



# Background subtraction in videos using LRMF and CWM algorithm

Wajiha Munir<sup>1</sup> · Adil Masood Siddiqui<sup>1</sup> · Muhammad Imran<sup>1</sup> · Imran Tauqir<sup>1</sup> · Nazish Zulfiqar<sup>1</sup> · Waseem Iqbal<sup>2</sup> · Awais Ahmad<sup>3</sup>

Received: 2 January 2021 / Accepted: 27 April 2021 / Published online: 17 May 2021  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

## Abstract

Background subtraction is a substantially important video processing task that aims at separating the foreground from a video to make the post-processing tasks efficient and relatively easier. Until now, several different techniques have been proposed for this task, but most of them cannot perform well for the videos having variations in both the foreground and the background. In this paper, a novel background subtraction technique is proposed that aims at progressively fitting a particular subspace for the background that is obtained from  $L_1$ -low-rank matrix factorization using the cyclic weighted median algorithm and a certain distribution of a mixture of Gaussian of noise for the foreground. The expectation maximization algorithm is applied to optimize the Gaussian mixture model. Furthermore, to eliminate the camera jitter effects, the affine transformation operator is involved to align the successive frames. Finally, the effectiveness of the proposed method is augmented using a subsampling technique that can accelerate the proposed method to execute on an average more than 250 frames per second while maintaining good performance in terms of accuracy. The performance of the proposed method is compared with other state-of-the-art methods and it was concluded that the proposed method performs well in terms of F-measure and computational complexity.

**Keywords** Background subtraction · Cyclic weighted median (CWM)algorithm · Low-rank matrix factorization (LRMF) · Gaussian mixture model(GMM) · Expectation maximization (EM)algorithm

✉ Muhammad Imran  
m.imran@mcs.edu.pk

Wajiha Munir  
wajihamunir93@gmail.com

Adil Masood Siddiqui  
dradil@mcs.edu.pk

Imran Tauqir  
imrantqr@mcs.edu.pk

Nazish Zulfiqar  
nazishzulfiqar70@gmail.com

Waseem Iqbal  
waseem.iqbal@mcs.edu.pk

Awais Ahmad  
awais.ahmad@mail.au.edu.pk

<sup>1</sup> Department of Electrical Engineering, MCS, National University of Sciences and Technology (NUST), Islamabad, Pakistan

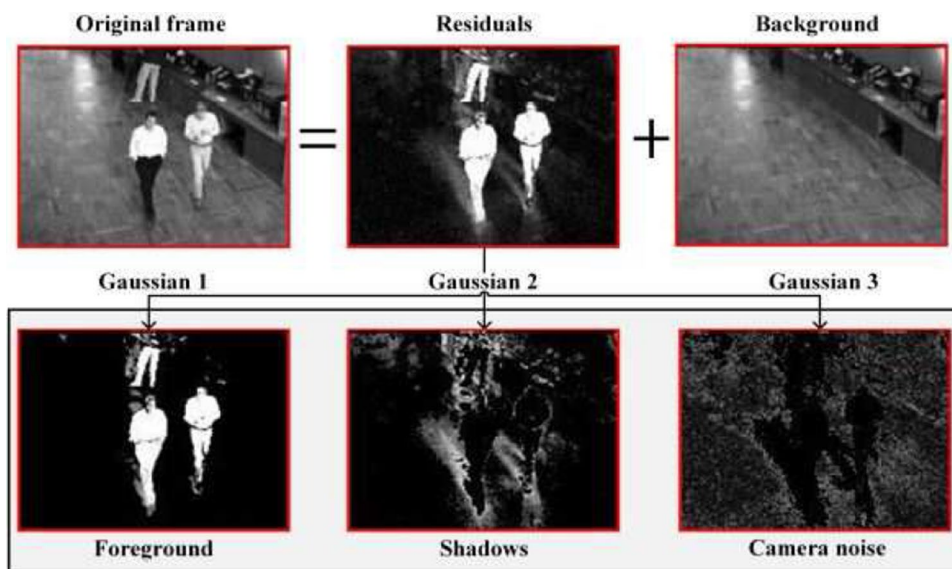
<sup>2</sup> Department of Information Security, MCS, National University of Sciences and Technology (NUST), Islamabad, Pakistan

<sup>3</sup> Department of Computer Science, Air University, Islamabad, Pakistan

## 1 Introduction

With the recent advances in the domain of computer vision and the development of digital inexpensive cameras, the need for both indoor and outdoor video surveillance systems has been stimulated. Video processing is a substantially important branch of image processing and computer vision focused on extracting information from real scene videos. Among several other video processing techniques, background subtraction has attained great importance as a developing research area, during the past few years [1–4]. Background subtraction aims at dividing the pixels of each frame of a video into two complementary sets, i.e., the background pixels corresponding to the stationary part of the frame and the foreground ones that correspond to the moving part. This makes the subsequent post-processing tasks efficient and relatively easier. This makes the subsequent post-processing tasks efficient and relatively easier. Most of the background subtraction techniques aim at choosing such algorithms and task scheduling methods that make the real-time processing of videos efficient and

**Fig. 1** Background subtraction results for a bootstrap sequence [12]. The first row contains the original frame, noise, and the background extracted by [12] (in left-to-right direction). The second row contains three noise components that correspond to the foreground, i.e., the moving object, its shadow, and certain feeble camera noise



accurate [1]. Furthermore, processing a high-resolution video stream efficiently is also a challenging task [2]. Recently, techniques are being proposed that focus on limiting the number of historical frames and thus making the background subtraction process suitable for the real-time implementation [3]. Background subtraction is a widely used technique that finds its application particularly in surveillance videos, object detection and tracking [5], detection of urban traffic [6], long-term monitoring of a scene [7], video compression [8], etc. Some critical challenges faced in the implementation of a background subtraction method are:

- i. Selection of the initial frames.
- ii. Development of a model for Background.
- iii. Selection of an effective threshold for the classification of pixels.
- iv. Updating process for the Background or the threshold or both.

Nowadays, the amount of videos obtained from surveillance cameras disseminated throughout the world is increasing dramatically. These continuously evolving videos not only made the evaluation of the background subtraction for these videos crucial but also generated an urge to build some real-time techniques to manage this increasing video data.

Recently, multiple background subtraction methods have been developed [9, 10] that gradually updates the low-rank structure underlying the background by considering one frame at a time incrementally, thus decreasing the computation time. These methods can effectively carry out background subtraction for real-time videos by

significantly speeding up the calculations performed for this task. However, there are still some evident defects in these techniques when employed for real-time videos. On one hand, these approaches mostly neglect the dynamic jitters of the camera including rotation, scaling, and illumination changes that occur frequently in the real life, while assuming that the background possesses a low-rank structure. On the other hand, most of the current methods make use of a fixed loss term for their models such as  $L_1$  or  $L_2$  norm losses. These loss terms make an implicit assumption for the noises, i.e., the foreground part of the videos, by considering them to have a fixed probability distribution such as Laplacian or Gaussian. Although, in real scenarios, there are always dramatic variations in the foreground over time, therefore, this assumption for noise to have a fixed probability distribution deviates from real scenarios. In some cases, multiple modalities of noises are contained in the foreground, as shown in Fig. 1. Such cases require taking into account more complex models for noise. In practice, neglecting such an essential intuition about the diversity in the foreground of a video does not allow the current methods to be robust against the noise variations that occur in the real-time videos.

The organization of the rest of the paper is as follows: In Sect. 2, a brief review of various already existing background subtraction methods is given. Section 3 covers the proposed methodology and the related algorithms. In Sect. 4, the results acquired by exploiting the proposed algorithm on various video datasets are explained. Finally, the conclusion of the proposed work is given in Sect. 5.

## 2 Related work

Over the last couple of years, various research studies have been presented on background subtraction due to which it has attained great importance as a developing research area [13–16]. Some of these recent studies have been as referenced for the development of the proposed model. Gaussian parameter compression techniques [14] as well as combination of Mixture of Gaussian (MoG) and compressive sensing (CS) [16] are recently being used. Furthermore, parallel implementation strategies for algorithms are also being utilized for object detection in UAV-sourced videos [15]. As described earlier, background subtraction is a widely used technique particularly in surveillance videos, object tracking and detection, traffic or crowd monitoring, etc. where the main focus is to separate the moving objects, i.e., foreground from the stationary background [17]. However, the detection of moving objects in videos or other applications is yet a challenging task owing to several issues such as shadow, varying illumination, occlusion, background motion, camera jitter, as well as several different types of ambiguities like atmospheric disturbances or noise, object overlapping outliers, etc. [18]. To overcome these challenges, statistical models are the most effective ones. Statistical models may be non-parametric or parametric. Some of the most commonly applied non-parametric methods are the Kernel Density Estimation (KDE) and Eigenvalue Decomposition, but these methods have an extensive memory requirement, as well as a high computational complexity [19]. Conversely, parametric statistical models depend on the use of statistical distributions for background modeling. One of the most popular statistical models is finite GMM, capable of coping with slight illumination changes as well as moving background with small repetitive motion [19]. As with non-parametric methods, parametric methods have their limitations such as the learning parameters need to be set automatically, have to cope with complex dynamic background, have to dissociate shadows from object, etc. [19]. To accommodate these challenges and overcome the aforementioned limitations, [20] proposed an improvised version of GMM for the detection of moving objects. It uses the Gaussian elements for modeling the intensity values of a block of pixels and compensates for the learning rate limitation using a dynamic learning rate. Considering the pixel block, rather than a single pixel value, reduced the computation time almost four times, keeping the performance nearly similar to earlier methods [20]. Recently, matrix decomposition methods, for instance, Robust Principal Component Analysis (RPCA), have become an efficient framework for background subtraction. These methods aim to decompose a matrix into low-rank (for

background) and sparse (for foreground) components. However, in some scenarios, as the size of input data increases and because of the absence of sparsity constraints, these methods show weak performance as they cannot handle the real-time challenges, resulting in inaccurate foreground areas. To address the aforementioned problem, an online single unified optimization framework for simultaneous detection of the foreground as well as learning of background is proposed in [21]. This method has better performance, as it provides a more reliable and efficient low-rank component, but it cannot be used for moving cameras. Although RPCA provides a good framework for background subtraction, it still has a very high computational complexity and huge memory requirements because of its batch optimization. To solve this issue, online RPCA (OR-PCA) [22] is developed which can process such high-dimensional data through stochastic manners. However, the sparse component obtained by OR-PCA cannot always handle numerous background modeling challenges, which degrades the performance of the system. To overcome these challenges, [23] presented a multi-feature based OR-PCA scheme. Integration of multiple features into OR-PCA not only augments the quality of detected foreground but also enhances the quantitative performance of this technique, as compared to single feature OR-PCA and RPCA through PCP-based methods [23]. However, when OR-PCA is applied to real sequences which have dynamically changing background, the performance of OR-PCA is also reduced. Therefore, there is a need for enhancement in OR-PCA to cope up with the increased complexity and variety of videos. In [24], an online algorithm built on Incremental Nonnegative Matrix Factorization (INMF) is presented, which resolves the problems encountered in OR-PCA using non-negative and structured sparsity constraints. In complex scenes, this algorithm reduces the number of missed and false detection. Subspace Learning methods such as Matrix Completion (MC) and RPCA have been explored and attained significant attention during the last few years [12, 25]. These methods are based on low-rank modeling and are meant to reduce the dimensionality in a very-high-dimensional space. Unfortunately, there are some prevalent challenges with most of the conventional matrix decomposition algorithms based on MC and RPCA. First, these methods use batch processing; second, for each iteration of the optimization process, all the frames have to be accessed. As a result, they have an extensive memory requirement and are computationally inefficient. The method proposed in [26] considers the sequence of images as constructed from a low-rank matrix for background and a dynamic tree-structured sparse matrix for the foreground. It solves the decomposition by the use of approximated RPCA which is extended to make it capable of handling

camera motion. This method decreases the complexity, requires less time for computation, and does not require huge memory for larger videos. Similarly, to estimate a robust background model, in [25], a robust spatiotemporal low-rank matrix completion (SLMC) algorithm is presented for dynamic videos. In the proposed method, spectral graphs are regularized for encoding spatiotemporal constraints. Furthermore, SLMC algorithm is extended to Spatiotemporal RPCA (SRPCA) with dynamic frames extraction. Together, these algorithms make the process robust and accurate but in SRPCA as both the foreground and background are optimized at the same time, the computational time is increased. Furthermore, impressive achievements in deep learning have enthused the researchers to apply deep neural networks for the task of background subtraction [27–29]. Usually, in a CNN, some of the essential information may be lost by the first convolution layer and is not available for the bottom layers of CNN, thus limiting its performance in multiple feature extracting. To overcome this limitation, in [27], a CNN combining multi-scale representation is studied and a multi-scale fully convolutional network architecture is proposed for background subtraction that takes advantage of multiple layer features. Similarly, in [28], a background subtraction method based on depth data in SBM-RGBD is proposed that is capable of achieving more accurate results than the traditional methods. The major difference between the depth data and traditional data is the distance information provided by the depth sensors. In this method, the impact of edge noise and absent pixels is reduced by applying a preprocessing method. Mostly, the current available CNN-based foreground methods make use of 2-D CNN, and hence, they fail to account for the temporal features present in the image sequences. These temporal features are beneficial in improving the performance of IR foreground detection, because the IR images lack rich spatial features. To solve this problem, a background subtraction method based on the 3-D convolutional network is proposed in [29] named as MFC3-D (multi-scale 3-D Fully Convolutional Network). The proposed network can effectively learn the deep and hierarchical multi-scale spatial-temporal features, hence, can perform well for foreground detection in IR videos. Although these deep learning-based algorithms have a better performance, their drawback is that they are increasingly dependent on costly hardware resources owing to the demanding training process. Because of the restricted computational resources and high real-time demands, these approaches are not realistic for visual surveillance. Moreover, mostly, the deep learning-based algorithms are supervised algorithms which means that they need a ground truth to train the model. These ground truths are constructed by either a human expert or by other unsupervised background

subtraction methods. However, the background subtraction methods for video surveillance have to be unsupervised. Thus, the CNN-based background subtraction methods may not be much useful when it comes to practical applications. Considering the pros and cons of all the techniques reviewed in the literature survey, in this work, we have proposed a technique that is computationally efficient and has improved accuracy in terms of detected background and foreground, with less number of false detections.

To mitigate the limitations of existing techniques, an innovative technique for background subtraction has been proposed in this paper. The proposed technique progressively fits a particular subspace for the background that is obtained from  $L_1$ -low-rank matrix factorization (LRMF) using cyclic weighted median (CWM) and a certain distribution of a mixture of Gaussian (MoG) of noise for the foreground. This fit is achieved by regularization of the background and foreground information that is acquired from the preceding frames. In comparison to the conventional methods that used a fixed noise distribution for all the frames of the video, a separate MoG distribution is used to model the noise or foreground for each frame of the video. The expectation maximization (EM) algorithm is applied to optimize the Gaussian mixture model (GMM). To eliminate the camera jitter effects, the affine transformation operator is involved that acts to align the successive frames. Finally, the effectiveness of the proposed method is augmented using a subsampling technique that can accelerate the proposed method to execute on an average more than 250 frames per second while maintaining good performance in accuracy.

### 3 Proposed methodology

In this section, the proposed algorithm for background subtraction is presented in detail. The gist of the proposed work is that for each new video frame  $x_t$ , the aim is to progressively fit a particular subspace for the background that is obtained from  $L_1$ -LRMF and a certain MoG distribution of noise for the foreground. This fit is achieved by regularization of the background and foreground information that is acquired from the preceding frames.

#### 3.1 LRMF to obtain background subspace

Low-rank matrix factorization is considered to be a significantly important technique in data science. The main idea of matrix factorization is that sometimes the data contain latent structures by uncovering which a compressed representation of the data can be obtained. Matrix factorization provides a unified method for dimensionality reduction, matrix completion, and clustering by factorization of the original data



matrix into low-rank matrices. The  $L_1$ -norm LRMF problem can be formulated as follows. Let  $\mathbf{X}$  be the video data matrix, such that  $\mathbf{X} = (x_1, x_2, \dots, x_n)$  where  $\mathbf{X} \in R_{d \times n}$ ,  $d$  is the dimensionality, and  $n$  is the number of data. Each column of  $\mathbf{X}$ , i.e.,  $x_i$  represents a video frame having  $d$ -dimension. The missing entries in the matrix  $\mathbf{X}$  are represented by an indicator matrix  $\mathbf{W}$ , where  $\mathbf{W} \in R_{d \times n}$ . The elements of  $\mathbf{W}$  i.e  $w_{ij}$  are taken in such a way that it is zero when the corresponding element is missing and one otherwise [30].

Given  $\mathbf{X}$  and  $\mathbf{W}$ , it is possible to formulate a general LRMF problem [31] as

$$\min_{\mathbf{U}, \mathbf{V}} \|\mathbf{W} \odot (\mathbf{X} - \mathbf{UV}^T)\|_{L_1}, \tag{1}$$

where  $\mathbf{U}$  and  $\mathbf{V}$  denote the basis and coefficient matrices respectively. Furthermore,  $\mathbf{U} = [u_1, u_2, \dots, u_k]$ ,  $\mathbf{V} = [v_1, v_2, \dots, v_k]$  and  $\mathbf{U} \in R^{d \times r}$  and  $\mathbf{V} \in R^{n \times r}$ , with  $r \ll \min(d, n)$  and  $\odot$  denotes the Hadamard product, i.e., component-wise multiplication, different from the common matrix product. Here,  $r \ll \min(d, n)$  basically indicates the property of low rank of  $\mathbf{U} \mathbf{V}^T$ . Under the framework of Maximum-Likelihood Estimation (MLE), Eq. (1) can also be understood as

$$x_{ij} = (\bar{u}_i)^T \bar{v}_j + e_{ij}, \tag{2}$$

where  $\bar{u}_i$  is the  $i$ th row vector of  $\mathbf{U}$ ,  $\bar{v}_j$  is the  $j$ th row vector of  $\mathbf{V}$ , and  $e_{ij}$  is the noise element embedded in  $x_{ij}$ .

To make the proposed model robust to complex noises present in the real-time videos, it is possible to better model the term  $e_{ij}$  as a parametric probability distribution. Among other probability distributions, MoG possess a strong capability to approximate to general distributions [31]; therefore, it is selected here to adapt flexibly to the real cases. In particular, it is assumed that each video frame  $x_{ij}$  follows:

$$x_{ij} \sim \sum_{k=1}^K \pi_k N(x_{ij} | (\bar{u}_i)^T \bar{v}_j, \sigma_k^2). \tag{3}$$

For Eq. (1), unfortunately, it is somehow difficult to solve the  $L_1$ -norm minimization because of two reasons. On one hand, its optimization is non-convex which generally makes the finding of a global minimum a bit difficult, and in case of missing entries, it is even proven to be an NP-hard problem [32]. On the other hand, standard optimization tools can hardly find an effective closed-form iteration formula [33], because  $L_1$ -norm minimization is non-smooth. In the proposed technique, we have made use of the simple cyclic coordinate descent algorithm [34] which shows outstanding performance on  $L_1$ -norm LRMF.

The core idea of the cyclic coordinate descent algorithm is to break the fundamental complex minimization problem into a sequence of simple elementary subproblems. Each of these subproblems, having only one scalar parameter, is then recursively optimized. Being convex optimization problems,

each of them can be readily solved using a weighted median filter, which eliminates the need for time-consuming inner loops for numerical optimization. Moreover, the recursive employment of weighted median filter makes the method more robust to the missing entries as well as the outliers to a large extent.

### 3.2 CWM algorithm for solving $L_1$ -norm LRMF problem

The main idea of CWM algorithm to solve the minimization in Eq. (1) is to apply recursively the weighted median filter [35] to update each element of  $\mathbf{U} = [u_1, u_2, \dots, u_k]$ ,  $\mathbf{V} = [v_1, v_2, \dots, v_k]$ , such that  $\mathbf{U} \in R^{d \times r}$  and  $\mathbf{V} \in R^{n \times r}$ . The algorithm can be described by the following steps:

*Step 1* To update each element  $v_{ij}$  of  $\mathbf{V}$  where ( $i=1, \dots, k$ ) and ( $j=1, \dots, n$ ), the weighted median filter is applied cyclically while keeping all other components of  $\mathbf{U}$  and  $\mathbf{V}$  fixed. This is done by solving

$$v_{ij}^* = \operatorname{argmin}_{v_{ij}} \|\mathbf{w}_j \odot e_j^i - \mathbf{w}_j \odot \mathbf{u}_i v_{ij}\|_{L_1}, \tag{4}$$

where  $w_j$  represents to the  $j$ th column vector of  $\mathbf{W}$  and  $e_j^i$  represents the  $j$ th column vector of  $\mathbf{E}_i$  defined as

$$\mathbf{E}_i = \mathbf{X} - \sum_{j \neq i} \mathbf{u}_i \mathbf{v}_j^T. \tag{5}$$

*Step 2* In the next step, apply cyclically the weighted median filter for updating each element  $u_{ij}$  of  $\mathbf{U}$  keeping all other components of  $\mathbf{U}$  and  $\mathbf{V}$  fixed. This can be done by solving the following minimization problem.

$$u_{ij}^* = \operatorname{argmin}_{u_{ij}} \|\tilde{\mathbf{w}}_j \odot \tilde{\mathbf{e}}_j^i - \tilde{\mathbf{w}}_j \odot \mathbf{v}_i u_{ij}\|_{L_1}, \tag{6}$$

where  $\tilde{\mathbf{w}}_j$  represents to the  $j$ th row vector of  $\mathbf{W}$  and  $\tilde{\mathbf{e}}_j^i$  represents the  $j$ th row vector of  $\mathbf{E}_i$ .

The factorized matrices  $\mathbf{U}$  and  $\mathbf{V}$  can be recursively updated via iterative implementation of the above procedures till the fulfilment of the termination condition.

In step 1 of the algorithm, the initial values of  $\mathbf{U}$  and  $\mathbf{V}$  are obtained from PCA [37], performed prior to the  $L_1$ -LRMF. For the termination condition of step 4, as in the iteration process, the objective function of Eq. 1 is decreasing monotonically; therefore, the algorithm terminates either when the rate of updating  $\mathbf{U}$  and  $\mathbf{V}$  is below some predetermined threshold or when the maximum number of iterations has been reached.

#### 3.2.1 GMM for foreground modeling

In the proposed technique, as discussed earlier, rather than using a fixed distribution of noise for all the frames in a video, noise or foreground  $e_{ij}$  of each frame is modeled as a

separate MoG distribution, that is regularized by a penalty to enforce its parameters close to those estimated from the preceding frames. This penalty may also be reformulated as the conjugate prior for MoG of the current frame, by encoding the knowledge of noise learned previously. As the MoG can effectively approximate an extensive range of distributions, the proposed method is capable of finely adapting to the variations in the foreground, even for the noises having complex dynamic structures. The expectation maximization (EM) algorithm is then used to iteratively approximate the maximum likelihood (ML) estimates of the mixture model parameters. One very useful property of the EM algorithm is that for each subsequent iteration, the maximum likelihood of the data increases strictly, which implies that it is guaranteed to approach a saddle point or local maximum. The EM algorithm, as the name suggests, essentially involves two steps. The first step being the E-step or Expectation step and the second step being the M-step or Maximization step. Iterative repetition of these two steps until convergence of the algorithm gives the maximum-likelihood estimate. The following steps explain the working of the EM algorithm.

- i. Initialize the model parameters, that is, means  $\mu_k$ , covariances  $\Sigma_k$ , and the mixture weights  $\pi_k$  where  $k = 1, 2, \dots, K$ . Estimate the initial value of log-likelihood.
- ii. After initialization, the second step is the expectation step. In this step, using the current parameters, the responsibilities are evaluated as

$$\gamma_{ik}^t = \frac{\pi_k N(x_i^t | (\bar{u}_i)^T \bar{v}, \sigma_k^2)}{\sum_{k=1}^K \pi_k N(x_i^t | (\bar{u}_i)^T \bar{v}, \sigma_k^2)}, \tag{7}$$

where  $\gamma_{ik}^t$  corresponds to the latent (hidden) variable for the  $k$ th Gaussian,  $x_i^t$  denotes the  $i$ th pixel of the newly coming frame  $x^t$ .

- iii. The next step is the maximization step. This step involves the recalculation of the model parameters  $\pi_k$ ,  $\sigma_k^2$ , and  $\bar{v}$  using the currently obtained values. This recalculation is done using the following equations:

$$\pi_k = \pi_k^{t-1} - \frac{\bar{N}}{N} (\pi_k^{t-1} - \bar{\pi}_k), \tag{8}$$

$$\sigma_k^2 = \sigma_k^{t-1^2} - \frac{\bar{N}_k}{N_k} (\sigma_k^{t-1^2} - \bar{\sigma}_k^2), \tag{9}$$

where  $\bar{N} = d$ ;  $\bar{N}_k = \sum_{i=1}^d \gamma_{ik}^t$ ;  $\bar{\pi}_k = \frac{\bar{N}_k}{\bar{N}}$ ;  
 $\bar{\sigma}_k^2 = \frac{1}{\bar{N}_k} \sum_{i=1}^d (\gamma_{ik}^t (x_i^t - (\bar{u}_i)^T \bar{v})^2)$ ;  
 $N = N^{t-1} + \bar{N}$  and  $N_k = N_k^{t-1} + \bar{N}_k$

$$\bar{v} = (\bar{U}^T \text{diag}(w^t)^2 \bar{U})^{-1} U^T \text{diag}(w^t)^2 x^t, \tag{10}$$

- iv. Finally, evaluate the log-likelihood [12]

$$\ln p(\mathbf{X}|\mathbf{U}, \mathbf{V}, \mathbf{\Sigma}, \mathbf{\Pi}) = \sum_{n=1}^N \left\{ \sum_{k=1}^K \pi_k N(x_{ij} | (\bar{u}_i)^T \bar{v}_j, \sigma_k^2) \right\}, \tag{11}$$

where  $\Pi = \{ \pi_k \}_{k=1}^K$ . If either the parameters or the log-likelihood has converged, it means that the desired results have been achieved, if not, then return to step 2 and iterate until convergence. After the convergence of the EM algorithm, the fitted model obtained is utilized to update the current subspace  $U_t$  using the updating rules [12] defined as follows:

$$\mathbf{u}_i^t = \bar{\mathbf{A}}_i^t \bar{\mathbf{b}}_i^t, \tag{12}$$

$$\bar{\mathbf{A}}_i^t = \frac{1}{\rho} \left( \bar{\mathbf{A}}_i^{t-1} - \frac{w_i^{t^2} \bar{\mathbf{A}}_i^{t-1} \bar{v}^t (\bar{v}^t)^T \bar{\mathbf{A}}_i^{t-1}}{\rho + w_i^{t^2} (\bar{v}^t)^T \bar{\mathbf{A}}_i^{t-1} \bar{v}^t} \right), \tag{13}$$

$$\bar{\mathbf{b}}_i^t = \rho \bar{\mathbf{b}}_i^{t-1} + w_i^{t^2} x_i^t \bar{v}^t, \tag{14}$$

where  $\bar{\mathbf{A}}_i^t$  and  $\bar{\mathbf{b}}_i^t$  denote the model variables that are used as background prior.  $\bar{\mathbf{A}}_i^t$  represents a semi-definite matrix which makes it easy to learn the subspace  $\mathbf{U}$ .  $w_i^t$  represents  $i$ th element of indicator matrix  $\mathbf{W}$ ,

defined as  $w_i^t = \sqrt{\sum_{k=1}^K \frac{\gamma_{ik}^t}{2\sigma_k^2}}$  and  $\rho$  controls the strength of the priors. Its value is set in such a way that allows the subspace to slightly lean to the current frame. In the proposed work, it is set to 0.98.

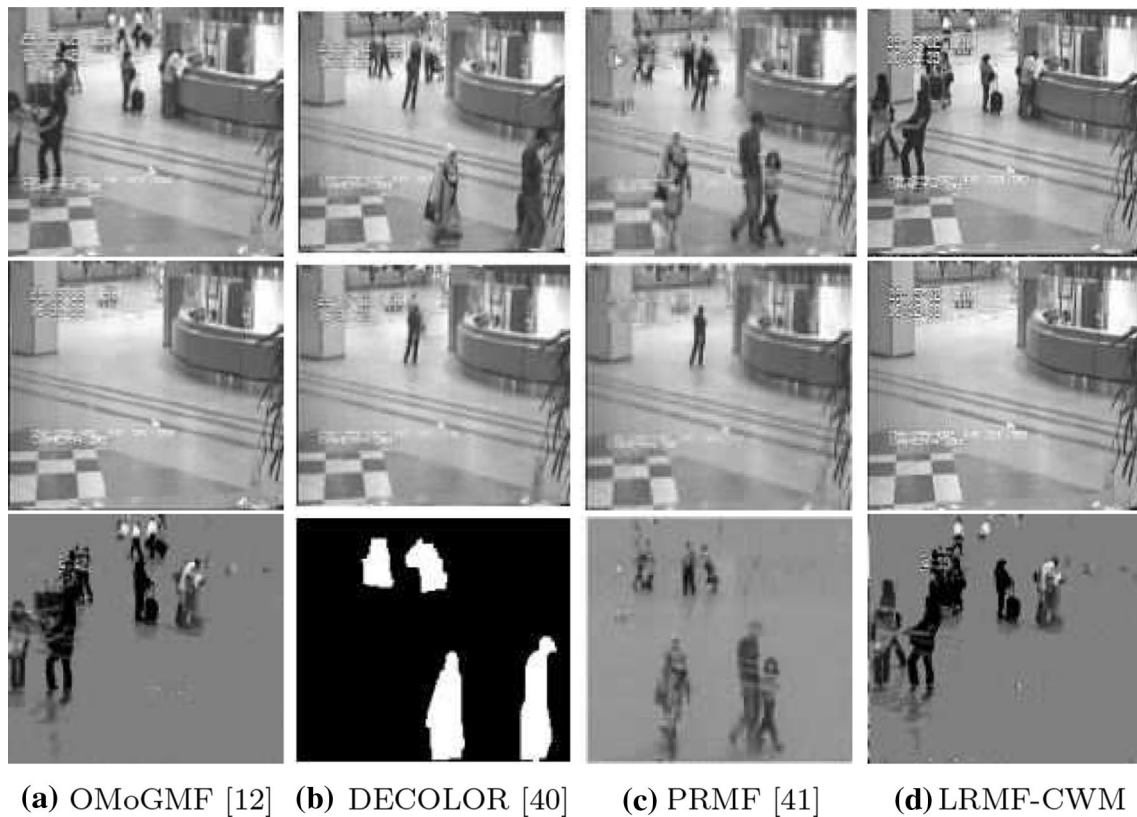
### 4 Performance evaluation

For the performance evaluation of the proposed technique, we conducted background subtraction experiments on three different datasets selected from the Li dataset<sup>1</sup> which have static as well as dynamic backgrounds, i.e.

- i. Airport video without camera jitter effect.
- ii. An unaligned face with different illuminations.
- iii. Synthetically transformed airport video with camera jitter effect.

All these experiments were implemented on a computer with Intel Core m3 CPU and 8GB RAM and Windows10 operating system. The proposed work is implemented on MATLAB R2015.

<sup>1</sup> <http://www.changedetection.net/>.



**Fig. 2** Airport dataset: from left to right OMoGMF [12], DECOLOR [40], PRMF [41], LRMF-CWM. First row: Original frame; second row: background; third row: foreground

### 4.1 Quantitative evaluation

The quantitative metrics used to assess the performance is F-measure which is expressed as

$$F\text{-measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{15}$$

Recall is a measure that tells how many of the true positives are identified correctly, whereas precision tells us that out of all the positive classifications, how many are actually correct. In Eq. (15), precision and recall are estimated as follows:

$$\text{Precision} = \frac{|S_f \cap S_{gt}|}{|S_f|} \tag{16}$$

$$\text{Recall} = \frac{|S_f \cap S_{gt}|}{|S_{gt}|} \tag{17}$$

where  $S_f$  is the support set for the foreground estimated from the proposed method and  $S_{gt}$  is the set of ground truth.

The quantitative analysis is performed by comparing the results obtained by proposed technique and OMoGMF

**Table 1** Comparison of F-Measure for airport dataset without jitter effect

Dataset	OMoGMF [12]	DECOLOR [40]	PRMF [41]	Proposed
Airport (without jitter effect)	61.9	63.98	65.87	64.1

**Table 2** Comparison of F-Measure for airport dataset with jitter effect

Dataset	OMoGMF [12]	t-GRASTA [43]	Proposed
Airport (jitter effect)	76.1	66.24	77.4

method proposed in [12]. This comparison of F-Measure is shown in Tables 1 and 2, respectively. In Table 1, this comparison is made for airport dataset without jitter effect, whereas in Table 2, comparison is done for airport dataset with jitter effect. It can be seen that in both cases, our proposed technique performs best. For qualitative comparison, we have compared the proposed technique with four already existing techniques. In these experiments, we have taken into account the two major problems that were

**Fig. 3** Airport dataset with jitter effect. First row: misaligned frame; second row: aligned frame; third row: background; fourth row: foreground



**(a)** OMoGMF [12] **(b)** t-GRASTA [43] **(c)** LRMF-CWM

not solved in the existing literature, i.e., the dynamic background such as illumination changes, waving trees, ripples in water, etc., and the camera jitter effects like translation, rotation, scaling, or a combination of all these, thereby making proposed technique more robust for dynamic background changes. Figure 2 illustrates the acquired results by comparing it with three different already existing techniques, i.e., OMoGMF [12], DECOLOR [40], and PRMF [41].

To demonstrate the effectiveness of the proposed technique for the camera jitter effects, we took the same Airport sequence, but this time, it has camera jitter effects in it. Each successive frame is either rotated or translated by a certain amount as compared to the previous frame. The results of applying the proposed technique, OMoGMF [12] and t-GRASTA [43] on such a video sequence are presented in Fig. 3. Under certain low-rank assumption, it is possible to reconstruct a larger low-rank matrix from a fewer number of its entries [38]. Stimulated by some previous efforts [39] on this issue, the efficiency of the proposed method is further improved by appending the subsampling technique as used in [12]. The introduction of a subsampling technique can accelerate the execution process to, on average, more than 250 frames per second. Furthermore, this add on does not affect the accuracy of the proposed method, i.e., a good performance in accuracy is maintained. In these experiments, the subsampling rate is 0.01. The results obtained using the

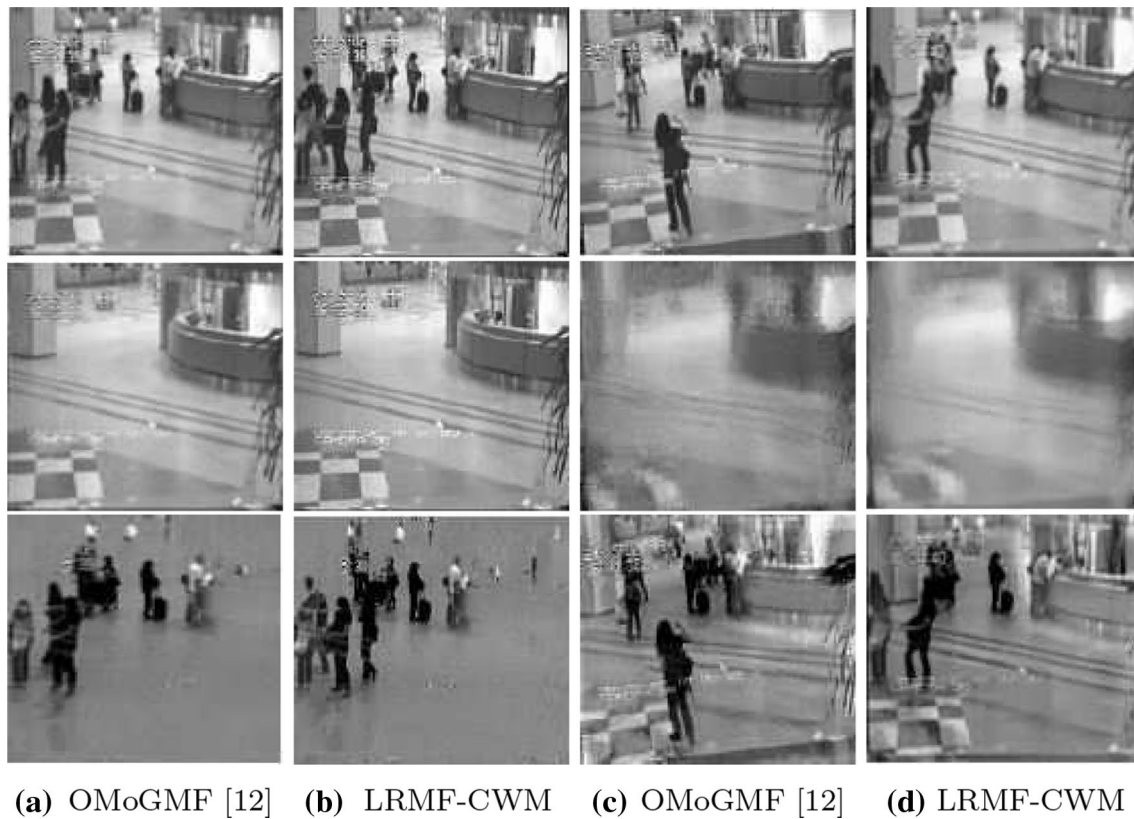
subsampling technique on the three video sequences are illustrated in Fig. 4.

To show the effectiveness of our proposed alignment approach using affine transformation, we have tested the proposed approach for aligning the frames of “Dummy” dataset shown in Fig. 5. This dataset contains multiple images that are not only misaligned but also suffer from illumination variation and block occlusion. The results obtained by applying the proposed technique and those proposed in [42, 43] and [12] are shown in Fig. 5. It can be seen that in some images, the face has become enlarged during the alignment process. Furthermore, there are darker images which means that the effect of illumination variation has not been removed effectively. In the proposed technique, it can be seen that the image alignment is improved and the illumination variation has been minimized.

## 4.2 Computational complexity

The computational complexity of the proposed technique is reduced as compared to the other state-of-the-art methods used for background subtraction. This is basically due to the CWM method that we have employed for the LRMF using  $L_1$ -norm. If we take  $n$  and  $d$  as the size and dimensionality of the input data matrix, respectively, then, for the proposed method, the computational complexity is of the order  $O(d+n)$ , whereas that of other state-of-the-art algorithms





**Fig. 4** For the Airport dataset, using subsampling rate=0.01. **a** and **b** are without jitter effect, whereas **c** and **d** are with jitter effect

**Table 3** Comparison of the computational time (in seconds) of OMoGMF [12] and proposed technique (without subsampling)

Techniques	Without subsampling		
OMoGMF [12] Proposed	Airport	Dummy	Airport (jitter effect)
	14.288	0.9968	14.866
	13.884	0.8636	13.288

**Table 4** Comparison of the computational time (in seconds) of OMoGMF [12] and proposed technique (with subsampling)

Techniques	With subsampling		
OMoGMF [12] Proposed	Airport	Dummy	Airport (jitter effect)
	2.712	0.544	3.232
	1.890	0.204	2.246

is  $O(dn)$  [36]. A comparison of computational time of our proposed algorithm with that of OMoGMF [12] algorithm is made in Table 3 and Table 4. It can be seen from Table 3, the computation time of the proposed technique is lessened as compared to the OMoGMF method [12]. The

computational time for the Airport dataset is reduced from 14.288 to 13.884 s, from 0.9968 to 0.8636 s for the Dummy dataset and from 14.866 to 13.288 s for the airport dataset with camera jitter effects. To further enhance the efficiency of the proposed technique, subsampling is embedded into the calculation. The subsampling rate is taken to be 0.01, i.e., 1%. This can accelerate the method to execute on an average of more than 250 frames per second while maintaining a good performance accuracy. Similarly, from Table 4, it is obvious that the proposed technique has reduced computation time as compared to OMoGMF [12] even with the subsampling technique.

## 5 Conclusion

In this paper, a technique has been proposed which aims at making background subtraction available for real-time videos both in terms of speed and accuracy. It gradually fits a specific subspace for the background that is obtained from  $L_1$ -LRMF using CWM and a certain MoG distribution of noise for the foreground. This fit is achieved by regularization of the background and foreground information that is acquired from the preceding frames. As opposed to conventional methods which used a fixed distribution of noise



(a) Original frames



(b) RASL [42]



(c) t-GRASTA [43]



(d) OMoGMF [12]



(e) LRMF-CWM

**Fig. 5** a Images in “Dummy” dataset containing misaligned images with block occlusion and illumination variation. Frames with block occlusion removed and illumination effect minimized using **b** RASL [42], **c** t-GRASTA [43], **d** OMoGMF [12], and **e** LRMF-CWM

for all frames in a video, in this technique, a separate MoG distribution is utilized to model the noise or foreground for each frame of the video. The EM algorithm is used for the optimization of GMM. To eliminate the camera jitter effects, the affine transformation operator is involved that acts to align the successive frames. The efficiency of the proposed method is augmented using a subsampling technique that can accelerate the proposed method to execute on an average

more than 250 frames per second while maintaining good performance in accuracy.

**Acknowledgements** Special thanks to National University of Sciences and Technology, Islamabad, Pakistan for supporting this research work.

## References

1. Szwoch, G., Ellwart, D., Czyżewski, A.: Parallel implementation of background subtraction algorithms for real-time video processing on a supercomputer platform. *J. Real-Time Image Proc.* **11**, 111–125 (2016)
2. Erichson, N.B., Brunton, S.L., Kutz, J.N.: Compressed dynamic mode decomposition for background modeling. *J. Real-Time Image Proc.* **16**, 1479–1492 (2019)
3. Cocorullo, G., Corsonello, P., Frustaci, F., et al.: Multimodal background subtraction for high-performance embedded systems. *J. Real-Time Image Proc.* **16**, 1407–1423 (2019)
4. Sepúlveda, J., Velastin, S.A.: Evaluation of background subtraction algorithms using MuHAVi, a multicamera human action video dataset, vol. 12, no. 6 (2014)
5. Beleznai, C., Fruhstuck, B., Bischof, H.: Multiple object tracking using local PCA. In: 18th International Conference on Pattern Recognition (ICPR’06), vol. 3, pp. 79–82. IEEE (2006)
6. Cheung, S.C.S., Kamath, C.: Robust background subtraction with foreground validation for urban traffic video. *EURASIP J. Adv. Signal Process.* **14**(726261), 2330–2340 (2005)
7. Senior, A.W., Tian, Y., Lu, M.: Interactive motion analysis for video surveillance and long term scene monitoring. In: Asian Conference on Computer Vision, pp. 164–174. Springer, Berlin, Heidelberg (2010)
8. Cao, W., Wang, Y., Sun, J., Meng, D., Yang, C., Cichocki, A., Xu, Z.: A novel tensor robust PCA approach for background subtraction from compressive measurements. *CoRR arXiv:abs/1503.01868* (2016)
9. Wang, N., Yao, T., Wang, J., Yeung, D.Y.: A probabilistic approach to robust matrix factorization. In: European Conference on Computer Vision. Springer, Berlin, Heidelberg, pp. 126–139 (2012)
10. He, J., Balzano, L., Szlam, A.: Incremental gradient on the Grassmannian for online foreground and background separation in subsampled video. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1568–1575. IEEE (2012)
11. Xu, J., Ithapu, V.K., Mukherjee, L., Reh, J.M., Singh, V.: GOSUS: Grassmannian online subspace updates with structured-sparsity. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3376–3383 (2013)
12. Yong, H., Meng, D., Zuo, W., Zhang, L.: Robust online matrix factorization for dynamic background subtraction. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(7), 1726–1740 (2018)
13. Wang, X., Liu, L., Li, G., Dong, X., Zhao, P., Feng, X.: Background subtraction on depth videos with convolutional neural networks. In: 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–7. IEEE (2018)
14. Ratnayake, K., Amer, A.: Embedded architecture for noise-adaptive video object detection using parameter-compressed background modeling. *J. Real-Time Image Proc.* **13**, 397–414 (2017)
15. Jaiswal, D., Kumar, P.: Real-time implementation of moving object detection in UAV videos using GPUs. *J. Real-Time Image Proc.* **17**, 1301–1317 (2020)
16. Mabrouk, L., Huet, S., Houzet, D., et al.: Efficient adaptive load balancing approach for compressive background subtraction

- algorithm on heterogeneous CPU-GPU platforms. *J. Real-Time Image Proc.* **17**, 1567–1583 (2020)
17. Wang, B., Dudek, P.: A fast-tuning background subtraction algorithm. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 395–398 (2014)
  18. Yadav, D.K., Sharma, L., Bharti, S.K.: Moving object detection in real-time visual surveillance using background subtraction technique. In: *2014 14th International Conference on Hybrid Intelligent Systems*, pp. 79–84. IEEE (2014)
  19. Boulmerka, A., Allili, M.S.: Background modeling in videos revisited using finite mixtures of generalized Gaussians and spatial information. In: *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 3660–3664. IEEE (2015)
  20. Ali, S.T., Goyal, K., Singhai, J.: Moving object detection using self-adaptive Gaussian mixture model for real time applications. In: *2017 International Conference on Recent Innovations in Signal processing and Embedded Systems (RISE)*, pp. 153–156. IEEE (2017)
  21. Javed, S., Ho Oh, S., Sobral, A., Bouwmans, T., Ki Jung, S.: Background subtraction via superpixel-based online matrix decomposition with structured foreground constraints. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 90–98 (2015)
  22. Feng, J., Xu, H., Yan, S.: Online robust PCA via stochastic optimization. In: *Advances in Neural Information Processing Systems*, pp. 404–412 (2013)
  23. Javed, S., Sobral, A., Bouwmans, T., Jung S.K.: OR-PCA with dynamic feature selection for robust background subtraction. In: *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, pp. 86–91 (2015)
  24. Chen, R., Li, H.: Online algorithm for foreground detection based on incremental nonnegative matrix factorization. In: *2016 2nd International Conference on Control, Automation and Robotics (ICCAR)*, pp. 312–317. IEEE (2016)
  25. Javed, S., Mahmood, A., Bouwmans, T., Jung, S.K.: Spatiotemporal low-rank modeling for complex scene background initialization. *IEEE Trans. Circuits Syst. Video Technol.* **28**(6), 1315–1329 (2016)
  26. Ebadi, S.E., Ones, V.G., Izquierdo, E.: Dynamic tree-structured sparse RPCA via column subset selection for background modeling and foreground detection. In: *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 3972–3976. IEEE (2016)
  27. Zeng, D., Zhu, M.: Background subtraction using multiscale fully convolutional network. *IEEE Access* **2018**(6), 16010–16021 (2018)
  28. Wang, X., Liu, L., Li, G., Dong, X., Zhao, P., Feng, X.: Background subtraction on depth videos with convolutional neural networks. In: *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7. IEEE (2018)
  29. Wang, Y., Zhu, L., Yu, Z.: Foreground detection for infrared videos with multiscale 3-D fully convolutional network. *IEEE Geosci. Remote Sens. Lett.* **16**(5), 712–716 (2019)
  30. Buchanan, A.M., Fitzgibbon, A.W.: Damped newton algorithms for matrix factorization with missing data. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, vol. 2, pp. 316–322. IEEE (2005)
  31. Meng, D., De La Torre, F.: Robust matrix factorization with unknown noise. *IEEE Int. Conf. Comput. Vis.* **2013**, 1337–1344 (2013)
  32. Gillis, N., Glineur, F.: Low-rank matrix approximation with weights or missing data is NP-hard. *SIAM J. Matrix Anal. Appl.* **32**(4), 1149–1165 (2011)
  33. Eriksson, A., Van Den Hengel, A.: Efficient computation of robust low-rank matrix approximations in the presence of missing data using the  $L_1$  norm. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 771–778. IEEE (2010)
  34. Tseng, P.: Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optim. Theory Appl.* **109**(3), 475–494 (2001)
  35. Brownrigg, D.R.K.: The weighted median filter. *Commun. ACM* **27**(8), 807–818 (1984)
  36. Meng, D., Xu, Z., Zhang, L., Zhao, J.: A cyclic weighted median method for  $L_1$  low-rank matrix factorization with missing entries. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 27, no. 1 (2013)
  37. Candès, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? *J. ACM (JACM)* **58**(3), 1–37 (2011)
  38. Candès, E.J., Recht, B.: Exact matrix completion via convex optimization. *Found. Comput. Math.* **9**(6), 717–772 (2009)
  39. He, J., Balzano, L., Szelam, A.: Incremental gradient on the Grassmannian for online foreground and background separation in subsampled video. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1568–1575. IEEE (2012)
  40. Zhou, X., Yang, C., Yu, W.: Moving object detection by detecting contiguous outliers in the low-rank representation. *Pattern Anal. Mach. Intell. IEEE Trans.* **35**(3), 597–610 (2013)
  41. Wang, N., Yao, T., Wang, J., Yeung, D.Y.: A probabilistic approach to robust matrix factorization. In: *European Conference on Computer Vision*, pp. 126–139. Springer, Berlin, Heidelberg (2012)
  42. Peng, Y., Ganesh, A., Wright, J., Xu, W., Ma, Y.: RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(11), 2233–2246 (2012)
  43. He, J., Zhang, D., Balzano, L., Tao, T.: Iterative Grassmannian optimization for robust image alignment. *Image Vis. Comput.* **32**(10), 800–813 (2014)
- Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
- Wajiha Munir** is currently doing her MS from National University of Sciences and Technology (NUST). Her research interest is in the field of image and signal processing.
- Adil Masood Siddiqui** completed his bachelor in telecommunications engineering, masters in electronics and telecommunication and Ph.D. in electrical from University of Engineering and technology in 1994, 2005 and 2009 respectively. His research interest includes image registration, image de-noising, image enhancement and defogging. After successfully completing Ph.D., he rejoined MCS (NUST) in 2009, where he is currently teaching various courses related to signal and image processing to undergraduate and postgraduate students and supervising postgraduate (MS and PhD) thesis besides other academic and research activities. He has contributed over 40 research publications which include 21 Journals.
- Muhammad Imran** received his Masters and Ph.D. degrees in Electrical Engineering from NUST, in 2011 and 2014, respectively. He received his BE degree in Information and Communication System Engineering from NUST, in 2007. His research interests include model/Controller order reduction, Control Systems theory, Communication Systems and Signal Processing. He is an active researcher and produced many publications in well reputed journals and conferences. Currently he is performing duties of Associate HoD of Electrical Engineering Department at MCS.

**Imran Tauqir** is a faculty member at MCS (NUST) since 2009. He is an active researcher and produced many publications in well reputed journals and conferences.

**Waseem Iqbal** is a faculty member at MCS (NUST) since 2012. He is an active researcher and produced many publications in well reputed journals and conferences.

**Nazish Zulfiqar** has completed her MS degree from College of Signals, National University of Sciences and Technology, Islamabad. Her research interest includes image and signal processing.