

Moving Object Detection under Discontinuous Change in Illumination Using Tensor Low-Rank and Invariant Sparse Decomposition

Moein Shakeri, Hong Zhang
 Department of Computing Science, University of Alberta
 Edmonton, Alberta, Canada
 {shakeri, hzhang}@ualberta.ca

Abstract

Although low-rank and sparse decomposition based methods have been successfully applied to the problem of moving object detection using structured sparsity-inducing norms, they are still vulnerable to significant illumination changes that arise in certain applications. We are interested in moving object detection in applications involving time-lapse image sequences for which current methods mistakenly group moving objects and illumination changes into foreground. Our method relies on the multilinear (tensor) data low-rank and sparse decomposition framework to address the weaknesses of existing methods. The key to our proposed method is to create first a set of prior maps that can characterize the changes in the image sequence due to illumination. We show that they can be detected by a k -support norm. To deal with concurrent, two types of changes, we employ two regularization terms, one for detecting moving objects and the other for accounting for illumination changes, in the tensor low-rank and sparse decomposition formulation. Through comprehensive experiments using challenging datasets, we show that our method demonstrates a remarkable ability to detect moving objects under discontinuous change in illumination, and outperforms the state-of-the-art solutions to this challenging problem.

1. Introduction

Moving object detection in an image sequence captured under uncontrolled illumination conditions is a common problem in computer vision applications such as visual surveillance [26], traffic monitoring [5], and social signal processing [28]. Although moving object detection and background subtraction is a well established area of research and many solutions have been proposed, still most of the existing solutions are vulnerable to complex illumination changes that frequently occur in practical situations,



Figure 1: First row of each sequence: images captured in a industrial or wildlife monitoring system. Second row: results of our proposed method to detect foreground objects.

especially when the changes are discontinuous in time. In such cases, current methods are often not able to distinguish between illumination changes (including those due to shadow), and changes caused by moving objects in the scene. In general, outdoor illumination conditions are uncontrolled, making moving object detection a difficult and challenging problem. This is a common problem for many surveillance systems in industrial or wildlife monitoring areas in which a motion triggered camera or a time-lapse photography system is employed for detecting objects of interest over time. Fig. 1 shows four image sequences under discontinuous changes in illumination, which illustrate these applications. Due to significant and complex changes in illumination and independent changes of the moving objects between images of the sequences, detection of the moving objects is extremely challenging. The second row of each sequence in Fig. 1 shows the sample results of our proposed method with detected moving objects.

Among the leading methods for the problem addressed in this paper is a group based on low-rank and sparse decomposition. This group of methods exploit the fact that the background in an image sequence can be described as a low-rank matrix whose columns are image pixels that are correlated [20, 3]. However, image sequences with moving objects under discontinuous change in illumination and object location using the timer-lapse photography are qual-

itatively different from regular frame-rate video sequences. While some existing solutions are able to handle the discontinuity in object location with limited success, there is a need to improve their ability to distinguish between moving objects and changes due to illumination.

Taking the idea of using the low-rank components of a matrix to capture the image background, the most recent development relies on tensors, which are higher dimensional data structures than $2D$ -matrices. Since the real world data are ubiquitously multi-dimensional, tensors are often more appropriate than $2D$ -matrices to capture higher order relations in data. It is not surprising that tensor low-rank methods have been successfully developed with promising results on real-time video sequences. However, such methods have yet to be studied for detecting moving objects under discontinuous changes in illumination and object position, such as those found in time-lapse image sequences.

In this paper, we propose a solution to the problem of moving object detection within the tensor low-rank framework that specifically addresses the problem of discontinuous changes in illumination and object location. We formulate the problem in a unified framework named tensor low-rank and invariant sparse decomposition (TLISD). To separate illumination changes from moving objects, first we compute multiple prior maps as illumination invariant representations of each image to build our tensor data structure. These prior maps provide us with information about the effect of illumination in different parts of an image. We show that by defining two specific penalty terms using these prior maps, our proposed method is able to decompose an image into background, illumination changes and foreground objects, with a significant boost in performance of moving object detection.

The main contributions are as follows.

- We propose to use multiple priors to model the effect of illumination in natural images by exploiting invariance properties of color image chromaticity.
- We make use of the priors in a tensor representation for the problem of moving object detection.
- We propose a low-rank tensor decomposition using group sparsity and k-support norm as two regularization terms to separate moving objects and illumination variations that undergo discontinuous changes.
- We introduce an extended illumination change dataset with over 80k real images captured by motion trigger cameras in industrial and wildlife monitoring systems.

2. Related Work

One successful approach to moving object detection attempts to decompose a matrix D representing an image sequence into a low-rank matrix L and sparse matrix S , so as to recover the background and the foreground [2]. The

problem is initially solved by the robust principal component analysis (RPCA). Since the foreground objects are described by the sparse matrix S , we can categorize existing methods by the types of constraints on S . The first group of these methods use l_1 -norm to constrain S [3, 34, 29] and solve the following convex optimization.

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1 \quad s.t. \quad D = L + S \quad (1)$$

where $\|L\|_*$ denotes the nuclear norm of matrix L , and $\|S\|_1$ is the l_1 -norm of S .

The second group of methods used the additional prior knowledge on the spatial continuity of objects to constrain sparse matrix S and improve the detection accuracy [10, 6]. Using spatial continuity (e.g., $l_{2,1}$ -norm in [10]) to enforce the block-sparsity of the foreground, results become more stable than conventional RPCA in the presence of illumination changes. However, it remains a challenge to handle moving shadows or significant changes in illumination. Furthermore, the position of an object in a time-lapse image sequence is discontinuous from one image to another so that the continuity assumption is invalid as a way to separate moving objects and changes in illumination.

The third group of methods also imposed the connectivity constraint on S [32, 30, 35, 21, 31, 17] using other formulations than the second group. For example, Liu *et al.* [17] attempted to use a structured sparsity norm [19] and a motion saliency map, to improve the accuracy of moving object segmentation under sudden illumination changes. However, this method still cannot handle shadows and severe illumination changes, especially in time-lapse sequences with independent object locations among the images in the sequence that change similarly to shadow and illumination. In general, although the low-rank framework is well-known to be robust against moderate illumination changes in frame-rate sequences, the existing methods are still not able to handle discontinuous change in illumination and shadow, especially in time-lapse sequences.

To effectively separate discontinuous changes due to moving objects and those due to illumination, Shakeri *et al.* [23] proposed a method called LISD. This method relies on an illumination regularization term combined with the standard low-rank framework to explicitly separate the sparse outliers into sparse foreground objects and illumination changes. Although this regularization term can significantly improve the performance of object detection under significant illumination changes, LISD assumes a) the invariant representation [22] of all images in a sequence are modeled by only one invariant direction and b) all illumination variations are removed in the invariant representation of images, which are not strictly valid in practice.

Recently, multi-way or tensor data analysis has attracted much attention and has been successfully used in many applications. Formally and without loss of generality, denote

a 3-way tensor by $\mathcal{D} \in R^{n_1 \times n_2 \times n_3}$. Tensor low-rank methods attempt to decompose $\mathcal{D} \in R^{n_1 \times n_2 \times n_3}$ into a low-rank tensor \mathcal{L} and an additional sparse tensor \mathcal{S} [8]. This decomposition is applicable in solving many computer vision problems, including moving object detection. One of the most recent methods relevant to our research is proposed by Lu *et al.* [18]. A tensor nuclear norm was used to estimate the rank of tensor data and RPCA was extended from 2D to 3D to formulate the following tensor robust PCA (TRPCA):

$$\min_{\mathcal{L}, \mathcal{S}} \|\mathcal{L}\|_* + \lambda \|\mathcal{S}\|_1 \quad s.t. \quad \mathcal{D} = \mathcal{L} + \mathcal{S} \quad (2)$$

They showed that the tensor nuclear norm on tensor data can capture higher order relations in data. Tensor data is used for background subtraction and foreground detection [24, 13, 12, 4, 16] by stacking two dimensional images into a three dimensional data structure, using which tensor decomposition can capture moving object due to the continuity of object positions in the third dimension. Obviously, this approach only works for frame-rate sequences with continuous foreground motion, but is not applicable to time-lapse image sequences with discontinuous changes in both object location and illumination.

In this paper, we introduce a new formulation for moving object detection under the framework of tensor low-rank representation and invariant sparse outliers. We first build a set of prior maps for each image in the image sequence and treat it as a tensor. These prior maps enable us to use two regularization terms to distinguish between moving objects and illumination changes. We demonstrate that their use within our proposed method significantly improves the performance of moving object detection in the case of discontinuous changes in illumination, a problem that most of the existing methods cannot handle effectively.

3. Tensor Low-Rank and Invariant Sparse Decomposition

Our proposed formulation seeks to decompose tensor data \mathcal{D} into a low-rank tensor \mathcal{L} , an illumination change tensor \mathcal{C} , and a sparse foreground tensor \mathcal{S} as follows.

$$\mathcal{D} = \mathcal{L} + \mathcal{S} + \mathcal{C} \quad (3)$$

In (3), both \mathcal{S} and \mathcal{C} are stochastic in time-lapse image sequences due to discontinuous change in object locations and illumination changes, and separating them is an ill-posed problem. To solve this issue, we compute a set of prior maps using multiple representations of an image, which are more robust against illumination change than RGB images. These prior maps enable us to find higher order relations between the different invariant representations and the intensity images, in both space and time. These relations are exploited as the basis for separating \mathcal{S} from \mathcal{C} as will be detailed in Section 3.1. It is worth mentioning that on one hand, illumination changes are related to the material in a scene, which

is invariant in all frames leading to a correlation between them. On the other hand, these changes are also related to the source of lighting, which is not necessarily correlated between frames. Consequently, illumination changes should be accounted for by both the low-rank part and the sparse part in an image decomposition. In our method, we model the highly correlated part of illumination with the low-rank tensor \mathcal{L} as background, and we model the independent changes in illumination as the foreground, while recognizing that uncorrelated illumination changes are not necessarily sparse. To accomplish such illumination modeling, we propose to use a balanced norm or k -support norm. We introduce our formulation in details in Section 3.2, and we describe a solution to the formulation in Section 3.3.

3.1. Generation of Prior Maps and Tensor Data \mathcal{D}

In this section we focus on obtaining the prior information that will enable us to distinguish between moving objects and illumination changes in our proposed formulation. In the case of discontinuous change in illumination, which is common in time-lapse image sequences, variation of shadows and illumination are unstructured phenomena and they are often mistakenly considered by many methods as moving objects. We address this problem through creating illumination-invariant and shadow-free images, a problem that has been well studied.

One of the most popular methods for this problem is proposed by Finlayson *et al.* [7], which computes the two-vector log-chromaticity χ' using red, green and blue channels. [7] showed that with changing illumination, χ' moves along a straight line e roughly. Projecting the vector χ' onto the vector orthogonal to e , which is called invariant direction, an invariant representation $I = \chi' e^\perp$ can be computed. This method works well when the assumption defined above hold true but in practice this assumption never holds exactly, i.e., χ' does not move along a straight line. As a result, the correspond invariant representation is flawed and can lead to sub-optimal performance.

Fig. 2 shows an example of the variability of the illumination invariant direction in an image sequence and its impact on generating a illumination-invariant image representation. Fig. 2(a) shows the invariant directions of an image sequence of 200 frames while illumination changes (blue line), one direction for each image, varying mostly between -4° and 13° . Fig. 2(c) shows a selected image from the sequence, which is image 11 and corresponds to the red line in Fig. 2(a). The invariant direction for this image is found to be 13° while the average invariant direction of the sequence is around 5° , when we assume χ' moves exactly along a straight line. Fig. 2(d) compares the two invariant representations created with invariant directions of 5° and 13° , respectively, and Fig. 2(e) shows the detected foreground objects using these two different representations from the

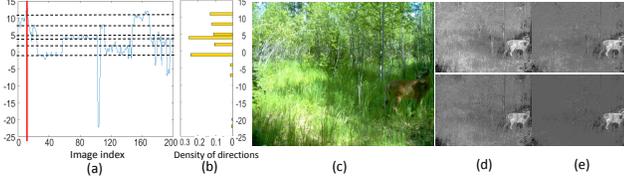


Figure 2: (a) Best invariant direction of each image in a sequence (y -axis: angle of the invariant directions e^\perp in degrees), (b) Dominant directions (yellow bars) after clustering, (c) 11^{th} image in the sequence as shown with a red line in (a), where its best invariant direction is 13° , (d) The first and the second rows show the invariant representations of the selected image using the average direction of the sequence (5°) and its best direction (13°), respectively. (e) Obtained outliers of the invariant representations.

RPCA method where the use of the optimal invariant direction (13°) produces much more desirable result than that of the sub-optimal direction (5°). This example clearly shows the importance of the choice of the invariant direction in creating the invariant representations, and the undesirable outcome when these representations are created with a sub-optimal invariant direction.

Our idea to account for the difference in the invariant direction among the images in the sequence, is to first estimate the image-specific invariant directions for the sequence, and then use a clustering algorithm to identify the dominant directions (dotted lines in Fig. 2(a) or the dominant yellow bars in Fig. 2(b)). Subsequently, for each image, we create multiple invariant representations, one for each dominant direction, and these multiple representations serve as multiple prior maps for the image. In particular, for each image, we first use the method in [7] to determine its best invariant direction. With n_2 images in an image sequence, this results in n_2 invariant directions where $n_2 = 200$ in Fig. 2(a). Second, we use k-means to identify $k = 10$ clusters of the n_2 invariant directions. Third, we choose the centroid of a cluster as a dominant invariant direction if the cluster has support by at least 10% of the images (yellow bars in Fig. 2(b)). By definition, there are no more than 10 dominant directions.

Now, to construct the tensor $\mathcal{D} \in R^{n_1 \times n_2 \times n_3}$ formally

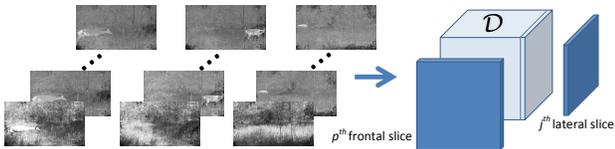


Figure 3: Right: sample images with their corresponding illumination invariant representations as prior maps. Left: Tensor \mathcal{D} . Frontal slices show p^{th} representation of the images in the sequence. Lateral slices show different representation of each image in the sequence.

(see Fig. 3), let $\mathcal{D}(:, :, 1)$ be an observed image sequence in our problem, where each column of $\mathcal{D}(:, :, 1)$ is a vectorized image from the sequence with n_1 pixels, and n_2 is the number of images in the sequence. p^{th} frontal slice $\mathcal{D}(:, :, p)$, $p = 2, \dots, n_3$ is a corresponding prior map, generated with a dominant invariant direction. Based on this tensor data structure, we are ready to present our new tensor low-rank and invariant sparse decomposition (TLISD) to extract the invariant sparse outliers as moving objects.

3.2. TLISD Formulation

As mentioned in Section 2, to detect moving objects under discontinuous illumination change in a sequence, current low-rank methods are insufficient when changes due to illumination and moving shadows are easily lumped with moving objects as the sparse outliers in the low-rank formulation. To separate real changes due to moving objects from those due to illumination, we use multiple prior illumination-invariant maps, introduced in Section 3.1, as constraints on real changes and illumination changes. In particular, real changes should appear in all frontal slices. Furthermore, lateral slices are completely independent from each other in a time-lapse sequence, but the different representations in each lateral slice (see Fig. 3) are from one image and therefore, the locations of real changes should be exactly the same in each lateral slice. Now, based on these observations, real changes in each frame should satisfy the group sparsity constraint, which is modeled with the minimization of the $l_{1,1,2}$ -norm defined as:

$$\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \|\mathcal{S}_{i,j,:}\|_2 \quad (4)$$

As discussed, illumination changes in an image sequence should be accounted for by both the low-rank part and the sparse part. The highly correlated part of illumination can be modeled with the low-rank tensor \mathcal{L} as background, but the independent changes in illumination are grouped as the foreground. To capture these uncorrelated illumination and shadow changes, and separate them from real changes, we recognize that they are not necessarily sparse. Fig. 4 shows two samples extracted illumination changes using our proposed method. Based on Fig. 4, it is easy to understand that illumination changes are on entire image and so, those uncorrelated changes are not completely sparse. These properties can be conveniently modeled with the k -support norm [1], which is a balanced norm and defined as:

$$\|\mathcal{C}_{:, :, p}\|_k^{sp} = \left(\sum_{m=1}^{k-r-1} (|c|_m^\downarrow)^2 + \frac{1}{r+1} \left(\sum_{m=k-r}^d |c|_m^\downarrow \right)^2 \right)^{\frac{1}{2}} \quad (5)$$

where $\mathcal{C}_{:, :, p}$ and $|c|_m^\downarrow$ denote the p^{th} frontal slice of \mathcal{C} and the m^{th} largest element in $|c|$, respectively. $r \in \{0; 1; \dots; k-1\}$ is an integer that is computed automatically by Algorithm 2 in the supplementary material. $c =$



Figure 4: Two sample images and their corresponding illumination changes captured by our proposed method

$vec(\mathcal{C}_{:,:,p})$ represents the vector constructed by concatenating the columns of $\mathcal{C}_{:,:,p}$ and $d = n_1 \times n_2$ is the dimension of the frontal slice. The k -support norm has two terms: l_2 -norm penalty for the large component, and l_1 -norm penalty for the small components. k is a parameter of the cardinality to achieve a balance between the l_2 -norm and the l_1 -norm ($k = n_1$ in our experiments). The k -support norm provides an appropriate trade-off between model sparsity and algorithmic stability [1], and yields more stable solutions than the l_1 -norm [14]. In this paper we show that the k -support norm can estimate the illumination changes in an image sequence accurately. Joining of this norm and (4) as two constraints in one optimization framework enables us to separate real changes from illumination changes.

To summarize, we propose the tensor low-rank and invariant sparse decomposition (TLISD) method, as follows.

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{S}, \mathcal{C}} \quad & \|\mathcal{L}\|_* + \lambda_1 \|\mathcal{S}\|_{1,1,2} + \lambda_2 (\|\mathcal{C}\|_k^{sp})^2 \\ \text{s.t.} \quad & \mathcal{D} = \mathcal{L} + \mathcal{S} + \mathcal{C} \end{aligned} \quad (6)$$

where $\|\mathcal{L}\|_*$ is the tensor nuclear norm, i.e. the average of the nuclear norm of all the frontal slices ($\|\mathcal{L}\|_* = \frac{1}{n_3} \sum_{p=1}^{n_3} \|\mathcal{L}_{:,:,p}\|_*$), and it approximates the rank of \mathcal{L} . \mathcal{S} and \mathcal{C} are detected moving objects and illumination changes, respectively.

3.3. Optimization Algorithm

In order to solve (6), we use the standard inexact augmented Lagrangian method (ALM) with the augmented Lagrangian function $\mathcal{H}(\mathcal{L}, \mathcal{S}, \mathcal{C}, \mathcal{Y}; \mu)$ whose main steps are described in this section for completeness.

$$\begin{aligned} \mathcal{H}(\mathcal{L}, \mathcal{S}, \mathcal{C}, \mathcal{Y}; \mu) = & \|\mathcal{L}\|_* + \lambda_1 \|\mathcal{S}\|_{1,1,2} + \lambda_2 (\|\mathcal{C}\|_k^{sp})^2 \\ & + \langle \mathcal{Y}, \mathcal{D} - \mathcal{L} - \mathcal{S} - \mathcal{C} \rangle + \frac{\mu}{2} \|\mathcal{D} - \mathcal{L} - \mathcal{S} - \mathcal{C}\|_F^2 \end{aligned} \quad (7)$$

where \mathcal{Y} is a Lagrangian multiplier, μ is a positive auto-adjusted scalar, and $\langle A, B \rangle = \text{trace}(A^T B)$. $\lambda_1 = 1/\sqrt{\max(n_1, n_2)n_3}$ and λ_2 is a positive scalar. Now we solve the problem through alternately updating \mathcal{L} , \mathcal{S} , and \mathcal{C} in each iteration to minimize $\mathcal{H}(\mathcal{L}, \mathcal{S}, \mathcal{C}, \mathcal{Y}; \mu)$ with other variables fixed until convergence as follows.

$$\mathcal{L}^{t+1} \leftarrow \min_{\mathcal{L}} \|\mathcal{L}\|_* + \frac{\mu}{2} \|\mathcal{L}^t - (\mathcal{D} - \mathcal{S}^t - \mathcal{C}^t + \frac{\mathcal{Y}^t}{\mu})\|_F^2 \quad (8)$$

$$\mathcal{S}^{t+1} \leftarrow \min_{\mathcal{S}} \lambda_1 \|\mathcal{S}\|_{1,1,2} + \frac{\mu}{2} \|\mathcal{S}^t - (\mathcal{D} - \mathcal{L}^{t+1} - \mathcal{C}^t + \frac{\mathcal{Y}^t}{\mu})\|_F^2 \quad (9)$$

$$\mathcal{C}^{t+1} \leftarrow \min_{\mathcal{C}} \lambda_2 (\|\mathcal{C}\|_k^{sp})^2 + \frac{\mu}{2} \|\mathcal{C}^t - (\mathcal{D} - \mathcal{L}^{t+1} - \mathcal{S}^{t+1} + \frac{\mathcal{Y}^t}{\mu})\|_F^2 \quad (10)$$

$$\mathcal{Y}^{t+1} = \mathcal{Y}^t + \mu(\mathcal{D} - \mathcal{L}^{t+1} - \mathcal{C}^{t+1} - \mathcal{S}^{t+1}) \quad (11)$$

where $\mu = \min(\rho\mu, \mu_{max})$. Both (8) and (9) have closed form solutions in [18] and [33] respectively, and (10) has an efficient solution in [14]. The error is computed as $\|\mathcal{D} - \mathcal{L}^t - \mathcal{S}^t - \mathcal{C}^t\|_F / \|\mathcal{D}\|_F$. The loop stops when the error falls below a threshold (10^{-5} in our experiments). Details of the solutions can be found in the supplementary material.

3.4. Time Complexity

In this work, we use ADMM to update \mathcal{L} and \mathcal{S} , which have closed form solutions. In these two steps the main cost per-iteration lies in the update of \mathcal{L}_{t+1} , which requires computing FFT and n_3 SVDs of $n_1 \times n_2$ matrices. Thus, time complexity of the first two steps per-iteration is $O(n_1 n_2 n_3 \log n_3 + n_{(1)} n_{(2)}^2 n_3)$, where $n_{(1)} = \max(n_1, n_2)$ and $n_{(2)} = \min(n_1, n_2)$ [18]. To update \mathcal{C}_{t+1} , we use an efficient solution based on binary search where the time complexity is reduced to $O((n_1 n_2 + k) \log(n_1 n_2))$ for each frontal slice per-iteration [14]. Therefore, the total time complexity of the optimization problem (6) is $O(n_1 n_2 n_3 \log n_3 + n_{(1)} n_{(2)}^2 n_3 + (n_1 n_2 + k) n_3 \log(n_1 n_2))$.

4. Experimental Results and Discussion

In this section, we provide an experimental evaluation of our proposed method, TLISD. We first evaluate the effect of each term in (6) and their λ coefficients. Then, we evaluate TLISD on benchmark frame-rate image sequences or those that are captured via time-lapse or motion-triggered photography. We also introduce a new dataset captured by industrial security cameras and wildlife monitoring systems during three years, and evaluate our method on this dataset.

4.1. Experiment Setup

Existing datasets¹: We evaluate our TLISD method on eleven selected sequences from the CDnet dataset [9], Wallflower dataset [27], I2R dataset [15], and ICD [23], which include illumination change and moving shadows.

Extended Illumination Change (EIC) dataset: Due to the lack of a comprehensive dataset with various illumination and shadow changes in a real environment, we have created a new benchmark dataset called EIC with around 80k images in 15 sequences, captured via available surveillance systems in wildlife and industrial applications. Particularly, ten sequences are captured via wildlife monitoring systems, and five sequences from industrial applications, with three railway sequences and two construction site sequences. Six sample sequences of this dataset are shown in Fig. 8. All sequences can be found in the supplementary material.

Evaluation metric: For quantitative evaluation, pixel-level F-measure = $2 \frac{\text{recall} \times \text{precision}}{\text{recall} + \text{precision}}$ is used. We also compare the different methods in execution time in seconds.

¹<https://sites.google.com/site/backgroundsubtraction/test-sequences>

4.2. Algorithm Evaluation: The effect of term \mathcal{C}

In the first set of experiments, we evaluate the effect of term \mathcal{C} in TLISD when we set different values for λ_1 , in comparison with TLISD without term \mathcal{C} , where (6) becomes

$$\min_{\mathcal{L}, \mathcal{S}} \|\mathcal{L}\|_* + \lambda_1 \|\mathcal{S}\|_{1,1,2} \quad s.t. \quad \mathcal{D} = \mathcal{L} + \mathcal{S} \quad (12)$$

Fig. 5(a) shows (12) can achieve around 70% accuracy with a well-tuned $\lambda_1 = 0.002$. Although the result shows the importance of multiple priors and the effect of group sparsity on them, the accuracy of (12) is still far below the accuracy of proposed TLISD by at least 10%, even with a well-tuned λ_1 . Fig. 5(a) also shows that adding term \mathcal{C} and k -support norm increases the robustness of our algorithm against tuning λ_1 . In fact, in (12) all illumination variations would be assigned to either of \mathcal{L} or \mathcal{S} . In this case, those variations should be assigned to the background (\mathcal{L}); however, they do not actually belong to background (e.g. moving shadows). As a result, the rank would be increased to absorb these changes into \mathcal{L} and naturally some parts of the moving objects \mathcal{S} would be also absorbed into the background. Fig. 5(b) supports the conclusion and shows the obtained rank through the iterations of the optimization. Between iterations 15 and 20, the rank of our method with-

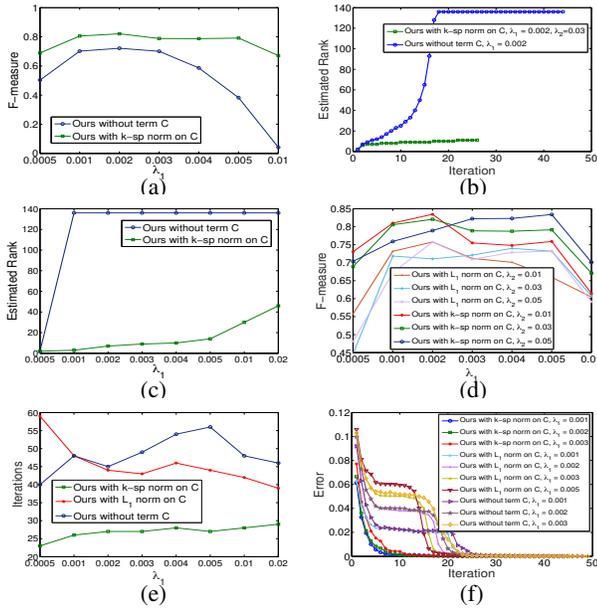


Figure 5: Self evaluation of TLISD. (a) Average F-measure with different values for λ_1 on all ICD sequences between TLISD and (12), (b) Estimated rank of TLISD and (12) through iterations on sequence “Wildlife3”, (c) Estimated rank of sequence “Wildlife3” with different values for λ_1 , (d) Average F-measure with different values for λ_1 and λ_2 on all ICD sequences between TLISD and (13), (e) Average number of iterations to converge TLISD, (12) and (13) on all ICD sequences, (f) Convergence curves of minimization error for TLISD, (12) and (13) on sequence “Wildlife3”.

out term \mathcal{C} significantly increases to absorb all variations into \mathcal{L} , and to complete the conclusion, Fig. 5(f) shows that around the same iterations, the residual error of the method without term \mathcal{C} is significantly reduced. This means, illumination variations and shadow changes must grouped into either of \mathcal{L} or \mathcal{S} , for (12) to converge. Estimated rank in Fig 5(c) shows the proof of this concept. Obviously, with a very small λ_1 , the estimated rank of \mathcal{L} for (12) is small and all illumination variations are easily lumped with moving objects in \mathcal{S} . This causes less accuracy and sometimes even cannot provide meaningful results. In contrast, TLISD can estimate a balanced rank and classify illumination variations into term \mathcal{C} with k -support-norm on it instead of increasing the rank to absorb them into \mathcal{L} .

To justify the use of k -support norm on \mathcal{C} in TLISD, we also compare the method with the other potential term on \mathcal{C} , which is l_1 -norm to absorb outliers, i.e., define (6) as

$$\min_{\mathcal{L}, \mathcal{S}, \mathcal{C}} \|\mathcal{L}\|_* + \lambda_1 \|\mathcal{S}\|_{1,1,2} + \lambda_2 \|\mathcal{C}\|_1 \quad s.t. \quad \mathcal{D} = \mathcal{L} + \mathcal{S} + \mathcal{C} \quad (13)$$

For this experiment, we evaluate our method with both l_1 and k -support norms on \mathcal{C} under different values of λ_1 and λ_2 . Fig. 5(d) illustrates the accuracy of our method with either of regularizers. Although l_1 -norm can increase the accuracy and robustness of the moving object detection in comparison with (12) that we showed in Fig. 5(a), the obtained accuracy is still less than TLISD. In addition, the number of iterations to converge, for both (12) and (13) is much more than that of in TLISD. Fig. 5(e) shows the average number of iterations for all three possible methods with different setup for λ_1 on all ICD sequences. For both TLISD and (13), $\lambda_2 = 0.03$, which produces robust results over different values of λ_1 (refer to Fig. 5(d)). As discussed in Section 3, illumination changes are not necessarily sparse and can be found throughout an image. Therefore, l_1 -norm is not a suitable regularizer to capture illumination changes. In such cases, the same issue as (12) happens when the optimizer increases the rank to minimize the residual error. Fig. 5(f) shows the error of all three methods through iterations. For (13), the same pattern as (12) is seen to decrease the error while the rank increases through optimization.

4.3. Evaluation on Benchmark Sequences

In this section we evaluate our method on the eleven benchmark sequences described in Section 4.1. Fig. 6 shows the qualitative results of TLISD on “Cubile” and “Backdoor”. The second and the third columns of Figs. 6(a) and (b) illustrate the first frontal slice of \mathcal{C} and \mathcal{S} , corresponding to illumination changes and moving objects, respectively. The high-quality of our detection result \mathcal{S} is clearly visible.

Figs. 7(a) and (b) show qualitative results of our method on two sample sequences of ICD, which has the most challenging conditions in terms of illumination changes. To appreciate the significant variations of illumination we show

two images from each sequence. The second and the third rows of each sub-figure show the first frontal slice of \mathcal{C} and \mathcal{S} , respectively. The results show the proposed method can accurately separate the changes caused by illumination and shadows from real changes.

We then compare TLISD quantitatively with two on-line and eight related RPCA batch methods. From online methods we select GMM [36] as a baseline method and GRATA [11] as an online method that uses the framework of low-rank and sparse decomposition. Also among batch methods, we select SSGoDec [34], PRMF [29], PCP [3], Markov BRMF [30], DECOLOR [35], LSD [17], ILISD [23], and TRPCA [18]. For all the competing methods we use their original settings through LRS Library [25], which resulted in the best performance. For quantitative evaluation of RPCA-related methods, a threshold criterion is required to get the binary foreground mask. Similarly, we adopt the same threshold strategy as in [25]. In TLISD, $\lambda_1 = 1/\sqrt{\max(n_1, n_2)n_3}$ (similar to TRPCA) and $\lambda_2 = 0.03$. Table 1 shows the performance of TLISD in comparison with the competing methods in terms of F-measure. For all the sequences TLISD ranked among the top two of all methods, and achieves the best average F-measure in comparison with all other methods. Although DECOLOR, LSD, and ILISD work relatively well, Only ILISD is comparable with our method due to the use of illumination regularization terms in ILISD. This evaluation shows the effectiveness of multiple prior maps and k -support norm as two regularization terms for separating moving objects from illumination changes, and boosting the overall performance of object detection.

4.4. Evaluation of TLISD on EIC Dataset

In this section, we evaluate TLISD on the introduced EIC dataset. Six sample sequences of ELC are shown in Fig. 8. To understand the significant variations of illumination and shadow, we show two images from each sequence in Figs. 8(a) and (b). Columns (c) and (d) show the first frontal slices of \mathcal{C} and \mathcal{S} obtained by TLISD for the images in column (b), in order to capture illumination changes and to detect moving objects. Table 2 show the capability of TLISD in comparison with the four best competitive methods (based on Table 1) in terms of F-measure, where TLISD can outperform the other methods by a clear performance margin. Fig. 9 also compares TLISD with ILISD (the second best method in Table. 2) qualitatively. This

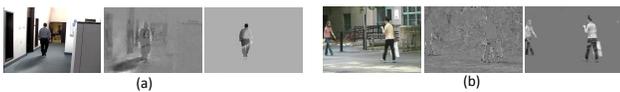


Figure 6: Columns from left to right show sample image, illumination changes, and detected moving objects for (a) cubicle and (b) backdoor sequences

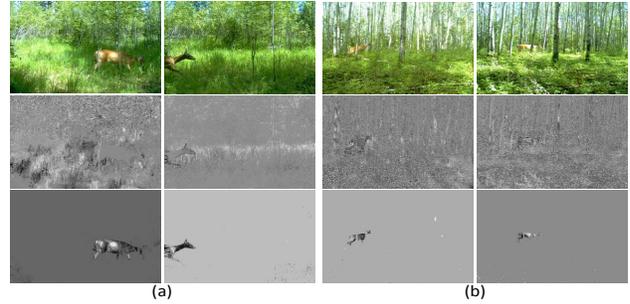


Figure 7: First row: two sample images from (a) Wildlife1, (b) Wildlife3 sequences. Second row: illumination changes obtained from the first frontal slice of \mathcal{C} . Third row: detected objects from the first frontal slice \mathcal{S} .

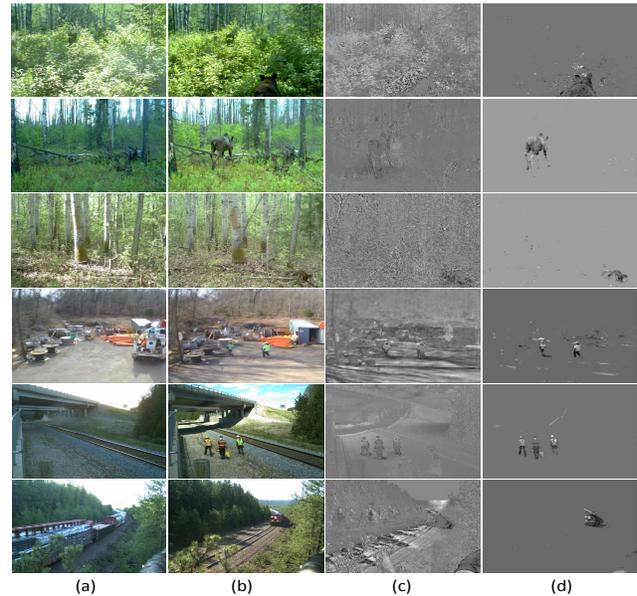


Figure 8: Columns (a) and (b): two sample images of each sequence, (c) and (d): illumination changes captured in \mathcal{C} , and detected objects of images in (b), respectively

qualitative comparison shows that one prior map only is not sufficient for removing the effect of illumination variations and shadow. As discussed in Section 3.1, due to the variation in the invariant direction for images in a sequence, in some conditions separating illumination changes and shadows from real changes is roughly impossible and selecting multiple prior maps is essential. More results on all sequences can be found in the supplementary material.

4.5. Execution Time of TLISD

Based on Tables 1 and 2, since ILISD is the only method with comparable results to ours, we examine our proposed method and ILISD in terms of computation time. Table. 3 compares the execution time of both methods on seven sequences. Regarding the computation time of the proposed method, our tensor-based method needs more time than [23]

Sequence	Backdoor	CopyMachine	Cubicle	PeopleInShade	LightSwitch	Lobby	Wildlife1	Wildlife2	Wildlife3	WinterStreet	MovingSunlight
GMM [36]	0.6512	0.5298	0.3410	0.3305	0.4946	0.3441	0.2374	0.2880	0.0635	0.1183	0.0717
GRATA [11]	0.6822	0.6490	0.4113	0.5288	0.5631	0.6727	0.3147	0.3814	0.2235	0.2276	0.1714
SSGoDec [34]	0.6611	0.5401	0.3035	0.2258	0.3804	0.0831	0.2912	0.2430	0.0951	0.1215	0.2824
PRMF [29]	0.7251	0.6834	0.3397	0.5163	0.2922	0.6256	0.2718	0.3991	0.07012	0.2108	0.2932
DECOLOR [35]	0.7656	0.7511	0.5503	0.5559	0.5782	<u>0.7983</u>	0.3401	0.3634	0.1202	0.4490	0.3699
PCP [3]	0.7594	0.6798	0.4978	0.6583	0.8375	0.6240	0.5855	0.6542	0.3003	0.1938	0.3445
BRMF [30]	0.6291	0.3293	0.3746	0.3313	0.2872	0.3161	0.2743	0.2812	0.0735	0.0872	0.2408
LSD [17]	0.7603	0.8174	0.4233	0.6168	0.6640	0.7313	0.6471	0.3790	0.0871	0.1604	0.3593
ILISD [23]	<u>0.8150</u>	<u>0.8179</u>	<u>0.6887</u>	0.8010	0.7128	0.7849	<u>0.8033</u>	<u>0.7277</u>	<u>0.7398</u>	<u>0.6931</u>	<u>0.6475</u>
TRPCA [18]	0.7022	0.6805	0.5329	0.5683	0.6924	0.6176	0.4382	0.3926	0.2854	0.2721	0.3018
TLISD	0.8276	0.8445	0.7350	<u>0.7961</u>	<u>0.7429</u>	0.8012	0.8862	0.8065	0.8010	0.7092	0.7122

Table 1: Comparison of F-measure score between our proposed method and other compared methods on benchmark real-time sequences (best F-measure: bold, second best F-measure: underline)



Figure 9: Comparison of qualitative results between TLISD and ILISD on four sequences of EIC dataset. Top to bottom: Sample Image, Ground Truth, ILISD, and TLISD

Sequence	Wildlife4	Wildlife5	Wildlife6	Railway1	Railway2	Industrial area1
PCP [3]	0.4150	0.4016	0.3092	0.3634	0.4086	0.2869
DECOLOR [35]	0.3475	0.2010	0.2604	0.2853	0.3021	0.3242
ILISD [23]	0.6020	0.6104	0.6170	0.5983	0.5414	0.5626
TRPCA [18]	0.2934	0.3082	0.2855	0.3447	0.2805	0.2914
TLISD	0.7508	0.8049	0.7522	0.7241	0.7116	0.7035

Table 2: Comparison of F-measure score between our proposed method and other compared methods on EIC dataset

for each iteration, which is normal due to use of the tensor structure. However, the number of iterations in our method is less than that of [23]. Fig. 10 shows the number of iterations to converge for both ILISD and TLISD methods. ILISD [23] has two independent optimization formulae: one for providing a prior map and the other for separating moving objects from illumination changes, and they have independent numbers of iterations to converge. After convergence, the optimized values are interchangeably used in an outer loop, and hence the total number of iterations is much more than that of our method which involves one optimization formula. As discussed in Section 3.4, the dominant time in our method is SVD decomposition for frontal slices, which are independent from each other, and so can be solved in parallel on a GPU to speed up the computation. Therefore, the total time of our method is at least comparable with ILISD and can be even faster due to the fewer number of iterations.

5. Conclusions

In this paper, we have proposed a novel method based on tensor low-rank and invariant sparse decomposition to detect moving objects under discontinuous changes in illumination, which frequently happen in video surveillance applications. In our proposed method, first we compute a set of illumination invariant representations for each image as prior maps, which provide us with cues for extracting moving objects. Then we model illumination changes in an image sequence using a k -support norm and derive a new formulation to effectively capture illumination changes and separate them from detected foregrounds. Currently, many surveillance systems, especially security and wildlife monitoring cameras, use motion triggered sensors and capture image sequences with significant illumination changes. Our proposed method can solve the problem with a performance that is superior to the state-of-the-art solutions. Our method is also able to extract natural outdoor illumination as labeled data for learning-based methods, which can be an effective alternative to optimization based methods such as ours, but with a sequential formulation, to detect illumination changes and moving objects from image sequences.

Acknowledgment

This research is supported in part by NSERC through its Discovery Grant and Strategic Network Grant (NCRN) programs.

Sequence	Backdoor	Lobby	Cubicle	Wildlife1	Wildlife2	Wildlife3	MovingSunlight
ILISD	0.49	0.53	0.74	1.24	1.33	1.18	2.2
TLISD	0.98	2.38	1.79	2.52	4.26	4.08	5.16

Table 3: Comparison of execution time (in sec.) per image

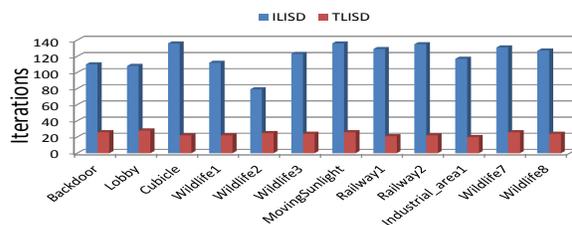


Figure 10: Number of iterations to converge ILISD and TLISD methods on twelve sequences

References

- [1] Andreas Argyriou, Rina Foygel, and Nathan Srebro. Sparse prediction with the k-support norm. In *Advances in Neural Information Processing Systems*, pages 1457–1465, 2012.
- [2] Thierry Bouwmans and El Hadi Zahzah. Robust pca via principal component pursuit: A review for a comparative evaluation in video surveillance. *Computer Vision and Image Understanding*, 122:22–34, 2014.
- [3] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.
- [4] Wenfei Cao, Yao Wang, Jian Sun, Deyu Meng, Can Yang, Andrzej Cichocki, and Zongben Xu. Total variation regularized tensor rpca for background subtraction from compressive measurements. *IEEE Transactions on Image Processing*, 25(9):4075–4090, 2016.
- [5] Bo-Hao Chen and Shih-Chia Huang. An advanced moving object detection algorithm for automatic traffic monitoring in real-world limited bandwidth networks. *IEEE Transactions on Multimedia*, 16(3):837–847, 2014.
- [6] Xinyi Cui, Junzhou Huang, Shaoting Zhang, and Dimitris N Metaxas. Background subtraction using low rank and group sparsity constraints. In *European Conference on Computer Vision (ECCV)*, pages 612–625. Springer, 2012.
- [7] Graham D Finlayson, Mark S Drew, and Cheng Lu. Entropy minimization for shadow removal. *International Journal of Computer Vision (IJCV)*, 85(1):35–57, 2009.
- [8] Donald Goldfarb and Zhiwei Qin. Robust low-rank tensor recovery: Models and algorithms. *SIAM Journal on Matrix Analysis and Applications*, 35(1):225–253, 2014.
- [9] Nil Goyette, Pierre-Marc Jodoin, Fatih Porikli, Janusz Konrad, and Prakash Ishwar. Changedetection.net: A new change detection benchmark dataset. In *Computer Vision and Pattern Recognition Workshops (CVPRW)*, *IEEE Computer Society Conference on*, pages 1–8. IEEE, 2012.
- [10] Charles Guyon, Thierry Bouwmans, and El-Hadi Zahzah. Foreground detection based on low-rank and block-sparse matrix decomposition. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 1225–1228. IEEE, 2012.
- [11] Jun He, Laura Balzano, and Arthur Szlam. Incremental gradient on the grassmannian for online foreground and background separation in subsampled video. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1568–1575. IEEE, 2012.
- [12] Wenrui Hu, Yehui Yang, Wensheng Zhang, and Yuan Xie. Moving object detection using tensor-based low-rank and saliently fused-sparse decomposition. *IEEE Transactions on Image Processing (TIP)*, 26(2):724–737, 2017.
- [13] Sajid Javed, Thierry Bouwmans, and Soon Ki Jung. Sbm1td: stationary background model initialization based on low-rank tensor decomposition. In *Proceedings of the Symposium on Applied Computing*, pages 195–200. ACM, 2017.
- [14] Hanjiang Lai, Yan Pan, Canyi Lu, Yong Tang, and Shuicheng Yan. Efficient k-support matrix pursuit. In *European Conference on Computer Vision (ECCV)*, pages 617–631. Springer, 2014.
- [15] Liyuan Li, Weimin Huang, Irene Yu-Hua Gu, and Qi Tian. Statistical modeling of complex backgrounds for foreground object detection. *IEEE Transactions on Image Processing (TIP)*, 13(11):1459–1472, 2004.
- [16] Ping Li, Jiashi Feng, Xiaojie Jin, Luming Zhang, Xianghua Xu, and Shuicheng Yan. Online robust low-rank tensor modeling for streaming data analysis. *IEEE transactions on neural networks and learning systems*, (99):1–15, 2018.
- [17] Xin Liu, Guoying Zhao, Jiawen Yao, and Chun Qi. Background subtraction based on low-rank and structured sparse decomposition. *IEEE Transactions on Image Processing (TIP)*, 24(8):2502–2514, 2015.
- [18] Canyi Lu, Jiashi Feng, Yudong Chen, Wei Liu, Zhouchen Lin, and Shuicheng Yan. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5249–5257, 2016.
- [19] Julien Mairal, Rodolphe Jenatton, Francis R Bach, and Guillaume R Obozinski. Network flow algorithms for structured sparsity. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1558–1566, 2010.
- [20] Nuria M Oliver, Barbara Rosario, and Alex P Pentland. A bayesian computer vision system for modeling human interactions. *IEEE transactions on pattern analysis and machine intelligence (PAMI)*, 22(8):831–843, 2000.
- [21] Moein Shakeri and Hong Zhang. Corola: a sequential solution to moving object detection using low-rank approximation. *Computer Vision and Image Understanding*, 146:27–39, 2016.
- [22] Moein Shakeri and Hong Zhang. Illumination invariant representation of natural images for visual place recognition. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 466–472. IEEE, 2016.
- [23] Moein Shakeri and Hong Zhang. Moving object detection in time-lapse or motion trigger image sequences using low-rank and invariant sparse decomposition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (ICCV)*, pages 5123–5131, 2017.
- [24] Andrews Sobral, Christopher Baker, Thierry Bouwmans, and El-hadi Zahzah. Incremental and multi-feature tensor subspace learning applied for background modeling and subtraction. In *International Conference Image Analysis and Recognition*, pages 94–103. Springer, 2014.
- [25] Andrews Sobral, Thierry Bouwmans, and El-hadi Zahzah. Lrslibrary: Low-rank and sparse tools for background modeling and subtraction in videos. In *Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*. CRC Press.
- [26] YingLi Tian, Andrew Senior, and Max Lu. Robust and efficient foreground analysis in complex surveillance videos. *Machine vision and applications*, 23(5):967–983, 2012.
- [27] Kentaro Toyama, John Krumm, Barry Brumitt, and Brian Meyers. Wallflower: Principles and practice of background maintenance. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on (ICCV)*, volume 1, pages 255–261. IEEE, 1999.

- [28] Alessandro Vinciarelli, Maja Pantic, and Hervé Bourlard. Social signal processing: Survey of an emerging domain. *Image and vision computing*, 27(12):1743–1759, 2009.
- [29] Naiyan Wang, Tiansheng Yao, Jingdong Wang, and Dit Yeung. A probabilistic approach to robust matrix factorization. In *European Conference on Computer Vision (ECCV)*, pages 126–139. Springer, 2012.
- [30] Naiyan Wang and Dit Yeung. Bayesian robust matrix factorization for image and video processing. In *International Conference on Computer Vision (ICCV)*, pages 1785–1792, 2013.
- [31] Bo Xin, Yuan Tian, Yizhou Wang, and Wen Gao. Background subtraction via generalized fused lasso foreground modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4676–4684, 2015.
- [32] Jia Xu, Vamsi K Ithapu, Lopamudra Mukherjee, James M Rehg, and Vikas Singh. Gosus: Grassmannian online subspace updates with structured-sparsity. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 3376–3383, 2013.
- [33] Zemin Zhang, Gregory Ely, Shuchin Aeron, Ning Hao, and Misha Kilmer. Novel methods for multilinear data completion and de-noising based on tensor-svd. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3842–3849, 2014.
- [34] Tianyi Zhou and Dacheng Tao. Godec: Randomized low-rank & sparse matrix decomposition in noisy case. In *International conference on machine learning (ICML)*. Omnipress, 2011.
- [35] Xiaowei Zhou, Can Yang, and Weichuan Yu. Moving object detection by detecting contiguous outliers in the low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 35(3):597–610, 2013.
- [36] Zoran Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *null*, pages 28–31. IEEE, 2004.