EFFICIENT BACKGROUND SUBTRACTION WITH LOW-RANK AND SPARSE MATRIX DECOMPOSITION

Salehe Erfanian Ebadi [*], *Valia Guerra Ones* [†], *Ebroul Izquierdo* [*]

[*] Queen Mary University of London, [†] Delft University of Technology

ABSTRACT

Decomposition of a video scene into background and foreground is an old problem, for which novel approaches in the last years have been proposed. The robust subspace approach based on a low-rank plus sparse matrix decomposition has shown a great ability to identify static parts from moving objects in video sequences. However, those models are still insufficient in realistic environments. In this paper, we propose a modified approximated robust PCA algorithm that can handle moving cameras and takes advantage of the block sparse structure of the pixels corresponding to the moving objects. Additionally, we propose a novel SVD-free algorithm for the case of rank-1 background that outperforms current state-of-the-art methods in computation cost/time as well as performance. Finally, experiments and numerical results evaluating the proposed methods are demonstrated.

Index Terms— Background subtraction, robust principal component analysis, matrix decomposition, low-rank, sparse.

1. INTRODUCTION

The separation of locally moving or deforming image areas from a static or globally moving background is a basic video processing task with manifold applications. The research in this paper addresses this fundamental task by leveraging and building on recent developments in the field of *Robust Principal Component Analysis* (RPCA). Specifically, the work reported here has been inspired by the critical breakthrough accomplished by Candès *et al.* [1], where the authors provided a practical solution for the long-standing problem of recovering the low-rank and sparse parts of a large matrix made up of the sum of these two components. In particular context of video processing, the 2-dimensional matrix A of size $m \times n$ stores pixel information of a video sequence I_j, $j = 1, \ldots, n$, or a set of images by concatenating each frame I_j as a column A_j in A. Then, the background part of the video sequence is modeled by the low-rank matrix, while the locally deforming parts constitute the sparse matrix component. More specifically they minimize a surrogate model using the ℓ_1 and nuclear norms with the convex optimization problem named *Principal Component Pursuit* (PCP) that with a high probability the solution of (1) can recover the low-rank (background) and the sparse (foreground) parts of the original matrix.

$$\arg\min \|L\|_* + \lambda \|S\|_1 \quad \text{subject to} \quad A = L + S \qquad (1)$$

where L and S are matrices of the same size as A, and $\|\cdot\|_*$ and $\|\cdot\|_1$ are the nuclear norm (which is the ℓ_1-norm of the singular values) and the ℓ_1-norm respectively, and λ is a balanced parameter (which according to [1] is assigned as $\frac{1}{\sqrt{max(m,n)}}$). PCP has led to impressive results in background modeling, foreground detection, removal of shadows and specularities in images, and face alignment for recognition. Although this formulation leads to a computationally feasible solution, the complexity is still high involving the calculation of many *Singular Value Decompositions* (SVD) for a very large matrix.

In this paper for practical reasons, we assume that the matrix A is decomposable; i.e., the matrix A is close to a matrix that can be written as the sum of a low-rank matrix L with singular vectors that are not spiky, and a sparse matrix S with a uniform and random pattern of sparsity. This paper addresses a number of critical issues and limitations of RPCA which are: embedding global motion parameters in the model, i.e., estimation of global motion parameters simultaneously with the foreground/background separation task; considering matrix block-sparsity rather than generic matrix sparsity as natural feature in video processing applications; and more critically providing an extremely efficient algorithm to solve the low-rank/sparse decomposition task. The first model aims video sequences captured by a moving camera, by estimating the global motion parameters while performing the targeted background/foreground separation task. The second model exploits the fact that in video processing applications the sparsity of the sparse matrix has a very special structure. In other words, the non-zero matrix entries are not randomly distributed but they build small blocks within the sparse matrix. Finally, the last solution targets the fact that RPCA approaches are computationally expensive. The proposed model introduces an extremely efficient "SVD-free" technique that can be applied in most background/foreground separation tasks.

2. RELATED WORK

Many algorithms have been developed to address and solve the problems in background modeling and foreground detection. Zhou *et al.* [2] proposed a method called DECOLOR in which they segmented moving objects from an image sequence by incorporating a Markov Random Fields (MRF) framework, and solving a non-convex penalty. Since DECOLOR minimizes a non-convex energy via alternating optimization, it converges to a local optimum with results depending on initialization of the foreground support, while PCP always minimizes its energy globally.

In another work, this time addressing the complexity issue, Zhou and Tao [3] proposed the approximated RPCA (called GoDec). GoDec aims at providing an approximate solution of the low-rank/sparse decomposition in presence of noise, that converges to a local minimum, when the exact and unique decomposition does not exist in more realistic situations. They estimate the low-rank part L and the sparse part S of a large matrix containing additive noise part G as:

$$\min_{L,S} \|A - L - S\|_F^2 \quad \text{such that} \quad rank(L) \leq k, card(S) \leq \kappa \quad (2)$$

where $\|\cdot\|_F$ is the Frobenius norm. However in the optimization process the rank of L and cardinality of S are fixed which imposes limitations to the decomposition of unconstrained real-world video

sequences. Moreover the hard-thresholding towards S requires sorting all its entries' magnitudes and thus is computationally expensive.

Recently, some works have been developed in which block-sparsity structure in the optimization process has been exploited [4], [5]. In these models the matrix S contains mostly zero columns, with several non-zero ones corresponding to foreground elements. In image processing applications this assumption cannot be made, since a whole column representing a frame cannot be zero in the sparse component. Furthermore, assuming that most columns of S are zero contradicts the definition of sparse matrix. When a whole column in the sparse matrix is zero it means the information in that column is assigned to the low-rank subspace. Moreover, if the video sequence contains foreground objects in all the frames this assumption does not help. In the next sections we present a number of solutions for the aforementioned critical issues with RPCA based solutions.

3. τ-DECOMPOSITION

The robust subspace learning models via matrix decomposition in the literature can mostly handle video sequences captured by static cameras. In this section building on [1], [6], and [3] an extension of the approximated RPCA model is proposed with the introduction of domain transformations into the optimization task, to compensate for background motion which is caused by camera movement. Basically, it is assumed that the columns of L are linearly dependent up to a certain parametric transformation. Given a data matrix A whose columns are the frames of a video sequence, captured by a moving camera, we write the decomposition of matrix A as:

$$A \circ \tau = L + S + G \tag{3}$$

where G is a matrix that contains the incomplete information and corruption by outliers it the original video sequence, e.g., Gaussian noise. $A_j \circ \tau_j$ denotes the j-th frame after transformation parameterized by the vector $\tau_j \in \mathbb{R}^\rho$ where ρ is the number of parameters fully describing the global motion model. Therefore, $\rho = 4$ corresponds to similarity, $\rho = 6$ to affine, and $\rho = 8$ to projective transformation. The i-th geometric transformation is comprised of a parameter vector τ_i, $i = 1, \ldots, n$ where different spatial transformations can be considered. We use 2D parametric transforms to model translation, rotation, and planar deformation of the background. Finally we use the multi-resolution incremental refinement described in [7], to estimate these motion parameters. We propose a computationally-cheaper (compared to [6]) algorithm based on an approximated RPCA formulation.

Given the data matrix A and the Lagrange tuning parameter λ the following optimization function recovers a low-rank matrix L, a sparse matrix S, and the motion parameter vector τ such that $A \circ \tau \approx L + S$:

$$\underset{\substack{L,S,\tau \\ rank(L)\leq k}}{\arg\min} \ \|A \circ \tau - L - S\|_F + \lambda \|S\|_1 \tag{4}$$

The first summand guarantees the approximations of the decomposition (minimizing the residual) and the second favors the sparse matrix solution S with many zero elements (i.e. sparse enough). The parameter λ controls the contribution of each summand to the function to be minimized. λ needs to be manually set depending on the problem to be solved and increases the model's flexibility and generalizability to different scenarios. The model is tested using variations of this parameter in our experiments with the *Receiver Operating Characteristic* (ROC) performance evaluation. We follow an alternating strategy minimizing the function for three parameters L, S, and τ one at a time until the solution reaches convergence (in an iterative process) to a local optima:

$$\tau^t = \underset{\tau}{\arg\min} \|A \circ \tau - L^{t-1} - S^{t-1}\|_F^2 \tag{5}$$

$$L^t = \underset{rank(L)\leq k}{\arg\min} \|A \circ \tau^t - L - S^{t-1}\|_F^2 \tag{6}$$

$$S^t = \underset{S}{\arg\min} \|A \circ \tau^t - L^t - S\|_F^2 + \lambda \|S\|_1 \tag{7}$$

The problem (5) can be written as a weighted least squares minimization where the solutions τ_i have a closed-form. To calculate the rank-k matrix that is the nearest estimate of the matrix $A \circ \tau^t - S^{t-1}$ in (6), SVD gives a closed-form solution as:

$$L^t = \sum_{i=1}^{k} \sigma_i U_i V_i^T$$

with the coefficients σ_i and the vectors U_i and V_i are the singular values, and the left and right singular vectors of the matrix $A \circ \tau^t - S^{t-1}$, respectively. Finally in (7) the matrix S^t is updated using the parameter λ acting as a tuning parameter in the matrix $A \circ \tau^t - L^t$; i.e. the elements of the matrix $A \circ \tau^t - L^t \leq \lambda$ are considered zero.

4. BLOCK-SPARSITY

The formulation of background modeling/foreground detection problem using the optimization function (4) favors solutions where the matrix S is sufficiently sparse. But this information does not take into account the structure of sparsity in S, and therefore does not yield good results when the sparse pattern involves for example clutters of non-zero entries representing foreground objects. Indeed, in real-world video sequences the foreground pixels do not appear as in a sparse matrix at random and scattered; strictly speaking they appear in regions of pixels corresponding to foreground objects in a scene. It would make sense if the block-sparsity is imposed on the pixels of each video frame rather than a whole column (whole frame) in the matrix S. This mathematical solution would favor solutions where the zero elements of the matrix S appear in blocks, where each block can represent the natural shape of a foreground object. We define $mat(\cdot)$ a mapping operator from the m-dimensional space into the $w \times h$ matrix as $\mathbb{R}^m \rightarrow \mathbb{R}^{w \times h}$. In other words $mat(A_j)$ is equal to video frame I_j. Hence we solve the minimization problem for block-sparse matrices $mat(S_j)$. Given a data matrix A whose columns are the frames of a video sequence captured by a moving camera and a Lagrange parameter λ, we minimize the following optimization problem that recovers the background and foreground of the sequence with the matrices L and S, respectively:

$$\underset{\substack{S,\tau \\ rank(L)\leq k}}{\arg\min} \ \|A \circ \tau - L - S\|_F^2 + \lambda \sum_{j=1}^{n} \|mat(S_j)\|_{2,1} \tag{8}$$

where $\ell_{2,1}$-norm is defined as $\|A\|_{2,1} = \sum_j \|A_j\|_2$ which is the ℓ_1-norm of the vector formed by taking the ℓ_2-norms of the columns of the underlying matrix. The optimization procedure is similar to the alternating strategy between three sub-problems in the previous section.

$$\tau^t = \underset{\tau}{\arg\min} \|A \circ \tau - L^{t-1} - S^{t-1}\|_F^2 \tag{9}$$

$$L^t = \underset{rank(L)\leq k}{\arg\min} \|A \circ \tau^t - L - S^{t-1}\|_F^2 \tag{10}$$

$$S^t = \underset{S}{\arg\min} \|A \circ \tau^t - L^t - S\|_F^2 + \lambda \sum_{j=1}^{n} \|mat(S_j)\|_{2,1} \tag{11}$$

We have found a closed form expression for the solution of the minimization problem (11) that facilitates the algorithm.

5. SVD-FREE SOLUTION TO RPCA

Considering a particular case, where the background in a video sequence does not change, and it can be described by a rank-1 matrix, the optimization problem of approximated RPCA algorithm is:

$$\underset{\substack{S,L,\tau \\ rank(L)=1 \\ card(S)\leq\kappa}}{\arg\min} \ \|A\circ\tau - L - S\|_F \tag{12}$$

Note that the Lagrange parameter λ is left out, and instead the number of non-zero elements $card(S)$ is being fixed. The cardinality κ acts as a hard-thresholding parameter that controls the quality of the reconstruction of A using the matrices L and S. The rank-1 restriction for L imposed in the optimization problem yields to solutions where the columns of the matrix L can be written as $L_j \leftarrow \alpha L_1$, $j = 1,\ldots,n$, where L_1 is the first column of L and α is a scalar. This is based on the fact that not all the columns of a rank-1 matrix are necessarily equal, and rather the columns are linearly dependent (by a scalar factor). Based on this fact, we assume a particular rank-1 matrix L where all the column vectors are equal; i.e. $L = l\mathbb{1}^T$ where l is a vector of size m and $\mathbb{1}^T = (1,\ldots,1)$. Note that any matrix in the form $l\mathbb{1}^T$ is a rank-1 matrix but not all rank-1 matrices can be written by repeating the same vector in all the columns.

The main advantage of this special rank-1 matrix is that we prove the vector l can be calculated without computing a SVD; therefore, this algorithm converges much faster as a result, since the most expensive computation in the described algorithms is in SVD calculation step. The optimization model is as below:

$$\underset{\substack{S,l,\tau \\ card(S)\leq\kappa}}{\arg\min} \ \|A\circ\tau - l\mathbb{1}^T - S\|_F \tag{13}$$

Similar to previous sections the optimization process includes an alternating strategy as below:

$$\tau^t = \underset{\tau}{\arg\min}\, \|A\circ\tau - l^{t-1}\mathbb{1}^T - S^{t-1}\|_F^2 \tag{14}$$

$$L^t = \underset{l}{\arg\min}\, \|A\circ\tau^t - l\mathbb{1}^T - S^{t-1}\|_F^2 \tag{15}$$

$$S^t = \underset{card(S)\leq\kappa}{\arg\min}\, \|A\circ\tau^t - l^t\mathbb{1}^T - S\|_F^2 \tag{16}$$

The matrix S^t that solves the optimization problem (16) is the matrix with zero elements in the positions corresponding to the fits κ smallest elements of the matrix $|A\circ\tau^t - l^t\mathbb{1}^T|$.

6. RESULTS

The proposed algorithms in sections 3, 4, and 5 were implemented and tested in MATLAB *8.3.0.532 (R2014a)* on a 64-bit PC with Intel Core i7-4770 CPU @3.40GHz (single core) and 32GB of RAM. A C++ implementation that performs similarly to that of MATLAB has also been developed. For the evaluations, 7 challenging background subtraction datasets ([8], [9], [10], [11], [12], [13], [14]) are tested with our models. They vary in resolution, quality, frame number, and general scene scenarios which guarantees an unbiased evaluation. A complete description of challenges is available in [15].

Figure 1 shows the results for a sequence captured by a moving camera. The marked green region corresponds to the recovered rank-1 background across the whole sequence with the motion parameters, which is visible in the selected frame.

To demonstrate the block-sparse model described in this paper, figure 2 shows a comparison between our block-sparse model and the RPCA-LBD model [4], [5] for a complex video sequence with dynamic background with water rippling. Columns (b) and (d) correspond to the low-rank and sparse parts obtained by the RPCA-LBD for $\lambda = 0.6710$ (as tuned by authors in their original paper). Our results were obtained with model parameters $rank(L) = 1$, and $lambda = 0.03$ in columns (c) and (e).

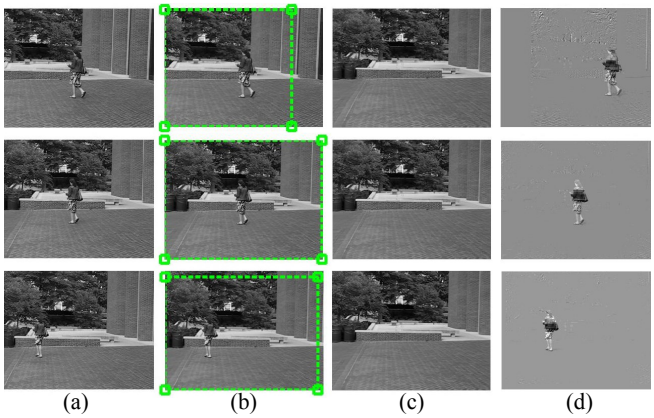


Fig. 1: Decomposition of a sequence with moving camera. (a) Original video frames 1, 20, and 40. (b) Reconstruction $A = L + S$. (c) Motion-compensated extracted L. (d) Motion-compensated extracted S.

Table 1: Time performance comparison (CPU sec.).

	CDW	BMC	CM	SAI	i2R	MuHAVi-MAS
# Frames	6049	591	500	600	15462	466
RPCA	1931.12	116.94	65.67	454.59	646.52	380.85
GoDec	1874.82	49.67	44.26	376.91	480.33	203.17
Our Model	358.91	20.82	17.38	117.69	209.84	70.27

We also evaluate our method for the task of foreground segmentation. The accuracy of foreground detection is measured by comparing the calculated foreground support with the binary ground-truth images. Figure 3 shows the unrefined segmentation results for a general surveillance sequence. Notice the accuracy and coherence of the segmentation in our results as compared to that of GoDec. The proposed algorithm can handle objects that occupy large portions of the frame as well as small objects (such as pedestrians in this scene) equally well simultaneously. The *Receiver Operating Characteristic* (ROC) curves obtained with Precision-Recall values for six datasets in figure 4 show the performance of our method against GoDec for varying thresholds. The results here guarantee superior performance for all datasets in segmentation accuracy. The Precision and Recall values are calculated with the number of pixels that are classified with regard to ground-truth as True-Positive tp, True-Negative tn, False-Positive fp, and False-Negative fn.

$$Precision = \frac{tp}{tp+fp}$$

$$Recall = \frac{tp}{tp+fn} = \frac{\text{\#correctly classified foreground pixels}}{\text{\#foreground pixels in GT}}$$

The CPU time consumption of our SVD-free model, for performing the decomposition task is shown in table 1 against Original RPCA [1], and approximated RPCA (GoDec) [3]. All the algorithms were run with 5 iterations and the tuning parameters were chosen to obtain maximal segmentation accuracy.

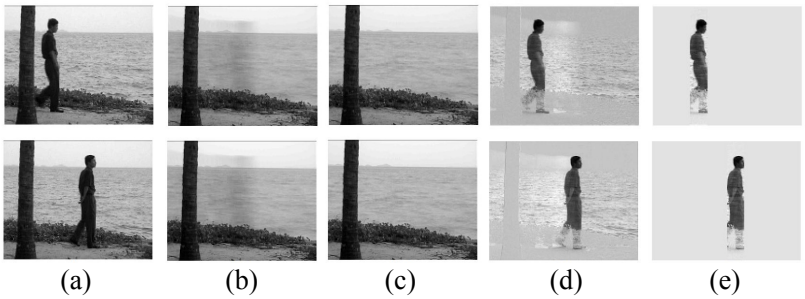


Fig. 2: RPCA-LBD algorithm vs. our block-sparse model. (a) Original video for frames 24 and 48. (b) L, RPCA-LBD. (c) L, our block-sparse. (d) S, RPCA-LBD. (e) S, our block-sparse.

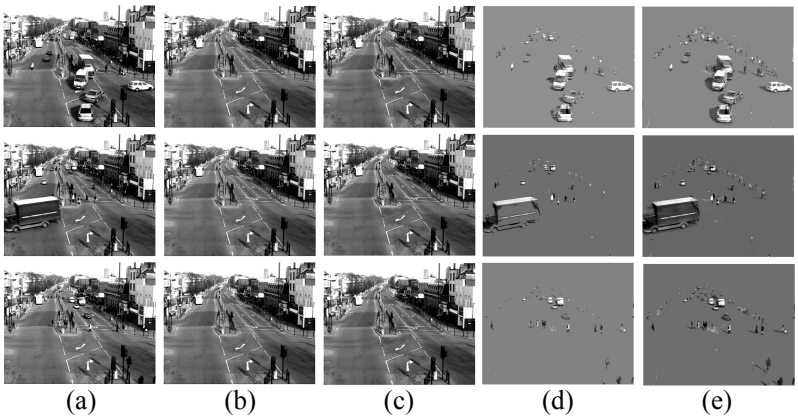
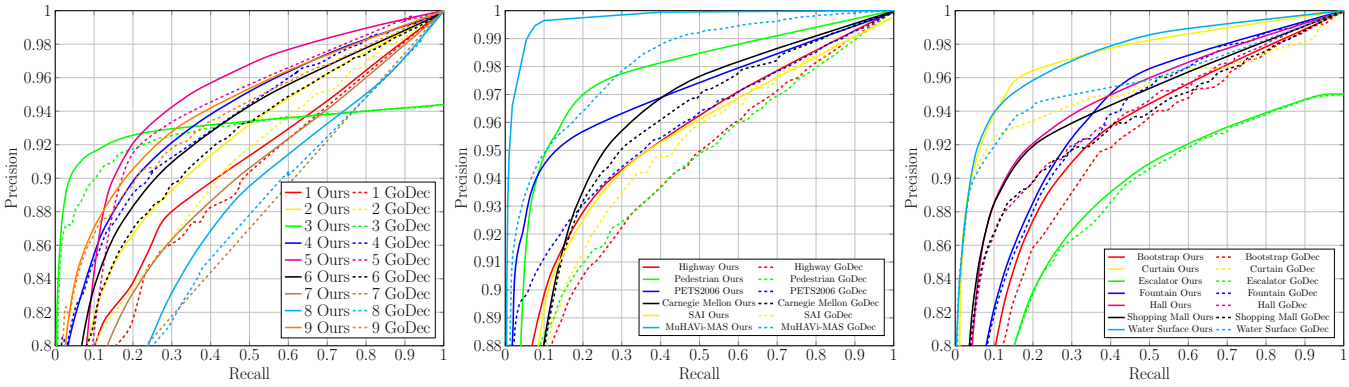


Fig. 3: Segmentation results comparison. (a) Original frames, (b) L GoDec, (c) L ours, (d) S GoDec, (e) S ours.



(a) BMC (b) CDW, CM, SAI, and MuHAVi-MAS (c) i2R

Fig. 4: Receiver Operating Characteristic (ROC) curves for the performance of our method vs. GoDec with varying thresholds.

7. CONCLUSION

In this article we proposed a number of improvements on the approximated RPCA algorithm as well as a novel SVD-free solution to the optimization problem for fast computation. The solutions presented in this paper aim to solve issues that arise with RPCA-based methods in foreground/background segmentation of general video sequences. Our proposed method can handle camera movement, various foreground object sizes, and slow-moving foreground pixels as well as sudden and gradual illumination changes in a scene. The qualitative and quantitative segmentation results outperform current state-of-the-art methods. Our SVD-free solution achieves more than double the amount of speed-up in computation time for the same performance target compared to its counterpart. In future we would like to move towards unconstrained cases where the captured video could have any motion, parametric transformation, quality, motion blur, or deformation of scene elements.

8. REFERENCES

[1] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright, "Robust principal component analysis?," *J. ACM*, vol. 58, no. 3, pp. 11:1–11:37, June 2011. 1, 2, 3

[2] Xiaowei Zhou, Can Yang, and Weichuan Yu, "DECOLOR: Moving object detection by detecting contiguous outliers in the low-rank representation.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 597–610, 2013. 1

[3] Tianyi Zhou and Dacheng Tao, "GoDec: Randomized low-rank and sparse matrix decomposition in noisy case," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, Lise Getoor and Tobias Scheffer, Eds. June 2011, ICML '11, pp. 33–40, ACM. 1, 2, 3

[4] Gongguo Tang and Arye Nehorai, "Robust principal component analysis based on low-rank and block-sparse matrix decomposition," in *45th Annual Conference on Information Sciences and Systems, CISS, The John Hopkins University, Baltimore, MD, USA, 23-25 March 2011*, 2011, pp. 1–5. 2, 3

[5] C Guyon, T Bouwmans, and E Zahzah, "Foreground detection based on low-rank and block-sparse matrix decomposition," in *International Conference on Image Processing, ICIP 2012*, 2012. 2, 3

[6] YiGang Peng, Arvind Ganesh, John Wright, Wenli Xu, and Yi Ma, "RASL: robust alignment by sparse and low-rank decomposition for linearly correlated images," *IEEE Transactions on Pattern Analysis and Machine Intellignece*, vol. 34, no. 11, pp. 2233–2246, 2012. 2

[7] Richard Szeliski, *Computer Vision: Algorithms and Applications*, Springer-Verlag New York, Inc., New York, NY, USA, 1st edition, 2010. 2

[8] Antoine Vacavant, Laure Tougne, Thierry Chateau, and Lionel Robinault, "Background Models Challenge, Workshop of ACCV 2012," Springer, Nov. 2012, http://liris.cnrs.fr/publis/?id=5905. 3

[9] Yi Wang, Pierre-Marc Jodoin, Fatih Porikli, Janusz Konrad, Yannick Benezeth, and Prakash Ishwar, "CDnet 2014: An Expanded Change Detection Benchmark Dataset," in *IEEE CVPR Change Detection workshop*, United States, June 2014, p. 8 p., https://hal-univ-bourgogne.archives-ouvertes.fr/hal-01018757. 3

[10] Yaser Sheikh and Mubarak Shah, "Bayesian modeling of dynamic scenes for object detection," *PAMI*, vol. 27, pp. 1778–1792, 2005. 3

[11] Sebastian Brutzer, Benjamin Höferlin, and Gunther Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *Computer Vision and Pattern Recognition (CVPR) IEEE*, 2011, pp. 1937–1944. 3

[12] Sanchit Singh, Sergio A. Velastin, and Hossein Ragheb, "MuHAVi: A multicamera human action video dataset for the evaluation of action recognition methods," in *Proceedings of the 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*, Washington, DC, USA, 2010, AVSS '10, pp. 48–55, IEEE Computer Society. 3

[13] 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*, vol. 13, no. 11, pp. 1459–1472, 2004. 3

[14] T.Brox and J.Malik, "Object segmentation by long term analysis of point trajectories," in *European Conference on Computer Vision (ECCV)*. Sept. 2010, Lecture Notes in Computer Science, Springer. 3

[15] D. D. Bloisi, *Background Modeling and Foreground Detection for Video Surveillance*, CRC Press, Taylor and Francis Group, July 2014. 3