



Background subtraction for aerial surveillance conditions

Francisco Sanchez-Fernandez, Philippe Brunet, Sidi-Mohammed Senouci

► To cite this version:

Francisco Sanchez-Fernandez, Philippe Brunet, Sidi-Mohammed Senouci. Background subtraction for aerial surveillance conditions. 14th International Conference on Innovations for Community Services (I4CS), Jun 2014, Reims, France. pp.28-33, 10.1109/I4CS.2014.6860549 . hal-02875098

HAL Id: hal-02875098

<https://hal.science/hal-02875098>

Submitted on 18 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Background Subtraction for Aerial Surveillance Conditions

Francisco Sanchez-Fernandez, Philippe Brunet, Sidi-Mohammed Senouci

DRIVE Lab, University of Burgundy.
Nevers, France.

e-mail: {Francisco.Sanchez-Fernandez, Philippe.Brunet, Sidi-Mohammed.Senouci}@u-bourgogne.fr

Abstract—The first step in a surveillance system is to create a representation of the environment. Background subtraction is widely used algorithm to define a part of an image that most time remains stationary in a video. In surveillance tasks, this model helps to recognize those outlier objects in an area under monitoring. Set up a background model on moving platforms (intelligent cars, UAVs, etc.) is a challenging task due camera motion when images are acquired. In this paper, we propose a method to support instabilities caused by aerial images fusing spatial and temporal information about image motion. We used frame difference as first approximation, then age of pixels is estimated. This latter gives us an invariability level of a pixel over time. Gradient direction of ages and an adaptive weight are used to reduce impact from camera motion on background modelling. We tested our proposed method simulating several conditions that impair aerial image acquisition such as intentional and unintentional camera motion. Experimental results show improved performance compared to algorithms GMM and KDE.

Keywords—Image processing, unmanned aerial vehicle, background modelling, background subtraction, mobile observer, moving objects.

I. INTRODUCTION

Visual surveillance has become an important task in the last years. Its main objective is to find changes over time in a sequence of images. Motion-related changes are most important but this task is linked with other higher-level tasks such as localization, identification and tracking of moving objects. Therefore, detecting changes is considered as a key-preprocessing step of videos. Activities in which visual surveillance is used includes protection of strategic facilities, pedestrian detection, analysis of suspicious behavior, etc.

When we analyze a sequence of images used in surveillance, we can always observe two main features. The first feature refers to a part of the image that remains temporarily constant, i.e. does not suffer any change unless sensor by which image is acquired, it is moved intentionally or not. The second feature is about absolute majority formed by intensity of certain pixels, above characteristics describe an important part in image processing called background of an image.

Background modelling or subtraction (BS) is an important component in a video surveillance system because it helps to distinguish moving or incoming objects in a sequence of images. The main objective of BS algorithms is to classify all moving

objects as foreground and then identify interesting areas of an image for further analysis. The simplest case for BS is assuming a fixed camera [1][2][3][4], however, this assumption rigorously limits the implementation of other computer vision algorithms. Nowadays, a lot of imagery sequences are done by platforms incorporating a mobile camera, like cellular phones, robots and unmanned aerial vehicles (UAVs). Therefore, it exists an increasingly need to develop algorithms to separate foreground objects from background using moving cameras for different tasks.

In any kind of images, there are many changes that happen over time and affect the performance of background modelling. It is vital for background model to be capable to efficiently handle changes. Above can be achieved by being invariant or by adapting to changes. These can be local, small persisting moving objects belonging to images like waving trees [5][6], or global, like illumination changes or camera motion [7][8]. Spatial and temporal information of pixels are two fundamental elements to understand background structure and these should complement each other to obtain a robust background model [9].

In this paper, we propose a new background subtraction method designed to deal with particular conditions of aerial images. Our method is capable to deal with motion of UAV to extract background information. We use spatial and temporal information to classify pixels between foreground and background. We used an adaptive method focused to weight spatial or temporal information depending if motion of UAV is smooth or not. Experimental results of sequences under different conditions demonstrate robustness of the proposed method. Evaluation and comparison with existing methods show that our proposed method provides same results in motionless sequences and improved performance when sequences are unstable (camera motion).

The rest of this work is organized as follows: Section 2 presents a review of several methods for background modelling. Formulation and description of the proposed approach is presented in Section 3. Section 4 describes evaluation method and features of each dataset. The discussion of results and conclusions are made in Section 5.

II. RELATED WORK

A big number of BS methods have been proposed in recent years, trying to retrieve a background model using different approaches. Basically, classification is divided into two types of algorithms [10] parametric and non-parametric models. Parametric models represent a probability density function (pdf) parametrically. It uses a specific statistical distribution and its associated parameters, based on previous known distributions [11]. In a different way, a non-parametric model does not need to fit data to a specific distribution, the pdf is directly estimated and thus avoids calculating all parameters related to a specific distribution. Sections below provide some works for each class.

A. Parametric background models

Many BS algorithms for fixed cameras work by comparing color or intensities of pixels amongst an incoming image and a reference. The first approaches of BS used frame difference [12][13] between two consecutive images, where major differences in intensity from the reference image were related to moving objects. Thresholding is applied to obtain background model. However, the background obtained through this approach is not suitable in some sequences with abrupt changes of illumination or motion related.

The basic statistical method for background modelling consists in representing the color intensity at each pixel by a single Gaussian distribution. The Gaussian mixture model (GMM)[10] is probably the most popular statistical technique for background modelling. Stauffer and Grimson [14] proposed a broad view to the single Gaussian approach, by using an adaptive multiple Gaussians method that model the pixel color as a mixture of Gaussians[15][16][17]. The model is constructed as follows:

$$P(X_t) = \sum_{i=1}^K w_{i,t} * \eta(X_t, \mu_{i,t}, \sigma_{i,t}) \quad (1)$$

The intensity of a pixel is modeled by a mixture of K Gaussian distributions (Typically, K is around 3 to 5). Every new pixel is checked against existing K Gaussians (η), if difference between pixel and Gaussian is less than 2.5σ , the weight for that distribution is updated. Several modifications to original model has been proposed to GMM model, e.g. [18][19][20] etc.

B. Non parametric background models

The main reference for non-parametric models is kernel density estimation (KDE) and proposed by [1][21]. Given a data sample $S = \{x_i\}_{i=1...n}$ with a specific pdf $p(x)$, an approximation $\hat{p}(x)$ is found with (2).

$$\hat{p}(x) = \frac{1}{n} \sum_{i=1}^n K_{\sigma}(x - x_i) \quad (2)$$

Where K_{σ} is a kernel function, centered at data points in sample space and uniformly weighted. Kernel density estimators have main property of asymptotically converge to any probability density function, this property is used to model background.

The kernel can be chosen from a wide range but typically a Gaussian kernel [22][23] is used for its continuity, differentiability, and locale properties. Differentiate between choosing the Gaussian as a kernel function and fitting the distribution to a Gaussian model is a critical part to understand non parametric models [21], KDE uses Gaussian only as a function to weight the data points and does not makes assumptions about shape of density function. In conclusion, a pixel is considered a background pixel if $\hat{p}(x) > \varepsilon$, where ε is a global threshold that is equal to whole image and can be adjusted to obtain a desired proportion of false positives [24][25].

The key motivation for using non parametric density estimation for background modeling lies on robustness presented at outdoor scenarios. This kind of scenarios always contains various disturbances [26] such as ocean waves, waving trees, rain, moving clouds, etc. Modeling such persistent motion requires a more flexible representation of the background probability distribution at each pixel. Since non parametric models are not based on a specific distribution, these can be updated according to characteristics of each image.

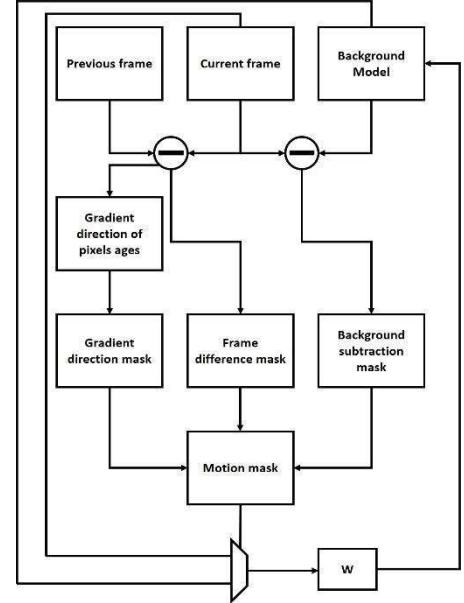


Figure 1. Framework of the proposed method.

III. ADAPTIVE BACKGROUND SUBTRACTION SOLUTION DESCRIPTION

Typical background subtraction techniques are inefficient on images taken from a mobile observer due to low temporal consistency of most pixels. We propose a combination of temporal and spatial information. The workflow of the proposed method is depicted in Fig. 1. The first part of our algorithm consists to do a basic approximation of background model using an adaptive method proposed by [27]. However, this model still includes outliers, which will be eliminated with further processing.

In a motionless sequence, the frame difference represents the easiest way to find all objects that move and are not part of the background model. So, we first identify all pixels that are not the same between two consecutive images, i.e. estimate the absolute difference between a past frame and an n th frame and then threshold to obtain a frame difference mask as follows:

$$M_{fd}(i, j) = \begin{cases} 1, & \text{if } |f_n(i, j) - f_{n-1}(i, j)| > \varepsilon_d \\ 0, & \text{else} \end{cases} \quad (3)$$

Frame difference mask is a binary image, where pixels that correspond to moving objects have the value 1, and remaining pixels are set to 0.

To prevent temporal variation on pixels generated by camera motion, the age of each pixel is used. Age refers to how many frames a pixel remains constant, we can infer that a pixel with a greater age has high probability to belong to the background model. Proposed method establish an upper limit of pixel age to eliminate excessive data storage. When a pixel reaches this limit, its age remains constant. Matrix containing the age of pixels is calculated according to (4) for each pair of consecutive images.

$$A_p(i, j) = \begin{cases} A_p(i, j) + 1, & \text{if } |f_n(i, j) - f_{n-1}(i, j)| < \varepsilon_d \\ 0, & \text{else} \end{cases} \quad (4)$$

Then, we estimate the gradient direction of A_p as follows:

$$\theta = \text{atan2}\left(\frac{\partial A_p}{\partial y}, \frac{\partial A_p}{\partial x}\right) \quad (5)$$

The gradient direction provides reliable information about moving objects. First, the areas of image with the maximum possible age have same direction because each point in this area represents a local maximum (its direction is equal to zero). The gradient directions of moving pixels due camera motion, also have same direction because all pixels move in the same way and represent a significant quantity of the distribution of directions. Therefore, gradient directions of moving objects are minority because each one has its own direction. Thresholding sorted histogram of gradient directions, we can separate pixels with

predominant gradient directions, which belong to the background. Gradient direction mask is computed as follows:

$$M_{gd}(i, j) = \begin{cases} 1, & \text{if } \theta(i, j) > \varepsilon_g \\ 0, & \text{else} \end{cases} \quad (6)$$

The background image is subtracted from the current image to obtain another representation of moving objects in the sequence. Background subtraction mask is formed as shown in (7).

$$M_{bs}(i, j) = \begin{cases} 1, & \text{if } |f_n(i, j) - B_{n-1}(i, j)| > \varepsilon_b \\ 0, & \text{else} \end{cases} \quad (7)$$

Finally, a function to decide which image to sample is required to update background model. Frame difference mask and background subtraction mask can perfectly describe background model when camera is static, otherwise, model still has outliers. Hence we add gradient direction mask to strengthen our sampling function shown in (8). We apply OR function because gradient direction mask also contains common areas with the other masks, especially those motionless.

$$M_m = (M_{fd} \& M_{bs}) | M_{gd} \quad (8)$$

The thresholds used in (3), (6) and (7) are calculated according to thresholding method in [28], it is used to split a histogram into two classes and minimize variance of data at each class.

To update the background model, we need to calculate an instantaneous background using the motion mask. A position where motion mask is 1, indicates that this pixel is part of a moving object then we sample last version of background model. Conversely, when motion mask is 0, current frame is sampled to keep model correctly updated. Updating rules above mentioned are shown in (9).

$$I_b = \begin{cases} f_n, & \text{if } M_m = 0 \\ B_{n-1}, & \text{if } M_m = 1 \end{cases} \quad (9)$$

The background model is a weighted average between instantaneous background and previous version of background model as follows:

$$B_n = \alpha I_b + (1 - \alpha) B_{n-1} \quad (10)$$

Weight used in (10) is estimated using an adaptive approach. If an image has high rate of static pixels, it means that we can prioritize instantaneous background and build model from it. Otherwise, the priority is the model that was built with previous versions. We should find the number of static pixels using the gradient direction mask, this is illustrated in (11) and (12).

$$s_p = \sum_{\substack{0 < i \leq m \\ 0 < j \leq n}} M_{gd}(i, j) \quad (11)$$

$$P_{sp} = \frac{s_p}{m \times n} \quad (12)$$

The parameter α is set in a smooth manner to mitigate e.g. influence of an abrupt camera motion or light changes. Weight update is done as follows:

$$\alpha_n = 0.9\alpha_{n-1} + 0.1P_{sp} \quad (13)$$

IV. EXPERIMENTAL RESULTS

We have tested the proposed method on several scenarios involving similar conditions in aerial images acquisition, including unintentional and intentional motion of camera. The proposed method is implemented using MATLAB software on an Intel Core i5 PC with 8 GB RAM.

The evaluation is done using dataset provided by [29], this dataset provides a realistic and diverse set of videos, including a wide range of detection challenges. We chose three categories to evaluate our algorithm since these categories represent various conditions that affect performance of the background subtraction algorithms in the aerial images. Baseline category has videos with subtle background motion, which can represent when a static UAV is acquiring images. In this category, background subtraction is easy to apply but not trivial, its main purpose is to serve as a reference. Camera jitter category is formed by sequences taken from a sensor affected by unintentional motion like climatological conditions and vibrations of vehicle. Motionless images are important in BS algorithms to avoid detection of fixed objects caused by unintentional motion of the camera. Lastly, PTZ category contains videos captured by pan-tilt-zoom cameras, which are commonly used for surveillance tasks. Images acquired from a PTZ camera can be represented as intentional motion of UAV, these sequences are challenging because motion must be removed to reduce misclassification of pixels.

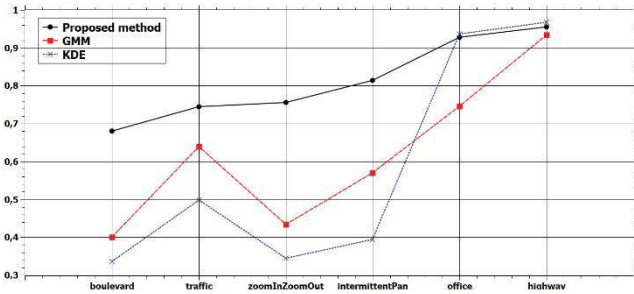


Figure 2. Precision of our proposed method, GMM and KDE for all sequences.

We compare our method against two commonly used methods, the Gaussian mixture model (GMM)[14] and kernel density estimation(KDE)[1]. To evaluate performance of methods, we compute precision of each method as follows:

$$Precision = \frac{True\ positives}{True\ positives + False\ Positives} \quad (14)$$

Where:

True positives: number of pixels that are correctly detected as background;

False positives: number of pixels that are wrongly detected as background.

To obtain true and false positives, results of each method are compared to ground truth, i.e. correct classification result. The ground truth is also provided with each sequence in the dataset. We test all methods in 6 different sequences, 2 sequences for each category aforementioned. The precision of each sequence are showed in Fig. 2.

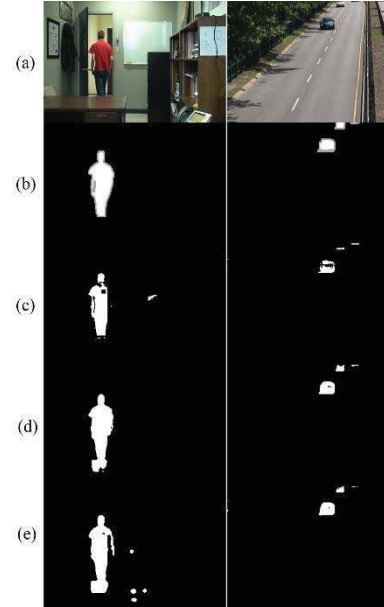


Figure 3. Results of baseline sequences “office” and “highway”: (a) original image, (b) ground truth, (c) proposed method, (d) GMM, (e) KDE.

The sequences “office” and “highway” are part of baseline category, which represent the simplest case for BS. In these sequences we obtained highest results. All methods show similar

performances due to their constant background with subtle motion; results of this category are shown in Fig. 3. The sequences “zoomInZoomOut” and “intermittentPan” show low temporal consistency between consecutive images because the pictures are taken by a PTZ camera, which is constantly panning and zooming. However, proposed method outperforms the GMM and KDE significantly in all two sequences, BS results are given in Fig. 4. Finally, in Fig. 5, sequences “boulevard” and “traffic” affected with camera jitter, show similar precision than previous sequences.

V. CONCLUSIONS

This work presents a novel method for background modelling with an adaptive update algorithm. Three types of information are fused to obtain a robust background model. The proposed method models variations according to age of pixels, i.e. we used those pixels with low variability over time. We tested our proposed method in different sequences simulating aerial imagery conditions and from experimental results better performance than GMM and KDE algorithms in changing images is observed. Shadow removal can be considered as future work to increase performance of proposed algorithm, since in current evaluation, pixels that are part of shadows are considered false positives and this affects performance evaluation.

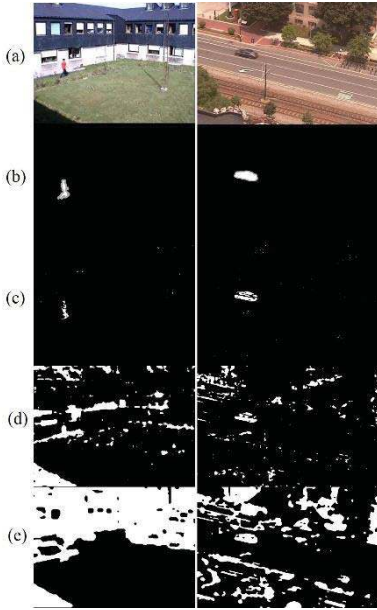


Figure 4. Results of moving sequences “zoomInZoomOut” and “intermittentPan” (a) original image, (b) ground truth, (c) proposed method, (d) GMM, (e) KDE.

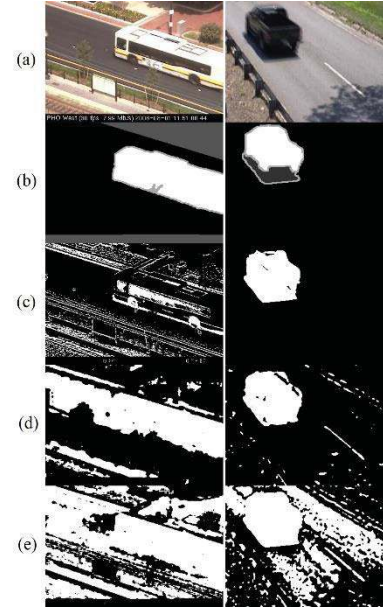


Figure 5. Results of unstable sequences “Boulevard” and “traffic”: (a) original image, (b) ground truth, (c) proposed method, (d) GMM, (e) KDE.

ACKNOWLEDGMENT

The authors would like to thank University of Burgundy for support this research. We also are very grateful to chagedetection.net [29] for providing dataset used to evaluate this work.

REFERENCES

- [1] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, “Background and foreground modeling using nonparametric kernel density estimation for visual surveillance,” *Proc. IEEE*, vol. 90, no. 7, pp. 1151–1163, Jul. 2002.
- [2] O. Javed, K. Shafique, and M. Shah, “A hierarchical approach to robust background subtraction using color and gradient information,” in *Workshop on Motion and Video Computing, 2002. Proceedings.*, 2002, pp. 22–27.
- [3] W. Long and Y.-H. Yang, “Stationary background generation: An alternative to the difference of two images,” *Pattern Recognit.*, vol. 23, no. 12, pp. 1351–1359, Jan. 1990.
- [4] A. Neri, S. Colonnese, G. Russo, and P. Talone, “Automatic moving object and background separation,” *Signal Processing*, vol. 66, no. 2, pp. 219–232, Apr. 1998.
- [5] A. A. Bhutta, I. N. Junejo, and H. Foroosh, “Selective subtraction when the scene cannot be learned,” in *2011 18th IEEE International Conference on Image Processing*, 2011, pp. 3273–3276.
- [6] B. Zhong, S. Liu, H. Yao, and B. Zhang, “Multi-resolution background subtraction for dynamic scenes,” in *2009 16th IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 3193–3196.
- [7] B. Yang, Y. Guo, Y. Ming, and A. Men, “An Effective Background Subtraction under a Continuously and Rapidly Varying Illumination,” in

- 2010 *Second International Conference on Future Networks*, 2010, pp. 16–19.
- [8] F.-C. Cheng, S.-C. Huang, and S.-J. Ruan, “Illumination-Sensitive Background Modeling Approach for Accurate Moving Object Detection,” *IEEE Trans. Broadcast.*, vol. 57, no. 4, pp. 794–801, Dec. 2011.
- [9] Y. Zhao, H. Gong, Y. Jia, and S.-C. Zhu, “Background modeling by subspace learning on spatio-temporal patches,” *Pattern Recognit. Lett.*, vol. 33, no. 9, pp. 1134–1147, Jul. 2012.
- [10] M. Hedayati, W. M. D. W. Zaki, and A. Hussain, “Real-time background subtraction for video surveillance: From research to reality,” in *2010 6th International Colloquium on Signal Processing & its Applications*, 2010, pp. 1–6.
- [11] H. T. Ngo, R. W. Ives, R. N. Rakvic, and R. P. Broussard, “Real-time video surveillance on an embedded, programmable platform,” *Microprocess. Microsyst.*, vol. 37, no. 6–7, pp. 562–571, Aug. 2013.
- [12] R. Jain, D. Militzer, and H. Nagel, “Separating non-stationary from stationary scene components in a sequence of real world TV-images,” in *Proceedings of the 5th international joint conference on Artificial intelligence*, 1977, pp. 612–618.
- [13] R. Jain and H.-H. Nagel, “On the Analysis of Accumulative Difference Pictures from Image Sequences of Real World Scenes,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 206–214, Apr. 1979.
- [14] C. Stauffer and W. E. L. Grimson, “Adaptive background mixture models for real-time tracking,” in *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, pp. 246–252.
- [15] E. Hayman and J. Eklundh, “Statistical background subtraction for a mobile observer,” in *Proceedings Ninth IEEE International Conference on Computer Vision*, 2003, pp. 67–74 vol.1.
- [16] N. I. Rao, “Background modeling from a free-moving camera by Multi-Layer Homography Algorithm,” in *2008 15th IEEE International Conference on Image Processing*, 2008, pp. 1572–1575.
- [17] A. Shimada, D. Arita, and R. Taniguchi, “Dynamic Control of Adaptive Mixture-of-Gaussians Background Model,” in *2006 IEEE International Conference on Video and Signal Based Surveillance*, 2006, pp. 5–5.
- [18] M. Alvar, A. Rodriguez-Calvo, A. Sanchez-Miralles, and A. Arranz, “Mixture of Merged Gaussian Algorithm using RTDENN,” *Mach. Vis. Appl.*, Oct. 2013.
- [19] C.-C. Chiu, M.-Y. Ku, and L.-W. Liang, “A Robust Object Segmentation System Using a Probability-Based Background Extraction Algorithm,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 4, pp. 518–528, Apr. 2010.
- [20] A. Bayona, J. C. SanMiguel, and J. M. Martinez, “Stationary foreground detection using background subtraction and temporal difference in video surveillance,” in *2010 IEEE International Conference on Image Processing*, 2010, pp. 4657–4660.
- [21] A. Elgammal, D. Harwood, and L. Davis, “Non-parametric model for background subtraction,” *Comput. Vision—ECCV 2000*, 2000.
- [22] T. Parag, A. Elgammal, and A. Mittal, “A Framework for Feature Selection for Background Subtraction,” in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR’06)*, 2006, vol. 2, pp. 1916–1923.
- [23] K. a Patwardhan, G. Sapiro, and V. Morellas, “Robust foreground detection in video using pixel layers,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 4, pp. 746–51, Apr. 2008.
- [24] S. Mubarak, “Bayesian modeling of dynamic scenes for object detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1778–1792, 2005.
- [25] Y. Sheikh, O. Javed, and T. Kanade, “Background Subtraction for Freely Moving Cameras,” in *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 1219–1225.
- [26] A. Mittal and N. Paragios, “Motion-based background subtraction using adaptive kernel density estimation,” in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, 2004, vol. 2, pp. 302–309.
- [27] W. Zhang, Q. M. J. Wu, and H. B. Yin, “Moving vehicles detection based on adaptive motion histogram,” *Digit. Signal Process.*, vol. 20, no. 3, pp. 793–805, May 2010.
- [28] N. Otsu, “A Threshold Selection Method from Gray-Level Histograms,” *IEEE Trans. Syst. Man. Cybern.*, vol. 9, no. 1, pp. 62–66, 1979.
- [29] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, “Changetection.net: A new change detection benchmark dataset,” in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2012, pp. 1–8.