

Improving NIC Algorithm Using Different Binary Structure Elements For Multi-modal Foreground Detection

Thien Huynh-The
Dept. of Computer Science &
Engineering,
Kyung Hee University
Yongin-si, Gyeonggi-do, Korea
thienht@oslab.khu.ac.kr

Cam-Hao Hua
Dept. of Computer Science &
Engineering,
Kyung Hee University
Yongin-si, Gyeonggi-do, Korea
hao.hua@oslab.khu.ac.kr

Sungyoung Lee
Dept. of Computer Science &
Engineering,
Kyung Hee University
Yongin-si, Gyeonggi-do, Korea
sylee@oslab.khu.ac.kr

ABSTRACT

This paper improves a remarkable background estimation algorithm, namely Neighbor-based Intensity Correction (NIC) which is used in the background subtraction technique for foreground detection. The algorithm has an efficient intensity correction scheme to update the current background based on calculating the standard deviation of two windows captured from the background and the input frame in which the windows are constructed by a squared-structure binary mask. Although the NIC algorithm achieved the comparative results with existing approaches on the foreground detection accuracy and processing speed, its performance in the multi-modal background including high-speed motion and camera jitter should be improved. In the original algorithm, we recognize that the shape of a binary mask further affects the updating performance besides the window size which was already analyzed. Various shapes are therefore recommended for the multi-modal background adaptation. Moreover, an adaptive threshold identified by referring several previous Otsu thresholds to cope with the high-speed motion challenge is proposed. Experimental results on some standard datasets such CAVIAR 2004, AVSS2007, PETS 2009, and CDNET 2014, demonstrate that the foreground detection accuracy is significantly boosted with 2.6-6.7% of the F-measure metric.

CCS Concepts

•Computing methodologies → Computer vision problems; *Motion capture*; Scene anomaly detection; Image manipulation; Image processing;

Keywords

Background Estimation, Foreground Detection, Adaptive Otsu Thresholding

1. INTRODUCTION

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

IMCOM '17, January 05-07, 2017, Beppu, Japan

© 2017 ACM. ISBN 978-1-4503-4888-1/17/01...\$15.00

DOI: <http://dx.doi.org/10.1145/3022227.3022337>

Despite the wide utilization of foreground detection in visual surveillance systems [7] for both indoor and outdoor scenarios, the authors aim to discuss some major issues consisted of the accuracy of foreground detection for multi-modal backgrounds. Compared with difference frame and optical flow approaches, background subtraction based techniques have some advantages for real-time implementation, however, detection accuracy mostly depends on the scene background that has to be estimated by a background modeling algorithm. Fundamentally, background is defined as the part of a scene that stays behind a main figure or object in a painting, photograph, etc., but in most researches, it is recognized as a reference image without human or objects. Some existing estimation models are capable of improving the terms of accuracy and computational cost with more frames in use, nevertheless, they are failed in harsh conditions such as dynamic background and camera jitter scenarios. A background updating scheme is commonly considered as an advanced solution for above challenges.

In spite of being recognized as one of the most simple schemes, statistical approaches cannot satisfy the real-time response due to the statistic accumulation process and as well as failure in involving objects as foreground in the intermittent object motion cases [13], [16], [10]. The most widely used background estimation algorithm is Gaussian Mixture Model (GMM) [18], in which the background is consecutively updated by an online approximation. The natural drawback of GMM is algorithm assumptions: (i) the background region is larger and more frequently observable than the foreground region and (ii) its variance is small enough. Several GMM based improvements have been recommended with an adaptive model component-parameter calculation. Some optimization techniques including online expectation-maximization [12] and particle swarm optimization algorithm [3], were further applied for learning model. Nevertheless, the fact that the parameter estimation in GMM based approaches is still a challenging task when applying them to practical systems.

Some researches have studied non-parameter background models to eliminate the parameter tuning task. In the codebook techniques, the background pixels were quantized as codewords to illustrate a compressed background model during several input frames [11]. Enhanced codebook based approaches [9], [14] improved the foreground detection accuracy in some special conditions such as dynamic background and camera jitter. A multilayer codebook model [6] was developed for non-stationary background removal and process-

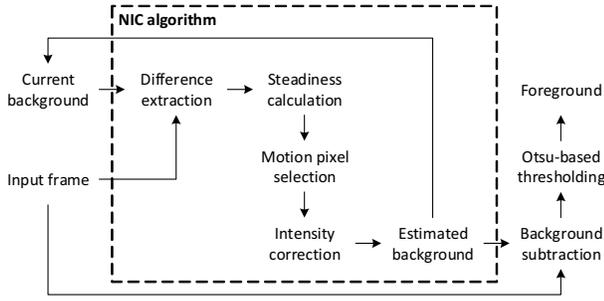


Figure 1: The workflow of the NIC-based foreground detection.

ing cost reduction. Requirements of a long duration and a high amount of memory for the background estimation step are the shortcomings of codebook based techniques. Kernel Density Estimation (KDE) [5], another non-parameter technique, estimates the probability density function on both global and local histograms to redefine the current pixel intensity. KDE was further developed for identifying whether a pixel belongs to the background or foreground site. A hybrid model [15], assembled by KDE and GMM, was successful to construct a probability density function for the background and motion model. Recently, Visual Background Extraction (ViBE) approach [2] determined whether a pixel belongs to the background based on the comparison between its intensity and random neighbouring values. Compared with other pixel based methods, ViBE reported an impressive performance of foreground detection accuracy and computational speed with a downscale version for the hardware embedment. However, ViBE is quite sensitive with rough scene models, e.g., camera jitter and highly dynamic background. A recursive algorithm, developed from the sigma-delta filter method, was a novel attempt to optimize the processing speed and memory utilization. An enhanced story of the sigma-delta algorithm was recommended by Toral et al. [19] to cope with dynamic speed motions. Recently, Neighbor-based Intensity Correction (NIC) [8] was introduced to use the neighbouring information to update the estimated background consecutively. NIC algorithm achieved a comparative performance with existing background modeling approaches, however, its accuracy was limited in multi-modal backgrounds.

In this paper, we improve the NIC algorithm [8] to efficiently detect the foreground in terms of multi-modal backgrounds. In the original algorithm, the current background is updated in general by comparing the standard deviation calculated from neighbouring pixels which are formed by a fixed-size square mask for all background challenges. Besides the mask size which was already investigated [8], the shape of mask further disturbs the estimation accuracy. Hence, the improvement on the use of different shapes such as rectangle, diagonal, and circle for the multi-modal background adaption of baseline, dynamic background, directional motion, and camera jitter is the main contribution of this paper. Moreover, an Otsu-based adaptive threshold is recommended to handle the high speed motion challenge in which the objects appear and disappear suddenly in a couple of frames. Compared to the original algorithm and the state-of-the-art approaches, our proposed improvements advance foreground detection accuracy on several standard datasets.

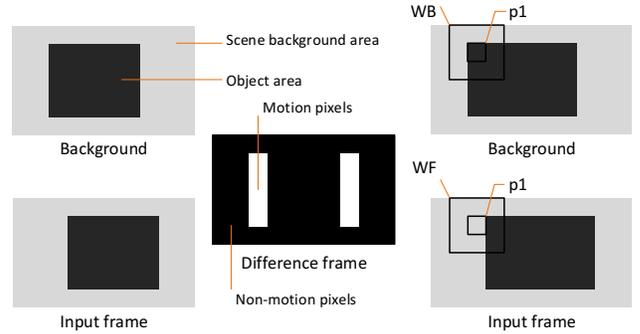


Figure 2: An illustration of capturing windows from the input frame and the current background.

2. THE METHODOLOGY

Foreground extraction is one of the most important stages in several video-based applications such as the object tracking and activity recognition which are usually perceived in real-time systems, such as Closed-Circuit Television (CCTV). However, the detection accuracy and processing speed are still challenges, especially in poorly visual conditions. In the research, we improve the NIC algorithm [8] for the background estimation and present an Otsu-based adaptive threshold for background subtraction. The workflow of the NIC based foreground detection approach is presented in Fig. 1 in which the improvements are performed in *Intensity correction* and *Otsu-based thresholding*.

2.1 The NIC algorithm

This section briefly reviews the NIC algorithm [8]. For the i^{th} frame ($\forall i \geq 2$), the difference frame, denoted D_i , between the estimated background B_{i-1} and the input frame F_i , is computed by the following equation:

$$D_i(x, y) = |F_i(x, y) - B_{i-1}(x, y)| \quad (1)$$

where (x, y) is pixel coordinates with $x \leq P$ and $y \leq Q$, $P \times Q$ is the image resolution. In practice, D may contain the information of object motion and noise, therefore, it should be separated into motion and non-motion areas by the constant threshold τ :

$$D_i^*(x, y) = \begin{cases} 1 & ; D_i(x, y) \geq \tau \\ 0 & ; D_i(x, y) < \tau \end{cases} \quad (2)$$

The output D_i^* is a binary image with 0-bit pixels representing the non-motion region and 1-bit pixels representing the motion region. In the next step, to eliminate outliers and minimize computational cost, the pixel intensity stability is studied in the whole of the image. A basic idea is that the stability of a pixel will be downgraded if its intensity is consecutively changed frame by frame and vice versa. Concretely, the stability is calculated as follows:

$$S_i(x, y) = \begin{cases} S_{i-1}(x, y) - 1 & ; \forall D_i^*(x, y) = 1 \\ S_{i-1}(x, y) + 1 & ; \forall D_i^*(x, y) = 0 \end{cases} \quad (3)$$

where S is the stability matrix which is initially a null/zero matrix, i.e., $S_1 = 0$. By crossing two conditions of motion region and negative stability, the motion pixels which need to be corrected are identified and grouped into the set P_i . The pixel selection in Fig 1 is processed by the following

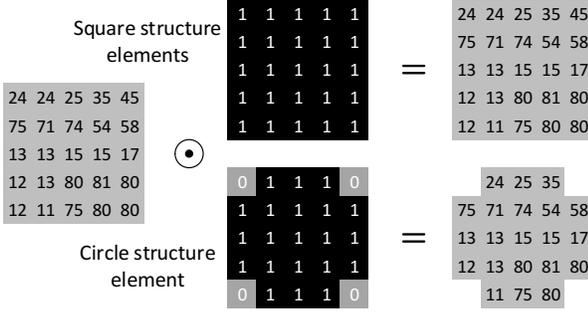


Figure 3: Multiplication operation results using two different structure elements.

expression:

$$P_i = \{(x, y) \mid [D_i^*(x, y) = 1] \cap [S_i(x, y) < 0]\} \quad (4)$$

The NIC algorithm is fully executed for the pixels in the set P_i . For each pixel in P_i as the central point, two windows containing neighbouring pixels of the background, denoted W_B , and the input frame, denoted W_F , are captured as the illustration in Fig 2. The neighbouring pixels are captured by a binary mask which is shaped in square. These pixels are mathematically determined by the multiplication operation:

$$W^* = W \odot M \quad (5)$$

where M is the binary mask, W is the window covering neighbouring pixels. This step is obviously described in Fig 3. In the original paper, $W^* = W$ as the illustration in Fig 3 with the square structure element as in Fig 3. The NIC's authors evaluated the algorithm with different sizes of the square mask, e.g., from (3×3) to (11×11) . Based on the experimental results [8], it is recognized that the size impact certainly make the influences on foreground detection accuracy and processing speed.

As the next step of NIC algorithm, the standard deviation is calculated for all pixels belonging to W^* by the following equation:

$$\sigma = \sqrt{\frac{1}{N} \sum_{n=1}^N \left(W^*(n) - \frac{1}{N} \sum_{n=1}^N W^*(n) \right)^2} \quad (6)$$

where N is the number of available pixels. For each pixel in set P , $\sigma_{(x,y)}^B$ and $\sigma_{(x,y)}^F$ of W_B^* and W_F^* are calculated. The intensity correction rule is executed based on comparing the standard deviations:

$$B_i(x, y) = \begin{cases} B_{i-1}(x, y) & ; \forall (x, y) \in P_i \mid \sigma_{(x,y)}^F \geq \sigma_{(x,y)}^B \\ F_i(x, y) & ; \forall (x, y) \in P_i \mid \sigma_{(x,y)}^F < \sigma_{(x,y)}^B \end{cases} \quad (7)$$

By this way, the background is successively updated at each input frame and then provided to the background subtraction stage.

2.2 Improvement with multi-shape for multi-modal backgrounds

In the NIC algorithm, pixel identification for the standard deviation calculation evidently alters the background updating performance. Besides the mask size that was evaluated in the original research [8], the shape also has a certain influence that should be put to the test. Five typical masks

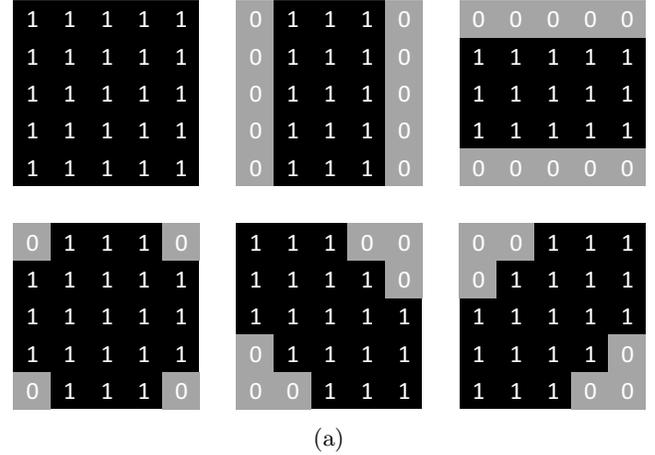


Figure 4: An illustration of using two different structure elements: (a) six structure elements (from left to right: square, vertical, horizontal, circle, and diagonal) and (b)

with different shapes, called structure elements (se), shown in Fig 4(a) are recommended to use in the NIC algorithm for multi-modal background challenge. Another example shown in Fig. 4(b) is particularly analysed in which the number and position of identified pixels for calculating the standard deviation are different. Simplifying an input frame only containing two groups, non-motion pixels with intensity g_0 and motion pixels with intensity g_1 , the standard deviation is calculated as follows (described in NIC's paper [8]):

$$\sigma = \sqrt{\frac{1}{n_0 + n_1} [n_0(g_0 - \mu)^2 + n_1(g_1 - \mu)^2]} \quad (8)$$

$$= \frac{\sqrt{n_0 n_1}}{n_0 + n_1} |g_0 - g_1|$$

where n_0 and n_1 are the numbers of non-motion and motion pixels. While the terms $|g_0 - g_1|$ and $(n_0 + n_1)$ are constant, the standard deviation depends on $\sqrt{n_0 n_1}$. As the illustration in Fig. 4(b), the output of intensity updating is false if the square structure element is utilized instead of the vertical structure element. In case of square structure element, the window captured from the background contains three non-motion pixels $n_0 = 3$ and six motion pixels $n_1 = 6$ while the window captured from the input frame has six non-motion

pixels $n_0 = 6$ and six motion pixels $n_0 = 3$. Following (8) that $\sigma_{p_2}^B = \sigma_{p_2}^F$ leads to $B_i(p_2) = B_{i-1}(p_2)$, i.e., the intensity of p_2 is preserved as g_1 instead of adjusting to g_0 . However, by using the vertical structure element, pixel p_2 is correctly updated from g_1 to g_0 because of $\sigma_{p_2}^F < \sigma_{p_2}^B$.

2.3 Otsu-based thresholding

The difference mask is calculated between the updated background and the current frame by re-using (1). An Otsu-based adaptive threshold is proposed for the foreground detection instead of the constant threshold τ . Fundamentally, Otsu method is formulated for image segmentation applications by the clustering-based thresholding. It calculates the optimal threshold to separate an image into the background area and the foreground area on the intra-class variance minimization, thus the classification error is thoroughly minimized. The Otsu threshold δ is defined as:

$$\delta_i = \arg \min_g (\sigma_\omega^2(g)) \quad (9)$$

where σ_ω^2 is the sum of weighted variances of two pixel classes at the intensity g . In the states of sudden arrival and departure of moving objects, the threshold might be inaccurately determined due to the existence of non-object motions as noise. This fact leads to wrongly identify moving objects in the scene. By referring m earlier values with the normalized weights \mathbf{w} , we smooth the current threshold:

$$\bar{\delta}_i = [\delta_i \quad \delta_{i-1} \quad \cdots \quad \delta_{i-m+1}] \times \mathbf{w}^T \quad (10)$$

where \mathbf{w} is the weight vector:

$$\mathbf{w} = \frac{[m^2 \quad (m-1)^2 \quad \cdots \quad 1]}{\sum_{k=1}^m k^2} \quad (11)$$

The thresholding process is implemented by replacing τ by $\bar{\delta}$ in (2). As the post-processing steps, some morphological operations, consisted of erosion and dilation, are usefully utilized to remove salt - pepper noise and small holes, to fuse narrow breaks and long thin gulfs, as well as fill gaps in the contour.

3. EXPERIMENTAL RESULTS AND CONCLUSION

We evaluate the NIC based improvement algorithm on six videos which are *Walk3* from CAVIAR [4], *PV_Medium* from AVSS 2007 [1], *View_006* from PETS 2009 [17], and *highway*, *canoe*, *badminton* from CDNET 2014 [20]. The experiments are simulated by MATLAB R2013a on a desktop PC operating Windows 7 with a 2.67 GHz Intel Core i5 CPU and 4GB RAM. The quality of estimated backgrounds in the statistical background scenes is evaluated by the Peak Signal to Noise Ratio (PSNR) and Structure Similarity (SSIM). As the principal experiment, the foreground detection accuracy is validated and further compared with some state-of-the-art approaches on quantitative performance metrics: True Positive Rate (TPR), False Positive Rate (FPR), Precision (PRE), and F-measure (F1). The authors set $\tau = 25$ and $m = 5$ as the default setting parameters. Three experiments are explained as follows:

- In the first experiment, the background estimation quality is measured for the statistical background challenge

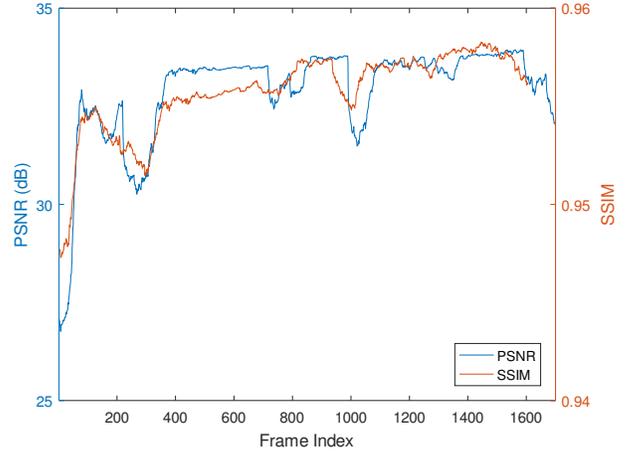


Figure 5: Background estimation results for *highway* sequence.

on the *highway* sequence using the circle structure element se4.

- Secondly, the foreground detection accuracy is benchmarked for all sequences representing challenges of the high-speed motion, shadow, camera jitter, and dynamic background.
- In the last one, the improved version is compared with original algorithm and some existing highlight approaches.

In the first experiment, the algorithm is benchmarked for the *highway* sequence representing the statistical background because obtaining the pure background in other sequences reflecting the dynamic background and camera jitter is almost impossible. As the PSNR and SSIM results are graphically shown in Fig. 5, the background quality is improved following the increment of the number of frames, however, the sudden appearance-disappearance of the objects in the high-speed motion scenario affects to estimation results at some frames. The Otsu-based adaptive thresholding cannot completely solve this critical problem.

In the next experiment, the foreground detection evaluation is performed for all video sequences using different structure elements. The visualization results of some samples are presented in the Fig. 6 and the quantitative results are graphically reported in Table 1. Compared with other structure elements, se4 (the circle shape) is most efficient for the camera jitter and the dynamic background challenge as well as with F1 measurement, e.g., 0.8264, 0.9440, and 0.9246 respectively corresponding to *PV_Medium*, *badminton*, and *canoe* sequence. The shadow problem and high-speed motion in *PV_Medium* and *highway* videos are adequately solved by either se5 or se6 with average F1 of 0.9484 and 0.8244, respectively. However, it should be noted that the remarkable result in the high-speed motion challenge also comes from the Otsu-based adaptive threshold. With the normal condition represented by the *View_006*, the differences in the accuracy of foreground detection from structure elements are insignificant. According to obtained results, the non-direction shaping mask se4 is strongly recommended for the multi-modal background challenge.

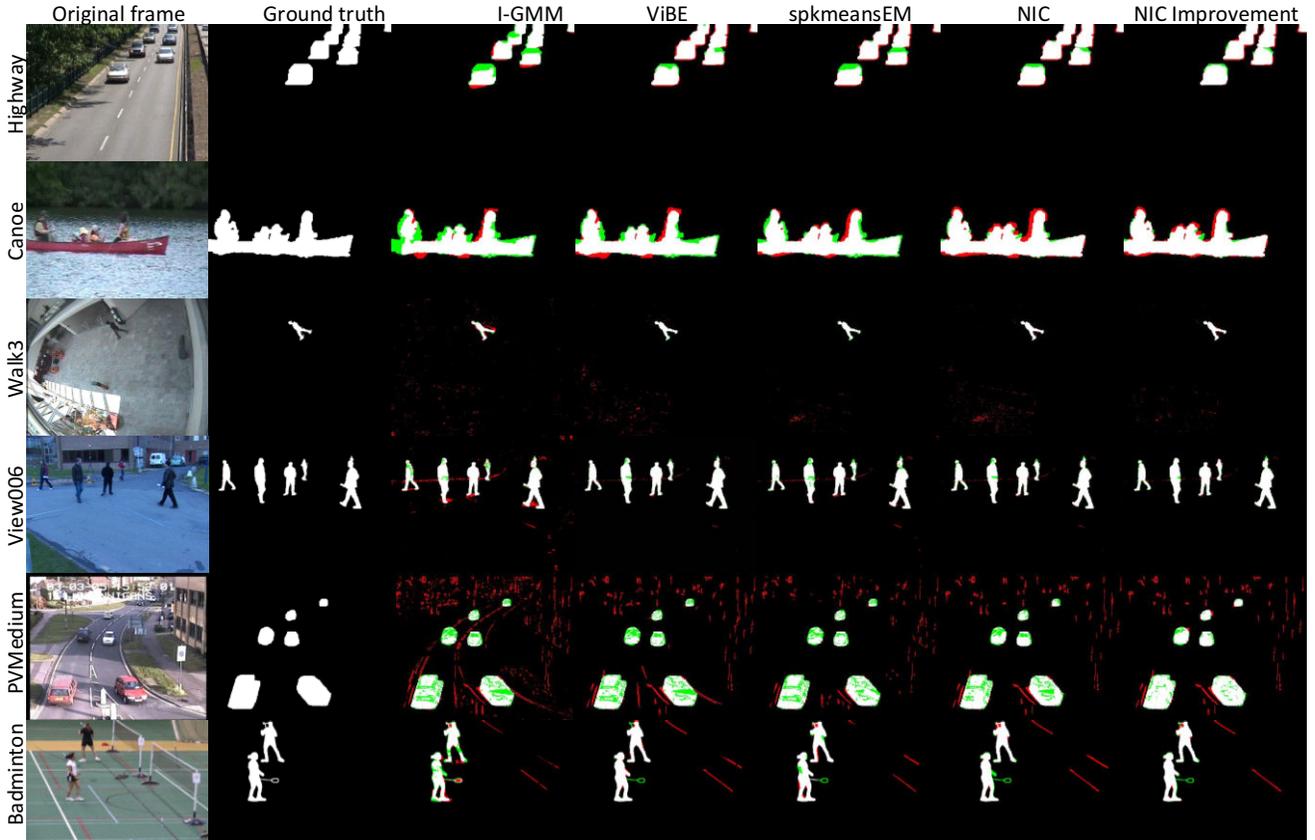


Figure 6: Visual foreground detection results of testing sequences with true positive pixel in white, true negative in black, false positive in red, and false negative in green.

The last experiment compares the NIC-based improvement algorithm with the original and other approaches including the improved adaptive GMM (I-GMM) [21], the original ViBE [2], and the spherical K-means Expectation-Maximization (spkmeansEM) [12]. All of the benchmarked algorithms are implemented by ourselves. The extracted foregrounds from all evaluated approaches are shown in Fig. 6 and the quantitative results are plotted in Table 2. It is important to note that *se4* is used for this experiment. Compared to the original algorithm, the detection accuracy of the NIC improvement is mostly boosted thanks to the new efficient mask and the adaptive threshold. TPRs are strongly enhanced for the *canoe*, *Walk3*, *View_006*, and *PV_Medium*, although the FPRs sometimes increases unfortunately in some of the videos. For all datasets, we reach the average F1 of 0.9108 which is higher than some testing approaches 2.99-6.70%. Compared with others, I-GMM frequently produces the poorest accuracy due to the natural limitation in parameter selection for each particular background challenge. Based on the results on the *highway* and *View_006* sample, ViBE is really effective for the high-speed motion and the normal background scenarios, however, its weakness is exposed in the tasks of camera jitter and dynamic background. By combining the spherical K-means clustering and the expectation-maximization algorithm, spkmeansEM produces comparative accuracy for all testing sequences.

4. CONCLUSIONS

We improved the NIC algorithm for multi-modal backgrounds by exploiting different structure elements for masking a window of pixels. In the background subtraction, an Otsu-based adaptive threshold is recommended against the high-speed motion challenge. The improvement is evaluated on six sequences representing some critical background challenges and then compared with existing highlight approaches in term of foreground detection accuracy. Based on the obtained results, the improvement with the circle mask, a non-directional shaping mask, mostly outperforms the original algorithm using the square element and other existing approaches 2.6-6.7% of higher accuracy, especially for the camera jitter and dynamic background. In the future, an automatically selective scheme for an appropriate structure element and the complexity evaluation need to be studied as well.

5. ACKNOWLEDGMENTS

This work was supported by the Industrial Core Technology Development Program, funded by the Korean Ministry of Trade, Industry and Energy (MOTIE), under grant number #10049079. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) NRF-2014R1A2A2A01003914

6. REFERENCES

Table 1: TPR, FPR, PRE, And F1 Results Of Foreground Detection With Different Structure Elements.

HIGHWAY				
	TPR	FPR	PRE	F1
se1	0.902	0.006	0.980	0.940
se2	0.910	0.009	0.970	0.939
se3	0.906	0.008	0.974	0.939
se4	0.896	0.005	0.985	0.938
se5	0.922	0.007	0.977	0.949
se6	0.916	0.005	0.983	0.949

CANOE				
	TPR	FPR	PRE	F1
se1	0.871	0.014	0.967	0.917
se2	0.860	0.012	0.972	0.912
se3	0.854	0.015	0.963	0.905
se4	0.885	0.014	0.969	0.925
se5	0.870	0.017	0.960	0.913
se6	0.874	0.010	0.976	0.922

WALK3				
	TPR	FPR	PRE	F1
se1	0.917	0.054	0.876	0.896
se2	0.905	0.049	0.883	0.894
se3	0.899	0.052	0.878	0.888
se4	0.932	0.048	0.889	0.910
se5	0.916	0.042	0.899	0.907
se6	0.920	0.045	0.895	0.907

VIEW006				
	TPR	FPR	PRE	F1
se1	0.873	0.021	0.948	0.909
se2	0.876	0.023	0.943	0.908
se3	0.914	0.031	0.928	0.921
se4	0.908	0.028	0.935	0.921
se5	0.888	0.025	0.939	0.913
se6	0.912	0.029	0.933	0.922

PV_MEDIUM				
	TPR	FPR	PRE	F1
se1	0.701	0.041	0.875	0.779
se2	0.719	0.048	0.860	0.783
se3	0.740	0.041	0.881	0.804
se4	0.806	0.059	0.848	0.826
se5	0.797	0.056	0.853	0.824
se6	0.789	0.051	0.864	0.824

BADMINTON				
	TPR	FPR	PRE	F1
se1	0.894	0.008	0.983	0.936
se2	0.882	0.006	0.986	0.931
se3	0.876	0.008	0.981	0.925
se4	0.908	0.008	0.983	0.944
se5	0.892	0.010	0.979	0.934
se6	0.896	0.009	0.980	0.936

- [1] AVSS2007. <http://www.eecs.qmul.ac.uk/andrea/avss2007d.html/>. 2007.
- [2] O. Barnich and M. V. Droogenbroeck. Vibe: A universal background subtraction algorithm for video sequences. *IEEE Trans. Image Process.*, 20(6):1709–1724, Jun 2011.
- [3] H. Bhaskar, L. Mihaylova, and A. Achim. Video foreground detection based on symmetric alpha-stable mixture models. *IEEE Trans. Circuits Syst. Video Technol.*, 20(8):1133–1138, Aug 2010.
- [4] CAVIAR2004. <http://groups.inf.ed.ac.uk/vision/caviar/caviardata1/>. 2004.
- [5] A. M. Elgammal, D. Harwood, and L. S. Davis. Non-parametric model for background subtraction. In *Proc. European Conf. Computer Vision (ECCV)*, volume 2, pages 751–767, 2012.
- [6] J.-M. Guo, C.-H. Hsia, Y.-F. Liu, M.-H. Shih, C.-H. Chang, and J.-Y. Wu. Fast background subtraction based on a multilayer codebook model for moving object detection. *IEEE Trans. Circuits Syst. Video Technol.*, 23(10):1809–1821, Oct 2013.
- [7] T. Huynh-The, O. Banos, B.-V. Le, D.-M. Bui, Y. Yoon, and S. Lee. Traffic behavior recognition using the pachinko allocation model. *Sensors*, 15(7):16040, 2015.
- [8] 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.
- [9] A. Ilyas, M. Scuturici, and S. Miguet. Real time foreground-background segmentation using a modified codebook model. In *Advanced Video and Signal Based Surveillance, 2009. AVSS '09. Sixth IEEE International Conference on*, pages 454–459, Sept 2009.
- [10] C. R. Jung. Efficient background subtraction and shadow removal for monochromatic video sequences. *IEEE Transactions on Multimedia*, 11(3):571–577, April 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] D. Li, L. Xu, and E. D. Goodman. Illumination-robust foreground detection in a video surveillance system. *IEEE Trans. Circuits Syst. Video Technol.*, 23(10):1637–1650, Oct 2013.
- [13] L. Li, W. Huang, I. Y.-H. Gu, and Q. Tian. Statistical modeling of complex backgrounds for foreground object detection. *IEEE Transactions on Image Processing*, 13(11):1459–1472, Nov 2004.
- [14] C. W. Lin, W. J. Liao, C. S. Chen, and Y. P. Hung. A spatiotemporal background extractor using a

Table 2: Foreground Detection Accuracy Results Of The NIC-based Improvement And Existing Highlight Approaches.

HIGHWAY				
	TPR	FPR	PRE	F1
I-GMM	0.892	0.005	0.982	0.935
ViBE	0.902	0.006	0.979	0.939
spkmeansEM	0.904	0.008	0.975	0.938
Original NIC	0.894	0.005	0.982	0.936
Improvement	0.896	0.005	0.985	0.938

CANOE				
	TPR	FPR	PRE	F1
I-GMM	0.794	0.015	0.962	0.870
ViBE	0.823	0.016	0.961	0.886
spkmeansEM	0.833	0.018	0.957	0.890
Original NIC	0.839	0.018	0.956	0.893
Improvement	0.885	0.014	0.969	0.925

WALK3				
	TPR	FPR	PRE	F1
I-GMM	0.786	0.024	0.930	0.852
ViBE	0.813	0.022	0.939	0.872
spkmeansEM	0.842	0.028	0.925	0.882
Original NIC	0.849	0.024	0.936	0.890
Improvement	0.932	0.048	0.889	0.910

VIEW006				
	TPR	FPR	PRE	F1
I-GMM	0.783	0.016	0.955	0.861
ViBE	0.838	0.021	0.945	0.889
spkmeansEM	0.813	0.025	0.935	0.873
Original NIC	0.823	0.020	0.947	0.881
Improvement	0.908	0.028	0.935	0.921

PV_MEDIUM				
	TPR	FPR	PRE	F1
I-GMM	0.589	0.066	0.786	0.673
ViBE	0.649	0.046	0.853	0.737
spkmeansEM	0.711	0.051	0.853	0.776
Original NIC	0.707	0.045	0.865	0.778
Improvement	0.806	0.059	0.848	0.826

BADMINTON				
	TPR	FPR	PRE	F1
I-GMM	0.788	0.010	0.976	0.872
ViBE	0.862	0.008	0.982	0.918
spkmeansEM	0.876	0.007	0.984	0.927
Original NIC	0.883	0.007	0.984	0.931
Improvement	0.908	0.008	0.983	0.944

single-layer codebook model. In *Advanced Video and Signal Based Surveillance (AVSS), 2014 11th IEEE International Conference on*, pages 259–264, Aug 2014.

- [15] Z. Liu, K. Huang, and T. Tan. Foreground object detection using top-down information based on em framework. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(9):4204–4217, Sep 2012.
- [16] J. Peng and J. WeiDong. Statistical background subtraction with adaptive threshold. In *Image and Signal Processing (CISP), 2012 5th International Congress on*, pages 123–127, Oct 2012.
- [17] PETS2009. <http://ftp.pets.rdg.ac.uk/pub/pets2009/>. 2009.
- [18] 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.
- [19] S. Toral, M. Vargas, F. Barrero, and M. G. Ortega. Improved sigma-delta background estimation for vehicle detection. *IET Electron. Lett.*, 45(1):32–34, Jan 2009.
- [20] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar. Cdnnet 2014: An expanded change detection benchmark dataset. In *CVPRW, 2014 IEEE Conf.*, pages 393–400, June 2014.
- [21] Z. Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *ICPR, 2004 Int. Conf.*, volume 2, pages 28–31 Vol.2, Aug 2004.