A Co-occurrence Background Model with Hypothesis on Degradation Modification for Object Detection in Strong Background Changes

Wenjun Zhou*, Shun'ichi Kaneko*, Manabu Hashimoto†, Yutaka Satoh‡ and Dong Liang§
*Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan 060–0814
Email: zhouwenjun@hce.ist.hokudai.ac.jp, kaneko@ssi.ist.hokudai.ac.jp
†Chukyo University, Japan
‡National Institute of Advanced Industrial Science and Technology, Japan
§Nanjing University of Aeronautics and Astronautics, China

***Abstract*—Object detection has become an indispensable part of video processing and current background models are sensitive to background changes. In this paper, we propose a novel background model using an algorithm called Co-occurrence Pixel-block Pairs (CPB) against background changes, such as illumination changes and background motion. We utilize the co-occurrence "pixel to block" structure to extract the spatial-temporal information of each pixel to build background model, and then employ an efficient evaluation strategy to identify the current state of each pixel, which is named as correlation dependent decision function. Furthermore, we also introduce a Hypothesis on Degradation Modification (HoD) into CPB structure to reinforce the robustness of CPB. Experimental results obtained from the dataset of the PETS 2001, AIST-Indoor, SBMnet and CDW-2012 databases show that our models can detect objects robustly in strong background changes.**

I. Introduction

As a pre-processing approach utilized in many computer vision applications, object detection plays an important role in various tasks [1]–[3]. However, object detection is faced with many practical challenges [4], especially the background changes, not least of which are those related to *illumination changes*, e.g. variable sunlight or lights being switched on and off indoors, and then *background motions*, e.g. the swaying motion of trees, fleeting cloud and moving waves on the water. Fig. 1 shows the typical examples of these challenges.

To handle such challenges, previous statical methods(such as Gaussian mixture model (GMM) [5] and Kernel Density Estimation (KDE) [6]) have been proposed, in which the intensity of each pixel is independently analyzed in the temporal domain and then the current frame is subtracted. However, such kind of methods is difficult to solve illumination changes with the intensity varies rapidly and significantly. Because a target pixel shares a similar change with its neighboring pixels, recent many local feature based methods [7]–[9] have been put forward for background modeling. However, such local feature based background models is susceptible to be affected by the dynamic motion of background, thus losing the robustness.

In this paper, we propose a robust object detection method against strong background changes and our work has already

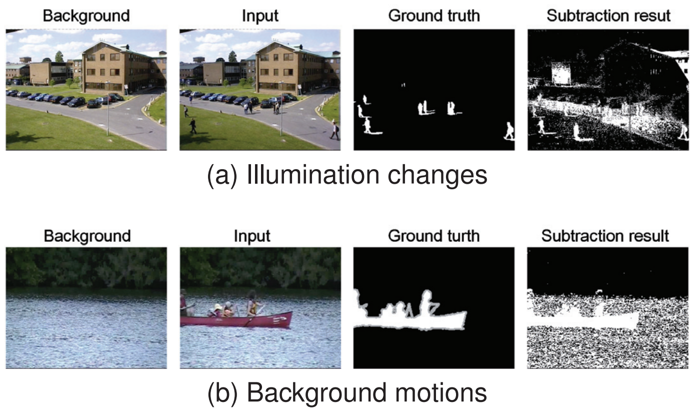


Fig. 1. The typical example results of illumination challenges and background motions by using the static frame difference approach, (a) Illumination changes: one sequence with the light intensity typically varies during day. (b) Background motions: one sequence with the water rippling.

been partially described in [10]. Here, we will introduce the new contents and give more detailed explanations. Our contributions of this paper are as follows: (1) we first successfully employ a co-occurrence "pixel to block" structure model to acquire the spatial-temporal information of each pixel from background; (2) then a novel evaluation strategy named correlation dependent decision function is introduced for foreground detection; (3) Based on CPB, we propose a Hypothesis on Degradation Modification (HoD) to adapt the dynamic changes in scenes to better distinguish objects. In this paper, Sections II introduces the working mechanism of CPB. Section III gives one introduction of Hypothesis on Degradation Modification (HoD). Section IV analyzes the experimental results from the dataset of the PETS 2001 [11], AIST-Indoor, SBMnet [12] and CDW-2012 [13]. Conclusions are discussed in Section V.

II. Methodology

The proposed CPB background model includes two processes: training process and detecting process. Fig. 2 is the working mechanism of CPB. As an extension from the "pixel

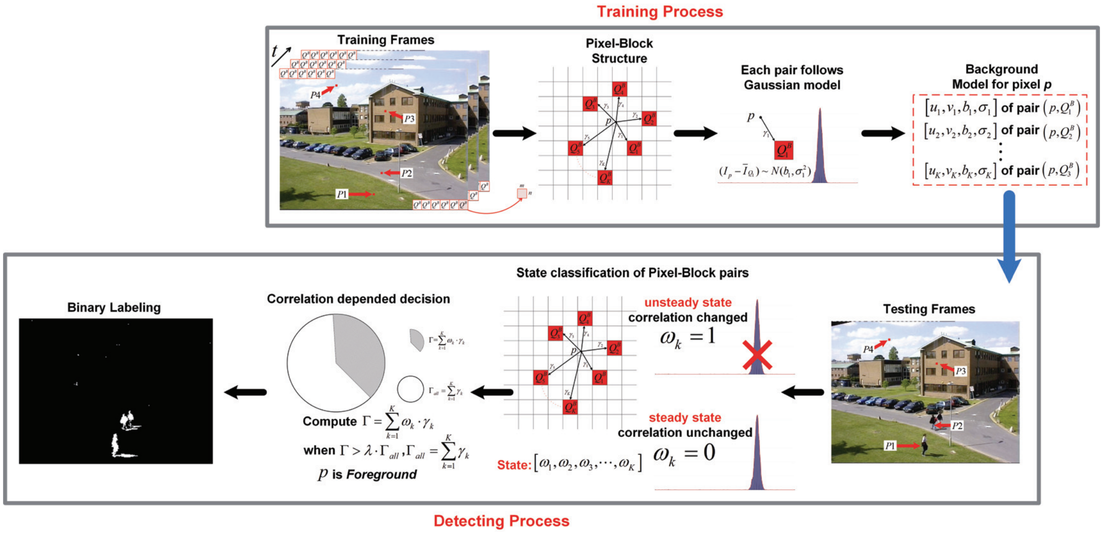


Fig. 2. Overview of working mechanism of CPB.

to pixel" structure that work [14] to estimate the target pixel p with other pixels one by one and then to select the suitable supporting pixels for the target pixel p, in CPB we compare the target pixel p with the Q^B as block, and define $\{Q_k^B\}_{k=1,2,...,K} = \{Q_1^B, Q_2^B, ..., Q_K^B\}$ to denote a supporting block set for the target pixel p. In this work, we utilize the Pearson's product-moment correlation coefficient to select the supporting blocks $\{Q_k^B\}$ for each target pixel p:

$$\{Q_k^B\}_{k=1,2,...,K} = \{Q^B|\gamma(p, Q^B) \text{is the } K \text{ highest}\}, \quad (1)$$

where

$$\gamma(p, Q_k^B) = \frac{C_{p,\bar{Q}_k}}{\sigma_p \cdot \sigma_{\bar{Q}_k}} \quad (2)$$

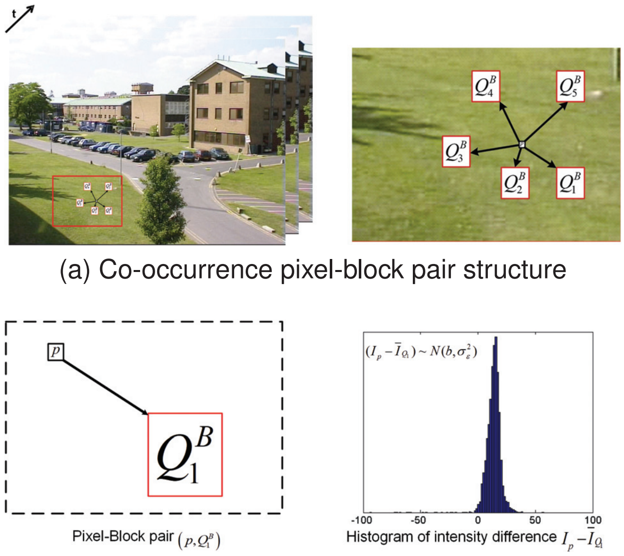
and $C_{p,\bar{Q}_k}$ is the intensity covariance between target pixel p and its k-th supporting block Q_k^B from a set of training frames, σ_p and $\sigma_{\bar{Q}_k}$ are the standard deviations in the pixel and the block, respectively.

For each pixel p, it is expected to own one or more blocks Q^B that maintain a stable relation in the difference $I_p - \bar{I}_Q$ throughout the whole training frames as shown in Fig.3, where $\bar{I}_Q$ is the average intensity of block Q^B. When such relation maintains steady as time goes by, the deviation between the target pixel p and its supporting block Q^B would be follow a single Gaussian distribution. This relation is called as "Co-occurrence between intensity" as shown in Fig.3(b) and we can utilize this knowledge to design the background model for the characteristics in background pixels.

A. Co-occurrence background model

We build a co-occurrence model using the single Gaussian distribution for the selected K pixel-block pairs:

$$\Delta_k \sim N(b_k, \sigma_k^2) \quad \Delta_k = I_p - \bar{I}_{Q_k}, \quad (3)$$



(a) Co-occurrence pixel-block pair structure

(b) Pairwise statistical model of pixl-block pair (p, Q_1^B)

Fig. 3. Basic structure of co-occurrence pixel to block pair.

where I_p is the intensity of the pixel p at t frame and $\bar{I}_{Q_k}$ is the average intensity of the block Q_k^B at t frame. In CPB, each pixel-block pair(p, Q_k^B) owns an unique Gaussian and we record two parameters that the differential increment b_k and the standard deviation σ_k as model as Fig.2 shows.

Where, b_k is defined as the following expression:

$$b_k = \frac{1}{T}\sum_{t=1}^{T} \Delta_k \quad (4)$$

and the variance estimation is defined as follows:

$$\sigma_k^2 = \frac{1}{T}\sum_{t=1}^{T}(\Delta_k - b_k)^2, \quad (5)$$

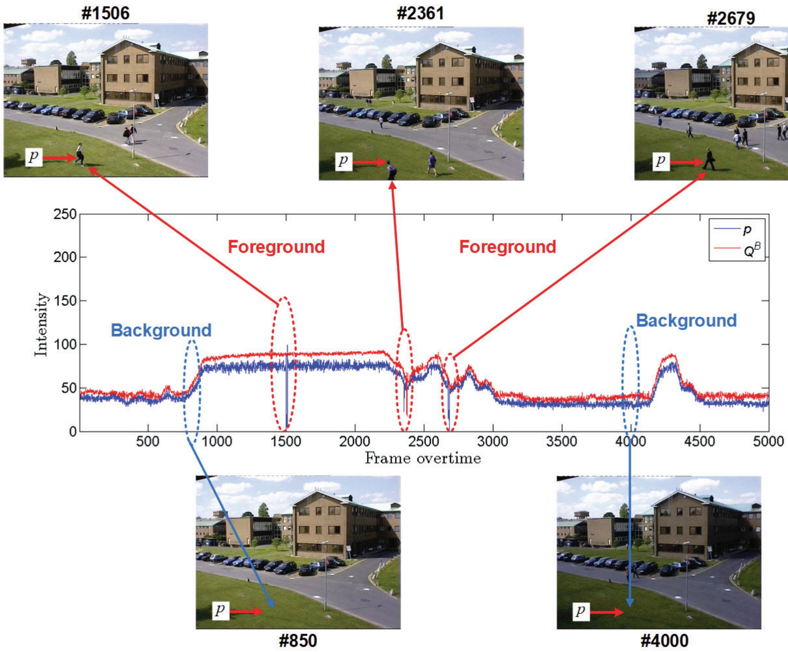


Fig. 4. Co-occurrence intensity changes between target pixel p and supporting block Q^B.

where T is the sequence of frames.

B. Correlation dependent decision

Based on the co-occurrence background model built above, CPB can acquire the spatial-temporal information of target pixel p and then compare the difference between target pixel p and supporting block Q_k^B to judge the state of target pixel p as shown in Fig.4. Once the co-occurrence relation appears an outlier, such situation would be regarded as an unsteady state of pixel-block pair(p, Q_k^B) at current frame, thereby we could estimate target pixel p as foreground.

For each pair (p, Q_k^B), a binary function for identifying its steady or unsteady state can be defined as follows:

$$\omega_k = \begin{cases} 1 & if \;\; \left|(p - Q_k^B) - b_k\right| \geq \eta \cdot \sigma_k \\ 0 & \text{otherwise} \end{cases}, \tag{6}$$

where$\left|(p - Q_k^B) - b_k\right|$ represents a bias in the intensity difference between the real value and the modeled parameter b_k to estimate the steady or unsteady state of each pair (p, Q_k^B), where η is a constant for setting some significant level in this statistical test procedure and ω_k presents a logical judgment: *the steady state* with 0 or *the unsteady state* with 1 for each pair, respectively. To define an efficient decision function for target pixel, here we introduce γ_k of the k-th elemental pair (p, Q_k^B) as a weight in the weighted summation of the products $\omega_k \cdot \gamma_k$ based on the previous decision proposed in [14], [15]. The definition is realized as Γ as follows:

$$\Gamma = \sum_{k=1}^{K} \omega_k \cdot \gamma_k \tag{7}$$

with two following significances: first, Γ could count up the unsteady pairs; second, the maximum value of Γ could be possibly obtained in the case that all of K elemental pairs are in the unsteady state and it is also a relative value with respect to the target pixel. Furthermore, Γ would not miss to count any high γ_k in the summation to lead a wrong decision. To realize relative decision making on Γ, we can have the following possible maximum value of it.

$$\Gamma_{all} = \sum_{k=1}^{K} \gamma_k. \tag{8}$$

With the consideration of mentioned above, by use of Γ_{all}, we can define the following evaluation criterion to classify the target pixel into the foreground class as: **if** $\Gamma > \lambda \cdot \Gamma_{all}$, **then** p is *foreground* and λ is a threshold parameter.

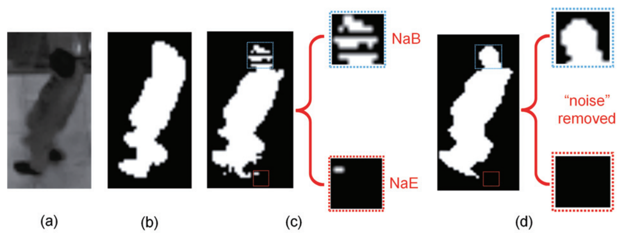


Fig. 5. Example of hypothesized noise. (a) Raw data. (b) Ground truth. (c) Description of hypothesized noise. (d) Modified result.

III. Hypothesis on Degradation Modification

A. Hypothesis on degradation

In practice, after a long utilization of initial CPB background model in an unlearned sequence, the expected relative

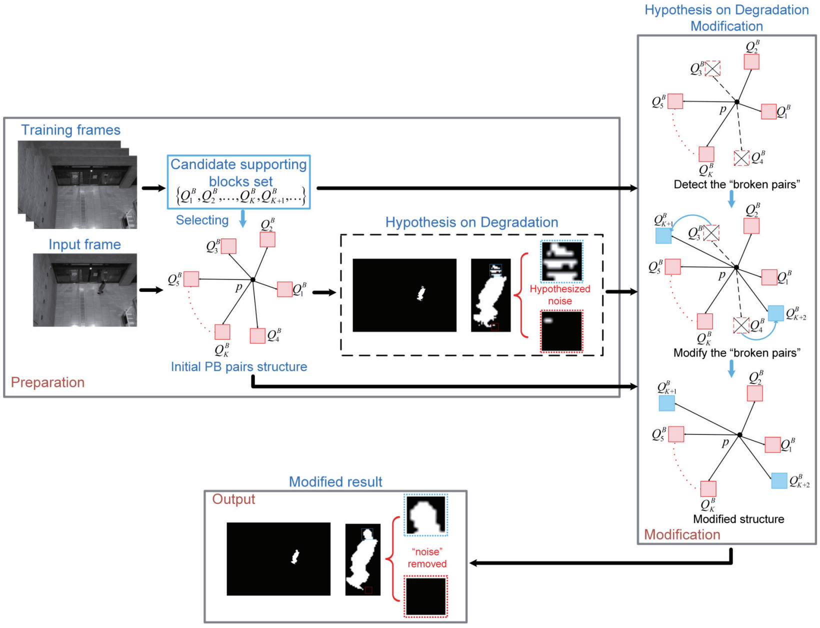


Fig. 6. Overview of HoD Modification.

relation of the pixel-block pair might be broken. In other words, initial CPB model might generate a degradation with the passage of time, then some "noise" might arise in detecting process. Here, we define such assumption as "Hypothesis on Degradation" and name the "noise " in detecting process as "hypothetical noise": (1) the hole surrounded by the detected foreground pixels, which is estimated as the background and we named it 'NaB'; (2) the dot surrounded by the non-detected pixels, which is estimated as the event and we named it 'NaE'. Fig. 5 shows an example of the hypothetical noise using AIST-Indoor dataset provided by the National Institute of Advanced Industrial Science and Technology in Japan. To reinforce the merits of CPB background model, we introduce a tactic named Hypothesis on degradation modification (HoD) into the CPB structure to remove the hypothetical noise. Fig. 6 describes an overview of the proposed HoD. Note that HoD is not one post-processing technique, in this study, HoD is an update approach of model structure to reinforce the robustness of CPB, and it is also a feasible on-line mode of CPB structure in future.

B. Broken pixel-block pairs detection

As shown in Fig. 6, first we need to detect the broken elemental pairs in pixel-block structure of the hypothetical noise. In this study, we assume that the larger γ (mentioned in Section II-A) could hold a higher weight in the trained pixel-block structure and such pair would be more likely to affect the state of pixels. Thus it is obvious that the pairs with large γ in unsteady state might cause a decision of NaE, whereas the pairs with large γ in steady state might cause a decision of NaB. With the above assumption, we propose a weight-based decision function to detect the broken pair:

$$\textit{if}\ \ \gamma_m \geq \bar{\gamma},\ \ \textit{then}\ (p, Q_m^B)\ \textit{is broken} \tag{9}$$

where (p, Q_m^B) is the pair, which is in unsteady state of NaE or steady state of NaB. Depending on the noise is NaE or NaB, the threshold $\bar{\gamma}$ owns different definition. In the case of NaE, it is defined by use of the total number of unsteady pairs $M = \sum_{k=1}^{K} \omega_k$ as follows:

$$\bar{\gamma} = \frac{1}{M} \sum_{k=1}^{K} \gamma_k \cdot \omega_k = \frac{1}{M} \mathit{\Gamma}. \tag{10}$$

In the other hand, for NaB case, it is defined as follows:

$$\bar{\gamma} = \frac{1}{K-M} \sum_{k=1}^{K} \gamma_k \cdot (1-\omega_k) = \frac{1}{K-M} (\mathit{\Gamma}_{all} - \mathit{\Gamma}). \tag{11}$$

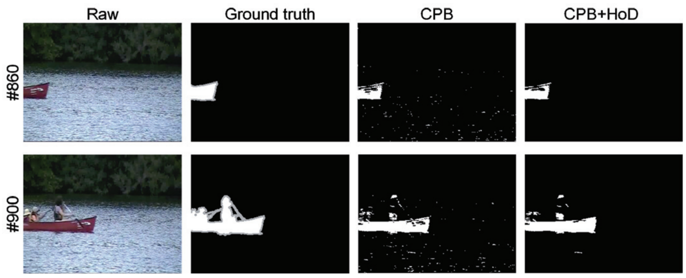
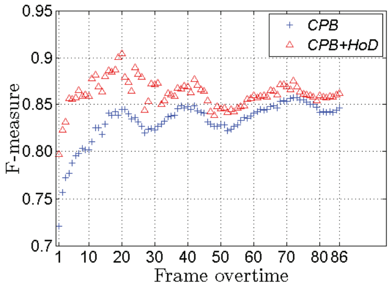


Fig. 7. The typical results of CPB and CPB+HoD in the sequence *canoe*.

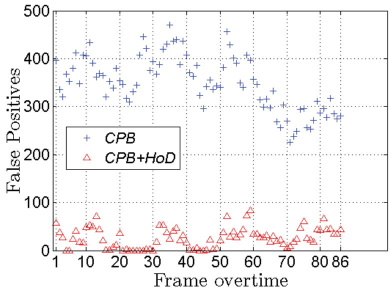
There is a slight difference in the above definitions, and then we record these broken pairs for the next process.

C. Structure modification

Then, we try to exchange the broken pair by new one which is kept as a spare pair in the training process and remove the hypothetical noise by using the modified pixel-block structure as shown in Fig. 6.



(a) F-measure comparison



(b) False Positives comparison

Fig. 8. Comparison between CPB and CPB+HoD in the sequence *canoe* overtime.

IV. Experiments

At first, to prove the performance of HoD, we compare the results of CPB and CPB+HoD in the sequence *canoe*, which is one typical scene with water rippling [13]. We select the first 300 frames for training and then at detecting process, the frame #845 to #930 with the continuous movement of the canoe, a total of 86 frames are selected as the testing frames. Fig.7 shows the typical results of CPB and CPB+HoD and Fig. 8 illustrates the *F-measure* and *False Positives* comparison between CPB and CPB+HoD in the sequence *canoe* overtime. From Fig. 7 and Fig. 8, it is clearly that with the help of HoD, CPB+HoD has a significant improvement over CPB and further restrained the noise in scene. These results suggest that HoD can effectively suppress the degradation in CPB with the passage of time.

On the other hand, we also consider four datasets to evaluate the proposed methods: (1) PETS2001 dataset [11]: one typical sequence of gradual illumination changes. (2) AIST-Indoor dataset: the sequence with sudden illumination change, which contains the strong sudden light changes when the auto-door opening, in such moment it is difficult to detect true foreground from the scene. AIST-Indoor dataset is provided by the National Institute of Advanced Industrial Science and Technology in Japan. (3) SBMnet dataset [12]: one sequence *advertisementBoard* with strong background motion is selected from SBMnet dataset for testing, and this sequence contains an ever-changing advertising board in the scene. (4) CDW-2012 dataset [13]: one typical sequence *canoe* with water rippling is selected from the CDW-2012 dataset.

We compare the proposed CPB and CPB+HoD with five different foreground detection methods: GMM [5] and KDE [6], which are two well-known traditional algorithms, and three state of the art techniques IMBS [16], T2FMRF-UV [17] and SuBSENSE [9], especially SuBSENSE is one of the top-ranked methods in CDW-2012 dataset at present and T2FMRF-UV is a foreground extraction algorithm specifically for dynamic backgrounds. In contrast to the methods with complex strategies [9], [16], [17], CPB is a low-complexity algorithm that is more easily realized. The parameters for GMM, KDE, IMBS, T2FMRF-UV and SuBSENSE were set by using the tool bgslibrary [18]. In experiments, we set each block as 8×8 pixels for CPB and the parameters are discussed in [10] with details as shown in Table. I.

TABLE I
Parameters Setting of CPB

Number of supporting blocks K	20
Gaussian model threshold η	2.5
Correlation dependent decision threshold λ	0.5

To evaluate the methods in pixel level, we utilize three common analysis measurements: *Precision*, *Recall*, and *F-measure*. These metrics are widely used to estimate the quality of background subtraction methods [4], [19]. For further evaluating our CPB and CPB+HoD, we introduce the peak signal-to-noise ratio (PSNR) as our metric [20], which can be used o measure the quality of the estimated resulted compared with the background truth [21]. The definition of *PSNR* is calculated as follows:

$$PSNR = 10 \cdot \log_{10}\left(\frac{255^2}{MSE}\right), \tag{12}$$

where *MSE* is the mean square error.

Examples of the foreground detection results are presents in Fig. 9. Table II lists the evaluation of these approaches in pixel level. *F-measure* is the comprehensive evaluation for foreground detection in pixel level and it should be as large as possible. Compared with above foreground detection results, *F-measure* of our methods is better than most of the other methods. Especially, CPB+HoD is quite efficient in extracting foreground from sequences that suffers from sudden illumination change and background motion challenges. Further more, it is should be noted that CPB and CPB+HoD can lead high *Precision* and *PSNR* in most testing sequences as the results shown in Table. II, that means our algorithm is robust against noise to detect foreground in strong background changes. The processing time for detection and modification was close to 1.5 seconds and 0.1 seconds respectively with frame size of 320×240 in MATLAB platform (Intel E3 3.5GHZ and 16G).

V. Conclusions

In this paper, we developed a new method for object detection based on co-occurrence pixel-block structure under strong background changes. It was designed to handle the problem of strong background changes in reality. Furthermore,

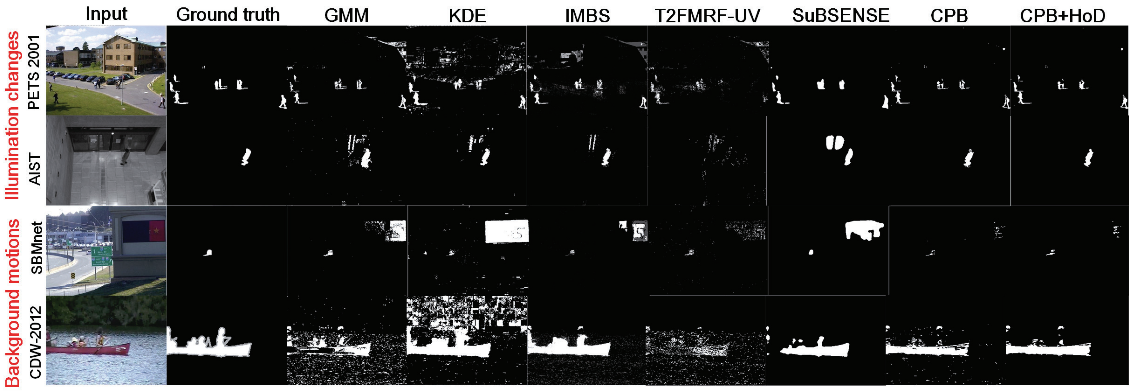


Fig. 9. The representative results in different challenging sequences.

TABLE II
PERFORMANCE EVALUATION FOR FOREGROUND DETECTION OVER FOUR DATASETS

Datasets	**PETS 2001**				**AIST**				**SBMnet**				**CDW-2012**			
Methods	Precision	Recall	F-measure	PSNR	Precision	Recall	F-measure	PSNR	Precision	Recall	F-measure	PSNR	Precision	Recall	F-measure	PSNR
GMM [5]	0.6465	0.9508	0.7697	39.46	0.6523	0.9207	0.7636	40.57	0.5151	0.5196	0.5174	26.92	0.6748	0.7024	0.6883	21.71
KDE [6]	0.5181	0.8836	0.6531	17.77	0.5896	0.6944	0.6377	38.16	0.4962	0.4856	0.4909	21.67	0.6584	0.8630	0.7468	17.22
IMBS [16]	0.5162	0.8841	0.6518	16.20	0.5760	0.6923	0.6288	36.36	0.5095	0.5118	0.5107	30.09	0.7315	0.8911	0.8035	21.60
T2FMRF-UV [17]	0.5818	0.8365	0.6863	34.94	0.6382	0.5818	0.6087	45.65	0.5508	0.5179	0.5338	35.38	0.6797	0.6114	0.6438	23.49
SuBSENSE [9]	0.9008	0.8840	0.8923	54.11	0.5864	0.7047	0.6401	37.14	0.5018	0.5033	0.5025	27.62	0.9766	0.7649	0.8573	30.80
CPB	0.9566	0.7517	0.8418	56.05	0.8651	0.8181	0.8409	53.14	0.7653	0.5118	0.6133	36.64	0.9283	0.7730	0.8436	32.61
CPB+HoD	0.9652	0.7562	0.8480	56.39	0.8668	0.8227	0.8442	53.31	0.7973	0.5214	0.6350	37.39	0.9809	0.7830	0.8708	34.16

* Note that red entries indicate the best in *F-measure*, and blue entries indicate the second best.

we realized a novel modification approach named hypothesis-on-degradation modification (HoD) for CPB to defense the possible degradation in practice and it is also a feasible on-line mode of CPB structure in future. Experimental results show the interest of the proposed methods. The future work is to develop an on-line mode of CPB structure by using the hypothesis on degradation modification (HoD).

REFERENCES

[1] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer vision and image understanding*, vol. 104, no. 2, pp. 90–126, 2006.

[2] S.-C. S. Cheung and C. Kamath, "Robust background subtraction with foreground validation for urban traffic video," *EURASIP Journal on Advances in Signal Processing*, vol. 2005, no. 14, p. 726261, 2005.

[3] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *Acm computing surveys (CSUR)*, vol. 38, no. 4, p. 13, 2006.

[4] A. Vacavant, T. Chateau, A. Wilhelm, and L. Lequièvre, "A benchmark dataset for outdoor foreground/background extraction," in *Asian Conference on Computer Vision*. Springer, 2012, pp. 291–300.

[5] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, vol. 2. IEEE, 1999, pp. 246–252.

[6] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proceedings of the IEEE*, vol. 90, no. 7, pp. 1151–1163, 2002.

[7] P.-M. Jodoin, M. Mignotte, and J. Konrad, "Statistical background subtraction using spatial cues," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 12, pp. 1758–1763, 2007.

[8] O. Barnich and M. Van Droogenbroeck, "Vibe: A universal background subtraction algorithm for video sequences," *IEEE Transactions on Image processing*, vol. 20, no. 6, pp. 1709–1724, 2011.

[9] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "Subsense: A universal change detection method with local adaptive sensitivity," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 359–373, 2015.

[10] W. Zhou, S. Kaneko, D. Liang, M. Hashimoto, and Y. Satoh, "Background subtraction based on co-occurrence pixel-block pairs for robust object detection in dynamic scenes," *IIEEJ transactions on image electronics and visual computing*, vol. 5, no. 2, pp. 146–159, 2017.

[11] "Performance evaluation of tracking and surveillance dataset 2001," *http://ftp.pets.rdg.ac.uk/pub/PETS2001*.

[12] "A dataset for testing background estimation algorithms," *http://scenebackgroundmodeling.net/*.

[13] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Changedetection. net: A new change detection benchmark dataset," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on.* IEEE, 2012, pp. 1–8.

[14] D. Liang, S. Kaneko, M. Hashimoto, K. Iwata, and X. Zhao, "Co-occurrence probability-based pixel pairs background model for robust object detection in dynamic scenes," *Pattern Recognition*, vol. 48, no. 4, pp. 1374–1390, 2015.

[15] S. Y. Elhabian, K. M. El-Sayed, and S. H. Ahmed, "Moving object detection in spatial domain using background removal techniques-state-of-art," *Recent patents on computer science*, vol. 1, no. 1, pp. 32–54, 2008.

[16] D. Bloisi and L. Iocchi, "Independent multimodal background subtraction." in *CompIMAGE*, 2012, pp. 39–44.

[17] Z. Zhao, T. Bouwmans, X. Zhang, and Y. Fang, "A fuzzy background modeling approach for motion detection in dynamic backgrounds," *Multimedia and signal processing*, vol. 346, pp. 177–185, 2012.

[18] A. Sobral, "Bgslibrary: An opencv c++ background subtraction library," in *IX Workshop de Visao Computacional (WVC2013)*, vol. 7, 2013.

[19] S. Brutzer, B. Höferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on.* IEEE, 2011, pp. 1937–1944.

[20] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of psnr in image/video quality assessment," *Electronics letters*, vol. 44, no. 13, pp. 800–801, 2008.

[21] 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*, 2016.