Real-time detection of moving objects in a video sequence by using data fusion algorithm

Chao Tang¹, Huosheng Hu², Miaohui Zhang³, Wen-Jian Wang⁴, Xiao-Feng Wang¹, Feng Cao⁴ and Wei Li⁵

Abstract
The moving object detection and tracking technology has been widely deployed in visual surveillance for security, which is, however, an extremely challenge to achieve real-time performance owing to environmental noise, background complexity and illumination variation. This paper proposes a novel data fusion approach to attack this problem, which combines an entropy-based Canny (EC) operator with the local and global optical flow (LGOF) method, namely EC-LGOF. Its operation contains four steps. The EC operator firstly computes the contour of moving objects in a video sequence, and the LGOF method then establishes the motion vector field. Thirdly, the minimum error threshold selection (METS) method is employed to distinguish the moving object from the background. Finally, edge information fuses temporal information concerning the optic flow to label the moving objects. Experiments are conducted and the results are given to show the feasibility and effectiveness of the proposed method.

Keywords
Moving object detection, edge detection, optic flow, Canny operation, security surveillance

Introduction
Moving object detection (MOD) is a preliminary phase in video analysis and has many real-world applications, such as video surveillance (Kaliraj and Manimaran, 2016; Ojha and Sakhare, 2015; Sun et al., 2016a; Verma et al., 2015), traffic monitoring (Chen and Huang, 2014; Hu et al., 2014; Huang and Chen, 2013; Lee et al., 2015; Shukla and Saini, 2015), video registration (Yang et al., 2017) and animal/people tracking (Huang et al., 2013; Jalal et al., 2017; Mirzamohammad et al., 2014; Siebel and Maybank, 2002; Zhang et al., 2014). Despite much effort in the past decade, challenging situations encountered in real-life environment (e.g. dynamic backgrounds, irregular object movements, low contrast, high sensor noise, camouflage, shelters, bad weather, and so on) make the design of a MOD algorithm that is robust under a wide variety of scenes remain to be an open problem (Hu et al., 2017). MOD in image sequences aims at detecting regions corresponding to moving objects such as vehicles and humans. Detecting moving regions provides a focus of attention for later processes such as tracking and behavior analysis because only these regions need be considered in the later processes. At present, most segmentation methods use either temporal or spatial information in the image sequence (Hu et al., 2004). In general, the methods of MOD can be categorized into three approaches: temporal differencing, background subtraction and optical flow.

The temporal differencing methodology is based on frame difference, which uses the differences between consecutive frames in a video sequence to detect moving regions (Ojha and Sakhare, 2015). Paul et al. (2017) proposed using a fixed number of alternate frames centered around a given frame in time instead of generating difference images using the traditional continuous frame difference approach. This approach helps reduce computational complexity without compromising on the quality of difference images. Al-Smadi et al. (2016) used a dynamic threshold value to estimate the global variance of the motion-accumulated variations of pixel intensity, based on the standard deviation of cumulative frame differencing (CFD). Radzi et al. (2014) proposed a technique to extract moving objects based on temporal differencing, ghost removal and shadow removal using normalized cross correlation (NCC) while using a non-static Pan-Tilt-Zoom (PTZ) camera.

¹Department of Computer Science and Technology, Hefei University, China
²School of Computer Science and Electronic Engineering, University of Essex, UK
³School of Computer and Information Technology, Shanxi University, China
⁴School of Computer and Information Engineering, Xiamen University of Technology, China

Corresponding author:
Miaohui Zhang, Institute of Energy, Jiangxi Academy of Sciences, Changdong Road 7777, Nanchang, 330096, China.
Email: zhangmiaohui1987@163.com
Background subtraction methods calculate the difference image $D(x, y)$ between an image $I$, and its corresponding background model $I_{\text{back}}$ to detect moving objects. Anandhalli and Baligar (2015) proposed an improvised background subtraction model for real-time tracking that solves the problems of shadow detection. Alvar et al. (2014) proposed an efficient model for background subtraction that processes only some pixels of each image. This model could result in a significant reduction in computation time for subsequent image analysis. In order to improve the performance of background subtraction algorithms, Sun et al. (2016b) integrated it with frame difference algorithms in order to effectively suppress noise and the ‘hole’. Heikilä and Pietikäinen (2006) present a novel and efficient texture-based method for modeling the background and detecting moving objects from a video sequence. Each pixel is modeled as a group of adaptive local binary pattern histograms that are calculated over a circular region around the pixel. Stauffer and Grimson (2000) proposed an adaptive background subtraction method that models each pixel as a mixture of Gaussians and uses an online approximant to update the model. The Gaussian distributions are then evaluated to determine which are most likely to result from a background process. This yields a stable, real-time outdoor tracker that reliably deals with lighting changes, repetitive motions from clutter, and long-term scene changes.

The optical flow method computes an optical flow image area based on the optical characteristics of a flow distribution image (Ramasubramanian et al., 2015). Sengar and Mukhopadhyay (2016) employed the normalized self-adaptive optical flow (NSOF) for efficient moving object area detection in a video sequence. The NSOF first performed smoothing on each frame of a given video data using a Gaussian filter, and then determined the optical flow field through an optical flow algorithm; it filtered out noise using the adaptive threshold approach, following which normalization, morphological operations and a self-adaptive window were finally applied to identify the areas occupied by the moving object.

Xin et al. (2014) proposed another self-adaptive optical flow method in. The method first estimated the original optical flow field using an optical flow algorithm, then enhanced the objects by a local mean algorithm, and finally filters out noise through a self-adaptive threshold algorithm. In Lu et al. (2010), a real-time motion detection algorithm was proposed based on the integration of optical flow and the inter-frame block-based histogram correlation method to achieve better performance. Zhou and Zhang (2005) introduced a novel approach to detect moving objects in a noisy background. Their approach combines a modified adaptive Gaussian mixture model (GMM) for background subtraction and optical flow methods supported by temporal differencing in order to achieve robust and accurate extraction of the shapes of moving objects.

In general, the frame difference method is less computationally complex and easier to implement than both background subtraction and optical flow methods, but cannot effectively extract the complete shapes of certain types of moving objects. The background subtraction method can achieve high accuracy and fast detection speed, but is difficult to implement owing to the complexity of the actual scene as well as a variety of noise and interference. The optical flow method can detect a moving object even when the camera moves, but requires high computational power and is very sensitive to noise. It is clear that a combination of these methods could overcome the limitations of individual methods.

In this paper, an entropy-based Canny (EC) operator is combined with the local and global optical flow (LGOF) method to solve the problems mentioned above. The proposed approach firstly uses an EC operator to acquire the boundary information of moving objects, and then obtain the motion vector field of the moving objects. Information from both processes is then fused for segmentation of moving objects.

The rest of this paper is organized as follows. Section 2 describes the proposed EC-LGOF method, including an edge detection algorithm, the optical flow algorithm, the binary process, data fusion and morphologic operation. Experimental results are presented in Section 3 to show the feasibility and performance of the proposed method. Finally, a brief conclusion and future extension are given in Section 4.

**Proposed MOD method**

**Overview of system**

Figure 1 shows the procedure of the EC-LGOF method proposed in this research. As can be seen, it consists of edge detection, optical flow, data fusion and morphologic operation. The EC Canny operator is used in edge detection for simplicity and effectiveness. The LGOF method can be used to determine optical flow, and quickly provide a dense optical flow vector of moving objects. The binary process employs the minimum error-threshold-selection method, and determines the threshold used to distinguish the background from the moving object in a self-adaptive manner. Owing to noise, the optical flow method cannot detect the accurate boundaries of moving objects. The edge detection algorithm mentioned above can solve this problem.

Moreover, the edge image acquired by the EC operator can be considered as a space gradient and the optical flow image as a time gradient. Combining space gradient information with time gradient information can yield more accurate information of moving objects. During the data fusion process, the AND operator is used between an edge binary image and an optical flow binary image. In order to obtain the exact contours of a moving object, morphologic operations such as Close and Hole Filling are implemented. Finally, moving objects are extracted from the image.

**EC operator**

In general, edges in an image contain important information that reflect its structural attributes, and edge detection is the first step in image processing. Of the various edge detection algorithms available, Canny edge detection algorithm, developed in 1986, has been widely used in research and engineering projects because of its impressive performance. However, it has certain defects, and especially cannot adaptively set a threshold. In Wang and Li (2015), an EC edge detection
algorithm, which combines the maximum entropy method with Otsu’s method, was proposed to determine high and low thresholds for the Canny edge detection algorithm. The EC algorithm consists of four steps:

Step 1: Noise elimination:

Let \( f(x,y) \) denote the input image and \( g(x,y) \) denote the Gaussian function

\[
g(x,y) = e^{-\frac{x^2 + y^2}{2\sigma^2}}
\]

where \( \sigma \) is the standard deviation of the Gaussian function, and controls the degree of smoothness. The image following smoothing is given by

\[
m(x,y) = \frac{g(x,y)f(x,y)}{-} \quad \text{(2)}
\]

Step 2: Calculation of magnitude and direction of gradient:

The magnitude and direction of the gradient are then computed as follows

\[
n(x,y) = \sqrt{\left(\frac{\partial m}{\partial x}\right)^2 + \left(\frac{\partial m}{\partial y}\right)^2} \quad \text{(3)}
\]

\[
\theta(x,y) = \arctan\left(\frac{\partial m}{\partial x} / \frac{\partial m}{\partial y}\right) \quad \text{(4)}
\]

where \( n(x,y) \) and \( \theta(x,y) \) are arrays, each of which are identical in size to the image.

Step 3: Non-maximum suppression (NMS):

Considering that there is only one response to a single edge, we need to refine the value of the image gradient to satisfy the third criterion of Canny edge detection. For each pixel with non-zero gradient strength, two adjacent pixels are searched along its gradient direction. If the gradient intensity value is less than the corresponding value of adjacent pixels in its gradient direction, then this cannot be the edge point, and then set its edge strength to 0; if not, it is chosen to be candidate edge point (Li and Ding, 2011).

Step 4: Connecting edges with dual threshold:

As mentioned above, Canny edge detection algorithm cannot adaptively set its threshold. In Wang and Li (2015), the maximum entropy method was combined with Otsu’s method to determine high and low thresholds for the Canny edge detection algorithm. The maximum entropy method uses entropy, an important concept in information theory pertaining to image segmentation. In the Canny algorithm, we need to provide two thresholds to obtain the edge. Select a low threshold \( k \) and a high threshold \( m \) in the gray-level set \([0, L-1]\). The image is divided into three categories, \( C_0, C_1 \) and \( C_2 \). \( C_0 \) consists of non-edge pixels, \( C_1 \) is a set of pixels that are not necessarily edges, and \( C_2 \) is the set of edge pixels. \( C_0 \) consists of pixels whose gray values are in the set \([0, k]\), \( C_1 \) of pixels with values in the set \([k + 1, m]\), and \( C_2 \) consists of pixels whose values are in \([m + 1, L-1]\). The entropies of \( C_0, C_1 \) and \( C_2 \) are defined as follows

\[
\begin{align*}
E_0 &= -\sum_{i=0}^{k} p_i \log p_i \\
E_1 &= -\sum_{i=k+1}^{m} p_i \log p_i \\
E_2 &= -\sum_{i=m+1}^{L-1} p_i \log p_i
\end{align*} \quad \text{(5)}
\]

where \( p_i \) is probability of gray level \( I \) and defined as follows

\[
p_i = \frac{n_i}{N} \quad (i = 0, 1, \ldots, L-1) \quad \text{(6)}
\]

The entropy of the entire image is

\[
E_k = -\sum_{i=0}^{L-1} p_i \log p_i \quad \text{(7)}
\]
We can calculate inter-cluster entropy instead of intra-cluster variance

$$\phi = \frac{(E_0 - E_1)^2 g(E_1 - E_2)^2}{((E_0 - E_1)^2 + (E_1 - E_2)^2)^2}$$ (8)

Using equation (7), we can obtain two thresholds for gray levels $k$ and $m$. $k$ represents the low threshold and $m$ the high threshold.

**The local and global optical flow method**

Differential methods are the most widely used technique for optic flow computation in image sequences. They can be classified into local and global methods. Local methods are often more robust against noise whereas global techniques yield dense flow fields.

Let us consider an image sequence $I(x, y, t)$, where $(x, y)$ denotes location within a rectangular image domain, and $t \in [0, T]$ denotes time. It is common to smoothen the image sequence prior to differentiation by convolving each frame with some Gaussian $K_\sigma(x, y)$ of standard deviation $\sigma$

$$f(x, y, t) := K_\sigma * I(x, y, t)$$ (9)

Many differential methods for optical flow are based on the assumption that the gray image values of objects in subsequent frames do not change over time

$$f(x + u, y + v, t + 1) = f(x, y, t)$$ (10)

where the displacement field $(u, v)^T(x, y, t)$ is called the optical flow. For small displacements, we may perform the 1st-order Taylor expansion to yield the optical flow constraint

$$f_x u + f_y v + f_t = 0$$ (11)

Evidently, this single equation is not sufficient to uniquely compute the two unknown variables $u$ and $v$ (the aperture problem). In order to cope with the problem, Lucas and Kanade (1981) proposed the assumption that the unknown optical flow vector is constant within some neighborhood of size $\rho$. In this case, it is possible to determine the two variables $u$ and $v$ at some location $(x, y, t)$ from a weighted least squares fit by minimizing the function

$$E_{LK}(u, v) := K_\rho((f_x u + f_y v + f_t)^2)$$ (12)

where the standard deviation $\rho$ of the Gaussian serves as an integration scale over which the main contribution of the least square fit is computed. A minimum $(u, v)$ of $E_{LK}$ satisfies $\partial_u E_{LK} = 0$ and $\partial_v E_{LK} = 0$. This gives the linear system

$$\begin{pmatrix} K_\rho * (f_x^2) & K_\rho * (f_x f_y) \\ K_\rho * (f_x f_y) & K_\rho * (f_y^2) \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} -K_\rho * (f_x f_t) \\ -K_\rho * (f_y f_t) \end{pmatrix}$$ (13)

which can be solved provided that its system matrix is invertible.

In order to end up with dense flow estimates, the optical flow constraint may be embedded into a regularization framework. Horn and Schunck (1981) have pioneered this class of global differential methods by selecting the unknown functions $u(x, y, t)$ and $v(x, y, t)$ as the minimizers of the global energy function

$$E_{HS}(u, v) = \int \alpha((f_x u + f_y v + f_t)^2 + \alpha(|\nabla u|^2 + |\nabla v|^2))dxdy$$ (14)

where the smoothness weight $\alpha > 0$ serves as regularization parameter.

Minimizing this convex function and then solving its corresponding Euler–Lagrange equations with reflecting boundary conditions.

$$0 = \Delta u - \frac{1}{\alpha}(f_x^2 u + f_x f_y v + f_t f_y)$$ (15)

$$0 = \Delta v - \frac{1}{\alpha}(f_y^2 v + f_x f_y v + f_t f_x)$$ (16)

The spatial Laplace operator is given by

$$\Delta := \partial_{xx} + \partial_{yy}$$ (17)

Bruhn et al. (2005) proposed a combined local-global optical flow (LGOF) approach to incorporate the advantages of both paradigms. In order to design the LGOF method, let us firstly reformulate the previous approaches by using notations

$$w := (u, v, 1)^T$$ (18)

$$|\nabla w|^2 := |\nabla u|^2 + |\nabla v|^2$$ (19)

$$\nabla w := (f_x, f_y, f_t)^T$$ (20)

$$J_{\rho}(\nabla w) := K_\rho * (\nabla w \nabla w^T)$$ (21)

It becomes evident that the Lucas–Kanade method minimizes the quadratic form

$$E_{LK}(w) = w^T J_{\rho}(\nabla w) w$$ (22)

whereas the Horn–Schunck technique minimizes the function

$$E_{HS}(w) = \int \alpha w^T J_{\rho}(\nabla w) w + \alpha |w|^2 dxdy$$ (23)

This terminology suggests a natural way to extend the Horn–Schunck function to the desired CLG function. We simply replace matrix $J_{\rho}(\nabla w)$ by the structure tensor $J_{\rho}(\nabla w)$ at some integration scale $\rho > 0$. Thus, we propose to minimize the function

$$E_{CLG}(w) = \int \alpha w^T J_{\rho}(\nabla w) w + \alpha |w|^2 dxdy$$ (24)

Its minimizing flow field $(u, v)$ satisfies the Euler–Lagrange equations

$$0 = \Delta u - \frac{1}{\alpha}(K_\rho * (f_x^2) u + K_\rho * (f_x f_y) v + K_\rho * (f_t f_y))$$ (25)
0 = Δv − \( \frac{1}{t} (K_\rho \ast (f_i^3 \ast \mu + K_\rho \ast (f_j \ast \mu + K_\rho \ast (f_j \ast \mu))) \) (26)

Combining the temporal extended variant of both the Lucas–Kanade and the Horn–Schunck methods, we obtain a spatio-temporal version of our CLG function given by

\[
E_{C_LG}(w) = \int_{[0,1]} \left( w^T J_f \nabla f + \alpha \nabla w^T \nabla w \right) dx dy dt
\]

(27)

where convolutions with Gaussians are now to be understood in a spatio-temporal manner, and

\[
|\nabla w|^2 = |\nabla u|^2 + |\nabla v|^2
\]

(28)

**EC-LGOF method**

The main flow of the EC-LGOF MOD algorithm can be summarized as follows:

- **Step 1: Image Preprocessing:** The two adjacent frames of the input image sequence are preprocessed by a Gaussian filter with a standard deviation of 1.5 to remove noise from the images

\[
f(x, y) = K_{1.5} * I(x, y)
\]

(29)

- **Step 2: Image edge detection:** The EC edge detection algorithm is used to obtain the edge information of the object in the image and the edge image \( f_{EC}(x, y) \), which provides information concerning the spatial gradient of the image

\[
f_{EC}(x, y) := EC(f(x, y))
\]

(30)

- **Step 3: Optical flow field computation:** The flow field of the entire image can be calculated by using the LGOF algorithm. We can hence obtain the optical flow image \( f_{LF}(x, y) \), which contains information concerning the time gradient of the image

\[
f_{LF}(x, y) := LGOF(f(x, y), f_{t-1}(x, y))
\]

(31)

- **Step 4: Binarization of image:** To dynamically select the binarization segmentation threshold, both Otsu’s method and the minimum-error-threshold-selection (METS) method are deployed to linearize the edge image and the optical flow image in order to obtain \( f_{BW1}(x, y) \) and \( f_{BW2}(x, y) \), respectively, as shown in equations (32) and (33)

\[
f_{BW1}(x, y) := OTSU(f_{EC}(x, y))
\]

(32)

\[
f_{BW2}(x, y) := METS(f_{LF}(x, y))
\]

(33)

- **Step 5: Data fusion:** Following the above operation, the space gradient binary image and the time gradient binary image are acquired. Combining the space gradient information with time gradient information can give us the more accurately information of the moving objects, so in the data fusion, the AND operator is used between the edge binary image and the optical flow binary image. The process can be simplified as follows

\[
f_{BW3}(x, y) := \begin{cases} 1, & f_{BW1}(x, y) = 1 \text{ and } f_{BW2}(x, y) = 1 \\ 0, & \text{otherwise} \end{cases}
\]

(34)

- **Step 6: Morphological processing of image:** \( f_{BW3}(x, y) \) may have a small non-target area; thus, we need to process the image using erosion, dilation, and the opening and closing operations, as shown in equations (35) to (38)

\[
f_E(x, y) := f_{BW3}(x, y) \odot S
\]

(35)

\[
f_D(x, y) := f_E(x, y) \oplus S
\]

(36)

\[
f_{OO}(x, y) := (f_D(x, y) \odot S) \oplus S
\]

(37)

\[
f_{CO}(x, y) := (f_{OO}(x, y) \oplus S) \odot S
\]

(38)

where S is a structural element, \( \odot \) is the corrosion operator, and \( \oplus \) is the expansion operator.

- **Step 7: Filling the image region and labeling the connected region:** Following corrosion, expansion, the open operation, the closed algorithm, and other steps of morphological processing, most moving objects are detected. In the two consecutive frames, the moving object easily causes the internal ‘holes’. In order to remove the ‘holes’ of the moving target, we need such processing as filling processing (FP) and connected region labeling (CRL). We can then extract the image of the moving target object (x, y) as shown in equations (39) and (40)

\[
f_{FP}(x, y) := FP(f_{CO}(x, y))
\]

(39)

\[
Object(x, y) := CRL(f_{FP}(x, y))
\]

(40)

**Experimental setup and analysis of results**

The proposed method is implemented in this section to show its feasibility and performance. The experimental results are compared with state-of-the-art methods. All experiments were carried out in MATLAB 2016a under a Windows 7 environment on a PC (an Intel (R) core(TM) i3-4030U processor with 1.90 GHz and 4 GB RAM). The selection of filter function is the key factor in the EC edge detection method. In the two-dimensional case, EC operator adopts two-dimensional Gaussian function derivatives as an image smoothing filter to suppress image noise. \( \sigma \) is the standard deviation of the
Gaussian function, which controls the degree of smoothness. The size of $\sigma$ is 1.5 in this experiment. The EC algorithm is adaptive to the selection of high and low thresholds. In different cases, the threshold can be adjusted adaptively according to the image, so the detection result is more accurate. The LGOF method has two important parameters: $\rho$ and $\alpha$. $\rho$ is size of neighborhood in LGOF. A sufficiently large value for $\rho$ is very successful in rendering the LGOF method robust under noise. $\alpha$ is the smoothness weight that serves as regularizations parameter. Larger values for $\alpha$ result in a stronger penalizations of large flow gradients and lead to smoother flow fields. In this experiment, $\rho$ is set to 4.5 and $\alpha$ is set to 950, which can achieve good performance.

Dataset
Three databases, CAVIAR (Fisher, 2001), PET (The University of Reading, 2001) and KHT (Laptev and Caputo, 2005), are used to test the proposed method. The CAVIAR database consists of three datasets: INRIA entrance hall, shopping mall front view and shopping mall corridor view. The INRIA entrance hall dataset has six types of events — Browsing, Fighting, Group meeting, Leaving bags, Rest, and Walking — for a total of 28 video sequences. The PETS visual surveillance database consists of a set of datasets: PETS2000, PETS2001, PETS2002, PETS2006, PETS2007 and PETS2009.

PETS2001 was used to assess the proposed method because the scenario involved rapid changes in illumination when sunlight was blocked by a cloud. The people moving appeared small in the scene. Another challenge posed by this video was that a tree was moving owing to wind. Sample video frames from the CAVIAR INRIA entrance hall dataset and the PETS2001 dataset are shown in Figure 2 and Figure 3, respectively.

The KTH database, shown in Figure 4, consists of images of 25 subjects performing six actions in four types of scenarios, recorded via a static camera. Actions were performed with single subjects visible in a frame, with multiple executions of an action in a sequence. The actions performed were walking, jogging, running, boxing, waving and clapping.

Experiments on the outdoor dataset
Figure 5 shows the results of MOD on the outdoor dataset. Figure 5(a) and Figure 5(b) shows two consecutive frames from an outdoor video sequence, where a car is moving from the left to the right along a road. Figure 6(a) shows the
optical flow motion vector image between frames No. 3006 and No. 3007 in the video sequence. There is clearly a lot of noise in the motion vector image, and the grass waving in the wind was also detected. It is easy to find that the optical flow method has the following shortcomings. Firstly, sometimes, even if there is no movement, the optical flow can also be observed when the external illumination changes. In addition, the actual movement is often not observed in areas lacking enough gray scale changes. Secondly, in the accurate segmentation, optical flow method also needs to use the spatial characteristics such as color, grayscale and edge to improve the segmentation accuracy. Figure 6(b) shows the difference in results between frames No. 3006 and No. 3007 in the video sequence, where the contours of the moving objects are vague. It can be seen that the frame difference method, which is to detect the moving objects from the difference between the existing frame and the reference frame, causes the lack of most of the real targets in the detection results. Because the light changes in this video are less, the frame difference has better detection effect than optical flow method, but there are still a small number of holes.

Figure 7(a) shows that the proposed EC-LGOF method is able to precisely detect the contours of the objects. More specifically, the moving object in binary image is shown in Figure 7(a) and the segmented result with the minimum rectangle outside the moving objects is shown in Figure 7(b). It is clear that the proposed EC-LGOF method achieves impressive anti-noise performance after combining space gradient information with time gradient information. It has the anti-interference ability for small changes in the scene. The effect of the waving grass in Figure 6(a) and Figure 6(b) was eliminated. The EC-LGOF algorithm combines the advantages of the Optical flow method and the frame difference method, and the morphological filtering can effectively suppress the cavity, and the detection effect is better.

Experiments on the indoor dataset

Figures 8(a) and 8(b) show two consecutive frame images from the indoor video sequence, where a person was walking from the right to the left in a hall. Figure 9(a) shows the optical flow motion vector image between frames No. 1324 and No. 1325 in the video sequence. There is clearly a lot of noise in optical flow motion vector image. Figure 9(b) shows the frame difference result between frames No. 1324 and No. 1325 in the video sequence. As shown in Figure 9(b), the contours of the moving objects are vague.

Figure 10 shows the MOD results of the EC-LGOF method for the indoor test sequence. As is evident, the EC-LGOF method handled both situations successfully, and
could clearly detect the moving objects. The EC-LGOF method adopts the edge detection operator—EC, which can make the greatest response to a particular direction. It can detect the contour of the objects precisely. As shown in Figure 10(a), EC-LGOF method which combines the space gradient information with the time gradient information has good anti-noise performance.

**Experiments on the KHT dataset**

Figure 11 shows the results of moving human-body detection by the EC-LGOF method for the human action test sequence. We see that the EC-LGOF method successfully handled this situation. It was able to detect walking, running and jogging clearly. Since clapping, boxing and waving are local human body actions, human-body detection generated “holes.” However, local hand movements were not detected.

**Comparison with other methods**

This section assesses the following parameters for performance analysis: Let $TP$ be the true positive pixels, $TN$ the true negative pixels, $FN$ the false negative, and $FP$ the false positive pixels, respectively. The average value of Precision ($P$) and Recall ($R$) were computed from equations (41) and (42) below

\[
\text{Precision}_{\text{average}} = \frac{\sum_i TP}{\sum_i (TP + FP)} \quad (41)
\]
Recall average = \frac{\sum TP}{\sum(TP + FN)} \quad (42)

F\_Score\_average = \frac{(1 + \alpha) \text{Precision\_average} \times \text{Recall\_average}}{\text{Precision\_average} + \text{Recall\_average}} \quad (43)

where S denotes total number of video sequences. The F-measure (F-Score) was computed by calculating the harmonic mean of Precision and Recall.

A few more experiments were performed to compare the proposed method with other existing MOD methods, such as Horn and Schunck’s optical flow method (HS), Lucas and Kanade’s method (LK), the three-frame subtraction method (TFS), the five-frame difference method (FFS), the background subtraction-based statistical average method (BS-SAM), the background subtraction-based Gaussian mixture model (BS-GMM), and other classical methods.

The comparative results are shown in Table 1. As Table 1 shows, the optical flow method (HS, LK), which has the low anti-noise performance, has the low recognition rate in this scene, and the frame difference method (TFS, FFS) does not have good performance either. Optical-flow-based methods can be used to detect independently moving objects even in the presence of camera motion. However, most flow computation methods are computationally complex and very sensitive to noise. Frame differencing is very adaptive to dynamic environments, but generally does a poor job of extracting all the relevant pixels, for example, there may be holes left inside moving entities. At the same time, the detection effect of background subtraction is not ideal. It is simple, but extremely sensitive to changes in dynamic scenes derived from lighting and extraneous events, and so forth. Therefore, it is highly dependent on a good background model to reduce the influence of

Figure 8. Two consecutive frames from CAVIAR-Walk.
(a) Frame No. 1324
(b) Frame No. 1325

Figure 9. The results of other methods.
(a) Optical flow image.
(b) Frame difference result.
these changes, as part of environment modeling. Compared with the above other methods, the EC-LGOF method has the high recognition rate for moving objects. As is evident, the proposed method yielded the highest recognition rate of 93%. Hence, the experimental results proved that the proposed EC-LGOF is an effective method for MOD in both indoor and outdoor environments. Owing to frame-based processing using grayscale format, the proposed real-time detection system takes much less execution time (the EC-LGOF method takes 0.0451 seconds per frame). It depicts that the proposed method can be able to process approximately 22 frames per second. So, it is suitable for real-time based online video surveillance applications. Owing to the limitation of the hardware, the proposed methods has processed approximately 22 frames per second with existing hardware. If we have real-time

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC-LGOF</td>
<td>0.93</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td>HS</td>
<td>0.76</td>
<td>0.78</td>
<td>0.77</td>
</tr>
<tr>
<td>LK</td>
<td>0.73</td>
<td>0.67</td>
<td>0.70</td>
</tr>
<tr>
<td>TFS</td>
<td>0.80</td>
<td>0.81</td>
<td>0.80</td>
</tr>
<tr>
<td>FFS</td>
<td>0.82</td>
<td>0.79</td>
<td>0.80</td>
</tr>
<tr>
<td>BS-SAM</td>
<td>0.83</td>
<td>0.84</td>
<td>0.83</td>
</tr>
<tr>
<td>BS-GMM</td>
<td>0.85</td>
<td>0.84</td>
<td>0.84</td>
</tr>
<tr>
<td>Lee and Lee (2014)</td>
<td>0.91</td>
<td>0.78</td>
<td>0.84</td>
</tr>
<tr>
<td>Zhou et al. (2013)</td>
<td>0.80</td>
<td>0.98</td>
<td>0.88</td>
</tr>
<tr>
<td>Haque et al. (2008)</td>
<td>0.87</td>
<td>0.95</td>
<td>0.91</td>
</tr>
<tr>
<td>Stauffer and Grimson (1999)</td>
<td>0.79</td>
<td>0.74</td>
<td>0.76</td>
</tr>
<tr>
<td>Jung (2009)</td>
<td>0.81</td>
<td>0.99</td>
<td>0.89</td>
</tr>
</tbody>
</table>
based fast processing machine, then the processing speed can be increased.

Conclusion

This paper proposed a novel MOD method based on multi-information fusion technology. The proposed EC-LGOF method contains both spatial and the temporal gradient information. Compared with the traditional MOD technology, the proposed method is more accurate, efficient and robust. Although a high detection rate was achieved in experiments during the course of this study, our future work will be focused on many challenging problems that remain, including the deployment of the advanced computer hardware.

Declaration of conflicting interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is supported by the National Natural Science Foundation of China (61602220, 61673249, 61672204, 41401521), the Excellent Talents Training Funded Project of Universities of Anhui Province(gxzx2017099 gxzyQZD 2017076), the National Natural Science Research Project of Universities of Anhui Province (KJ2014A212, KJ2015A206.), the Research Fund Project of Talents Hefei University (13RC01, 15RC07), the Natural Science Foundation of Jingxi province of China (20161BAB21057), and the Natural Science Foundation of Fujian province of China (2016J01325, 2015J05015).

References


