Making Background Subtraction Robust to Various Illumination Changes

Keigo Takahara, Takashi Toriu and Thi Thi Zin

Graduate School of Engineering, Osaka City University, Sugimoto 3-3-138, Sumiyoshi-ku, Osaka, Japan

Summary
Background subtraction is an important step used to segment moving regions in surveillance videos. Modern background subtraction techniques can handle gradual illumination changes but can easily be confused by rapid ones. In particular, varying illuminations cause significant changes in the representation of a scene in different spaces, which in turn results in the high levels of failure in such conditions. Especially, sudden illumination changes often falsely labeled as foreground objects, which may severely degrade the accuracy of object localization and detection. Thus in this paper, we propose a robust background modeling technique that overcomes this limitation by employing adaptive-length recursive hybrid median filters. This algorithm can achieve significantly better image quality than fixed length standard median filters when the images are corrupted by impulsive noise making our approach extremely robust to illumination changes, whether slow or fast. The performance of the proposed algorithm is compared with slow median filters, rapid median filters, and hybrid median filters by showing its effectiveness for occlusion handling in real time scenarios.

Key words:
Background subtraction, background estimation, illumination changing, object extraction.

1. Introduction

Background subtraction is a critical component of many applications, ranging from video surveillance to augmented reality. State-of-the-art algorithms can handle progressive illumination changes but, as shown in Fig. 1, remain vulnerable to sudden changes. Shadows cast by moving objects can easily be misinterpreted as additional objects. Background subtraction detects moving objects from the difference between the current frame and a background image. To obtain accurate detection of moving objects, the background image must be a representation of the scene with no moving objects and must be updated regularly so as to adapt to the varying lighting conditions and geometry settings [1]. Many background subtraction methods have been proposed in the literatures such as running Gaussian average, temporal median filter, mixture of Gaussians, kernel density estimation, etc. The major problems exist in these methods are either computation expensive or memory expensive. Temporal median refers the median of previous frames in a video sequence to establish a statistical background model for background subtraction. Lo and Velastin [2] first presented the temporal median background update technique for congestion detection system of underground platform. Cucchiara et al. [3] pointed out that temporal median filter provides an adequate background model which immediately reflects sudden scene change. Temporal median filter offers acceptable accuracy while achieving a high frame rate and having limited memory requirements [1]. Therefore, it has become one of most popular background subtraction methods [4-7].

The basic process of this method includes three steps. Firstly, a background model is established according to the temporal sequence of the frames. This background model provides a statistical description of the entire background scene. Secondly, the moving objects are detected based on the difference between the current frame and the background model to identify pixels in the video frame that cannot be adequately explained by the background model, and outputs them as a binary candidate foreground mask. Finally, the background model is updated periodically to adapt the variety of the monitoring scene. Although the approach is significant, it is difficult to extract the moving objects due to many factors [8], such as noise, motion changes of background, abnormal motion changes of the interested objects, cast shadows, and etc. There is a wide variety of techniques trying to improve the performance of extraction, not only the accuracy but also the speed. Even though many background subtraction algorithms have been proposed in the literature, the problem of identifying moving objects in complex environment is still far from being completely solved. There are several problems that a good background subtraction algorithm must solve correctly.

Manuscript received March 5, 2011
Manuscript revised March 20, 2011

Fig. 1 Two different input images of the same scene.

A good background model should specifically address several important properties. These include:

(i) Noise tolerance: the background model should exhibit appropriate noise immunity.
(ii) Dynamic background: the background model should be feature adaptability to dynamic background movements.
(iii) Clutter motion: the background model should not be sensitivity to repetitive clutter motion.
(iv) Convenient implementation: the background model should be able to be set up fast and reliably.

According to Cheung and Kamath [9], the existing background subtraction methods can be broadly classified into two as: (i) non-recursive and (ii) recursive. A non-recursive technique estimates the background based on a sliding-window approach. The *L* observed video frames are stored in a buffer, considering the existing pixel variations in the buffer the background image will be estimated. Since in practice the buffer size is fixed as time passes and more video frames come along the initial frames of the buffer are discarded which makes these techniques adaptive to scene changes depending on their buffer size. However, in the case of adapting to slow moving objects or coping with transient stops of certain objects in the scene the non-recursive techniques require large amount of memory for storing the appropriate buffer. With a fixed buffer size this problem can partially be solved by reducing the frame rate as they are stored. Some of the commonly-used non-recursive techniques are: frame differencing; median filter; linear predictive filter; non-parametric model.

On the contrary the recursive techniques instead of maintaining a buffer to estimate the background they try to update the background model recursively using either a single or multiple model(s) as each input frame is observed. Therefore, even the very first input frames are capable to leave an effect on new input video frames which makes the algorithm adapt with periodical motions such as flickering, shaking leaves, etc. Recursive methods need less storage in comparison with non-recursive methods but possible errors stay visible for longer time in the background model. The majority of schemes use exponential weighting or forgetting factors to determine the proportion of contribution of past observations. Some representative recursive techniques include: Approximated median filter; Kalman filter; Mixture of Gaussians.

Unfortunately, none of the existing background models can achieve desirable performance on all of the above mentioned criteria. Therefore we propose a recursive Hybrid Median Filter (HMF) background modeling technique based on slow and rapid recursive median filters. It is shown that every Recursive Median (RM) filter has an equivalent implementation as a HMF. Recasting the RM filter into this new form implies easier analysis, a more intuitive description and an extremely fast implementation. This paper also addresses the problem of real time background maintenance in complex environment. Rather than relying upon the distribution of the pixel value, two RM background models is presented to conserve the original and the current background separately. Moreover, through analyzing properties of object motion in image pixels and the background subtraction results, an adaptive HMF background update model is developed to select and maintain the suitable background model under different conditions. Experiments are performed using video sequences under different conditions indoor and outdoor. The results show that the proposed algorithm is effective and efficient in accurate background maintenance in complex environment. In particular, it is highly computationally cost effective and thus provides enough time for further target detection, tracking and classification.

The remainder of this paper is organized as follows. Section 2 outlines the overview of the proposed method. Section 3 describes the multiple layer background modeling algorithm. Section 4 and 5 contain the experimental results under various conditions and conclusion.

2. Proposed Method

The flow diagram of the algorithm is shown in Fig. 2. There are four major modules: preprocessing module, multiple layer background modeling, background subtraction and foreground object segmentation. In the preprocessing module the task of pixel level motion detection is carried out by using Gaussian kernel which identifies each pixel's changing character over a period of time by frame-to-frame difference for further analysis in multiple layer background modeling module. Fusing the detection result of pixel and the background subtraction results, the multiple layer background update model will establish and maintain the new hybrid recursive median filter background layer under different conditions based on two recursive median filters namely; Slow Median Filter (SMF) and Rapid Median Filter (RMF). In background subtraction step, each video frame is compared against the reference HMF background pixels in the current frame that deviate significantly from the background layer will be detected. In the end, a foreground object segmentation unit based on connected blob extraction and image down sampling is used to segment the moving objects. In our proposed system the output is a binary foreground mask $F_t(x, y)$ at time t with $F_t(x, y) = 1$ indicating a foreground pixel detected at location (x, y). There are three inputs to the system

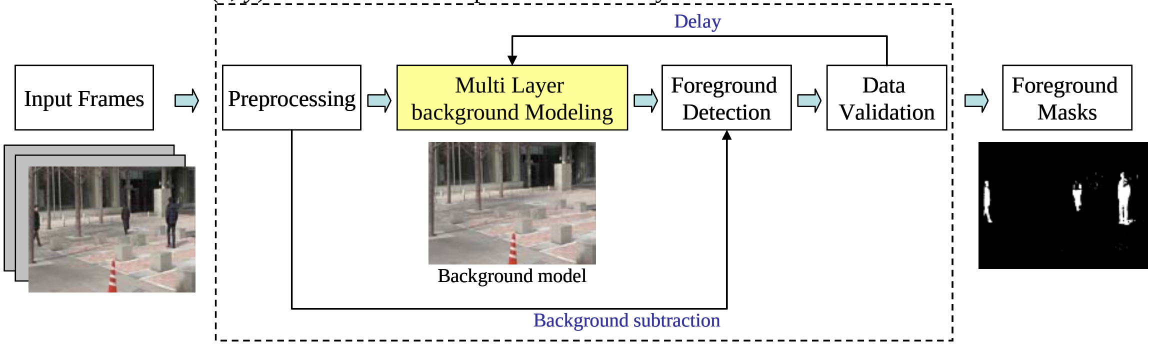


Fig. 2 Overview Flow Diagram of Proposed Method.

(i) $I_t(x, y)$ is the video frame at time t;
(ii) $S_t(x, y)$ is the binary foreground mask from a slow-median background subtraction algorithm;
(iii) $R_t(x, y)$ denotes the foreground mask obtained from rapid median filter background update by dynamic thresholding on the normal statistics of the difference between $I_t(x, y)$ and $I_{t-1}(x, y)$.

2.1 Preprocessing Module

In this module, we firstly use simple temporal and spatial smoothing to reduce camera noise. Smoothing can also be used to remove transient environmental noise. Then to ensure real-time capabilities, we have to decide on the frame-size and frame-rate which are the determining factors of the data processing rate. Another key issue in preprocessing is the data format used by the particular background subtraction algorithm. Most of the algorithms handle only luminance intensity, which is one scalar value per each pixel. However, color image, in either RGB or HSV color space, is becoming more popular these days. In the case of a mismatch, some time will be spent on converting the output data from the driver of the camera to the required input data type for the algorithm. The input to our algorithm is a time series of spatially registered and time-synchronized color images obtained by a static camera.

At first all pixels stored in a matrix (m by n), where m is equal to the total number of pixels in the image and n is equal to the number of frames for training sequence. Each image should be converted from two dimensional to one dimensional to get m. For the first image, all brightness values for all pixels are stored in the first column in the matrix. The same process is repeated for the second image in the second column and so on. Then, all brightness values for each pixel in each row are summed and divided by the number of frames for training sequences. The result is chosen to represent the initial background model for that pixel and so on.

Color Model: The input to our algorithm is a time series of spatially registered and time-synchronized color images obtained by a static camera in the RGB color space. This allows us to separate the luminance and chroma components by our camera hardware. The observation at pixel p at time t can then be written as: $p_t = (R_t, G_t, B_t)$. HSV color space also explicitly separates chromaticity and luminosity and so we use H value to form a new pixel representation $p_t = (R_t, G_t, B_t, H_t)$ to set a mathematical formulation for shadow detection [10-11]. At this point, we still have a major inconvenience in the model. Shadows are not translated as being part of the background and we definitely do not want them to be considered as an object of interest. To remedy this, we have chosen to classify shadows as regions in the image that differ in H but S, V rest unchanged. Since the H component is only sensible to illumination changes, it is in fact redundant for foreground or background object discrimination.

Mixture of Gaussian: We decided to use the Mixture of Gaussian (MoG) method to maintain a density function at pixel level. This choice is made for each pixel after studying the behavior of the camera sensor. However, to enable real-time computations, we assumed that the density function of each channel have only a single distribution which is obtained by considering only the mean value of the distribution.

3. Multi Layer Background Modeling Algorithms

The main problem of the background subtraction approach to moving object detection is its extreme sensitivity to dynamic scene changes due to lighting and extraneous events. While the background model eventually adapts to these "holes", they generate false alarms for a short period of time. Therefore, it is highly desirable to construct an approach to motion detection based on a background model that automatically adapts to changes in a self-organizing manner and without *a priori* knowledge. We propose to adopt a recursive median filter method based on finite automaton logical rules. The recursive techniques maintain a single background model that is updated with each new video frame. These techniques are generally computationally efficient and have minimal memory requirements. Throughout the remainder of this paper $p_t^c(x,y)$ and $B_t^c(x,y)$ are used to denote the value of channel c in RGB and Hue channels of the pixel at location (x; y) at time t for the incoming video sequence and the background model, respectively. The pixel location (x, y) is dropped when the spatial location is irrelevant.

$$B_t^c(x,y)=\begin{cases}(1-\beta_t)B_{t-1}^c+\alpha_t\Delta(p_t^c,p_{t-1}^c) & if\ \ p_{t-1}^c>B_{t-1}^c,\\(1-\beta_t)B_{t-1}^c-\alpha_t\Delta(p_t^c,p_{t-1}^c) & if\ \ p_{t-1}^c<B_{t-1}^c,\\B_{t-1}^c & if\ \ p_{t-1}^c=B_{t-1}^c.\end{cases}\quad(1)$$

where Δ ( ) is kernel function , α_t and β_t are the learning rate and forgetting rate schedules, respectively. In currently existing methods, both parametric and non-parametric, the learning rates are selected to be constant and have small values. This makes the convergence of the pixel model to be slow. In this aspect, McFarlane and Schofield [12] use a recursive filter to estimate the median using the update Eq.(1) in the case of α_t =1 , β_t = 0 and $\Delta(p_t^c,p_{t-1}^c)=1$. This update scheme becomes as shown in the following.

$$B_t^c=\begin{cases}B_{t-1}^c+1, & if\ \ p_{t-1}^c>B_{t-1}^c,\\B_{t-1}^c-1 & if\ \ p_{t-1}^c<B_{t-1}^c,\\B_{t-1}^c & if\ \ p_{t-1}^c=B_{t-1}^c.\end{cases}\qquad(2)$$

The recursive background model according to the update scheme in Eq.(2) is termed in this paper as Slow Median Filter (SMF).

In order to speed up rapidly the modeling convergence, in the proposed method we build a schedule for learning the background model at each pixel based on its history. At early stages the learning occurs faster (α_t =1) and by time it decreases and converges to the target rate ($\alpha_t\rightarrow\alpha_t(0)$). The forgetting rate schedule is used to account for removing those values that have occurred long time ago and no longer exist in the background. These schedules will make the adaptive learning process converge faster, without compromising the stability and memory requirements of the system.

Also training these rates independently for each pixel based on spatial changes in the scene makes the convergence more effective for different situations. This learning schedule is shown in equation:

$$\alpha_t(0)=(\max\{\Delta(p_t^c,p_{t-1}^c)\})\big/m,$$

where m is the maximum of channel values which are 255 when c belong to RGB and 360 for c in HSV and $\Delta(p_t^c,p_{t-1}^c)=\left|p_t^c-p_{t-1}^c\right|$. Thus we have adopted four background update schemes three in RGB and one in HSV. The corresponding background model is named as rapid median filter (RMF).

We now introduce the Hybrid Median Filter (HMF) based on the two background models SMF and RMF. In this method, a detection stage based on the application of two background subtraction methods at different frame rate is applied. The two models are based on the GMM employing one model for short-term detection (updating it every frame) and another for long-term detection (updating it every n frames). Rapid median filter background is adapted faster and the scene changes are introduced more quickly on it. On the other hand, slow median filter background is adapted to the changes of the scene at a lower learning rate. Then, the foreground masks of the two models are computed at every frame and a combination of them is performed and branded as hybrid median filter as shown in Table. 1.

4. Experimental results

In order to have a quantitative evaluation of the performance, we have selected ten frames at regular intervals from each test sequence, and manually highlighted all the moving objects in them. These “ground-truth” frames are selected randomly from the ten test sequences to minimize the effect of the initial adaptation of the algorithms. This sampling rate allows the persons to move a reasonable distance, making each ground-truth frame sufficiently different from others. In the manual annotation, we highlight only the pixels belonging to stationary objects and persons that are actually moving at that frame. The ground-truth frames showing only the moving objects are shown in Fig. 3(a-ii). Our comparison of three algorithms is performed using a diverse set of 5 outdoor indoor video sequences. One sample sequence using SMF, RMF and HMF HMF are shown in Fig. 3.

Table 1: Determination of update conditions in HMF

Output of RMF	BG	BG	FG	FG

Output of SMF	BG	FG	BG	FG
The value of HMF	The value of input frame	The value of RMF	The value of SMF	No change

BG: Background, FG: Foreground

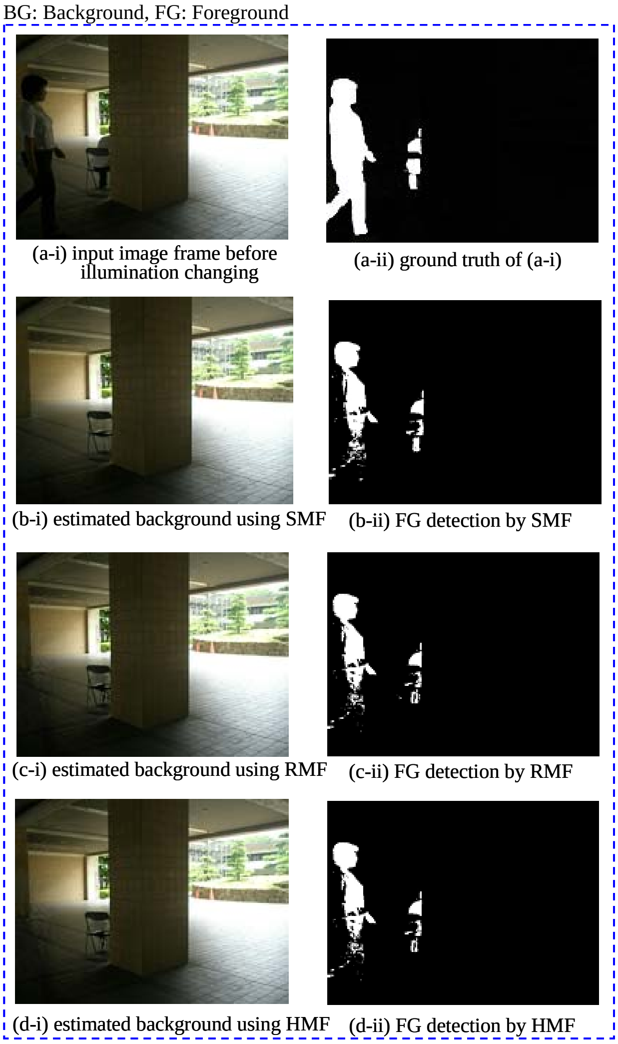


Fig. 3 Comparison of detected FG regions (no illumination change) using SMF, RMF and HMF.

The outdoor sequences present a significant challenge as they contain moving background elements, objects moving at varying speeds, and objects of varying sizes, campus. The indoor sequences are also challenging which exhibit examples of shadows and varying lighting conditions. For evaluating the proposed techniques we make use of the aforementioned videos taken in an international airport and the university campus since these cover a range of interesting scenarios and are all real footage. Some more examples are shown in Fig. 4, Fig. 5 and Fig. 6, respectively. We use two information retrieval measurements, recall and precision, to quantify how well each algorithm matches the ground-truth. They are defined in our context as follows:

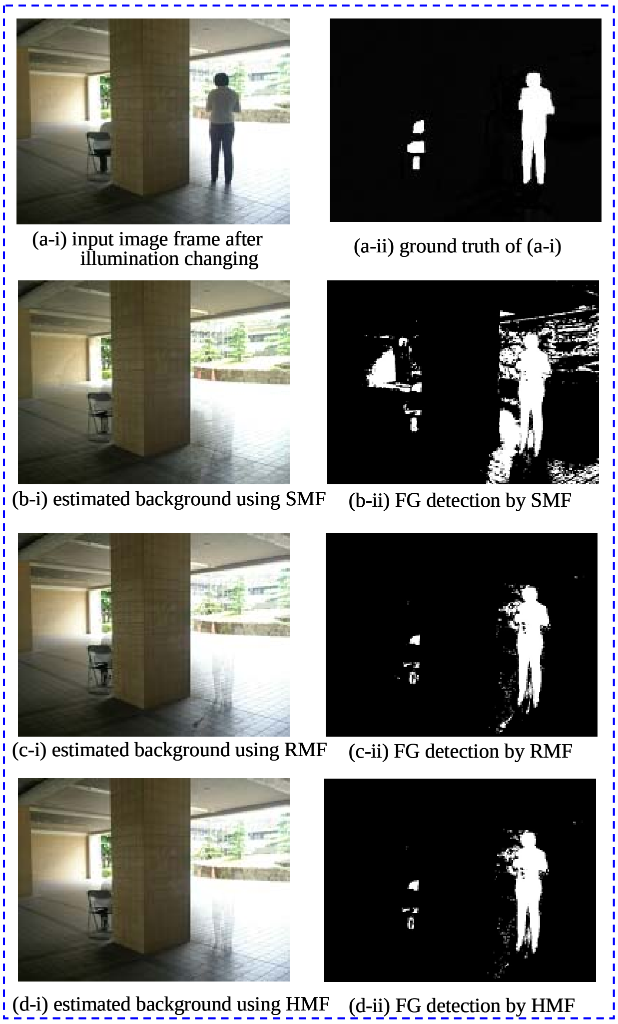


Fig. 4 Comparison of detected FG regions (illumination changing) using SMF, RMF and HMF.

$$\text{Recall} = \frac{\text{No. of FG pixels correctly identified by the algorithm}}{\text{No. of FG pixels in ground - truth}}$$

$$\text{Precision} = \frac{\text{No. of FG pixels correctly identified by the algorithm}}{\text{No. of FG pixels detected by the algorithm}}$$

Recall and precision values are both within the range of 0 and 1. When applied to the entire sequence, the recall and precision reported are averages over all the measured frames. Typically, there is a trade-off between recall and precision— recall usually increases with the number of foreground pixels detected, which in turn may lead to a decrease in precision. A good background algorithm should attain as high a recall value as possible without sacrificing precision (Fig. 7).

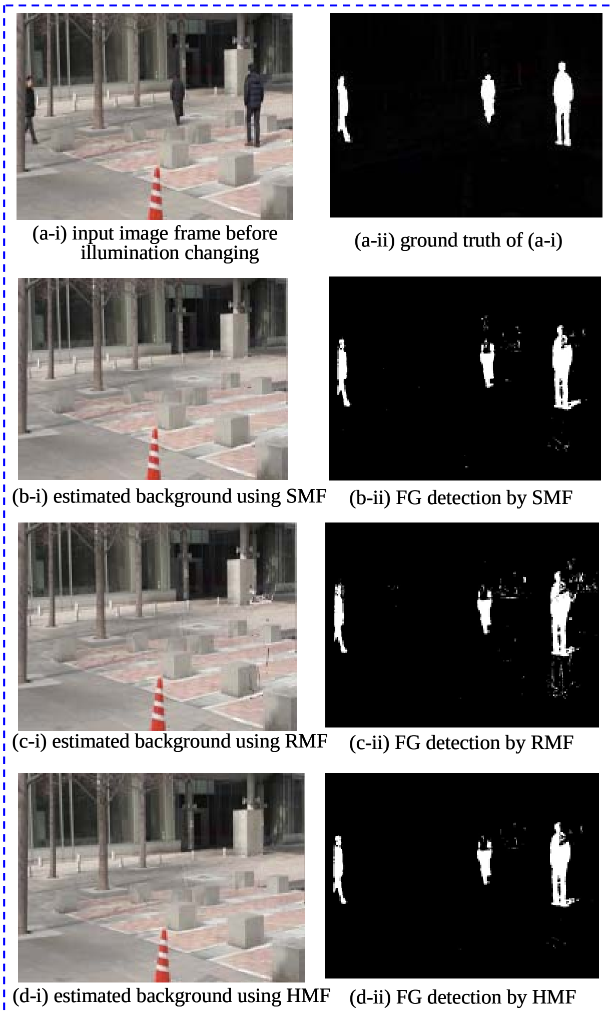


Fig. 5 Outdoor video sequence (illumination is not changed): comparison of detected FG regions using SMF, RMF and HMF.

In our experiments, we vary the parameters in each algorithm to obtain different recall-precision operating points. The resulting graphs for the test sequences are shown in Fig. 8(a) and Fig. 8(b). There are four plots containing three graphs one for each of three models, SMF, RMF and HMF. The first plot corresponds to precision rates. The second plot corresponds to recall rates. The third and fourth plots show both recall-precision curves for cases without and with illumination changes.
Based on the measurements shown in Figure 5 and visual examination on the resulting foreground masks, we observe that with the appropriate parameters, HMF achieves the best precision and recall.

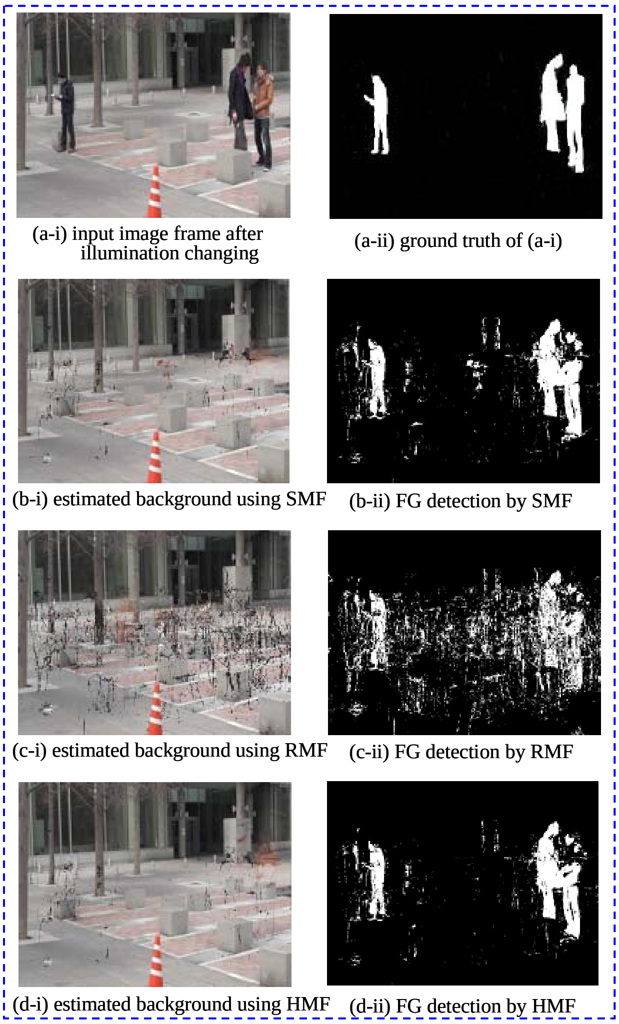
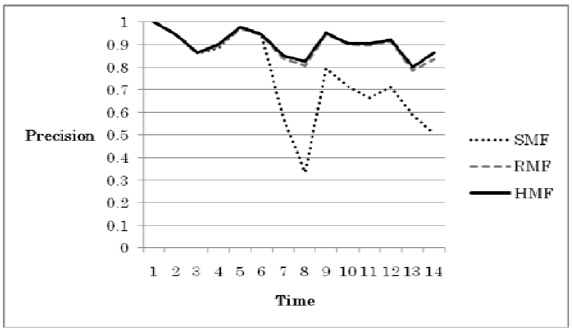


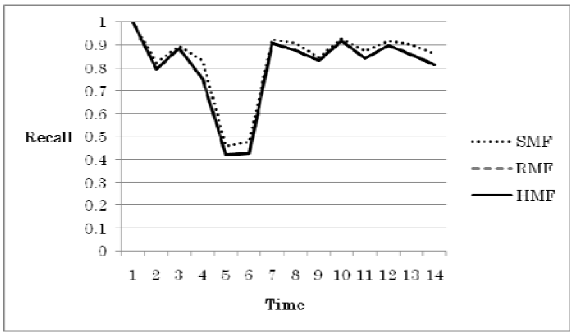
Fig. 6 Outdoor video sequence under illumination changing: comparison of detected FG regions using SMF, RMF and HMF.

5. Conclusions

In this paper, a recursive hybrid median filter learning scheme for background and foreground modeling is presented. The adaptive learning and forgetting rates proposed here make the generated models adapt to gradual and sudden changes. The model is automatically updated based on two median filter models and add to the accuracy of the overall performance. The proposed method is performed by competitively comparing these models to achieve temporal coherence. The experimental results show that the system converges reasonably fast to the underlying models and has produced encouraging results for more robust, reliable and applicable approach for real time background modeling for surveillance on both short and long time period through many video sequences used to test it.

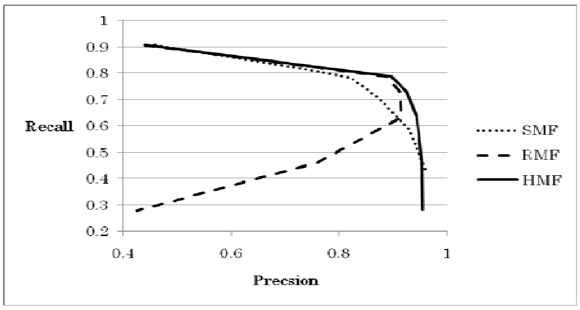


(a) precision rate

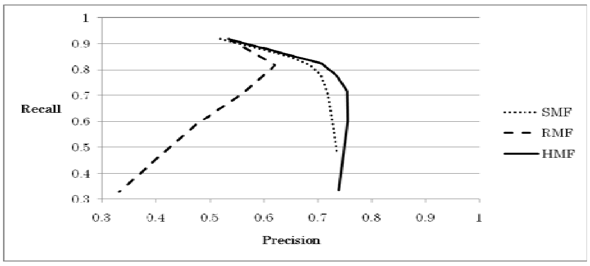


(b) recall rate

Fig. 7 Detected result on a video sequence under illumination changing.



(b) no illumination changing situation



(b) illumination changing situation

Fig. 8 Precision vs. recall rates.

Acknowledgments

This work is partially supported by SCOPE: Strategic Information and Communications R&D Promotion Program (10103768). We thank Scope project members and the students of Physical Electronics and Informatics, for their participations in producing tested videos.

References

[1] M. Piccardi, “Background subtraction techniques: a review,” *in Proc. of IEEE Intl. Conf on Systems, Man and Cybernetics*, vol. 4, pp. 3099-3104, 2004.

[2] B.P.L Lo and S.A. Velastin, “Automatic congestion detection system for underground platforms,” *Proc. of ISIMP 2001*, pp. 158-161, May. 2001.

[3] R. Cucchiara, C. Cranna, M. Piccardi and A. Prati, “Detecting moving objects, ghosts and shadows in video streams,” *IEEE Trans on Pattern Anal. and Machine Intell.*, vol. 25, no. 10, pp. 1337-1442, 2003.

[4] Sen-Ching S.Cheung and Chandrika Kamath “Robust techniques for background subtraction in urban traffic video” *in the 9th IEEE Intl. Workshop on Video Surveillance at IEEE Intl. Conf. on Computer Vision (ICCV 2009)*, 2009.

[5] Mohamed, S.S. Tahir, N.M. Adnan, R. “Background modeling and background subtraction performance for object detection”, *in 6th Intl. Conf. on Signal Processing and Its Applications (CSPA2010)*, 2010.

[6] B. Yang, Y. Guo, Y. Ming, A. Men, “An Effective Background Subtraction under a continuously and rapidly Varying Illumination,” *in Second Intl. Conf. on Future Networks (ICFN’10)*, 2010.

[7] A. Bayona, “Comparative Evaluation of Stationary Foreground Object detection Algorithms Based on Back ground Subtraction Techniques,” *Proc. of 6th IEEE Intl. Conf. on Advanced Video and Signal Based Surveillance (AVSS ’09)*, 2009.

[8] F. Y. Hu, Y. N. Zhang, L. Yao, and J. Q. Sun. “A New Method of Moving Object Detection and Shadow Removing,” *Journal of Electronic(China)*, vol. 24, no.4, pp. 528-536, 2007.

[9] S.C. Cheung and C. Kamath, “Robust techniques for background subtraction in urban traffic video,” *Video Communications and Image Processing, SPIE Electronic Imaging, UCRL Conf. San Jose*, vol. 200706, Jan. 2004.
[10] Prati, A., Mikic, I., Trivedi, M.M., Cucchiara, R., Detecting moving shadows: algorithms and evaluation, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, pp. 918-923, 2003.
[11] Cucchiara, R., Piccardi, M., Prati, A., Detecting moving objects, ghosts and shadows invideo streams. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp.1337-1342, 2003.
[12] N. McFarlane and C. Schofield. Segmentation and tracking of piglets in images. *Machine Vision and Applications*, vol. 8, no. 3. pp. 187-193, May 1995.

Keigo Takahara received the B.E. in Information Engineering from Osaka City University, Osaka, Japan in 2011. His research interests include image processing and background modeling.

Takashi Toriu received the B.Sc. in 1975, M.Sc. and Ph.D. degree in physics from Kyoto University, Kyoto, Japan, in 1977 and 1980, respectively. He was a researcher in Fujitsu Laboratories Ltd. from 1982 to 2002, and now he is a Professor of Osaka City University. His research interests are in the areas of image processing, computer vision, and especially in modeling of human visual attention. He is a member of IEEE, IEICE, IPSJ, ITE and IEEJ.

Thi Thi Zin received the B.Sc. degree (with honor) in Mathematics in 1995 from Yangon University, Myanmar and the M.I.Sc degree in Computational Mathematics in 1999 from University of Computer Studies, Yangon, Myanmar. She received her Master and Ph.D. degrees in Information Engineering from Osaka City University, Osaka, Japan, in 2004 and 2007, respectively. From 2007 to 2009, she was a Postdoctoral Research Fellow of Japan Society for the Promotion of Science (JSPS). She is now a specially appointed Assistant Professor of Graduate School of Eng, Osaka City University. Her research interests include human behavior understanding, ITS, and image recognition. She is a member of IEEE.