use the three-level low-low band image (LL_3) that does not only bring a great image size transfer computation, but also the slow motion of the real moving objects may disappear. After dealing with the background subtraction, Alsaqre et al. [7] used a local pre-process method to smooth the image with reducing noise and other small fluctuations. However, this approach is unable to reduce the post-processing computation. Sugandi et al. [8,9] proposed a method for detecting and tracking objects by using a low resolution image with the 2×2 average filter (2×2 AF), which is generated by replacing each pixel value of the original image with the average value of its neighbors and itself. They mentioned that the low resolution image is insensitive to illumination changes and can reduce the small movement like moving leaves of trees in the background. Although this method can deal with small movement, these low resolution images become more blurred than the LL-band image generated by using DWT.

To overcome the above-mentioned problems, we propose a new approach, direct LL-mask band scheme (DLLBS), for detecting and tracking moving objects by using lifting-based wavelet coefficient DWT [9]. In DLLBS, we can select only the LL-mask band of modified lifting-based DWT. Unlike the conventional DWT method to process row and column dimensions separately by low-pass filter and down-sampling, the LL-mask band of modified lifting-based DWT can be used to directly calculate the LL-band image. Our proposed method can reduce the image transfer computing cost and remove fake motion that does not belong to the real moving object. For objects occlusion, a new approach, characteristic point recognition (CPR), was proposed. Combined with DLLBS and CPR, it can have accurate object tracking for various types of occlusions. Furthermore, it can retain a better slow motion of objects than that of the low resolution method [8,9] and provide effective and complete moving object regions.

2. Discrete Wavelet Transform and Low Resolution Technique.

Due to the imitation of video acquisition systems and transmission channels, images are often corrupted by noise. This degradation leads to a significant reduction in image quality, which makes more difficult the tasks that require acute computer vision, such as object tracking and recognition. Several methods have been proposed to remove noises or fake motion and

reduce computing cost. DWT [3] and the low resolution technique [8,9] are two important methods that are briefly described in the following sub-sections.

2.1. **Discrete wavelet transform method.** Wavelet transform [10] was proposed in the mid-1980s and it has been used in various fields, such as signal processing, image processing, computer vision, image compression, biochemistry and medicine, etc. It provides an extremely flexible, multi-resolution image, for image processing, and can decompose an original image into different subband images, including low and high frequencies, so it is possible choose the specific, relevant resolution data, or subband images [10-15].

A 2-D DWT of an image is illustrated in Figure 1(a). When the original image is decomposed into four-subband images, it must deal with row and column directions separately. Firstly, the high-pass filter, G, and the low-pass filter, H, are are used to analyze each row's data and are then down-sampled by 2, to produce the high and low frequency components of the row. The high- and the low-pass filters are applied again for each of the high- and low-frequency components of the column and are then down-sampled by 2. By way of this processing, the four-subband images, HH, HL, LH and LL, are generated. Each subband image has its own features. The low-frequency information is preserved in the LL-band and the high-frequency information is preserved in the HH-, HL-, and LH-bands. The LL-subband image can be further decomposed, in the same way, to produce a second level subband image. Using 2-D DWT, an image can be decomposed into any levels of subband images, as shown in Figure 1.

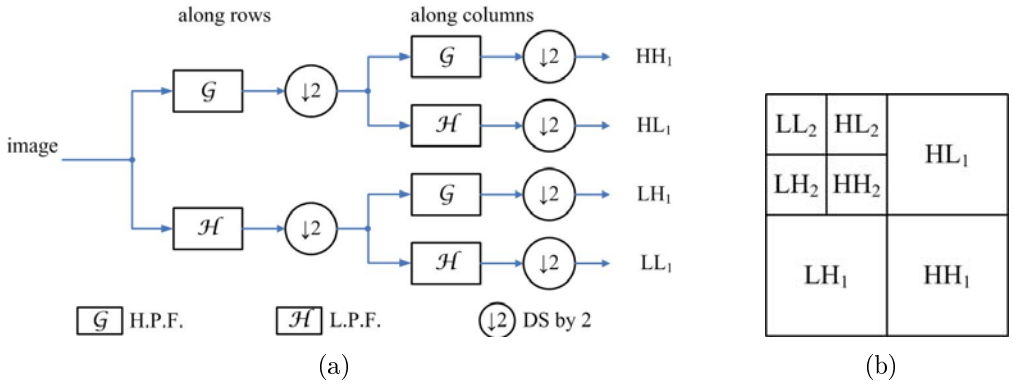


FIGURE 1. Diagrams of DWT image decomposition: (a) the 1-level 2-D analysis DWT image decomposition process, (b) the 2-level 2-D analysis DWT subband

Cheng et al. [3] used 2-D DWT for the detection and tracking of moving objects. Only the LL-band image was used for detecting the motion of a moving object. Because noises are restricted to the high frequency component, the computing cost for post-processing is reduced by using the LL-band image. This method copes with noise, or fake motion, effectively, however a conventional DWT scheme involves complicated calculations, when an original image is decomposed into the LL-band image. Moreover, the use of an LL-band image, to deal with the fake motion, can cause discontinuity of moving object detection regions.

2.2. **Low resolution method.** Sugandi et al. [8,9] proposed a simple method by using the low resolution concept to deal with the fake motion such as moving leaves of trees. The low resolution image is generated by replacing each pixel value of an original image with the average value of its four neighbor pixels and itself as shown in Figure 2. It also provides a flexible multi-resolution image like the DWT. Nevertheless, the low resolution

images generated by using the 2×2 average filter method are more blurred than that by using the DWT method, as shown in Figure 3. The average filtering is a low pass filter which denoises the image and performs restoration by the noise reduction spatial domain. It may reduce the preciseness of post-processing operation (such as occlusion and object identification), because the post-processing depends on the correct location of the moving object detecting and accuracy moving object data.

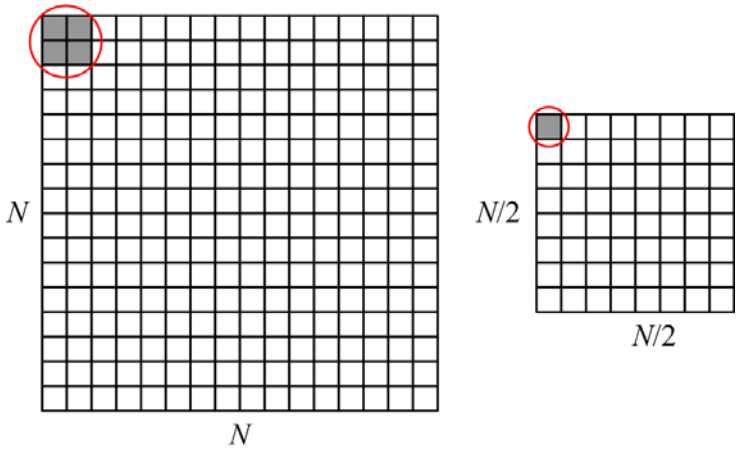


FIGURE 2. Diagram of the 2×2 average filter method

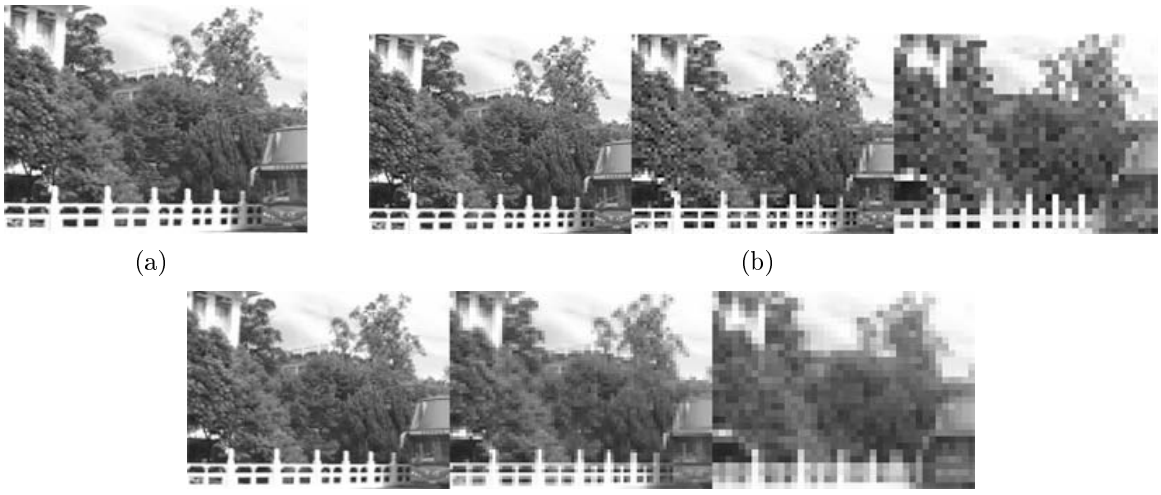


FIGURE 3. Comparisons of low resolution images: (a) the original image (320×240), (b) each subband image with LL-band DWT from left to right as 160×120, 80×60 and 40×30, respectively, (c) each resolution image with the 2×2 average filter method from left to right as 160×120, 80×60 and 40×30, respectively

3. **Direct LL-Mask Band Scheme.** In order to detect and track the moving object more accurately, we propose a new method called direct LL-mask band scheme (DLLBS) that is based on the 2-D lifting-based discrete wavelet transform (LDWT) [10]. It does not only retain the features of the flexibilities for multi-resolution, but also does not cause high computing cost when using it for finding different subband images. In addition, it

preserves more image quality of the low resolution image than that of the low resolution method [8,9].

3.1. **Direct LL-mask band.** In 2-D LDWT, the computation needs a large transpose memory requirement and has a long critical path. The modified lifting-based DWT (direct LL-mask band) has many advanced features such as short critical path, high speed operation, regular signal coding, and independent subband processing. The derivation coefficient of the 2-D direct LL-mask band is based on the 2-D 5/3 integer LDWT. For LL-band wavelet coefficient computation speed and simplicity considerations, this mask is 5×5, are used to perform spatial filtering tasks.

According to the 2-D 5/3 LDWT, the LL-band coefficients of the direct LL-mask band can be expressed as follows:

$$\begin{aligned} \mathrm{LL}(i,j) = {} & (9/16)x(2i,2j) + (1/64)\sum_{u=0}^{1}\sum_{v=0}^{1} x(2i-2+4u, 2j-2+4v) \\ & +(1/16)\sum_{u=0}^{1}\sum_{v=0}^{1} x(2i-1+2u, 2j-1+2v) \\ & +(-1/32)\sum_{u=0}^{1}\sum_{v=0}^{1} x(2i-1+2u, 2j-2+4v) \\ & +(-1/32)\sum_{u=0}^{1}\sum_{v=0}^{1} x(2i-2+4u, 2j-1+2v) \\ & +(3/16)\sum_{u=0}^{1}[x(2i-1+2u, 2j) + P(2i, 2j-1+2u)] \\ & +(-3/32)\sum_{u=0}^{1}[x(2i-2+4u, 2j) + x(2i, 2j-2+4u)] \end{aligned} \tag{1}$$

β	α	δ	α	β
α	γ	ε	γ	α
δ	α	ζ	α	δ
α	γ	ε	γ	α
β	α	δ	α	β

FIGURE 4. The subband mask coefficients of LL

The mask as shown in Figure 4 can be obtained via (1), where $\alpha = -1/32$, $\beta = 1/64$, $\gamma = 1/16$, $\delta = -3/32$, $\varepsilon = 3/16$ and $\zeta = 9/16$. The complexity of the modified lifting-based DWT is further reduced by employing the symmetric feature of the mask. First, the initial horizontal scan LL(0,0). The next coefficient can be calculated as LL(0,1), where the variable $\mathrm{XM}_{\mathrm{H+1}}$ denotes the repeated part after the first horizontal coefficient. The general form of the first horizontal step can be expressed as:

$$\begin{aligned} \mathrm{LL}(i,1) = {} & \beta \times x(i,j+2) + \delta \times x(i,j+4) + \alpha \times x(i,j+5) \\ & +\beta \times x(i,j+6) + \alpha \times x(i+1,j+2) + \varepsilon \times x(i+1,j+4) \\ & +\gamma \times x(i+1,j+5) + \alpha \times x(i+1,j+6) + \delta \times x(i+2,j+2) \\ & +\zeta \times x(i+2,j+4) + \varepsilon \times x(i+2,j+5) + \delta \times x(i+2,j+6) \\ & +\alpha \times x(i+3,j+2) + \varepsilon \times x(i+3,j+4) + \gamma \times x(i+3,j+5) \\ & +\alpha \times x(i+3,j+6) + \beta \times x(i+4,j+2) + \delta \times x(i+4,j+4) \\ & +\alpha \times x(i+4,j+5) + \beta \times x(i+4,j+6) + XM_{H+1}, \end{aligned} \tag{2}$$

where $i = 0 \sim N-1$, and

$$XM_{H+1} = \alpha \times x(i,3) + \gamma \times x(i+1,3) + \varepsilon \times x(i+2,3) + \gamma \times x(i+3,3) + \alpha \times x(i+4,3). \quad (3)$$

The next coefficient can be calculated as LL(0,2), where the variable $\mathrm{XM_{H+n}}$ denotes the repeated part after the second horizontal coefficient. From LL(0,2), the general form can be expressed as:

$$\begin{aligned} \mathrm{LL}(i,j+2) = {} & \delta \times x(i,2j+6) + \alpha \times x(i,2j+7) + \beta \times x(i,2j+8) \\ & + \varepsilon \times x(i+1,2j+6) + \gamma \times x(i+1,2j+7) + \alpha \times x(i+1,2j+8) \\ & + \zeta \times x(i+2,2j+6) + \varepsilon \times x(i+2,2j+7) + \delta \times x(i+2,2j+8) \\ & + \varepsilon \times x(i+3,2j+6) + \gamma \times x(i+3,2j+7) + \alpha \times x(i+3,2j+8) \\ & + \delta \times x(i+4,2j+6) + \alpha \times x(i+4,2j+7) + \beta \times x(i+4,2j+8) \\ & + XM_{H+n}, \end{aligned} \quad (4)$$

where $i = 0 \sim N-1$, $j = 0 \sim N-2$, and

$$\begin{aligned} XM_{H+n} = {} & \beta \times x(i,2j+4) + \alpha \times x(i,2j+5) + \alpha \times x(i+1,2j+4) \\ & + \gamma \times x(i+1,2j+5) + \delta \times x(i+2,2j+4) + \varepsilon \times x(i+2,2j+5) \\ & + \alpha \times x(i+3,2j+4) + \gamma \times x(i+3,2j+5) + \beta \times x(i+4,2j+4) \\ & + \alpha \times x(i+4,2j+5). \end{aligned} \quad (5)$$

The vertical scan can be done in the same way, where LL(0,0) is the same as that horizontal in LL(0,0). The next coefficient can be calculated as LL(1,0), where the variable $\mathrm{XM_{V+1}}$ denotes the repeated part after the vertical first coefficient. The general form of the first vertical step can be expressed as:

$$\begin{aligned} \mathrm{LL}(1,j) = {} & \beta \times x(2i,j) + \alpha \times x(2i,j+1) + \delta \times x(2i,j+2) + \alpha \times x(2i,j+3) \\ & + \beta \times x(2i,j+4) + \delta \times x(2i+4,j) + \varepsilon \times x(2i+4,j+1) \\ & + \zeta \times x(2i+4,j+2) + \varepsilon \times x(2i+4,j+3) + \delta \times x(2i+4,j+4) \\ & + \alpha \times x(2i+5,j) + \gamma \times x(2i+5,j+1) + \varepsilon \times x(2i+5,j+2) \\ & + \gamma \times x(2i+5,j+3) + \alpha \times x(2i+5,j+4) + \beta \times x(2i+6,j) \\ & + \alpha \times x(2i+6,j+1) + \delta \times x(2i+6,j+2) + \alpha \times x(2i+6,j+3) \\ & + \beta \times x(2i+6,j+4) + XM_{V+1}, \end{aligned} \quad (6)$$

where $i = 0$, $j = 0 \sim N-1$, and

$$XM_{V+1} = \alpha \times x(3,j) + \gamma \times x(3,j+1) + \varepsilon \times x(3,j+2) + \gamma \times x(3,j+3) + \alpha \times x(3,j+4). \quad (7)$$

Next, the second vertical scan is calculated by the method.

$$\begin{aligned} \mathrm{LL}(i+2,j) = {} & \delta \times x(2i+6,j) + \varepsilon \times x(2i+6,j+1) + \zeta \times x(2i+6,j+2) \\ & + \varepsilon \times x(2i+6,j+3) + \delta \times x(2i+6,j+4) + \varepsilon \times x(2i+7,j+2) \\ & + \gamma \times x(2i+7,j+1) + \varepsilon \times x(i,2j+7) + \gamma \times x(2i+7,j+3) \\ & + \alpha \times x(2i+7,j+4) + \beta \times x(i,2j+8) + \alpha \times x(2i+8,j+1) \\ & + \delta \times x(2i+8,j+2) + \alpha \times x(2i+8,j+3) + \beta \times x(2i+8,j+4) \\ & + XM_{V+n}, \end{aligned} \quad (8)$$

where $i = 0 \sim N-1$, $j = 0 \sim N-2$, and

$$\begin{aligned} XM_{V+n} = {} & \beta \times x(2i+4,j) + \alpha \times x(2i+4,j+1) + \delta \times x(2i+4,j+2) \\ & + \alpha \times x(2i+4,j+3) + \beta \times x(2i+4,j+4) + \beta \times x(2i+5,j) \\ & + \gamma \times x(2i+5,j+1) + \varepsilon \times x(2i+5,j+2) + \gamma \times x(2i+5,j+3) \\ & + \alpha \times x(2i+5,j+4). \end{aligned} \quad (9)$$

Finally, the diagonal oriented scan can be derived as:

$$\begin{aligned} \mathrm{LL}(1,1) = {} & \beta \times x(2,2) + \alpha \times x(2,5) + \beta \times x(2,6) + \zeta \times x(4,4) + \varepsilon \times x(4,5) \\ & + \alpha \times x(5,2) + \varepsilon \times x(5,4) + \gamma \times x(5,5) + \alpha \times x(5,6) + \beta \times x(6,2) \\ & + \delta \times x(6,4) + \alpha \times x(6,5) + \beta \times x(6,6) + XM_{D+1}, \end{aligned} \quad (10)$$

where the variable $\mathrm{XM_{D+1}}$ denotes the repeated part after the first diagonal scan.

Next the LL(2,2) is calculated as:

$$\begin{aligned}\mathrm{LL}(2,2) = {} & \varepsilon \times x(6,5) + \zeta \times x(6,6) + \varepsilon \times x(6,7) + \gamma \times x(7,5) \\ & + \varepsilon \times x(7,6) + \gamma \times x(7,7) + \alpha \times x(7,8) + \alpha \times x(8,5) \\ & + \delta \times x(8,6) + \alpha \times x(8,7) + \beta \times x(8,8) + XM_{D+n},\end{aligned} \tag{11}$$

where the variable $\mathrm{XM_{D+2}}$ denotes the repeated part after the first diagonal scan. The variable $\mathrm{XM_{D+1}}$ denotes the repeated part after the first diagonal scan. The general form of $\mathrm{XM_{D+n}}$ can be expressed as:

$$\begin{aligned}XM_{D+n} = {} & \beta \times x(2i+6, 2i+6) + \alpha \times x(2i+6, 2i+7) + \delta \times x(2i+6, 2i+8) \\ & + \alpha \times x(2i+6, 2i+9) + \beta \times x(2i+6, 2i+10) + \alpha \times x(2i+7, 2i+6) \\ & + \gamma \times x(2i+7, 2i+7) + \varepsilon \times x(2i+7, 2i+8) + \gamma \times x(2i+7, 2i+9) \\ & + \alpha \times x(2i+7, 2i+10) + \delta \times x(2i+8, 2i+6) + \varepsilon \times x(2i+8, 2i+7) \\ & + \delta \times x(2i+8, 2i+10) + \alpha \times x(2i+9, 2i+6) + \gamma \times x(2i+9, 2i+7) \\ & + \beta \times x(2i+10, 2i+6) + \alpha \times x(2i+10, 2i+7).\end{aligned} \tag{12}$$

The general form of the rest part can be expressed as:

$$\begin{aligned}\mathrm{LL}(i+1, j+1) = {} & \zeta \times x(2i+8, 2i+8) + \varepsilon \times x(2i+8, 2i+9) + \varepsilon \times x(2i+9, 2i+8) \\ & + \gamma \times x(2i+9, 2i+9) + \alpha \times x(2i+9, 2i+10) \\ & + \delta \times x(2i+10, 2i+8) + \alpha \times x(2i+10, 2i+9) \\ & + \beta \times x(2i+10, 2i+10) + XM_{D+n},\end{aligned} \tag{13}$$

where $i = 1 \sim N-1$, $j = 1 \sim N-1$.

The discussion above shows that the complexity of the proposed SMDWT can be significantly reduced by exploiting the symmetric feature of the masks, and the four-matrix frameworks, HH, HL, LH and LL can be individually employed for any specific applications. Table 1 shows the analysis results of the complexity between the conventional 2-D Lifting DWT [16] and the conventional 2-D Daubechies DWT [24] (the length of the filter is 4) and the 2-D SMDWT scheme for obtaining the LL-band images. Suppose an image is of size $N \times N$, the ratio ((proposed method/reference method) $\times$ 100%) between LDWT [16] and SMDWT at all levels are equivalent to 4.2%; the ratio between D4DWT [24] and SMDWT at all levels are 1.6%. The conventional 2-D Daubechies DWT scheme [24] requires to conduct a 1-D horizontal DWT to obtain low- and high-frequency subband images, and then gone through another 1-D vertical DWT for generating four-subband images. In [24], an image is transformed into four-subband images with the Daubechies coefficients DWT. The overall computational complexity (C) can be evaluated as below:

$$\mathrm{C} = 16 \times N^2 \times J \times (1 - 4^{-L})/3 \tag{14}$$

where J and L denote the length of the filter, and the number of level decompositions, respectively.

In [16], it transforms an image into four-subband images with the lifting coefficients DWT. The overall computational complexity can be evaluated as below:

$$\mathrm{C} = \sum_{i=1}^{L} \left(\frac{N^2}{2^{2L-1}} \times 6 + \frac{N^2}{2^{2L}} \times 12 \right) \tag{15}$$

On the other hand, the SMDWT directly transforms an original image into four-subband images using the four derived masks which does not need to process the row and column data separately. In addition, it simply calculates even pixels for every row

and column during the transforming process. The overall computational complexity of the SMDWT is as below.

$$\mathrm{C} = \sum_{i=1}^{L} \frac{N^2}{4^i} \tag{16}$$

TABLE 1. Complexity comparisons among various 2-D DWT approaches

Level(s)	①LL of LDWT [16]	②LL of D4DWT [24]	◎LL of SMDWT	Ratio(◎/①) ×100%	Ratio(◎/②) ×100%
level1	$6 \times N^2$	$16 \times N^2$	$1/4 \times N^2$	4.2%	1.6%
level1+level2	$15/2 \times N^2$	$20 \times N^2$	$5/16 \times N^2$	4.2%	1.6%
level1+level2+level3	$63/8 \times N^2$	$21 \times N^2$	$21/64 \times N^2$	4.2%	1.6%

3.2. **Detection and tracking flow.** The pre-processing flowchart for the proposed DLLBS moving object detection and tracking system is shown in Figure 5. Fristly, prior to color converting RGB data to YCbCr data, the double-change-detection method [16] is used to detect the moving objects. In order to decrease the holes left inside the moving entities, this system uses three continuous frames (F_{t-1}, F_t and F_{t+1}) for the detection of a moving object mask. These three continuous frames are decomposed into LL_2-band frames (LL_{2t-1}, LL_{2t} and LL_{2t+1}), using modified lifting-based DWT. After most of the noises and fake motions have been moved into the high-frequency subband, as shown in Figure 6, post-processing can proceed, using these three LL_2-band frames. Binary masks, B_{t-1} and B_t, can be obtained by computing the binary values of these three successive LL_2-band frames (between LL_{2t-1}, LL_{2t} and LL_{2t+1}) and the threshold value, T, in (14).

$$\begin{aligned} \mathrm{B}_{t-1}(i,j) &= \begin{cases} 1, & \text{if } |\mathrm{LL}_{2t-1}(i,j) - \mathrm{LL}_{2t}(i,j)| > T \\ 0, & \text{otherwise,} \end{cases} \\ \mathrm{B}_t(i,j) &= \begin{cases} 1, & \text{if } |\mathrm{LL}_{2t}(i,j) - \mathrm{LL}_{2t+1}(i,j)| > T \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{17}$$

The motion mask (MM_t) is generated using the union operation (logical OR) of B_{t-1} and B_t. The function is represented as follows:

$$\mathrm{MM}_t = \mathrm{B}_t \cup \mathrm{B}_{t-1}. \tag{18}$$

Holes may still exist in the motion masks, because some motion pixels are so tiny that they are incorrectly identified as non-motion pixels. In order to increase the robustness of the motion mask (MM_t), a morphological *closing* method [17] is used to fill these holes. Firstly, a dilation operator is used to fill the middle of the isolated pixels in the motion masks. It is defined as follows:

$$\mathrm{F}_t(i,j) = \begin{cases} 1, & \text{if one or more pixels of the adjacent pixels of motion mask} \\ & \mathrm{MM}_t(i,j) \text{ are 1,} \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

An erosion operator is then used to eliminate redundant pixels in the motion mask boundary, as follows:

$$\mathrm{MMR}_t(i,j) = \begin{cases} 0, & \text{if one or more pixels of the adjacent pixels of motion mask} \\ & \mathrm{F}_t(i,j) \text{ are 0,} \\ 1, & \text{otherwise.} \end{cases} \tag{20}$$

Eight neighbors of the motion mask MMR_t image are scanned, pixel-by-pixel, from top left to bottom right (raster scan). After extracting the connected component, it produces

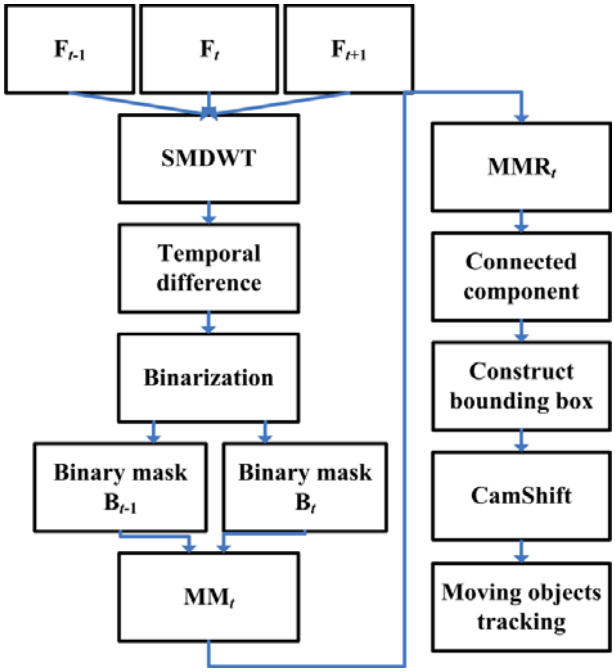
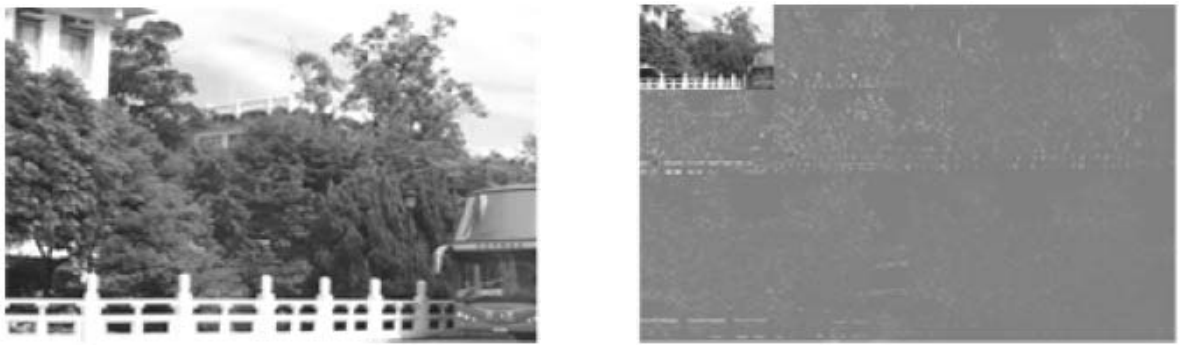


FIGURE 5. The pre-processing flowchart of the moving object detection and tracking based on DLLBS



(a) (b)

FIGURE 6. After most of the noises and fake motions are removed using direct LL-mask band (a) the original image, (b) LL_2-band image

several moving objects. This work uses a region-based tracking algorithm [3,18,19], to track the motion of the moving object.

When there is more than one moving object in the scene, connected component labeling is used to label each moving object and track each moving object, individually. The labeling of the components, based on pixel connectivity (intensity), [20] is performed by scanning an image and its groups, pixel by pixel from top left to bottom right. Connected pixel regions are identified by comparing the eight neighbors in each scan. If the pixel has at least one neighbor with the same label, it is labeled as a neighbor. In this way, each moving object is identified. The boundary of the moving object is defined by using a rectangular box to track the moving object. The bounding box is defined according to the object's motion mask. The bounding box uses the minimum and maximum values of row and column coordinates of the motion mask. In order to track moving objects in the original image, it is necessary to transform the coordinates from the LL_2 image size back

to the original image, according to the spatial relationship of the DWT, as follows:

$$O(i,j) = \mathrm{LL}_n(i \times 2^n, j \times 2^n). \tag{21}$$

where $n = 0 \sim l$ and l is the number of levels.

In block-matching motion estimation, the motion vector is the displacement of a block with the minimum distortion from the reference block. A CamShift block-matching algorithm determines the motion vector by identifying the block with the minimum distortion, using fast diamond-arc-hexagon search patterns, in the search area [21].

3.3. **Occlusion handling for multiple objects tracking.** In the post-processing, occlusion handling is a major problem in a video surveillance system. The most popular color space is the RGB color space [22]. If the multiple objects bounding boxes are occluded, the object bounding boxes are merged into the occlusion bounding box. Here we propose a new approach for occlusion in multiple objects tracking, called characteristic point recognition (CPR). Figure 8 shows the operation flowchart of CPR. CPR uses bounding boxes during pre-processing of DLLBS. For each tracked individual, the system will detect whether it makes occlusion with other object or group. It can obtain the RGB information from the video capture device directly to calculate the color information of the moving pixels. Owing to the information of moving pixels the size of the inter-frame difference image (1/16 of the original image) is with the central pixels.

To recognize every object, it uses the bounding box to find the characteristic point (CP). CP represents the central point of the bounding box as shown in the following equation:

$$\mathrm{Cs}^q[n] = \mathrm{B}_n\{(x_1, y_1), (x_2, y_2), \ldots, (x_q, y_q)\} \tag{22}$$

where $\mathrm{Cs}^q[n]$ is an array to store the CP of every object, n the label of the object, q the amount of CP, B_n the bounding box of every object, and (x, y) the color information indexed by the position of CP. Therefore CP expresses the feature of the object. We would like to focus on each object bounding box in order to select one CP or more.

At first, the CP of every object is stored in the buffer when the first frame is input, and is regarded as the initial sample. In latter frames, the CP is matched with the sample. In other words, the CP of 1 to n matches with the CP of the sample as shown in the following equation:

$$\mathrm{Cd}^q[n] = abs\left\{((\mathrm{Cs}^q[n] - \mathrm{Cm}^q[N])_{\mathrm{R}}, (\mathrm{Cs}^q[n] - \mathrm{Cm}^q[N])_{\mathrm{G}}, (\mathrm{Cs}^q[n] - \mathrm{Cm}^q[N])_{\mathrm{B}})_N\right\}, \tag{23}$$

where $\mathrm{Cm}^q[N]$ is a sample array to store the CP, N the label of the sample, and $\mathrm{Cd}^q[n]$ the absolute values obtained from the difference between $\mathrm{Cs}^q[n]$ and $\mathrm{Cm}^q[N]$.

After the match step, $\mathrm{Cd}^q[n]$ stores the sample N which is identical to the object n as shown in (24):

$$\mathrm{L}[n] = N, \tag{24}$$

where $\mathrm{L}[n]$ is the label N to label object n. Therefore, the object is recognized and labeled as sample N.

However, the objects of a frame may disappear or be occluded in latter frames. In order to hold the information of the object, the CP of the object has to be retained. Hence, we must know the object which has ever occurred when the object appears again in some frames. Because the CP may be changed by the environmental factors, the buffer has to be updated whenever a new frame is input in order to obtain the latest CP. If a new object appears, the CP of the new object should be added into the buffer to update the CP information. The CPR flowchart is shown in Figure 7.

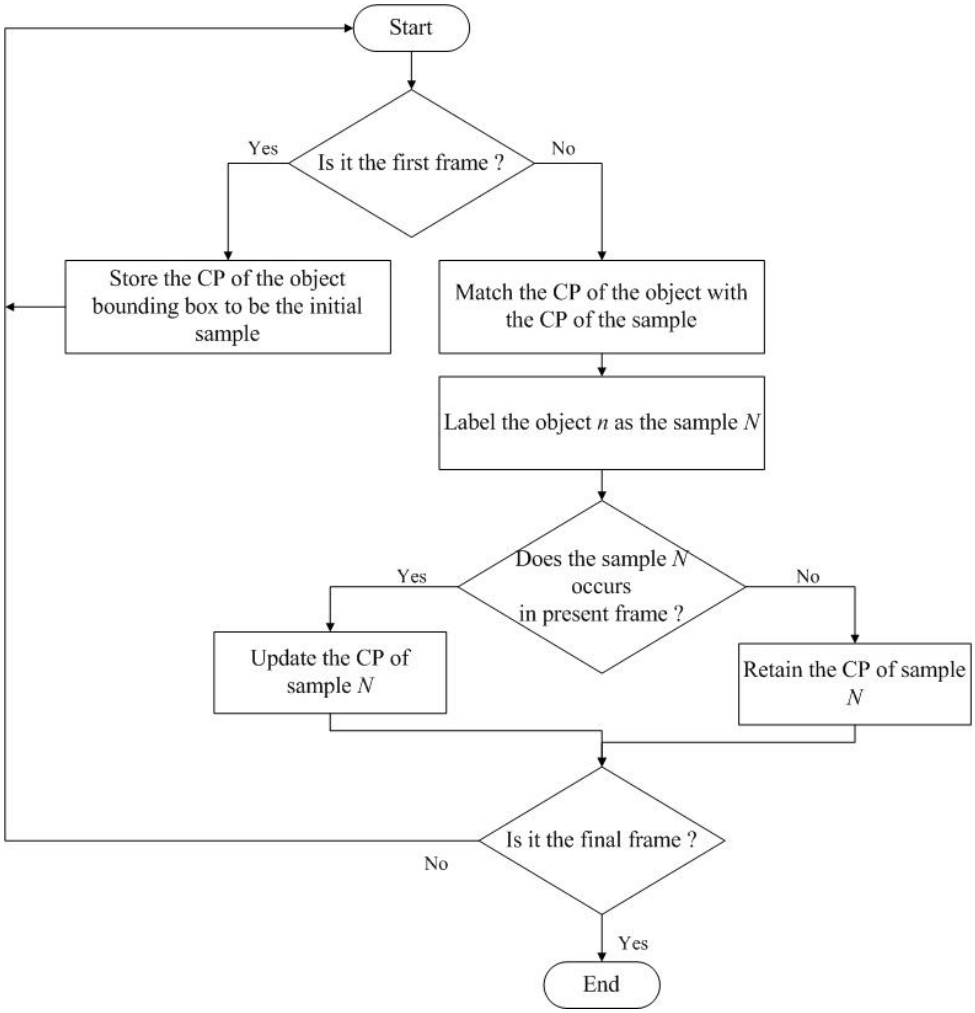
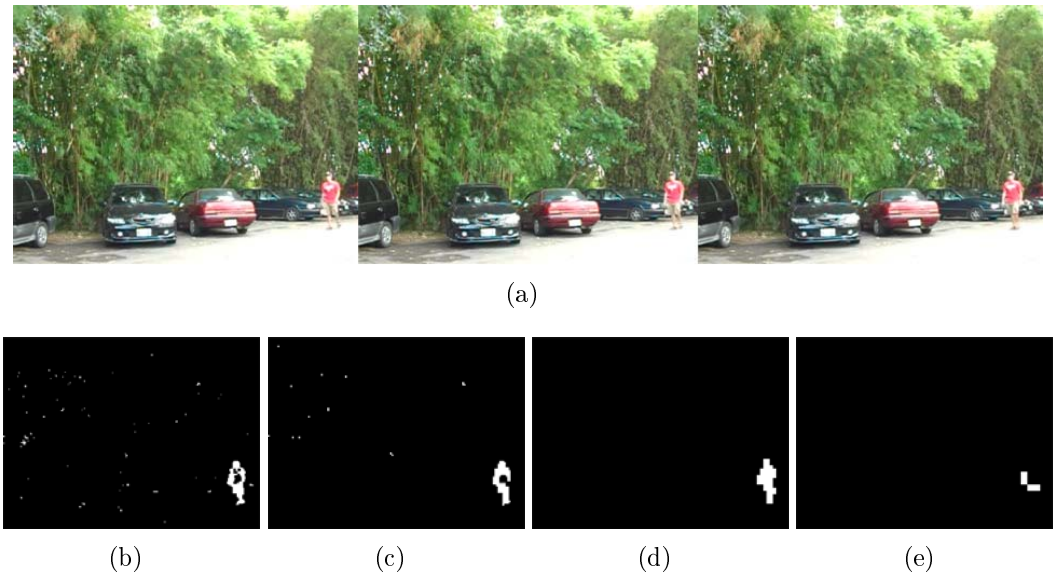


FIGURE 7. CPR flowchart

4. **Experimental Results.** In this work, the experimental results from several different environments, including indoor (day and night) and outdoor (day and night) environments, using a statistical video system, are demonstrated. The original image frame sizes were 320×240 and 640×480 and the format of color image frame is 24 bit, in an RGB system. All gray level frames were used to transfer the RGB system to the YCbCr system, for the detection of the motion of a moving object and the LL_2 (for 320×240) and the LL_3 (for 640×480) image size of 80×60, generated using SMDWT, from the original image, were used for the proposed moving object detection and tracking system. The experimental environment was established using an Intel 2.83 GHz Core 2 Quad CPU, 2 GB RAM, Microsoft Windows XP SP3 and a Borland C^{++} Builder (BCB) 6.0. BCB as a software development platform. The software includes verification of algorithms and image processing for the detection of moving objects.

4.1. **Dealing with noise issues.** Many difficulties were incorporated, such as fake motion and Gaussian noise, in the background. Different LL-band images, including one-level, two-level, three-level, and multi-level LL-band images, were used to deal with noise and their effectiveness was compared. It is suggested that an image that successfully eliminated noise has no other motion mask, besides moving object motion masks, as shown in Figures 8 and 9. Table 2 shows the average (Figures 8 and 9) success rate for the

elimination of noise, for each level of LL-band image. The first row is in the indoor environment and the second row in the outdoor environment. Each level of LL-band image is effective when dealing with indoor noises, like Gaussian noise, produced by random noise and statistical noise. However, when dealing with outdoor noise, such as moving leaves in trees, the LL_1-band image is not so effective, because these outdoor noises are sometimes too large to be eliminated completely.



(a)

(b) (c) (d) (e)

FIGURE 8. Moving object detection in an outdoor environment, with fake motion: (a) the original image, with three consecutive frames, (b) the temporal difference results for the original image, (c) the temporal difference results for the LL_1-band image, (d) the temporal difference results for the LL_2-band image, (e) the temporal difference results for the LL_3-band image

TABLE 2. The moving objects detection and tracking results

Resolution	[1]DLLBS	[2]LS	[3]2×2 AFS	[4]DSS
Level	Accuracy rate	Accuracy rate	Accuracy rate	Accuracy rate
LL_1 (160×120)	99.54 %	99.54 %	99.07 %	98.15 %
LL_2 (80×60)	99.07 %	99.07 %	93.07 %	81.94 %
LL_3 (40×30)	95.83 %	95.83 %	86.11 %	63.89 %

[1]DLLBS: direct LL-mask band scheme; [2]LS: lifting scheme; [3]2×2 AFS: 2×2 average filter scheme; [4]DSS: down-sampled scheme; [5]Accuracy rate: successful moving object tracking/original frames.

4.2. **Moving object tracking.** A moving object region is considered to be complete, if it is successful, as shown in Figure 10(a). In Figure 10(b), the moving object regions contain only a part of a moving object, so this is a failure to track. Figure 11(a) shows the original frame, without detecting, or tracking moving objects. Without DLLBS, many masks caused by noise are tracked and, even if the moving objects are tracked, those moving regions are fragmented, as shown in Figure 11(b). Using DLLBS, these noises can be filtered out, as shown in Figure 11(c). There is still incomplete generation of moving object regions, using the LL_1-band image, because the relevance of these pixels, in the LL_1-band image is ignored. When using a three-level resolution image to detect the

(a)

(b) (c) (d) (e)

FIGURE 9. Moving object detection in an indoor environment, with Gaussian noise: (a) the original image, with three consecutive frames, (b) the temporal difference results for the original image, (c) the temporal difference results for the LL_1-band image, (d) the temporal difference results for the LL_2-band image, (e) the temporal difference results for the LL_3-band image

moving objects, moving object regions are incompletely generated, because the LL_3-band image causes the disappearance of too many slow movements that apply to the moving object, as shown in Figure 11(e). Finally, the results of the LL_2-band image, in Figure 11(d), show that the two-level band image has a better tracking region and can also cope with noises and fake motion effectively, as shown in Table 2.

A 2×2 average filter scheme (AFS) was used for the substitution of the original DLLBS block system, to demonstrate the moving object, however this produceda a more blurred image than the DLLBS technique. The accuracy rates for object tracking with the 2×2 AFS are shown in Tables 2-4. The contrasts between Tables 2 and 3, for any resolution, are easily seen; the LL-band image generated by the DLLBS is more successful than the low-resolution image generated by the 2×2 AFS [8,9]. Several experiments were performed, in order to to prove the feasibility of the proposed approach for the detection and tracking of a moving object and for the case of obstruction. The performance of the detection method was tested by analyzing 16 video sequences, containing moving objects, with various backgrounds, and by simulating several conditions for moving objects, such as a single object in day time (indoor/outdoor), a single object at night (outdoor) and multiple objects in day time (outdoor) environments. All the test sequences were stored in the Microsoft AVI format, with a raw file of resolution 320×240 and 640×480, and a frame rate of 30 fps, as shown in Figure 12.

16 test sequences were established at our campus, in different environments, such as daytime, night-time, rainy day, fast movement, slow movement, and obstruction, as shown in Figure 12. Compared with other methods (2×2 AFS and DSS), DLLBS achieves good separation of spatially localized details, such as edges and singularities. Because such details are typically abundant in natural images and convey a significant part of the information embedded therein, DWT is used for image denoising. From Tables 3-5, it can be seen that some objects are not correctly identified, in the test frame of the sequences.

(a)

(b)

FIGURE 10. Examples of (a) successful tracking of a moving object and (b) failure to track a moving object

TABLE 3. Single moving object processing (without occlusion)

Sequence 1 ∼ 9	DLLBS		2 × 2 AFS [8,9]		DSS [23]	
	Accuracy rate	Detection + Tracking	Accuracy rate	Detection + Tracking	Accuracy rate	Detection + Tracking
Average	90.53 %	54.7 FPS	79.48 %	60.3 FPS	71.53 %	60.1 FPS

TABLE 4. Multiple moving objects processing (with occlusion)

Sequence 10 ∼ 13	DLLBS		2×2 AFS [8,9]		DSS [23]	
	Accuracy rate	Detection + Tracking + occlusion	Accuracy rate	Detection + Tracking + occlusion	Accuracy rate	Detection + Tracking + occlusion
Average	90.84 %	54.5 FPS	84.69 %	60.5 FPS	83.52 %	58.9 FPS

TABLE 5. Multiple moving objects processing (with occlusion)

Sequence 14 ∼ 16	DLLBS		2 × 2 AFS [8,9]		DSS [23]	
	Accuracy rate	Detection + Tracking + occlusion	Accuracy rate	Detection + Tracking + occlusion	Accuracy rate	Detection + Tracking + occlusion
Average	85.10 %	14.1 FPS	74.57 %	16.6 FPS	42.51 %	15.3 FPS

According to Tables 4 and 5, the proposed low-resolution method outperforms other methods [8,9,23] with respect to image size, accuracy rate, and number of frames per second. The low-resolution method provides a framework for the significant reduction of the spatial domain. The radar plot, in Figure 13, is used to visualize the characteristics of DLLBS. The three design criteria considered in Tables 4 and 5 are shown in the radar plot. The closer the point is to the center, the worse is the performance.

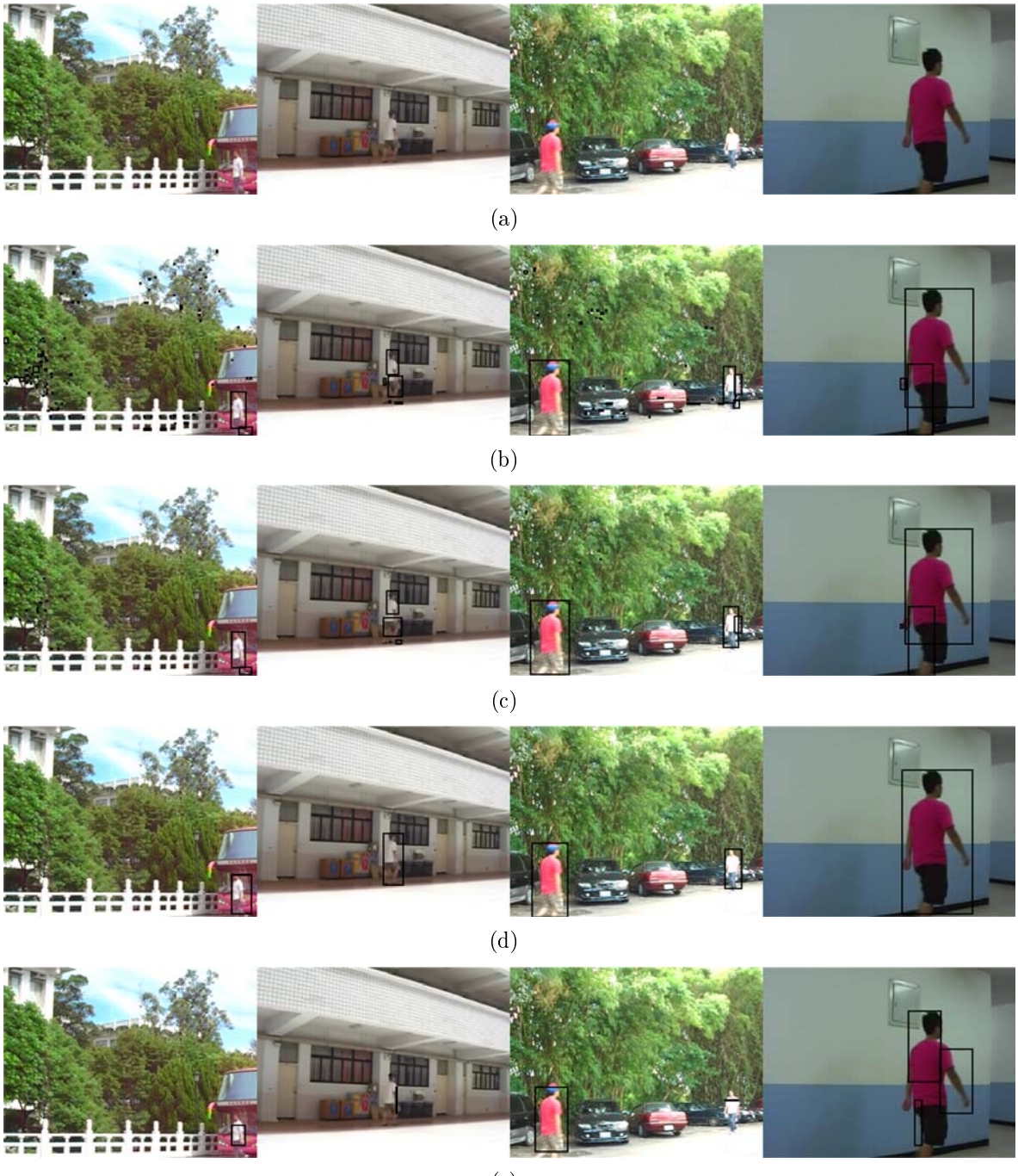


FIGURE 11. Results of tracking moving objects in various environments: (a) original frames without region-based object tracking, (b) original frames with region-based object tracking, (c) LL_1-band frames with region-based object tracking, (d) LL_2-band frames with region-based object tracking, (e) LL_3-band frames with region-based object tracking

5. **Conclusions.** This paper proposes DLLBS, for the detection and tracking of a moving object. The method is able to detect and track moving objects, using a low-resolution technique, in indoor and outdoor environments, with statistical video systems. The proposed DLLBS not only overcomes the drawbacks of highly complex computation and slow speed inherent to conventional DWT, but also preserves the wavelet features of the

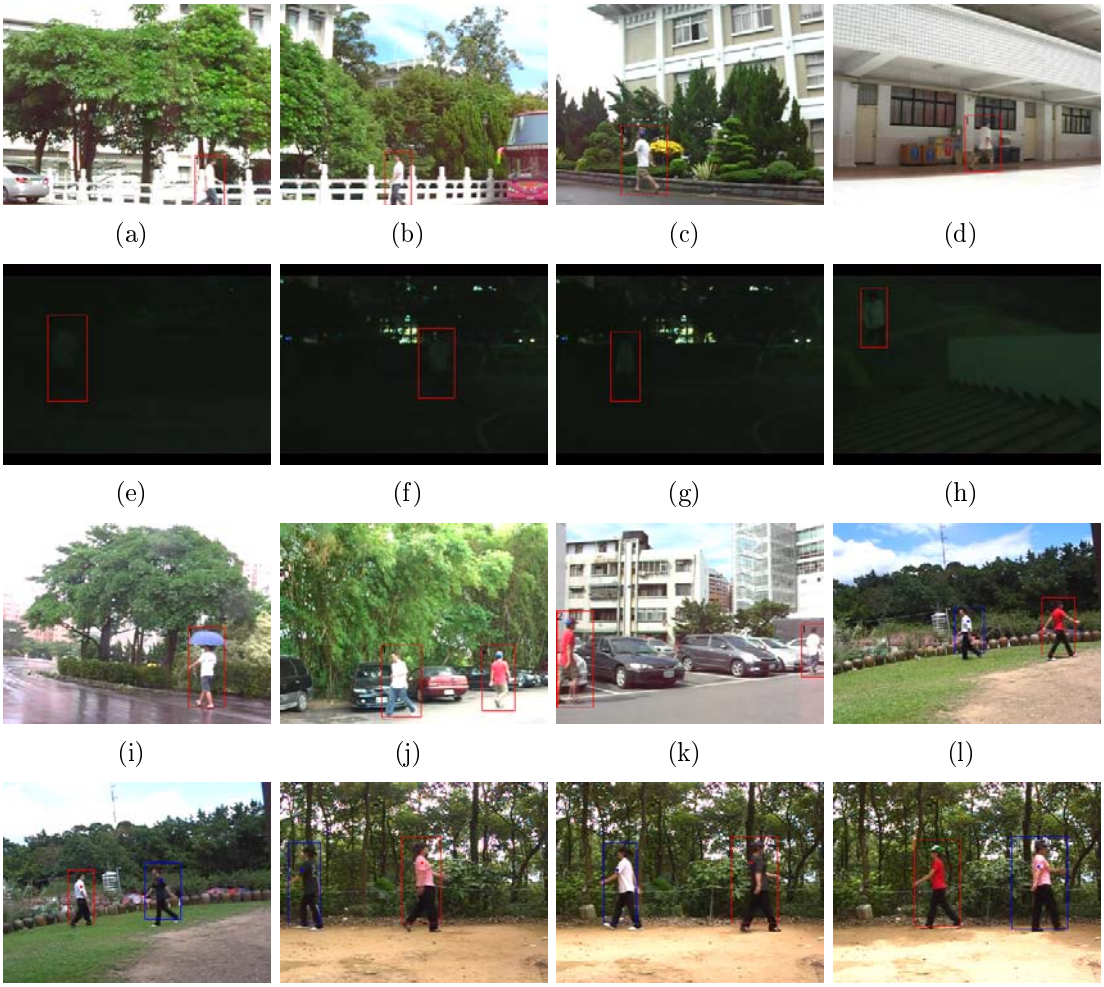


FIGURE 12. Test sequences: (a)-(p) are sequences 1-13 (320×240) and 14-16 (640×480), (a)-(c) show the single moving object in the outdoor, (d) single moving object in the indoor, (e) single moving object in the outdoor (fast movement to slow movement), (f) single moving object in the outdoor (fast movement), (g) single moving object in the outdoor (slow movement), (h) single moving object in the indoor (zoom-out to zoom-in), (i) single moving object in the outdoor (rainy day), (j) multiple moving object in the outdoor (occlusion), (k) multiple moving object in the outdoor, (l) multiple moving object in the outdoor (occlusion), (m) multiple moving object in the outdoor (occlusion), (n) multiple moving object in the outdoor (occlusion), (o) multiple moving object in the outdoor (occlusion), (p) multiple moving object in the outdoor (occlusion)

flexible multi-resolution image and is capapble of dealing with noises and fake motion, such as moving leaves in trees. In real-word applications, the experimental results demonstrate that the 2-D LL_2-band (for 320×240) and the 2-D LL_3-band (for 640×480) can effectively track moving objects in any environments (day and night), using region-based tracking, as well as coping with noise issues. In situations where obstruction occurs, the CPR technique was proposed. With a combination of DLLBS and CPR can accurately track various types of obstructed movement. This DLLBS is suitable for real-time video

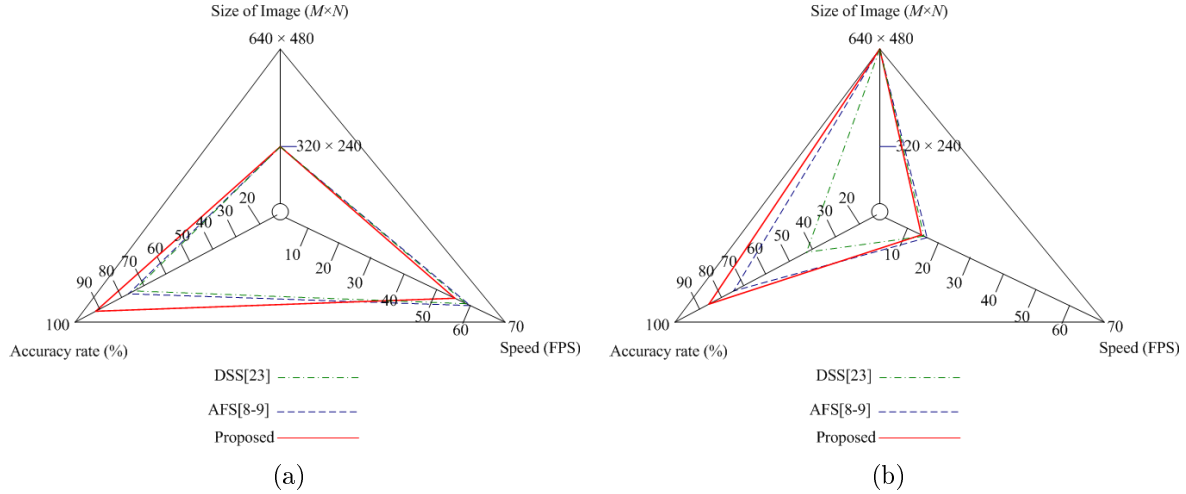


FIGURE 13. Rader plot showing a comparison with other methods: (a) for 320×240, (b) for 640×480

surveillance system applications, such as object classification and the descriptive behavior of objects.

Acknowledgement. This research work was partially supported by the National Science Council of Taiwan, under grant number NSC-99-2221-E-032-028.

REFERENCES

[1] W.-M. Hu, T.-N. Tan, L. Wang and S. Maybank, A survey on visual surveillance of object motion and behaviors, *IEEE Trans. on Systems, Man, and Cybernetics – Part C: Applications and Reviews*, vol.34, no.3, pp.334-352, 2004.

[2] N. Jacobs and R. Pless, Time scales in video surveillance, *IEEE Trans. on Circuits and Systems for Video Technology*, vol.18, no.8, pp.1106-1113, 2008.

[3] F.-H. Cheng and Y.-L. Chen, Real time multiple objects tracking and identification based on discrete wavelet transform, *Pattern Recognition*, vol.39, no.3, pp.1126-1139, 2006.

[4] K.-Q. Huang, L.-S. Wang, T.-I. Tan and S. Maybank, A real-time objects detecting and tracking system for outdoor night surveillance, *Pattern Recognition*, vol.41, no.1, pp.423-444, 2008.

[5] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley Longman Publish Co., Inc., Boston, MA, 2001.

[6] R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt and L. Wixson, A system for video surveillance and monitoring, *Technical Report, CMU-RI-TR-00-12*, 2000.

[7] F. E. Alsaqre and Y. Baozong, Multiple moving objects tracking for video surveillance system, *Proc. of IEEE International Conference on Signal Processing*, vol.2, pp.1301-1305, 2004.

[8] B. Sugandi, H. Kim, J. K. Tan and S. Ishikawa, Tracking of moving objects by using a low resolution image, *The 2nd International Conference on Innovative Computing, Information and Control*, Kumamoto, Japan, pp.408-408, 2007.

[9] B. Sugandi, H. Kim, J. K. Tan and S. Ishikawa, Real time tracking and identification of moving persons by using a camera in outdoor environment, *International Journal of Innovative Computing, Information and Control*, vol.5, no.5, pp.1179-1188, 2009.

[10] I. Daubechies and W. Sweldens, Factoring wavelet transforms into lifting steps, *The Journal of Fourier Analysis and Applications*, vol.4, no.3, pp.247-269, 1998.

[11] S. G. Mallat, A theory for multi-resolution signal decomposition: The wavelet representation, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.11, no.7, pp.674-693, 1989.

[12] W. Ge, L.-Q. Gao and Q. Sun, A method of multi-scale edge detection based on lifting scheme and fusion rule, *Proc. of International Conference on Wavelet Analysis and Pattern Recognition*, vol.2, pp.952-955, 2007.

[13] H.-H. Liu, X.-H. Chen, Y.-G. Chen and C.-S. Xie, Double change detection method for moving-object segmentation based on clustering, *IEEE International Symposium on Circuits and Systems*, pp.5027-5030, 2006.
[14] J. Ahmed, M. N. Jafri and J. Ahmad, Target tracking in an image sequence using wavelet features and neural network, *IEEE TENCON*, pp.1-6, 2005.
[15] F. A. Tab, G. Naghdy and A. Mertins, Multiresolution video object extraction fitted to scalable wavelet-based object coding, *IET Image Processing*, vol.1, no.1, pp.21-38, 2007.
[16] J.-C. Huang, T.-S. Su, L.-J. Wang and W.-S. Hsieh, Double-change-detection method for wavelet-based moving-object segmentation, *IET Electronics Letters*, vol.40, no.13, pp.798-799, 2004.
[17] C.-C. Hsieh and S.-S. Hsu, A simple and fast surveillance system for human tracking and behavior analysis, *Proc. of IEEE Conference on Signal-Image Technologies and Internet-Based System*, pp.812-828, 2007.
[18] S. J. Mckenna, Tracking groups of people, *Computer Vision and Image Understanding*, vol.80, no.1, pp.42-56, 2000.
[19] D.-T. Chen and J. Yang, Robust object tracking via online dynamic spatial bias appearance models, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.29, no.12, pp.2157-2169, 2007.
[20] L. He, Y. Chao, K. Suzuki and K. Wu, Fast connected-component labeling, *Pattern Recognition*, vol.42, no.9, pp.1977-1987, 2009.
[21] J.-S. Chiang, H.-T. Lin and C.-H. Hsia, Novel fast block motion estimation using diamond-arc-hexagon search patterns, *Journal of the Chinese Institute of Engineers*, vol.31, no.6, pp.955-966, 2008.
[22] W.-M. Hu, X. Zhou, M. Hu and S. Maybank, Occlusion reasoning for tracking multiple people, *IEEE Trans. on Circuits and Systems for Video Technology*, vol.19, no.1, pp.114-121, 2009.
[23] S. Cvetkovic, P. Bakker, J. Schirris and P. H. N. de With, Background estimation and adaptation model with light-change removal for heavily down-sampled video surveillance signals, *Proc. of IEEE International Conference on Image Processing*, pp.1829-1832, 2006.
[24] G. K. Kharate, V. H. Patil and N. L. Bhale, Selection of mother wavelet for image compression on basis of nature of image, *Journal of Multimedia*, vol.2, no.6, pp.44-51, 2007.