A Noisy Videos Background Subtraction Algorithm Based on Dictionary Learning

Huaxin Xiao[1], Yu Liu[1], Shuren Tan[1], Jiang Duan[2] and Maojun Zhang[1]
[1] College of Information System and Management, National University of Defense Technology, Changsha, PR China
[e-mail: huaxinxiao@hotmail.com]
[2] School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu, PR China
[e-mail: duanj711@gmail.com]
*Corresponding author: Huaxin Xiao

Received January 23, 2014; revised March 31, 2014; revised April 24, 2014; accepted April 28, 2014 ; published June 27, 2014

Abstract

Most background subtraction methods focus on dynamic and complex scenes without considering robustness against noise. This paper proposes a background subtraction algorithm based on dictionary learning and sparse coding for handling low light conditions. The proposed method formulates background modeling as the linear and sparse combination of atoms in the dictionary. The background subtraction is considered as the difference between sparse representations of the current frame and the background model. Assuming that the projection of the noise over the dictionary is irregular and random guarantees the adaptability of the approach in large noisy scenes. Experimental results divided in simulated large noise and realistic low light conditions show the promising robustness of the proposed approach compared with other competing methods.

Keywords: Dictionary learning, background subtraction, sparse representation, low light, noisy videos

This research was partially supported by Research Fund for Doctoral Program of Higher Education of China under project No. 20134307120040 and National University of Defense and Technology under project No.JC13-05-02.

http://dx.doi.org/10.3837/tiis.2014.06.008

1. Introduction

The continuous improvement of equipment manufacturing and computer processing capability led to the wide application of intelligent video surveillance technology to industry, defense, transportation, and other fields. One of the goals of this technology is to simulate the function of the human visual system, such as object tracking, classification, and behavior understanding, in an arbitrary scene. However, these smart applications are based on motion detection that correctly detects moving targets and exactly segments them. Three methods of motion detection have been developed in previous literature: optic flow, frame difference, and background subtraction.

The optic flow method [1, 23, 24] assigns a velocity vector to each pixel of image that forms an optic flow field. The optic flow field is an approximate estimate of the true motion field that reflects the gray changing trend of pixels. This method can detect motion object without any information of scene or with a stationary camera. However, sensitivity of light and complexity of computation restrict its application in video surveillance systems [2].

Frame difference is based on a threshold difference between the previous and the current frames. This method efficiently performs in computational terms and grants a prompt object motion detection between two frames [3]. Nevertheless, it suffers two well-known drawbacks [4] caused by frame rate and object speed: foreground aperture and ghosting. Moreover, it lacks the flexibility to handle the dynamic and complex scene.

Background subtraction [5-10, 14-16, 30] establishes a background model of the monitored scene through a suitable method and then calculates the difference between the current frame and the background model, which segments the foreground area from the scene. It can solve issues of frame difference and robustly perform in the dynamic scene by using the background update procedure. A large number of algorithms have been developed to represent the statistical model of the background [25]. These methods perform at the level of pixel and ignore the correlation of spatial information. Wren et al. [5] independently modeled the background at each pixel location with a Gaussian probability density function. Later, Friedman and Russell [6] used three Gaussian distributions that correspond to the road, shadow, and vehicle to model the traffic surveillance system. Stauffer and Grimson [7] extended this opinion by employing a mixture of multiple Gaussian distributions to model the pixels in the scene. It has been proved to be a popular solution to the modeling of complex background. When the assumptions imposed by the selected model in parametric methods fail, the non-parametric approaches are a better choice. In non-parametric approaches, a kernel is created around each of the previous samples, and the density is estimated using an average over the kernels. Elgammal et al. [8] proposed a normal kernel that can deal with arbitrary shapes of the density function. Unlike the methods of statistical background model, Oliver et al. [9] considered the spatial configuration that captured the Eigen backgrounds by eigenvalue decomposition based on the whole image. Later Monnet et al. [10] proposed an incremental principal component analysis (PCA) method to predict model states, which can be used to capture motion characteristics of backgrounds. In practical applications, a robust PCA approach [11–13] was proposed that is more effective than the incremental PCA method. The spatial correlated approaches can effectively deal with the brightness and other global changes. In addition, employing compressive sensing theory in solving background subtraction has been successful in recent years. Cevher et al. [14] assumed that the majority of the pixels in a frame belong to the background. Thus, the foreground is sparse after background subtraction.

Subsequently, Zhao et al. [15] further developed this idea by adding an assumption that the background also has a sparse representation and learning a dictionary to characterize the background changes.

1.1 Contribution and Organization

The aforementioned methods are mainly for the complex and dynamic scene in the background, such as rain, snow, waves and shaking trees, without considering low light or noisy environment. Large noise, low value and small differences in grey level are the typical characteristics of low light images. Excessive large noise and low grey value would bring negative influence on detection, which leads to the existing motion detection methods perform inappropriately.

In this paper, we propose a robust background subtraction method based on dictionary learning and sparse coding to handle the large noise condition. Firstly, this paper formulates the background modeling step as a sparse representation problem and regards the background subtraction as the sparse projection over the dictionary. Then it detects the foreground as the difference between the reconstructed image and the background model.

Secondly, different from the assumptions of [14, 15], we put forward a significant assumption that statistical noise is typically distributed through the larger space anisotropically. Then analysis and certify this assumption in the latter section. Based on this assumption, the proposed method can remove the influence of large noise distinctly and perform robust under different large noise and low light environments as a result of sparse representation.

The rest of this paper is organized as follows: Section 2 describes the basic principle of the proposed method based on three assumptions. Section 3 presents the mathematical formulation of the proposed method. Section 4 shows the comparison of the experimental results with those of the existing methods on public testing datasets with simulated noise and realistic low light videos. Section 5 concludes and discusses future possible research direction.

2. Basic principle

According to the approximate description of the proposed method in Section 1, the proposed approach can be divided into three parts: background modeling by dictionary learning and sparse coding, sparse representation of the current frame and foreground detection. This Section will introduce the principles of these parts based on the following assumptions that theoretically provide a reasonable explanation of the proposed method.

In the framework of background subtraction, the current frame I can be linearly decomposed as follows:

$$I = I_B + I_F \tag{1}$$

where I_B is the background model and I_F is the foreground candidate.

The background model I_B is the most critical step to the success of background subtraction approach. This model is established with the linear and sparse combination of the atoms in the dictionary D, which is based on the idea of background modeling with basis vectors [9]:

$$I_B = D \times \alpha \tag{2}$$

where α is the sparse coefficient.

Compared with the eigenvalue decomposition [9], the sparse decomposition over a redundant dictionary is more effective in applying signal processing. The background can be represented sparsely by projecting on the atoms of the dictionary. This process leads to the first assumption similar to [15]:

Assumption 1: The background of an arbitrary scene can be sparsely and linearly represented by the atoms of the dictionary.

Sparse representation always aims that the reconstruction signal can be as close as possible to the original one. When a moving target enters into the scene, it changes the structure of the background, and the original sparse representation will not be the same. In other words, when the test frame with moving objects is presented by the subspace spanned by pure background bases, the unchanged area of the scene can be well recovered. By contrast, the changed area would be reconstructed with a deviation of the projection on such subspace. Measuring this deviation satisfies the purpose of detection. The second assumption is proposed based on the above-mentioned analysis:

Assumption 2: The foreground leads to the changing of the background and greatly transforms the projection over the dictionary.

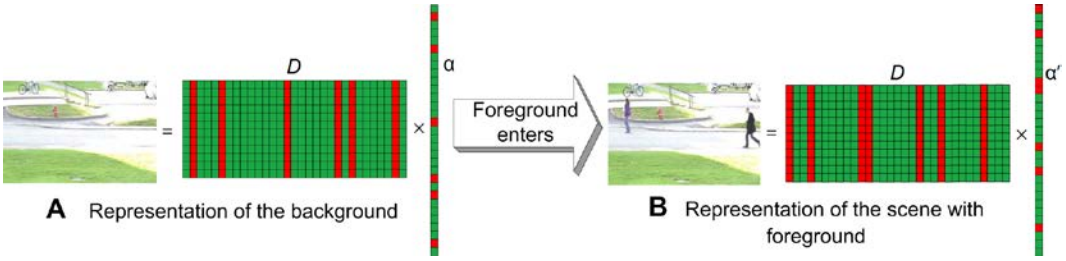

Fig. 1. Process of sparse coefficients changing when a foreground enters the scene. The dictionary D and sparse coefficients α are employed to represent the pure background model and the scene with foreground. **Fig. 1A** is the sparse representation of the pure background as described in Assumption 1. Then, the foreground breaks the original equation, as shown in **Fig. 1B**. The dictionary D can be obtained by employing the learning algorithms such as K-SVD and Online Dictionary Learning, and the coefficient α is a sparse coding problem.

Fig. 1 shows the process of sparse coefficients changing when a foreground enters into the scene. **Fig. 1A** shows the sparse representation of the pure background as described in Assumption 1. Then, the foreground breaks the original equation, as shown in **Fig. 1B**.

The two predominant sources of noise in digital image acquisition are the stochastic nature of the photon counting and the intrinsic thermal and electronic fluctuations of the acquisition devices [17]. Under the normal illumination circumstance, the second noise is the primary component. When the light decreases, the rapid increase of the first noise brings a large number of noise to the captured images. When the noise flashing level is very large, the existing detection methods become ineffective. To guarantee the adaptability of the proposed method under low light condition, a noise assumption is proposed:

Assumption 3: The projection of the noise over the dictionary is irregular and random.

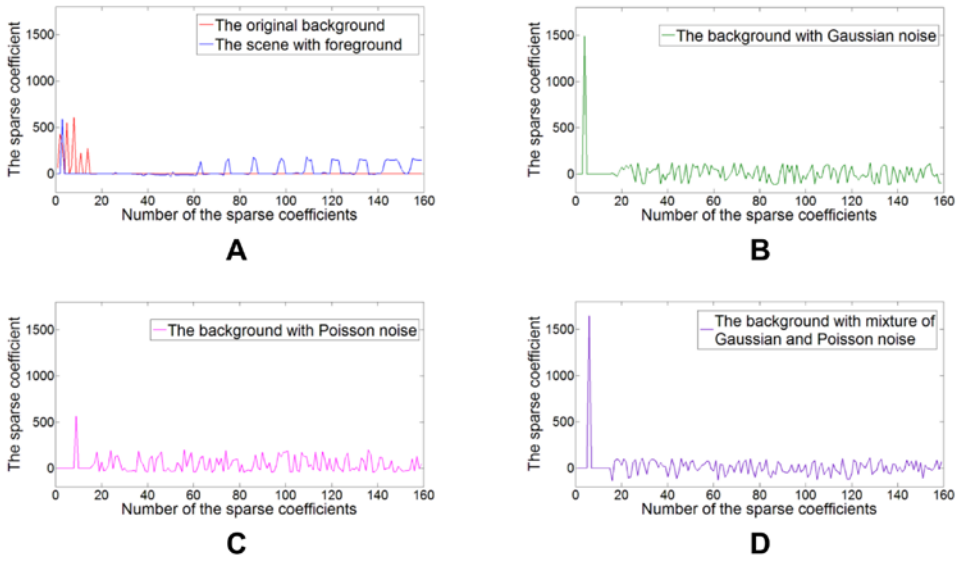

Fig. 2. Comparison of sparse coefficients. The red curve in **Fig. 2A** is the sparse coefficient of the original background projected on the dictionary, and the blue one is the scene with foreground. The other three curves in **Fig. 2B-D** are the background with Gaussian white noise ($\sigma = 400$), Poisson noise ($\alpha = 500$) and mixture of both ($\sigma = 250$, $\alpha = 250$).

Fig. 2 shows the comparison of the sparse coefficients under different circumstances in a certain background. The red curve in **Fig. 2A** represents the sparse coefficients of the original background projected on the learned dictionary, and the blue one is a case with foreground entering. When these two curves are compared, the foreground significantly and regularly changes the sparse coefficients. Regardless of the types of foreground, it always has a certain structure that presents the regular coefficients over the bases. The other three curves in **Fig. 2B-D** are the background with Gaussian white noise ($\sigma = 400$), Poisson noise ($\alpha = 500$) and mixture of both ($\sigma = 250$, $\alpha = 250$). These three curves randomly and confusedly reflect the sparse coefficients of the noise distributes on the dictionary as described in Assumption 3. Regardless of the types of noise, the randomness and anisotropy of noise determine the disorder of the distribution on the whole dictionary. Thus, when reconstructing an image through the sparse model, only several atoms in the dictionary are selected to represent the original signal. Most of the noise can be effectively removed. These factors ensure the proposed method is suitable for handling large noise environments.

3. Proposed method

The three assumptions described in Section 2 are the bases of the proposed method. First, according to Assumption 1, dictionary learning is applied to obtain the basis vectors of the scene. Then, sparse coding is combined with dictionary learning to model the background of the scene. For an arbitrary frame, the proposed method projects it on the learned dictionary to acquire the sparse representation. Finally, the difference between the sparse representations of the background model and the current frame are regarded as the detection criteria. **Fig. 3** shows the flowchart about the detailed process of the proposed method.

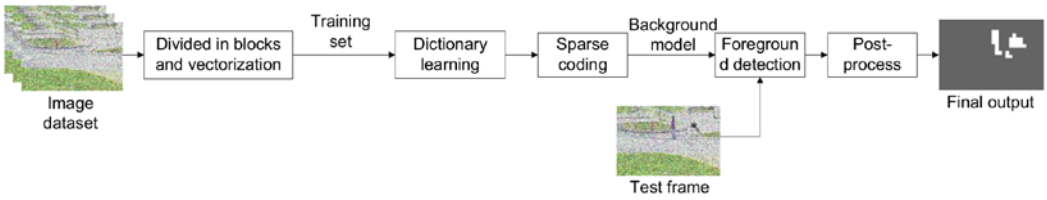

Fig. 3. Flowchart of the proposed method.

3.1 Background modeling

In (2), the background model is formulated as the linear and sparse combination of the atoms in the dictionary D. Dictionary has been proved very effective for signal reconstruction and classification in audio and image processing domains [18]. Compared with the traditional signal decomposition methods such as wavelet and PCA, dictionary learning does not emphasize the orthogonality of bases. Thus, it represents the signal as having better adaptability and flexibility.

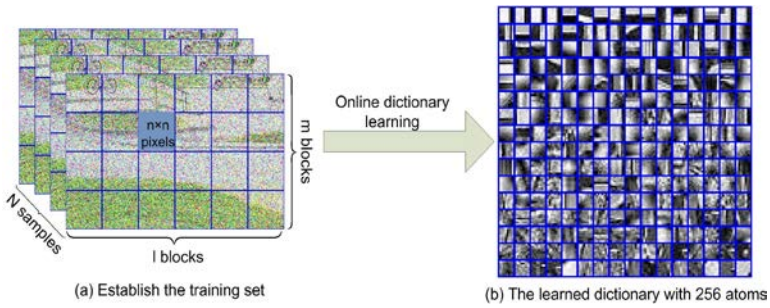

Fig. 4. (a) Training set established with N samples. Each image is divided into $m \times l$ blocks of size $n \times n$ pixels. (b) Learned dictionary with 256 atoms by Online Dictionary Learning [18].

The background frames without foreground are extracted from the surveillance video to form a training set with N samples. **Fig. 4(a)** shows that each of the collected images is divided into $m \times l$ blocks of size $n \times n$ pixels. The jth image block of the ith sample can be vectorized as $\vec{x}_i^j$. Then the jth image block of each sample is combined and it consists of a training set of $X_j = \{\vec{x}_i^j \,|\, i = 1, \cdots, N\}$. Its dictionary D_j satisfies the following formula [18, 19]:

$$D_j = \arg\min_{D_j} \sum_{i=1}^{N} \min_{\alpha_i} (\left\| \vec{x}_i^j - D_j \alpha_i \right\|_2^2 + \lambda \left\| \alpha_i \right\|_1) \tag{3}$$

where α_i is the ith sparse coefficient and λ is a regularization parameter.

The Online Dictionary Learning algorithm [18] is used in this study to solve Formula (3). The learned dictionary with 256 atoms is shown in **Fig. 4(b)**. The algorithm adopts the stochastic gradient descent method in each loop to choose a vector $\vec{x}_i^j$ that is regarded as x_t from X_j and t is the times of the repeat. It applies sparse coding based on the previous $t-1$ loops to obtain the tth decomposition coefficient α_t. The formula is as follows:

$$\alpha_t = \arg\min_{\alpha \in \mathbf{R}^k} \frac{1}{2} \left\| x_t - D_{t-1} \alpha \right\|_2^2 + \lambda \left\| \alpha \right\|_1 \tag{4}$$

Sparse coding is a class of methods that automatically choose good basis vectors for unlabeled data. The Least Angle Regression algorithm [20] can solve Formula (4), especially when the solution is sufficiently sparse. Furthermore, the solution is precise and does not rely on the correlation of atoms in the dictionary unless the solution is not unique. Then the dictionary $D_{t-1}=[d_1,\cdots,d_k]$ is updated column by column and a new dictionary D_t is obtained. The update rules are as follows:

$$\begin{cases} u_j \leftarrow \dfrac{1}{A_{jj}}\left(b_j - D\mathrm{a}_j\right)+d_j \\ d_j \leftarrow \dfrac{1}{\max(\left\|u_j\right\|_2,1)}u_j \end{cases} \tag{5}$$

where $A=[\mathrm{a}_1,\cdots,\mathrm{a}_k]=\sum_{i=1}^{t}\alpha_i\alpha_i^T$ and $B=[b_1,\cdots,b_k]=\sum_{i=1}^{t}x_i\alpha_i^T$. When the background of a scene changes, the above-mentioned update rules can be used to update the background model, thereby ensuring robustness.

Sparse coding and dictionary updating are alternately performed until the times of iteration are achieved. This algorithm is simple, fast, and suitable for large-scale image processing. Mairal et al. [18] have shown that the algorithm can converge to a fixed point. The above-mentioned method is applied onto each block and then the whole image dictionary D and sparse coefficients α are obtained. The process of background modeling is then completed with (2), as described in **Algorithm 1**.

Algorithm 1: Background modeling

Input: $m\times l$ (number of blocks), T (number of iterations), λ (regularization parameter), k (number of atoms), D_i^0 (initial dictionary for ith block).

for $i=1$ to $m\times l$ **do**

  for $t=1$ to T **do**

    (a) Sparse coding: fixed D_i^{t-1}, employ LARS algorithm [20] to solve (4) and get the sparse coefficients α_{t-1};

    (b) Dictionary update: update the dictionary $D_i^{t-1}=[d_1,\cdots,d_k]$ column by column with (5) and obtain a new dictionary D_i^t;

  end

  Compute the sparse coefficients α_i over the final dictionary D_i;

end

Compute the background model with (2): $I_B=D\times\alpha=[D_1,\cdots,D_{m\times l}]\times\begin{bmatrix}\alpha_1^T\\ \vdots \\ \alpha_{m\times l}^T\end{bmatrix}$;

Output: the background model I_B.

3.2 Foreground detection

After the background model is established, the next step is to detect the foreground. Similar to the process of background modeling, the sparse coefficients α' on the dictionary D for an arbitrary frame I can be obtained by sparse coding. Referring to the idea of the background subtraction method, the foreground is detected by the differences between the sparse representation of the current frame I and the background model I_B. Thus, the foreground that enters the monitored scene can be presented as follows:

$$I_F = I - I_B = D\alpha' - D\alpha \tag{6}$$

The differences of I_F are calculated in blocks, and they are summed up as vector Δ:

$$\Delta = \left\{ \sum_{p=1}^{n^2} I_F^j(p) \middle| j = 1, 2, \cdots, m \times l \right\} \tag{7}$$

where $I_F^j(p)$ is the pth pixel of the jth block in I_F.

Then the threshold region T is used to judge $\Delta(j)$. The structure of the jth block in this region does not change, i.e., no foreground accesses. By contrary, if an object enters the scene, then $\Delta(j)$ is set to 0. This study assumes that the data in the Δ approximately follow the Gaussian distribution. Therefore, the upper and lower limits of the threshold region T are set with 3σ criterion:

$$\begin{cases} T_{\max} = \mu + 3\sigma \\ T_{\min} = \mu - 3\sigma \end{cases} \tag{8}$$

where μ and σ are the mean and variance respectively of the differences between images in the training set and background model.

However, avoiding an isolated point that appears in the detection results is difficult for one time of judgment with certain threshold. The results of the previous threshold judgment are post-processed with weight coefficients. Given that the pixels in an object are monolithic, the image block can be determined according to its neighbor:

$$\Delta'(j) = \left(1 - SSIM_j\right) * \left(\Delta(j) + \sum_{k \in neighbor(j)} \Delta(k) \right) \tag{9}$$

where $neighbour(j)$ is the 3×3 neighborhood of the jth block. $SSIM_j$ is the value of Structural Similarity Index Measurement [21] between the ith block of the current frame and the background model. $1 - SSIM_j$ is the weight coefficient that adequately uses the structure information. If the block of the current frame is similar to the one in the background model, the $1 - SSIM_j$ would be very small. The $\Delta'(j)$ would then be low and the block would not be regarded as the foreground. Formula (9) can enhance the effect of foreground segmentation. The detailed algorithm about foreground detection is described in **Algorithm 2**.

Algorithm 2: Foreground detection

Input: I (test frame), $m \times l$ (number of blocks), D_j (learned dictionary for the jth block), $T_{\min}$ (minimum of threshold region), $T_{\max}$ (maximum of threshold region).

for $j = 1$ to $m \times l$ **do**

Compute the sparse coefficients α'_j of the jth block in I over the dictionary D_j;

$I_F^j \leftarrow D_j \alpha'_j - D_j \alpha_j$;

$\Delta(j) \leftarrow \sum_{p=1}^{n^2} I_F^j(p)$;

if $\Delta(j) \geq T_{\max}$ or $\Delta(j) \leq T_{\min}$

$\Delta(j) \leftarrow 0$;

end

end

for $j = 1$ to $m \times l$ **do**

Compute the $SSIM_j$ between the jth block of I and I_B;

Compute new score $\Delta'(j)$ with (9);

if $\Delta'(j) \geq T_{\max}$ or $\Delta'(j) \leq T_{\min}$

$\Delta'(j) \leftarrow 0$;

end

end

Select the zero value in Δ' as the index of the block to identify the foreground;

Output: the foreground I_F.

4. Experiments

To show the qualitative and quantitative performance of the proposed method, it has been tested under different levels of light and noise conditions. The experiments are implemented in two parts: on the public testing datasets [22] and on realistic low light videos. The realistic videos are converted to 360×240 size, similar to the size of datasets in [22] to provide an even comparison.

4.1 Implementation details

Different levels of Gaussian noise, Poisson noise, or mixture of both are added to the images in the public testing dataset [22]. The following equation presents the process of artificial noise:

$$nI = \alpha y + n \tag{10}$$

where nI is the pixel value of the noise image and α is the scale factor. y and n obey the

distribution of Poisson $P(\lambda)$ and Gaussian $N(\mu,\sigma^2)$ respectively. λ is the pixel value of the original image. μ and σ are the mean and variance of the Gaussian noise. The different degrees of noise images can be obtained by adjusting the parameters of (10). For the low light video, the illumination of the environment was recorded when taking the video.

In this study, the image block is treated as a basic processing unit, and the size of block has a certain effect on the computing speed, detection results, and recovered image effects. The different performances on the block size are shown in **Fig. 5**. Smaller blocks can maintain a better precision of detection results, whereas a larger block size can guarantee the accuracy, as shown in the second row of **Fig. 5**. Precision and accuracy are trade-off parameters, and simultaneously ensuring both at a high level is difficult. After testing and comparison, 12×12 is chosen as the block size. Such a block size deals with a frame in 1.5 seconds to 2.0 seconds on the MATLAB implementation of the proposed algorithm. The execution time is recorded by a machine with 2.2GHz Pentium E2200 processor and 2GB of RAM.

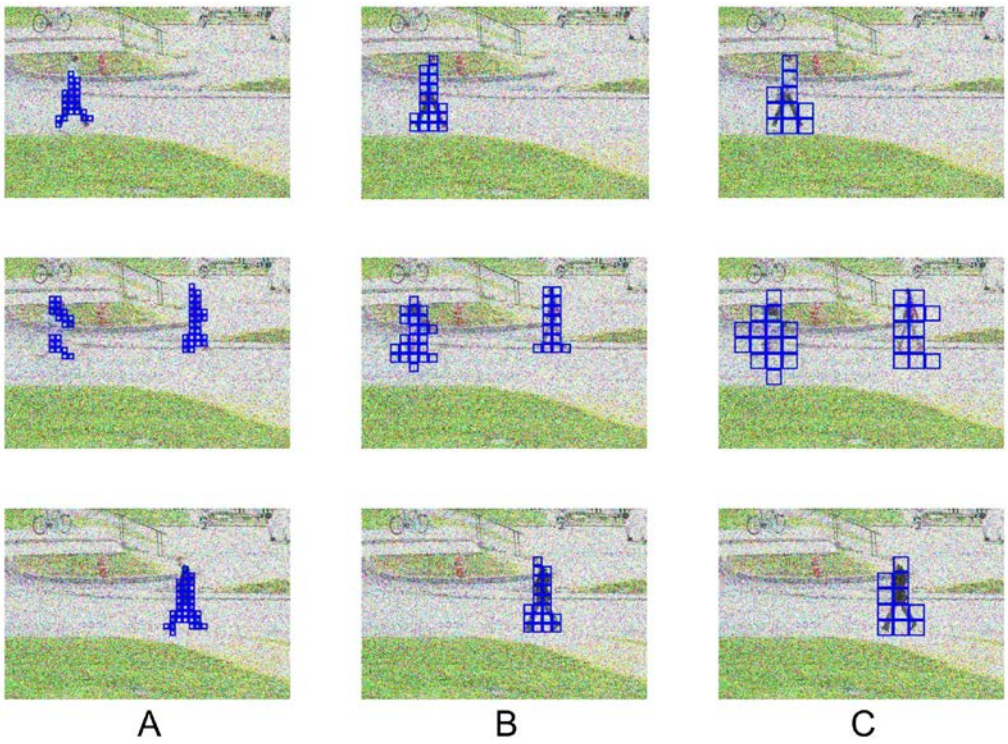

Fig. 5. Different sizes of block comparison. Noise was added to the *Pedestrians* dataset [22] with a mixture of additive Gaussian white noise ($\sigma = 50$) and Poisson noise ($\alpha = 50$). The detection results are presented with blue boxes. **Fig. 5A-C**: Detection results of 8×8 , 12×12 , and 20×20 block size.

4.2 Results on public testing dataset

The proposed method is compared with the competing background subtraction algorithms: Mixture of Gaussian model [7], Non-parametric model [8], and ViBe [28]. The robustness of the proposed approach is then confirmed under different types of large noise.

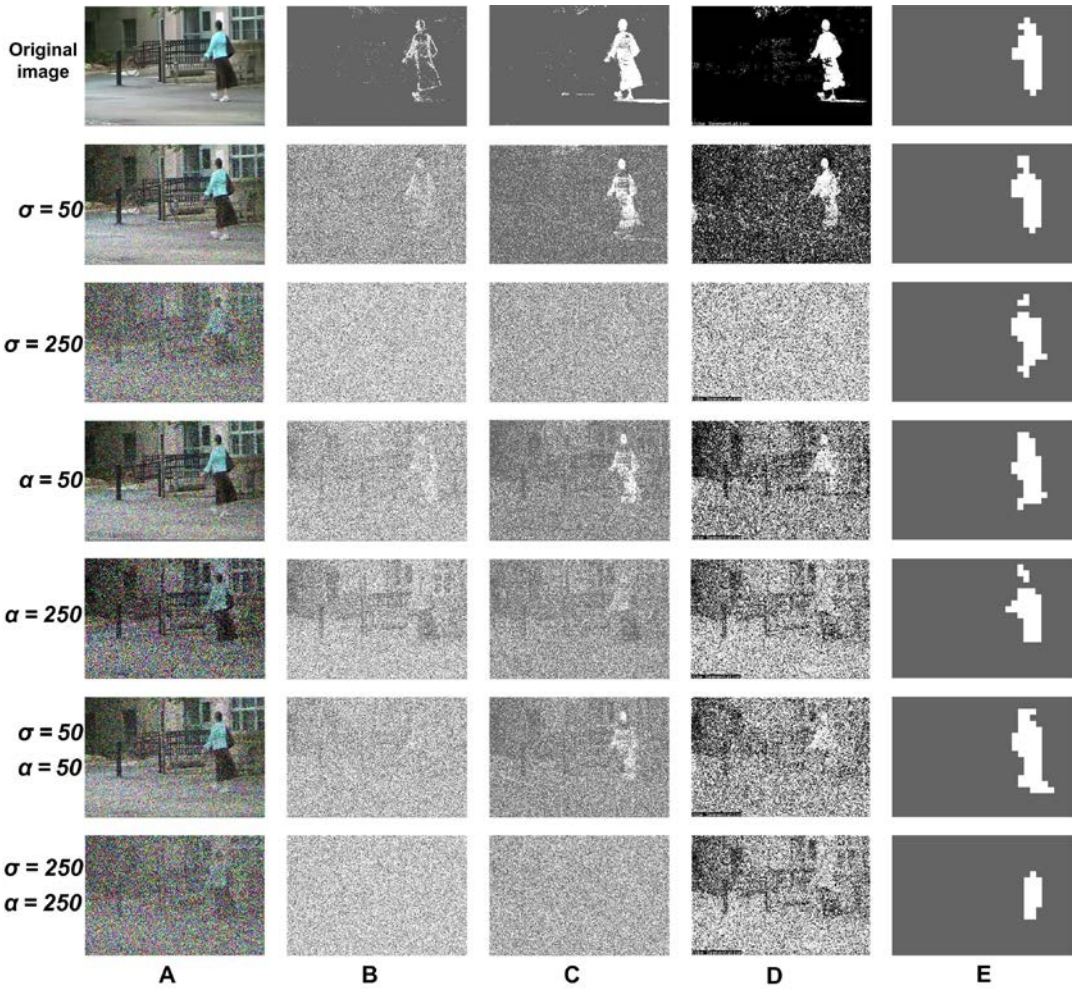

Fig. 6. Detection results of different methods with different types of noise on *Backdoor* dataset [22]. Different levels of Gaussian white noise (σ =50 and 250), Poisson noise (α =50 and 250) and mixture of both ($\sigma = 50, \alpha = 50$ and $\sigma = 250, \alpha = 250$) are added to the original image. **Fig. 6A**: Test images with different types of noise. **Fig. 6B-E**: Detection results of using mixture of Gaussian model [7], non-parametric model [8], ViBe [28] and the proposed method.

Fig. 6 shows the comparison between competing background subtraction algorithms and the proposed method. Different levels of Gaussian white noise (σ =50 and 250), Poisson noise (α =50 and 250) and mixture of both ($\sigma = 50, \alpha = 50$ and $\sigma = 250, \alpha = 250$) are added to the original image. The results show that in none noise condition, non-parametric [8] and Vibe [28] have a better exact detection result. However, when adding a certain degree noise to the original image, the results of the compared methods are seriously affected by the noise. With the noise continuing to rise, the compared methods lose efficacy completely because the background model assumptions fail. By contrary, the proposed method performs robust and handles different noise well. The proposed method was tested on various datasets in [22] and the detection results were presented on one of the datasets.

Fig. 7. Detection results under different types of noise on *Bus Station* dataset [22]. **Fig. 7A**: Original images with ground truth. **Fig. 7B-D**: Images and results with Gaussian white noise ($\sigma = 250$), Poisson noise ($\alpha = 250$), and the mixture of both ($\sigma = 150, \alpha = 250$).

In **Fig. 7**, different types of noise are added on original test images. Disregard to the noise type, the proposed approach can perform a stable and robust detection work. However, since the dictionary is learned from the background images, the recovered blocks are close to the background when the foreground has similar colors. This circumstance increases the difficulty in detection process. In the second row of **Fig. 7**, the color of the right-most pedestrian's cloth is identical to the color of the dustbin which leads to a fail detection. By contrast, if the background and foreground colors have visible differences (the third row of **Fig. 7**), the proposed method can properly detect the person/object.

To evaluate the quantitative performance of the proposed method, three quantitative metrics were adopted in this study [30]:

$$Recall = \frac{tp}{tp + fn} \tag{11}$$

$$Precision = \frac{tp}{tp + fp} \tag{12}$$

$$F - measure = \frac{2 * Recall * Precision}{Recall + Precision} \tag{13}$$

where tp is the number of pixels correctly classified as foreground. $tp + fn$ and $tp + fp$ are the number of pixels detected as foreground pixels by ground truth and the proposed method, respectively.

One hundred frames with foreground from *Backdoor* and *Bus station* dataset [22] were selected to calculate the quantitative metrics. The results are shown in **Table 1**, **2** and **3**. In **Table 1**, the compared methods have a higher value of Recall because of false detection affected by noise, whereas the Precision of the proposed method obtains a better performance in **Table 2**. F-measure is considered as a single measure, that is, the weighted harmonic mean of Recall and Precision in (13). **Table 3** shows that the proposed method has a satisfactory quantitative performance regardless of the dataset or noise level.

Table 1. Quantitative metric of *Recall*

Dataset [22]	Noise level	MOG [7]	KDE [8]	ViBe [28]	Proposed
Backdoor	No noise	0.5601	0.8525	**0.8677**	0.8504
	$\alpha = 20$	0.6404	0.6502	**0.7998**	0.7739
	$\alpha = 50$	**0.8060**	0.5154	0.7329	0.7322
	$\alpha = 80$	0.6754	0.4731	0.7125	**0.7208**
Bus station	No noise	0.7504	**0.8571**	0.8362	0.8398
	$\alpha = 20$	0.3308	0.4314	**0.7756**	0.7200
	$\alpha = 50$	0.3631	0.4052	0.5628	**0.6168**
	$\alpha = 80$	0.4871	0.4015	0.4492	**0.5801**

Table 2. Quantitative metric of *Precision*

Dataset [22]	Noise level	MOG [7]	KDE [8]	ViBe [28]	Proposed
Backdoor	No noise	0.8837	0.7077	**0.8969**	0.8913
	$\alpha = 20$	0.2629	0.3189	0.1532	**0.8876**
	$\alpha = 50$	0.1122	0.1757	0.1180	**0.8633**
	$\alpha = 80$	0.1046	0.1276	0.0923	**0.8460**
Bus station	No noise	0.8934	0.7925	**0.9262**	0.8657
	$\alpha = 20$	0.1252	0.1060	0.0844	**0.8525**
	$\alpha = 50$	0.0396	0.0685	0.0457	**0.8617**
	$\alpha = 80$	0.0345	0.0495	0.0372	**0.8339**

Table 3. Quantitative metric of *F-measure*

Dataset [22]	Noise level	MOG [7]	KDE [8]	ViBe [28]	Proposed
Backdoor	No noise	0.6856	0.7734	**0.8821**	0.8704
	$\alpha = 20$	0.3728	0.4279	0.2571	**0.8269**
	$\alpha = 50$	0.1970	0.2621	0.2033	**0.7924**
	$\alpha = 80$	0.1811	0.2010	0.1634	**0.7784**
Bus station	No noise	0.8157	0.8235	**0.8789**	0.8526
	$\alpha = 20$	0.1816	0.1702	0.1522	**0.7807**
	$\alpha = 50$	0.0714	0.1172	0.0845	**0.7190**
	$\alpha = 80$	0.0644	0.0881	0.0687	**0.6842**

4.3 Results on our low light video

The proposed method is employed on the low light video taken in this study. The image sensor is SONY IMX 104 CMOS. Similar to the Section 4.1, the methods of [7], [8] and [28] are compared with the proposed method and the detection results are shown under different low illumination environments.

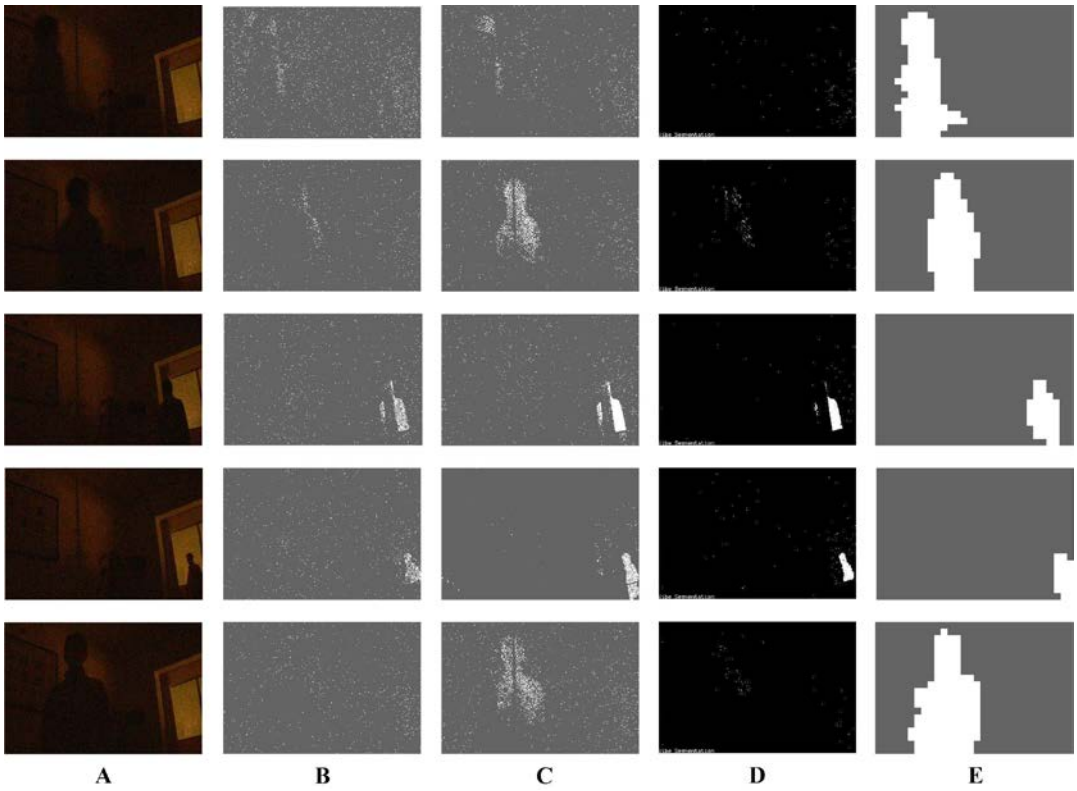

Fig. 8. Detection results of different methods under 0.1-0.5 **lx** environment. **Fig. 8A**: Test frames extracted from the low light video. **Fig. 8B-E**: Detection results of using mixture of Gaussian model [7], non-parametric model [8], ViBe [28] and the proposed method.

In **Fig. 8**, the mixture of Gaussian [7], non-parametric model [8] and ViBe [28] are compared with the proposed method under realistic low light condition. The illumination of the environment in **Fig. 8** is about 0.1-0.5 **lx**. There are obvious noise appearing in the captured images. The noise in **Fig. 8** corresponds the mixture of Gaussian ($\sigma = 50$) and Poisson ($\alpha = 50$) noise approximately. With the reducing of the illumination, the noise captured in the low light video increases exponentially. While the light of the scene decreases from right to left of the scene, the performances of the compared methods are degenerate. When the moving object is in the left of the scene (the third and four row of the **Fig. 8**), the methods of [7], [8] and [28] can detect the object. However, the compared methods perform poorly when it appears in the left of scene (the first and fifth row of the **Fig. 8**). Meanwhile, the proposed method behaves robust and well no matter where the foreground is. The method proposed in this study is robust regardless of the artificial large noise or realistic low light circumstances, as shown in **Fig. 6**.

Fig. 9 shows the detection results of different methods under lower light. The illumination

of environment in **Fig. 9** is about 0.01-0.05 **lx**. In order to facilitate observation, the intensity of the brightness is increased, which also increases the noise level. Similar to the **Fig. 8**, the methods of [7], [8] and [28] are also compared with the proposed approach. The proposed method still behaves more robust than the compared methods.

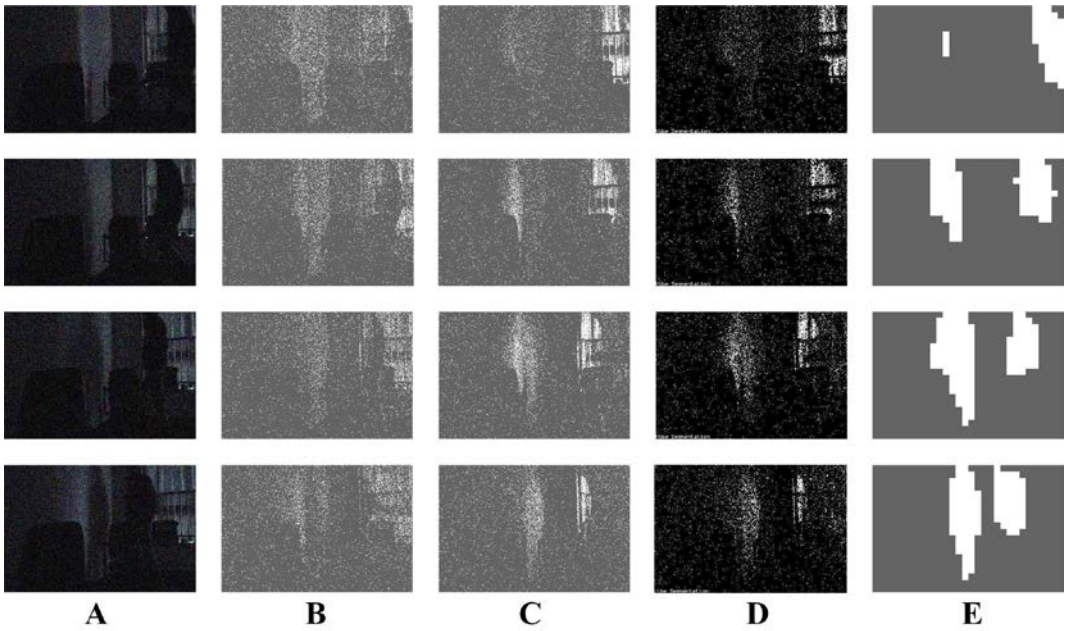

Fig. 9. Detection results of different methods under 0.01-0.05 **lx** environment. **Fig. 9A**: Test frames extracted from the low light video. **Fig. 9B-E**: Detection results of using mixture of Gaussian model [7], non-parametric model [8], ViBe [28] and the proposed method.

5. Conclusion

Most background subtraction methods highlight the capability of handling dynamic scenes [10, 15, 29] but ignore low light circumstances. Large noise caused by low light will greatly affect the traditional algorithms and lead to their poor performances, as shown in **Fig. 6** and **8**. This paper proposes a robust background subtraction algorithm based on dictionary learning and sparse coding to handle the large noise condition. The proposed method can achieve a satisfactory detection performance that is not influenced by the large statistical noise with different types and scales. The proposed method would poorly work when the variance of noise is larger than 500. In this case, distinguishing it from others is difficult for the human visual system.

In the proposed method, the whole image is divided into a group of blocks within which the motion detection is independently dealt with. Thus, the result is calculated as an inaccurate mosaicking output when compared with the pixel-level background subtraction methods. This study will focus on the precise detection of the proposed method, which will be further refined in the future. This study will also be a promising topic of investigation in the future.

References

[1] J. L. Barron, D. J. Fleet and S. S. Beauchemin, "Performance of optical flow techniques," *International journal of computer vision*, vol. 12, no. 1, pp. 43-77, Feb. 1994. Article (CrossRef Link)

[2] S. S. Beauchemin and J. L. Barron, "The computation of optical flow," *ACM Computing Surveys (CSUR)*, vol. 27, no. 3, pp. 433-466. Sep. 1995. Article (CrossRef Link)

[3] D. A. Migliore, M. Matteucci and M. Naccari M, "A revaluation of frame difference in fast and robust motion detection," in *Proc. of ACM VSSN*, pp. 215-218, Oct. 2006. Article (CrossRef Link)

[4] H. Lee, S. Hong, and E. Kim, "Probabilistic Background Subtraction in a Video-based Recognition System," *KSII Transactions on Internet & Information Systems*, vol. 5, no. 4, May, 2011.

[5] Article (CrossRef Link)

[6] C. R. Wren, A. Azarbayejani, T. Darrell and A. P. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780-785, Jul., 1997. Article (CrossRef Link)

[7] N. Friedman and S. Russell S, "Image segmentation in video sequences: A probabilistic approach," in *Proc. of the 13th conference on Uncertainty in artificial intelligence*, pp. 175-181, Aug., 1997. Article (CrossRef Link)

[8] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. of IEEE CVPR*, vol. 2, Jun., 1999. Article (CrossRef Link)

[9] A. Elgammal, D. Harwood and L. Davis, "Non-parametric model for background subtraction," in *Proc. of Computer Vision-ECCV*, pp. 751-767, Jul., 2000. Article (CrossRef Link)

[10] N. M. Oliver, B. Rosario and A. P. Pentland, "A Bayesian computer vision system for modeling human interactions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*,vol. 22, no. 8, pp. 831-843. Aug., 2000. Article (CrossRef Link)

[11] A. Monnet, A. Mittal, N. Paragios and V. Ramesh, "Background modeling and subtraction of dynamic scenes," in *Proc. of IEEE ICCV*, pp. 1305-1312, Oct., 2003. Article (CrossRef Link)

[12] F. De La Torre and M. J. Black, "A framework for robust subspace learning," *International Journal of Computer Vision*, vol. 54, no. 1-3, pp. 117-142, Aug., 2003. Article (CrossRef Link)

[13] Q. Ke and T. Kanade, "Robust L1 norm factorization in the presence of outliers and missing data by alternative convex programming," in *Proc. of IEEE CVPR*, vol. 1, pp. 739-746. Jun. 2005. Article (CrossRef Link)

[14] E. J. Candès, X. Li, Y. Ma and J. Wright, "Robust principal component analysis?," *Journal of the ACM*, vol. 58, no. 3, pp. 11, May, 2011. Article (CrossRef Link)

[15] V. Cevher, A. Sankaranarayanan, M. F. Duarte, D. Reddy, R. G. Baraniuk and R. Chellappa, "Compressive sensing for background subtraction," in *Proc. of Computer Vision-ECCV*, pp. 155-168, Oct., 2008. Article (CrossRef Link)

[16] Z. Cong, W. Xiaogang and C. Wai-Kuen. "Background subtraction via robust dictionary learning," *EURASIP Journal on Image and Video Processing*, 2011. Article (CrossRef Link)

[17] R. Sivalingam, A. D'Souza, M. Bazakos, R. Miezianko, V. Morellas and N. Papanikolopoulos. "Dictionary learning for robust background modeling," in *Proc. of IEEE ICRA*, pp. 4234-4239, May, 2011. Article (CrossRef Link)

[18] F. Luisier, T. Blu and M. Unser, "Image denoising in mixed Poisson-Gaussian noise," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 696-708, Mar., 2011. Article (CrossRef Link)

[19] J. Mairal, F. Bach, J. Ponce and G. Sapiro, "Online learning for matrix factorization and sparse coding," *The Journal of Machine Learning Researc*, vol. 11, pp. 19-60, Mar., 2010. Article (CrossRef Link)

[20] M. Aharon, M. Elad and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no.11, pp. 4311-4322, Nov. 2006. Article (CrossRef Link)

[21] B. Efron, T. Hastie, I. Johnstone and R. Tibshirani, "Least angle regression," *The Annals of statistics*, vol. 32, no. 2, pp. 407-499, 2004. Article (CrossRef Link)

[22] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, Apr., 2004. Article (CrossRef Link)
[23] Dataset available from: http://www.changedetection.net/
[24] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial intelligence* vol. 17, no. 1, pp. 185-203, Nov., 1981. Article (CrossRef Link)
[25] J. J. Koenderink, "Optic flow," *Vision research*, vol. 26, no. 1, pp. 161-179, 1986. Article (CrossRef Link)
[26] L. Li, W. Huang, I Y H Gu and Q Tian, "Statistical modeling of complex backgrounds for foreground object detection," *IEEE Transactions on Image Processing*, vol. 13, no. 11, pp. 1459-1472, Nov., 2004. Article (CrossRef Link)
[27] P. M. Jodoin, M. Mignotte and J. Konrad, "Statistical background subtraction using spatial cues," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 12, pp.1758-1763, Dec., 2007. Article (CrossRef Link)
[28] J. Huang, X. Huang and D. Metaxas, "Learning with dynamic group sparsity," in *Proc. of IEEE ICCV*, pp. 64-71, Sep., 2009. Article (CrossRef Link)
[29] O. Barnich and M. V. Droogenbroeck, "Vibe: A universal background subtraction algorithm for video sequences," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011. Article (CrossRef Link)
[30] B. F. Wu and J. H. Juang, "Real-Time Vehicle Detector with Dynamic Segmentation and Rule-based Tracking Reasoning for Complex Traffic Conditions," *KSII Transactions on Internet & Information Systems*, vol. 5, no. 12, May. 2011. Article (CrossRef Link)
[31] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1168–1177, Jul. 2008. Article (CrossRef Link)

Huaxin Xiao received his BS degree in automation from the University of Electronic Science and Technology of China. He is currently pursuing his MS degree in control science and engineering from the National University of Defense Technology, Changsha, China. His research interests include sparse representation and computer vision.

Yu Liu received his BS degree from Northwestern Polytechnical University, Xi'an, China in 2005. He then received his MSc on image processing and PhD on computer graphics from the University of East Anglia, Norwich, UK, in 2007 and 2011, respectively. He is currently a lecturer in the department of system engineering, National University of Defense Technology. His research interests include image/video processing, computer graphics, and visual haptic technology.

Shuren Tan received the Bachelor, MS and PhD degrees from the Department of System Engineering at the National University of Defense Technology , Changsha, China in 1993, 1996 and 2011, respectively. He is currently an Associate Professor of System Engineering at the National University of Defense Technology. His research interests include computational imaging, computer vision, and signal processing.

Jiang Duan received his BS degree from Southwest Jiaotong University, Chengdu, China in 2002. He then received his PhD on image processing from the University of Nottingham, England, UK in 2006. He is currently a professor in the school of economic information engineering, Southwestern University of Finance and Economics. His research interests include image processing, computer vision, and information engineering.

Maojun Zhang received his BS and PhD degrees in system engineering from National the University of Defense Technology, Changsha, China, in 1992 and 1997, respectively. He is currently a professor in the department of system engineering, National University of Defense Technology. His research interests include computer vision, information system engineering, system simulation, and virtual reality technology.