This ICCV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the version available on IEEE Xplore.

Moving Object Detection in Time-Lapse or Motion Trigger Image Sequences Using Low-rank and Invariant Sparse Decomposition

Moein Shakeri, Hong Zhang
Department of Computing Science, University of Alberta
Edmonton, Alberta, Canada, T6G2E8
{shakeri,hzhang}@ualberta.ca

Abstract

Low-rank and sparse representation based methods have attracted wide attention in background subtraction and moving object detection, where moving objects in the scene are modeled as pixel-wise sparse outliers. Since in real scenarios moving objects are also structurally sparse, recently researchers have attempted to extract moving objects using structured sparse outliers. Although existing methods with structured sparsity-inducing norms produce promising results, they are still vulnerable to various illumination changes that frequently occur in real environments, specifically for time-lapse image sequences where assumptions about sparsity between images such as group sparsity are not valid. In this paper, we first introduce a prior map obtained by illumination invariant representation of images. Next, we propose a low-rank and invariant sparse decomposition using the prior map to detect moving objects under significant illumination changes. Experiments on challenging benchmark datasets demonstrate the superior performance of our proposed method under complex illumination changes.

1. Introduction

Moving object segmentation from an image sequence or a video stream is a fundamental problem in various applications of computer vision such as visual surveillance [20], traffic monitoring [5], object-based video encoding and social signal processing [22] where the accuracy of segmentation can significantly affect the overall performance of the application. However, current solutions are vulnerable to various illumination changes frequently occurring in real environments and are often not able to distinguish between changes caused by illumination and those caused by moving objects in the scene. Currently, many surveillance systems, specifically those that use security cameras and wildlife monitoring cameras, capture a scene using a motion trigger sensor or timer-lapse photography in order to detect moving objects of interest over time. Since captured images by these cameras are in different time of a day with different illumination and weather conditions, their processing is challenging. Fig. 1 shows an example of this kind of images and illustrates the problem of object detection under significant illumination changes. The first row in Fig.1 shows selected images captured by a motion-trigger camera for wildlife monitoring and the second row shows the results of our method described in this paper to detect moving objects from the images.

Figure 1. first row: selected images from an image sequence captured by a motion triggered camera for wildlife monitoring. second row: detected foreground objects of our method.

Recent years have seen the development of a new group of methods, based on low-rank and sparse decomposition, under one major assumption that images in a sequence are correlated. Methods in this group follow the basic idea from [17] where the principal component analysis (PCA) for background modeling was proposed. Extending this idea, current methods exploit the fact that the background model in an image sequence can be defined as a low-rank matrix by those pixels that are temporally correlated [4]. Many methods have been proposed based on this idea for background subtraction and foreground detection [10, 7, 18]. But, moving object detection in a time lapse video is different from that in a regular video due to the discontinuous object motion and significant lighting change. While existing solutions can handle discontinuous object motion, they are not able to distinguish between moving objects and background changes due to illumination.

In this paper, we offer a solution to the problem within the low-rank approximation (LRA) framework that specifically addresses this challenge. We formulate the problem in a unified framework named Low-rank and Invariant Sparse Decomposition (LISD). Since changes due to illumination and shadow are easily lumped with moving objects and detected as the sparse outliers in the low-rank formulation, first we compute a prior map using illumination invariant representation of images to provide information about the effect of illumination. Then we define two penalty terms based on the prior map to decompose an image into three parts: the background model, illumination changes, and moving objects. The key to our solution is incorporating the prior information in the LRA framework. The corresponding optimization problem can be readily solved with existing tools. We also propose an iterative version of LISD (ILISD) to improve the performance of LISD by updating the prior map. Since we use two representations (grayscale and illumination invariant representations), the prior map in ILISD is updated iteratively from the results of each representation that is used as a constraint in another representation. We also propose a new dataset consists of time-lapse videos which are challenging to exiting methods and we use them to demonstrate the superiority of our solution.

The remainder of the paper is organized as follows. Related works on moving object detection using low-rank and sparse decomposition are summarized in Section 2. Section 3 explains the details of our low-rank and invariant sparse decomposition method for foreground detection in time-lapse videos under significant illumination changes. Experimental results and discussion are presented in Section 4, and concluding remarks in Section 5.

2. Moving object detection using low-rank and sparse decomposition

Recently, one approach to moving object detection attempts to decompose a matrix D of the observed image sequence into a low-rank matrix L and a sparse matrix S so as to recover background and foreground [1]. The problem can be solved by the well known robust principal component analysis (RPCA), which has been widely studied. Based on different constraints on S, RPCA methods can be categorized into different groups. Candes *et al.* [4] used l_1-norm to constrain the sparse matrix by the following convex optimization.

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

where $\|L\|_*$ denotes the nuclear norm of matrix L - i.e., the sum of its singular values - and $\|S\|_1$ is the l_1-norm of S. Following this approach, Zhou *et al.* [27] proposed approximated RPCA methods GoDec and semi-soft GoDec (SS-GoDec) to accelerate the decomposition. Wang *et al.* [23] proposed a propabilistic matrix factorization (PRMF) using Laplace error and Gaussian prior, which correspond to an l_1 loss and l_2 regularizer, respectively.

To improve the performance of moving object detection, some other constraints have been recently imposed on sparse matrix S using prior knowledge of spatial continuity of objects [10, 7]. Guyon *et al.* [10] proposed the low-rank and block sparse matrix decomposition (RPCA-LBD) using $l_{2,1}$-norm as a spatial continuity to enforce the block-sparsity of the foreground. Although the method is more robust than conventional RPCA in the presence of illumination changes, the block-sparsity property is unable to model sparse outliers or filter out significant illumination changes and moving shadows. Besides, in the case of time-lapse video or low frame rate image sequences, where consecutive frames are captured with a large time-interval the position of an object in each frame is discontinuous from other frames and $l_{2,1}$-norm cannot handle the situation.

Another group of methods used the connectivity constraint on moving objects [26, 24, 28, 18, 25, 15]. Xu *et al.* [26] proposed an online subspace update method GOSUS that defines an objective function with a superpixel method to achieve sparsity of the groups. Wang *et al.* [24] proposed a full Bayesian robust matrix factorization (BRMF). They further extended it by assuming that the outliers form clusters with close within-group spatial proximity which correspond to moving objects. This is achieved by placing a first-order Markov random field (MRF) [12], and the method is referred to as Markov BRMF or MBRMF. Zhou *et al.* [28] proposed DECOLOR by assuming that the moving objects are also small. Under this constraint, a sequential implementation of DECOLOR to moving object detection (COROLA) is proposed [18]. Due to use of GMM, COROLA can deal with background changes. Although COROLA improves the accuracy of moving object detection compared to DECOLOR, it is still not able to handle severe illumination changes and moving shadow, especially in a low frame-rate image sequence. Following the connectivity constraint, Liu *et al.* [15] proposed a method using a structured sparsity norm [16] based on 3×3 overlapping-patch groups. Since the foreground is usually spatially contiguous in each image, computing the maximum values of each group promotes the structural distribution of sparse outliers during the minimization. They also used a motion saliency map to distinguish the foreground object from background motion. Using this saliency map, the method is robust in the case of background motion and sudden illumination change in the image sequence. However [15] cannot handle severe illumination changes or moving shadows in time-lapse videos where the foreground objects are completely stochastic as are shadow and illumination changes.

In this paper, we introduce a prior map for outliers and we use this prior information in a new formulation for moving object detection under the framework of low-rank representation and invariant sparse outliers. Due to use of this prior map in our formulation as a penalty term, the proposed method significantly improves foreground detection in the case of moving shadows and severe illumination changes.

3. Low-rank and invariant sparse decomposition

Our proposed formulation seeks to decompose a data matrix D into a low-rank background matrix L, sparse illumination change matrix C, and sparse foreground matrix S as follows.

$$D = L + C + S \tag{2}$$

In (2), C and S are considered as outliers. Since both of them are stochastic in time-lapse video or low frame rate image sequences, separating them is an ill-posed problem. We address this challenge by using an illumination invariant representation of an image, which serves as a prior for outliers in our formulation. This prior enables us to have a pattern for estimating C and S through the optimization as will be detailed in Section 3.1. Then in Section 3.2 we introduce our formulation to detect moving objects under significant illumination changes and in Section 3.3 we describe a solution to the formulation.

3.1. Initialization of the prior map

In this section we focus on obtaining the prior information, which enables us to distinguish between moving objects and illumination changes in our proposed formulation. In the case of time-lapse images, shadows and illumination changes are unstructured phenomena and most of the time they are mistakenly considered as moving objects. Illumination invariant and shadow free images have been well studied and many methods have been proposed. One of the most popular methods for this task is proposed by Finlayson *et al.* [8]. This method assumes the camera sensor sensitivities are Dirac delta functions and illumination can be modeled by Planck's law. For removing the effect of illumination, [8] computes the two-vector log-chromaticity χ' using red, green and blue channels. Finlayson *et al.* [8] showed that by 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, we obtain the invariant representation $I = \chi' e^{\perp}$. The best direction for e can be found by minimizing Shannon's entropy [8].

Although this method works with the mentioned assumptions for some real images, in case of significant illumination changes, specially if the assumptions do not hold, χ' necessarily does not move along a straight line. This issue causes two major problems in the invariant representation I. First, χ' vectors of the same material under different illumination are not projected to the same location in the orthogonal vector and therefore, the method cannot remove

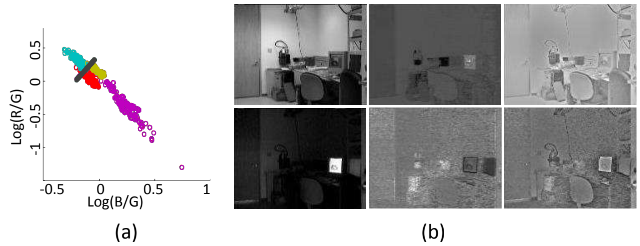

Figure 2. (a):log-chromaticity vectors of pixels from one material in different illumination condition, (b) Columns from left to right: two images with extreme illumination changes, their corresponding invariant image I, and their corresponding final invariant representation I_{inv}

the effect of illumination accurately. Secondly, in the process of projection onto the orthogonal direction of illumination variation, some pixels with the same log-chromaticity but from different objects are projected to the same location in the orthogonal vector, and the invariant representation removes much meaningful information about the image, especially around edges.

Although the first issue would be problematic for an individual image, if we have an image sequence, corresponding pixels of the images in invariant representation are correlated to each other and therefore those pixels can be captured in a low-rank matrix. Fig. 2(a) shows the details of this concept. Four different locations but from one material are selected. Sample points with the same color show log-chromaticity of corresponding pixels from the selected locations in a sequence of images with different illumination. Assuming that the camera is fixed, the invariant direction between images is roughly similar. Black circles in Fig. 2(a) show the projected pixels of the same material from all images to the average invariant directions of all images, where corresponding pixels of all images with different illumination are projected to one coordinate or are close to each other in invariant representation. In other words, the corresponding pixels of all images under different illumination are correlated.

To alleviate the effect of the second issue for preserving the structural information of images, we extract invariant features $\tilde{I}$ from each image using Wiener filter, which has been used successfully for face recognition in [6]. Wiener filter decomposes a signal into its components from two stationary processes with different autocorrelation functions, where the Fourier transform of the autocorrelation function is the power spectrum density in the frequency domain. [6] showed that this method retains features at every frequency. We add $\tilde{I}$ to the invariant image I. This final invariant representation is called I_{inv}. Fig. 2(b) shows the effect of adding the invariant features $\tilde{I}$ to the invariant representation I. First column shows two images from one scene with the light switch on/off and the second column shows the corresponding invariant image I. Last column illustrates I_{inv}, the results of adding invariant features $\tilde{I}$ to the invariant image I. To combine $\tilde{I}$ and I, we use simple weighted averaging similar to [19].

Now, to construct the prior map formally, let $D \in R^{m\times n}$ be an observed matrix (an image sequence in our problem), where each column of matrix D is a vectorized image from the sequence with m pixels, and n is the total number of images in the sequence. Then the following function convert all D_i images $i = 1, 2, .., n$ to the invariant representation D_{inv}.

$$D_{inv} = \Omega(D) \tag{3}$$

where $D_{inv} \in R^{m\times n}$ be a matrix of all vectorized invariant representations I_{inv_t}, $t = 1, 2, ...n$. We can decompose matrix D_{inv} into low-rank matrix L_{inv} and sparse matrix S_{inv} using optimization, so that all illumination variations are absorbed into the low-rank matrix L_{inv}.

$$\min_{L_{inv},S_{inv}} \|L_{inv}\|_* + \lambda_{inv}\|S_{inv}\|_1 \;\; s.t.\, D_{inv} = L_{inv} + S_{inv} \tag{4}$$

To solve (4) we use inexact augmented Lagrangian multiplier (ALM) [13]. Optimization problem (4) can account for most of the illumination and shadow changes with the low-rank part and for the moving objects with the sparse part S_{inv}. Finally, we can use S_{inv} to build the prior map Φ as follows.

$$\Phi = \frac{1}{1 + e^{-\alpha(|S_{inv}| - \sigma)}} \tag{5}$$

where σ shows the standard deviation of corresponding pixels in D_{inv}, and α is a constant. We use the prior map Φ, to define two penalty terms in the LRA framework to extract the invariant sparse outliers as moving object, as will be explained in the next section.

3.2. Low-rank and invariant sparse decomposition

To detect moving objects in time-lapse videos under severe illumination changes, standard low-rank method is insufficient because we need to separate illumination changes and moving shadows from real changes and both of them are sparse outliers. To do so, we define a constraint based on the prior knowledge from illumination invariant representation introduced in the previous section. In particular, real changes should be included in the subspace that is orthogonal to the illumination change subspace. Since outliers are completely independent in different frames of a low framerate image sequence, real changes in the ith frame should satisfy the following properties.

$$(\Phi_i^{\perp})^T|S_i| = 0, \quad \Phi_i^T|C_i| = 0 \tag{6}$$

where $\Phi_i^{\perp} = [1]_{m\times 1} - \Phi_i$, is the complement of Φ_i. Sparse S and C are the detected objects and illumination changes in the grayscale domain.

To formalize the prior knowledge from illumination invariant representation on the outliers, we propose Low-rank and Invariant Sparse Decomposition (LISD) method, as follows.

$$\begin{aligned}\min_{L,S,C} \; & \|L\|_* + \lambda(\|S\|_1 + \|C\|_1) + \gamma\Psi(S, C, \Phi) \\ & s.t. \;\; D = L + S + C\end{aligned} \tag{7}$$

where $\|L\|_*$ is the nuclear norm, i.e. the sum of the singular values, and it approximates the rank of L. S and C are detected foreground and illumination changes, respectively. $\Psi(S, C, \Phi) = \sum_i(\Phi_i^{\perp})^T|S_i| + \sum_i \Phi_i^T|C_i|$, is the geometric constraint function. To make the problem more tractable, the geometric constraint Ψ can be relaxed to the penalty terms $\Sigma_i\|G_iC_i\|_F^2$, and $\Sigma_i\|G_i^{\perp}S_i\|_F^2$ so that (7) becomes

$$\begin{aligned}\min_{L,S,C} \; & \|L\|_* + \lambda(\|S\|_1 + \|C\|_1) + \gamma\Sigma_i(\|G_iC_i\|_F^2 + \|G_i^{\perp}S_i\|_F^2) \\ & s.t. \;\; D = L + S + C\end{aligned} \tag{8}$$

where λ and γ are positive parameters and $G_i = diag[\sqrt{\Phi_{1i}}; \sqrt{\Phi_{2i}}; ...; \sqrt{\Phi_{mi}}]$.

3.3. Optimization algorithm

In order to solve (8), we use inexact ALM method [13], and start by computing the augmented Lagrangian function $\mathcal{L}(L, S, C, Y; \mu)$, given by

$$\begin{aligned}\mathcal{L}(L, S, C, Y; \mu) = & \|L\|_* + \lambda(\|S\|_1 + \|C\|_1) + \gamma\Sigma_i(\|G_iC_i\|_F^2 \\ & + \|G_i^{\perp}S_i\|_F^2) + < Y, D - L - S - C > \\ & + \frac{\mu}{2}\|D - L - S - C\|_F^2 \\ = & \|L\|_* + \lambda(\|S\|_1 + \|C\|_1) - \frac{1}{2\mu}\|Y\|_F^2 \\ & + h(L, S, C, Y, \mu)\end{aligned} \tag{9}$$

where $< A, B >= trace(A^TB)$, μ is a positive scalar, Y is a Lagrangian multiplier matrix, and $h(L, S, C, Y, \mu) = \Sigma_i(\frac{\mu}{2}\|D_i - L_i - S_i - C_i + \frac{Y_i}{\mu}\|_F^2 + \gamma\|G_iC_i\|_F^2 + \gamma\|G_i^{\perp}S_i\|_F^2)$ is a quadratic function. We optimize (9) by updating each of the variables L, S, C, and Y in turn, iteratively until convergence.

Updating L^{k+1}: From (9), the augmented Lagrangian reduces to the following form:

$$L^{k+1} = \arg\min_L \|L\|_* + \frac{\mu}{2}\|L^k - (D - S^k - C^k - \frac{Y^k}{\mu})\|_F^2 \tag{10}$$

The subproblem (10) has the closed-form solution by applying the singular value thresholding algorithm [3], with the soft-thresholding shrinkage operator $\mathcal{S}_\epsilon(x)$, which is defined as $\mathcal{S}_\epsilon(x) = max(0, x - \epsilon)$, where $x \geq 0$ and $\epsilon \geq 0$.

Updating S^{k+1}: From (9), the augmented Lagrangian reduces to

$$\min_S \lambda\|S\|_1 + h(L, S, C, \mu) \tag{11}$$

Since $h(L, S, C, \mu)$ is a quadratic function, it is convenient to use the linearization technique of the LADMAP method [14] to update S^{k+1} by replacing the quadratic term h with its first order approximation, computed at iteration k and add a proximal term giving the following update.

$$S^{k+1} = \arg\min_S \lambda\|S\|_1 + \Sigma_i(\frac{\eta\mu}{2}\|S_i - S_i^k + [-\mu(D_i - L_i^{k+1} - S_i^k - C_i^k + \frac{Y_i^k}{\mu}) + 2\gamma(G_i^{\perp})^T G_i^{\perp} S_i^k]/(\eta\mu)\|_F^2) \quad (12)$$

Updating C^{k+1}: From (9), the augmented Lagrangian reduces to

$$\min_S \lambda\|C\|_1 + h(L, S, C, \mu) \quad (13)$$

Similar to (11) we use the LADMAP method to update C^{k+1} by giving the following update

$$C^{k+1} = \arg\min_C \lambda\|C\|_1 + \Sigma_i(\frac{\eta\mu}{2}\|C_i - C_i^k + [-\mu(D_i - L_i^{k+1} - S_i^{k+1} - C_i^k + \frac{Y^k}{\mu}) + 2\gamma G_i^T G_i C_i^k]/(\eta\mu)\|_F^2) \quad (14)$$

The error is computed as $\|D - L^k - S^k - C^k\|_F / \|D\|_F$. The loop stops when the error reaches the value lower than 10^{-5}.

3.4. Updating the prior map

In our proposed LISD method, we first compute a prior map from illumination invariant representation of images and then use the map to separate foreground from background and illumination changes in the grayscale representation of images. Although LISD provides satisfactory results in our experiments, still we can improve the performance by updating the prior map (5) iteratively. We refer to this iterative version as Iterative LISD (ILISD). In ILISD, the first step is exactly similar to LISD where we compute the prior map in one representation and use it in another representation. Ideally, moving object in both representations should build similar prior maps. Using this assumption, we compute the second prior map Φ_{inv} from the result of LISD to use in illumination invariant representation. To do it, we rewrite (4) similar to (7) as follows.

$$\begin{aligned} \min_{L_{inv}, S_{inv}, C_{inv}} \quad & \|L_{inv}\|_* + \lambda_{inv}(\|S_{inv}\|_1 + \|C_{inv}\|_1) \\ & + \gamma\Psi(S_{inv}, C_{inv}, \Phi_{inv}) \\ s.t. \quad & D_{inv} = L_{inv} + S_{inv} + C_{inv} \end{aligned} \quad (15)$$

where Ψ is defined the same as (7) but for illumination invariant representation. Φ_{inv} is computed similar to (5) using the obtained S from LISD method. To solve (15) we use the same idea that we have described on grayscale images from (9) to (14). Generally speaking in ILISD the obtained map from each representation is used into another representation in the next iteration until convergence. The convergence criterion is $\|S^{j+1} - S^j\|_F / \|S^j\|_F < 10^{-5}$. All details about ILISD are described in Algorithm 1.

Algorithm 1 Iterative Low-rank and Invariant Sparse Decomposition via Inexact ALM (ILISD)

Input: Observation matrix D, $\Phi_{inv} = [1]_{m \times n}$, Parameters λ,γ,η,
1: computing invariant representation D_{inv} according to Section 3.1
2: **while** not converged **do**
3: $[L_{inv}^{j+1}, S_{inv}^{j+1}, C_{inv}^{j+1}] = LISD(D_{inv}, \Phi_{inv}, \lambda, \gamma, \mu_{inv})$
4: compute Φ according to (5) using S_{inv}^{j+1}
5: $[L^{j+1}, S^{j+1}, C^{j+1}] = LISD(D, \Phi, \lambda, \gamma, \mu)$
6: compute Φ_{inv} according to (5) using S^{j+1}; $\quad j = j + 1$
7: **end while**
Output L^j, S^j, C^j

function $[L^k, S^k, C^k] = LISD(D, \Phi, \lambda, \gamma, \mu)$
8: **while** not converged **do**
9: $(U, \Sigma, V) = svd(D - S^k - C^k + \mu^{-1} Y^k)$ //lines 9-10 solve (10)
10: $L^{k+1} = U\mathcal{S}_{(1/\mu)}(\Sigma)V^T$
11: Compute for all coulmns i //lines 11-13 solve (12)
12: $tempS_i = S_i^k + [\mu(D_i - L_i^{k+1} - S_i^k - C_i^k + \mu^{-1} Y_i^k) - 2\gamma(G_i^{\perp})^T G_i^{\perp} S_i^k]/(\eta\mu)]$
13: $S^{k+1} = \mathcal{S}_{\lambda/(\eta\mu)}(tempS)$, $tempS = [tempS_1, ..., tempS_n]$
14: Compute for all coulmns i //lines 14-16 solve (14)
15: $tempC_i = C_i^k + [\mu(D_i - L_i^{k+1} - S_i^{k+1} - C_i^k + \mu^{-1} Y_i^k) - 2\gamma G_i^T G_i C_i^k]/(\eta\mu)]$
16: $C^{k+1} = \mathcal{S}_{\lambda/(\eta\mu)}(tempC)$, $tempC = [tempC_1, ..., tempC_n]$
17: $Y = Y + \mu(D - L^{k+1} - S^{k+1} - C^{k+1})$
18: $\mu = \rho\mu$; $k = k + 1$
19: **end while**
Output L^k, S^k, C^k

4. Experimental Results and Discussion

Our main application of interest is moving object detection in time-lapse videos with varing illumination. Therefore, we evaluate our method under two increasingly difficult conditions. First, we use datasets that contain moving objects and significant illumination changes or shadows but in real-time sequences with continuous object motion. Secondly, we use a challenging dataset that contains moving objects, illumination, and shadows, where images are captured via time-lapse or motion-trigger photography with large inter-image time intervals. In this case, the position of an object between two consecutive images may not be continuous. This is a common phenomenon in many long-term surveillance applications such as wildlife monitoring. Since real benchmark datasets only contain the first condition, we have built a new dataset which contains the second condition and use it in this paper. Then we perform two sets of experiments on benchmark and the newly proposed dataset.

4.1. Experiment Setup

Benchmark datasets: We evaluate our proposed method on selected sequences from the CDnet dataset [9], Wallflower dataset [21], and I2R dataset [11]. Since the goal of the experiments is to illustrate the ability of our method to detect real changes from illumination changes, we select sequences with varying illumination or moving shadows. From CDnet dataset four sequences are in this category de-

Figure 3. Selected images from each sequence of ICD.

Dataset	Sequences	Size $\times$ Number of frames
Illumination Changes Dataset (ICD)	Wildlife1	[508,358] $\times$ 194
	Wildlife2	[508,358] $\times$ 225
	Wildlife3	[508,358] $\times$ 136
	WinterStreet	[460,240] $\times$ 75
	MovingSunlight	[640,360] $\times$ 237

Table 1. Details of all sequences of ICD

picting indoor and outdoor scenes exhibiting moderate illumination changes and moving shadows. We also use sequences "Camouflage" and "LightSwitch" from Wallflower dataset and image sequence "Lobby" from I2R dataset, which include images with global and sudden illumination changes.

Illumination change dataset (ICD): In this paper we introduce the dataset that we have built, which includes five image sequences with severe illumination changes. Selected images from these sequences are shown in Fig. 3, and the number of images and the image size of each sequence are described in Table 1. Some of these images are without any object and just illumination change. The sequences of ICD are divided into two groups. The first three sequences are captured using a motion triggered camera for the wildlife monitoring application on different days. The last two sequences are taken with time-lapse photography of a large time interval to record changes of a scene that take place over time, which is common for many surveillance applications. Moving objects in the first sequence are under extreme sunlight or heavy shadow. Color of the objects in the second sequence are similar to the background or shadow and since illumination is changing, separating them is a difficult task. The third sequence shows objects with different size under varying illumination. The fourth sequence shows global illumination changes with moving shadows and the last row shows the sequence of images with moving objects while a strong moving sunbeam changes illumination of the scene. For each sequence among those frames that moving objects are available, we selected 15 frames with highest illumination changes. The ground truths of these selected frames were generated manually.

Evaluation metric: For quantitative evaluation, the performance metric pixel-level $F\!-\!measure = 2\frac{recall \times precision}{recall + precision}$ is used [2].

4.2. Evaluation on benchmark datasets

In the first set of experiments we use the sequences from benchmark datasets corresponding to Section 4.1 to evaluate the proposed method. We compare LISD as an intermediate results of our method and ILISD with the six related RPCA algorithms, namely SemiSoft GoDec (SSGoDec) [27], PRMF [23], PCP [4], Markov BRMF [24], DECOLOR [28], and LSD [15].

Table 2 shows performance of LISD and ILISD in comparison with the competing methods in terms of *F-measure*. The proposed method obtains the best average *F-measure* against all the other methods, and for the all sequences our method ranked among the top two of all methods. The first four rows of Table 2 are from CDnet dataset and our method has superior performance. The last three rows of Table 2 are from "Wallflower" and "I2R" datasets. For the "Camouflage" sequence a large object comes to the scene and therefore the global illumination is changed. In this case, only DECOLOR, LSD and our method detect the foreground object relatively well. LSD uses a structured sparsity term by selecting a maximum value of outliers in a specific neighborhood for pixels in each iteration. So it can keep the connectivity of outliers and classifies the foreground better than our method. Although we can use the structured sparsity term in our formulation instead of l_1-norm, the solution becomes significantly slow. For two sequences "Camouflage" and "LightSwitch" only one frame has ground-truth and the results are based on just one frame and cannot be reliable for the whole sequence; however, our method still is in the second place. For the "lobby" sequence ground-truth is available for some selected frames, but none of them show the ground-truth while illumination is changing. In this sequence the accuracy of our method is still in the second place and the accuracy of DECOLOR is a little better than ours.

4.3. Evaluation on ICD

In the second set of experiments we evaluate our proposed method on the sequences from ICD which has the most challenging condition and compare them with competing methods. Fig. 4 shows the results of our method to de-

Sequence	SSGoDec	PRMF	Decolor	PCP	BRMF	LSD	LISD	ILISD
Backdoor	0.6611	0.7251	0.7656	0.7594	0.6291	0.7603	<u>0.8015</u>	**0.8150**
CopyMachine	0.5401	0.6834	0.7511	0.6798	0.3293	<u>0.8174</u>	0.7832	**0.8179**
Cubicle	0.3035	0.3397	0.5503	0.4978	0.3746	0.4232	**0.7201**	<u>0.6887</u>
PeopleInShade	0.2258	0.5163	0.5559	0.6583	0.3313	0.6168	<u>0.6733</u>	**0.8010**
Camouflage	0.6452	0.6048	0.8125	0.3388	0.6048	**0.9456**	0.8605	<u>0.8663</u>
LightSwitch	0.3804	0.2922	0.5782	**0.8375**	0.2872	0.6640	0.6904	<u>0.7128</u>
Lobby	0.0831	0.6256	**0.7983**	0.6240	0.3161	0.7313	0.7830	<u>0.7849</u>

Table 2. Comparison of F-measure score between our proposed method and other compared methods on benchmark sequences (Best F-measure: Bold, Second best F-measure: Underline)

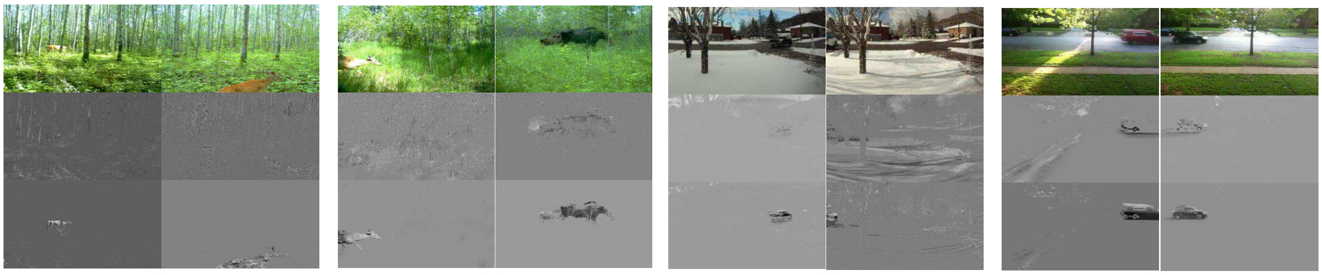

Figure 4. First row: two sample images with different illumination for each sequence. Second row: sparse outliers C. Third row: detected objects S.

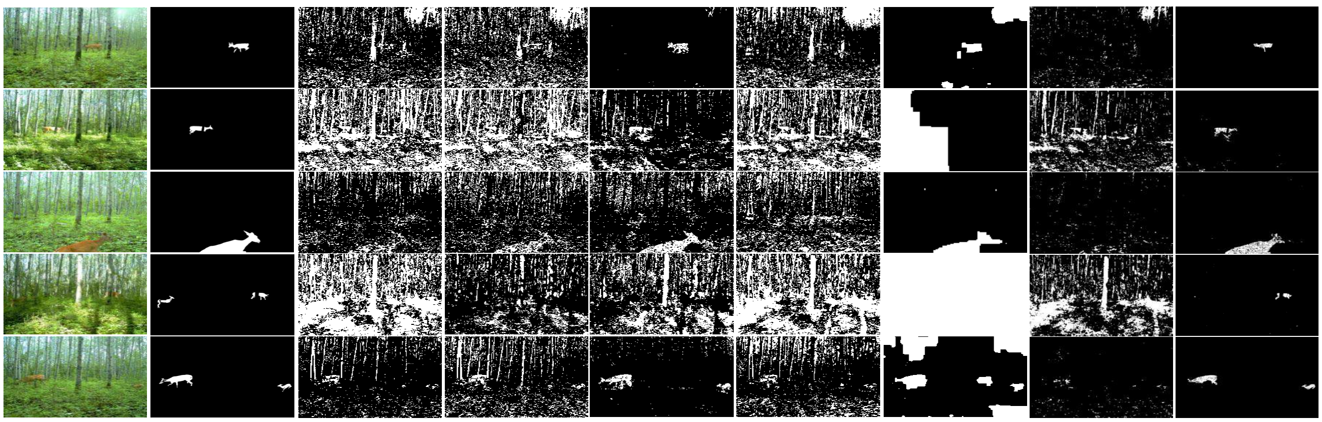

Figure 5. Comparison of qualitative results between our method and six competing methods on selected images of sequence "Wildlife3".

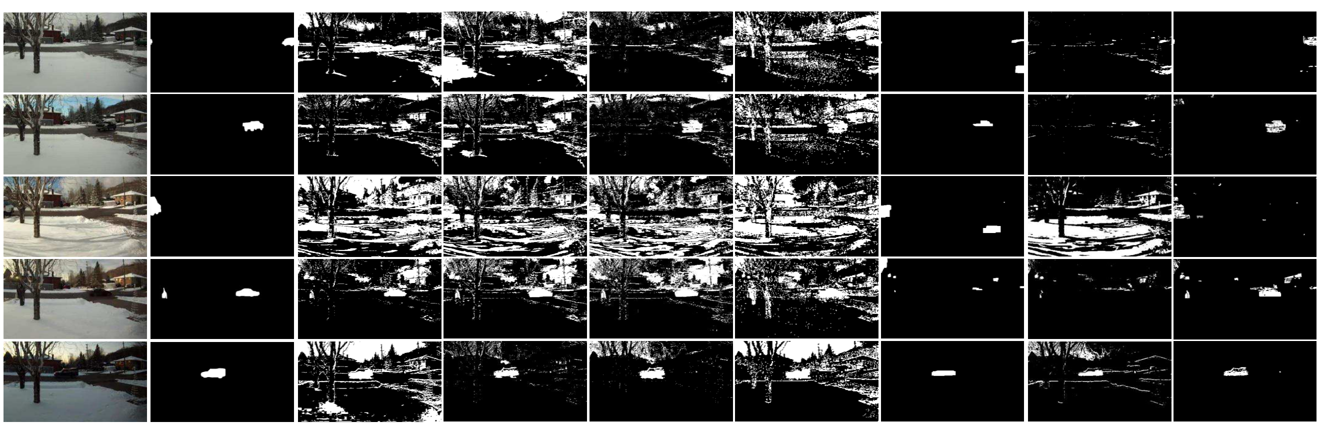

Figure 6. Comparison of qualitative results between our method and six competing methods on selected images of sequence "WinterStreet".

tect objects and separating them from illumination changes. In the first row of Fig. 4, two samples per sequence are shown where real changes and illumination changes occur at the same time. The second and the third rows show the sparse outliers of C and S, respectively. Based on our experiments most of illumination changes can be classified as the outliers C and the real changes are separated into matrix S. To show the capability of the proposed method, we compare qualitatively and quantitatively the results of our method with the results of the competing six sparse decomposition methods. As a sample, we show the comparison of qualitative results on selected images of sequence "Wildlife3". Since the illumination variations in the time-lapse image sequences are significant, we show five images of the sequence in Fig. 5 to provide a better comparison. The second and the fourth rows depict heavy illumination changes, and all competing methods fail to detect objects. In the first, the third, and the last rows of Fig. 5, where the illumination is relatively unchanged, PCP and DECOLOR can show relatively meaningful results. However, both of

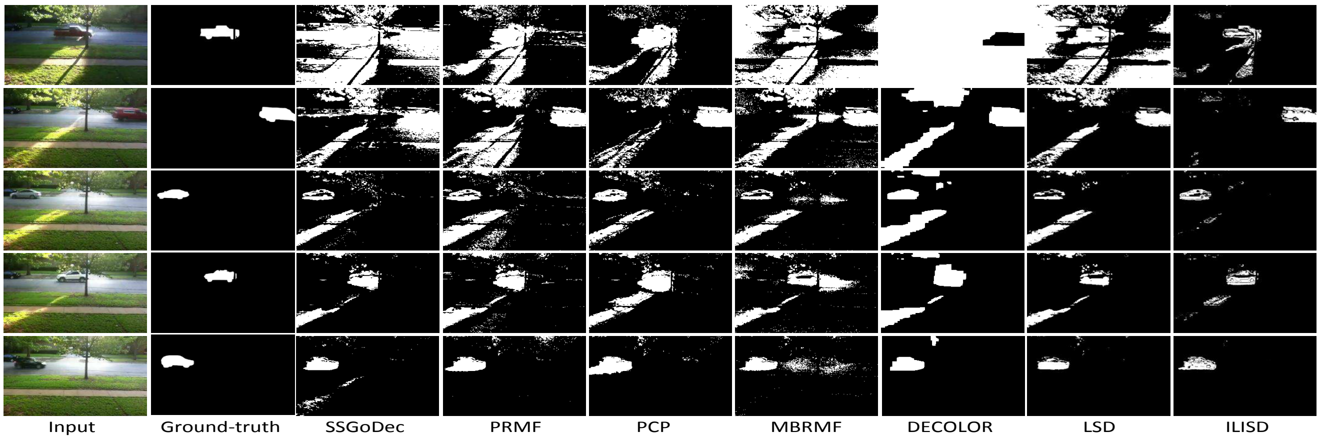

Figure 7. Comparison of qualitative results between our method and competing methods on selected images of sequence "MovingSunlight"

Sequence	SSGoDec	PRMF	PCP	MBRMF	DECOLOR	LSD	LISD	ILISD
Wildlife1	0.2826 / 0.2113	0.2586 / 0.2000	0.5968 / 0.2042	0.2679 / 0.2117	0.3409 / 0.2834	0.6480 / 0.1302	0.7747 / 0.0557	**0.8033** / 0.0416
Wildlife2	0.2585 / 0.1369	0.4141 / 0.2324	0.6430 / 0.0996	0.2654 / 0.1410	0.3517 / 0.2200	0.3899 / 0.1659	0.7168 / 0.0625	**0.7277** / 0.0298
Wildlife3	0.0753 / 0.0722	0.0754 / 0.0745	0.3124 / 0.2656	0.0510 / 0.0441	0.1019 / 0.0929	0.0871 / 0.0825	0.7318 / 0.1269	**0.7398** / 0.1234
WinterStreet	0.1120 / 0.0752	0.1677 / 0.1433	0.1766 / 0.1021	0.0871 / 0.0440	0.4575 / 0.2509	0.1604 / 0.1086	0.6869 / 0.0824	**0.6931** / 0.0928
MovingSunlight	0.2926 / 0.1927	0.2925 / 0.1732	0.3451 / 0.1387	0.2426 / 0.1403	0.3466 / 0.2590	0.3593 / 0.2426	0.6163 / 0.1393	**0.6475** / 0.1601

Table 3. Comparison of F-measure score between our proposed method and other compared methods on ICD sequences

them are not reliable over the entire sequence.

We also compare the results of ILISD with the results of other competing methods on two more sequences "WinterStreet" and "MovingSunlight". For the first one, global illumination changes and for the second one, sunbeam is moving. The comparison results of these two sequences are shown in Figs. 6 and 7, respectively. Fig. 6 shows that only DECOLOR can be comparable with our method for sequence "WinterStreet" and all other methods fail. As Fig. 7 shows, although all competing methods can detect the foreground in the last row the same as our method, all of them make many false positives and even cannot show meaningful results for the first four rows.

For quantitative evaluation on all sequences of ICD, Table 3 shows the *F-measure* of the competing methods on the five sequences in ICD. For each sequence we also compute standard deviations of all results that can show the reliability of each method for clear conclusions on the performance of the proposed method. The numerical results demonstrate that our method can provide better performance in handling such illumination changes than other competing methods in time-lapse videos.

5. Conclusion

In this paper, we have proposed a novel method named LISD to detect moving objects in a time-lapse video using the framework of low-rank and sparse decomposition. In our proposed method, first a prior map is built based on an illumination invariant representation and then the obtained prior map is used into the proposed low-rank and invariant sparse decomposition framework to extract foreground under severe illumination changes. We also have proposed an iterative version of LISD by updating the prior map in one representation and impose it as a constraint into the LISD formulation with another representation. Based on our extensive experiments on real data sequences from public datasets, we are able to establish that LISD and ILISD achieve the best performance in comparison with all evaluated methods including the state-of-the-art methods. We have also constructed novel datasets involving time-lapse sequences with significant illumination changes, which will be publicly upon publication of this paper. Currently, many security and wildlife monitoring cameras use a motion trigger sensors or timer-lapse photography. The problem of moving object detection in these cases cannot be solved by existing methods and we proposed a new formulation to solve the problem with a great improvement in comparison with the existing methods.

Despite its satisfactory performance in all our experiments, a challenge facing our proposed method is dynamic background. The reason is our proposed method uses l_1-norm for outliers without any structural constraint. In the future, we plan to develop a version of LISD that can work with dynamic background.

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) through the NSERC Canadian Field Robotics Network (NCFRN) and by Alberta Innovates Technology Future (AITF).

References

[1] T. Bouwmans and E. H. 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. 2

[2] S. Brutzer, B. Höferlin, and G. Heidemann. Evaluation of background subtraction techniques for video surveillance. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1937–1944. IEEE, 2011. 6

[3] J.-F. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010. 4

[4] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011. 1, 2, 6

[5] B.-H. Chen and S.-C. 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. 1

[6] L.-H. Chen, Y.-H. Yang, C.-S. Chen, and M.-Y. Cheng. Illumination invariant feature extraction based on natural images statisticstaking face images as an example. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 681–688. IEEE, 2011. 3

[7] X. Cui, J. Huang, S. Zhang, and D. N. Metaxas. Background subtraction using low rank and group sparsity constraints. In *European Conference on Computer Vision*, pages 612–625. Springer, 2012. 1, 2

[8] G. D. Finlayson, M. S. Drew, and C. Lu. Entropy minimization for shadow removal. *International Journal of Computer Vision*, 85(1):35–57, 2009. 3

[9] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar. Changedetection. net: A new change detection benchmark dataset. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 1–8. IEEE, 2012. 5

[10] C. Guyon, T. Bouwmans, and E.-H. 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. 1, 2

[11] L. Li, W. Huang, I. Y.-H. Gu, and Q. Tian. Statistical modeling of complex backgrounds for foreground object detection. *IEEE Transactions on Image Processing*, 13(11):1459–1472, 2004. 5

[12] S. Z. Li. *Markov random field modeling in image analysis*. Springer Science & Business Media, 2009. 2

[13] Z. Lin, M. Chen, and Y. Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2010. 4

[14] Z. Lin, R. Liu, and Z. Su. Linearized alternating direction method with adaptive penalty for low-rank representation. In *Advances in neural information processing systems*, pages 612–620, 2011. 5

[15] X. Liu, G. Zhao, J. Yao, and C. Qi. Background subtraction based on low-rank and structured sparse decomposition. *IEEE Transactions on Image Processing*, 24(8):2502–2514, 2015. 2, 6

[16] J. Mairal, R. Jenatton, F. R. Bach, and G. R. Obozinski. Network flow algorithms for structured sparsity. In *Advances in Neural Information Processing Systems*, pages 1558–1566, 2010. 2

[17] N. M. Oliver, B. Rosario, and A. P. Pentland. A bayesian computer vision system for modeling human interactions. *IEEE transactions on pattern analysis and machine intelligence*, 22(8):831–843, 2000. 1

[18] M. Shakeri and H. Zhang. Corola: a sequential solution to moving object detection using low-rank approximation. *Computer Vision and Image Understanding*, 146:27–39, 2016. 1, 2

[19] M. Shakeri and H. Zhang. Illumination invariant representation of natural images for visual place recognition. In *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, pages 466–472. IEEE, 2016. 4

[20] Y. Tian, A. Senior, and M. Lu. Robust and efficient foreground analysis in complex surveillance videos. *Machine vision and applications*, 23(5):967–983, 2012. 1

[21] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 1, pages 255–261. IEEE, 1999. 5

[22] A. Vinciarelli, M. Pantic, and H. Bourlard. Social signal processing: Survey of an emerging domain. *Image and vision computing*, 27(12):1743–1759, 2009. 1

[23] N. Wang, T. Yao, J. Wang, and D.-Y. Yeung. A probabilistic approach to robust matrix factorization. In *European Conference on Computer Vision*, pages 126–139. Springer, 2012. 2, 6

[24] N. Wang and D.-Y. Yeung. Bayesian robust matrix factorization for image and video processing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1785–1792, 2013. 2, 6

[25] B. Xin, Y. Tian, Y. Wang, and W. Gao. Background subtraction via generalized fused lasso foreground modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4676–4684, 2015. 2

[26] J. Xu, V. K. Ithapu, L. Mukherjee, J. M. Rehg, and V. Singh. Gosus: Grassmannian online subspace updates with structured-sparsity. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3376–3383, 2013. 2

[27] T. Zhou and D. Tao. Godec: Randomized low-rank & sparse matrix decomposition in noisy case. In *International conference on machine learning*. Omnipress, 2011. 2, 6

[28] X. Zhou, C. Yang, and W. Yu. Moving object detection by detecting contiguous outliers in the low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3):597–610, 2013. 2, 6