# FOREGROUND ESTIMATION VIA STRUCTURED GAUSSIAN SCALE MIXTURE MODELLING

[1]Sindhia L, [2]Narentharan A, [3]Keerthana N, [4]Keerthana B
[1] Assistant professor, [234]Student
[1234] Sri Shakthi Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India.

**Abstract-** Video Analysis is a field within Computer Vision that involves the automatic interpretation of digital video using computer algorithms. Developing algorithms for the computer to perform the task has been highly evasive and is now an active research field. Extracting the background and foreground parts from video has important applications in video surveillance. The background parts are supposed to be the case of stationary part and the foreground are thinly dispersed, Robust Principal Component Analysis (RPCA) is the existing method for background and foreground parts estimation. In realistic complex events, the conventional $l_1$ norm thinly dispersed regularizer often fails to well characterize the varying thinly dispersity of the foreground components. To select the thinly dispersed regularize parameters adaptively according to the local statistics is critical to the success of the RPCA framework for background subtraction. Structure Gaussian scale mixture (SGSM) is used to model the thinly dispersed component. In this method, the input frame is divided into many similar regions by super-pixel segmentation. By characterizing the set of thinly dispersed coefficients in each homogeneous region with the same SGSM prior, the local dependencies among the thinly dispersed coefficients can be effectively exploited. Experiment output shows that the proposed SGSM method performs better than most of current background subtraction methods in terms of implementation and speed.

**Keywords- Background subtraction method, Structured Gaussian scale mixture Model, Super pixel segmentation.**

## I. INTRODUCTION

Video object tracking [41] is a significant assignment inside the domain of Computer Vision. Detection of moving objects in video frame has broad applications like vehicle navigation, augmented reality, scene understanding, video surveillance, computer interaction and military [22]. One of the commonly used approaches for moving object detection is background subtraction [14], which aims to separate moving objects from background images to generate the masks of objects in motion [11]. Considering the object's initial state in very first frame of an image sequence, it aims to track the position of object in the remaining frames. Some of the challenging factors in this approach are illumination change, geometric deformation, occlusion, bad weather and background clutter [5].

Various algorithms have been proposed for robust object tracking [11] in recent years that aims on constructing effective model by capturing high-level structural cues or low-level vision cues. Low-level vision cues are captured by several operators such as color, texture, gradient, local histogram or local binary patterns [28]. Since it has limited performance. They are not effective in representing targets. For example, Haar-like features have the ability in handling severe appearance and illumination changes [26]. High-level structural cues generally convey prior knowledge about target objects such as contour or geometry. For example, an object silhouette to track non-rigid target objects during occlusion situations. They also need to about object's shape to obtain preferable performance.

Considering the images in the background are correlated, Robust Principal Object Analysis (RPCA) [4] used for subtracting the background. In this method each frame is converted into a column of data at first and then the resulting matrix is changed into low rank background matrix and sparse foreground matrix [42]. This RPCA-based method can exploit the temporal correlations of the background. Though it describes the correlations among each object in the background the classic sparse model which corresponds independent and identical distributed Laplace model which cannot accurately model the moving objects in complex scenarios [37]. Markov Random Field (MRF) model is used to eliminate this downfall in RPCA model. Some improved RPCA model tried to overcome the scale issue problem i.e., universal standardization of parameter cannot handle the foreground objects of varying sizes [9].

To utilize the structural correlations among every pixels of the video frames, spatially-consistent methods have also been developed [43]. For more effective low-rank decomposition, super pixel-based online decomposition method was proposed [15]. Minimum Spanning Tree (MST) has been introduced for imposing smoothness constrained on similar pixels. Supposing that the real-world backgrounds generally span numerous manifolds, spatial temporal sparse spectral [8] clustering regularization for RPCA has also been proposed [23]. By imposing sequentially on low-dimensional multiple manifolds, this method handles the dynamic background efficiently [31].

Rest of the paper is organized as follows, Section I contains the introduction of Foreground Estimation Via Structured Gaussian Scale Mixture Modelling, Section II contains the literature survey, Section III contain the Existing methods of background subtraction methods, Section IV contains methodology, Section IV results and experiments, section VI conclusion of research work.

## II. RECENT RELATED RESEARCHES: A SURVEY

T.S.F. Haines [14] have anticipated a Dirichlet process Gaussian mixture model to estimate a per-pixel background distribution which is followed by probabilistic regularization.it is a non-parametric Bayesian method that automatically estimates the number of mixture components forming as more information becomes available and handles multi-modal Dynamic background with regular changes .combining information between pixels only as a regularization step does not fully exploit the information available and so a rigorous method of spatial information.

R. Wang [35] has introduce flux tensor with split Gaussian models that exploits the benefits of fusing a motion computation method based on spatio temporal tensor formulation and a multi-cue appearance comparison. This hybrid system can handle challenges such as shadows, illumination variations, noise and occlusions. Its model foreground and background separately, and use adaptively changing number of Gaussians for the background model. The final multi-cue object level classification distinguishes stopped objects form background revealed by removed objects and thus reduces false positives

X. Liu [24] have anticipated a low-rank method. In this method moving objects in the scene are modeled as pixel-wised sparse outliers.it introduce a class of structured sparsity-inducing norms to model moving objects in videos.it propose a saliency measurement to dynamically estimate the support of the foreground. A matrix composed of the observed video frames can be decomposed into a low-rank matrix representing the background and a sparse matrix consisting of the foreground objects treated as the sparse foreground. The total variation penalty on the sparse deviations was employed to better handle noisy data. After the decomposition, many foreground candidates will be detected and then a motion saliency map will be used to remove background motion and weigh the motion salient groups.

J. Yang [36] have projected a motion-assisted matrix restoration (MAMR) model for foreground-background separation from video clips. A dense motion field is estimated for each frame to facilitate efficient. The method is quite versatile for surveillance videos with different types of motions and lighting conditions and outperforms many other state-of-the-art methods.

Mohand Saïd Allili [21] combining temporal and spatial information for robust online background subtraction (BS) in videos. Temporal information is modelled by coupling finite mixtures of generalized Gaussian distributions. Spatial information is modelled by combining multiscale inter-frame correlation analysis and histogram matching. Obtained results have shown that the approach surpasses several state-of-the-art methods on the aforementioned challenges while maintaining comparable computational time.

Hajer Fradi [3] it proposes incorporating a uniform motion model into GMM background subtraction. By considering these both cues, high accuracy of foreground segmentation is obtained. This approach has been experimentally validated showing better segmentation.

## III. EXISTING WORKS

This portion, we analyze the works interconnected to the Gaussian Scale Mixture -BS (background subtraction) method. The scale mixture models and Robust Principal Component Analysis -BS (background subtraction) methods that have been used in the applications of image refurbishment. In the Robust Principal Component Analysis –BS (background subtraction) methods [9], the observed K video frames of size l × m formulated as matrix H, is decomposed as $H = W + Y$ where $H \in \mathbb{R}^{k\times K} = \{h1, \ldots\ldots h_N\}$ is formed by vectoring each frame as a column of the matrix, $W \in \mathbb{R}^{k\times K} = \{w1, \ldots\ldots w_N\}$and $Y \in \mathbb{R}^{k\times K} = \{y1, \ldots\ldots y_N\}$represents the foreground and background components, respectively, and $q = l \times m$ Denote by $h_{n,t}$the $n$th column of the matrix $H$ and by $H_{n,t}$ the $n$th pixel at the $t$th frame $h_t$. The background images have strong interdependences, the background matrix W should be modest rank, whereas the foreground matrix $Y$ containing moving segments e.g., bike or people are usually thinly distributed. Thus, the estimation of $W$ and $Y$ from $H$ can be considered as the RPCA problem

$$\min_{W,Y} rank(W) + \lambda\|Y\|_0, \qquad y.t. H = W + Y, \tag{1}$$

Where rank (·) indicates the rank of a matrix. As both the rank and the $l_0$ of above Equation are nonconvex and NP hard, they are often replaced with their convex substitutes

$$\min_{W,Y}\|W\|_* + \lambda\|Y\|_1, \qquad y.t. H = W + Y, \tag{2}$$

Where $\|.\|_*$ is the sum of the singular values. By choosing a suitable Lagrange multiplier, the contrived optimization problem can be changed into a non-contrived one

$$\min_{W,Y}\|H - W - Y\|_2^2 + \eta\|W\|_* + \lambda\|Y\|_1, \tag{3}$$

Which can be implemented by alternative optimization. In the Robust Principal Component Analysis model (RPCA), the choosing of the parameter is deprecator to the accomplishment of this model. There is a scale issue of the moving objects, it is very difficult to select a global regularization parameter λ that are appropriate for moving objects. The $l_1$ norm thinly dispersed regularization cannot accomplish the interdependences between the nearest pixels. The scale problem is being recovered by two-phases of Robust Principal Component Analysis methods have been used, which produce roughly detected moving objects and then the regularization parameters are being set based on the assumed motion prominence of foregrounds. The performance of two-phases of the method highly depend on the perfection of the beginning detection of the foregrounds and still cannot satisfy performance. The complication of the mathematical calculations of these methods are very risky. As an ideal statistical model, GSM model [37] is used to model the thinly dispersed coding coefficients for image refurbishment. Similarly, in the LSM model has been used for mixed noise removal. Outstanding success by the GSM/LSM modeling for thinly dispersed representation coefficients, in this paper we propose to model the thinly dispersed moving objects with the Gaussian Scale Mixture models, which will solve the scale-issues of Robust Principal Component Analysis model.

## IV. METHODOLOGY

### a) Foreground Estimation via Gaussian Mixture Model

This segment, it presents a MAP (Maximum a Posterior) estimator for evaluating $W$ and $Y$ from the perceived $H$. Using the MAP estimator, the evaluation of $W$ and $Y$ from $H$ can be developed as

$$(W, Y) = argmax \ \log P(H|W, Y) + \log P(W) + \log P(Y) \tag{4}$$

The probability term $P(H|W,Y)$ can be distinguished by the zero-mean Gaussian distribution with variance $\sigma^2{}_{w}$.

$$P(H|W,Y) = \frac{1}{\sqrt{2\pi\sigma_\omega}} \exp(-\frac{\|H-W-Y\|_F^2}{2\sigma_\omega^2}) \quad (5)$$

It can be demonstrated as

$$P(W) \propto \frac{1}{c}\exp(-\eta\|W\|_*), \quad (6)$$

where c is constant. It is easy to validate that if $P(Y)$ is modeled with an independent and identically dispensed zero-mean Laplacian model, $P(Y_{n,t}) \propto \frac{1}{2\alpha_{i,t}} \exp(-\frac{Y_{n,t}}{\alpha_{n,t}})$,

The MAP of Eq. (7) can be demonstrated as

$$(W,Y) = arg\min_{W,Y} \frac{1}{2}\|H-W-Y\|_F^2 + \eta\sigma_\omega^2\|W\|_* + \sum_{n,t}\lambda_{n,t}|Y_{n,t}| \quad (7)$$

Where $\lambda_{n,t} = \sigma^2/\alpha_{n,t}$ , $\alpha_{n,t}$ is the standard deviation of $Y_{n,t}$ .

Each foreground pixels with the Gaussian Scale Mixture (GSM) models. In GSM modeling, each pixel $Y_{n,t}$ is disintegrated into a product of a random Gaussian variable $\pi_{n,t}$ and a positive hidden multiplier $\alpha_{n,t}$, i.e., $Y_{n,t} = \alpha_{n,t}\ \pi_{n,t}$ . Each foreground pixel $Y_{n,t}$ is then modeled as a zero-mean Gaussian distribution of standard deviation $\alpha_{n,t}$ . By staggering a prior distribution $P(\alpha_{n,t})$ [25]over $\alpha_{n,t}$ and assuming that $\alpha_{n,t}$ and $Y_{n,t}$ are independent, the GSM modeling of Y can be developed

$$P(Y) = \prod_{n,t} P(Y_{n,t}), \qquad P(Y_{n,t}) = \int_0^\infty P(Y_{n,t}|\alpha_{n,t})P(\alpha_{n,t})\, d\theta_{n,t} \quad (8)$$

It has been known that the GSM model can well exhibit various scant dispensation, such as Laplacian, Generalized Gaussian, for an appropriate $P(\alpha_{n,t})$. We propose to use GSM to distinguish the sparse moving objects.

Generally, for most alternatives of, there is no inquisitive expression of P(Y) and thus it is hard to evaluate $P(Y_{n,t})$.with MAP. Although, such issue can be avoided by joint evaluation of $P(Y_{n,t}, \alpha_{n,t})$. By replacing $P(Y_{n,t})$ With. In the MAP estimator of Eq. (7), we obtain

$$(W,Y,Z) = argmax\ \log P(H|W,Y) + \log P(W) + \log P(Y|Z) + \log P(Z) \quad (9)$$

Where Z = [α1, α2, ⋯, αN] ∈ $\mathbb{R}^{q\times N}$ indicate the matrix of the positive multipliers. By exchange the prior of P (W) and the GSM preceding of $P(Y,Z)$ into this MAP estimator, we can acquire the following discriminatory function

$$(W,Y,Z) = \underset{W,Y,Z}{\text{argmin}} \frac{1}{2\sigma_\omega^2}\|H-W-Y\|_F^2 + \eta\|W\|_* + \sum_t\sum_n \frac{Y_{n,t}^2}{2\theta_{n,t}^2} + 2\sum_t\sum_n \log\theta_{n,t}, \quad (10)$$

Where we presumed that each pixel $Y_{n,t}$ are independent and $P(Y_{n,t}). = \frac{1}{\theta_{n,t}}$is used.

Y = Z ʘ B, where ʘ denotes the pixel-wise product and B = [π1, π2, ⋯, πN] ∈$\mathbb{R}^{p\times N}$is the matrix representation of the Gaussian random variables.

$P(\alpha_{n,t})$. = 1/$\alpha_{n,t}$. is numerical unstable, we present a small constant ε in $P(\alpha_n)$., a $P(\alpha_n). = \frac{1}{(\theta_n+\epsilon)}$Then, Eq. (10) can be rewritten as

$$(W,Z,B) = arg\min_{W,Z,B} \|H - W - Z \odot B\|_F^2 + 2\eta\theta_\omega^2 \|W\|_* + 4\sigma_\omega^2 \sum_t \sum_n \log(\theta_{n,t} + \varepsilon) + \sigma_\omega^2 \|B\|_F^2 \quad (11)$$

## b. Foreground Estimation via Structured Gaussian Scale Mixture Modeling

The proposed equitable function of Equation. (11), each foreground pixel$Y_{n,t}$are assumed to be independent and parallel distributed. Although, it is widely known that the adjacent pixels frequently have rugged connections. Pixels intimacy to the equivalent object or correspondent field should be characterized with the alike preparatory, i.e., the same$\alpha_{n,t}$ . To make utilize the association amid the adjacent pixels, we further extend the GSM model of Eq. (11) into structured GSM model,

$$P(Y) = \prod_t \prod_k \prod_{v\varepsilon J_k} P(Y_{v,t}), \quad (12)$$

$$P(Y_{v,t}) = \int_0^\infty P(Y_{v,t}|\alpha_{k,t})\, d\theta_{k,t}, \quad v\epsilon J_{k,t}, \quad (13)$$

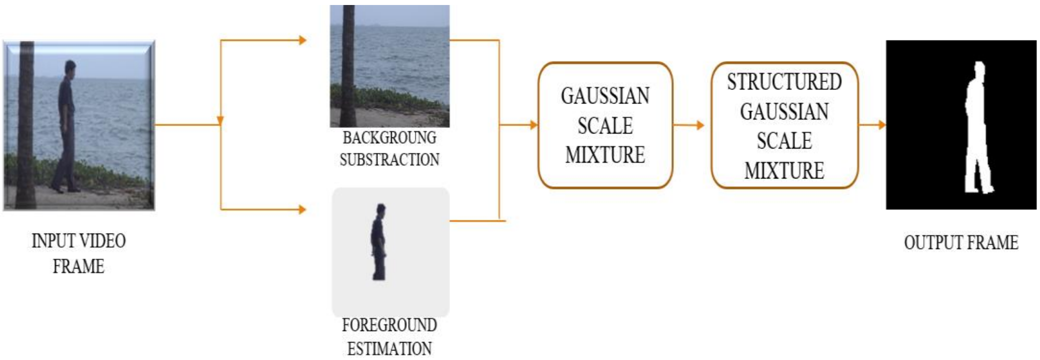
Where we have divided each foreground frame $Y_t$ into P component, indicated as $J_{k,t}$. Each component is conceded as an equivalent domain. Then, the foreground pixels in each $J_{k,t}$ are distinguished by the same second-order statistic $\alpha_{n,t}$. A pragmatic theme of Eq. (12) is that the foreground frame $Y_t$is concealed, and thus we cannot acquire the fragmentation and calculate the prior of Eq. (12). To circumvent such problem, we prefer to apply the fragmentation using the perceived frame $dt$. In this journal, we use the well-organized Entropy Rate super pixel Segmentation (ERS) method [23] to chunk each input frame $dt$ into K correspondent regions. An illustration of the fragmentation is exhibit in Fig. 1, from which we can see that the pixels in the frame can be well segmented into disparate congruent regions. Based on the pixel sorting using $dt$, we can sort the foreground pixels into K classes, symbolized as $J_{k,t}$.

By exchanging the organized GSM model of Eq. (13) into the MAP estimator of Eq. (8), the assembled GSM model-based foreground exemplary can be invented as

$$(W,Z,B) = arg\min_{W,Z,B} \sum_t \sum_k \sum_{v\in J_{k,t}} (H_{v,t} - W_{v,t} - \alpha_{k,t}\pi_{v,t})^2 + 2\eta\sigma_\omega^2 + \|W\|_* + 4\sigma_\omega^2 \sum_t \sum_k |J_{k,t}| \log \alpha_{k,t} + \epsilon) + \sigma_\omega^2 \|B\|_F^2, \quad (14)$$

Where the initial expression is the data expression symbolize in the pixel-wise structure, $J_{k,t}$ indicate the k-th groups of the pixels of the t fixture, $J_{k,t}|$ signifies the number of pixels in the sort of $J_{k,t}$ , and $\alpha_{k,t}$ represents the standard deviation of the pixels belong to $J_{k,t}$. By distributing the alike Gaussian parameter $\alpha_{k,t}$ over the sort of pixels, the reliance's between the adjacent pixels can be utilized, showing to further upgrade of the foreground estimations.

## Proposed method block diagram



## OPTIMIZATION ALGORITHM

Same as the standard RPCA problem, the proposed objective function of Equation. (14) Can be calculated by alternative updating the estimates of the background component W and the foreground component **Y**. For an initial estimate of **W**, we solve for the foreground Y by alternatively optimizing the positive multipliers **Z** and the Gaussian variables **B**, and update W with fixed **Y = Z ʘ B.**

### a. Solving the Y-Sub problem

Different from the standard RPCA problem, the estimate of the sparse component **Y** is obtained by alternatively estimating the positive multipliers **z** and the Gaussian variables **B**. For a given **W, Y= Z ʘ B** can be solved by minimizing

$$(W, Z, B) = arg\min_{Z,B} \sum_t \sum_k \sum_{v \in J_{k,t}} (H_{v,t} - W_{v,t} - \alpha_{k,t}\pi_{v,t})^2 + 4\sigma_\omega^2 \sum_t \sum_k |J_{k,t}| \log \alpha_{k,t} + \epsilon) + \sigma_\omega^2 \|B\|_F^2, \quad (15)$$

Note that solving Eq. (15) equals to solving a sequence of the following minimization problem for each frame, t =1, 2, ⋯, N

$$(W, Z, \alpha) = arg\min_{Z,\alpha} \sum_t \sum_k \sum_{v \in J_{k,t}} (H_{v,t} - W_{v,t} - \alpha_{k,t}\pi_{v,t})^2 + 4\sigma_\omega^2 \sum_t \sum_k |J_{k,t}| \log \alpha_{k,t} + \epsilon) + \sigma_\omega^2 \|B\|_F^2, \quad (16)$$

### b. Solving the W-Sub problem

For an incentive evaluation of Y, the background element W can be acquired by resolving

$$W= arg\min_{W} ||H\text{-}Y\text{-}W||^2_F + 2\eta\sigma^2_w||W|| \quad (17)$$

Which is a low-rank matrix estimation issue and can be resolved in a closed-form. Nevertheless, in each repetition of Singular Value Thresholding (SVT), the SVD (singular value decomposition) has to be conducted on the complete illustrative matrix, which is extremely slow and memory swallowing. Consequently, the modernizing of W using SVT is not acceptable for feasible application of background subtraction. In this journal, we arrogate the stochastic optimization approach of to solve the sub problem Eq. (17), which can notably minimize the computational complexity and the memory utilization. The fundamental purpose of the stochastic optimization algorithm is to demonstrate the nuclear norm of C whose grade is upper bounded by m as

$$||W||_* = \quad inf\, C\in R^{N\times m}, X\in R^{N\times m} \left\{\tfrac{1}{2}(||C||^2_F + ||X||^2_F)\; y.t.\; W=CX^T \right\} \quad (18)$$

Subsequently the low-rank estimation issue is reconstructed as a low-rank matrix factorization problem by treating C ∈ $\mathbb{R}^{n\times m}$ as the premise of the low-dimensional subdivision and V ∈ $\mathbb{R}^{n\times m}$ as the variation coefficients of W regarding C. With

Eq. (18), the background element W can then be evaluated by diminishing

$$(C, X) = argmin\,(C, X)\, ||H\text{-}CX^T\text{-}Y||^2_F + \eta\sigma^2_w(||C||^2_F + ||X||^2_F) \quad (19)$$

Where $X_t$ is the $t^{th}$ column of X. we can be seen that each $x_t$ can be resolved discretely construct on a perceived frame $d_t$.

## Algorithm

**Step1:** initialize the parameters $\eta$, $\sigma^2_{w,}$ $A_0 \in \mathbb{R}^{r*r} = 0$ and B0 ∈ $\mathbb{R}^{p*r=0.}$

**Step2**: create outer loop to access each input frame.

**Step3**: create inner loop to solve the equations mentioned in the proposed method.

**Step4**: update the initialized parameters $\eta$, $\sigma^2{}_{w,}$ $A_0 \in \mathbb{R}^{r*r} = 0$ and $B0 \in R^{p*r} = 0$.

**Step5**: subtract the background components and estimate the foreground components.

## V. RESULT AND DISCUSSION

In this section, we look into the performance of the proposed Structured GSM modeling-based background subtraction method on several real video sequences from the Perception Test Images Sequences (PITS) [12] and the change detection dataset [36]. To show the effects of the pixel grouping on the performance of the SGSM-BS method, we have also implemented a baseline of the proposed SGSM-BS method, i.e., enforcing the pixels $Y_{k,t}$ within a block to share the same regularization parameter $\alpha_{k,t}$. Now, we divide each frame into multiple non-overlapping blocks of size m ×m (m = 5 in our implementation) and model the foreground pixels in a block with the same Gaussian parameter $\alpha_{k,t}$. We denote this method as SGSM-BS-block. We also implemented the proposed Gaussian Scale Mixture based method without considering the correlations among neighboring pixels, which solves the objective function of Eq. (11). Similar to Eq. (12), Eq. (11) can also be solved by alternative optimization. The major parameters of the proposed method are empirically set as: r = 15, η = 400 √p, σ2 w = 1.05×10−3. Note that the parameters are fixed for all the test sequences, and the pixels of the test videos are normalized to the range of [0,1]. To calculate the performance of the proposed method, the criteria of recall along with precision are employed:

$$Precision = \frac{TP}{TP+FP} \qquad (19)$$

$$Recall = \frac{TP}{TP+FN} \qquad (20)$$

Where TP (true positives) is the total numbers of pixels correctly classified as foreground, FP (false positives) is the total numbers of pixels incorrectly classified as foreground, and FN (false negatives) is the total numbers of foreground pixels incorrectly classified as background. As the harmonic mean of the precision and recall.

### A.Perception Test Images Sequences (PITS)

The widely used Perception Test Images Sequences dataset contains 9 real sequences, that covers multiple scenarios including static background with short-time lingering objects (Hall, Shopping Mall and Bootstrap), dynamic background (Water Surface, Fountain, Campus, Escalator and Curtain) and sudden illumination changes (Lobby). The frame numbers of the test videos are between 523 and 3584, and only 20 ground truth frames of the foreground parts are provided for evaluation. To demonstrate the effectiveness of the proposed method, we compare the proposed method with several well-known methods, including the Detecting Contiguous outliers in the Low-Rank Representation (DECOLOR) method [42], the Smoothness and Arbitrariness Constraints (SAC) method [13], the Motion Saliency Detection (MSD) method5 , the Principle Component Pursuit (PCP) method Component Pursuit (PCP) method [38] the Probabilistic Robust Matrix Factorization (PRMF) method, and several recently proposed state-of-the-art RPCA based methods: Low-Rank and, and several recently proposed state-of-the-art RPCA based methods: Low-Rank and Structured Spare Decomposition (LRSSD) method[23], the TV regularized RPCA (TVRPCA) method [5], and the Modified Linear Regression with Basis Selection (MLRBS) method [34]. Two pixel-based methods, i.e., Pixel-Based Adaptive Segmenter (PBAS) [15] and the improved Gaussian Mixture Model (GMM) [43] are also included.

In Fig 1 Background subtraction results of 9 videos from PTIS dataset. (a) Original frame; (b) Background obtained by the proposed SGSM-BS method; (c) Ground truth foreground mask; Foreground mask obtained from (d) DECOLOR [42]; (e) SAC [13]; (f) PCP [38]; (g) PRMF [17]; (h) PBAS [15]; (i) GMM [43]; (j) Proposed GSM-BS; (k)Proposed SGSM-BS-block; (l)Proposed SGSM-BS. (White represents correctly detected foreground, red represents missing pixels, and blue represents false alarm pixels).

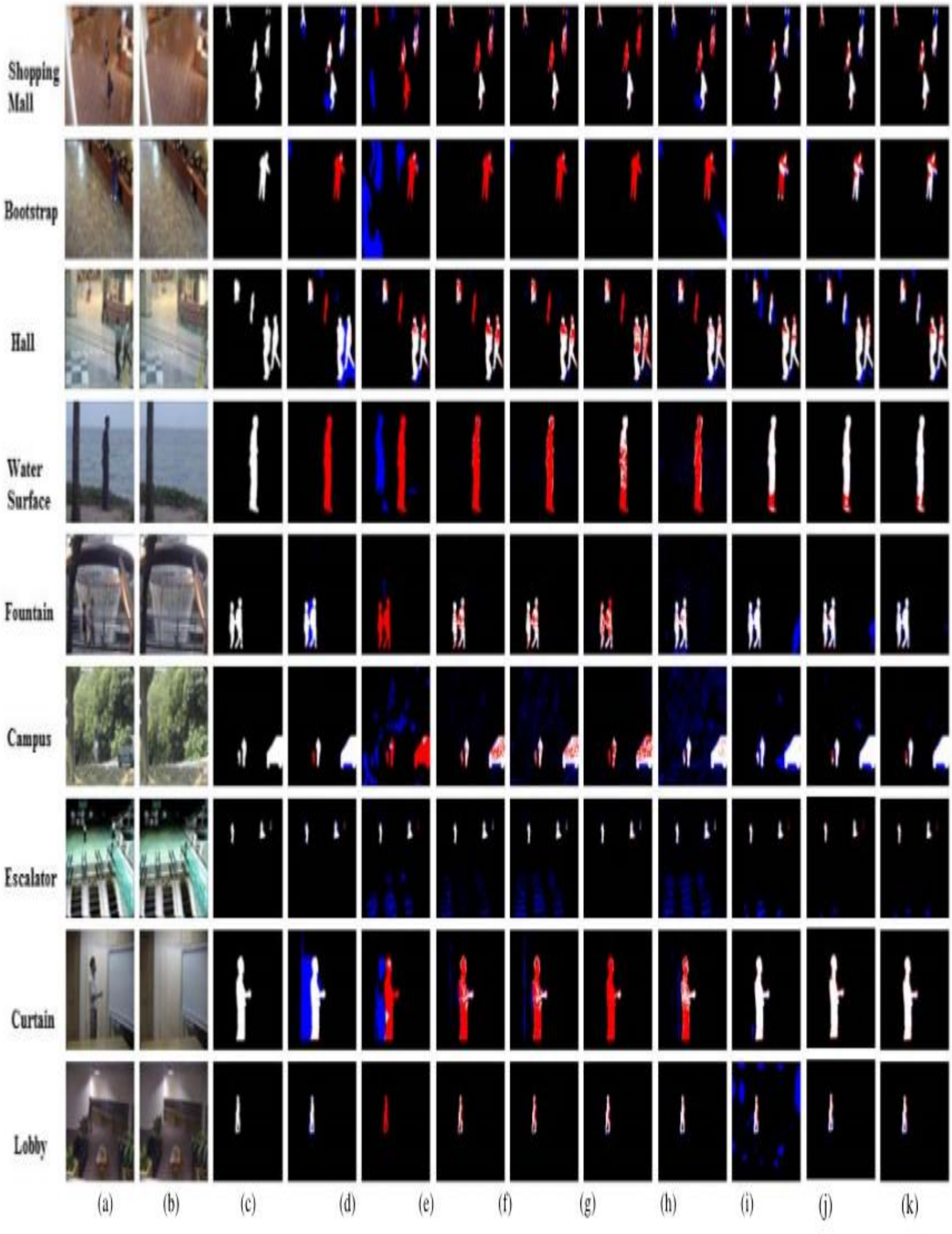



**Fig 1 Background subtraction results of 9 videos from PTIS dataset.**

**Table 1**

**Performance Of F-Measure (%) On Perception test Images Sequences Dataset**

| Video / method | Shopping Mall | Lobby | Curtain | Fountain | Campus | Hall | Bootstrap | Water Surface | Average |
|---|---|---|---|---|---|---|---|---|---|
| DECOLOR | 68.06 | 56.52 | 81.58 | 82.76 | 73.29 | 53.20 | 56.86 | 79.65 | 69.71 |
| SAC | 74.07 | 80.29 | 89.76 | 75.44 | 67.79 | 66.73 | 68.41 | 87.96 | 74.89 |
| MSD | 73.34 | 46.88 | 30.25 | 63.31 | 66.75 | 49.87 | 54.62 | 27.74 | 53.25 |
| PCP | 66.78 | 62.18 | 81.28 | 63.74 | 40.69 | 50.25 | 56.78 | 39.76 | 58.33 |
| PRMF | 69.68 | 52.13 | 58.94 | 53.80 | 36.58 | 58.88 | 59.62 | 35.40 | 51.42 |
| LRSSD | 73.62 | 73.13 | 83.57 | 83.71 | 76.13 | 72.22 | 58.42 | 90.50 | 75.93 |
| MLRBS | 74.74 | 70.93 | 87.25 | 86.52 | 70.17 | 67.64 | 69.72 | 87.24 | 76.72 |
| TVRPCA | - | 75.00 | - | 80.00 | 77.00 | 69.00 | 69.00 | 88.00 | - |
| PBAS | 71.77 | 56.74 | 72.48 | 62.04 | 69.50 | 61.92 | 55.38 | 89.02 | 63.33 |
| GMM | 67.25 | 42.62 | 30.52 | 50.06 | 34.77 | 42.17 | 66.36 | 32.88 | 46.34 |
| GSM-BS | 79.28 | 52.97 | 92.35 | 67.54 | 64.68 | 76.28 | 78.43 | 91.78 | 74.376 |
| SGSM-BS | 79.89 | 86.02 | 93.54 | 87.16 | 83.15 | 77.59 | 78.75 | 92.88 | 83.41 |

**Table 2**

**Performance Of F-Measure (%) On Dynamic backgrounds category Of The 2014 Cd Dataset**

| Video / Method | Boats | Cano | Fall | Fountain_1 | Founatin_2 | Overpass | Average |
|---|---|---|---|---|---|---|---|
| DECOLOR | 49.48 | 6.67 | 54.20 | 2.71 | 75.22 | 87.72 | 46.00 |
| PCP | 34.15 | 37.09 | 34.90 | 4.31 | 45.92 | 47.92 | 34.05 |
| LRSSD | 79.71 | 83.51 | 74.27 | 33.15 | 85.36 | 74.39 | 71.73 |
| MLRBS | 81.61 | 82.89 | 73.60 | 36.66 | 94.17 | 88.49 | 76.24 |
| TVRPCA | 39.00 | 86.00 | 51.00 | 12.00 | 72.00 | 77.00 | 61.00 |
| SuBSENSE | 69.32 | 79.23 | 86.61 | 75.31 | 94.41 | 85.72 | 81.70 |
| SOBS | 89.57 | 95.24 | 27.75 | 11.58 | 81.77 | 88.37 | 65.71 |
| PBAS | 36.11 | 71.96 | 87.14 | 41.73 | 93.55 | 79.25 | 68.29 |
| GMM | 74.74 | 81.51 | 42.38 | 8.15 | 79.17 | 42.17 | 54.68 |
| GSM-BS | 88.05 | 91.49 | 85.04 | 63.74 | 79.35 | 88.56 | 82.71 |
| SGSM-BS | 88.73 | 92.39 | 86.44 | 67.66 | 87.77 | 85.96 | 84.83 |

## B.Change Detection Dataset

As one of the most difficult detection benchmark dataset, the 2014 Change Detection (2014 CD) dataset has been proposed [36]. In the 2014 CD dataset, more than 70,000 frames have been captured and manually annotated. We have conducted experiments on all the categories of the CD dataset, except the PTZ category. As the videos of PTZ category were captured by the pan-tilt-zoom cameras and zooming mode, the backgrounds of the videos of PTZ category do not have the low-rank property, and thus the low-rank and sparse based methods (including our method) fail to deal with videos. In this section, we first present the comparison studies on three typical categories and then report the average results on all the categories except the PTZ category.

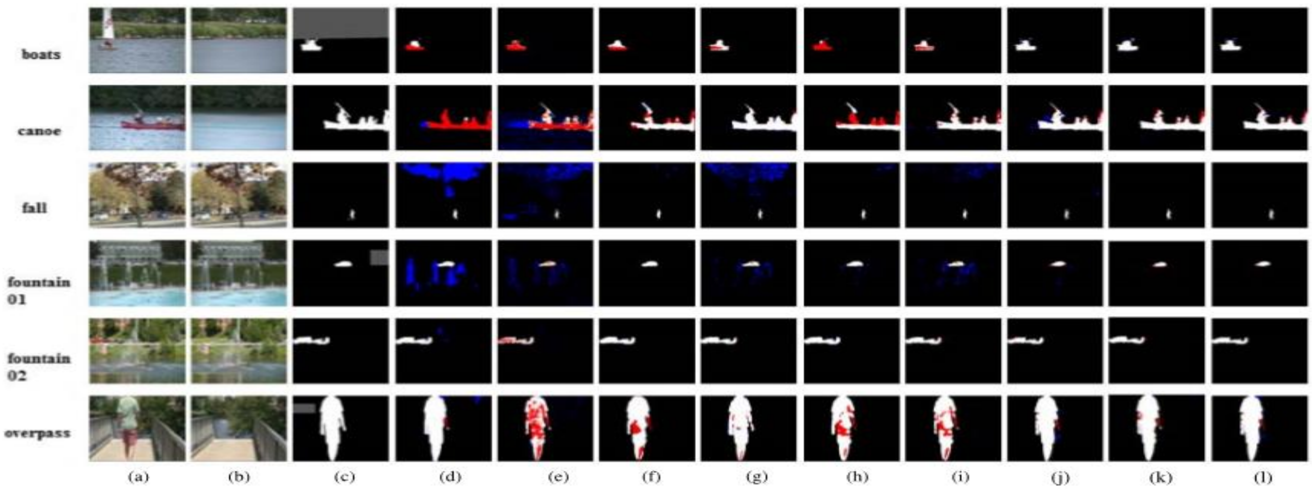



**Fig 2 Dynamic Background category of 2014 CD dataset.**

Fig 2 background subtraction results of 6 videos from Dynamic Background category of 2014 CD dataset. (a) Original frame; (b) Background obtained from the proposed SGSM-BS method; (c) Ground truth foreground mask; Foreground mask obtained from (d) DECOLOR [41]; (e) PCP [23]; (f) SuBSENSE; (g) SOBS[26]; (h) PBAS [15]; (i) GMM [43]; (j) Proposed GSM-BS; (k)Proposed SGSM-BS-block; (l)Proposed SGSM-BS. (White represents correctly detected foreground, red represents missing pixels, and blue represents false alarm pixels.

## C.Parameters Selection

The proposed method, there are only four free parameters needed to be tuned i.e., the regularization parameter η = γ1 √p, the variance of the approximation error σ2 w, the bilateral random projections number r and the number of super pixels K. the proposed method is insensitive to the parameter γ1. In our implementation, we set η = 400 √p. In general, the performance of the proposed method is robust to the values of σ2 w in the range of [0.9,1.2] ×10−3. In our implementation, we set σ2 w =1.05×10 −3. For the tradeoff between the performance and the computational complexity, we set r =15 in our implementation. We have also conducted experiments to verify the effects of the number of super pixels on the performance of the proposed SGSM-BS method.

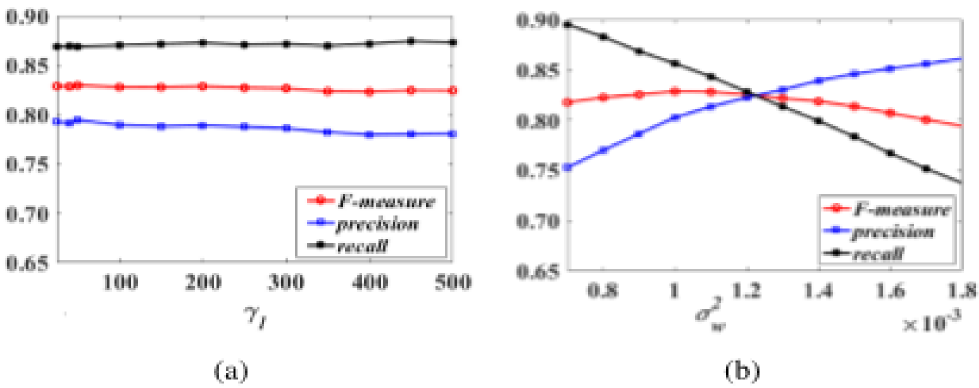



**Fig 3 The F-measure, precision, and recall curves as function of the regularization parameter γ1 & $\sigma^2_\omega$**

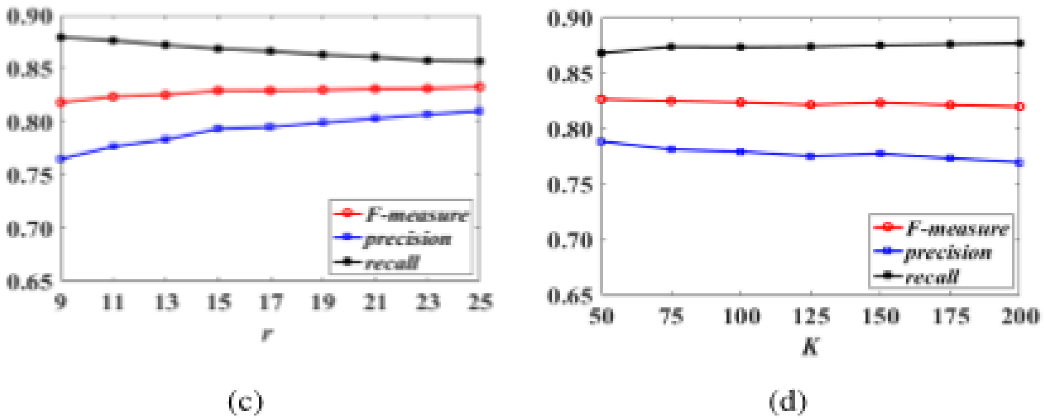



**Fig 4 The F-measure, precision, and recall curves as function of the number of bilateral random projections r and super pixels K.**

Fig 3 and Fig 4 the F-measure, precision, and recall curves as functions of (a) the regularization parameter γ1, (b) the regularization parameter σ2 w, (c) the number of bilateral random projections r, and the number of super pixels K.

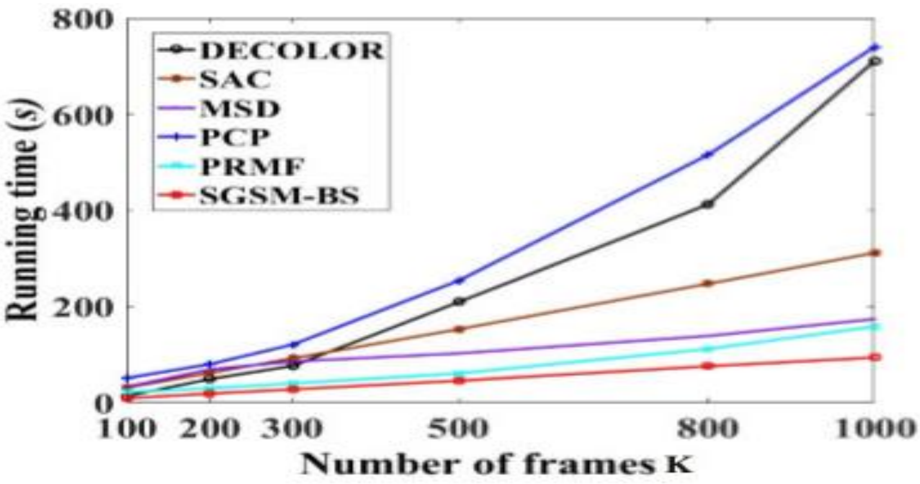



**Fig 5 The running time (seconds) as a function of the number of frames N**

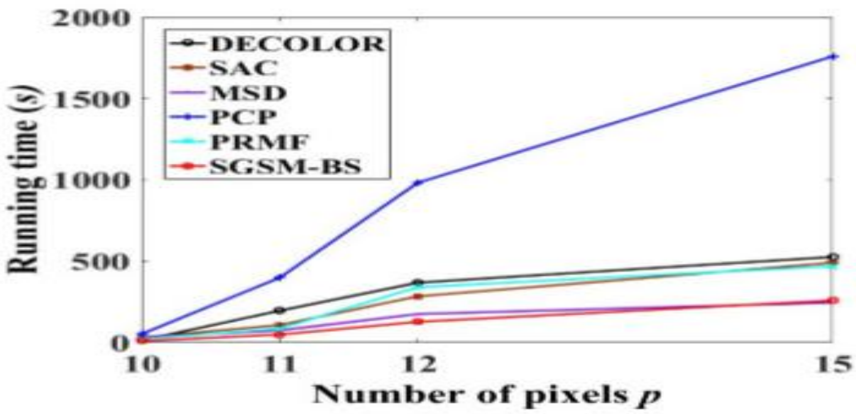



**Fig 6 The running time (seconds) as a function of the number of pixels p in log domain**

The running time (seconds) curves as functions of the number of frames K, and the number of pixels p in log domain. Fig 5 The running time (seconds) as a function of the number of frames K; Fig 6 The running time (seconds) as a function of the number of pixels p in log domain.

## VI. CONCLUSION

Thus, the RPCA framework has been used for background subtraction successful. The conventional $\ell_1$ norm based sparse regularizer cannot characterize the foreground components in varying sparsity. Compared with the ℓ norm, the GSM model based sparse regularizer can jointly estimating the regularizer parameters and the unknown sparse coefficients from the observed video, leading to significant improvements. Also, we further extend the GSM models to structured GSM models by considering the correlations among the neighboring pixels. By modelling the pixels inside each homogeneous region with the same GSM model, the foreground estimation accuracy can be improved. Experimental results on various scenarios show that the proposed method can perform much better than most of existing background subtraction methods in means of both performance and speed.

## REFERENCES

[1] Andrews. D.F. and C. L. Mallows, "Scale mixtures of normal distributions," J. Roy. Statist. Soc. Ser. B, vol. 36, no. 1, pp. 99–102, 1974.

[2] Barnich. O and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," IEEE Trans. Image Process., vol. 20, no. 6, pp. 1709–1724, Jun. 2011.

[3] Bouwmans. T, A. Sobral, S. Javed, S. K. Jung, and E.-H. Zahzah, "Decomposition into Low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset," Comput. Sci. Rev., vol. 23, pp. 1–71, Feb. 2016.

[4] Candès. E.J., X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" J. ACM, vol. 58, no. 3, 2011, Art. no. 11.

[5] Cao.X, L. Yang, and X. Guo, "Total variation regularized RPCA for irregularly moving object detection under dynamic background," IEEE Trans. Cybern., vol. 46, no. 4, pp. 1014–1027, Apr. 2016.

[6] Chen.M, X. Wei, Q. Yang, Q. Li, G. Wang, and M.-H. Yang, "Spatiotemporal GMM for background subtraction with superpixel hierarchy," IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 6, pp. 1518–1525, Jun. 2017.

[7] Cordts et al. M., "The cityscapes dataset for semantic urban scene understanding," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2016, pp. 3213–3223.

[8] Evangelio. R.H, M. Pätzold, and T. Sikora, "Splitting Gaussians in mixture models," in Proc. IEEE 9th Int. Conf. AVSS, Sep. 2012, pp. 300–305.

[9] Gao. Z, L.-F. Cheong, and Y.-X. Wang, "Block-sparse RPCA for salient motion detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 10, pp. 1975–1987, Oct. 2014.

[10] Garrigues. P and B. A. Olshausen, "Group sparse coding with a Laplacian scale mixture prior," in Proc. Adv. Neural Inf. Process. Syst., 2010, pp. 676–684.

[11] Gemignani. G and A. Rozza, "A robust approach for the background subtraction based on multi-layered self-organizing maps," IEEE Trans. Image Process., vol. 25, no. 11, pp. 5239–5251, Nov. 2016.

[12] Goyette. N, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Changedetection.Net: A new change detection benchmark dataset," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, Jun. 2012, pp. 1–8.

[13] Guo. X, X. Wang, L. Yang, X. Cao, and Y. Ma, "Robust foreground detection using smoothness and arbitrariness constraints," in Proc. Eur. Conf. Comput. Vis, 2014, pp. 535–550.

[14] Haines. T.S.F and T. Xiang, "Background subtraction with Dirichlet processes," in Proc. Eur. Conf. Comput. Vis., 2012, pp. 99–113.

[15] Hofmann. M, P. Tiefenbacher, and G. Rigoll, "Background segmentation with feedback: The pixel-based adaptive segmenter," in Proc. IEEEConf. Comput. Vis. Pattern Recognit. Workshops, Jun. 2012, pp. 38–43.

[16] Hu. W, Y. Yang, W. Zhang, and Y. Xie, "Moving object detection using tensor-based low-rank and saliently fused-sparsedecomposition," IEEE Trans. Image Process., vol. 26, no. 2, pp. 724–737, Feb. 2017.

[17] Javed. S, A. Mahmood, T. Bouwmans, and S. K. Jung, "Background– foreground modeling based on spatiotemporal sparse subspace clustering," IEEE Trans. Image Process., vol. 26, no. 12, pp. 5840–5854, Dec. 2017.

[18] Javed. S, T. Bouwmans, and S. K. Jung, "Improving OR-PCA via smoothed spatially-consistent low-rank modeling for background subtraction," in Proc. ACM Symp. Appl. Comput., 2017, pp. 89–94. [31] S. Javed, A. Mahmood, T. Bouwmans, and S. K. Jung, "Spatiotemporal low-rank modeling for complex scene background initialization," IEEE Trans. Circuits Syst. Video Technol., vol. 28, no. 6, pp. 1315–1329, Jun. 2016.

[19] Li. L, R. Socher, and L. Fei-Fei, "Towards total scene understanding: Classification, annotation and segmentation in an automatic framework," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2009, pp. 2036–2043.

[20] Li. L, P. Wang, Q. Hu, and S. Cai, "Efficient background modeling based on sparse representation and outlier iterative removal," IEEE Trans. Circuits Syst. Video Technol., vol. 25, no. 2, pp. 278–289, Feb. 2016.

[21] Lin. S, H. F. Cheng, W. Li, Z. Huang, P. Hui, and C. Peylo, "Ubii: Physical world interaction through augmented reality," IEEE Trans. Mobile Comput., vol. 16, no. 3, pp. 872–885, Mar. 2017.

[22] Liu. N, H. Wu, and L. Lin, "Hierarchical ensemble of background models for PTZ-based video surveillance," IEEE Trans. Cybern., vol. 45, no. 1, pp. 89–102, Jan. 2015.

[23] Liu. X, G. Zhao, J. Yao, and C. Qi, “Background subtraction based on low-rank and structured sparse decomposition,” IEEE Trans. Image Process., vol. 24, no. 8, pp. 2502–2514, Aug. 2015.
[24] Liu. Y, O. Tuzel, S. Ramalingam, and R. Chellappa, “Entropy rate superpixel segmentation,” in Proc. IEEE Conf. CVPR, Jun. 2011, pp. 2097–2104.
[25] Liu et al. X, “Background subtraction using spatio-temporal group sparsity recovery,” IEEE Trans. Circuits Syst. Video Technol., to be published.
[26] Maddalena. L and A. Petrosino, “The SOBS algorithm: What are the limits?” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, Jun. 2012, pp. 21–26.
[27] Matsushita. Y, Y.-W. Tai, and I. S. Kweon, “Fast randomized singular value thresholding for nuclear norm minimization,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 4484–4493.
[28] Roy. S and A. Ghosh, “Real-time adaptive Histogram Min-Max Bucket (HMMB) model for background subtraction,” IEEE Trans. Circuits Syst. Video Technol., to be published.
[29] Sobral. A, S. Javed, S. K. Jung, T. Bouwmans, and E.-H. Zahzah, “Online stochastic tensor decomposition for background subtraction in multispectral video sequences,” in Proc. IEEE Int. Conf. Comput. Vis. Workshops, Dec. 2015, pp. 946–953.
[30] Socher. R, and L. Fei-Fei, “Towards total scene understanding: Classification, annotation and segmentation in an automatic framework,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2009, pp. 2036–2043.
[31] Stauffer. C and W. E. L. Grimson, “Adaptive background mixture models for real-time tracking,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., vol. 2, Jun. 1999, pp. 246–252.
[32] Them et al. K, “Sensitivity enhancement in magnetic particle imaging by background subtraction,” IEEE Trans. Med. Imag., vol. 35, no. 3, pp. 893–900, Mar. 2016.
[33] Unzueta. M, M. Nieto, A. Cortés, J. Barandiaran, O. Otaegui, and P. Sánchez, “Adaptive multicue background subtraction for robust vehicle counting and classification,” IEEETrans. Intell. Transp. Syst., vol. 13, no. 2, pp. 527–540, Jun. 2012.
[34] Wang. N, T. Yao, J. Wang, and D.-Y. Yeung, “A probabilistic approach to robust matrix factorization,” in Proc. Eur. Conf. Comput. Vis, 2012, pp. 126–139.
[35] Wang. R, F. Bunyak, G. Seetharaman, and K. Palaniappan, “Static and moving object detection using flux tensor with split Gaussian models,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, Jun. 2014, pp. 420–424.
[36] Wang. Y, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, “CDnet 2014: “An expanded change detection benchmark dataset,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, Jun. 2014, pp. 393–400.
[37] Xin. B, Y. Tian, Y. Wang, and W. Gao, “Background subtraction via generalized fused lasso foreground modeling,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2015, pp. 4676–4684.
[38] Xue. Y, X. Guo, and X. Cao, “Motion saliency detection using low-rank and sparse decomposition,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Mar. 2012, pp. 1485–1488.
[39] Yang. J, X. Sun, K. Li, C. Hou, and Y. Wang, “Foreground– background separation from video clips via motion-assisted matrix restoration,” IEEE Trans. Circuits Syst. Video Technol., vol. 25, no. 11, pp. 1721–1734, Nov. 2015.
[40] Ye. X, J. Yang, X. Sun, K. Li, C. Hou, and Y. Wang, “Foreground– background separation from video clips via motion-assisted matrix restoration,” IEEE Trans. Circuits Syst. Video Technol., vol. 25, no. 11, pp. 1721–1734, Nov. 2015.
[41] Yilmaz. A, O. Javed, and M. Shah, “Object tracking: A survey,” ACM Comput. Surv., vol. 38, no. 4, 2006, Art. no. 13.
[42] Zhou. X, C. Yang, and W. Yu, “Moving object detection by detecting contiguous outliers in the low-rank representation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 3, pp. 597–610, Mar. 2013.
[43] Zivkovic. Z, “Improved adaptive Gaussian mixture model for background subtraction,” in Proc. 17th Int. Conf. Pattern Recognit., vol. 2, Aug. 2004, pp. 28–31.