ADM-HIPaR: An Efficient Background Subtraction Approach

Thien Huynh-The, Sungyoung Lee, Cam-Hao Hua
Department of Computer Science and Engineering
Kyung Hee University
1732 Deokyoungdae-ro, Giheung-gu, Yongin-si, Gyeonggi-do, 446-701, South Korea.
thienht@oslab.khu.ac.kr, sylee@oslab.khu.ac.kr, hao.hua@oslab.khu.ac.kr

Abstract

This paper presents a novel background subtraction method that is flexible for various background scenarios. The method includes automated-directional masking (ADM) algorithm for adaptive background modeling and historical intensity pattern reference (HIPaR) algorithm for foreground segmentation. By selecting an appropriate mask in a set based on directional feature, ADM updates background smoothly and precisely following a boundary-based strategy with an intensity correction rule. In order to segment foreground, HIPaR refers intensity patterns of previous backgrounds and input frames and then compares their mean difference with a checking threshold to make foreground decision. Experimental results prove that our proposed ADM-HIPaR outperforms other state-of-the-art methods in terms of foreground detection accuracy.

1. Introduction

Foreground detection is the prior stage for numerous multimedia applications of video surveillance such as traffic monitoring and human-computer interaction. Many background subtraction based approaches have been studied to reach performance in terms of accuracy and real-time processing, however, some of which are usually fragile under hash conditions of artifacts, shadow, dynamic background, camera jitter, etc due to poorly estimated background [2, 3, 18]. Background estimation is therefore realized as an important component that directly affects to the accuracy of foreground detection at first and overall system so far.

As basic models, luminance- and histogram-statistical approaches have been presented to determine whether a pixel belongs to foreground or background. Some advanced techniques such as Gaussian Mixture Model (GMM) [19, 21], Kernel Density Estimation (KDE) [5], and Principle Component Analysis (PCA) [16] have been proposed to accurately detect foreground for multi-modal background. However, GMM takes a drawback in estimating model parameters and KDE cannot handle concurrently moving objects at various velocities. A hybrid model [13] combining GMM and KDE is developed to control the above problem. To eliminate the parameter tuning task, some non-parametric background models have been studied. Codebook technique [11] allows to quantize background pixels to codewords for building a compressed background model in a number of frames. Requirements of high memory consuming and expensive frame resource are major shortcomings of codebook-based approaches [10, 12].

Artificial intelligence including fuzzy-based and neural-network techniques such as Fuzzy Mixture of Gaussian [4] and Self-Organizing Neural Network [14, 6] have been studied as well for background modeling. They build a flexible probabilistic background model by an unsupervised learning to detect moving objects using block-similarity matching. Recently, pixel-wise advanced background estimation algorithms classify pixels by comparing its intensity with randomly neighbouring pixels in Visual Background Extraction (ViBE) [1] and shaped neighbouring pixels in Neighbor-Intensity Correction (NIC) [9]. Using a fixed-squared mask for shaping neighbouring pixels is not robust and thus suppresses the multi-modal background estimation performance.

This work presents a novel background subtraction method, namely ADM-HIPaR, which is a combination of two proposed algorithms, one is automated-directional masking and another is historical intensity pattern reference. Based on directional feature calculated for each edge pixel belonging to motion boundary, a particular mask in a predefined set is selected to capture neighbouring-intensity patterns of two adjacent frames. According to intensity pattern comparison, background update operation is done by an efficient correction rule which is improved from NIC [9]. As a result, the background is estimated more accurate if compared to existing works. In the background subtraction stage, we retrieve the squared intensity patterns
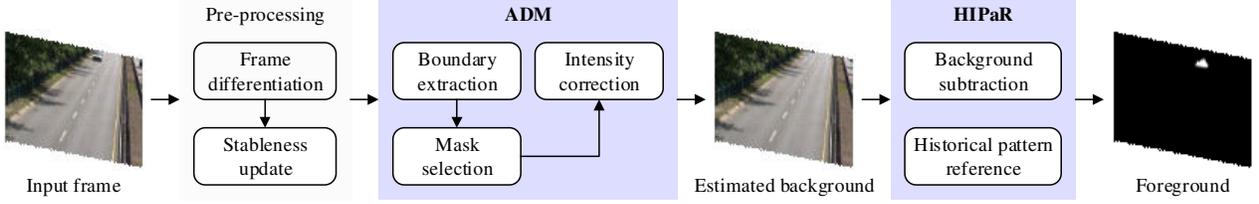
Figure 1. The workflow of our proposed method with Automated -Directional Masking (ADM) for background estimation and Historical Intensity Pattern Reference (HIPaR) for foreground detection.
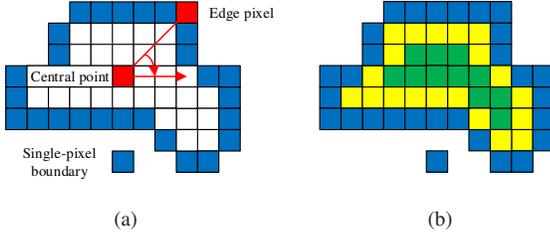
Figure 2. The illustrations of: (a) directional feature calculation for edge pixels denoted by blue color, (b) boundary-based strategy processing from blue to yellow and green.

which are extracted from backgrounds and input frames in the spatio-temporal dimension, to execute foreground decision by comparing the pattern difference with a checking threshold. Experimental results demonstrate the superiority of our proposed ADM-HIPaR over recent background subtraction approaches on the CDnet dataset.

2. Methodology

This section describes our contribution of two proposed algorithms that are automated-directional masking for adaptive background estimation and historical intensity pattern reference for accurate foreground detection as presented in Figure 1.

2.1. Automated-Directional Masking For Background Estimation

Following NIC algorithm for background estimation, an initial background is updated by analyzing neighbouring-intensity patterns surrounding each motion pixel. Two patterns, independently captured from a background and an input frame using a binary mask, are compared through standard deviations to perform an intensity correction rule. Since using a fixed-squared mask for various background models is analyzed as the major limitation of [9], the main idea of this work is to automatically select an appropriate binary mask based on directional feature of boundary edge. This leads to significant accuracy enhancement for foreground detection.

Given a set of motion-differencing pixels $\mathcal{M}$, extracted from a difference image $D(t) = |F(t) - F(t-1)|$, which needs to be modified to retrieve truth background intensity, we at first trace boundaries as illustrated in Figure 2(a). Compared to [9] where intensity adjustment is done for pixels in coordinate ordering, the proposed boundary-based strategy as shown in Figure 2(b) smoothly produces a high-quality background. For each boundary B, we calculate directional feature as an angle between each edge pixel $p \in B$ and a central point c as

$$\theta(p, c) = \tan^{-1}\left(\frac{y_p - y_c}{x_p - x_c}\right) \qquad (1)$$

where (x_p, y_p) and (x_c, y_c) are locations of p and c respectively. We quantize $\theta(p, c)$ to four classes constructed by eight directional sectors, denoted $S_{1\to8}$. Sector definition is graphically illustrated in Figure 3(a) and expressed as

$$
\begin{array}{ll}
S_1: & \theta(p, c) < 22.5^0 \,|\, 337.5^0 \leq \theta(p, c) < 360^0 \\
S_2: & 22.5^0 \leq \theta(p, c) < 67.5^0 \\
S_3: & 67.5^0 \leq \theta(p, c) < 112.5^0 \\
S_4: & 112.5^0 \leq \theta(p, c) < 157.5^0 \\
S_5: & 157.5^0 \leq \theta(p, c) < 202.5^0 \\
S_6: & 202.5^0 \leq \theta(p, c) < 247.5^0 \\
S_7: & 247.5^0 \leq \theta(p, c) < 292.5^0 \\
S_8: & 292.5^0 \leq \theta(p, c) < 337.5^0
\end{array}
\qquad (2)
$$

Based on a sector which $\theta(p, c)$ should belong to, a corresponding mask $H(p)$ is particularly identified for each edge pixel p:

$$
H(p) = \begin{cases}
H_{15} & \text{if} \quad \theta(p, c) \in \{S_1, S_5\} \\
H_{26} & \text{if} \quad \theta(p, c) \in \{S_2, S_6\} \\
H_{37} & \text{if} \quad \theta(p, c) \in \{S_3, S_7\} \\
H_{48} & \text{if} \quad \theta(p, c) \in \{S_4, S_8\}
\end{cases}
\qquad (3)
$$

There are totally five masks which are horizontal H_{15}, vertical H_{37}, bottom-left to top-right diagonal H_{48}, top-left to bottom-right diagonal H_{26}, and circle H_c as shown in Figure 3. It is noted that the circle one in Figure 3(a), is used for the special case of single-pixel boundaries in Figure 2.

Our proposed ADM algorithm plays the most important role in the background estimation workflow. We assume
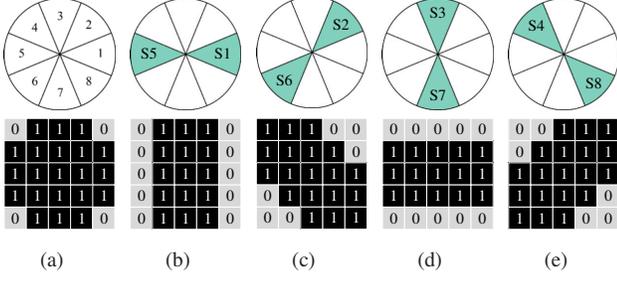
Figure 3. The illustrations of: (a) sector partition (top) and circle mask H_c (bottom) for single-pixel boundaries, (b)-(e) four binary masks corresponding to eight directional sectors including H_{15} for $\theta \in \{S_1, S_5\}$, H_{26} for $\theta \in \{S_2, S_6\}$, H_{37} for $\theta \in \{S_3, S_7\}$, and H_{48} for $\theta \in \{S_4, S_8\}$, respectively.

the first frame in a video sequence is the initial background $J(1) = F(t=1)$. Since background motion and moving objects may be available in background J beforehand, they need to be discarded to obtain a background which is as *pure* possible. A difference image $D(t)$ is extracted at t^{th} frame ($\forall t \geq 2$). By global thresholding, a motion-differencing set $\mathcal{M}$ is defined as

$$\mathcal{M}(t) = \{(x,y) \,|\, D(t) \geq \tau\} \quad (4)$$

where τ is the constant threshold to identify the difference between two adjacent frames. To exclude outliers, pixel stableness is measured for all of pixels by monitoring intensity changes between two frames as

$$\psi(t) = \begin{cases} \psi(t-1)+1 & \forall (x,y)\,|\,D(t) < \tau \\ \psi(t-1)-1 & \forall (x,y)\,|\,D(t) \geq \tau \end{cases} \quad (5)$$

where $\psi(1) = 0$ for initialization. This coefficient of a pixel will be small if its intensity is successively changed and vice versa. To filter pixels for applying intensity correction, we intersect two condition as follows

$$(x,y)\,|\,\mathcal{M}(t) \cap (\psi(t) < 0) \quad (6)$$

Eq. (6) refers that only pixels in $\mathcal{M}(t)$ with negative stableness are requested to update intensity following boundary-based strategy. Our proposed ADM, consisted of directional feature calculation by Eq. (1), sector mapping by Eq. (2), and mask selection by Eq. (3), is then applied to each edge pixel to choose an appropriate mask. By doing multiplication operation between H and neighbouring pixels, two patterns, denoted $\mathcal{F}(t)$ and $\mathcal{F}(t-1)$, are extracted from $F(t)$ and $F(t-1)$ respectively. Due to the difference between two intensity patterns, particularly the ratio between a number of motion and non-motion pixels, we compare standard deviation which is calculated as

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}|g(i)-\mu|^2} \quad (7)$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n}g(i) \quad (8)$$

where g is pixel intensity and n is the number of pixels in mask. Value of a motion-differencing pixel is updated in J to a new intensity of corresponding location in either $F(t)$ or $F(t-1)$ by a correction rule

$$J(t) = \begin{cases} F(t) & \text{if} \quad \sigma_{\mathcal{F}(t-1)} \geq \sigma_{\mathcal{F}(t)} \\ F(t-1) & \text{if} \quad \sigma_{\mathcal{F}(t-1)} < \sigma_{\mathcal{F}(t)} \end{cases} \quad (9)$$

where $\sigma_{\mathcal{F}(t-1)}$ and $\sigma_{\mathcal{F}(t)}$ are the standard deviations calculated from $\mathcal{F}(t-1)$ and $\mathcal{F}(t)$ respectively. Compared to [9] where a pixel is updated from background intensity to current frame intensity only if $\sigma_{\mathcal{J}(t-1)} < \sigma_{\mathcal{F}(t)}$, where $\sigma_{\mathcal{J}(t-1)}$ is the standard deviation of background pattern $\mathcal{J}(t-1)$, the event of object motionlessness for a while is able to handle with the new rule.

2.2. Historical Intensity Pattern Reference For Foreground Detection

By subtraction operation, we extract frame differencing between the current frame $F(t)$ and the updated background $J(t)$ by $\mathcal{D}(t) = |F(t) - J(t)|$. Let $\mathcal{N}$ be a set of candidate pixels for foreground classification

$$\mathcal{N}(t) = \{(x,y)\,|\,\mathcal{D}(t) \geq \delta\} \quad (10)$$

where δ is the intensity threshold to filter foreground pixels. To verify whether a pixel belongs to foreground, we propose historical intensity pattern reference algorithm (HIPaR) that allows to segment foreground more accurately. We retrieve squared intensity patterns of the estimated backgrounds $\{J(t-1), J(t)\}$ and the input frames $\{F(t-1), F(t)\}$, denoted $\mathcal{J}(t-1,t)$ and $\mathcal{F}(t-1,t)$ respectively, surrounding a pixel $(x,y) \in \mathcal{N}(t)$. Referring intensity of neighbouring pixels aims to handle frequent background motions as well. By an comparison operator, foreground decision for each pixel is adopted as follows

$$(x,y) = \begin{cases} 1 & \text{if} \quad \Delta_\mu \geq \gamma \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

where γ is the threshold that aims to check the pattern difference. $\Delta_\mu = |\mu_{\mathcal{J}(t-1,t)} - \mu_{\mathcal{F}(t-1,t)}|$ is the mean difference in which $\mu_{\mathcal{J}}$ and $\mu_{\mathcal{F}}$ can be calculated using Eq. (8). As a post-processing step, opening and closing morphological operations are utilized to remove noise and fill holes.

3. Experimental results

In this section, we evaluate and analyze foreground detection performance of the proposed method with various parameter configurations on several testing video sequences. A method comparison in terms of accuracy is further presented to prove the efficiency of ADM-HIPaR against to other state-of-the-art methods.
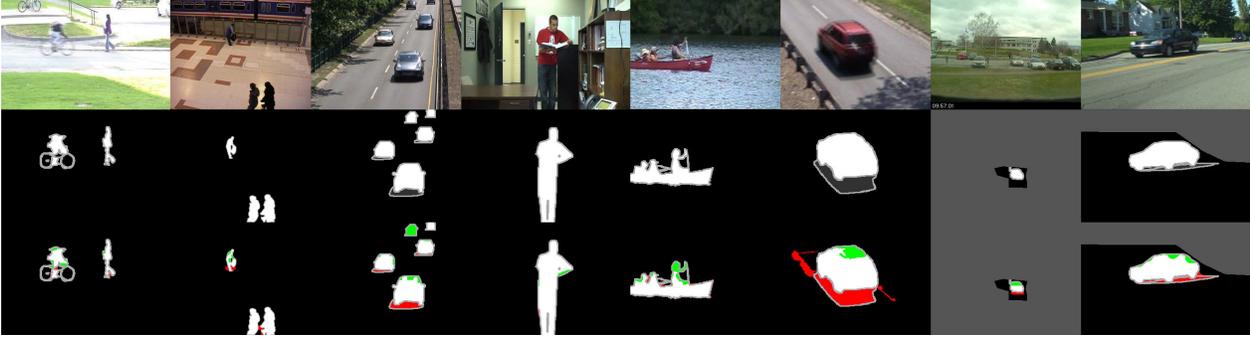
Figure 4. Visual results of foreground detection with true positive (TP) pixels in white, true negative (TN) pixels in black, false positive (FP) pixels in red, and false negative (FN) pixels in green on testing videos (Top to bottom: input frame, ground truth, and detected foreground. Left to right: pedestrians, PETS2006, highway, office, canoe, traffic, parking, and bungalows).

3.1. Dataset

We evaluate our proposed ADM-HIPaR model on several video sequences in the ChangeDetection dataset [20]. Totally eight videos are chosen from five categories: *highway*, *office*, *pedestrians*, and *PETS2006* from Baseline, *canoe* from Dynamic Background, *traffic* from Camera Jitter, *parking* from Intermittent Object Motion, and *bungalows* from Shadow. Some samples in Baseline contain subtle background motion, isolated shadows, abandoned objects, and pedestrians (stop for a short while and then go away). *Canoe* presents a scene with dynamic background motion, e.g., leaf and water motion. *Traffic* is captured in outdoor by a strongly vibrating camera. For much more challenging, *parking* and *bungalows* samples contain background objects moving away and hard shadow objects with intermittent shades.

3.2. Metric

Foreground detection performance is quantitatively benchmarked using Recall (Re), Specificity (Sp), False Positive Rate (FPR), False Negative Rate (FNR), Percentage of Wrong Classifications (PWC), F-Measure (F1), and Precision (Pre) [7]. Larger Re, Sp, F1, and Pre indicate superior detection performance while remains should converge to zero of wrong classification.

3.3. Evaluation

The default parameters for all testing videos is configured as follows: $\tau = 15$, $\delta = 30$, and $\gamma = 25$. Based on the quantitative results reported in Table 1, we reach the high foreground detection accuracy for the most of video sequences, except *traffic* and *parking*. Concerning the baseline category, the average F-measure is observed at 0.9191. Because of the strong vibration in *traffic*, estimated backgrounds are unstable leading to poor quality of segmented foreground. Containing more background objects causes

(a)

(b)

(c)
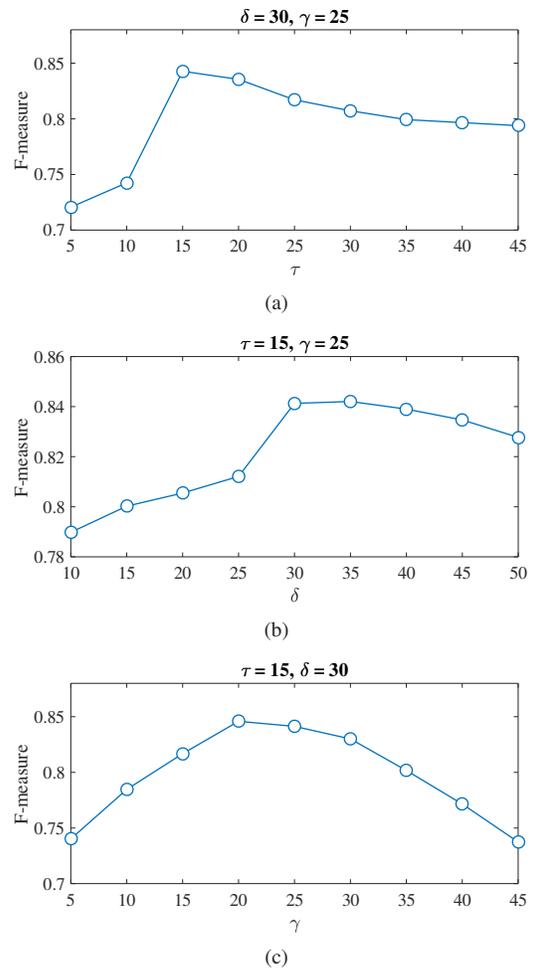
Figure 5. Average F-measures of ADM-HIPaR for all testing videos when varying parameters (a) the difference threshold τ in Eq. (4), (b) the intensity threshold δ in Eq. (10), and (c) the pattern checking threshold γ in Eq. (11).
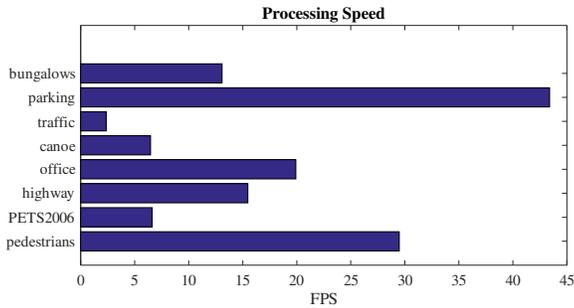
Figure 6. Average processing speeds of all video sequences expressed in FPS (Frames Per Second).

foreground misunderstanding in *parking*. The visual results of foreground detection are presented in Fig. 4 where binary classification results are highlighted in different colors. We also evaluate the proposed method under various parameter configurations. According to the average F-measure scores plotted in Fig. 5, we recommend to use $15 \leq \tau \leq 20$, $30 \leq \delta \leq 40$, and $20 \leq \gamma \leq 25$ to obtain high accuracy in background estimation as well as foreground detection.

In the next experiment, we compare our proposed approach with two baseline algorithms as Kernel Density Estimation (KDE) [5], Gaussian Mixture Model (GMM) [21] and several state-of-the-art approaches such as Spatial Coherent Self-Organizing Background Subtraction (SC-SOBS) [14], Pixel-Based Adaptive Segmenter (PBAS) [8], Spectral-360 [17], Graph Cut algorithm (GraphCutDiff) [15], Neighbor-based Intensity Correction (NIC) [9] on eight video samples. Based on the quantitative results reported in Table 2, it can be seen that our ADM-HIPaR mostly outperforms other methods, especially with Recall, False Negative Rate, and F-measure metrics. For the remaining metrics, our results have insignificant deviation of performance with those of the best score methods. Compared to NIC using a fixed-squared mask, ADM-HIPaR impressively improves 5.43%, 8.76%, and 7.88% of Recall, Precision, and F-measure respectively.

In order to analyze the computational cost of the proposed method, we use a profiling tool in Matlab 2014b to measure the time required for background estimation and foreground detection. The experiment is performed on a notebook operating Windows 10 with a 2.70 GHz i7-5700HQ processor and 16-GB RAM. According to the results shown in Fig. 6, *traffic* and *canoe* samples require more time for background estimation because of vibration and dynamic background motion.

4. Conclusions

In this paper, we propose an efficient background subtraction method, namely ADM-HIPaR, that greatly im-

proves foreground detection accuracy on several challenging video sequences. A background is flexibly estimated by our proposed automated-directional masking algorithm and an improved correction rule. To segment foreground, we introduce historical intensity pattern reference algorithm where the foreground decision is adopted by comparing the pattern difference of background and input frame with a checking threshold. We benchmark our proposed ADM-HIPaR method on various parameter configurations in terms of background detection accuracy and processing speed. Experimental results demonstrate that the proposed method is more powerful than recent state-of-the-art approaches.

Future work will concentrate on handling hard shadow problem, as well as automatically selecting best-score parameters for particular background scenarios.

5. Acknowledgements

References

[1] O. Barnich and M. V. Droogenbroeck. Vibe: A universal background subtraction algorithm for video sequences. *IEEE Transactions on Image Processing*, 20(6):1709–1724, June 2011.

[2] Y. Benezeth, P.-M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger. Comparative study of background subtraction algorithms. *Journal of Electronic Imaging*, 19(3):033003–033003–12, 2010.

[3] T. Bouwmans, J. Gonzàlez, C. Shan, M. Piccardi, and L. Davis. Special issue on background modeling for foreground detection in real-world dynamic scenes. *Machine Vision and Applications*, 25(5):1101–1103, 2014.

[4] F. El Baf, T. Bouwmans, and B. Vachon. *Type-2 Fuzzy Mixture of Gaussians Model: Application to Background Modeling*, pages 772–781. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[5] A. Elgammal, D. Harwood, and L. Davis. *Non-parametric Model for Background Subtraction*, pages 751–767. Springer Berlin Heidelberg, Berlin, Heidelberg, 2000.

[6] G. Gemignani and A. Rozza. A robust approach for the background subtraction based on multi-layered self-organizing maps. *IEEE Transactions on Image Processing*, 25(11):5239–5251, Nov 2016.

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

[8] M. Hofmann, P. Tiefenbacher, and G. Rigoll. Background segmentation with feedback: The pixel-based adaptive segmenter. In *2012 IEEE Computer Society Conference on*

Sample	Re	Sp	FPR	FNR	PWC	Pre	F1
pedestrians	0.9283	0.9995	0.0005	0.0717	0.1194	0.9491	0.9386
PETS2006	0.9383	0.9967	0.0033	0.0617	0.4180	0.8049	0.8665
highway	0.9341	0.9924	0.0076	0.0659	1.1077	0.8853	0.9090
office	0.9848	0.9954	0.0046	0.0152	0.5321	0.9409	0.9623
canoe	0.8861	0.9991	0.0009	0.1139	0.4890	0.9734	0.9277
traffic	0.7853	0.9724	0.0276	0.2147	3.9294	0.6534	0.7133
parking	0.7020	0.9488	0.0512	0.2980	7.0278	0.5347	0.6071
bungalows	0.9356	0.9773	0.0227	0.0644	2.5203	0.7244	0.8166
Average	0.8868	0.9852	0.0148	0.1132	2.0180	0.8083	0.8426

Table 1. Foreground detection performance of the proposed ADM-HIPaR.

Method	Re	Sp	FPR	FNR	PWC	Pre	F1
DKE [5]	0.7974	0.9852	0.0148	0.2026	2.5760	0.7993	0.7803
I-GMM [21]	0.7996	0.9874	0.0126	0.2004	2.3700	0.8121	0.7945
SC-SOBS [14]	0.8585	0.9851	0.0149	0.1415	2.2609	0.8000	0.8230
GraphCutDiff [15]	0.6762	0.8945	0.1055	0.3238	12.1194	0.6401	0.5180
PBAS [8]	0.7866	**0.9924**	**0.0076**	0.2134	2.0436	**0.8640**	0.7717
Spectral [17]	0.8108	0.9911	0.0089	0.1892	**2.0033**	0.8529	0.8091
NIC [9]	0.8411	0.9827	0.0173	0.1589	2.5593	0.7432	0.7810
Our ADM-HIPaR	**0.8868**	0.9852	0.0148	**0.1132**	2.0180	0.8083	**0.8426**

Table 2. Comparison of the proposed ADM-HIPaR to several state-of-the-art methods using seven performance metrics.

Computer Vision and Pattern Recognition Workshops, pages 38–43, June 2012.

[9] T. Huynh-The, O. Banos, S. Lee, B. H. Kang, E. S. Kim, and T. Le-Tien. Nic: A robust background extraction algorithm for foreground detection in dynamic scenes. *IEEE Transactions on Circuits and Systems for Video Technology*, PP(99):1–1, 2016.

[10] A. Ilyas, M. Scuturici, and S. Miguet. Real time foreground-background segmentation using a modified codebook model. In *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 454–459, Sept 2009.

[11] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis. Real-time foreground-background segmentation using codebook model. *Real-Time Imaging*, 11(3):172–185, June 2005.

[12] C. W. Lin, W. J. Liao, C. S. Chen, and Y. P. Hung. A spatiotemporal background extractor using a single-layer codebook model. In *2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 259–264, Aug 2014.

[13] Z. Liu, K. Huang, and T. Tan. Foreground object detection using top-down information based on em framework. *IEEE Transactions on Image Processing*, 21(9):4204–4217, Sept 2012.

[14] L. Maddalena and A. Petrosino. The sobs algorithm: What are the limits? In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 21–26, June 2012.

[15] A. Miron and A. Badii. Change detection based on graph cuts. In *2015 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 273–276, Sept 2015.

[16] 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, Aug 2000.

[17] M. Sedky, M. Moniri, and C. C. Chibelushi. Spectral-360: A physics-based technique for change detection. In *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 405–408, June 2014.

[18] A. Sobral and A. Vacavant. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Computer Vision and Image Understanding*, 122:4 – 21, 2014.

[19] C. Stauffer and W. E. L. Grimson. Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):747–757, Aug. 2000.

[20] Y. Wang, P. M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar. Cdnet 2014: An expanded change detection benchmark dataset. In *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 393–400, June 2014.

[21] Z. Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 2, pages 28–31 Vol.2, Aug 2004.