# BACKGROUND EXTRACTION FROM VIDEO SEQUENCES VIA MOTION-ASSISTED MATRIX COMPLETION

*Jingyu Yang*[1], *Xin Sun*[1], *Xinchen Ye*[1], *Kun Li*[2]

[1]School of Electronic Information Engineering
[2]School of Computer Science and Technology
Tianjin University, Tianjin 300072, China

## ABSTRACT

Background extraction from video sequences is a useful and important technique in video surveillance. This paper proposes a motion-assisted matrix completion model for background extraction from video sequences. A binary motion map is first calculated for each frame by optical flow. By excluding areas associated with moving objects with the binary motion maps, the background extraction is formulated into a motion-assisted matrix completion (MAMC) problem. Experimental results show that our method not only extracts promising backgrounds but also outperforms many state-of-the-art methods in distinguishing moving objects on challenging datasets.

***Index Terms*—** Background extraction, optical flow, motion detection, matrix completion, video surveillance

## 1. INTRODUCTION

Video analysis, [1, 2] is of crucial importance to dig out interesting information from mass data in videos. Background extraction, as a prerequisite technique to detect interesting objects, has been used in many applications such as motion detection, object recognition, and video coding. Background extraction from video sequences is to estimate an image that contains only the background of the captured scene rather than any interesting fore-object.

Background extraction techniques, as currently an active area of research, have made significant progress over the recent years. Many algorithms and models in terms of extracting background are proposed. The Gaussian mixture model (GMM) [3, 4] is the most popular pixel-level method. However, GMM algorithm is sensitive to light condition. Self-organization background subtraction (SOBS) [5] achieves robust detection for scenes with gradual illumination variations, but its model bears a high computational complexity burden. Barnich *et al.* put forward a powerful pixel-based background extraction algorithm named ViBe [6]. Recently, the robust principal components analysis (RPCA) [7] is essentially a matrix recovery problem. It is proved that, under some suitable assumptions, both the low-rank and the sparse components can be exactly recovered by solving a convex programming. However, these RPCA-based methods [7, 8, 9] do not seriously address complex motion characteristics in video sequences, and do not perform well in the videos which contain slowly-moving objects, camouflages, and large foreground objects. Methods above are all faced with several challenges arising from the practical video surveillance, for example, under changeable illumination condition, the background can hardly adapt to the environment [10] and camouflage: moving objects are difficultly identified, resulting in wrong classification.

In this paper, we propose a new background extraction method via motion-assisted matrix completion. The main idea is to incorporate motion information into the matrix completion framework to facilitate the separation of the foreground from the background. To this end, an optical flow estimation method is used to extract motion information from the video sequence. Binary motion maps are created by thresholding the estimated optical flow. The background extraction is formulated into a motion-assisted matrix completion (MAMC) problem by zeroing out areas associated with moving objects with the binary motion maps. The model is solved by the alternating direction method under the augmentation Lagrangian multiplier (ADM-ALM) framework. Experiments show that our method achieves consistently better performance than many state-of-the-art methods. The evaluations on challenging datasets demonstrate our method is quite versatile for surveillance videos with different types of motions and lighting conditions.

## 2. BACKGROUND EXTRACTION VIA MOTION-ASSISTED MATRIX COMPLETION

The powerful robust principal components analysis (RPCA) developed by Candes *et al.* [7] is suitable for background modeling in video analysis. However, the extracted background images present smearing artifacts on the regions where moving objects ever appear. These annoying artifacts are not only visually unpleasant, but also can deteriorate the performances of following modules, e.g., recognition or coding.

We observe that background is occluded by moving objects in most frames of the analyzed sequences. In the RPCA-based background extraction method, nuclear norm and $\ell_1$ norm are used to separate background from moving fore-objects under the convex optimization framework. Essentially, the nuclear-norm term describes low-frequency components along the temporal while the $\ell_1$ norm addresses the high-frequency components. However, the high-frequency components can leak into extracted background images for areas that are dominated by moving objects. The leakage as smearing artifacts present in extracted background cannot be well handled by adjusting the weights between the two regularization terms. In summary, the RPCA-based method suffers from the above artifacts due to the unawareness of motion information.

### 2.1. Proposed Background Extraction MAMC Frameworks

We propose a motion-assisted matrix completion for accurate background extraction, to overcome the aforementioned shortcoming of RPCA, and eliminate the deficiency of other methods in the presence of lighting conditions, camouflage, and different types of motion. The key idea is to assign the reliabilities of the observed pixels that would contribute to the background. The background is to be extracted from the $K$ frames of surveillance video sequences denoted by $\{f_k\}_{k=0}^{K-1}$ of size $M \times N$. For easy mathematical manipulation, let $\mathbf{f}_k$ be the vector form of frame $f_k$ with the size $MN \times 1$. Then, we represent the video sequences with matrix $\mathbf{F} = (\mathbf{f}_0, \mathbf{f}_1, ..., \mathbf{f}_{K-1})$ of size $MN \times K$, where each column is the vector form of a video frame. The recovered background component and complementary error in $\mathbf{F}$, are denoted by $\mathbf{B}$ and $\mathbf{E}$ respectively. The aim is to separate $\mathbf{B}$ and $\mathbf{E}$ from $\mathbf{F}$. Denote a matrix, named motion map, as $\mathbf{W} \in [0,1]^{MN\times K}$ whose element $w_{ik}$ represents the likelihood of pixel $f_{ik}$ in $\mathbf{F}$ that belongs to the motion region (hence foreground). We propose the model for background extraction as the following convex optimization:

$$\min_{\mathbf{B},\mathbf{E}} ||\mathbf{B}||_* + \lambda||\mathbf{E}||_1, \text{ subject to } \mathbf{W}\circ\mathbf{F} = \mathbf{W}\circ(\mathbf{B}+\mathbf{E}), \quad (1)$$

where $||\cdot||_*$ and $||\cdot||_1$ denote the nuclear norm and $\ell_1$ norm of a matrix, respectively, and "$\circ$" denotes element-wise multiplication of two matrices. Like previous methods, it is reasonable to assume the background to stay motionless in most practical surveillance applications (otherwise a global motion should be compensated). Under this assumption, any area with motion should not be considered as a part of background. Therefore, the motion map $\mathbf{W}$ is constructed from motion information (see Sec.2.2 ). By incorporating motion information, areas dominated by moving objects are suppressed while the background that appears at only a few frames has more chances to be recovered in the final results.

Model (1) extends the classic matrix recovery model by taking the likelihood of observed data into consideration. When $\mathbf{W}$ is an all-one matrix, Model (1) turns into the classic matrix recovery model. When $\mathbf{W}$ is a binary matrix (called binary motion map), i.e., $\mathbf{W} \in \{0,1\}^{MN\times K}$, it becomes the following matrix completion model:

$$\min_{\mathbf{B},\mathbf{E}} ||\mathbf{B}||_* + \lambda||\mathbf{E}||_1, \text{ subject to } P_\Omega(\mathbf{F}) = P_\Omega(\mathbf{B}+\mathbf{E}), \quad (2)$$

where $\Omega := \{\mathbf{Z}|\mathbf{Z} = \mathbf{W}\circ\mathbf{X}, \mathbf{X} \in R^{MN\times K}\}$ denotes the linear subspace of entries in the observed matrix that belong to background for sure, and $\mathcal{P}_\Omega(\cdot)$ is the associated projection operator. Note that whether a pixel belongs to the background or not is a binary decision. Model (2) is stronger to prevent moving objects from appearing in the recovered background than the more general Model (1). Therefore, we will develop our proposed method with Model (2), which is referred to as motion assisted matrix completion (MAMC).

### 2.2. Construction of Binary Motion Map W

In the proposed model, the motion map $\mathbf{W}$ is constructed from motion information. We use the optical flow method in [11] to extract a dense motion field $(\mathbf{O}_k^x, \mathbf{O}_k^y)$ for video frame $f$, where $\mathbf{O}_k^x$ and $\mathbf{O}_k^y$ are the horizontal component and vertical component of the motion field, respectively. Both $\mathbf{O}_k^x$ and $\mathbf{O}_k^y$ are in the vector form with the same organization as $\mathbf{f}_k$. We define $\mathbf{O}^x$ of size $MN \times K$ as the matrix form of horizontal motion fields for all frames in $\mathbf{F}$ by stacking $\mathbf{O}_k^x$, $k = 0, 1, ..., K-1$ as columns. Similarly, $\mathbf{O}^y$ is defined for vertical motion fields. The binary motion map $\mathbf{W}$ is constructed as follows

$$w_{ik} = \begin{cases} 0, \sqrt{(O_{ik}^x)^2 + (O_{ik}^y)^2} \geq \tau, \\ 1, \text{otherwise}, \end{cases} \quad (3)$$

where $\tau$ is a threshold to determine the values of entries. The threshold should be appropriate: too large a threshold would lead to underestimating of motion (hence smearing artifacts around moving objects) while too small a threshold would result in overestimating of motion (hence incompletely recovered background). The threshold is set at one in our experiments as a matter of experience.

### 2.3. The ADM-ALM Algorithm to Solve the MAMC Model

The MAMC model is essentially a convex optimization problem that can be solved by the alternate direction method under the framework of augmented Lagrange multipliers method[12, 13]. The idea of ALM framework is to convert the original constrained optimization problem (2) to the minimization of an augmented Lagrangian function. The augmented Lagrangian function of problem (2) is given by

$$L(\mathbf{B},\mathbf{E},\mathbf{Y},\mu) = ||\mathbf{B}||_* + \lambda||\mathbf{E}||_1 + \langle \mathbf{Y}, \mathcal{P}_\Omega(\mathbf{F}-\mathbf{B}-\mathbf{E})\rangle + \frac{\mu}{2}||\mathcal{P}_\Omega(\mathbf{F}-\mathbf{B}-\mathbf{E})||_F^2 \tag{4}$$

where $\mu$ is a penalty parameter, $\langle\cdot,\cdot\rangle$ denotes the matrix inner product, and $||\cdot||_F$ denotes the matrix Frobenius norm. Instead of optimizing $\mathbf{E}$ and $\mathbf{B}$ simultaneously, the ADM solves $\mathbf{E}$ and $\mathbf{B}$ alternatively:

$$\begin{cases} \mathbf{E}_{j+1} = \arg\min_{\mathbf{E}} \lambda||\mathbf{E}||_1 - \langle \mathbf{Y}_j, \mathcal{P}_\Omega(\mathbf{E})\rangle \\ \qquad + \frac{\mu_k}{2}||\mathcal{P}_\Omega(F-\mathbf{B}_j-\mathbf{E})||_F^2 \\ \mathbf{B}_{j+1} = \arg\min_{\mathbf{B}} ||\mathbf{B}||_* - < \mathbf{Y}_j, \mathcal{P}_\Omega(\mathbf{B}) > \\ \qquad + \frac{\mu_k}{2}||\mathcal{P}_\Omega(F-\mathbf{B}-\mathbf{E}_{j+1})||_F^2 \\ \mathbf{Y}_{j+1} = \mathbf{Y}_j + \mu_j \mathcal{P}_\Omega(F-\mathbf{B}_{j+1}-\mathbf{E}_{j+1}) \\ \mu_{j+1} = \rho\mu_j \end{cases} \tag{5}$$

The solution of $\mathbf{E}$ has the following closed-form

$$\mathbf{E}_{j+1} = \text{shrink}\left(\frac{1}{\mu_j}\mathbf{Y}_j + \mathcal{P}_\Omega(\mathbf{F}-\mathbf{B}_j), \frac{\lambda}{\mu_k}\right) \tag{6}$$

where $\text{shrink}(\cdot,\cdot)$ is the soft-thresholding function that is in an element-wise manner applied on the matrix. The $\mathbf{B}$ subproblem in (5) does not have a closed-form solution, and we resort to the accelerated proximal gradient algorithm given as:

$$\begin{cases} (\mathbf{U}_i, \mathbf{S}_i, \mathbf{V}_i) = \text{svd}\left(\frac{1}{\mu_j}\mathbf{Y}_j + \mathcal{P}_\Omega(\mathbf{F}) - \mathbf{E}_{j+1} + P_{\overline{\Omega}}(\mathbf{Z}_i)\right) \\ \mathbf{B}_{i+1} = \mathbf{U}_i \text{shrink}\left(\mathbf{S}_i, \frac{1}{\mu_k}\right)\mathbf{V}_i^T \\ \mathbf{Z}_{i+1} = \mathbf{B}_{i+1} + \frac{t_i - 1}{t_{i+1}}(\mathbf{B}_{i+1}-\mathbf{B}_i) \\ t_{i+1} = 0.5(1+\sqrt{1+4t_i^2}) \end{cases} \tag{7}$$

where $t_i$ is a positive sequence with $t_0 = 1$, $\overline{\Omega}$ denotes the complementary of $\Omega$, and $\text{svd}(\cdot)$ denotes the singular value decomposition of a matrix. The solution of Model (2), denoted by $\mathbf{B}^*$, is obtained after the convergence of the iterative procedure (5). $\mathbf{B}^*$ contains a background image for each frame. We take the average of all backgrounds in $\mathbf{B}^*$ as the final extracted background image; while $\mathbf{E}$ contains foreground information, light variation noise, and so on.

## 3. EXPERIMENTAL RESULTS

In this section, the proposed method is evaluated on four datasets of surveillance videos, i.e., *Cars* (320×240), *Browser1* (384×288), *Hallmonitor* (352×288), *Cell_phone_theft* (720×576). Besides background extraction, we also apply our method to motion detection to distinguish moving objects and compare with other five methods.

***Experiments on Background Extraction***: Fig. 1 presents the extracted background for the four videos by the proposed method, the RPCA-based method [7], and the SOBS algorithm [5]. Compared with other two methods, our method provides the best performance. The background recovered by our method is a quite reliable representation of the real background while the ones extracted by the RPCA-based method and the SOBS algorithm present severe smearing artifacts along the trajectories of moving objects as shown in the red rectangles. Moreover, for *Browser1* where moving objects are small, it is observed that the ghosting artifacts appear around the pavement. In general, our MAMC method indeed creates the best extracted background as indicated in the test results.

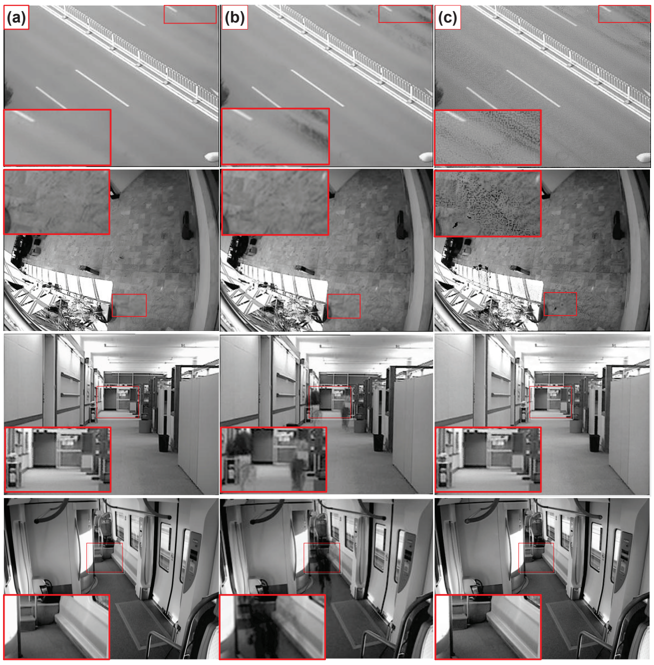


**Fig. 1**. Background extraction results for four test sequences: (a) Ours, (b) RPCA [7], and (c) SOBS [5]. Regions are highlighted in red rectangles for better visual inspection.

***Experiments on Motion Detection***: We apply our background extraction method to motion detection, and compare with five methods, i.e., GMM [15], SOBS [5], FBM [14], ViBe [6], and RPCA [7]. All the compared methods are tuned to yield their best results. For benchmarking, ground truth motion segmentation data of video sequences are manually created. For all the methods, we perform the same post processing (dilation morphological operations with a 3 × 3 square mask) on the obtained foreground binary maps.

In terms of objective comparison, we measure the correctness of the motion detection results with three metrics, namely Precision ($P$), Recall ($R$), and F-measure ($F_1$), defined as follows:

$$\begin{cases} R = tp/(tp+fn) \\ P = tp/(tp+fp) \\ F_1 = (2\times R\times P)/(R+P) \end{cases} \tag{8}$$

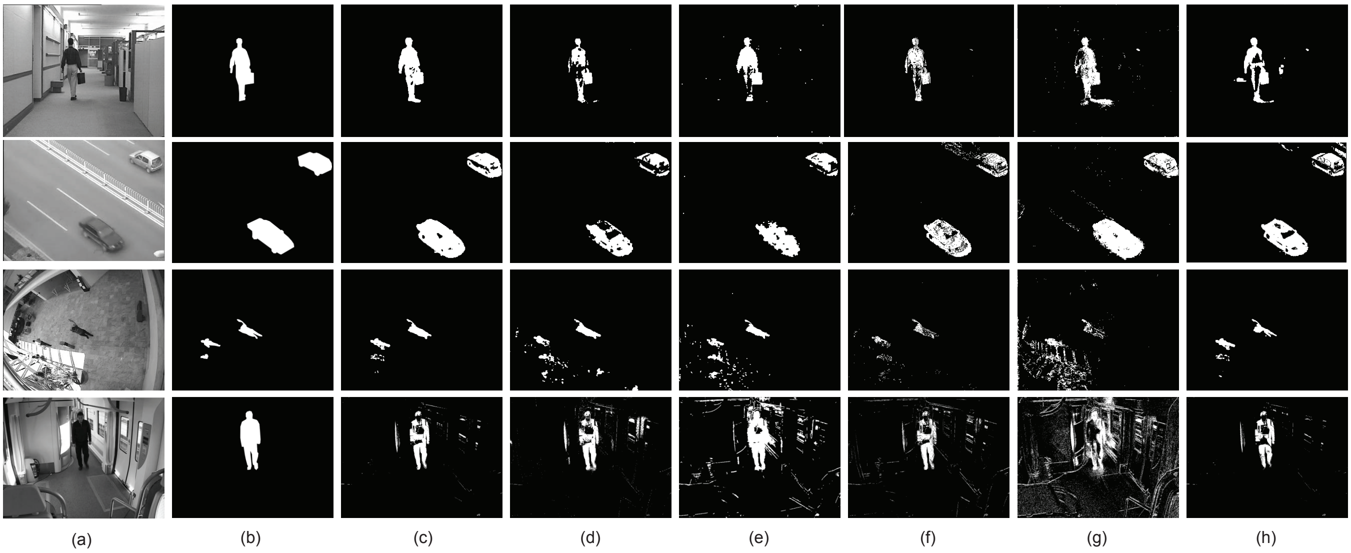

**Fig. 2**. Visual comparison on motion segmentation results: (a) origin frame, (b) ground truth, (c) ours, (d) ViBe [6], (e) SOBS [5], (f) FBM [14], (g) GMM [15], and (h) RPCA [7].

**Table 1**. Quantitative motion detection results on four datasets in terms of precision ($P$), recall ($R$), and F-measure ($F_1$).

| | ***Hallmonitor*** | | | ***Cars*** | | | ***Browser1*** | | | ***Cell_phone_theft*** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Recall | Precision | $F_1$ | Recall | Precision | $F_1$ | Recall | Precision | $F_1$ | Recall | Precision | $F_1$ |
| **Ours** | **0.94** | 0.94 | **0.94** | **0.87** | 0.81 | **0.84** | **0.85** | **0.88** | **0.86** | 0.76 | **0.93** | **0.84** |
| **ViBe[6]** | 0.61 | 0.97 | 0.75 | 0.58 | 0.85 | 0.69 | 0.49 | 0.72 | 0.58 | 0.69 | 0.64 | 0.67 |
| **SOBS[5]** | 0.77 | 0.93 | 0.84 | 0.66 | **0.91** | 0.77 | 0.76 | 0.71 | 0.62 | **0.90** | 0.46 | 0.61 |
| **FBM[14]** | 0.61 | **0.99** | 0.75 | 0.62 | 0.79 | 0.70 | 0.36 | 0.59 | 0.45 | 0.60 | 0.63 | 0.61 |
| **GMM[15]** | 0.74 | 0.78 | 0.76 | 0.86 | 0.76 | 0.81 | 0.52 | 0.21 | 0.30 | 0.57 | 0.15 | 0.24 |
| **RPCA[7]** | 0.80 | 0.90 | 0.84 | 0.71 | 0.81 | 0.76 | 0.72 | 0.87 | 0.79 | 0.70 | 0.85 | 0.74 |

where $tp$ is correctly classified foreground pixels, $fn$ is the number of foreground pixels incorrectly classified as background, $fp$ stands for the total number of background pixels incorrectly classified as foreground. For all the three metrics, higher values imply better motion detection accuracy. As shown in Table 1, though some values of the three metrics are a little lower than other methods, our method obtains the best results for most cases.

Fig. 2 presents visual comparison of foreground detection for one typical frame in each sequence. Our result is the closest to ground truth motion segmentation, and outperforms other algorithms. For example, for *Hallmonitor* and *Cars*, the foregrounds (man and cars) are completely detected while excluding extra noises in background. For *Browser1* and *Cell_phone_theft*, these datasets contain intense lighting variations and wide waggling by the train. It is observed that our method successfully picks up the intact foreground while other methods cannot precisely identify the foreground people with considerable amount of noise.

***Running Time***: We report running time for *Cars* with 40 frames of size $320 \times 240$. The ADM-ALM algorithm is implemented in MATLAB (R2013a) running on a desktop with a 3.4 GHz Core4 i7 processor and 8 GB memory. The motion estimation takes about 100 seconds on average to process 40 frames. The ALM-ADM algorithm to solve the MAMC model takes 2.53 seconds to extract the background.

## 4. CONCLUSION

In this paper, we propose a motion-assisted matrix completion model for background extraction from video sequences. Our method extracts promising backgrounds and provides excellent performance in distinguishing moving objects. Experimental results show that our method is quite versatile for surveillance videos with different types of motions and lighting conditions. In future work, we will exploit fast motion extraction strategies to accelerate our method for practical applications.

## 5. REFERENCES


[1] Xiaowei Zhou, Can Yang, and Weichuan Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE TPAMI*, vol. 35, no. 3, pp. 597–610, 2013.

[2] Daniel D Doyle, Alan L Jennings, and Jonathan T Black, "Optical flow background estimation for real-time pan/tilt camera object tracking," *Measurement*, vol. 48, pp. 195–207, 2014.

[3] Chris Stauffer and W Eric L Grimson, "Adaptive background mixture models for real-time tracking," in *IEEE CVPR*, 1999, vol. 2.

[4] Chris Stauffer and W. Eric L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE TPAMI*, vol. 22, no. 8, pp. 747–757, 2000.

[5] Lucia Maddalena and Alfredo Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE TIP*, vol. 17, no. 7, pp. 1168–1177, 2008.

[6] Olivier Barnich and Marc Van Droogenbroeck, "Vibe: A universal background subtraction algorithm for video sequences," *IEEE TIP*, vol. 20, no. 6, pp. 1709–1724, 2011.

[7] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright, "Robust principal component analysis?," *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 11, 2011.

[8] Charles Guyon, Thierry Bouwmans, and El-Hadi Zahzah, "Foreground detection based on low-rank and block-sparse matrix decomposition," in *Proc. IEEE ICIP*, 2012, pp. 1225–1228.

[9] Zhi Gao, Loong-Fah Cheong, and Mo Shan, "Block-sparse rpca for consistent foreground detection," in *Proc. ECCV*, pp. 690–703. Springer, 2012.

[10] Sebastian Brutzer, Benjamin Hoferlin, and Gunther Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *IEEE CVPR*, 2011, pp. 1937–1944.

[11] Christopher Richard Wren, Ali Azarbayejani, Trevor Darrell, and Alex Paul Pentland, "Pfinder: Real-time tracking of the human body," *IEEE TPAMI*, vol. 19, no. 7, pp. 780–785, 1997.

[12] Zhouchen Lin, Minming Chen, and Yi Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

[13] Junfeng Yang and Yin Zhang, "Alternating direction algorithms for $\ell_1$-problems in compressive sensing," *SIAM journal on scientific computing*, vol. 33, no. 1, pp. 250–278, 2011.

[14] Zhenjie Zhao, Thierry Bouwmans, Xuebo Zhang, and Yongchun Fang, "A fuzzy background modeling approach for motion detection in dynamic backgrounds," in *Multimedia and Signal Processing*, pp. 177–185. 2012.

[15] Zoran Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *IEEE ICPR*, 2004, vol. 2, pp. 28–31.