Artif Life Robotics
DOI 10.1007/s10015-016-0347-9

ORIGINAL ARTICLE

Moving objects detection employing iterative update of the background

F. X. Arinto Setyawan[1] · Joo Kooi Tan[1] · Hyoungseop Kim[1] · Seiji Ishikawa[1]

Received: 7 April 2016 / Accepted: 22 November 2016

Abstract Detection of objects from a video is one of the basic issues in computer vision study. It is obvious that moving objects detection is particularly important, since they are those to which one should pay attention in walking, running, or driving a car. This paper proposes a method of detecting moving objects from a video as foreground objects by inferring backgrounds frame by frame. The proposed method can cope with various changes of a scene including large dynamical change of a scene in a video taken by a stationary/moving camera. Experimental results show satisfactory performance of the proposed method.

Keywords Object detection · Background models · Normal distribution · Adaptive models · Moving camera · Computer vision

1 Introduction

It is one of the most important issues in computer vision study to develop a technique for detecting objects from a video. The detection of moving objects in a scene is particularly important, since they might collide with a human in walking outdoors or in driving a car. There are many methods of detecting moving objects from a video. They include optical flows detection, pattern matching, applying HOG followed by tracking, etc. They have, however, a drawback that they only inform us, where moving objects are in a given image by a bounding box and do not provide their shape [1–3] which is indispensable to further processing, such as object recognition or object motion analysis. Therefore, many researchers [e.g., 4–8] have studied background subtraction which gives directly the shape of an object. They assume a stationary camera when inferring the background images. Recently, some researchers assume a moving camera [e.g., 9–13]. These proposals are, however, not very strong at sudden dynamic change of the background, including sudden illumination change.

This paper proposes a method of detecting moving objects from a video based on sequential background inference. The method infers the background images frame by frame and detects a set of pixels different from the background image as foreground images. They are expected to provide a moving object or moving objects. Further processing, such as object recognition, may clear what they are, but it is out of the scope of the present paper. A video-taking camera can be stationary or mobile in the proposed method. However, in the latter case, the camera motion is assumed to be slow. The main difference of the proposed method from existent methods is that the proposed method can cope with not only small disturbance in the background but also large change of gray values in a video. The proposed method can be applied to an automatic surveillance system indoors/outdoors, where large dynamical change may occur. The proposed method and some experimental results are given in the following sections.

This work was presented in part at the 21st International Symposium on Artificial Life and Robotics, Beppu, Oita, January 20–22, 2016.

✉ Seiji Ishikawa
ishikawa@ss10.cntl.kyutech.ac.jp;
etheltan@ss10.cntl.kyutech.ac.jp

[1] Department of Control Engineering, Kyushu Institute of Technology, Sensuicho 1-1, Tobata, Kitakyushu 804-8550, Japan

Published online: 20 January 2017

2 The proposed method

Given a video, a background model BGM is defined which contains the background model at every sample time denoted by BGM_t ($t=0,1,\ldots,T-1$). BGM_t is a set of normal distributions each of which gives a gray value distribution of every pixel on an image frame.

In the proposed method, the mean and the variance of the normal distribution $N_{t,h}$ vary according to the pixel h on the next frame f_{t+1} if it is a foreground pixel or a background pixel. The method has two strategies on the adaptation of the normal distribution according to the promptness of illumination change. If the illumination change is gradual, the model changes gradually; if it changes suddenly, the model also changes in a prompt way.

2.1 A background model

Let us denote a sample time by t ($t=0, 1, \ldots, T-1$) and an image frame at time t by f_t. Let us also denote a pixel at the position (m,n) ($m=0,1,\ldots,M-1$; $n=0,1,\ldots,N-1$) on f_t by $p_{t,h}$ ($h=Mn+m$) and its gray value by $f_{t,h}$. The background model of the proposed method represents the gray values which a pixel $p_{t,h}$ on frame f_t takes by a normal distribution $N_{t,h}(f_{t,h};\, \mu_{t,h},\sigma^2_{t,h}) \equiv N(f;\mu,\sigma^2)_{t,h} \equiv N_{t,h}$ [5]. Here, μ is the mean and σ^2 is the variance of the normal distribution. This signifies that, even if gray value $f_{t,h}$ varies under the following condition

$$\left|f_{t,h} - \mu_{t-1,h}\right| < \theta_h \sigma_{t-1,h}, \tag{1}$$

pixel $p_{t,h}$ is regarded as a background pixel. The mean value and the variance depend on time and pixel location. The threshold θ_h ($h = 0, 1, 2, \ldots, MN - 1$), in general, depends on pixel location and it is determined experimentally.

The proposed background model BGM is defined as follows:

$$\begin{aligned} \mathrm{BGM} &= \{\mathrm{BGM}_t \mid t = 0, 1, \ldots, T-1\} \\ \mathrm{BGM}_t &= \{N(f;\mu,\, \sigma^2)_{t,h} \mid h = 0, 1, \ldots, \mathrm{MN}-1\} \end{aligned} \tag{2}$$

with respect to the initial background model, the mean, μ_0, of the normal distribution at each pixel on the initial image frame is defined by the gray value of the pixel, and the variance, σ_0^2, is assumed a certain constant (actually, constant 1 is chosen in the performed experiment). This strategy works even if the initial image contains a moving object, since the proposed method repeats the update of the background model and the moving object disappears after some frames. If it remains in the images, it becomes part of the background.

In an outdoor environment, the background can vary depending on weather condition, for example. The proposed BGM can eliminate small fluctuation of the background gray values caused by weather changes, such as wind, rain, and illumination.

2.2 Extracting moving objects

By comparing the present image frame f_t and the background model BGM_{t-1} using Eq. (1), pixels $p_{t,h}$ ($h=0, 1, \ldots, \mathrm{MN}-1$) are divided into a set of foreground pixels and a set of background pixels, which are denoted by FG_t and BG_t, respectively. The pixels in FG_t give candidates of moving objects. Unlike the existent moving objects detection methods which do not use background subtraction, FG_t provides the shape of moving objects directly. This is the main advantage of the use of background images. It is, however, noted that further processing needs to be done to know what object it is.

2.3 An adaptive background model

The proposed method makes the background model at time t, BGM_t, adapt to varying environment by the parameter tuning of a single normal distribution. The method employs a single normal distribution, for simplicity, instead of employing the Gaussian mixture model [7, 14], since the emphasis is placed rather on the adaptability of the method to large dynamical change of a scene. The proposed method performs the update of a background model by tuning the average and the variance of the background model BGM_t. The average tuning is done by the following formula:

$$\mu_{t+1,h} = \begin{cases} \alpha f_{t+1,h} + (1-\alpha)\mu_{t,h} & \cdots\ p_{t+1,h} \in \mathrm{BG}_{t+1} \\ (1-\beta) f_{t+1,h} + \beta \mu_{t,h} & \cdots\ \left(p_{t+1,h} \in \mathrm{FG}_{t+1}\right) \cap \left(\left|\overline{f_{t+1,h}} - \overline{f_{t,h}}\right| < \omega\right), \\ \beta f_{t+1,h} + (1-\beta)\mu_{t,h} & \cdots\ \left(p_{t+1,h} \in FG_{t+1}\right) \cap \left(\left|\overline{f_{t+1,h}} - \overline{f_{t,h}}\right| \geq \omega\right) \end{cases} \tag{3}$$

$$\alpha = cN(f_{t+1,h};\mu_{t,h}, \sigma^2_{t,h}), \tag{4}$$

$$\beta = 1/(1 + kF_{t+1,h})\ . \tag{5}$$

Here, BG_{t+1} and FG_{t+1} are the set of background pixels and the set of foreground pixels, respectively, at time $t+1$: α and β are the parameters defined by Eqs. (4) and (5): $N(-;-,-)$ in Eq. (4) is a normal distribution and c is a normalizing constant to make the maximum value of α 1. α is close to 1, if the gray value of the pixel on a newly fed frame, $p_{t+1,h}$, does not change much, i.e., $f_{t+1,h} \approx \mu_{t,h}$. $F_{t+1,h}$ in Eq. (5) is the number of the most recent successive frames on which $p_{t,h}$ was judged a foreground pixel, and k is a constant to tune the influence of $F_{t+1,h}$. The longer pixel $p_{t,h}$ stays on foreground images, the smaller β becomes. The average gray value of image frame f_t and that of image frame f_{t+1} are denoted by $\overline{f_{t,h}}$ and $\overline{f_{t+1,h}}$, respectively: Parameter ω in Eq. (3) is a threshold determined by experiment.

The amount $\left|\overline{f_{t+1,h}} - \overline{f_{t,h}}\right|$ signifies the overall change of the gray values on successive image frames. If it is less than a specified value ω, a larger weight is given to the present average as defined in Eq. (3), whereas, if it is larger than ω, a larger weight is placed to the gray values of a newly fed image as in Eq. (3).

On the other hand, the variance tuning is done by

$$\sigma^2_{t+1,h} = \begin{cases} \alpha\,(f_{t+1,h} - \mu_{t,h})^2 + (1-\alpha)\sigma^2_{t,h} & \cdots\ p_{t+1,h} \in \mathrm{BG}_{t+1} \\ (1-\beta)(f_{t+1,h} - \mu_{t,h})^2 + \beta\sigma^2_{t,h} & \cdots\ p_{t+1,h} \in \mathrm{FG}_{t+1} \end{cases}, \tag{6}$$

$$\alpha = cN(f_{t+1,h};\mu_{t,h}, \sigma^2_{t,h}), \tag{7}$$

$$\beta = 1/(1 + kF_{t+1,h})\ . \tag{8}$$

It is noted that it is of no use to change the variance of a background model largely according to the overall large change of the gray values in a fed image, since Eq. (3: bottom) holds with a very small number of successive image frames, say, just a single frame, whereas Eq. (3: middle) holds with most of the fed image frames, as sudden large change is rare.

The background update strategy given by Eq. (3) signifies that Eq. (3: top & middle) are employed for the update of the model, if a fed scene contains gradual or small change in its gray values. On the other hand, Eq. (3: top & bottom) are used for the update, if the gray values of the image change largely. The degree of the gray value change is known by the change of the average value of the gray values of a scene. The judgment is done by use of parameter ω.

Since parameter k is a small number in Eq. (3: bottom), $\beta \approx 1$ with a newly fed image frame ($F_{t+1,h}$=1). Then, Eq. (3: bottom) becomes $\mu_{t+1,h} \approx f_{t+1,h}$. This means that, if large dynamical scene change has occurred, an updated background model is almost a copy of the fed image frame.

The update model given by Eq. (3) adapts to sudden large change in a scene, such as the sudden change of the weather, turning on/off the light in the room, or sudden appearance or disappearance of a large vehicle obstructing a camera view.

2.4 Background model creation with a moving camera

It is desirable that a background model is also created even if a camera moves when taking a video, such as using a hand-held camera. It is then necessary to know camera motion, which can be computed from observation of the background motion. In case of a stationary camera, the update algorithm of a background model given by Eqs. (3–5), and (6–8) directly applies to each pixel on a given image frame without difficulty, because the background is stationary and the location a pixel specifies does not change in the time lapse. In case of a moving camera, however, camera motion must be computed in advance to update the background model at time t to obtain the background model at time $t+1$. For this purpose, pixel-to-pixel correspondence between successive image frames is found computationally.

In the proposed method, camera motion is described by a 2D projective transform. Although it is approximate description, it works satisfactorily in the performed experiment. This may be because the distance between a camera and foreground objects is large enough to do the approximation.

The following procedure realizes the update of the background model in the case of a moving camera [13].

1. Extract feature points on image frame f_{t+1} using the Harris corner detector.
2. Find their corresponding points on image frame f_t using the Lucas–Kanade tracker.
3. Compute a 2-D projective transform $T_{t+1\to t}$employing the set of above feature point pairs (the feature points in the background image are the present concern: those in the foreground image are discarded as outliers by RANSAC).
4. Compute corresponding points of all the pixels of f_{t+1} on f_t using $T_{t+1\to t}$.
5. Compute the mean and the variance of every point obtained at step 4 from its nearest 4 pixels' normal distributions by use of bilinear interpolation to define the normal distribution at the point.
6. Assign the normal distribution of the point as the normal distribution of the pixel on f_{t+1} corresponding to the point.

3 Experimental results

The proposed method was applied to some real images to extract moving objects. Experiments employing the background model given by Eqs. (3–5), and (6–8) were done with respect to a stationary camera case and a moving camera case. The specifications of the used PC are OS: Windows 7 Enterprise, CPU: Intel core 2 Duo E7500 2.93 GHz, and memory: 4 GB. The parameter value in Eq. (1) is $\theta_h \equiv \theta = 25$ in the stationary camera case, whereas $\theta_h \equiv \theta = 30$ in the moving camera case. In Eqs. (3)–(5), $\omega = 100$ and $k = 0.1$ in both cases. They are experimentally chosen.

3.1 Stationary camera case

The proposed method was applied to three videos captured outdoors. In the first and the second videos, a person with an umbrella walks in the garden in the windy and rainy weather. In the third video, a car and a human pass in the rain and wind. It is noted that, since all these videos do not contain large illumination change, Eq. (3: top & middle) are practically employed for the update.

The results employing the second video (labeled 1; 750 frames) and the results using the third video (labeled 2; 225 frames) are shown in Fig. 1. In Fig. 1, (a) is the original image, (b) the result of the moving object detection, and (c) the result of evaluation. Because of strong wind and rainfall, the leaves of the trees in the background are swaying and raindrops are observed in the videos. However, as shown in (b), the background is almost removed satisfactorily and the foreground pixels are well detected. The swaying leaves and the rain drops were not detected, because their pixel intensities were within the threshold value. The foreground objects in the bottom image of (b) include a car and a pedestrian who is seen above the detected car.

Having obtained the ground truth image manually from the video, the results were evaluated employing *recall*, *precision*, and *F* value each defined by

$$recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{9}$$

$$precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{10}$$

$$F = 2 \cdot \frac{recall \cdot precision}{recall + precision}. \tag{11}$$

Here, TP means true positive, FN false negative, and FP false positive. In Fig. 1c, the TP area is indicated by red, the FN area by blue and the FP area by green.

As the result of having applied the proposed method to the three outdoor video images, the recall was 64.3%, the precision was 93.7% and the *F* value was 75.8 in average. The high value of the precision indicates that the gray value fluctuation in the background is well absorbed by the proposed background model.

We have also performed the background subtraction method employing the three outdoor video images and obtained recall: 91.4%, precision: 38.95% and *F* value: 53.7 in average. In this case, the small movement in the background was not effectively removed, resulting in the detection of noisy pixels and hence lower precision.

The effect of the background model given by Eq. (3) is shown in Fig. 2. In Fig. 2, (a) initial six image frames (frames 1, 2, 4, 5, 8, and 10) are chosen from a video (261 frames) in the time lapse and a room light is turned off at frame 4 (denoted by *f*_4), and it continues to *f*_10, (b) The

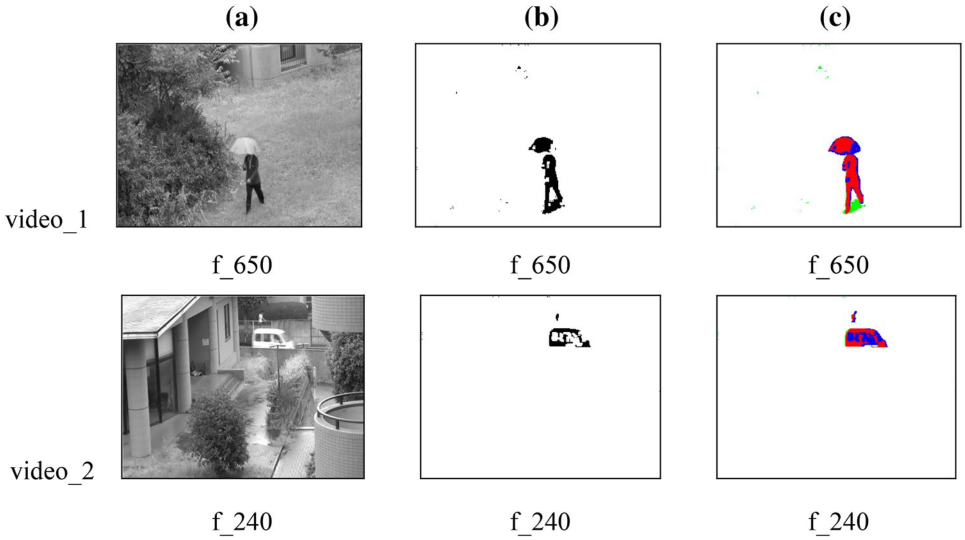

Fig. 1 Experimental results with the stationary camera case: **a** original images, **b** detected moving objects, and **c** result of evaluation (f_*i* means frame *i*)

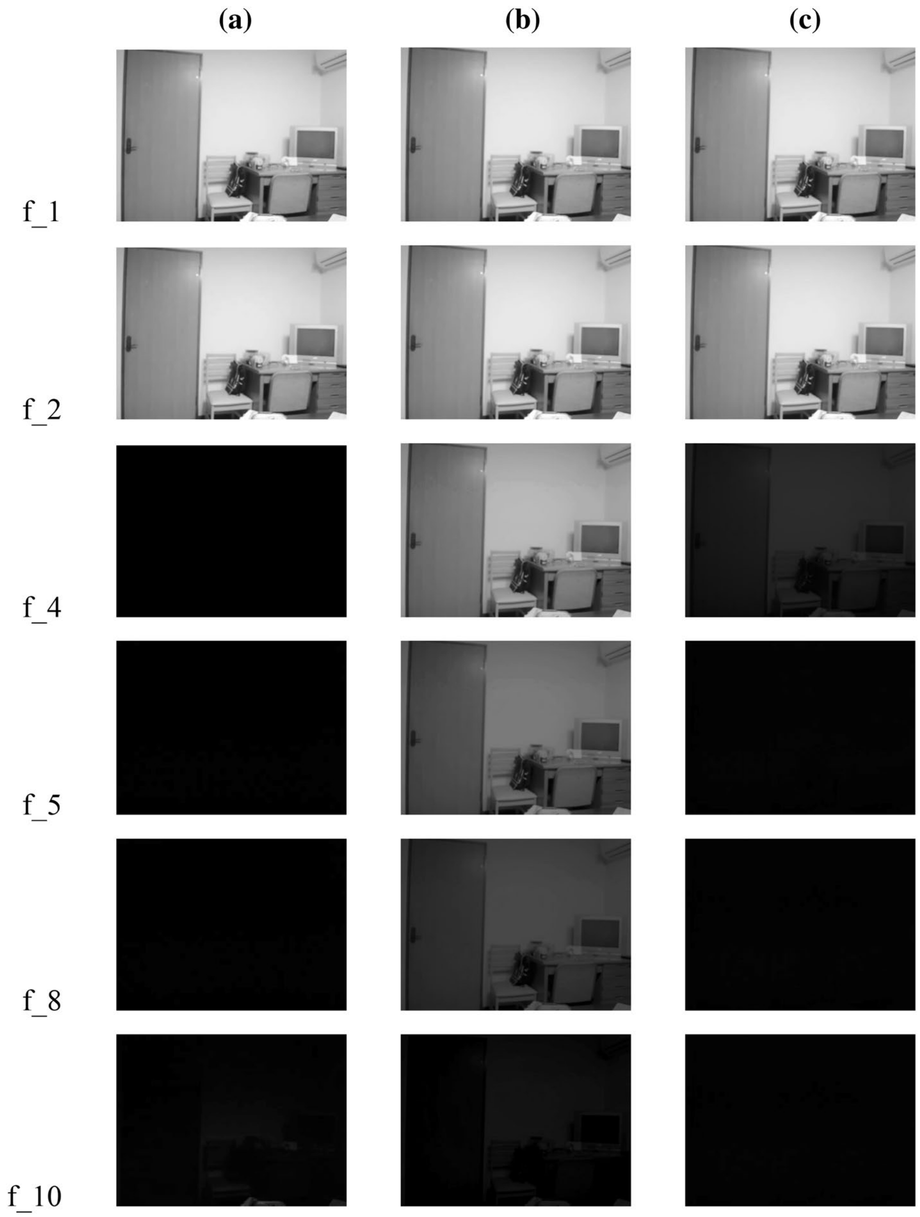

Fig. 2 Comparison of two background models: **a** original images, **b** background images updated by Eq. (3), and **c** background images updated by Eq. (3)

background images are updated by Eq. (3: top & middle), where the background changes from *f*_4 to *f*_10 gradually, and (c) on the other hand, the background images are updated employing Eq. (3) [including Eq. (3: bottom)], in which case the background image changes to dark promptly at *f*_4 and it continues to *f*_10.

3.2 Moving camera case

The proposed method was also applied to the videos taken by a hand-held moving camera. A result of a person detection under large illumination change is shown in Fig. 3. In the video used in this experiment, a room light was turned off and then turned on while a person walks in a room. The video contains 191 frames. Frames 60, 70, 80, 90, 100, 110, 120, 130, 140, and 150 are shown in Fig. 3.

In Fig. 3, (a) is part of the original video where *f*_90 is the time when the room light was turned off and (b) is the background model employing Eq. (3). In (b), the background model changed dark suddenly at *f*_90 according to *f*_90 of (a). (c) is the result of moving object (foreground pixels) detection. They are indicated by red. By comparing (a) and (b), foreground objects are detected in *f*_60, *f*_70, *f*_130, *f*_140, and *f*_150 successfully. Some noises are also detected in *f*_80. This may come from the gray values change on the curtain caused by a just passed person.

In this way, the proposed method updates the background adaptively and extracts a foreground object satisfactorily.

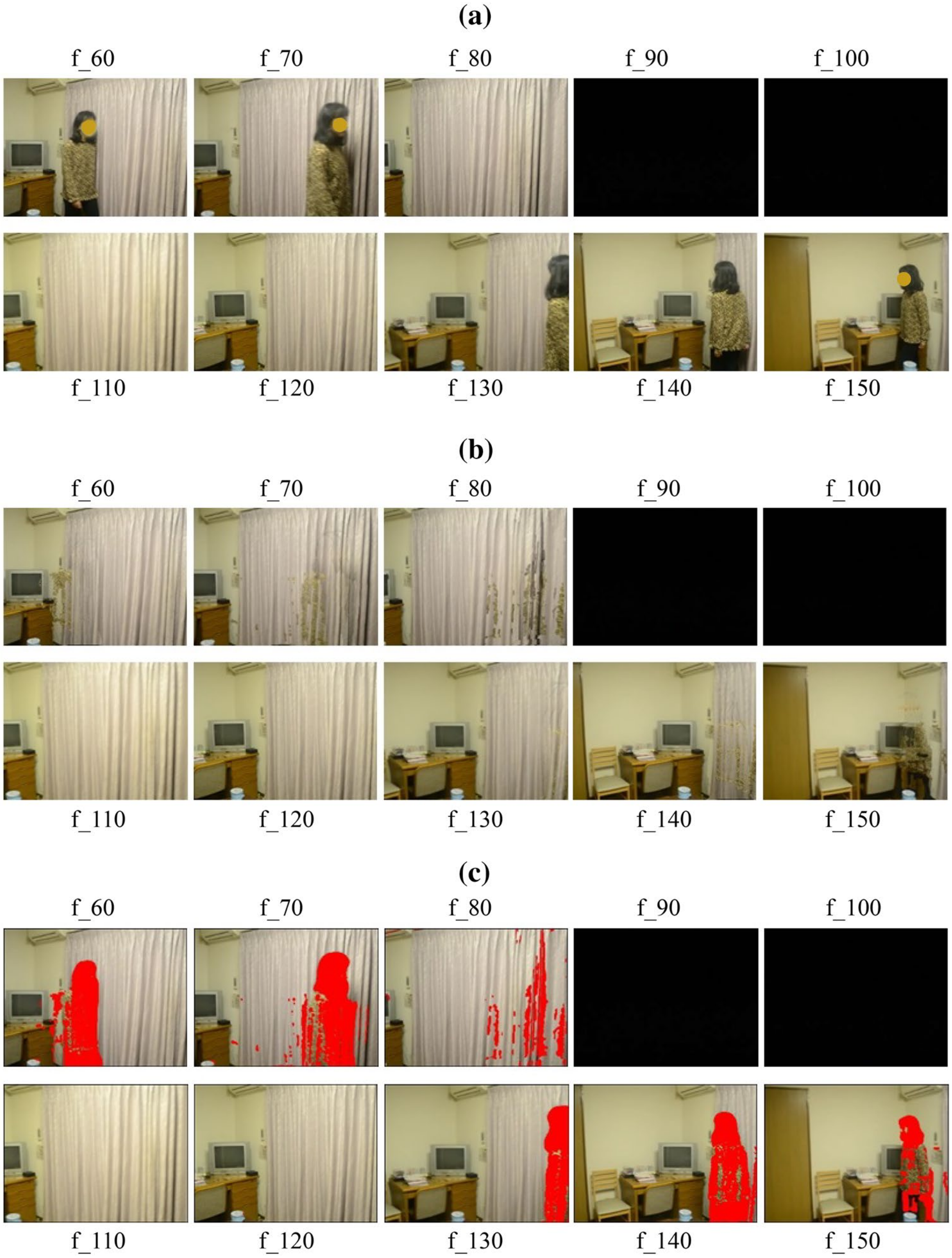

Fig. 3 Foreground object detection in the case of large illumination change: **a** part of the original image sequence, **b** background model employing Eqs. (3–8), and **c** result of moving object (foreground pixels) detection. Foreground pixels are indicated by *red*

4 Discussion and conclusion

A method of moving objects detection under dynamic background was proposed based on background subtraction. When making a background model, the proposed method can not only adapt to gradual scene change which most of the existent methods consider, but also adapt to sudden large change of the scene which makes the method different from others. The performance of the method was examined experimentally using some outdoor/indoor videos and satisfactory results were obtained. More number of videos containing various environments and events need to be employed to further examine the performance of the method, though.

The effectiveness of slow/prompt update of the background model defined by Eq. (3–5) was confirmed experimentally. The convergence of the background model is quick if large dynamical change is quick: It took just one or two frames to converge to a new background model when the room light was turned on/off. If the large change is a little slower, the method may iterate Eq. (3: middle & bottom) several times by comparing $\left|\overline{f_{t+1,h}} - \overline{f_{t,h}}\right|$ to ω before finally converging to a new background model.

Obviously, the number of the iteration reduces on quicker scene change.

Applications of the proposed method may include a surveillance system indoors/outdoors, where a large scene change may occur, such as large illumination change, scene change by a sudden camera movement, intrusion of a large object like a container car into a camera view, or the opposite case, etc.

Further improvement needs to be done to raise the precision of the foreground objects detection. In the proposed method, a single Gaussian was used with each pixel for simplicity to describe the background. The GMM considering the adaptation to large scene change remains to be developed. A method should also be taken into account which extracts the foreground pixels having the gray values similar to those in the background.

Acknowledgements This work was supported by JSPS KAKENHI Grant Number 16K01554.

References

1. Aslani S, Nasab HM (2013) Optical flow based moving object detection and tracking for traffic surveillance. Int J Elec Rob Electro Comm Eng 7:774–777
2. Prabhakar N et al (2012) Object tracking using frame differencing and template matching. Res J Appl Sci Eng Tech 4(24):5497–5501
3. Keerthana N, Ravichandran KS, Santhi B (2013) Detecting the moving object in dynamic backgrounds by using fuzzy-extreme learning machine. Int J Eng Tech 5(2):749–754
4. Bouwmans T (2011) Recent advance on statistical background modeling for foreground detection: a systematic survey. Recent Patent on Computer Science 4(3):147–176
5. Guyon C, Bouwmans T, Zahzah E (2012) Foreground detection based on low-rank and block-sparse matrix decomposition. Proc IEEE Int Conf Image Process (ICIP) 1225–1228
6. Oh SH, Javed S, Jung SK (2013) Foreground object detection and tracking for visual surveillance system: a hybrid approach. In: The 11th International Conference on Frontiers of Information Technology (FIT), pp 13–18
7. Zivkovic Z (2004) Improved adaptive Gaussian mixture model for background subtraction Proc. ICPR
8. Barnich O, Van Droogenbroeck M (2009) Vibe: a powerful random technique to estimate the background in video sequences. In: International conference on acoustics, speech, and signal processing (ICASSP 2009), pp 945–948
9. Szolgay D, Benois-Pineau J, Megret R et al (2011) Detection of moving foreground objects in video with strong motion. J Patt Anal Appl 14(3):311–328
10. Kwak S et al. (2011) Generalized background subtraction based on hybrid inference by belief propagation and Bayesian filtering. ICCV
11. ElQursh A, Elgammal A (2012) Online moving camera background subtraction. ECCV
12. Zamalieva D, Yilmaz A, Davis JW (2014) A multi-transformational model for background subtraction with moving cameras. ECCV
13. Setyawan FXA, Tan JK, Kim H, Ishikawa S (2014) Detecting moving objects from a video taken by a moving camera using sequential inference of background images. Artif Life Robot 19:291–298
14. Stauffer C, Grimson WEL (1995) Adaptive background mixture models for real time tracking. Proc CVPR 2:246–252