AUTOMATIC MOTION DETECTION USING FALSE BACKGROUND ELIMINATION

JIN KEUN SEO1 AND SUKHO LEE2

1 DEPARTMENT OF MATHEMATICS, YONSEI UNIVERSITY, SOUTH KOREA
E-mail address: seoj@yonsei.ac.kr

2 DIVISION OF COMPUTER INFORMATION ENGINEERING, DONGSEO UNIVERSITY, SOUTH KOREA
E-mail address: petrasuk@gmail.com

ABSTRACT. This work deals with automatic motion detection for surveillance tracking that aims to provide high-lighting movable objects which is discriminated from moving backgrounds such as moving trees, etc. For this aim, we perform a false background region detection together with an initial foreground detection. The false background detection detects the moving backgrounds, which become eliminated from the initial foreground detection. This false background detection is done by performing the bimodal segmentation on a deformed image, which is constructed using the information of the dominant colors in the background.

1. INTRODUCTION

Automatic motion detection is an important low-level computer vision algorithm which can be applied to automatic surveillance systems [1]-[3]. Nowadays, the need for automatic surveillance system is high, since a security guard watching many video channels may easily miss important events such as events which need immediate emergency assistance or crime prevention in a secluded area. A motion detection scheme should be robust against false alarms, which may happen due to the movement of objects in the background such as moving trees, or noise. Furthermore, due to low SNR (signal to noisy ratio) of night videos, conventional background subtraction method may not be robust against missing targets and false alarms in video security surveillance. Therefore, it is a crucial to develop a motion detection scheme which works in real-time and is robust against false regions.

This work deals with an automatic motion detection algorithm that aims to provide a high-lighting of movable objects. Normally, motion detection with stationary cameras are performed by background differencing techniques. Background differencing techniques are known as the most common and effective approach to segment the foreground image for stationary video camera. There are numerous variation of background differencing including background subtraction between current frame and background image, single or mixture of...
Gaussian background model [4], and so on. These background differencing techniques usually are combined with morphological operations such as dilation and erosion to deal with noisy environment, especially for night surveillance video. However, these morphological operation are not effective to reduce false positives and false negatives in the foreground detection stage. In this paper, we propose a segmentation method which is based on the bimodal segmentation together with a Laplacian smoothing term which combines automatically adaptive thresholding and morphological operation at the same time. Our method provides a novel real time motion detection algorithm.

NOMENCLATURE

\[ D: \text{The difference between the present and the average of past images} \]
\[ \Omega: \text{The domain of the given image } u_0 \]
\[ \phi: \text{The level set function} \]
\[ \Psi[\phi](r): \text{The integrand of the energy functional} \]
\[ D\phi: \text{The distributional gradient} \]
\[ E(\phi): \text{The energy functional with respect to the level set function} \]
\[ H(\cdot): \text{The Heaviside function} \]
\[ u_0: \text{The given image} \]

2. PREVIOUS WORK: LEVEL SET BASED BIMODAL SEGMENTATION

In [5], we have proposed the following energy functional:

\[
E(\phi) = +\lambda_1 \int_{\Omega} |u_0(r) - ave_{\{\phi \geq 0\}}|^2 \phi(r) H(\alpha + \phi(r)) \, dr \\
- \lambda_2 \int_{\Omega} |u_0(r) - ave_{\{\phi < 0\}}|^2 \phi(r) H(\alpha - \phi(r)) \, dr
\]

(2.1)

where \( \alpha \) is an arbitrary small positive value, and \( \Omega \) is the domain of the given image \( u_0 \). The minimization of this energy functional with regard to \( \phi \) results in a \( \phi \) function of which the zero level set becomes the contour that separates the object from the background. \( H(\cdot) \) is the Heaviside function which is defined as follows:

\[
H(\phi) = \begin{cases} 
1 & \text{if } \phi \geq 0 \\
0 & \text{if } \phi < 0 
\end{cases}
\]

(2.2)

The minimization is performed over the set of continuous functions \( \phi \) with \( \int_{\Omega} |D\phi| < M \) for a fixed number \( M \). Here, \( D\phi \) is the distributional gradient.

We explain the effect of minimizing the proposed functional for the following two regions: \( D_{|\phi| > \alpha} = \{ r \in \Omega : |\phi(r)| > \alpha \} \) and \( D_{|\phi| \leq \alpha} = \{ r \in \Omega : |\phi(r)| \leq \alpha \} \). For simplicity, and without loss of generalization, we let \( \lambda_1 = \lambda_2 = 1 \) afterward. We denote by \( \Psi[\phi](r) \) the integrand of the energy functional:

\[
\Psi[\phi](r) := +|u_0(r) - ave_{\{\phi \geq 0\}}|^2 \phi(r) H(\alpha + \phi) \\
-|u_0(r) - ave_{\{\phi < 0\}}|^2 \phi(r) H(\alpha - \phi),
\]

(2.3)

and observe the fact that in \( \Psi[\phi](r) \) the values of \( |u_0(r) - ave_{\{\phi \geq 0\}}|^2 \) and \( |u_0(r) - ave_{\{\phi < 0\}}|^2 \) depend only on the zero level set, while the values of \( \phi(r)H(\alpha + \phi) \) and \( \phi(r)H(\alpha - \phi) \) are controlled by all the level sets.

Now, consider the case \( r \in D_{|\phi| > \alpha} \). If \( \phi > \alpha \), then \( \Psi[\phi](r) = +|u_0(r) - ave_{\{\phi \geq 0\}}|^2 \phi(r) \), and therefore, \( \Psi[\phi](r) \) decreases as \( \phi(r) \) decreases, until \( \phi(r) \) reaches the boundary
Figure 1. Image adaptive segmentation results of images which have different intensities with bimodal segmentation scheme. (d)-(f) are the corresponding $\phi$ functions to (a)-(c) with $\alpha = 8$. ($\alpha$ can be any positive value) (h) is the segmentation result of the LSBS performed on (d).

$\alpha$. Here, $\alpha$ acts as a boundary across which $\Psi[\phi](r)$ increases abruptly if $\phi(r) = \alpha_- \rightarrow \alpha_+$. Otherwise, if $\phi < -\alpha$, then $\Psi[\phi](r)$ decreases as $\phi(r)$ increases and reaches the boundary $\phi(r) = -\alpha$. Therefore, every $\phi(r)$ with magnitude value $|\phi(r)| > \alpha$ experiences a change such that $|\phi(r)|$ decreases, and is forced to reside between $-\alpha \leq \phi(r) \leq \alpha$, i.e., every region $D_{|\phi| > \alpha}$ converts to $D_{|\phi| \leq \alpha}$.

Since the segmentation result depends on the competition of the colors, that is, on the relative relationship of the colors, the segmentation result is adaptive to the image. Figure 1. shows segmentation results on several gray-level images having different intensity values. Even though the intensity values vary for each image, they all have the same $\phi$ function in the steady state, and therefore, the segmentation results are the same.

The adaptive property of the bimodal segmentation makes it useful for motion detection. As mentioned in the introduction, the object is normally segmented by segmenting the background subtracted frame, where the threshold value varies from scene to scene. However, the bimodal segmentation algorithm segments the object without thresholding, and therefore the segmentation process becomes adaptive to the image and automatical, since there is no need to set an a priori known threshold value.

3. Proposed Scheme

For the surveillance tracking, the background subtraction method is the most effective tool in general. However, the direct application of the background subtraction method to motion detection is not appropriate in usual noisy environments such as changing illumination, raining, shadow, moving trees, etc. To deal with this noisy environment, we develop a novel segmentation based on a special region based segmentation using a deformed image $\hat{u}$ which makes the target object region to appear more apparent. The choice of the deformed image $\hat{u}$ may differ in each individual environment, and it is used to muffle various uncorrelated background scenes effectively. We introduce the deformed image $\hat{u}$ later, and first give some definitions which will be used throughout the paper.
Figure 2. Video sequence used for experiments. The moving object (car) is moving in the upper part of the image, while the shrub is also moving due to the wind. (a) Current frame (b) Background image. (c) Absolute difference image of (a) and (b)

3.1. Definitions and notations. To be precise, we fix some notations and introduce definitions. Let \( u(i, j, t) \) be a given 2D-grey scale image at the pixel \((i, j)\) at time \(t\) in the video sequence obtained by a surveillance camera. We assume that the domain of \( u \) is the set \( \Omega := \{(i, j) \in \mathbb{Z} \times \mathbb{Z} : 0 \leq i, j \leq N\} \), where \( \mathbb{Z} \) is the set of integers, and the range of \( u \) is contained in the set \( \{k \in \mathbb{Z} : 0 \leq k \leq M\} \). The histogram of \( u \), denoted by \( \tilde{u} \), is given by \( \tilde{u}(k, t) = \# \{(i, j) \in \Omega : u(i, j, t) = k\} \), which is the number of the set. Let \( D_u(i, j, t) = u(i, j, t) - \frac{1}{n} \sum_{k=1}^{n} u(i, j, t - k\Delta t) \) be the difference between the present image and the average of a few past images, where \( \Delta t \) is the time interval between consecutive frames. Formally, we can let \( D_u(i, j, t) = I(i, j, t) - B(i, j, t) \), where \( I(i, j, t) \) is the given grey image at the pixel \((i, j)\) at time \(t\). The corresponding background image, denoted by \( B(i, j, t) \), can be chosen as \( B(i, j, t + 1) = \alpha u(i, j, t) + (1 - \alpha)B(i, j, t) \), where \( \alpha \) is the update rate which takes account of slow illumination changes. Normally, \( 0 < \alpha < 0.5 \).

Here, \( n \) and \( m \) should be small integers to ease off the memory requirements. For a given variance \( \sigma \) and \( 0 < \theta < 1 \), we define the most dominant color \( \lambda_1^\sigma(t) \) by

\[
\lambda_1^\sigma(t) = \arg \sup_{\mu} \sum_{|k - \mu| < \sigma, |k - \lambda_1^\sigma(t)| > \sigma} \frac{h_u(k, t)}{|k - \mu|^2}
\]

Similarly, we can define the second dominant color \( \lambda_2^\sigma(t) \) by

\[
\lambda_2^\sigma(t) = \arg \sup_{\mu} \sum_{|k - \mu| < \sigma, \max(|k - \lambda_1^\sigma(t)|, |k - \lambda_2^\sigma(t)|) > \sigma} \frac{h_u(k, t)}{|k - \mu|^2}
\]

and the third dominant color \( \lambda_3^\sigma(t) \) as

\[
\lambda_3^\sigma(t) = \arg \sup_{\mu} \sum_{|k - \mu| < \sigma, \max(|k - \lambda_1^\sigma(t)|, |k - \lambda_2^\sigma(t)|, |k - \lambda_3^\sigma(t)|) > \sigma} \frac{h_u(k, t)}{|k - \mu|^2}
\]

and so on for \( \lambda_4^\sigma(t), \ldots \).

3.2. Detection of the Foreground. For the initial foreground detection, we use the time difference image \( D_u \) and the level set function \( \phi \). For the use of PDE (partial differential equation) based techniques, we often pretend that \( u \) and \( D_u \) are the corresponding continuous functions defined on the domain \( \Omega \). The proposed motion detection combines two different segmentations to reduce false alarm rates. In this section, the notation \( t \) will be
used for two different purposes but there will be no confusion from the context.

The initial foreground detection is performed by minimizing the following energy functional

\[ E(\phi) = \lambda_1 \int |D_u(r) - \text{ave}_{\phi \geq 0}|^2 \phi(r) H(\alpha + \phi) \, dr - \lambda_2 \int |D_u(r) - \text{ave}_{\phi < 0}|^2 \phi(r) H(\alpha - \phi) \, dr + \text{regularizion} \quad (3.1) \]

where \( \alpha \) is an arbitrary small positive value, \( H \) the Heaviside function, and \( \text{ave}_{\phi \geq 0} \) the average of \( D_u \) over the region \( \{\phi > 0\} \). The minimization of the above energy functional results in the region \( R_1 := \{\phi_1 > 0\} \), which we use as the initial foreground region. We compute a minimizer \( \phi_1 \) of (3.1) by solving the following parabolic equation:

\[ \frac{\partial \phi}{\partial t}(i, j) = S\phi(i, j) + \lambda_1 |D_u - \text{ave}_{\phi \geq 0}|^2 [H(\alpha + \phi) + \phi H'(\alpha + \phi)] - \lambda_2 |D_u - \text{ave}_{\phi < 0}|^2 [H(\alpha - \phi) - \phi H'(\alpha - \phi)] \quad (3.2) \]

where \( S\phi \) is used to eliminate false segmented region due to the noise. For fast computation, \( S\phi \) should eliminate false regions fast. The proposed \( S\phi \) is

\[ S\phi(i, j) = \begin{cases} \Delta H(\phi) & \text{if } |\Delta H(\phi)| \geq \frac{3}{4} \\ 0 & \text{otherwise} \end{cases} \quad (3.3) \]

where \( \Delta H(\phi) \) is the standard discrete Laplacian of \( H(\phi) \):

\[ \Delta H(\phi)(i, j) = \frac{1}{4} \left\{ H(\phi)(i + 1, j) + H(\phi)(i, j + 1) + H(\phi)(i - 1, j) + H(\phi)(i, j - 1) - 4H(\phi)(i, j) \right\}. \]

Note that \( \Delta H(\phi) \in \{0, \pm \frac{1}{4}, \pm \frac{1}{2}, \pm 1\} \), and having \( |\Delta H(\phi)(i, j)| \geq \frac{3}{4} \) are out of harmony, so they are considered as a false regions to be eliminated. The above steady state solution can be obtained through a few iteration steps, and therefore works in real-time.

However, there exists a trade-off between the elimination of false regions and detection of small-sized moving objects. Eliminating false regions too fast may also eliminate small-sized moving objects, or erode some of the moving object region. Therefore, the effect of \( S\phi \) should not be too large. In this paper, we want to eliminate false regions by the second segmentation explained next, rather than by the regularization term \( S\phi \). Therefore, we set \( \lambda_1 \) and \( \lambda_2 \) rather large to emphasize the effect of the bimodal segmentation.

Figure 2(a) shows a frame from a video sequence in which a car is moving (in the upper
part of the image) and a moving shrub exists in the background. Figure 2(b) shows the background image. Figure 2(c) is the result of direct background subtraction. Due to the movement of the shrub, many false regions appear in the image. Since there are many false regions in the difference image, a direct bimodal segmentation will also result in an image with many false regions even with the use of the regularization term.

3.3. **Detection of False Regions.** To detect false regions, we first construct a deformed image \( \hat{u} \) that is given as below. We fix \( 0 < \theta < 1 \) indicating the portion of image domain.

\[
\hat{u}(i, j, t) := \beta_1 \exp \left[ \sum_{m=1}^{m_0} \frac{-\beta_2}{|u(i, j, t) - \lambda_m(t)|^q} \right] \quad (\beta_1, \beta_2: \text{scaling parameters}),
\]

where \( m_0 \) is the largest \( m \) satisfying \( \sum_{|k - \lambda_m(t)| < \sigma} \frac{h_u(k, t)}{|k - \lambda_m(t)|^2} > \theta N^2 \), where \( N^2 \) is the size of the image. The deformed image \( \hat{u} \) has a large value at pixels which have brightness values similar to the dominant colors \( \lambda_m(t) \). Since the dominant colors often correspond to the colors of the background region, \( \hat{u} \) will indicate the background region. The segmented domain \( R_2 = \{ \phi_2 > 0 \} \) is obtained by

\[
\frac{\partial \phi}{\partial t}(i, j) = S(\phi)(i, j) + \lambda_1 \left[ \hat{u} - \overline{\Delta e \{ \phi \geq 0 \}} \right]^2 \left[ H(\alpha + \phi) + \phi H'(\alpha + \phi) \right] - \lambda_2 \left[ \hat{u} - \overline{\Delta e \{ \phi < 0 \}} \right]^2 \left[ H(\alpha - \phi) - \phi H'(\alpha - \phi) \right]
\]
where $\overline{\text{ave}}_{\{\phi \geq 0\}}$ is the average of $\hat{u}$ over the region $\{\phi \geq 0\}$. This is also a real time segmentation. This method is quite effective for indoor surveillance systems since it can advantage of various a priori information. This will extract the background region including moving backgrounds. Figure 3 shows the histogram of the image in Fig. 2(a). The three dominant brightness values are 85, 102, and 128. Figure 4(a)-(c) show the deformed images which are constructed using only one of the dominant colors, respectively. Figure 4(d) shows the image constructed using all the three dominant colors. Figure 5(b) shows the segmented region $R_2 = \{\phi_2 > 0\}$, which is obtained by taking the segmentation scheme on the image Figure 4(d). It can be observed that the segmented regions correspond to the moving regions of the shrub.

3.4. **Final Detected Motion Region.** After $R_2 = \{\phi_2 > 0\}$ is obtained, the final segmentation region is obtained from $R = R_1 \cap \hat{R}_2 = \{r : \phi_1(r,t) > 0, \phi_2(r,t) < 0\}$. Here, $\hat{R}_2$
is the complementary set of $R_2$. This corresponds to the region which can be obtained also by subtracting the false segmented foreground region $R_2$ from $R_1$. Figure 5(a) shows the region $R_1$ which is obtained by the initial foreground detection. Even with the Laplacian regularization term, many false regions can be observed which is due to the movement of the bush. Figure 5(b) shows the region $\hat{R}_2$. Figure 5(c) shows the $R = R_1 \setminus \hat{R}_2$ where it can be observed that moving regions in the background are well eliminated while the moving car region remains well detected. Figure 6 shows the experimental results with three more video sequences. The figures are in color, and the difference with the background image has been calculated by taking the difference in the color space. As can be observed, the mere bimodal segmentation detects false foreground regions as the background intensity is changing. However, with the proposed method the false foreground regions are almost eliminated, as the proposed method eliminates the false regions.

ACKNOWLEDGMENTS

This work was supported by the “Dongseo Frontier Project” Research Fund of 2011 by Dongseo University.

REFERENCES


