Illumination-Sensitive Background Modeling Approach for Accurate Moving Object Detection

Fan-Chieh Cheng, Shih-Chia Huang, and Shanq-Jang Ruan, *Member, IEEE*

***Abstract*—Background subtraction involves generating the background model from the video sequence to detect the foreground and object for many computer vision applications, including traffic security, human-machine interaction, object recognition, and so on. In general, many background subtraction approaches cannot update the current status of the background image in scenes with sudden illumination change. This is especially true in regard to motion detection when light is suddenly switched on or off. This paper proposes an illumination-sensitive background modeling approach to analyze the illumination change and detect moving objects. For the sudden illumination change, an illumination evaluation is used to determine two background candidates, including a light background image and a dark background image. Based on the background model and illumination evaluation, the binary mask of moving objects can be generated by the proposed thresholding function. Experimental results demonstrate the effectiveness of the proposed approach in providing a promising detection outcome and low computational cost.**

***Index Terms*—Background model, object detection, sudden illumination change, video surveillance.**

I. Introduction

VIDEO surveillance systems have significant implications in the defense against criminality and terrorist threats in both public and private sectors. In the design of an advanced video surveillance system, motion detection is usually a critical process responsible for determining the locations of moving objects. Furthermore, motion detection has been used for many computer vision applications, including driver assistance [1], human-machine interaction [2], face detection [3], collision prediction of pedestrians [4], remote image processing [5], detection of foreign bodies in food [6], event recognition of human actions [7], and so on.

According to a survey of related literature [8], motion detection approaches can be broadly classified as three categories: temporal difference [9], [10], optical flow [11], [12], and background subtraction [13], [14]. Temporal difference approaches detect moving objects by calculating the differences between pixels in consecutive frames of a video sequence. Although temporal difference approaches are adaptive to environments with sudden illumination change, some relevant pixels cannot be extracted; this results in holes inside moving entities. Optical flow methods usually use characteristics of flow-vectors over time to indicate moving regions in a video sequence. However, flow-vectors of moving objects only can be applied to illustrate streams of moving objects, thus detecting a sparse form of object regions. Out of these three categories, background subtraction received the most attention due to its computationally affordable implementation and its accurate detection of moving entities.

In general, moving objects can be detected by the background subtraction approaches [13], [14] which consist of the background model and the object thresholding function. However, the generated background model may not be applicable in some scenes with sudden illumination change. As a result, it is highly dependent on a good background model to reduce the influence of these changes. Therefore, we propose an illumination-sensitive background modeling approach based on an illumination evaluation in order to generate an accurate background model for motion detection. The remainder of this paper is organized as follows: Section II describes some representational motion detection approaches that are implemented in our experiments. Section III presents our approach in detail. In Section IV, we present the experimental results and discussions. Finally, our conclusion is provided in Section V.

II. Related Work

In this section, we describe three representative motion detection approaches, including a multiple temporal difference approach (MTD) [10], a simple statistical difference approach (SSD), and a multiple $\Sigma-\Delta$ estimation approach (MSDE) [14]. Note that each approach is described by assuming that $I_t(x, y)$ refers to an original video frame with size $M \times N$, where t represents the frame index and (x, y) represents the coordinate of each pixel.

A. MTD Approach

The MTD approach [10] doesn't maintain a background model but holds several previous reference frames to reduce holes inside moving entities for motion detection. According to the literature [10], there are seven previous reference frames be used to calculate the difference image. The previous frames can be held by the following function:

$$\text{if } count = 5n, \text{ then} \begin{cases} B_n = I_{t-1}, \\ n = n+1, \end{cases} \tag{1}$$

where B_n represents each reference frame, n represents the index of reference frame, and $count$ represents the short term counter which should be increased progressively $(count++)$ at each frame. Note that both n and $count$ are initialized to zero. Furthermore, n and $count$ should be reset to zero when n is

Manuscript received January 26, 2011; revised April 14, 2011; accepted June 07, 2011. Date of publication July 25, 2011; date of current version November 23, 2011. This work was supported by the National Science Council under the Grant NSC 99-2221-E-027-108 and NSC 99-2221-E-011-145.

F.-C. Cheng and S.-J. Ruan are with the Department of Electronic Engineering, National Taiwan University of Science and Technology (e-mail: d9802108@mail.ntust.edu.tw; sjruan@mail.ntust.edu.tw).

S.-C. Huang is with the Department of Electronic Engineering, National Taipei University of Technology (e-mail: schuang@ntut.edu.tw).

Digital Object Identifier 10.1109/TBC.2011.2160106

greater than $n_{\max}$, which is empirically set to six according to the literature [10].

After generating the reference frames, the absolute difference image $\Delta_t(x,y)$ can be calculated as follows:

$$\Delta_t(x,y) = \sum_{n=0}^{n_{\max}} \left| I_t(x,y) - B_n(x,y) \right|. \tag{2}$$

The binary mask of moving objects is calculated as follows:

$$D_t(x,y) = \begin{cases} 1, & \text{if } \Delta_t(x,y) > 250, \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

where $D_t(x,y)$ equals 1 to represent a motion pixel and equals 0 to represent a background pixel.

B. SSD Approach

The SSD approach consists of a simple background model by temporal average and an object thresholding function by temporal standard deviation. For each pixel, the temporal average $\mu(x,y)$ and temporal standard deviation $\sigma(x,y)$ are calculated as follows:

$$\mu(x,y) = \left((t-1)\mu(x,y) + I_t(x,y)\right)/t, \tag{4}$$

$$\sigma(x,y) = \sqrt{\frac{(t-1)\sigma(x,y)^2 + \left(I_t(x,y) - \mu(x,y)\right)^2}{t}}. \tag{5}$$

Note that both $\mu(x,y)$ and $\sigma(x,y)$ are initialized to zero. Based on the simple background model by equation (4), the absolute difference image $\Delta_t(x,y)$ can be calculated as follows:

$$\Delta(x,y) = |I_t(x,y) - \mu(x,y)|. \tag{6}$$

The binary mask of moving objects is calculated as follows:

$$D_t(x,y) = \begin{cases} 1, & \text{if } \Delta_t(x,y) > \lambda\sigma(x,y), \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

where λ represents the pre-defined parameter. According to the literature [13], λ is empirically set to 3 yields much better results.

C. MSDE Approach

The MSDE approach consists of a hybrid background model and an object thresholding function, which are both formulated by traditional sign function. Before calculating the hybrid background model, the different reference images must be generated at each frame. When the frame index t is a multiple of α_i, the different reference images can be formulated as follows:

$$b_t^i(x,y) = b_{t-1}^i(x,y) + \operatorname{sgn}\left(b_t^{i-1}(x,y) - b_{t-1}^i(x,y)\right), \tag{8}$$

where $b_t^i(x,y)$ represents the current frame of i-th reference image, $b_{t-1}^i(x,y)$ represents the previous frame of i-th reference image, and $b_t^{i-1}(x,y)$ represents the current frame of $(i-1)$-th reference image. Note that $b_t^0(x,y)$ is regarded as original video frame $I_t(x,y)$. Additionally, weighted variances can be calculated as

$$v_t^i(x,y) = v_{t-1}^i(x,y) + \operatorname{sgn}\left(N\Delta_t^i(x,y) - v_{t-1}^i(x,y)\right), \tag{9}$$

where $\Delta_t^i(x,y)$ represents the absolute difference between $I_t(x,y)$ and $b_t^i(x,y)$, $v_t^i(x,y)$ represents the current frame of i-th weighted variance, $v_{t-1}^i(x,y)$ represents the previous frame of i-th weighted variance, and N represents the pre-defined parameter which is empirically set to two [14].

Based on the use of reference images and weighted variances, the hybrid background model $B_t(x,y)$ can be calculated as

$$B_t(x,y) = \frac{\sum_{i=1}^{R} \alpha_i b_t^i(x,y)/v_t^i(x,y)}{\sum_{i=1}^{R} \alpha_i/v_t^i(x,y)}. \tag{10}$$

According to the literature [14], R is experimentally set to three and confidence values α_1 to α_3 are set to 1, 8, and 16, respectively. Finally, the binary mask of moving objects $D_t(x,y)$ can be generated by the same thresholding function of SDE approach that uses the first weighted variance for binarization.

D. Discussion

Since the MTD approach uses several temporal reference images to detect moving objects and adapt to sudden illumination change, holes are reduced inside moving entities of the binary mask. However, the detected objects may drag ghost artifacts due to use of several consecutive frames possibly involving moving objects.

Compared with the MTD approach, the SSD and MSDE approaches can generate the background model to reduce ghost artifacts. However, these approaches sometimes fail to detect moving objects because the luminance of the generated background model cannot be easily updated in environments with sudden illumination change.

To compensate for the limitations of these approaches, an approach must be developed to create an illumination-sensitive background model for object detection. To achieve this objective, an object detection approach should be proposed using the global video trend along with the fast thresholding function.

III. Proposed Method

In order to solve the problem as mentioned above, we propose a novel background subtraction approach, which includes a background model module, an illumination evaluation module, and an object detection module. The flowchart of the proposed background subtraction approach is shown in Fig. 1.

1) Background model: In general, the background model can be generated by temporal average. However, temporal average cannot adapt to the current luminance of the incoming video frame $I_t(x,y)$. To solve this problem, the background model can be generated by the running average expressed as follows:

$$B_t(x,y) = B_{t-1}(x,y) + \alpha\left(I_t(x,y) - B_{t-1}(x,y)\right), \tag{11}$$

where $B_t(x,y)$ is the current background model, $B_{t-1}(x,y)$ is the previous background model, $I_t(x,y)$ is the current video frame, and α represents the adaptive parameter. For the digital video sequence, the running average can be simplified by a constant calculation as follows:

$$B_t(x,y) = \begin{cases} B_{t-1}(x,y) + 1, & \text{if } B_{t-1}(x,y) < I_t(x,y), \\ B_{t-1}(x,y) - 1, & \text{if } B_{t-1}(x,y) > I_t(x,y). \end{cases} \tag{12}$$

Note that B_0 is set to I_0 for initialization.

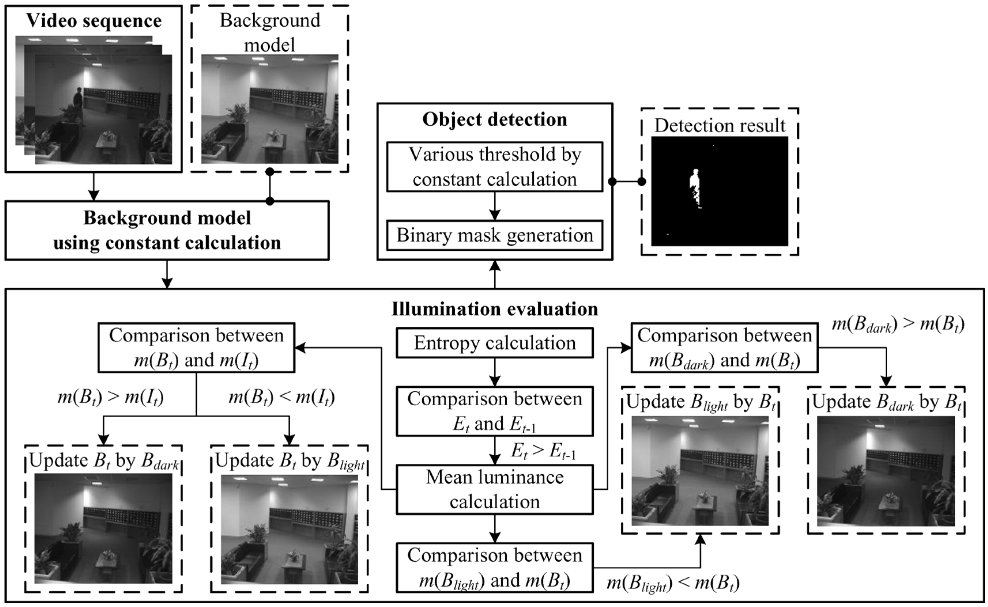

Fig. 1. Flowchart of the proposed background subtraction approach.

2) Illumination evaluation: According to entropy theory, a dark image possesses a low entropy value due to insufficient luminance of the image content, whereas a light image possesses a high entropy value because of sufficient luminance of the image content. Based on this characteristic, we use entropy formula to evaluate the change of illumination. For each incoming video frame $I_t(x,y)$, the probability density function (pdf) can be expressed as follows:

$$pdf(l) = n_l/(MN), \tag{13}$$

where l represents each luminance of the video frame, n_l represents the number of each luminance, and MN is the image resolution. Based on pdf, the entropy formula is expressed as follows:

$$E_t = -\sum_{l=l_{\min}}^{l_{\max}} pdf(l)\log\left(pdf(l)\right), \tag{14}$$

where E_t is the current entropy value, $l_{\min}$ is the minimum luminance of the current video frame $I_t(x,y)$, and $l_{\max}$ is the maximum luminance of the current video frame $I_t(x,y)$.

After calculating the entropy value, the illumination change can be measured as follows:

$$A_t = \begin{cases} 1, & \text{if } |E_t - E_{t-1}| > T, \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

where T represents the adjusted threshold which can be empirically set to 0.05. Note that A_t equals one to indicate the sudden illumination change. When A_t equals one, the background candidates can be updated by the following function:

$$B_t \text{ is used to update } \begin{cases} B_{light}, & \text{if } m(B_{light}) < m(B_t), \\ B_{dark}, & \text{if } m(B_{dark}) > m(B_t). \end{cases} \tag{16}$$

where B_{light} is the light background candidate, B_{dark} is the dark background candidate, $m(B_{light})$ is the mean luminance of B_{light}, $m(B_{dark})$ is the mean luminance of B_{dark}, and $m(B_t)$ is the mean luminance of B_t. The light and dark background candidates can be used to update the luminance of the current background model. When A_t equals one, the current background model can be updated by the following function:

$$B_t = \begin{cases} B_{light}, & \text{if } m(B_t) < m(I_t), \\ B_{dark}, & \text{otherwise,} \end{cases} \tag{17}$$

where $m(I_t)$ is the mean luminance of I_t. Note that both B_{light} and B_{dark} are set to B_0 for initialization.

3) Object detection: After generating the illumination-sensitive background model, absolute difference $\Delta_t(x,y)$ can be calculated between the incoming video frame $I_t(x,y)$ and the background model $B_t(x,y)$ as follows:

$$\Delta_t(x,y) = \left|B_t(x,y) - I_t(x,y)\right|. \tag{18}$$

For each video frame $I_t(x,y)$, the binary mask of moving objects $D_t(x,y)$ can be generated as

$$D_t(x,y) = \begin{cases} 1, & \text{if } \Delta_t(x,y) > V_t(x,y), \\ 0, & \text{otherwise,} \end{cases} \tag{19}$$

where $V_t(x,y)$ represents the threshold value, $D_t(x,y)$ equals 1 to represent a motion pixel and equals 0 to represent a background pixel. When $D_{t-1}(x,y)$ is regarded as a motion pixel, the threshold value can be calculated as follows:

$$V_t(x,y) = \begin{cases} V_{t-1}(x,y) + 1, & \text{if } V_{t-1}(x,y) < \Delta_t(x,y), \\ V_{t-1}(x,y) - 1, & \text{if } V_{t-1}(x,y) > \Delta_t(x,y). \end{cases} \tag{20}$$

Note that $V_0(x,y)$ is empirically set to twenty for initialization.

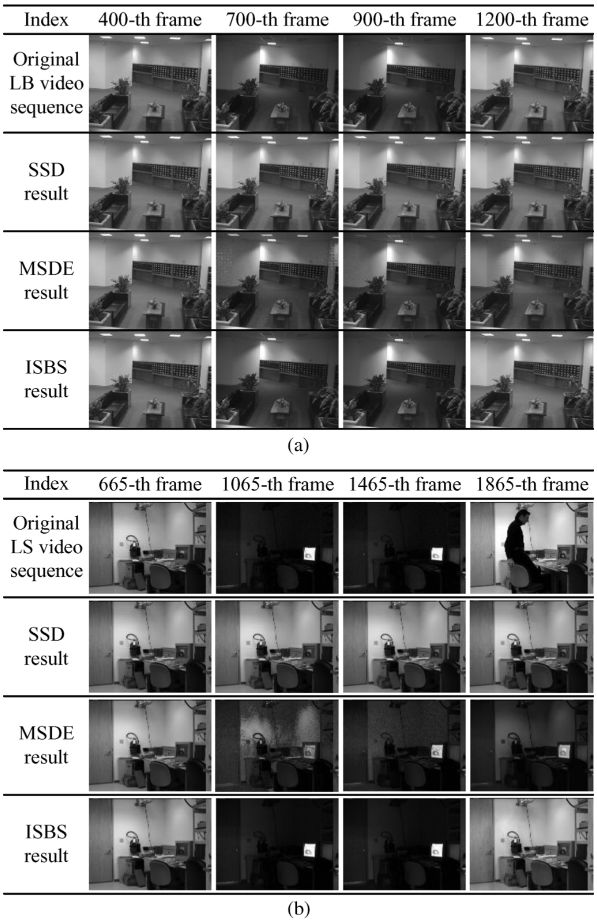

Fig. 2. Various background models are generated by each approach in the (a) LB and (b) LS video sequences.

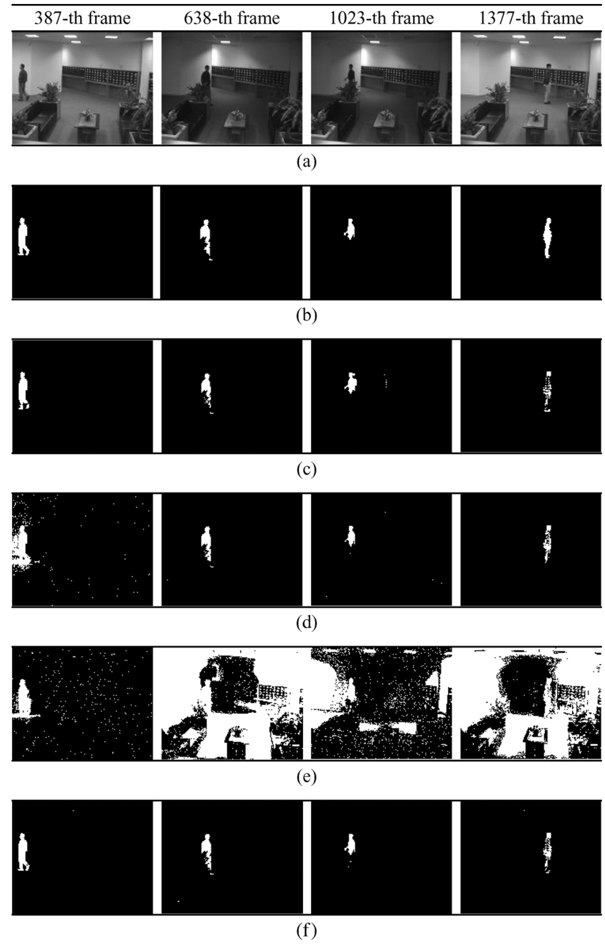

Fig. 3. LB video sequence. (a) Four original LB frames and (b) ground truths. The remaining four sub-pictures present the binary mask of moving objects generated by the (c) MTD [10], (d) SSD, (e) MSDE [14], (f), and ISBS approaches.

IV. Experimental Results

This section summarizes the experimental results for motion detection. Four motion detection approaches including the MTD approach [10], the SSD approach, the MSDE approach [14], and the proposed illumination-sensitive background subtraction (ISBS) approach are compared using a selection of scenes featuring sudden illumination change.

A. Qualitative Measurement

In general, illumination change usually acts as sources of complexity in environments. In order to test the performance of each background subtraction approach, the "lobby (LB)" and "light switch (LS)" sequences are used in our experiments. Fig. 2 shows the LB and LS video sequences and the generated background models obtained by the SSD approach, the MSDE approach [14], and the ISBS approach, respectively. According to Fig. 2, we can readily observe that sudden illumination changes occur in the original LB and LS sequences when lights are switched on/off. As a result, all compared approaches fail to generate an accurate background model when luminance is suddenly changed, with the only exception being the proposed ISBS approach.

Fig. 3 shows the LB video sequence, ground truths, and binary mask of moving objects obtained by the MTD [10], SSD, MSDE [14], and ISBS approaches, respectively. Since the MTD approach [10] doesn't maintain a background model but holds several previous frames for motion detection, the binary mask of moving objects can be generated at each frame as shown in Fig. 3(c). The SSD approach only calculates the temporal average at each pixel for background generation, thus the generated background model cannot deal with the illumination change. However, the SSD approach calculates the temporal standard deviation to determine the threshold value, which can be directly modified when the luminance change increases. As a result, Fig. 3(d) presents some comparable motion masks, but noises still exist in some frames. The MSDE approach [14] uses different time periods to generate several reference images for smooth generation of the background model. Unfortunately, this approach cannot adequately deal with the illumination change, thus causing the serious motion detection mistake shown in Fig. 3(e). Compared with other approaches, the proposed ISBS approach is more adaptable to the sudden illumination change because of its determination of light and dark background candidates, thus leading to the superior detection results shown in Fig. 3(f).

Fig. 4 shows the LS video sequence, ground truths, and binary mask of moving objects obtained by the MTD [10], SSD, MSDE [14], and ISBS approaches, respectively. In this case, the MTD approach [10] retains some reference frames that exhibit lower or higher luminance than that of the current video

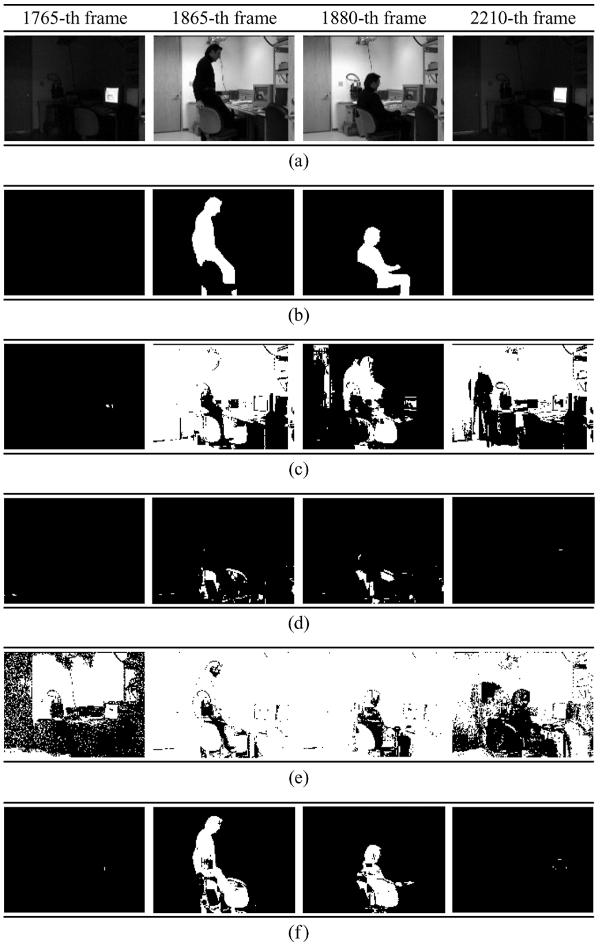

Fig. 4. LS video sequence. (a) Four original LS frames and (b) ground truths. The remaining four sub-pictures present the binary mask of moving objects generated by the (c) MTD [10], (d) SSD, (e) MSDE [14], (f), and ISBS approaches.

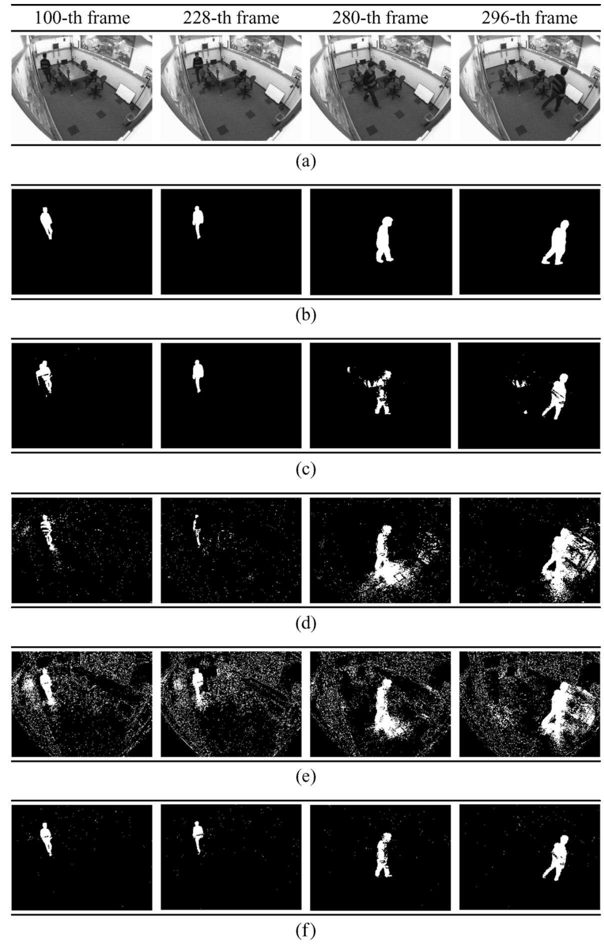

Fig. 5. IR video sequence. (a) Four original IR frames and (b) ground truths. The remaining four sub-pictures present the binary mask of moving objects generated by the (c) MTD [10], (d) SSD, (e) MSDE [14], (f), and ISBS approaches.

frame. Therefore, this approach doesn't work properly for motion detection as shown in Fig. 4(c). Although the SSD approach can generate the various threshold values through use of the temporal standard deviation, the difference between the simple background model and incoming video frame is too large to distribute the value of moving objects and background region displayed in Fig. 4(d). Again, the MSDE approach [14] still fails to generate the background model with illumination change, thus generating the inadequate detection results shown in Fig. 4(e). As indicated in Fig. 4(f), the ISBS approach presents a good detection result when compared with other approaches.

Fig. 5 shows the "intelligent room (IR)" video sequence, ground truths, and binary mask of moving objects obtained by the MTD [10], SSD, MSDE [14], and ISBS approaches, respectively. According to Fig. 5(a), one person is working in a room with constant illumination. Due to the low quality of device, some system noises are generated in this sequence. As indicated in Fig. 5(c), the MTD approach generates a motion mask with artificial ghost trails because the previous reference frames also involve the consecutive motion pixels of a moving object. According to Fig. 5(d) and (e), both SSD and MSDE approaches cannot deal with the system noises due to the inability of the threshold values of each pixel. Since the proposed ISBS approach only modifies the threshold value through adaption to moving objects, noise artifacts are reduced in the detection results shown in Fig. 5(f).

Fig. 6(a) shows the original "moved object (MO)" video sequence and Fig. 6(b) shows its ground truths, while other sub-pictures show the binary mask of moving objects obtained by the MTD [10], SSD, MSDE [14], and ISBS approaches, respectively. According to Fig. 6(a), one person enters the room and sits in a chair to make a phone call. As displayed in Fig. 6(c), the MTD approach generates the comparable motion masks but artificial ghost trails are still involved. According to Fig. 6(d) and (e), both SSD and MSDE approaches generate undesirable motion masks that involve apparent noise. Moreover, the person sits down for a long time, thus creating a misunderstanding between background and motion pixels at the 985-th frame. Compared with the other approaches, the proposed ISBS approach is adaptive to environmental fluctuations as shown in Fig. 6(f).

Fig. 7 shows the "airport (AP)" video sequence, ground truths, and binary mask of moving objects obtained by the MTD [10], SSD, MSDE [14], and ISBS approaches, respectively. As indicated in Fig. 7(a), the sizes of moving objects are different due to the depth of field. As displayed in Fig. 7(c), the MTD approach still generates artificial ghost trails in the detected motion masks. According to Fig. 7(d), the SSD approach can generate the comparable detection results. In contrast, the MSDE approach is capable of adequately detecting each moving object but causes the serious mistakes indicated in

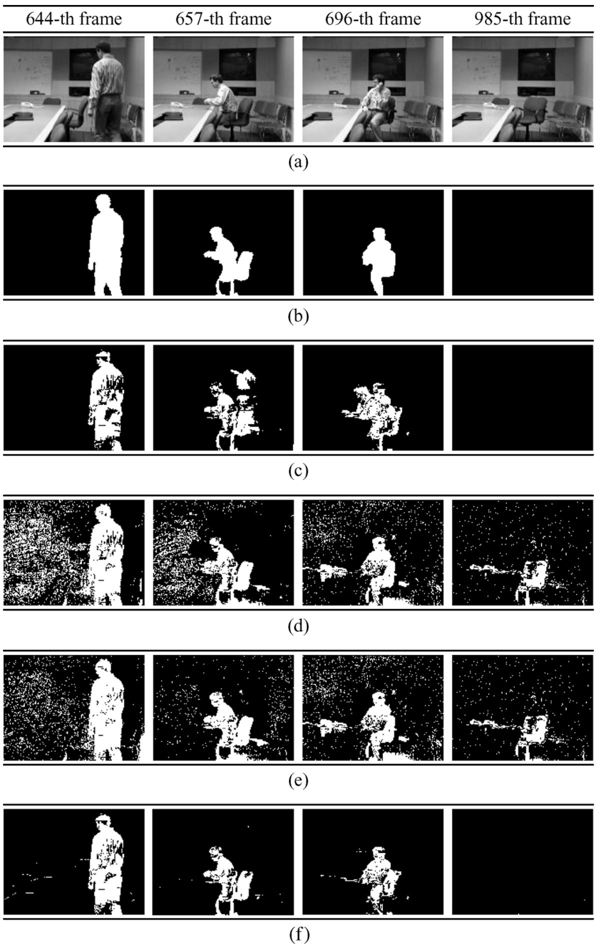

Fig. 6. MO video sequence. (a) Four original MO frames and (b) ground truths. The remaining four sub-pictures present the binary mask of moving objects generated by the (c) MTD [10], (d) SSD, (e) MSDE [14], (f), and ISBS approaches.

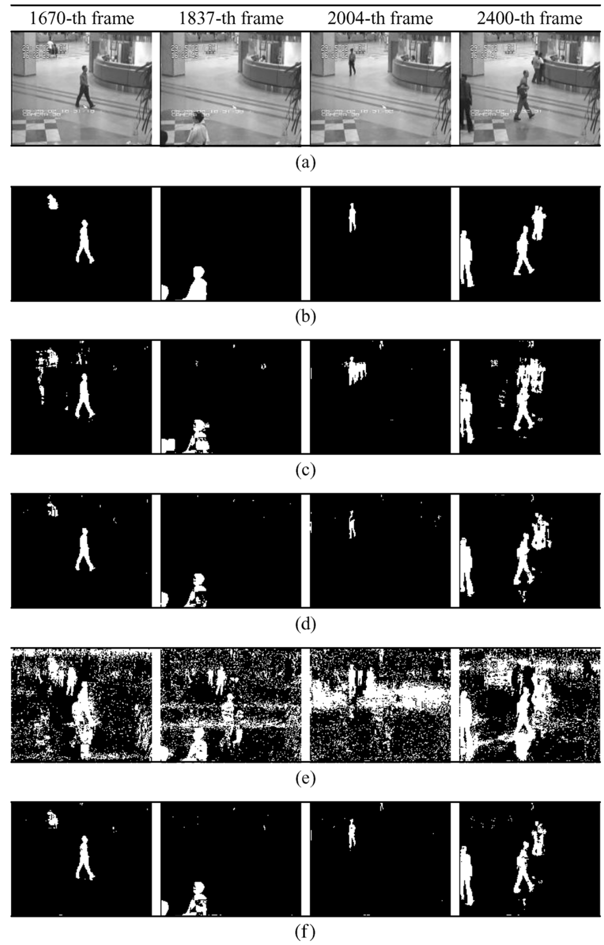

Fig. 7. AP video sequence. (a) Four original AP frames and (b) ground truths. The remaining four sub-pictures present the binary mask of moving objects generated by the (c) MTD [10], (d) SSD, (e) MSDE [14], (f), and ISBS approaches.

Fig. 7(e). As shown in Fig. 7(f), the proposed ISBS approach generates once again the desirable detection results.

Fig. 8 shows the "shopping mall (SM)" video sequence, ground truths, and binary mask of moving objects obtained by the MTD [10], SSD, MSDE [14], and ISBS approaches, respectively. Fig. 8(a) features many people shopping in a mall. With exception of the SSD and ISBS approaches, the use of the compared approaches still causes some problematic issues. Therefore, in addition to circumstances involving sudden illumination change, the proposed ISBS approach also works well for general situations.

B. Quantitative Measurement

In addition to qualitative assessment, quantitative measurement is also desirable in experiments. However, quantitative measurement of motion detection is not an easy task. In general, the accuracy of binary motion mask was usually assessed through use of $Recall$ and $Precision$ [7], [15]–[18]. The percentage of true positives in the detected motion mask can be evaluated using the $Recall$ function as follows:

$$Recall = \frac{\mathrm{tp}}{\mathrm{tp} + \mathrm{fn}}, \tag{21}$$

where tp is the total number of true positive pixels in the motion mask, fn is the total number of false negative pixels in the motion mask, and (tp+fn) indicates the total number of true positive pixels in the ground truth. On the other hand, $Precision$ presents the effect of false positives in the detected motion mask. The $Precision$ function can be expressed as

$$Precision = \frac{\mathrm{tp}}{\mathrm{tp} + \mathrm{fp}}, \tag{22}$$

where fp is the total number of false positive pixels in the motion mask and (tp+fp) indicates the total number of positive pixels in the detected motion mask.

However, a motion mask with all positives can generate the highest $Recall$ value and a motion mask with only one true positive can generate the highest $Precision$ value. Thus, simply using $Recall$ and $Precision$ alone cannot supply an objective measurement. To solve this problem, $Recall$ and $Precision$ can be mixed to form the hybrid metrics: F_1 and $Similarity$ [18]. The F_1 and $Similarity$ can be expressed as follows:

$$F_1 = \frac{2(Recall)(Precision)}{Recall + Precision}, \tag{23}$$

$$Similarity = \frac{\mathrm{tp}}{\mathrm{tp} + \mathrm{fp} + \mathrm{fn}}. \tag{24}$$

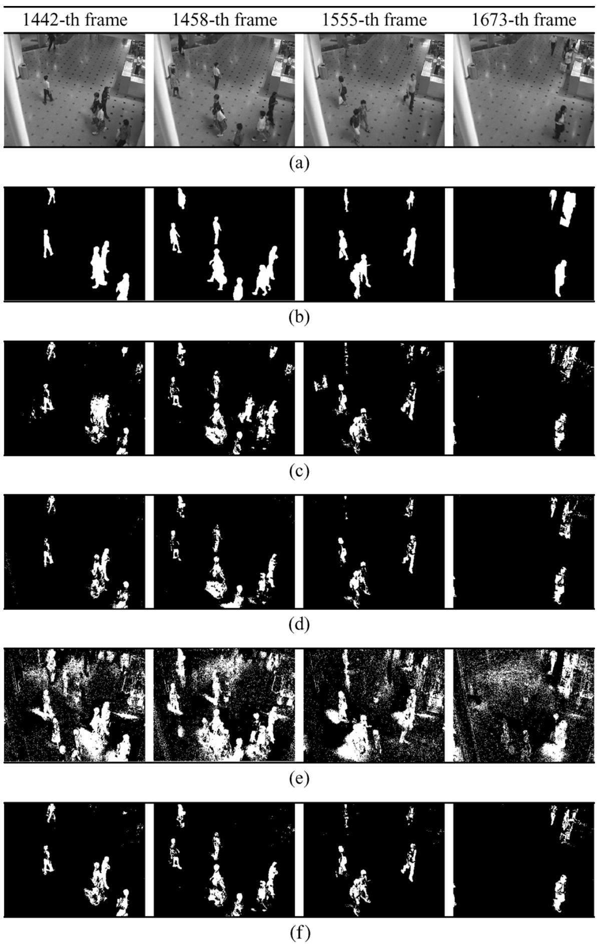

Fig. 8. SM video sequence. (a) Four original SM frames and (b) ground truths. The remaining four sub-pictures present the binary mask of moving objects generated by the (c) MTD [10], (d) SSD, (e) MSDE [14], (f), and ISBS approaches.

TABLE I
QUANTITATIVE MEASUREMENT FOR VARIOUS APPROACHES

Sequences	Metrics	MTD	SSD	MSDE	ISBS
LB	$Recall$	0.8197	0.8832	**0.9839**	0.9036
	$Precision$	0.9410	0.8348	0.1163	**0.9826**
	F_1	0.8609	0.8227	0.1754	**0.9339**
	$Similarity$	0.7739	0.7245	0.1162	**0.8899**
LS	$Recall$	0.6586	0.0355	0.5201	**0.7824**
	$Precision$	0.2125	0.1535	0.0585	**0.7282**
	F_1	0.2977	0.0577	0.1052	**0.7494**
	$Similarity$	0.1809	0.0300	0.0560	**0.5993**
IR	$Recall$	0.7408	0.6992	**0.9354**	0.8028
	$Precision$	0.7471	0.3475	0.1643	**0.8949**
	F_1	0.7293	0.4460	0.2759	**0.8436**
	$Similarity$	0.5805	0.2894	0.1621	**0.7310**
MO	$Recall$	0.7361	0.8356	**0.8955**	0.8106
	$Precision$	0.7876	0.5037	0.5696	**0.9480**
	F_1	0.7540	0.6193	0.6937	**0.8740**
	$Similarity$	0.6096	0.4495	0.5393	**0.7782**
AP	$Recall$	0.9168	0.7956	**0.9781**	0.9095
	$Precision$	0.6139	0.8687	0.1216	**0.8968**
	F_1	0.7195	0.8238	0.2087	**0.9002**
	$Similarity$	0.5697	0.7071	0.1211	**0.8206**
SM	$Recall$	0.6550	0.5825	**0.8890**	0.6763
	$Precision$	0.7399	0.8521	0.2351	**0.8693**
	F_1	0.6928	0.6910	0.3699	**0.7596**
	$Similarity$	0.5311	0.5288	0.2285	**0.6132**

TABLE II
COMPARISON OF FPS OF VARIOUS APPROACHES

Sequences	Resolution	MTD	SSD	MSDE	ISBS
LB	128×160	98	91	75	117
LS	120×160	103	97	81	119
IR	240×320	32	29	24	37
MO	120×160	102	92	83	118
AP	144×176	82	74	62	94
SM	256×320	27	26	21	34

Note that all metric-attained values range from 0 to 1, with higher values representing better accuracy. Table I lists the average $Recall$, $Precision$, F_1, and $Similarity$ of the detection masks generated by the MTD [10], SSD, MSDE [14], and proposed ISBS approaches for each tested video, respectively. This observation demonstrates that the proposed ISBS approach almost attains the highest accuracy among all the detection approaches.

C. Time Complexity

In order to further demonstrate the efficiency of the proposed ISBS approach, we evaluate the processing time of the motion detection approaches for each video. Note that all approaches are implemented by prototype C programming language on a laptop with a Core 2 Duo CPU, a 4 GB RAM, and running Windows 7 operation system.

Table II lists the frame per second (fps) of various approaches for the LB, LS, IR, MO, AP, and SM sequences, respectively. While observation of Table II demonstrates that all approaches can be easily implemented for real-time application, the proposed ISBS approach has significantly lower computational complexity than any of the other approaches.

V. CONCLUSION

This paper presents a novel background subtraction approach which features low computational complexity and high performance. Based on an illumination evaluation, the background model can be easily updated with each change in luminance level by the determination of background candidates. The proposed approach also calculates the adaptive threshold values in order to segment moving objects. Importantly, the proposed approach can increase the accuracy of object detection and reduce the processing time by constant calculation of the background model. Experimental results indicate that the proposed approach attains the most satisfactory outcome by both qualitative and quantitative evaluations compared with those of other approaches. Moreover, the analysis of time consumption also verifies that the proposed approach can be easily implemented in embedded systems with limited resources.

REFERENCES

[1] H. Cheng, N. Zheng, X. Zhang, J. Qin, and H. Wetering, "Interactive road situation analysis for driver assistance and safety warning systems: framework and algorithms," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 1, pp. 157–167, Mar. 2007.

[2] A. Licsar, T. Sziranyi, and L. Czuni, "Trainable blotch detection on high resolution archive films minimizing the human interaction," *Mach. Vis. Appl.*, vol. 21, no. 5, pp. 767–777, Aug. 2010.

[3] M. Castrillon, O. Deniz, C. Guerra, and M. Hernandez, "ENCARA2: Real-time detection of multiple faces at different resolutions in video streams," *J. Vis. Commun. Image R.*, vol. 18, no. 2, pp. 130–140, Apr. 2007.
[4] J. L. Castro, M. Delgado, J. Medina, and M. D. Ruiz-Lozano, "An expert fuzzy system for predicting object collisions. Its application for avoiding pedestrian accidents," *Expert Systems With Applications*, vol. 38, no. 1, pp. 486–494, Jan. 2011.
[5] M. Ding, Z. Tian, Z. Jin, M. Xu, and C. Cao, "Registration using robust kernel principal component for object-based change detection," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 761–765, Oct. 2010.
[6] G. Ginesu, D. D. Giusto, and V. Margner, "Detection of foreign bodies in food by thermal image processing," *IEEE Trans. Ind. Electron.*, vol. 51, no. 2, pp. 480–490, Apr. 2004.
[7] S. Park and J. Aggarwal, "A hierarchical Bayesian network for event recognition of human actions and interactions," *Multimedia Syst.*, vol. 10, no. 2, pp. 164–179, Aug. 2004.
[8] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 34, no. 3, pp. 334–352, Aug. 2004.
[9] C.-C. Chang, T.-L. Chia, and C.-K. Yang, "Modified temporal difference method for change detection," *Opt. Eng.*, vol. 44, no. 2, pp. 1–10, Feb. 2005.
[10] J.-E. Ha and W.-H. Lee, "Foreground objects detection using multiple difference images," *Opt. Eng.*, vol. 49, no. 4, p. 047201, Apr. 2010.
[11] F. Barranco, J. Diaz, E. Ros, and B. Pino, "Visual system based on artificial retina for motion detection," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 3, pp. 752–762, Jun. 2009.
[12] A. Doshi and A. G. Bors, "Smoothing of optical flow using robustified diffusion kernels," *Image Vis. Comput.*, vol. 28, no. 12, pp. 1575–1589, Dec. 2010.
[13] M. Oral and U. Deniz, "Centre of mass model—A novel approach to background modelling for segmentation of moving objects," *Image Vis. Comput.*, vol. 25, pp. 1365–1376, Aug. 2007.
[14] A. Manzanera and J. C. Richefeu, "A new motion detection algorithm based on $\Sigma-\Delta$ background estimation," *Pattern Recognit. Lett.*, vol. 28, pp. 320–328, Feb. 2007.
[15] P. Remagnino, T. Tan, and K. Baker, "Multi-agent visual surveillance of dynamic scenes," *Image Vis. Comput.*, vol. 16, no. 8, pp. 529–532, Jun. 1998.
[16] Z. Zhu, G. Xu, B. Yang, D. Shi, and X. Lin, "VISATRAM: A real-time vision system for automatic traffic monitoring," *Image Vis. Comput.*, vol. 18, no. 10, pp. 781–794, Jul. 2000.
[17] S. L. Dockstader and A. M. Tekalp, "Multiple camera tracking of interacting and occluded human motion," *Proc. IEEE*, vol. 89, no. 10, pp. 1441–1455, 2001.
[18] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE Trans. Image Process.*, vol. 17, pp. 1168–1177, Jul. 2008.

Fan-Chieh Cheng received the B.S. degree in electronic engineering from Huafan University, Taiwan (2005–2009). He is now a Ph.D. student at Department of Electronic Engineering, National Taiwan University of Science and Technology in Taipei, Taiwan.

He has published and presented more than ten articles in journals and at conferences. His research interests include digital image processing, video coding, and bus codec design, in particular, contrast enhancement, moving object detection, depth generation, super-resolution, motion estimation, and vehicular CAN bus transmission.

Shih-Chia Huang received the master's degree at the Computer Science Information Engineering Department of National Chiao Tung University in July 2005, and the doctoral degree at the Electrical Engineering Department, National Taiwan University, Taiwan, in January 2009.

He is an Assistant Professor at the Department of Electronic Engineering of National Taipei University of Technology, Taiwan. He has published and presented more than five papers in journals and at conferences, and also holds more than 24 U.S. and T.W. patents. His research interests include image and video coding, video transmission, video surveillance, error resilience and concealment techniques, digital signal processing, Java processor design, and embedded software and hardware co-design.

Shanq-Jang Ruan (S'01–M'02) received the B.S. degree in computer science and information engineering from Tamkang University, Taiwan (1992–1995), the M.S. degree in computer science and information engineering (1995–1997), and Ph.D. degree in electrical engineering (1999–2002) from National Taiwan University, respectively.

He served as an Electronic officer in R.O.C. Air Force during July 1997 and May 1999. From September 2001 to May 2002, he served as a software engineer in Avant! Corporation. From June 2002 to November 2003, he joined Synopsys as a senior software engineer. His research interests include power efficient circuits and systems, image processing, and electronic design automation. He was an assistant professor at Department of Electronic Engineering of National Taiwan University of Science and Technology from 2003 to 2007. He is currently an Associate Professor at Department of Electronic Engineering of National Taiwan University of Science and Technology.