Joint video frame set division and low-rank decomposition for background subtraction

Jiajun Wen, Yong Xu, *Member, IEEE,* Jinhui Tang, *Member, IEEE,* Yinwei Zhan, *Member, IEEE,* Zhihui Lai, and Xiaotang Guo

***Abstract*—The recently proposed Robust Principle Component Analysis (RPCA) has been successfully applied in background subtraction. However, low-rank decomposition makes sense on the condition that the foreground pixels (sparsity patterns) are uniformly located at the scene, which is not realistic in real-world applications. Our work reconstructs the input video data and aims to make the foreground pixels not only sparse in space but also sparse in "time". Therefore, we propose a Joint Video Frame Set Division and RPCA-based (JVFSD-RPCA) method for background subtraction. In addition, we use the motion as a priori knowledge which has not been considered in the current subspace-based methods. The proposed method consists of two phases. In the first phase, we propose a Lower Bound-based Within-Class Maximum Division (LBWCMD) method to divide the video frame set into several subsets. In this way, the successive frames are assigned to different subsets in which the foregrounds are located at the scene randomly. In the second phase, we augment each subset with the frames with a small quantity of motion. To evaluate the proposed method, the experiments are conducted on real-world and public datasets. The comparisons with the state-of-the-art background subtraction methods validate the superiority of our method.**

***Index Terms*—Within-class maximum division, Motion priori knowledge, Low-rank decomposition, Background subtraction.**

I. Introduction

FOR numerous computer vision tasks, such as indoor surveillance [1], anomaly detection [2], sports video analysis [3], traffic surveillance [4] etc., background subtraction has been a fundamental step to segment out the motion objects for high level vision understanding. Usually, the scene suffers from various influence including lighting changes and dynamic backgrounds. Owing to the complex environment and real-time requirement of the surveillance system, many methods [1], [6]-[9] had been proposed to overcome the above problems. These state-of-the-art methods work well under certain conditions. However, it is not an easy task to handle all the above problems by using a single method. Generally speaking, background subtraction methods can be classified into three categories, i.e. statistic-based methods [5]-[13], classification-based methods [14], [15] and subspace-based methods [1], [16]-[24].

The Gaussian background modeling [5] has been a classical statistic-based method and is very popular in the surveillance system. Later researches developed this model to the multi-Gaussian versions [6]-[9] for background subtraction. Since these methods are of the parametric-based type whose parameters are hard to learn and adjust when it comes to the complex environment, literatures [10], [11] preferred to the non-parametric methods for background pixel modeling. However, these methods cannot well deal with the continuous changing situations. Unlike the background subtraction method at the pixel level, [12] and [13] modeled the background at the region level. The region-based method is capable of handling the noise, illumination variations and dynamic environment. Since background subtraction can be viewed as a classification problem, neural network [14] and support vector machines [15] are also exploited for foreground detection. For these methods, a learning procedure is necessary before the detection stage. They show a good adaptation to the learned situations. However, they are not flexible to the new cases which had not been considered in the learning stage. Recent researches on subspace analysis consider that the background lies in a low dimensional subspace, i.e. eigenspaces. Such eigenspaces enable the algorithms to resist a variety of contamination. Tsai et al. [1] assumed that the background and foreground are two independent signals and extracted the foreground by independent component analysis. Cevher et al. [16] proposed to use compressive sensing to recover the region of interest. However, the theory requires that the foreground occupies a small portion of the scene. Early in 2000, the classical Principal Component Analysis (PCA) had been used in background modeling [17]. But it is vulnerable when the data is contaminated by the noise. Later, Torre et al. [18] extended classical PCA to develop the Robust Principal Component Analysis with M-estimation (RPCA-ME) which is more adaptive to noise corruption, alignment errors and occlusion. Unlike RPCA-ME, the Robust Principal Component Analysis (RPCA) model proposed in [19] treated background subtraction as a matrix decomposition problem. It is reported that RPCA is able to recover the low-rank and sparse components of a data matrix even though a quantity of entries of the matrix are contaminated with arbitrary noise intensity.

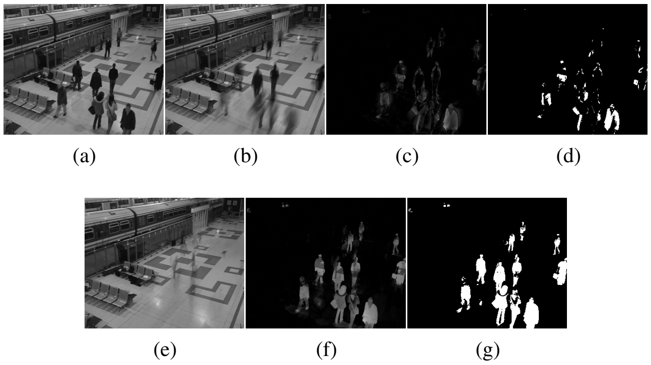
For a measurement matrix $M \in \Re^{n_1 \times n_2}$ with partial

Manuscript received X X, X; revised X X, X. This article is partly supported by 973 Program under Grant 2014CB347600, NSFC under grants Nos. 61370163, 61233011 and 61332011, as well as the Shenzhen Municipal Science and Technology Innovation Council (Nos. CXZZ20120613141657279, JCYJ20120613153352732, JCYJ20130329151843309 and JCYJ20130329151843309).

Jiajun Wen, Yong Xu (corresponding author), Zhihui Lai and Xiaotang Guo are with the Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055, China (e-mail: wenjiajun.hit@gmail.com; yongxu@ymail.com; lai_zhi_hui@163.com; guoxiaotang.hit@gmail.com).

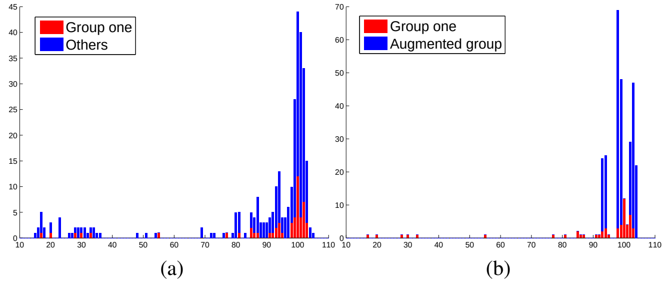
Jinhui Tang is with the School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: tangjh1981@acm.org)

Yinwei Zhan is with the Visual Information Processing R&D Center, Guangdong University of Technology, Guangzhou 510006, China (e-mail: ywzhan@gdut.edu.cn).


(a) (b) (c) (d)

(e) (f) (g)

Fig. 1: Results on a rush hour sequence. (a) is a frame from *Pets 2006 s7*. (b) and (c) are the low-rank and sparse components of RPCA. (d) is the binary image of (c). (e) and (f) are the low-rank and sparse components of the proposed method. (g) is the binary image of (f).



(a) (b)

Fig. 2: The statistical distribution of a pixel values of the sequence used in Fig. 1. (a) shows the statistical distribution of the pixel values. Group one is the statistics of the number of the pixel values from the frame subset obtained by our method. (b) is the statistical distribution of the same pixel values by using our augmented method. The augmented group is the statistics of the number of the pixel values from the augmented set obtained by the method in Section V.

missing entries or being contaminated by noise, there exist many realistic applications that need to recover its original signal and the corresponding noise signal. In other words, we are required to obtain the decomposition $M = L_0 + S_0$, where L_0 is a low-rank matrix, S_0 is a sparse matrix, and both components are allowed to have random intensity.

Let $\sigma_i(M)$ be the i-th singular value of M, $||M||_* := \sum_i \sigma_i(M)$ and $||M||_1 = \sum_{ij} |M_{ij}|$ denote the kernel norm and l_1 norm of matrix M respectively, then under rather weak assumption, RPCA is able to accurately recover low-rank matrix L_0 and sparse matrix S_0 by solving the following mathematical model.

$$\begin{aligned} &\text{minimize} ||L||_* + \lambda ||S||_1 \\ &\text{s. t. } L + S = M \end{aligned} \tag{1}$$

In the application of background subtraction, the foreground and background in each frame are referred to the sparse and low-rank components of RPCA model. As introduced in [19]-[24], RPCA has given a promising result in background subtraction. The online version of RPCA proposed by Qiu et al. [20] enabled the subspace-based background subtraction to be in real time. Mu et al. [21] exploited the random projection and SVD to accelerate the calculation efficiency with controllable loss of the performance. Bao et al. [22] proposed an inductive RPCA to handle gross corruptions and new data efficiently. Ding et al. [23] combined a Bayesian framework with RPCA to broaden the adaptation of the algorithm in a wide range of noise levels. Zhou et al. [24] proposed a 'Go Decomposition' model and used a bilateral random projection technique to acquire the low-rank and sparse components as well as the noise.

It should be noted that the RPCA model makes sense under two assumptions : (1) the genuine signal L_0 is of low-rank but not sparse; (2) the sparsity patterns should be uniformly distributed in the sparse matrix at random [19]. However, such assumptions may not be fully satisfied in real-world applications. With these reasons, in background subtraction, RPCA still has its own shortcomings as follows: (1) It requires the moving objects are uniformly located at random. It fails to detect the objects if they are always located at a limited region of the scene; (2) It does not make use of the motion message. For example, we find out that RPCA is not effective for the rush hour sequence as shown in Fig. 1 (d). Obviously, the rush hour sequence is full of 'noise' (namely the foreground) that largely decreases the quality of recovering the low-rank and sparse components. Considering the values of a pixel on this sequence, we run all the frames and obtain the statistical distribution of the pixel values, see in Fig. 2 (a). The values around 100 give a good estimation of this pixel. However, the values that are far away from 100 come from the moving objects and have side effect on the low-rank decomposition. Our question is: Can we make use of the motion message and change the value distribution of the pixel to the one shown in Fig. 2 (b)? If it works, it will be much easier and more effective to recover the low-rank and sparse components.

In this paper, unlike RPCA that processes the whole sequence for signal recovery, we take advantage of the motion message and devise a new framework to divide the video frames into several groups. In each group, the statistical distribution of pixel values is supposed to be contaminated by less 'noise' than ever before. The idea of the frame set division and frame set augmentation reduces the motion in each subset and provides more genuine background pixels, which facilitates RPCA to obtain better low-rank and sparse components. First, we estimate the position of the moving objects and ratio of the foreground pixel to all the pixels in each frame by using a simple background subtraction technique. Second, we propose a Lower Bound-based Within-Class Maximum Division (LBWCMD) method to divide the video frame set into several subsets based on the position information. Third, the frames will be ranked in ascending order based on the ratio of the foreground pixel to all the pixels and the top several frames will be added in each frame subset to construct the groups. Our results on the rush hour image are shown in Fig. 1 (e)(f)(g). More results are shown in Section VII. To summarize, our contributions are as follows:

TABLE I: Variables of our method

Ω	: Video frame set;
m	: Number of divisions of video frame set;
i	: A subscript, $i = 1, 2, \cdots, m$;
X_i	: The i-th basic set ($X_i \subset \Omega$);
P_i	: The i-th division set with motion ($P_i \subset \Omega$);
Q_i	: The i-th non-motion set ($Q_i \subset \Omega$);
Y_i	: The i-th augmented set with slight motion ($Y_i \subset \Omega$);
Z_i	: The i-th group set, $Z_i = X_i \bigcup Y_i$.

(1) We made use of the motion priori knowledge and joined the video frame set division and low-rank decomposition for background subtraction. The frame set division is beneficial to the recovery of low-rank and sparse components.

(2) A new method called LBWCMD reconstructed the video frame set to obtain several subsets. The proposed method assigns the successive frames to different subsets as much as possible so that the low-rank decomposition is conducted on each frame subset with less motion. In fact, as shown in Fig. 1 (c)(d), matrix M constructed by all the frames with highly dense motion is not good for low-rank decomposition.

(3) We had proposed a framework to accomplish the background subtraction with frame set division and low-rank decomposition. The framework will be stated in detail in the following sections.

The remainder of the paper is organized as follows. Section II gives an overview of the proposed problem and our solution. Section III shows the estimation of priori motion information. Section IV presents the proposed LBWCMD method in detail. Section V illustrates the augmented set construction and shows the implementation of the whole framework. Section VI discusses the parameter tuning. The experimental results are described in Section VII. We make a conclusion of this paper in Section VIII.

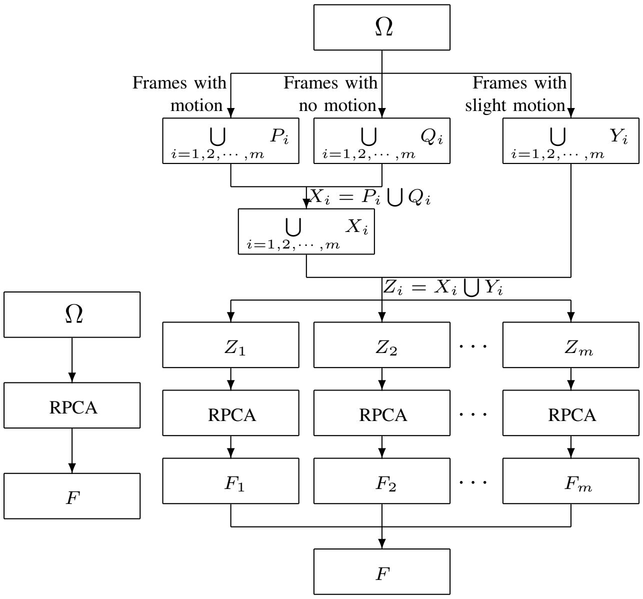
II. Method Overview

A. Problem Statement

The variables that we used in this problem are shown in Table I. For RPCA-based background subtraction [19], [21], [23] the authors stacked all video frames as column vectors one by one into matrix M. Then a low-rank decomposition algorithm is performed to recover the low-rank and sparse components, see the flow chart in Fig. 3 (a). It should be noted that, the highly dense motion appears in a local region has a great impact on the recovery results. It is hard to extract good low-rank and sparse components in such a situation. With this reason, a promising idea is to assign the successive frames to different frame subsets to alleviate such a phenomenon. Our problem is as follows. For a frame set Ω with $|\Omega| = n$, whether there is a finite division to Ω, namely $\Omega = \bigcup_{i=1,2,\cdots,m} X_i$, when RPCA is applied on $X_i (i = 1, 2, \cdots, m)$, the recovery of the low-rank and sparse components is better than that of the low-rank and sparse components decomposed from the frame set Ω. If it exists, how to obtain such a division?

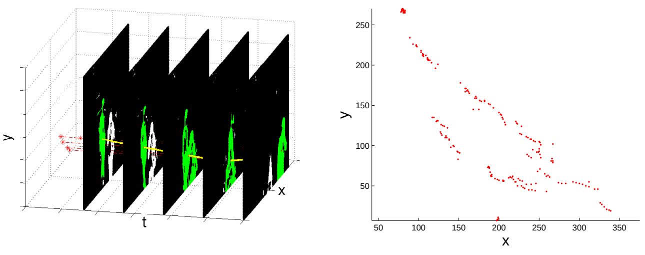
B. The Proposed Method

The theoretical finding in [19] pointed out that low-rank decomposition makes sense on the condition that the sparsity patterns are uniformly distributed in the sparse matrix at random [19]. However, realistic situations may not satisfy this condition. Therefore, when we divide a video frame set, our division principle is to let the moving objects be uniformly located at random in each subset. This is what the proposed LBWCMD (in Section IV) does to satisfy the above condition. It should be noted that if successive frames are assigned to the same subset, the sparsity patterns will be restricted in a limited region of the scene which is not good for low-rank decomposition. To this end, successive frames should be assigned to different subsets. Let's study the following division scheme: $\Omega = \bigcup_{i=1,2,\cdots,m} P_i$, $P_i \bigcap P_j = \varnothing$, $i \neq j$ and each element in P_i is selected from Ω using LBWCMD method. P_i can be regarded as division set. However, the scale of P_i is too small for running RPCA. We propose to make use of the motion priori knowledge to devise a new division scheme to make up such a deficiency, see in Fig. 3 (b). Firstly, we divide Ω into division set P_i and non-motion set Q_i, namely $\Omega = \bigcup_{i=1,2,\cdots,m} P_i \bigcup_{i=1,2,\cdots,m} Q_i$ where P_i consists of the frames with motion and Q_i consists of the frames with no motion. P_i and Q_i can produce a basic set $X_i = P_i \bigcup Q_i$. Then, we extract the frames with slight motion in Ω to obtain the augmented set Y_i, where the elements in Y_i are not included in X_i. Finally, the basic set and the augmented set collaborate together to construct group set $Z_i = X_i \bigcup Y_i$. RPCA will be performed on the group sets to run the recovery task. The division set P_i will be obtained by LBWCMD. Non-motion set Q_i and augmented set Y_i will be obtained by using the methods in Section III and V respectively.


Ω
RPCA
F
Frames with motion
Frames with no motion
Frames with slight motion
$\bigcup_{i=1,2,\cdots,m} P_i$
$\bigcup_{i=1,2,\cdots,m} Q_i$
$\bigcup_{i=1,2,\cdots,m} Y_i$
$X_i = P_i \bigcup Q_i$
$\bigcup_{i=1,2,\cdots,m} X_i$
$Z_i = X_i \bigcup Y_i$
Z_1
Z_2
Z_m
RPCA
F_1
F_2
F_m
F

(a) (b)

Fig. 3: Flow charts of RPCA and the proposed framework for background subtraction. (a) is the flow chart of RPCA for background subtraction. (b) is the flow chart of the proposed framework for background subtraction.



(a) (b)

Fig. 4: Centroids of the largest foreground region. (a) shows the largest foreground region in successive frames. (b) is the projections of the centroids on a plane.

III. Estimation of Priori Motion Information

The motion in the video has a great influence on the quality of low-rank decomposition. For the whole sequence, if the moving objects appear in a local region of the scene, then according to our analysis in Section II-B, the low-rank decomposition results will not be good. This is the reason why we use a new designed strategy to scatter the successive frames to different subsets. Before the division tasks, we need to evaluate the motion priori knowledge in advance. In this section, we estimate and utilize the ratio of the moving area and the centroids of the objects. To this end, a simple background subtraction technique [5] is used to extract the moving information.

We take the centroid of the largest foreground block in each frame as division criterion. The reasons are as follows: (1) The largest block in the foreground reflects the major movement in a frame. Fig. 4 (b) shows the projection of all the centroids of the largest foregrounds in a plane. It gives a clear picture of the motion distribution. (2) The utilization of the centroid of the largest foreground block can accelerate data processing. (3) Adjacent centroids can reveal the continuous movement of the object along the time axis, since the positions of an object in successive frames are normally the nearest.

By performing the simple background subtraction method [5], we can obtain a coarse motion estimation. Fig. 4 (a) shows that the foreground images are in alignment along the time axis. Both the green area and white area are the foregrounds in each frame, whereas the green part is the largest foreground block. The centroid of the largest foreground block can be viewed as a representation of the foreground. It roughly reflects the motion distribution of a video sequence, see in Fig. 4 (b). We summarize the procedure of the motion priori knowledge estimation as follows.

For n video frames $I^{(k)} \in \mathbb{N}^{h\times w}, (k = 1, 2, \cdots, n)$, let $B^{(k)}$, $F^{(k)}$ be the corresponding background and foreground. Let $\alpha \in (0, 1)$ be a weight coefficient.

Firstly, we obtain the foreground of the k-th frame using

$$F^{(k)} = |I^{(k)} - B^{(k-1)}| \tag{2}$$

and thresh the foreground into binary image FG using

$$FG(x,y) = \begin{cases} 255 & F^{(k)}(x,y) > T \\ 0 & F^{(k)}(x,y) \le T \end{cases} \tag{3}$$

TABLE II: Variables of the proposed models

U	: A division which divides the frames into different subsets;
Ψ_k	: The k-th subset of Ω;
A_{ik}^U	: The event that frame I_i is assigned to Ψ_k with division U;
Γ	: A centroid set of Ω;
n	: The number of the centroids of Γ;
x_i	: A centroid of the i-th frame of Ω;
Υ_k	: The k-th centroid set of Ψ_k;
y_{kj}	: The j-th centroid of Υ_k
H_Γ	: The convex set of Γ;
$H_{\Upsilon_k}^U$	: The convex set of Υ_k with division U;
r:	: a radius;
$N_{(x,r)}^U$	: The number of the centroids in a ball with centroid x and radius r.

where T is a predefined threshold. Update the corresponding background using

$$B^{(k)} = \alpha I^{(k)} + (1-\alpha)B^{k-1}. \tag{4}$$

Then we find the largest foreground block through optimizing

$$S = \arg\max_{FA_j} Area(FA_j), \tag{5}$$

where FA_j is the j-th foreground block of FG and $Area(x)$ is the total number of nonzero pixels in x. Calculate the center of S to obtain

$$c = (\frac{\sum xS(p,q)}{\sum S(p,q)}, \frac{\sum yS(p,q)}{\sum S(p,q)}) \tag{6}$$

and update the centroid set, namely

$$\Phi^{(k)} = \Phi^{(k-1)} \bigcup \{c\}. \tag{7}$$

Finally, we obtain the centroid set Φ^* as soon as we go through n video frames.

The basic set $X_i (i = 1, 2, \cdots, m)$ consists of the division set P_i and non-motion set Q_i. The non-motion set Q_i is easily obtained and it is used for reducing the ratio of contamination in the frame set. The reason for using the division set P_i is to scatter the successive frames as much as possible. We will discuss how to obtain P_i in the next section.

IV. LBWCMD Method

A. Model of Frame Set Division

To alleviate the influence of the motion on low-rank decomposition, we consider to reconstruct the frame set Ω to obtain the group sets $Z_i (i = 1, 2, \cdots, m)$ where $\Omega = \bigcup_{i=1,2,\cdots,m} Z_i$. As mentioned in Section II-B, the division set P_i constitutes an important part of the group set Z_i. Since the centroid set Φ of the video has been obtained, we divide the frames to acquire the division set P_i based on the locations of the centroids. To this end, the following two facts should be taken into account:

(1) The successive frames with motion should be assigned to different subsets as much as possible.

(2) The spatial distribution of the centroid set of each frame subset should be consistent with the spatial distribution of the centroid set of Ω.

The former tries to apportion the recovery work to each subset, while the latter guarantees the centroids in each subset come from the whole scene, not a local region of the scene. To start with, we list the variables used in the following models in Table II.

To model the first fact, we use a strategy U to divide the frame set Ω and we have $\Omega = \bigcup_{k=1,2,\cdots,m} \Psi_k$, $\Psi_k \bigcap \Psi_l = \varnothing(k \neq l)$. Then, a probability optimization is used to describe the problem, see the mathematical model (8). In other words, U_{opt} should maximize the combinational probability of A_{ip}^U and A_{jq}^U, namely the probability of successive frames being assigned to different subsets. For our problem, we constrain δ to a small value so as to control the distribution manner of the successive frames. If we set $\delta = 1$, it means the successive frames will be assigned to different subsets as much as possible.

$$U_{opt} = \arg\max_{U} \sum_{\substack{i \neq j \\ i,j \in \{1,2,\cdots,n\}}} P(A_{ip}^U, A_{jq}^U, p \neq q \,\big|\, |i-j| \leq \delta). \quad (8)$$

To model the second fact, not only the local similarity but also the global similarity between the spatial distributions of the centroid set Γ and Υ_k needs to be considered. Therefore, we put forward that the following two goals should be achieved as much as possible.

(a) The convex hull $H_{\Upsilon_k}^U$ of the centroid set Υ_k should occupy almost the same space as the convex hull H_Γ of the centroid set Γ;

(b) The density of the centroid set Υ_k should be almost the same as the density of the centroid set Γ.

For the first goal, our model is as follows. Let $\Gamma = \{x_i, i = 1, 2, \cdots, n\}$ be the centroid set of Ω. With division U, the centroid set Γ is divided into m subsets $\Upsilon_k = \{y_{kj}, j = 1, 2, \cdots, n_k\}(k = 1, 2, \cdots, m)$. That is, $\Gamma = \bigcup_{k=1,2,\cdots,m} \Upsilon_k, \Upsilon_k \bigcap \Upsilon_t = \varnothing(k \neq t), \sum_{k=1}^m n_k = n$. The convex hulls of Γ and its subset Υ_k are

$$H_\Gamma = \{\sum_{i=1}^n \eta_i x_i | x_i \in \Gamma, \sum_{i=1}^n \eta_i = 1, \eta_i \in [0,1]\}, \quad (9)$$

and

$$H_{\Upsilon_k}^U = \{\sum_{j=1}^{n_k} \xi_j y_{kj} | y_{kj} \in \Upsilon_k, \sum_{j=1}^{n_k} \xi_j = 1, \xi_j \in [0,1]\}. \quad (10)$$

Therefore, the first goal proposed for the second fact seeks the following optimization

$$U_{opt} = \arg\max_U \sum_{k=1}^m \frac{|H_\Gamma \bigcap H_{\Upsilon_k}^U|}{|H_\Gamma|}, \quad (11)$$

That is, U_{opt} should maximize the intersection of H_Γ and $H_{\Upsilon_k}^U$. However, (11) only measures a global property of the similarity between the centroid sets.

Our second goal serves as a complementary of the problem. Since Υ_k is a subset of Γ, we can find the centroids which simultaneous belong to Γ and Υ_k. Therefore, we have $x'_j = y_{kj}$, $x'_j \in \Gamma(j = 1, 2, \cdots, n_k)$. In Γ, we center a ball with radius r on x'_j, the number of the centroids in this ball are regarded as $N_{(x'_j,r)}^U$. Similarly, in Υ_k, the number of the centroids in the ball centered at y_{kj} with radius r is $N_{(y_{kj},r)}^U$. The second goal pursuits that the ratio $N_{(x'_j,r)}^U/|\Gamma|$ should be approximately equal to the ratio $N_{(y_{kj},r)}^U/|\Upsilon_k|$ as much as possible. Therefore, the second goal tries to optimize the following mathematical problem

$$U_{opt} = \arg\min_U \sum_{k=1}^m \sum_{j=1}^{n_k} \left| N_{(x'_j,r)}^U/|\Gamma| - N_{(y_{kj},r)}^U/|\Upsilon_k| \right|. \quad (12)$$

Though the total number of elements in Υ_k is smaller than that in Γ, the difference between the ratios can be as small as possible.

We incorporate global and local measurements into the same objective function. Then the mathematical model of the second fact is

$$U_{opt} = \arg\min_U \frac{\sum\limits_{k=1}^m \sum\limits_{j=1}^{n_k} \left| N_{(x'_j,r)}^U/|\Gamma| - N_{(y_{kj},r)}^U/|\Upsilon_k| \right|}{\sum\limits_{k=1}^m |H_\Gamma \bigcap H_{\Upsilon_k}^U|/|H_\Gamma|}. \quad (13)$$

The objective functions of (8) and (13) are difficult to optimize directly. However, this problem can be regarded as an optimization problem of the discrete centroids. Therefore, it is of importance to evaluate the relations between the centroids of a subset and the relations between the centroids of different subsets.

B. The Proposed Solution - LBWCMD

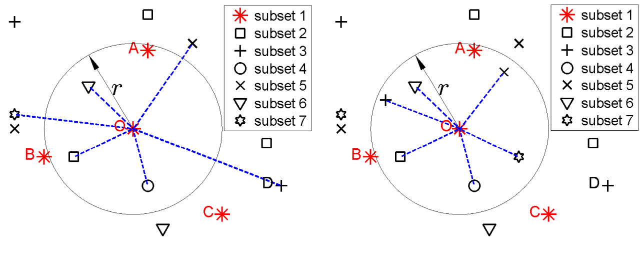
Based on the two facts we have mentioned, actually our problem is to maximize the distances between the centroids in a subset and minimize the distances between the centroids in different subsets. This problem is quite different from the clustering problems [26] we have seen commonly. In this paper, we propose a lower bound based measurement to describe the distance between the two centroids from the same subset or different subsets so as to obtain a reasonable solution of the problem. For the well known k-mean method [26], to accomplish the clustering task, it needs to answer the following two questions after the initialization. (a) Which sample in the class should be removed? (b) Which class should the removed sample be added to? Follow the same procedures as k-mean but rather different intrinsic principles, we demonstrate the proposed method as follows. First, we initialize m subsets of the centroid set Γ. Then, for the centroids in a subset, we have to tackle with the following two problems. (1) Which centroid should be the candidate that has to be transferred to another subset? (2) Which subset should the centroid be assigned to when the centroid has to be transferred?

For the first problem, we propose to use a ball centered at a centroid with radius r to determine whether this centroid should be excluded from the subset. Based on the first fact, adjacent centroids should be assigned to different subsets. Let n_i be the number of the centroids in the i-th subset $(i = 1, 2, \cdots, m)$. Let $x_{ij} = (x_{ij}^1, x_{ij}^2)$ be the j-th centroid in the i-th subset, $j = 1, 2, \cdots, n_i$. Define a ball set $\Lambda_{ij} = \{b \,\big|\, (b_1 - x_{ij}^1)^2 + (b_2 - x_{ij}^2)^2 \leq r^2\}$ for x_{ij}, where $b = (b_1, b_2)$ is a centroid inside the ball. A centroid x_{ij} should be removed only if the following two restrictions are satisfied.

(1) The first restriction is $\{x_{iq}\} \neq \varnothing$, $x_{iq} \in \Lambda_{ij}$ and $q \neq j$.

(2) The second restriction is $|\{x'_{pq}\}| < m - 1$, $x'_{pq} \in \Lambda_{ij}$,

$$x'_{pq} = \arg\min_q \|x_{ij} - x_{pq}\|, \quad (14)$$



(a) (b)

Fig. 5: The centroids are distributed to seven subsets. (a) shows the r-ball of centroid O contains centroid A from the same subset. (b) shows the situation which meets the first restriction but not the second one, namely r-ball of centroid O contains the centroids from all subsets.

Algorithm 1 Lower bound based within-class maximum division (LBWCMD)

```
Input: $X \in \mathbb{N}^{v \times n}$, number of subsets $m$, lower bound $r$, initial values for number of iterations $u = 0$, thresh $T = 20$, flag $b = 0$, constant $\tau = 5$
 1: Initialization:
      $$N_t(i) \leftarrow \begin{cases} 1 & i = 1, 2, \cdots, m-1 \\ n - (m-1) & i = m \end{cases},$$
      $$L(j) \leftarrow \begin{cases} j & j = 1, 2, \cdots, m-1 \\ m & j = m, m+1, \cdots, n \end{cases}.$$
 2: while $b$ equals to 0 and $u < T$ do
 3:   for $i = 1, 2, \cdots, m$ do
 4:     if $N_t(i) \leq 1$ then continue. end if
 5:     Extract samples with label $i$ from $X$ to obtain $Y_i$.
 6:     $N_{Y_i} = N_t(i)$;
 7:     for $j = 1, 2, \cdots, N_{Y_i}$ do
 8:       Extract the $j$-th sample from $Y_i$ to obtain $x_j$.
 9:       $d = \arg\min_{d_p} d_p$, s.t. $d_p = \|x_j - x_p\|$, $p = 1, 2, \cdots, N_{Y_i}$ and $p \neq j$.
10:       if $d \leq r$ then
11:         $d_c(k) = \arg\min_{d_k} d_k$, s.t. $d_k = \|x_j - y_{kq}\|, y_{kq}$ is the $q$-th sample from the $k$-th subset, $k = 1, 2, \cdots, m$, $k \neq i$.
12:         $s = \sum_{k=1,2,\cdots,i-1,i+1,\cdots,m} \delta(r - d_c(k))$,
              $$\delta(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}.$$
13:         $l = \arg\max_k d_c(k), k = 1, 2, \cdots, m$ and $k \neq i$.
14:         if $s < m - 1$ or $|N_{Y_i} - N_{Y_l}| < \tau$ then
15:           $N_t(i) \leftarrow N_t(i) - 1$.
16:           $L(j) \leftarrow l$.
17:           $N_t(l) \leftarrow N_t(l) + 1$.
18:         end if
19:       end if
20:     end for
21:   end for
22:   $d_g(k) = \arg\min_{d_k} d_k, d_k = \|y_{kp} - y_{kq}\|, y_{kj}$ is the $j$-th sample from the $k$-th subset, $k = 1, 2, \cdots, m$, $p, q = 1, 2, \cdots, N_t(k)$ and $p \neq q$.
23:   $s = \sum_{k=1,2,\cdots,m} \delta(d_g(k) - r)$.
24:   if $s$ is equal to $m$ then $b = 1$. end if
25:   $u \leftarrow u + 1$.
26: end while
Output: $N_t, L$.
```

and $p \neq i$.

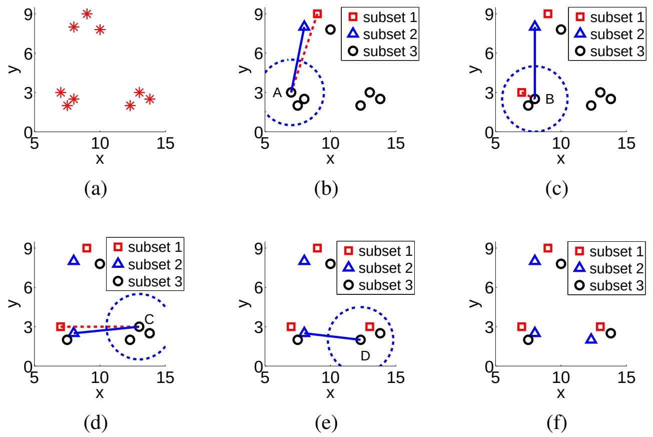
The first restriction shows that some centroids that are from the same subset as x_{ij} fall into r-ball of x_{ij}. Therefore, x_{ij} should be the candidate to be removed so as to maintain the space structure of the centroids in the i-th subset, that is, no more than two centroids from the same subset can coexist in a r-ball. However, the second restriction guarantees that at least one subset is available for x_{ij} to be transferred to, meanwhile the space structure of the centroids in this subset is well maintained. Optimization problem of (14) obtains all the centroids that are the nearest to x_{ij} from other subsets. The second restriction shows that these centroids are in the r-ball of x_{ij}. However, if all the subsets have the centroids (the nearest one to x_{ij} in each subset) in the r-ball of x_{ij}, namely $\{\{x'_{pq}, p = 1, 2, \cdots, m\} \bigcup x_{ij}\} \subset \Lambda_{ij}$, it is not necessary to transfer centroid x_{ij} to another subset. Because, in such a situation, two centroids of a subset will be in a r-ball. Therefore, we need additional restriction $|\{x'_{pq}\}| < m - 1$ and $p \neq i$. In Fig. 5 (a), seven subsets have been initialized. The r-ball of centroid O contains centroid A that comes from the same subset as centroid O. So centroid O is the candidate to be removed. In addition, subsets 1, 2, 4 and 6 have the centroids in r-ball of centroid O and not less than three subsets have no centroid in the r-ball. Hence, centroid O should be removed from subset 1. However, Fig. 5 (b) gives a situation which meets the first restriction but not the second one. Therefore, in such a situation, centroid O should not be removed.

For the second problem, based on the second fact, we need to guarantee the motion of each subset comes from the whole scene. Therefore, for the centroids in a subset, the within-class distances should be maximized. We propose to assign centroid x_{ij} to the p'-th subset with the following condition, $\{x_{p,q}\} \bigcap \Lambda_{i,j} = \varnothing$,

$$p' = \arg\max_{p} \min_{q} \|x_{i,j} - x_{p,q}\|, \tag{15}$$

and $p \neq i$. The formulas $\{x_{p,q}\} \bigcap \Lambda_{i,j} = \varnothing$ and $p \neq i$ try to exclude the existing subsets in which the centroids fall into r-ball of $x_{i,j}$. In order to determine an optimal subset that the selected centroid x_{ij} should be assigned to, we consider the remaining subsets which contain the centroids outside r-ball of $x_{i,j}$. We rank these subsets in descending order based on the distances between $x_{i,j}$ and each subset. The top ranking subset, namely the p'-th subset, is the best selection. The dashed lines in Fig. 5 (a) show the nearest distances between centroid O and other subsets. Among all the nearest distances, the one between centroid O and centroid D is the largest. Based on the within-class maximum principle, centroid O should be assigned to subset 3. The details of the method are described in Algorithm 1. The proposed LBWCMD method can assign successive frames to different subsets by a lower bound strategy which gives an approximate solution to (8). Moreover, this method ensures the spatial distribution of each centroid subset meets (13) as much as possible.

We now give a concise demo to demonstrate the main steps of LBWCMD method. Fig. 6 (a) shows several centroids for division. Let $m = 3$, $r = 2.5$. According to step 1 of LBWCMD, the i-th centroid is assigned to the i-th division, $i = 1, 2, \cdots, m-1$, and the rest of the centroids are assigned to the m-th subset as shown in Fig. 6 (b). Suppose that the



(a) (b) (c)

(d) (e) (f)

Fig. 6: A concise demo of LBWCMD. (a) is the original distribution of the centroids. We first initialize the states of the centroids as shown in (b). According to LBWCMD, centroids A, B, C and D become square, triangle, square and triangle markers respectively as shown in (c), (d), (e) and (f).

centroids with the square marker are from the first subset, the ones with the triangle marker are from the second subset and the ones with the circle marker are from the third subset. According to step 8 to 19 of Algorithm 1, centroid A is assigned to the farthest subset outside r-bound of centroid A. As shown by the dash line in Fig. 6 (b), centroid A should be assigned to the first subset. Then, we go through centroid B, C and D based on the same idea. Similarly, the division results are shown in Fig. 6 (c)(d)(e)(f) respectively. Fig. 6 (f) shows that the proposed method not only separates adjacent centroids to different subsets, but also enable each subset to maintain almost the same spacial distribution as the original centroid set.

By using the proposed LBWCMD method, we can map the corresponding frames to the division set $P_i (i = 1, 2, \cdots, m)$ based on the already known centroid subsets.

V. Augmented Set Construction

Although we have finished the construction of the basic sets. It is not yet ready to carry on RPCA on these basic sets. First, the scale of each basic set is much smaller than ever before, which results in insufficient number of frames for the low-rank decomposition. Second, there are still more possibilities for us to make full use of the motion priori knowledge. In this section, we propose to construct an augmented set as a supplementation for each basic set. To this end, the frames with a small quantity of motion will be selected. Let s_j be the area of the binary foreground f_j of the j-th frame, $j = 1, 2, \cdots, n$. Based on s_j, we sort the foreground in ascending order to obtain $\{f_{k_j} | s_{k_1} \leq s_{k_2} \leq \cdots \leq s_{k_n}, j = 1, 2, \cdots, n\}$, where the top $\lfloor \alpha \times n \rfloor$ frames are selected to construct set $M = \{f_{k_j} | j = 1, 2, \cdots, \lfloor \alpha \times n \rfloor\}$ that composed of the frames with a small quantity of motion. Let Y_i be the augmented set which is corresponding to basic set X_i, $i = 1, 2, \cdots, m$. Y_i includes the frames in M but not in X_i. Therefore, Y_i acts as a supplementation to the basic set. The details for constructing the augmented sets are summarized in Algorithm 2.

```
Algorithm 2 Augmented set construction (ASC)
Input: Basic set $X_i$, foreground set $\{f_j\}$ and its corresponding area $s_j, j = 1, 2, \cdots, n$, initial values for $\alpha, Y_i = \varnothing, i = 1, 2, \cdots, m$.
Sort $\{f_j\}$ based on $s_j$ in ascending order to obtain
  $\{f_{k_j} | s_{k_1} \leq s_{k_2} \leq \cdots \leq s_{k_n}, j = 1, 2, \cdots, n\}$.
Select the top $\lfloor \alpha \times n \rfloor$ frames to construct set $M = \{f_{k_j}\}$,
  $j = 1, 2, \cdots, \lfloor \alpha \times n \rfloor$.
for $j = 1, 2, \cdots, \lfloor \alpha \times n \rfloor$ do
  for $i = 1, 2, \cdots, m$ do
    if $f_{k_j} \notin X_i$ then
      $Y_i \leftarrow Y_i \bigcup \{f_{k_j}\}$.
    end if
  end for
end for
Output: $Y_i, i = 1, 2, \cdots, m$.
```

```
Algorithm 3 Joint Video Frame Set Division and RPCA-based background subtraction (JVFSD-RPCA)
Input: $n$ video frames $I^{(k)} \in \mathbb{N}^{h \times w}$, number of partitions $m$, lower bound $r$, initial values for $\alpha$.
1: Use the proposed method in Section III to obtain centroid set $\Phi$
   which consists of the centroids of major moving objects in all
   frames and set $Q$ with non-motion frames.
2: Put the centroids in $\Phi$ as column vectors of data matrix $X$. Take
   $X$ as input for Algorithm 1 (LBWCMD) to assign the centroids to
   $m$ divisions. Then we have division set $P_i, i = 1, 2, \cdots, m$.
3: Divide $Q$ into $m$ subsets $Q_i, i = 1, 2, \cdots, m$. Merge $P_i$ and $Q_i$
   to obtain basic set, namely $X_i = P_i \bigcup Q_i$. The index of frames in
   $X_i$ is $\{l_{ij} \,|\, j = 1, 2, \cdots, |X_i|;\ i = 1, 2, \cdots, m\}$.
4: Use Algorithm 2 (ASC) to construct the augmented set $Y_i$ and
   merge it with the basic set $X_i$ to obtain the group set $Z_i = X_i \bigcup Y_i$.
5: Run RPCA on each group set $Z_i$ to isolate the corresponding
   background and foreground for each frame.
Output: Background set $\{B_{l_{ij}}\}$ and foreground set $\{F_{l_{ij}}\}, j = 1, 2, \cdots, |X_i|;\ i = 1, 2, \cdots, m$.
```

Ultimately, we divide Ω into $Z_1, Z_2, \cdots, Z_m$, namely $\Omega = \bigcup_{i=1,2,\cdots,m} Z_i$, where $Z_i = X_i \bigcup Y_i$. The full name of the whole framework is Joint Video Frame Set Division and RPCA-based background subtraction (JVFSD-RPCA) which is described in Algorithm 3.

VI. Parameter Tuning of α

The proposed LBWCMD consists of three parameters, namely m, r and α. m indicates the number of divisions of video frame set. r describes the distance between the locations of two people. We will consider the values of these two parameters in the next Section. For the third parameter α, it reflects the number of frames with slight motion. The larger α is, the more these types of frames exist. Our rule for setting the value of this parameter is as follows. Define a step function

$$g(x) = \begin{cases} 1 & , x < 0 \\ 0 & , x \geq 0. \end{cases} \tag{16}$$

Let $\frac{Area(F^{(i)})}{wh}$ stand for the ratio of the foreground pixel to all the pixels in the frame, where $F^{(i)}$ is the i-th foreground which is estimated by the method used in Section III. Therefore, the ratio of the number of these types of frames to n is

$$f(x) = \frac{1}{n} \sum_{i=1}^{n} \left[g(\frac{Area(F^{(i)})}{wh} - \epsilon) \right]. \tag{17}$$

TABLE III: Parameter setting for the point set division

Point set	Value of m	Value of r
Decentralized assembled points	6	70
Straight walking points	4	3
Random walking points	4	5
Uniformly distributed points	6	28

Then, we define

$$\alpha = \begin{cases} f(x) & , f(x) > \delta \\ \delta & , f(x) \leq \delta. \end{cases} \tag{18}$$

It means that the proposed method pursues the frames with slight motion as much as possible.

VII. Experimental Results

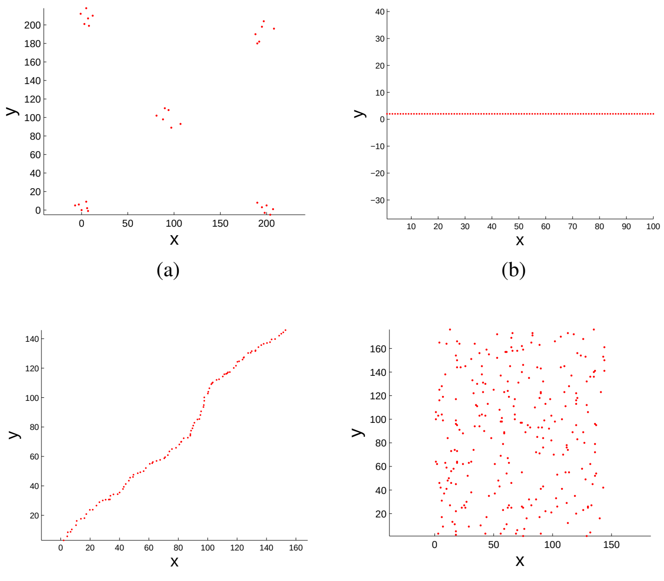
In this section, two groups of experiments were used to evaluate the effectiveness of JVFSD-RPCA and a detailed complexity analysis of the proposed method was given. Since the LBWCMD method is the core of our method, in the first group of experiment, we created four kinds of artificial point sets to test the performance of LBWCMD. In the second group of experiment, we compared JVFSD-RPCA with the state-of-the-art subspace-based background subtraction methods, including Robust Principle Component Analysis (RPCA) [19], Go Decomposition (GoDec) [24], Principle Component Analysis (PCA) [17] and Robust Principle Component Analysis with M-Estimation (RPCA-ME) [18]. Moreover, we also compared the proposed method with the statistic-based, classification-based and non-parametric methods, including Mahalanobis Distance (MD) [27], Improved Gaussian Mixture Model (G-MM) [28], Self-Organizing-based method (SOBS) [14] and Kernel Density Estimation (KDE) [10]. All the comparisons will be conducted on real-world and public sequences.

A. Experiments on Artificial Point Sets

For a certain point set, the proposed LBWCMD method can maximize the within-class distances while maintaining the spatial distribution in each subset. In this subsection, we created four artificial point sets to validate it. The point sets include the decentralized assembled points, straight walking points, random walking points and uniformly distributed points, which are shown in Fig. 7 (a)(b)(c)(d). We also used the objective function values of (11), (12) and (13) to evaluate the performance of LBWCMD.

The lower bound r plays an important role in LBWCMD. It decides the minimum distance between the two points from the same subset. If r is too small, the points cannot be assigned to different subsets as much as possible. On the contrary, if r is too big, there will be an unbalance number of the points in the subsets. Therefore, a proper value of r should be used. In this experiment, we used the parameters in Table III to divide the point sets.

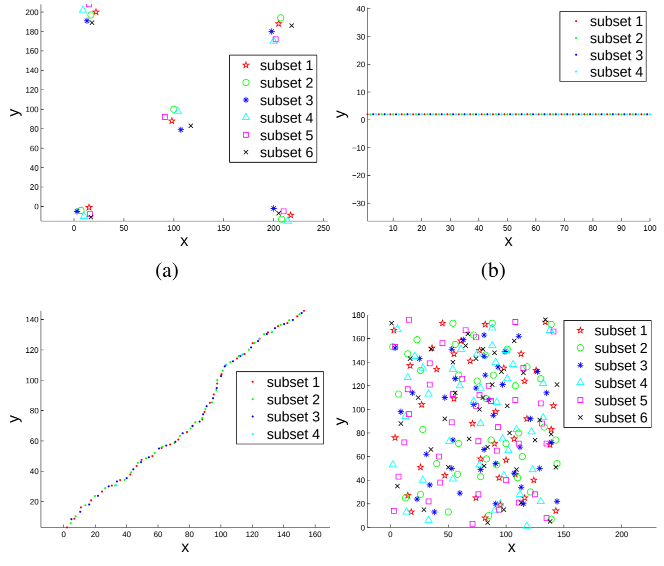
The division results are shown in Fig. 8. Obviously, the adjacent points can be assigned to different subsets while the spatial distribution of each subset is well maintained. We can see some numerical results in Table IV. It shows that the convex hull of each subset largely overlaps with that of the original point set except for subset 4 of the random walking point set. Because some of the points in the random walking point set are not well aligned. In addition to the global similarity, the local similarity related to the point sets are well obtained. A comprehensive evaluation of the effectiveness of LBWCMD is shown in the last row in Table IV. The smaller



(a) (b)

(c) (d)

Fig. 7: Artificial point sets for division. (a) shows the decentralized assembled points. (b) shows the straight walking points. (c) shows the random walking points. (d) shows the uniformly distributed points.



(a) (b)

(c) (d)

Fig. 8: Results of LBWCMD on the point sets. The division results of decentralized points, straight walking points, random walking points and uniformly distributed points are shown in (a), (b), (c) and (d) respectively.

TABLE IV: The effectiveness of LBWCMD on the point sets

		Decentralized assembled points	Straight walking points	Random walking points	Uniformly distributed points
Global similarity in (11) for all subsets	subset 1	84.90%	97.00%	91.75%	84.11%
	subset 2	88.49%	97.00%	86.31%	89.13%
	subset 3	80.12%	97.00%	85.70%	77.90%
	subset 4	86.63%	97.00%	24.14%	85.82%
	subset 5	82.87%	╲	╲	87.80%
	subset 6	84.54%	╲	╲	86.55%
Local similarity in (12)		0	2.88	2.19	6.71
Objective function value of (13)		0	0.74	0.76	1.31

the objective function value of (13) is, the better subsets we had obtained. Among all the objective function values, the one of the uniformly distributed point set is the worst. The main reason is that some badly aligned points are not well handled. Nevertheless, it is concluded from the overall results that the proposed method is feasible to make a good division to the point sets.

B. Experiments on video datasets

In this subsection, we conducted the experiments on 11 real-world and public sequences. The sequences include three challenging ones (*Fighting* and two *Walking with occlusion* sequences) captured by ourselves, *Intelligent room* [29], three sequences (*WaterSurface*, *Meeting room*, *Switch light*) from Li's Dataset [37], two sequences (*Highway*, *Pedestrians*) from Change Detection Dataset [31], *Dance bootstrapping* from Competition dataset [32] and *Pets 2006 S7* [25]. We compared JVFSD-RPCA with RPCA [19], GoDec [24], PCA [17], RPCA-ME [18], MD [27], SOBS [14], GMM [28] and KDE [10]. The source codes of RPCA, GoDec and RPCA-ME can be downloaded at [33], [34] and [35] respectively. [17] did not provide the source code of PCA for background subtraction. Hence, we implemented this code by ourselves. The source codes of MD, SOBS, GMM and KDE can be downloaded at the website of BGSlibrary [36].

In these experiments, all the ground truths of the frames were used to evaluate the performances of the methods. If the detected regions are included in the ground truth, they are true positive, otherwise they are false positive. We adopted the similarity measurement used in [37] to evaluate the performances of the methods. Let D be a detected region and G be the ground truth on the corresponding frame. The similarity measurement between D and G is defined as

$$S(D,G) = \frac{D\bigcap G}{D\bigcup G}. \tag{19}$$

If the detected foreground is exactly the same as the ground truth, the similarity approaches to 1. On the contrary, if the detected foreground has no overlap with the ground truth, the similarity approaches to 0. Therefore, we can conveniently evaluate the performance of JVFSD-RPCA and other methods with similarity measurement (19).

1) Parameters setting: For our method, three parameters need to be determined, namely m, r and α. We used $m = 10$ and $r = 6$ for all the sequences. α can be determined by (18) with empirical values $\epsilon = 0.01$ and $\delta = 0.05$. For RPCA, as mentioned in [19], the suggested value of λ is $1/\sqrt{\max(wh, n)}$, where w and h are the width and height of the frame. However, we discovered that the suggestion does not work in our experiments. To this end, we multiplied λ by a weight ρ so as to make a proper balance between the low-rank and sparse components. Hence we obtained λ by

$$\lambda = \frac{\rho}{\sqrt{\max(wh, n)}}, \tag{20}$$

where $\rho = 0.06$. For GoDec, we set the rank of measurement matrix to 2 and used the default values of the iteration parameters in the algorithm. For RPCA-ME and PCA, the number of principle components were set to 10. Since these subspace-based methods are of the batch types for dealing with the coming frames, we used every 300 frames to construct the measurement matrix for background subtraction. After the foreground had been extracted by the subspace-based methods, a threshold of 25 was adopted to convert the results into binary images. For MD method, the sensitivity, noise variance and learning rate were set to 100, 150 and 30, respectively. For SOBS method, the training sensitivity, learning rate in training phase and training steps were set to 245, 255 and 55 while the testing sensitivity and learning rate in testing phase were set to 130 and 62, respectively. For GMM method, three Gaussian models were used and the learning rate was set to 0.008. For KDE method, the window size was set to 100.

2) Experiments on Different Environments: In this subsection, we compared the performances of the methods mentioned above on different environments including four indoor sequences (*Fighting*, *Intelligent room*, *Meeting room* and *Dance bootstrapping*) and two outdoor sequences (*WaterSurface* and *Highway*). The *Fighting* sequence was captured by ourselves with the length of 300. We labeled the ground truths of the frames on every 10 frames. The resolution with 352×288 was used for the experiment. The *Intelligent room* sequence has a length of 300 and a resolution of 320×240. According to [29], the frames range from 82 to 299 are provided with ground truths when the person starts to walk into the room. The *WaterSurface* and *Meeting room* sequences contain dynamic backgrounds. The former is of the length of 633 and the latter is of the length of 2964. Both sequences are in the resolution of 160×128. According to [37], for each sequence, 20 ground truths are provided on the key frames to evaluate the performances of the algorithms. The *Highway* sequence has a length of 1700 and a resolution of 320×240. The ground truths are provided in the range of 470 to 1700. For the *Dance bootstrapping* sequence, it has a length of 747 with the resolution of 384×240. All the ground truths of the frames are provided. We ran 9 algorithms on 6 sequences. The results of the average similarities on the sequences are shown in Table V. The results in bold font highlight the highest average similarity among all the methods on the same sequence. The proposed JVFSD-RPCA outperforms other methods on four sequences. In another two sequences, the competitiveness of our method

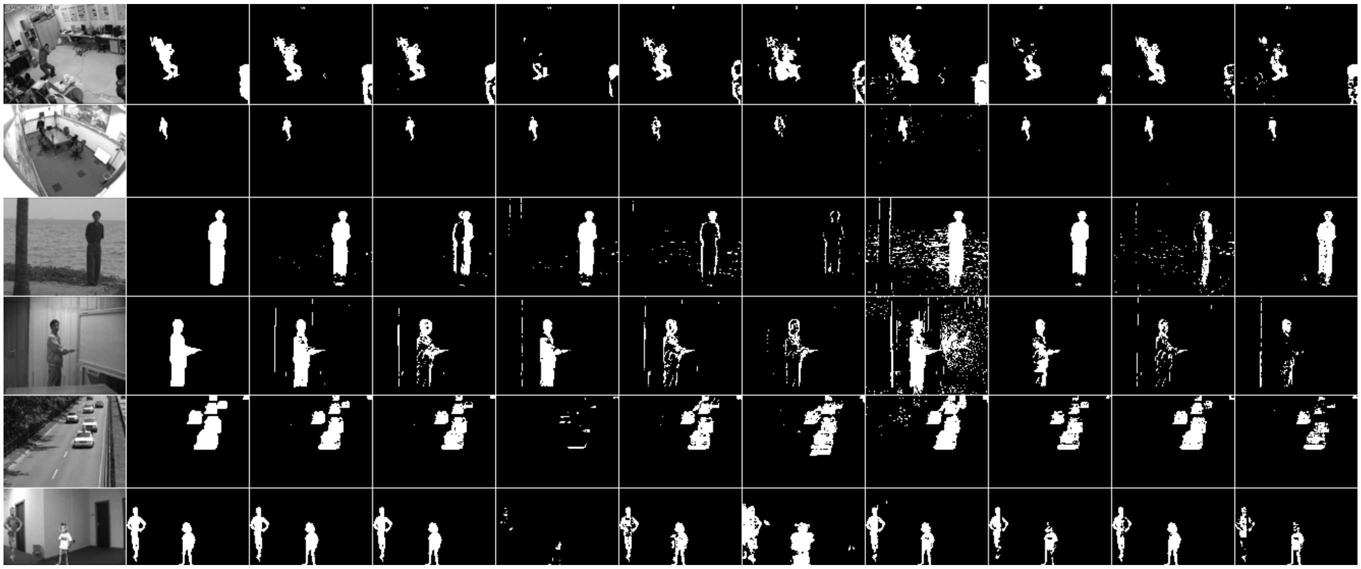


Fig. 9: Comparison results on different sequences. The first to the sixth rows show the detection results on 221-th frame of *Fighting*, 216-th frame of *Intelligent room*, 1615-th frame of *WaterSurface*, 23857-th frame of *Meeting room*, 1615-th frame of *Highway* and 338-th frame of *Dance bootstrapping* sequences. The first and the second columns shows the original image and its ground truth. The third to the eleventh columns show the results of JVFSD-RPCA, RPCA, GoDec, RPCA-ME, PCA, MD, SOBS, GMM and KDE respectively.

TABLE V: Average similarities (%) on different sequences

Sequence	JVFSD-RPCA	RPCA	GoDec	RPCA-ME	PCA	MD	SOBS	GMM	KDE
Fighting	**57.26**	48.27	17.32	41.37	37.69	35.74	22.61	36.04	42.09
Intelligent room	**64.33**	63.38	61.76	45.27	26.96	28.31	63.92	31.67	63.17
WaterSurface	**76.84**	32.08	76.12	20.70	9.18	50.49	76.46	32.13	74.67
Meeting room	70.40	42.00	**71.91**	37.47	22.37	45.62	51.36	38.75	41.90
Highway	**67.76**	64.60	15.17	54.25	48.06	46.07	53.98	67.10	47.10
Dance bootstrapping	89.09	**89.22**	15.57	74.56	46.61	79.70	51.17	76.06	46.79

is almost the same as GoDec on *Meeting room* sequence and RPCA on *Dance bootstrapping* sequence. Some examples of the foreground extraction results are shown in Fig. 9. The second column shows the ground truth of the corresponding frame. Our result is shown in the third column. We can see that the proposed method performs well on all the sequences. However, some of the algorithms, like RPCA-ME, PCA and GMM do not work well on *WaterSurface* and *Meeting room* sequences. RPCA works well on most of the sequences. However, if the moving objects stay for a while in a fixed location of the scene, RPCA can't extract the entire foreground of the object. Differ from RPCA, the proposed JVFSD-RPCA pursuits the sparsity not only in space but also in "time" and exploits the frames with slight motion to provide more genuine backgorund pixels for low-rank decomposition. With these situations, JVFSD-RPCA is much more competitive than RPCA. Form the eighth and eleven column of Fig. 9, we can find that the MD method is quite sensitive to the noises and KDE cannot perform well on *Meeting room*, *Highway* and *Dance bootstrapping*.

3) Experiments on strong lighting sequences: In this subsection, we evaluated the algorithms under strong lighting situations. Two sequences were used for testing. One is an outdoor *Pedestrians* sequence with strong lighting. The sequence has a length of 1099 and a resolution of 360×240. The other is an indoor *Switch light* sequence which first turns off the light and then turns on the light causing strong lighting changes. The sequence is with the length of 1546 and resolution of 160×128. *Pedestrians* sequence provides the ground truths in the range of 300 to 1099 and *Switch light* sequence provides 20 ground truths of the key frames for performance evaluation according to [37]. We performed all the algorithms on the two sequences. The average similarities results are shown in Table VI. Our method obtains better foreground extraction results than the other methods. See the second row in Fig. 10, RPCA-ME, PCA, SOBS and GMM do not obtain good results under the situation of *Switch light*. MD and KDE are much more sensitive to lighting changes comparing with other methods.

4) Experiments on challenging sequences with serious occlusion: To further evaluate the performances of all the algorithms, we used three challenging sequences for the experiments. Two *Walking with occlusion* sequences were captured by ourselves. Both sequences are of the length of 400 and have a resolution of 352×288. We labeled the ground truths on every 10 frames. The *Walking I* and *Walking II* sequences are shown in the first and second rows of Fig. 11. The third sequence is *Pets 2006 S7* with a length of 1200 and a resolution of 360×288. This sequence is the most challenging one among all the seven sequences in Pets 2006 Dataset. Many people appear in the scene. Moreover, the occlusion between

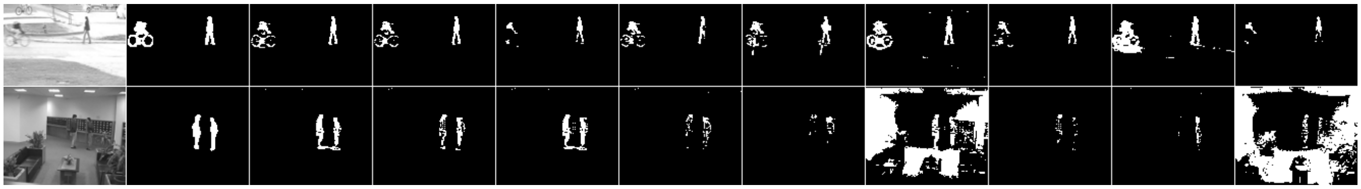


Fig. 10: Comparison results on strong lighting sequences. The first to the second rows show the detection results on 467-th frame of *Pedestrians* and 2507-th frame of *Switch light* sequences. The first and the second columns shows the original image and its ground truth. The third to the eleventh columns show the results of JVFSD-RPCA, RPCA, GoDec, RPCA-ME, PCA, MD, SOBS, GMM and KDE respectively.

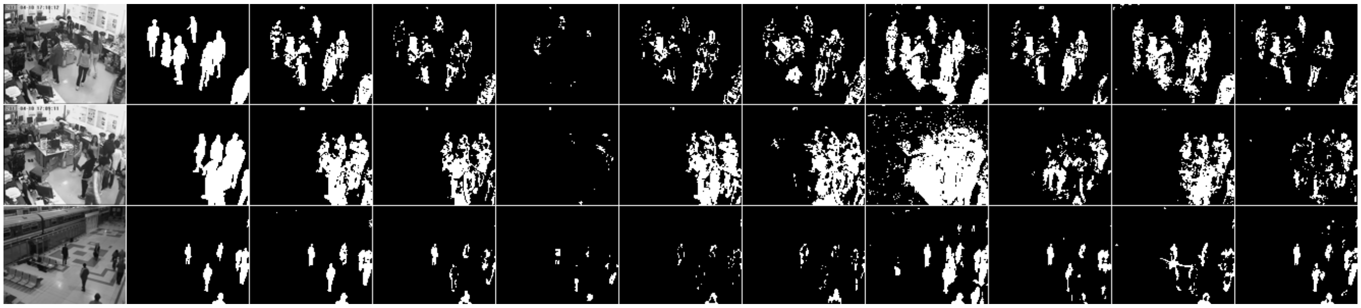


Fig. 11: Comparison results on challenging sequences with serious occlusion. The first to the third rows show the detection results on 211-th frame of *Walking I*, 171-th frame of *Walking II* and 841-th frame of *Pets 2006 S7* sequences. The first and the second columns shows the original image and its ground truth. The third to the eleventh columns show the results of JVFSD-RPCA, RPCA, GoDec, RPCA-ME, PCA, MD, SOBS, GMM and KDE respectively.

TABLE VI: Average similarities (%) on strong lighting sequences

Sequence	JVFSD-RPCA	RPCA	GoDec	RPCA-ME	PCA	MD	SOBS	GMM	KDE
Pedestrians	**73.66**	73.53	52.53	72.77	58.41	65.25	66.14	72.07	48.84
Switch light	**61.68**	57.23	59.32	38.23	29.23	11.09	23.14	36.20	8.36

TABLE VII: Average similarities (%) on challenging sequences

Sequence	JVFSD-RPCA	RPCA	GoDec	RPCA-ME	PCA	MD	SOBS	GMM	KDE
Walking I	**66.45**	52.68	16.02	49.28	43.46	47.67	53.98	47.74	48.43
Walking II	**68.89**	47.91	9.34	46.44	45.18	27.80	16.18	40.05	24.89
Pets 2006 S7	**71.82**	53.68	20.22	37.05	25.07	32.27	40.43	31.16	44.60

the people happens quite often. Since no ground truths of this sequence are provided, we labeled the ground truths on every 10 frames and totally used 120 ground truths for evaluation. The detailed testing results of all the methods on *Walking I*, *Walking II* and *Pets 2006 S7* sequences are shown in Fig. 12, 13 and 14. It can be seen from three similarity curves that JVFSD-RPCA outperforms other methods among most of the frames with ground truths. *Walking II* sequence has the most serious occlusion phenomenon among three challenging sequences. Even under such a situation, JVFSD-RPCA still can obtain a satisfactory result. However, the performances of other methods drop down as shown in Fig. 13. The average similarities of all the algorithms on three sequences are illustrated in Table VII which validate the effectiveness of JVFSD-RPCA on challenging sequences with serious occlusion. More vivid demonstration of the foreground extraction results can be found in Fig. 11.

We present the rational explanation of the experimental results as follows. The subspace-based methods take all the successive frames as adjacent column vectors of a matrix and try to obtain the sparse component or outlier of this matrix. It can be regarded as a global scheme for background subtraction. However, in the crowded sequence, the background does not appear frequently owing to the highly dense walking flow. The global scheme fails when it comes to this situation. Unlike the global scheme, the JVFSD-RPCA takes the motion priori knowledge into account, then the proposed LBWCMD method can assign the successive frames to different subsets. With this reason, the frames in each subsets are not continuous, which means the measurement matrix is not only sparse in space but also sparse in "time". This can alleviate the influence of the highly dense movement to some extent. In addition, we augmented each subset by using the frames with a small quantity of motion. This also can provide more genuine background pixels and prompt good recovery results. For the other methods, they usually need a training or learning phase to produce suitable parameters to detect the foreground. However, good parameters are very hard to obtain under the situation with serious occlusion.

TABLE VIII: Flops of the main steps in Algorithm 1

Main steps in Algorithm 1	Flops
9	$3vN_{Y_i} - 3v - 1$
11	$3vn - m + 1 - 3vN_{Y_i}$
22	$\frac{3}{2}v\sum_{i=1}^{m} N_{Y_i}^2 - \frac{3}{2}vn - m$

C. Complexity analysis of the proposed method

The proposed method is summarized in Algorithm 3, which consists of Algorithm 1 and Algorithm 2. For Algorithm 1, the flops of the main steps are listed in Table VIII, where n is the number of samples, v is the dimension of a sample, m is the number of subsets, N_{Y_i} is the number of samples of the i-th subset in a certain while loop and T is the maximum iteration times within while loop. Since the sample is the location of moving object, $v = 2$. Taking all the steps into consideration, the maximum complexity of Algorithm 1 is $O(Tmn^2)$. For Algorithm 2, the main complexity is produced in step 1 which consists of foreground pixel counting and foreground area sorting. The former costs $O(nwh)$ and the latter costs $O(n^2)$, where h and w are respectively the height and width of an image. However, in our problem, $wh \gg n$. Therefore, The complexity of Algorithm 2 is $O(nwh)$. For Algorithm 3, the complexity of step 1 is easy to evaluate, namely $O(nwh)$. Step 2 and 4's complexities are related to Algorithm 1 and 2, which have been discussed above. The complexity of step 3 is $O(m)$. Finally, we execute step 5 to run RPCA on each group set. The complexity of RPCA has been discussed in [21], namely $O(nwh\text{min}(n, wh))$. In summary, the complexity of the proposed method is $O(nwh\text{min}(n, wh))$. For other subspace-based methods, the complexity of GoDec, RPCA-ME and PCA is $O(nwh\bar{r})$ [24], $O(nwh\bar{k})$ [18] and $O(n^3)$ where $\bar{r}$ is the predefined low rank of measurement matrix and $\bar{k}$ is the number of principle components of measurement matrix, respectively.

Since the proposed method is based on RPCA, the preprocessing procedures have to be completed at first. Then RPCA is conducted on each group set $Z_i (i = 1, 2, \cdots, m)$ to extract the foregrounds. Although the complexity of the proposed method is of the same level as RPCA, the counting flops of the proposed method are more than RPCA. Based on the design of the proposed method as shown in Fig. 3 (b), the executions of RPCA on the group sets are independent actually. Therefore we can use parallel computing technique to accelerate the speed of the proposed method. For the sake of completeness, we compare the execution times of all the methods on Windows XP system with 3.4GHz CPU and 8GB RAM. The experiments were run 10 times over different sequences. The average execution times are listed in Table IX. The results in bold font highlight the longest execution time among all the methods on the same sequence. Obviously, the efficiency of GoDec is the lowest among all the methods. On the contrary, RPCA-ME holds the highest efficiency among all the methods. The proposed method runs about 1.5 seconds slower than RPCA on average. Since the proposed method has completed some optimization steps before RPCA, JVFSD-RPCA costs a little more time in background subtraction. Even though JVFSD-RPCA costs more time, experiments in 11 sequences show that our method is more competitive than the other methods.

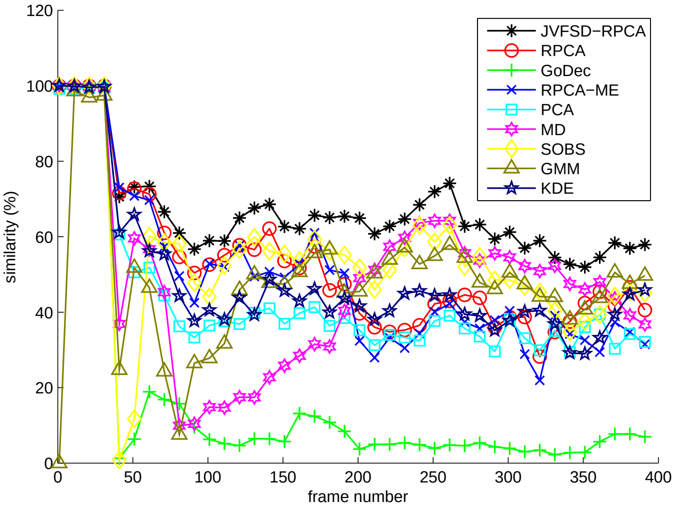


Fig. 12: Comparison results on *Walking I*

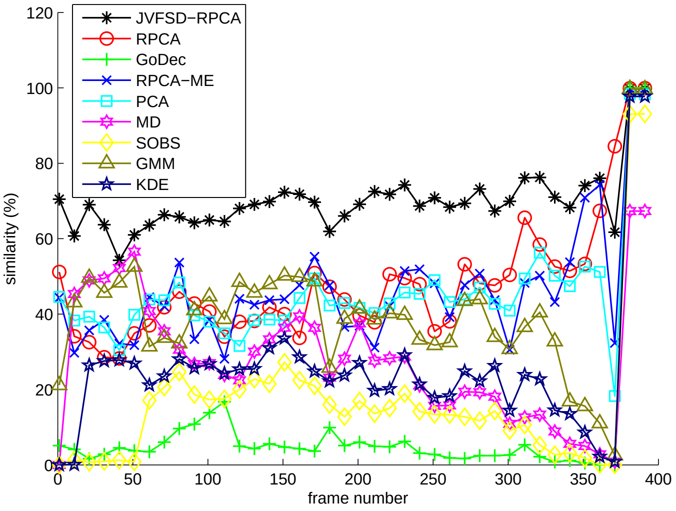


Fig. 13: Comparison results on *Walking II*

VIII. Conclusion

Unlike current subspace-based background subtraction methods, we took advantage of the motion priori knowledge and proposed a new JVFSD-RPCA method for background subtraction. The coarse motion estimation in Section III provides us the locations of the objects and the ratios of the foreground areas in the frames. Based on the centroids of the moving objects, we firstly developed LBWCMD method to divide the video frame set into different subsets for alleviating the influence of the highly dense movement. Then we constructed an augmented set using the frames with a small quantity of motion which facilitates us to obtain more genuine background pixels in each subset. By integrating the above two phases, we finally obtained the group sets, in which the moving objects are uniformly located at random. The proposed method makes the foreground pixels sparse not only in space but also in "time".

TABLE IX: Average execution times (s) of the methods on different sequences

Sequence	JVFSD-RPCA	RPCA	GoDec	PRCA-ME	PCA	MD	SOBS	GMM	KDE
Intelligent room	12.7	11.0	**17.9**	2.0	2.5	5.0	5.4	4.4	4.6
Walking I	19.8	19.3	**21.5**	2.8	3.4	7.3	7.7	6.7	6.9
WaterSurface	19.9	17.6	**25.6**	2.2	4.9	9.7	10.2	8.7	7.1
Dance bootstrapping	27.1	25.7	**33.2**	3.1	6.7	11.2	12.0	10.1	10.5

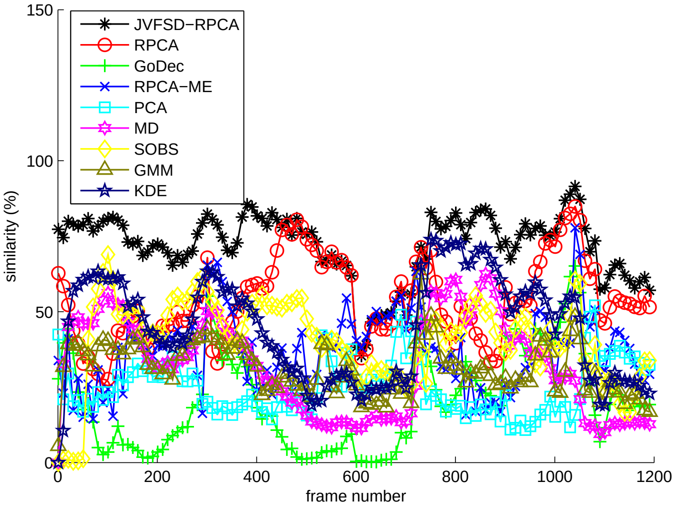


Fig. 14: Comparison results on *Pets 2006 S7*

Experiments on various challenging sequences validated the competitiveness of the proposed method comparing with the state-of-the-art background subtraction methods. The characteristics of the proposed method can be concluded from the experiments. First, LBWCMD can maintain the number of the frames in each subset at the same level. This is very important since the unbalanced scale of the subset will result in bad background extraction results. Second, JVFSD-RPCA works well in the situations of lighting changes or large size of occlusion such as sequence *Walking II*. However, some of the other methods fail in these situations. Third, the proposed method is an improved version of RPCA and it concentrates on enhancing the effectiveness of foreground detection results. Some optimization steps are completed before RPCA which results in more counting flops than RPCA. In order to make JVFSD-RPCA applicable, we can use the parallel computing technique to accelerate the speed of the proposed method. Experimental results in 11 sequences show that the proposed method is more competitive than the other methods. In the future, we will pay more attention to further improving the efficiency of the algorithm.

References

[1] D.-M. Tsai and S.-C. Lai, "Independent component analysis-based background subtraction for indoor surveillance," *IEEE Trans. on Image Processing*, vol. 18, pp. 158-167, Jan. 2009.

[2] T. Xiang and S. Gong, "Video behavior profiling for anomaly detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, No. 5, pp. 893-908, May. 2008.

[3] M. Xu, O. James, and G. Jones, "Tracking football players with multiple cameras," *2004 Int. Conf. on Image Processing*, pp. 24-27.

[4] L. Unzueta, M. Nieto, A. cortés, J. Barandiaran, O. Otaegui, and P. Sánchez, "Adaptive multicue background subtraction for robust vehicle counting and classification," *IEEE Trans. on Intelligent Transportation System*, vol. 30, No. 2, pp. 527-540, Jun. 2012.

[5] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: real-time tracking of the human body," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 780-785, Jul. 1997.

[6] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," *1999 IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 246-252.

[7] M. Harville, "A Framework for high-level feedback to adaptive, per-pixel, mixture-of-gaussian background models," *2002 European Conf. on Computer Vision*, pp. 543-560.

[8] F. E. BAF, T. BOUWMANS, and B. VACHON, "Fuzzy statistical modeling of dynamic backgrounds for moving object detection in infrared videos," *2009 IEEE Conf. on Computer Vision and Pattern Recognition Workshop*, pp. 60-65.

[9] J. K. Suhr, H. G. Jung, G. Li, and J. Kim, "Mixture of Gaussians-based background subtraction for bayer-pattern image sequences," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 21, pp. 365-370, Mar. 2011.

[10] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," *2000 European Conf. on Computer Vision*, pp. 751-767.

[11] C. Cuevas and N. García, "Efficient moving object detection for lightweight applications on smart cameras," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 23, No. 1, pp. 1-14, Jan. 2013.

[12] J.-M. Guo, Y.-F. Liu, C.-H. Hsia, M.-H. Shih, and C.-S. Hsu, "Hierarchical method for foreground detection using codebook model," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 21, NO. 6, pp. 804-815, Jun. 2011.

[13] V. Reddy, C. Sanderson, and B. C. Lovell, "Improved foreground detection via block-based classifier cascade with probabilistic decision integration," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 23, No. 1, pp. 83-93, Jun. 2013.

[14] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE Trans. on Image Processing*, vol. 17, pp. 1168-1177, Jul. 2008.

[15] B. Han and L. S. Davis, "Density-based multifeature background subtraction with support vector machine," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 1017-1023, May 2012.

[16] V. Cevher, A. Sankaranarayanan, M. F. Duarte, D. Reddy, R. G. Baraniuk, and R. Chellappa, "Compressive sensing for background subtraction," *2008 European Conf. on Computer Vision*, pp. 155-168.

[17] N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian computer vision system for modeling human interactions," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 831-843, Aug. 2000.

[18] F. D. L. Torre and M. J. Black, "Robust principal component analysis for computer vision," *2001 IEEE Int. Conf. on Computer Vision*, pp. 362-369.

[19] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis," *Journal of ACM*, vol. 58, No. 3, Article 11, May 2011.

[20] C. Qiu and N. Vaswani, "Real-time robust principal components' pursuit," *2010 48th Annual Allerton Conf. on Communication, Control, and Computing*, pp. 591-598.

[21] Y. Mu, J. Dong, X. Yuan, and S. Yan, "Accelerated low-rank visual recovery by random projection," *2011 IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2609-2616.

[22] B.-K. Bao and G. Liu, "Inductive robust principal component analysis," *IEEE Trans. on Image Processing*, Apr. 2012.

[23] X. Ding, Li. He, and L. Carin, "Bayesian robust principal component analysis," *IEEE Trans. on Image Processing*, vol. 20, pp. 3419-3430, Dec. 2011.

[24] T. Zhou and D. Tao, "GoDec: randomized low-rank & sparse matrix decomposition in noisy case," *2011 Int. Conf. on Machine Learning*, pp. 33-40.

[25] ftp://ftp.pets.rdg.ac.uk/pub/PETS2006/

[26] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*, Addison, 1974.

[27] Y. Benezeth, P. M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, “Comparative study of background subtraction algorithms,” *Journal of Electronic Imaging*, vol. 19, 033003, Sep. 2010.

[28] Z. Zivkovic, “Improved adaptive Gaussian mixture model for background subtraction,” *2004 Int. Conf. on Pattern Recognition*, pp. 28-31.

[29] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, “Detecting moving shadows: algorithms and evaluation,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 918-923, July. 2003.

[30] L. Li, W. Huang, I. Y.-H. Gu, and Q. Tian, “Statistical modeling of complex backgrounds for foreground object detection,” *IEEE Trans. on Image Processing*, vol. 13, pp. 1459-1472, Nov. 2004.

[31] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, “Changedetection.net: A new change detection benchmark dataset,” *IEEE Workshop on Change Detection (CDW-2012) at CVPR-2012*, pp. 16-21.

[32] http://mmc36.informatik.uni-augsburg.de/VSSN06_OSAC/#testvideo

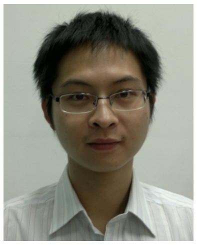
[33] http://perception.csl.illinois.edu/matrix-rank/sample_code.html#RPCA

[34] http://sites.google.com/site/godecomposition/code

[35] http://users.salleurl.edu/˜ftorre/papers/rpca2.html

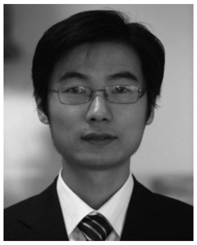
[36] S. Andrews, “An OpenCV C++ background subtraction library,” 2012 [Online] Available: https://code.google.com/p/bgslibrary/

[37] L. Li, W. Huang, I. Y.-H. Gu, and Q. Tian, “Statistical modeling of complex backgrounds for foreground object detection,” *IEEE Trans. on Image Processing*, vol. 13, pp. 1459-1472, Nov. 2004.



Jiajun Wen received the M. S. degree in applied computer technology at Guangdong University of Technology, Guangzhou, China, in 2010. He is currently a Ph. D. candidate in computer science and technology at Shenzhen graduate school, Harbin Institute of Technology (HIT). His current interests include pattern recognition and video analysis.

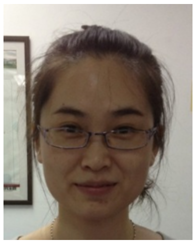
Yong Xu received his B.S. and M.S. degrees at Air Force Institute of Meteorology (China) in 1994 and 1997, respectively. He then received his Ph.D. degree in pattern recognition and intelligence system at the Nanjing University of Science and Technology (NUST) in 2005. From May 2005 to April 2007, he worked at Shenzhen graduate school, Harbin Institute of Technology (HIT) as a postdoctoral research fellow. Now he is a professor at Shenzhen graduate school, HIT. He also acts as a research assistant researcher at the HongKong Polytechnic University from August 2007 to June 2008. His current interests include pattern recognition, biometrics, and machine learning. He has published more than 60 scientific papers.



Jinhui Tang (S’03-M’08) received the B.E. and Ph.D. degrees from the University of Science and Technology of China, Hefei, China, in 2003 and 2008, respectively. After that, he worked as a Research Fellow with the School of Computing, National University of Singapore (NUS). From June 2006 to Feb. 2007, he worked as a Research Intern with Microsoft Research Asia. From January 2010 to April 2010, he was a Visiting Research Scientist with the Department of Information and Computer Science, University of California, Irvine, as a Research Scientist. In December 2010, he joined the School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing, China, where he is currently a Professor. He is the author of 60 journals and conference papers in these areas. His current research interests include semantic-based image retrieval, social media analysis, and multimedia data management. Dr. Tang is a member of the Association for Computing Machinery (ACM). He served as a guest editor of IEEE TRANSACTIONS ON MULTIMEDIA, ACM Transactions on Intelligent Systems and Technology, ACM/Springer Multimedia System Journal, Journal of Visual Communications and Image Representation, Neurocomputing, and Springer Multimedia Tools and Applications Journal. He is a Co-chair of Association for Computing Machinery (ACM) First International Workshop on Web-Scale Multimedia Corpus 2009, a Special Session Chair of ACM Third International Conference on Internet Multimedia Computing and Service 2011, a technical program committee member of over 30 international conferences, and a Reviewer of over 30 prestigious international journals. He was the recipient of the 2008 President Scholarship of Chinese Academy of Science and the Excellent Doctoral Dissertation Award of Anhui Province, and a corecipient of the Best Paper Award in ACM Multimedia 2007.

Yinwei Zhan received the BS degree in 1986 and the MS degree in 1988 in mathematics from Jilin University and the PhD degree in 1992 from Dalian University of Technology in mathematics. He was a postdoctoral researcher at Beijing Normal University from 1992 to 1994 and then an associate professor in Shantou University before being a postdoctoral researcher from Nov. 2001 to Dec. 2004 at CWI and Groningen University, the Netherlands. Since Jan. 2005, he has been a professor at Guangdong University of Technology, being the director of Visual Information Processing R&D Center. His research interests include computer vision and pattern recognition.

Zhihui Lai received the B.S degree in mathematics from South China Normal University, M.S degree from Jinan University, and the PhD degree in pattern recognition and intelligence system from Nanjing University of Science and Technology (NUST), China, in 2002, 2007 and 2011, respectively. He has been a research associate at Hong Kong Polytechnic University since 2010. Currently, he is also a Postdoctoral Fellow at Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology (HIT). His research interests include face recognition, image processing.



Xiaotang Guo is currently a M. S. candidate in computer science and technology at Shenzhen graduate school, Harbin Institute of Technology (HIT). His current interests include pattern recognition and video analysis.