Simultaneous Background/Foreground Segmentation and Contour Smoothing with Level Set based Partial Differential Equation for Intelligent Surveillance Systems over Network

Suk-ho Lee, Nam-seok Choi,
and Byung Gook Lee
Dongseo Univ., Dept. Multimedia Eng.
San 69-1 Churye-2-Dong, Sasang-gu
Busan 617-716, Korea
petrasuk@gmail.com

Moon Gi Kang
Yonsei Univ., Dept. of Electric and Electronic Eng.
134 Shinchon-dong, Seodaemun=gu
Seoul 120-749, Korea
mkang@yonsei.ac.kr

Abstract

In this paper, we propose a level set based energy functional, the minimization of which results in simultaneous background modeling, foreground segmentation, and contour smoothing. The simultaneous dealing of background modeling and foreground segmentation has the effect that the two processes constrain each other positively, such that a good estimate of the background can be obtained with a small number of frames, and a temporal change in the scene is reflected quickly in the construction of the background image. Furthermore, the simultaneous level set based contour smoothing eliminates spurious regions, and smooths the contour that encompasses the object, so that a good representation for the boundary of the object is obtained. The level set based approach makes it possible to derive a level set based Euler-Lagrangian equation, which can be directly implemented and works in real-time.

1. Introduction

Nowadays, there is a large demand for intelligent surveillance systems that can automatically detect the object. This is due to the fact that it is difficult for the observer who monitors a property or a room by a surveillance camera to look at the screen continuously (see Fig. 1). Therefore, algorithms which can automatically detect the moving object are the key technologies for intelligent surveillance systems. Background subtraction refers to the class of motion detection techniques that segment out moving objects by comparing an observed image with an estimate of the background image which is usually estimated from a given sequence taken of the same scene for a certain time interval. However, the difficulty in estimating the reference background image lies in the fact that all the frames in the given image sequence may contain moving objects, where ideally the reference background image should contain no moving objects in it. The difficulty increases when the memory of the surveillance system is limited, and only a small number of frames can be used in the modeling.

Another difficulty lies in the fact that there exists a tradeoff between the acquirement of a clean background image and the fast recognition of changes in the scene, i.e., either tails of the moving object become apparent in the background image or the background image is updated slowly. For a fast recognition of changes in the scene, it is crucial that only a small number of frames are used in the background modeling.

Algorithms that model the background image by modeling the color of each pixel with a single or a mixture of Gaussians [4]-[6] usually need more than hundreds of frames for training, and therefore, a change in the scene is reflected very slowly in the background image. Background modeling techniques that update the reference background image by blending the current background image with the current frame [1]-[3] also need many frames to reflect a change in the scene depending on the blending parameter.

The number of frames used in the background modeling can be reduced if the mutual dependence of the foreground segmentation and the background image modeling are taken into account. A variational approach in which the background image modeling and the foreground segmentation affect each other positively has been introduced in [7]. However, an Euler-Lagrange equation cannot be obtained directly from the variational form, and therefore, the implementation is based on half quadratic minimization which takes a lot of time, making it difficult for the algorithm to be applied for real-time applications.

In this paper, we propose an energy functional that for-

978-0-7695-3575-3/09 $25.00 © 2009 IEEE
DOI 10.1109/CISIS.2009.198

mulates the problem of simultaneous background modeling, foreground segmentation, and contour smoothing into a level set based energy minimization problem. In other words, the minimization of a single level set based energy functional deals with all the problems simultaneously. Due to the use of the level set, a level set based Euler-Lagrange equation can be derived directly from the energy functional, which makes the foreground-background segmentation and contour smoothing work in real-time. The simultaneous dealing of the background image modeling and the foreground segmentation makes it possible to obtain a clean background image with a relatively small number of frames. Due to the use of a small number of frames, a change in the scene is reflected fast in the formation of the reference background image, so that objects that starts(stops) moving are rapidly recognized as the foreground(background). Furthermore, the level set based contour smoothing eliminates spurious regions and noise to reduce the false alarm rate in surveillance systems, and smooths the contour that encompasses the moving object to obtain a better representation of the boundary of the object.

Besides being used as a partitioning operator that segments the foreground and the background, and as an auxiliary function used in the contour evolution, the level set function is also used as a weighting function in the modeling of the background image. The weighting is determined such that colors that are close to those in the background image are given a larger weight in the modeling of the next background image than colors that are not. This reduces erroneous segmentation results which can result from the uniform blending parameter that is used in temporal blending based background subtraction schemes[1]-[3].

2 Proposed Model

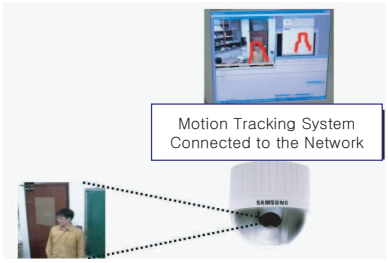
We introduce the following energy functional that formulates the problem of simultaneous background modeling, foreground segmentation and contour smoothing into a level set based minimization problem:

$$E(B,\phi) = \int_{\Delta t}\int_{\Omega} F(\phi)\left[\alpha - (B - I(t))^2\right] + \lambda|\nabla\phi|^2 \; d\mathbf{r}dt, \tag{1}$$

where B is the background image, ϕ is the level set function, $I(t)$ is the frame at time t, Ω is the domain of the image, Δt is a certain time interval along the temporal axis, α and λ are positive constant parameters, and $F(\phi)$ is a function of ϕ defined as follows:

$$F(\phi) = \begin{cases} k\frac{\phi-\phi_{min,\phi\geq 0}}{\phi_{max,\phi\geq 0}-\phi_{min,\phi\geq 0}} + (1-k), & \text{if } \phi \geq 0 \\ k\frac{\phi-\phi_{min,\phi<0}}{\phi_{max,\phi<0}-\phi_{min,\phi<0}}, & \text{if } \phi < 0 \end{cases} \tag{2}$$

(a)



(b)

Figure 1. (a) The main closed-circuit television(CCTV) control center in Gangnam district, Seoul, Korea. Security officers keep on watching hundreds of video channels displayed on monitor screens. (b) Intelligent surveillance system, that automatically obtains the contour of the intruding object.

where $\phi_{max,\phi\geq 0}$ and $\phi_{min,\phi\geq 0}$ are the maximum and the minimum values of $\phi(\mathbf{r})$ in $\{\mathbf{r}|\phi(\mathbf{r}) \geq 0\}$, respectively, and $\phi_{max,\phi<0}$ and $\phi_{min,\phi<0}$ are the maximum and the minimum values of $\phi(\mathbf{r})$ in $\{\mathbf{r}|\phi(\mathbf{r}) < 0\}$, respectively, and k is a constant lying in the interval $0 \leq k \leq 1$.

The energy functional is minimized with respect to the background image B and the level set function ϕ, which are the solutions being sought. The level set function is an auxiliary 3 dimensional function which domain has the size $[\Omega \times \Delta t]$. However, it is actually composed of several 2 dimensional frames, where each frame corresponds to each image frame in Δt.

The level set function plays several different roles in the scheme. First, the use of the level set function makes it possible to derive a level set based Euler-Lagrange equation from (1), so that a real-time working scheme can be implemented. Second, it acts as a partitioning operator which segments the image region into the foreground and the background region, where $\{\mathbf{r}|\phi(\mathbf{r}) \geq 0\}$ represents the foreground region and $\{\mathbf{r}|\phi(\mathbf{r}) < 0\}$, the background region. Third, the value of the level set function determines the value of $F(\phi)$ which acts as a weighting function that is used in the construction of the background image. Fourth, it is used in the smoothing of the contour that encompasses the moving object. Besides these, the level set function can also be used to obtain an adaptive thresholding used in the

segmentation of the foreground. In the following sections, it is shown how the minimization of (1) results in the background image modeling, foreground segmentation, and the contour smoothing.

3 Background Image Modeling

The equation for background image modeling is obtained by minimizing the energy functional in (1) with respect to the background image. The minimization is done directly by letting the gradient of the functional be zero. Keeping ϕ fixed and minimizing the energy $E(B, \phi)$ with respect to the background image B, gives:

$$\begin{aligned}
&\int_{\Delta t}\int_{\Omega} 2F(\phi)(B(\mathbf{r}) - I(\mathbf{r},t))d\mathbf{r}dt = 0 \\
\Leftrightarrow &\int_{\Delta t}\int_{\Omega} F(\phi)B(\mathbf{r})d\mathbf{r}dt = \int_{\Delta t}\int_{\Omega} F(\phi)I(\mathbf{r},t)d\mathbf{r}dt \\
\Leftrightarrow &\int_{\Omega} B(\mathbf{r})d\mathbf{r} = \frac{\int_{\Delta t}\int_{\Omega} F(\phi)I(\mathbf{r},t)d\mathbf{r}dt}{\int_{\Delta t} F(\phi)dt} \\
\Leftrightarrow &\int_{\Omega} B(\mathbf{r})d\mathbf{r} = \frac{\int_{\Omega}\int_{\Delta t} F(\phi)I(\mathbf{r},t)dtd\mathbf{r}}{\int_{\Delta t} F(\phi)dt}
\end{aligned}$$

One solution that satisfies the above equation is:

$$B(\mathbf{r}) = \frac{\int_{\Delta t} F(\phi)I(\mathbf{r},t)dt}{\int_{\Delta t} F(\phi)dt}. \quad (3)$$

We use (3) to construct the background image. It should be noticed that the integration in (3) is along the sequential axis, and not over the image domain. The brightness value of $B(\mathbf{r})$ for each pixel $\mathbf{r}$ is a weighted average of the brightness values $I(\mathbf{r}, t)$ at the same position $\mathbf{r}$ and different t in Δt, where $F(\phi)$ acts as the weighting function.

As can be observed from (2), $F(\phi)$ lies in the interval $0 \leq F(\phi) < k$, if ϕ is negative, while it lies in the interval $1-k \leq F(\phi) \leq 1$, if ϕ is positive. For example, if $k = 0.3$, then $0 \leq F(\phi) < 0.3$, if ϕ is negative, i.e., if the pixel $\mathbf{r}$ belongs to the foreground region, while $0.7 < F(\phi) \leq 1$, if the pixel $\mathbf{r}$ belongs to the background region. Therefore, if $k < 0.5$, current image intensities corresponding to pixels that are classified as the background region are taken more into account in the formation of the next background image than intensities of pixels that are classified as the foreground region. Normally, we use $k = 0.1$.

Furthermore, the weighting is adaptive based on the difference of the current image intensity and the background image intensity at the pixel $\mathbf{r}$. It is large at pixels where the difference of the intensity value of the background and the current image is small, and small at pixels where the difference is large. This is due to the fact that the value of $F(\phi)$ is large when ϕ is large, which again is large if $(B - I(t, \mathbf{r}))^2$ is small, and vice versa. Therefore, the weighted average is more weighted to the intensities in the current frame that are similar to the intensities of the current background image. In this way, the formation of the next background image becomes constrained by the former segmentation result and the intensity values of the former background image, and thus, is less affected by moving objects than conventional background modeling schemes. Therefore, moving objects leave less tails in the next background image even with a relatively small number of frames than conventional schemes.

On the other side, the use of a relatively small number of frames in the formation of the background image has the effect that the change of state of a certain object, i.e., the change from a static state to a moving state or vice versa, becomes reflected fast into the background image.

4 Foreground Segmentation

The classification of the background and the foreground region is determined by the parameter α which acts as a threshold value. The thresholding is done indirectly via the Euler-Lagrange equation relating the level set function. The Euler-Lagrange equation is obtained by keeping B fixed, and minimizing the energy functional in (1) with respect to ϕ.

Supposing that the background image B is fixed, and ignoring the regularization term($|\nabla\phi_{2D}|^2$) for a while, it can be seen from (1), that the integrand in (1) decreases as the value of $F(\phi)$ decreases, if the value $\alpha - (B - I(t, \mathbf{r}))^2$ at the pixel $\mathbf{r}$ and time t is positive. According to (2), the decrease of $F(\phi)$ indicates the decrease in the value of ϕ, and thus the minimization process makes ϕ go to $-\infty$. As a result, every pixel in the current frame at which the value $(B(\mathbf{r}) - I(\mathbf{r}))^2$ is smaller than α becomes classified in the region $\{\mathbf{r}|\phi(\mathbf{r}) < 0\}$, that is, the background region. Likewise, $\phi(\mathbf{r})$ becomes positive at the pixel $\mathbf{r}$ where $(B(\mathbf{r}) - I(\mathbf{r}))^2 \geq \alpha$, and thus, the pixel becomes classified in the region $\{\mathbf{r}|\phi(\mathbf{r}) \geq 0\}$, the foreground region.

The minimizer ϕ is in fact a 3 dimensional function in the domain $[\Omega \times \Delta t]$, and has to be computed using all the image frames in the time interval Δt. However, since B is being kept fixed, and the image frames in Δt are independent to each other, we find a 2 dimensional minimizer ϕ_{2D} function for each frame instead of the 3 dimensional minimizer ϕ function, and assume that the 3 dimensional minimizer is the stack of all ϕ_{2D} slices in Δt. Therefore, we formulate the problem as:

$$arg\ \min_{\phi_{2D}} \int_{\Delta t}\int_{\Omega} F(\phi_{2D})\left[\alpha - (B - I(t))^2\right] + \lambda|\nabla\phi_{2D}|^2 d\mathbf{r}dt, \quad (4)$$

for each frame in Δt. Then, the following Euler-Lagrange equation for ϕ_{2D} is deduced for each frame in Δt:

$$\frac{\partial\phi_{2D}}{\partial t} = F'(\phi_{2D})\left[(B - I(t))^2 - \alpha\right] + \lambda\nabla^2\phi_{2D}, \quad (5)$$

where ∇^2 is the laplacian operator and $F'(\phi_{2D})$ is the derivative of $F(\phi_{2D})$ with respect to ϕ_{2D}:

$$F'(\phi_{2D}) = \begin{cases} k\frac{1}{\phi_{max,\phi\geq 0}-\phi_{min,\phi\geq 0}}, & \text{if } \phi \geq 0 \\ k\frac{1}{\phi_{max,\phi<0}-\phi_{min,\phi<0}}, & \text{if } \phi < 0 \end{cases}$$

where $\phi_{max,\phi\geq 0}$, $\phi_{min,\phi\geq 0}$, $\phi_{max,\phi<0}$, and $\phi_{min,\phi<0}$ have been regarded as constants. The Euler-Lagrange equation is solved using the forward difference scheme with a prescribed number of iterations.

Here, $F'(\phi_{2D})$ acts as a normalizing function, that normalizes the value of $(B - I(t))^2 - \alpha$ with respect to the maximum and minimum values of ϕ in $\{\mathbf{r}|\phi(\mathbf{r}) \geq 0\}$ and $\{\mathbf{r}|\phi(\mathbf{r}) < 0\}$. By observing (5) and $F'(\phi_{2D})$, it can be seen that the same normalized result can be obtained regardless of the number of iterations, if the regularization term is omitted, and therefore, in this case, it is enough to solve (5) for one iteration, with the initial condition $\phi_{2D}(\mathbf{r}) = 0$ for all $\mathbf{r}$. Even with the regularization term, only a few iterations are required, since the normalization keeps the absolute value of $\phi_{2D}(\mathbf{r})$ small and since the laplacian operator smooths out $\phi_{2D}(\mathbf{r})$ very fast. The small number of iterations greatly saves the execution time.

5 Smoothing of the Contour and Removal of Spurious Regions

After the foreground segmentation via the ϕ function, the zero level contour of the ϕ function becomes the contour that represents the boundary of the foreground. A clean extraction of the boundary and the removal of spurious regions caused by noise are important for higher computer vision tasks, such as behavioral analysis, and also for the reduction of false alarm rate and memory allocation. To this end, some extra morphological post-processes such as opening and closing are used in conventional tracking schemes. However, with the proposed model, such a process is incorporated in the minimization of (1). The minimization of the regularization term($|\nabla\phi|^2$) of the integrand in (1) results in a smoothing of the level set function, which again results in the elimination of spurious regions and smoothing of the zero level contour.

6 Implementation of the Algorithm

Even though the algorithm uses several frames for the computation of the current reference background image and implements a partial differential equation, it can be executed in real-time. This is due to the fact that the number of iterations in implementing (5), normally, is less than 5 iterations, and the number of frames in Δt are small. The number of frames can be further reduced if the frame-rate decreases, that is, if a sampled version of the frames are taken from the video sequence. In the experiments, we usually used $5 \sim 10$ frames. The fundamental steps of the proposed algorithm are presented below.

Principle Steps of the Algorithm

1. At the initial step, an initial reference background image is constructed, e.g., by taking the average image of contiguous frames.
2. Using the initial reference background image computed in step 1, The ϕ_{2D} functions for every frame in the time interval Δt are computed using (5).
3. Compute the reference background image according to (3) using all the ϕ_{2D} functions in Δt.
4. Update Δt such that all the frames in Δt are shifted by one frame along the sequential axis.
5. Compute ϕ_{2D} functions for every frame in the time interval Δt using (5).
6. Repeat step 3–5.

7 Experimental Results

The video sequence which we used in the experiments contains moving objects in all frames. Figure 2 compares the segmented foreground images and the estimated background images obtained by different methods. The simple averaging, median, and the proposed method all use 10 sampled frames, which are sampled every second frame, i.e., have a frame rate of 15 fps. For Gaussian modeling based methods, 10 frames are too few and have been excluded from the comparison. As can be observed, the averaging method and the blending method leave "tails" of the moving objects in the background image, which again affects the foreground segmentation. The median method shows better results similar to the proposed algorithm, but the computational cost is larger due to the sorting process. In comparison, the proposed method leaves less "tails" than the averaging or blending methods, and due to the inherent smoothing of the contour, the boundary of the object is better defined as can be seen in Fig. 2(o). The processing time for each computation in Δt was about 0.04 seconds using frames of size 160×120. Figure 3 shows the enlarged figures of the contours in Fig. 2. It can be seen that the contour obtained with the proposed scheme is smooth and includes the object entirely. We computed the Dice coefficient for the object in Fig. 3 compared with a manually segmented region. Figure 4 compares the proposed method with the

blending method in the case that the moving object comes to a static state. It can be seen that the proposed scheme reflects the static object in the background image faster than the blending method. Figure 5 compares the case that the static object starts moving and becomes excluded from the background image. The proposed scheme obtains a "clean" background image faster than the blending method.

8 Acknowledgement

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund) (KRF-2008-331-D00566).

References

[1] J. Heikkila and O. Silven: A real-time system for monitoring of cyclists and pedestrians. Second IEEE Workshop on Visual Surveillance Fort Collins, Colorado (Jun. 1999) 74–81

[2] C. Wren, A. Azabayejani, T. Darrell and A. Pentland: Pfinder: Real-time tracking of the human body. IEEE Trans. on Pattern Analysis and Machine Intelligence. **19** (1997) 780–785

[3] G. Halevy and D.Weinshall: Motion of disturbances: detection and tracking of multibody non-rigid motion Machine Vision and Applications. **11** (1999) 122–137

[4] C. Stauffer and W.E.L. Grimson: Learning patterns of activity using real time tracking. IEEE Trans. on Pattern Analysis and Machine Intelligence. **22** (2000) 747–767

[5] X. Gao, T.E. Boult, F. Coetzee, and V. Ramesh: Error Analysis of Background Subtraction. In IEEE Int. Conf. on Computer Vision (2000)

[6] L. Li, W. Huang, I.Y.H. Gu, Q. Tian "Foreground object detection from videos containing complex background", ACM Multimedia, (2003)

[7] P. Kornprobst and R. Deriche: Image Sequence Analysis via Partial Differential Equations. Journal of Mathematical Imaging and Vision. **11** (1999) 5–26

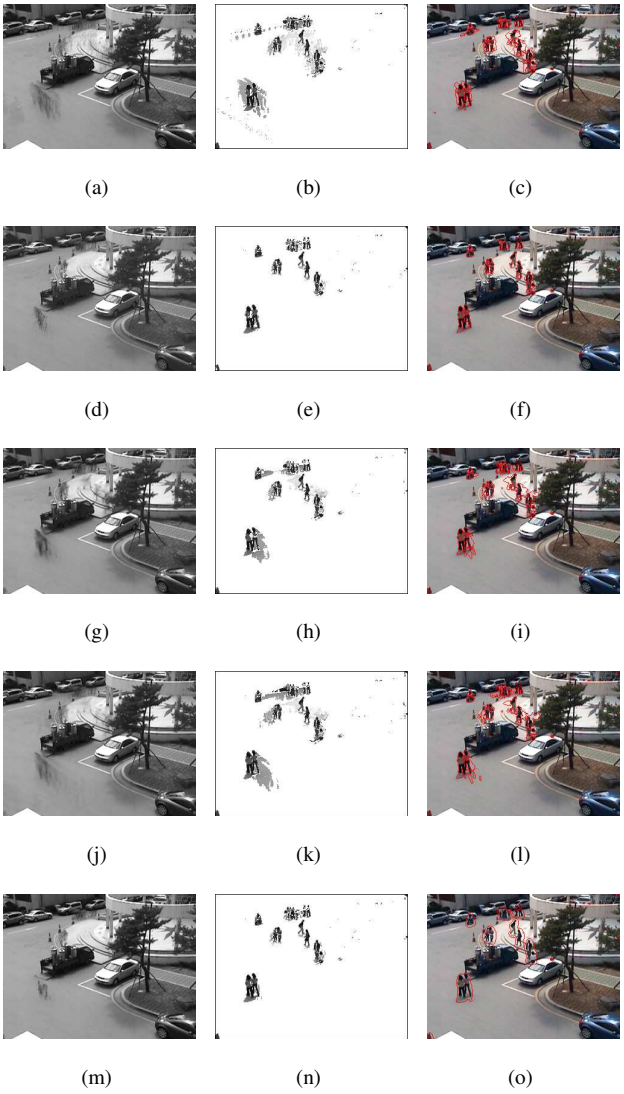


Figure 2. The first column shows the estimated background images obtained by (a) averaging (d) median (g) blending with blending parameter 0.1 (j) blending with blending parameter 0.05 (m) proposed method. The second column shows the corresponding segmented foreground, where the foregrounds in (b),(e),(h) and (n) have been obtained by thresholding with threshold value of 30, and the foreground in (j) has been obtained by the proposed method with $\alpha = 30$. The third column shows the corresponding zero level contours in red colors.

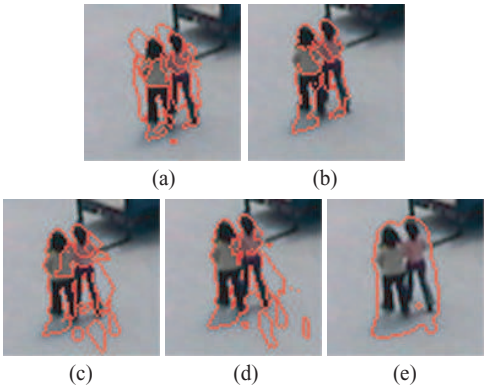


Figure 3. Enlarged figures of the contours in Fig. 3. (a) averaging $d_v = 0.75$ (b) median $d_v = 0.87$ (c) blending with blending parameter 0.1 $d_v = 0.79$(d) blending with blending parameter 0.05 $d_v = 0.82$(e) proposed method $d_v = 0.88$

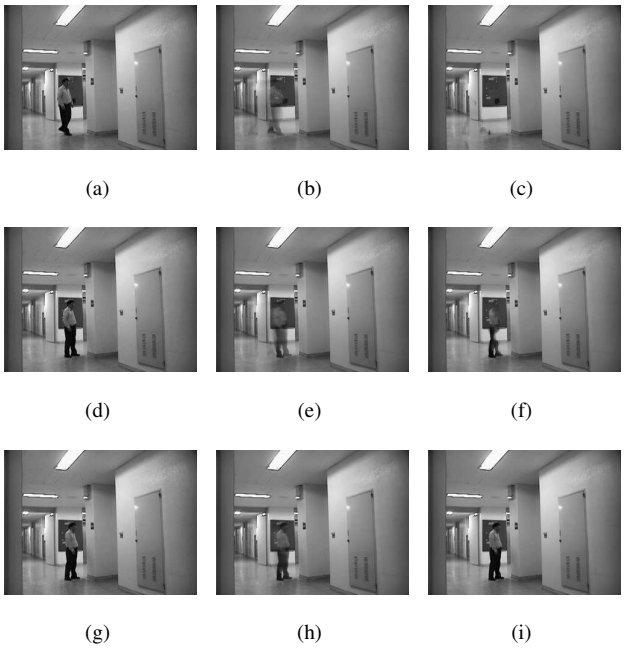


Figure 4. Comparison of the estimated background image when the moving object comes to a static state. The first column shows the frame image, the second shows the estimated background image obtained by the blending method, and the third shows that obtained by the proposed method. The first, second, and the third row correspond to the frame 226, 235, and 243, respectively.

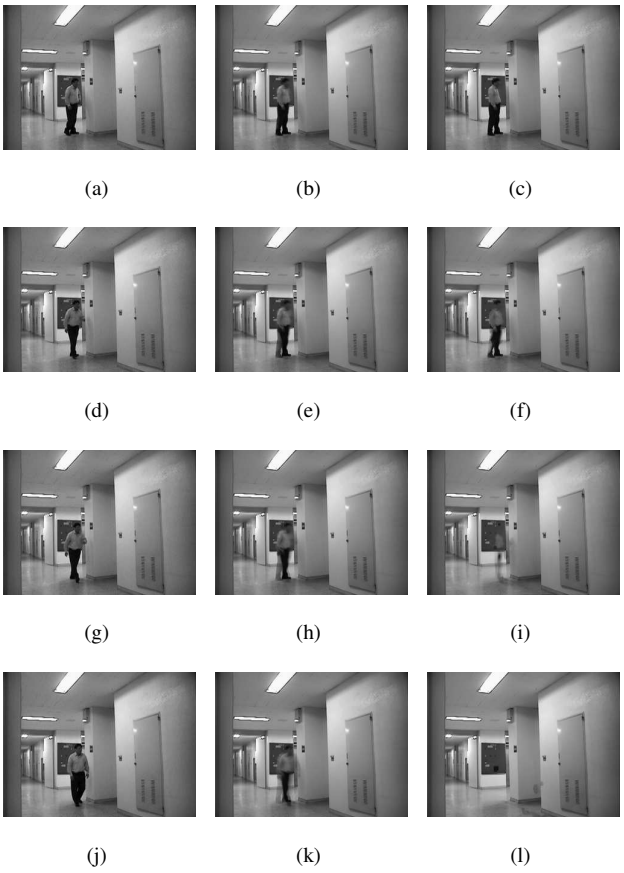


Figure 5. Comparison of the estimated background image when the static object starts moving. The first column shows the frame image, the second shows the estimated background image obtained by the blending method, and the third shows that obtained by the proposed method. The first, second, third, and the fourth rows correspond to the frame 313, 316, 321, and 325, respectively.