Signal Processing 160 (2019) 66–79

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Foreground detection based on co-occurrence background model with hypothesis on degradation modification in dynamic scenes

Wenjun Zhou[a,*], Shun'ichi Kaneko[a], Manabu Hashimoto[b], Yutaka Satoh[c], Dong Liang[d]

[a] *Graduate School of Information Science and Technology, Hokkaido University, Sapporo, 060–0814, Japan*
[b] *Chukyo University, Japan*
[c] *National Institute of Advanced Industrial Science and Technology, Japan*
[d] *Nanjing University of Aeronautics and Astronautics, China*

ARTICLE INFO

Article history:
Received 25 June 2018
Revised 15 February 2019
Accepted 18 February 2019
Available online 19 February 2019

Keywords:
Hypothesis on degradation modification (HoD)
Co-occurrence pixel-Block pairs (CPB)
Foreground detection
Illumination changes
Background motion

ABSTRACT

This work presents a Hypothesis on Degradation Modification (HoD) based on Co-occurrence Pixel-Block Pairs (CPB, which is proposed in our previous work) to further resist background changes for foreground detection, such as illumination changes and background motion. HoD provides CPB with a model update strategy that can be used for a long time. While further improving the robustness of CPB, it also stabilizes the efficiency of CPB over time. A key contribution of this work is it offers a robust background subtraction for foreground detection in dynamic scenes. The observation is robust to illumination changes and background motion and demonstrates the ability of HoD. Experimental results obtained from the datasets under different challenges of PETS 2001, AIST-Indoor, SBMnet and CDW-2012 databases prove that our algorithm has a good effectiveness for foreground detection.

1. Introduction

As a pre-processing approach utilized in many computer vision applications [1], foreground detection plays an important role in various tasks like video surveillance [2], traffic monitoring [3], scene background initialization [4,5] and object tracking [6,7]. As a basic understanding, foreground is any change in video sequences. In most cases, it should be the moving objects which are of interest to the person, such as pedestrians, vehicles, animals etc., and background must be the stationary objects that could vary in color and intensity under illumination changes over time [8]. Generally, foreground and background should be defined based on the ground truth.

As we know, one simple way to do background model is to acquire a background image without any moving objects. However, foreground detection is faced with many practical challenges [9,10], especially the background changes, not least of which is those related to *illumination changes*, e.g. variable sunlight or lights being switched on and off indoors, and *background motion*, e.g. the swaying motion of trees, fleeting cloud and moving waves on the water. The typical examples of these challenges are shown in Fig. 1.

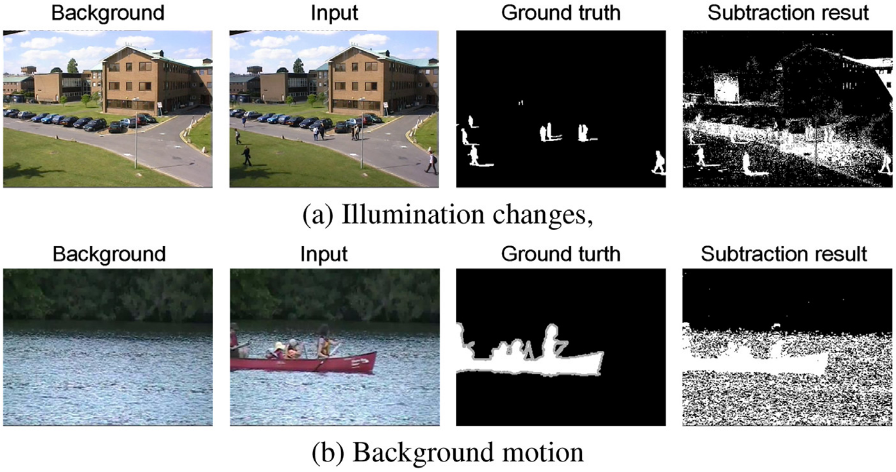
To handle such challenges, previous statical methods have been proposed, in which the intensity of each pixel is independently analyzed in the temporal domain and then the current frame is subtracted, such as Pfinder [11], using the Gaussian Mixture Model (GMM) to build a pixel-wise model for each pixel and Kernel Density Estimation (KDE) [12], a non-parametric method can detect objects in dynamic scenes. 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 [13–15] have been put forward for background modeling. Barnich et al. proposed ViBe [14], a method that involves comparing each pixel with a set of previous values located the same or neighborhood positions to evaluate whether a pixel belongs to the background. Subsense [15], a recent method following ViBe's strategy to build a non-parametric background model with the Local Binary Similarity Patterns (LBSP) [16] features. However, such local feature based background models is susceptible to be affected by the dynamic motion of background, thus losing the robustness.

Pattern classification technologies based on convolutional neural network (ConvNets) have also been used in background subtraction [17,18]. For example, Braham et al. [18] presented an algorithm based on spatial features learned with ConvNets for their scene-specific backgrounds. These methods using ConvNets

[*] Corresponding author.
E-mail address: zhouwenjun@hce.ist.hokudai.ac.jp (W. Zhou).

https://doi.org/10.1016/j.sigpro.2019.02.021
0165-1684/© 2019 Elsevier B.V. All rights reserved.



(a) Illumination changes,

(b) Background motion

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

can deal well with complicated backgrounds. However, a substantial amount of labeled data with teachers signals or ground truth is necessary for their training, which are generally of high cost and may not always be available [19]. In contrast, CPB is low cost in collecting the training data without any teachers signals.

To overcome above problems, this paper proposed a robust foreground detection method called as CPB against strong background changes, which has already been described in [20]. And in order to further resist background changes for foreground detection, we introduce the Hypothesis on Degradation Modification (HoD, which is introduced briefly in [21]) into CPB. Here, we will introduce HoD with new contents and give more detailed explanations for our work with new experimental results. Our contributions of this paper are as follows: (1) on the basis of CPB, this work proposes the Hypothesis on Degradation Modification (HoD) to further improve the robustness of CPB and stabilize the efficiency of CPB in the long-term use process; (2) HoD can help CPB to resist the interference under the adversarial training data [22], that is verified by the experiments in Section 4; (3) for the dynamic scenes, more experiments and analyses further validate the effectiveness of the combination of CPB and HoD.

The main contents of this paper are as follows. Sections 2 introduces the working mechanism of Co-occurrence Pixel-Block Pairs Background Model (CPB). Section 3 gives an introduction and explanation of Hypothesis on Degradation Modification (HoD) in details. Section 4 demonstrates the ability of HoD with experiments and analyzes the experimental results from the dataset of PETS 2001 [23], AIST-Indoor, SBMnet [24] and CDW-2012 [25] to demonstrate the robustness of CPB and CPB+HoD. Discussions and conclusions are described in Section 5 and Section 6, respectively.

2. Methodology

In our previous work [26], we proposed one “pixel to pixel” structure strategy 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, and this strategy is quite effective at dealing with background changes, however it suffers from the open problem of time consumption. In order to handle this problem, we need to find a strategy to avoid the defect of CP3 [26], therefore the approaches using superpixels for cost reduction in follow-up processing and for image matting [27–29] could be helpful. For reducing the processing cost, we introduced a “pixel-to-block” structure into CPB as an extension of “pixel-to-pixel” structure as shown in Fig. 2. Fig. 2, which illustrates the two processes of CPB: training process and detecting process. In this work, we compare the target pixel p with the Q^B as block, and define $\{Q^B_k\}_{k=1,2,\dots,K} = \{Q^B_1, Q^B_2, \dots, Q^B_K\}$ to denote a supporting block set for the target pixel p. We divide each frame (of size $U \times V$) into blocks Q^B of size $m \times n$. Hence, each block Q^B consists of $m \times n$ pixels:

$$Q^B = \left\{\begin{matrix} Q_{11} & Q_{12} & \dots & Q_{1n} \\ Q_{21} & Q_{22} & \dots & Q_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ Q_{m1} & Q_{m2} & \dots & Q_{mn} \end{matrix}\right\}. \tag{1}$$

Similar to [26], we assume that each pixel Q_{mn} that belongs to reference block Q^B is correlated with target pixel p. As a result, we expect one or more blocks Q^B to possess a stable intensity difference $I_p - \bar{I}_Q$ throughout the whole training frames ($\bar{I}_Q$ is the average intensity of block Q^B), even though pixel p and block Q^B can be at quite dissimilar positions, as shown in Fig. 3, in which the size of Q^B is set to 5×5. In theory, since a large part of computation cost can be reduced in the training process, CPB is expected mn times faster in the training than CP3 [26]. 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.

2.1. Supporting blocks selection

In this work, we utilize the Pearson’s product-moment correlation coefficient to select the supporting blocks $\{Q^B_k\}$ for each target pixel p:

$$\{Q^B_k\}_{k=1,2,\dots,K} = \left\{Q^B | \gamma\left(p, Q^B\right) \text{is the } K \text{ highest}\right\}, \tag{2}$$

where

$$\gamma\left(p, Q^B_k\right) = \frac{C_{p,\bar{Q}_k}}{\sigma_p \cdot \sigma_{\bar{Q}_k}} \tag{3}$$

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

In general, we can expect that if the pixel-block pair (p, Q^B_k) keeps a high correlation coefficient, then the supporting block Q_k can provide some reliability to estimate the current state of the target pixel p. Fig. 4 shows example layouts of the supporting

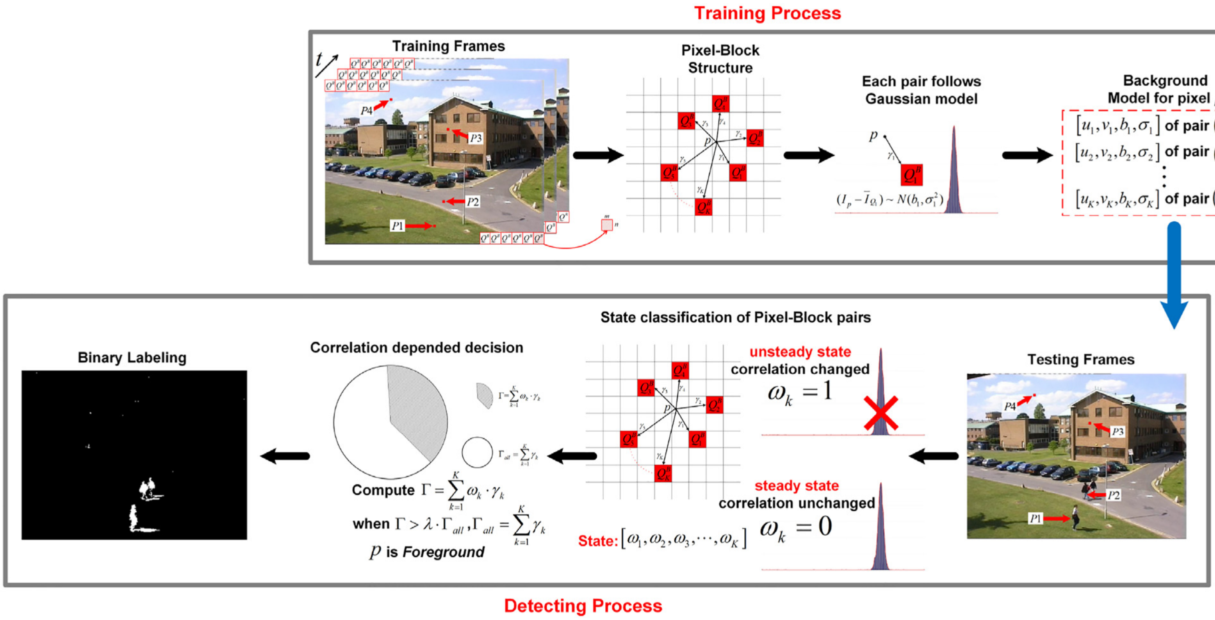

Training Process
Training Frames
t
P4
P3
P2
P1
Pixel-Block Structure
Each pair follows Gaussian model
p
$(I_p - \bar{I}_{Q_1}) \sim N(b_1, \sigma_1^2)$
Background Model for pixel p
$[u_1, v_1, b_1, \sigma_1]$ of pair (p, Q_1^B)
$[u_2, v_2, b_2, \sigma_2]$ of pair (p, Q_2^B)
$[u_K, v_K, b_K, \sigma_K]$ of pair (p, Q_3^B)
State classification of Pixel-Block pairs
Binary Labeling
Correlation depended decision
$\Gamma = \sum_{k=1}^{K} \omega_k \cdot \gamma_k$
$\Gamma_{all} = \sum_{k=1}^{K} \gamma_k$
Compute $\Gamma = \sum_{k=1}^{K} \omega_k \cdot \gamma_k$
when $\Gamma > \lambda \cdot \Gamma_{all}, \Gamma_{all} = \sum_{k=1}^{K} \gamma_k$
p is Foreground
unsteady state
correlation changed
$\omega_k = 1$
steady state
correlation unchanged
$\omega_k = 0$
State: $[\omega_1, \omega_2, \omega_3, \cdots, \omega_K]$
Testing Frames
Detecting Process

Fig. 2. Overview of working mechanism of CPB.

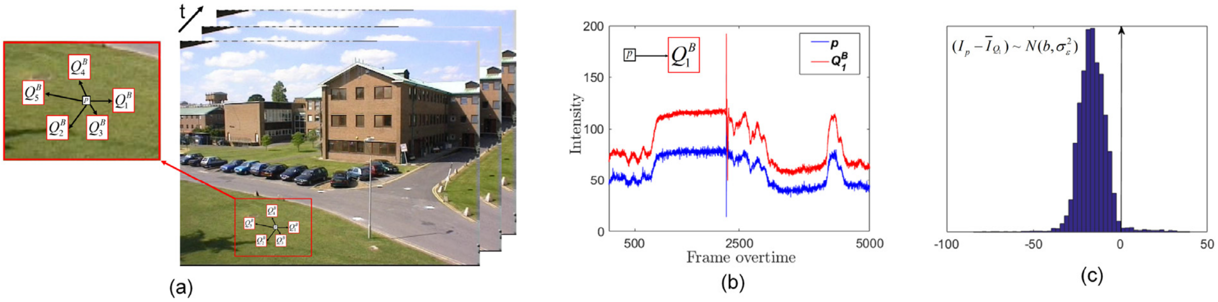


Fig. 3. Basic structure of co-occurrence pixel to block pair. (a) Co-occurrence pixel-block pair structure. (b) Correlation of pixel-block pair (p, Q_1^B). (c) Statistical model of pixel-block pair (p, Q_1^B).

blocks using $PETS2001 - dataset3 - camera1$ and the target pixels are selected from the four representative regions: “Grass,” “Road,” “Building,” “Sky,” respectively.

2.2. Co-occurrence background model

The work builds a co-occurrence model using the single Gaussian distribution for the selected K pixel-block pairs:

$$\Delta_k \sim N(b_k, \sigma_k^2) \quad \Delta_k = I_p - \bar{I}_{Q_k}, \tag{4}$$

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:

$$b_k = \frac{1}{T} \sum_{t=1}^{T} \Delta_k \tag{5}$$

and the variance estimation is defined as follows:

$$\sigma_k^2 = \frac{1}{T} \sum_{t=1}^{T} (\Delta_k - b_k)^2, \tag{6}$$

where T is the sequence of frames. Through the training process, the parameters σ_k and b_k are recorded as a model description for the next detecting stage and then the background model is built as a list consisting of $[u_k, v_k, b_k, \sigma_k]$ for supporting block set $\{Q_k^B\}_{k=1,2,\ldots,K}$, where (u_k, v_k) is the coordinate of supporting block.

2.3. 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. 5. 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. This knowledge can be realized as the following: the state F (unsteady) means p may be occluded by any foreground object, while the state B (steady) means that p may be exposed to the camera as it has been in the statistical training frames. In order to obtain any difference between these two states, for each pixel p, we introduce an index value as a “penalty” for violating the relationships authorized at the statistical training process. In other words, if the state F is associated with pixel p and the pixel value may also be changed, therefore we can utilize statistical tests in which the difference may belong to the registered distribution or be rejected as a value outside of the distribution.

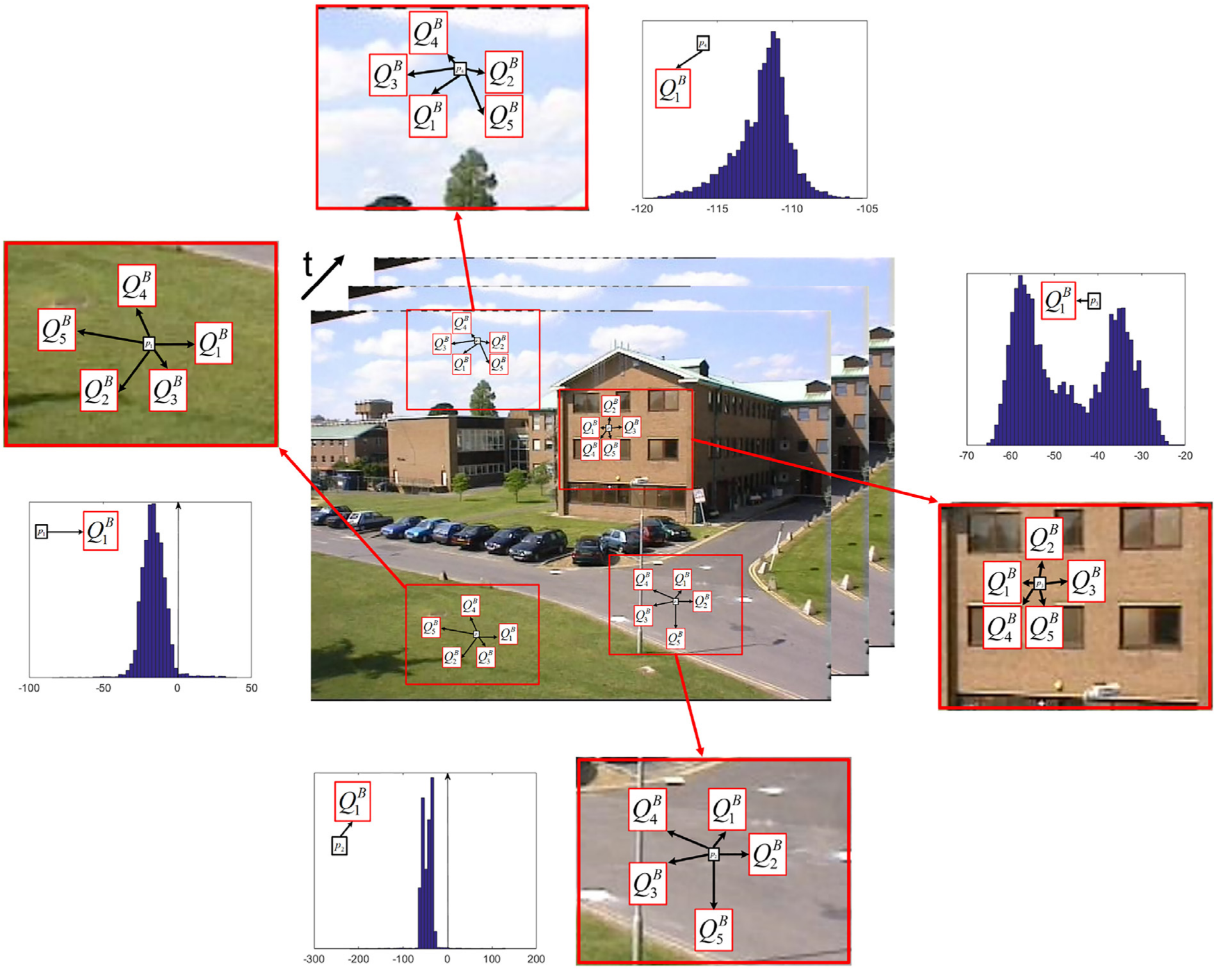


Fig. 4. Example layouts of pixel-block pairs for different position pixels $p_1(256, 483)$, $p_2(551, 432)$, $p_3(435, 168)$ and $p_4(250, 41)$, respectively, where $K = 5$ and the size of each block is 5×5, and examples of the correlation of pairs at different position, respectively.

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

$$\omega_k = \begin{cases} 1 & \text{if } \left|(p - Q_k^B) - b_k\right| \geq \eta \cdot \sigma_k \\ 0 & \text{otherwise} \end{cases}, \tag{7}$$

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 [26,30]. The definition is realized as Γ as follows:

$$\Gamma = \sum_{k=1}^{K} \omega_k \cdot \gamma_k \tag{8}$$

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 the 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.

$$\Gamma_{all} = \sum_{k=1}^{K} \gamma_k. \tag{9}$$

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 described in [20] with details. The decision function is shown in pseudo-code in Algorithm 1.

2.4. Training data selection

In this section, we want to give some comments on how to select the training data as an important step in our CPB's mechanism. We need an enough set of suitable data for training, and then CPB may train itself properly to detect expected foreground pixels. It has been a common and important problem in the algorithms that need any training data, such like IMBS [31] or SuBSENSE [15] as Fig. 12 in Section 4.2 to do this preparation. In this paper, since we use many databases that have their own ground truth frames and therefore we can see some types of the expected foreground pixels, such as walking peoples or vehicles, it is possible to select some frames as the training data, which do not include any excessive foreground pixels. But in any real tasks in which it is not reality to take high cost for making effective ground truth data, one may

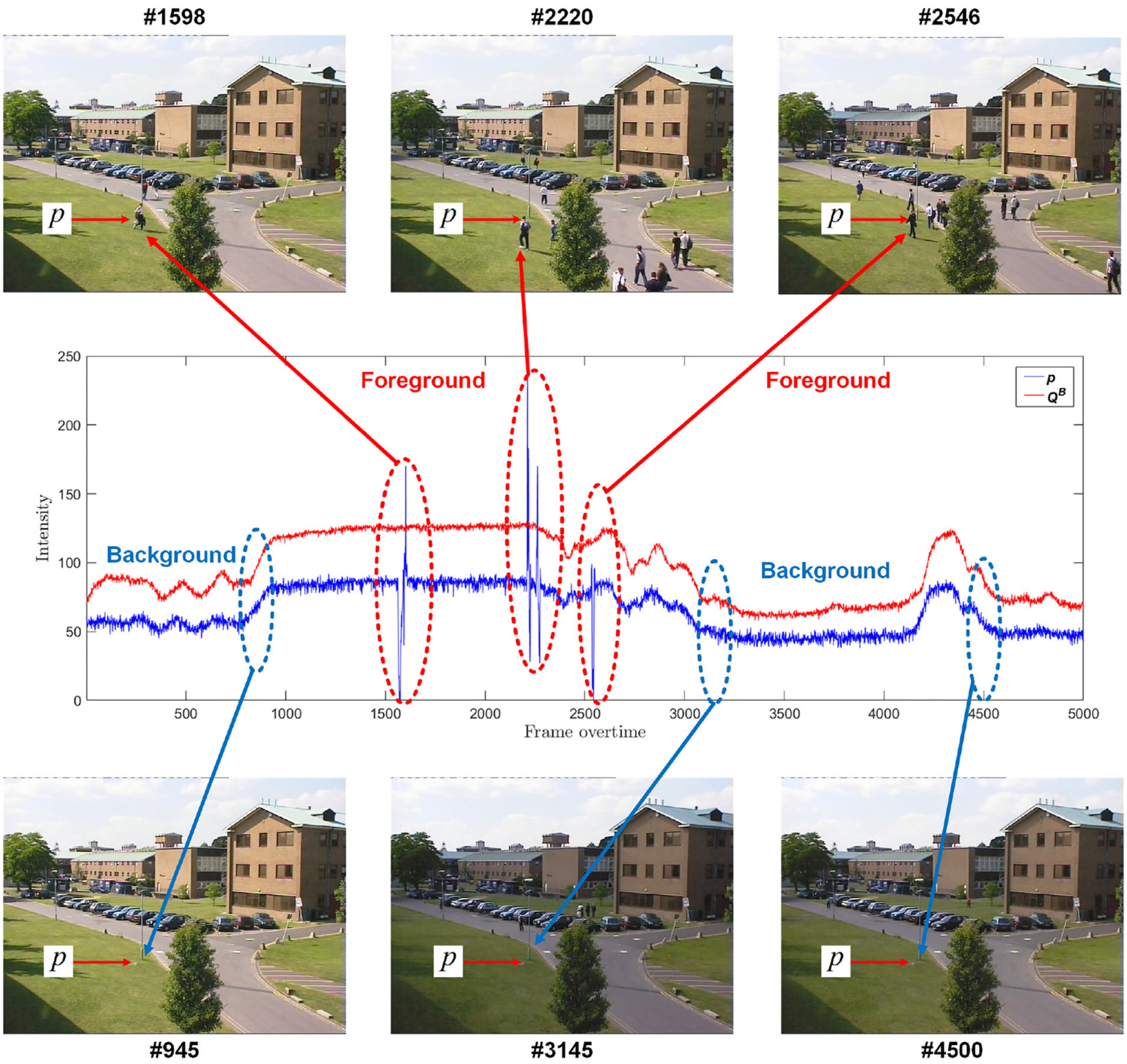


Fig. 5. Co-occurrence intensity changes between target pixel p and supporting block Q^B overtime.

have to make the training data or frames through implicit definition of foreground pixels and selecting the proper frames. Therefore, in our experimental demonstration, we prudently select the frames for training to avoid the emergence of adversarial data.

3. Hypothesis on degradation modification

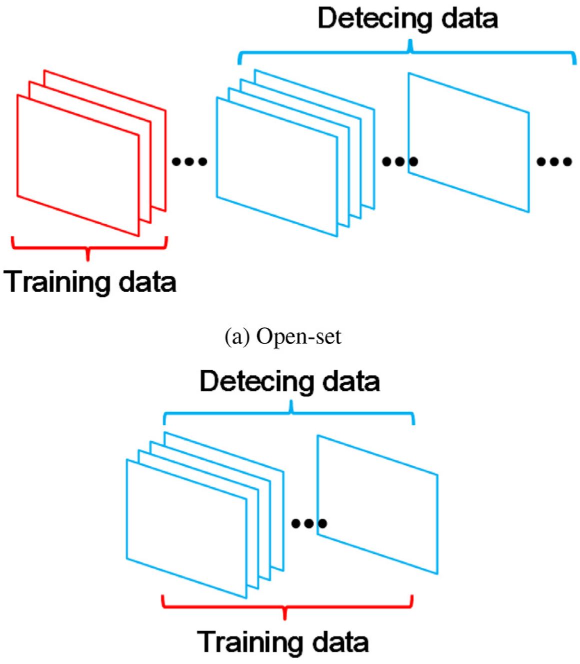
We have introduced the basic CPB algorithm for robust background subtraction, however, where the data for training and detecting are prepared in advance of the operations. In this work, we define two types of problems *open-set* (generalization problem), which is shown in Fig. 6 (a), where the data in training are known in advance but the data for detecting are unknown, continuable, and different from the training data. *Open-set* is different from the type *closed-set* (classification problem), where the detecting data can be selected from the same set of the training data and the also can be known in advance as shown in Fig. 6 (b). We may have some mechanism to modify the model to fix some errors which may be observed in *open-set* condition. In general, hypothesis of this paper follows two significances: (1) we assume that some “noise” may arise in detecting process due to a long time usage of initial CPB background model and we can not confirm such “noise” is true or not without any ground truth for verification in real applications; (2) second, we assume that after a prolonged using, the initial “Pixel to Block” structure can no longer adapt to the current, then resulting in errors. Then, based on the above assumptions, in this section, we intend to introduce a simple mechanism named Hypothesis on Degradation Modification (HoD) extended from CPB to adapt the background changes and reinforce the robustness of CPB to resist the “noise” in real applications.

3.1. Hypothesis on degradation

In practice, after a long time utilization of initial CPB background model in an unlearned sequence, the expected relative 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. 7 shows an example of the

```
Input:
    Testing frame; Parameters: $\eta$ and $\lambda$ ;
Output:
    The state of pixel $p$;
    for each pixel $p$ do
        Load $[\mu_k, \nu_k, b_k, \sigma_k]$ for each pair $(p, Q_k^B)$;
        Estimate the state of each pair;
        for $k = 1, 2, \dots, K$ do
            if $\left|(p - Q_k^B) - b_k\right| \geq \eta \cdot \sigma_k$ then
                $\omega_k = 1$
            else
                $\omega_k = 0$
            end if
        end for
        compute $\Gamma$
        if $\Gamma \geq \lambda \cdot \Gamma_{all}$, where $\Gamma_{all} = \sum_{k=1}^{K} \gamma_k$ then
            $p$ is foreground
        else
            $p$ is background
        end if
        return the state of pixel $p$
    end for
```

Algorithm 1. Correlation dependent decision.



(a) Open-set

(b) Closed-set

Fig. 6. Descriptions of open-set and closed-set conditions.

hypothetical noise using *copyMachine* from CDW-2012 dataset [25]. 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. 8 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. Moreover, we also can clearly notice that HoD is a self-checking mode, which is completely different from the retraining mode. In HoD mode, it costs less time and consumes less data cost over a period of usage, and is more efficient than the retraining mode.

3.2. Broken pixel-block pairs detection

As shown in Fig. 8, 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 2.1) 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 on NaE, whereas the pairs with large γ in steady state might cause a decision on NaB. With the above assumption, we propose a weight-based decision function to detect the broken pair:

$$\textit{if}\ \ \gamma_m \geq \bar{\gamma},\ \ \textit{then}\ (p, Q_m^B)\ \textit{is broken} \tag{10}$$

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:

$$\bar{\gamma} = \frac{1}{M}\sum_{k=1}^{K} \gamma_k \cdot \omega_k = \frac{1}{M}\Gamma. \tag{11}$$

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

$$\bar{\gamma} = \frac{1}{K-M}\sum_{k=1}^{K} \gamma_k \cdot (1 - \omega_k) = \frac{1}{K-M}(\Gamma_{all} - \Gamma). \tag{12}$$

There is a slight difference in the above definitions, and then we record these broken pairs for the next process. This process is shown in pseudo-code in Algorithm 2.

3.3. 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

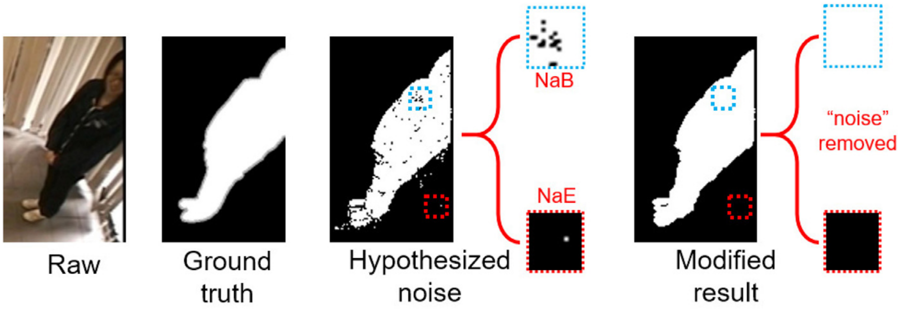


Fig. 7. Description of hypothesized noise.

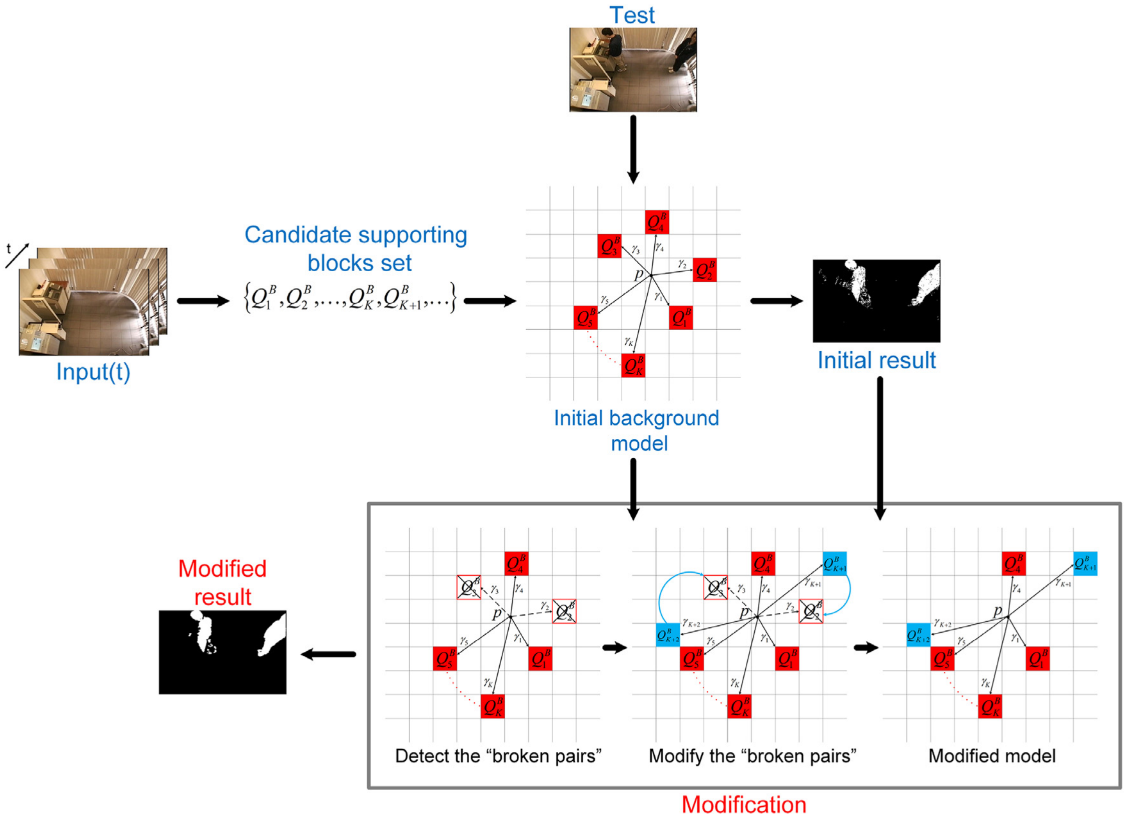

Test
t
Input(t)
Candidate supporting blocks set
$\{Q_1^B, Q_2^B, \ldots, Q_K^B, Q_{K+1}^B, \ldots\}$
Initial background model
Initial result
Modified result
Detect the "broken pairs"
Modify the "broken pairs"
Modified model
Modification

Fig. 8. Overview of HoD Modification.

```
Input: Initial model structure of the hypothetical noise $p$;
Output: Broken pixel-block pairs
  Detect the broken pixel-block pairs;
  for each pair $(p, Q_m^B)$ do
    if $\gamma_m \geq \bar{\gamma}$ then
      $(p, Q_m^B)$ is broken pair
    else
      $(p, Q_m^B)$ is stable pair
    end if
    return the state of pair $(p, Q_m^B)$
  end for
  Record the broken pixel-block pairs.
```

Algorithm 2. Broken pixel-block pairs detection.

hypothesized noise by using the modified pixel-block structure as shown in Fig. 8.

In general, HoD is a new strategy for the background model update. It is not only applied on the top of CPB, but also can be applicable to the other pixel-correlation based algorithms (such as ViBe [14] based on random neighboring pixels, SuBSENSE [15] based on local binary similarity patterns features or our previous work CP3 [26]), HoD provides a new and natural thought: the structure of background model can be updated by the designed correlation weight, which is discussed in details in Section 3 and the validity of HoD is proved in Section 4.1 and 4.2.

4. Experiments

4.1. Verification of HoD's performance

In order to verify the performance of HoD in *open-set* condition, we compare the results of CPB and CPB+HoD in the sequence *canoe*, which is a typical scene with rippling water [25]. In this experiment, 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. 9 shows the typical results of CPB and CPB+HoD and Fig. 10 illustrates the *F-measure* and *False Positives* comparison between CPB and CPB+HoD in the sequence *canoe* overtime. From Figs. 9 and 10, it is clear that with the help of HoD, CPB+HoD has a significant improvement over CPB and further restrained the noise in scene. Table 1 illustrates a change in False Positives of ♯860 and ♯900 between CPB and CPB+HoD, from the table we can note that HoD greatly restrains the noise in dynamic scene. These results suggest that HoD can effectively suppress the degradation in CPB with the passage of time.

Table 1
A change in False Positives of ♯860 and ♯900.

Methods	The number of False Positives	
	♯860	♯900
CPB	320	350
CPB+HoD	**1**	**33**

4.2. Ability of HoD under adversarial data

To get an idea of what adversarial looks like, consider one demo from "Explaining and Harnessing Adversarial Examples" [22]: inputting a panda image, and adding some perturbation that has been evaluated to make the image be recognized as a gibbon with high confidence. Similarly, we can define the following training data as the adversarial data:

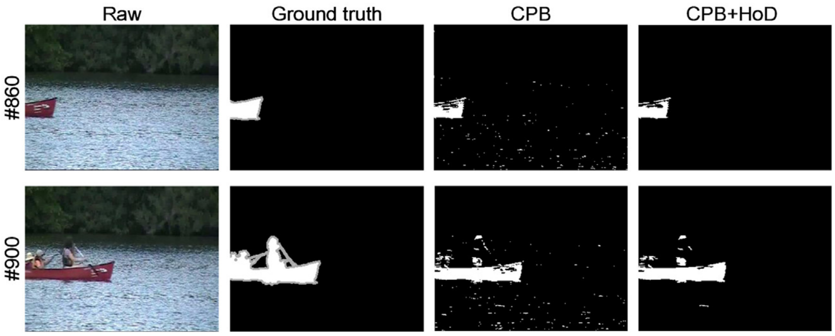
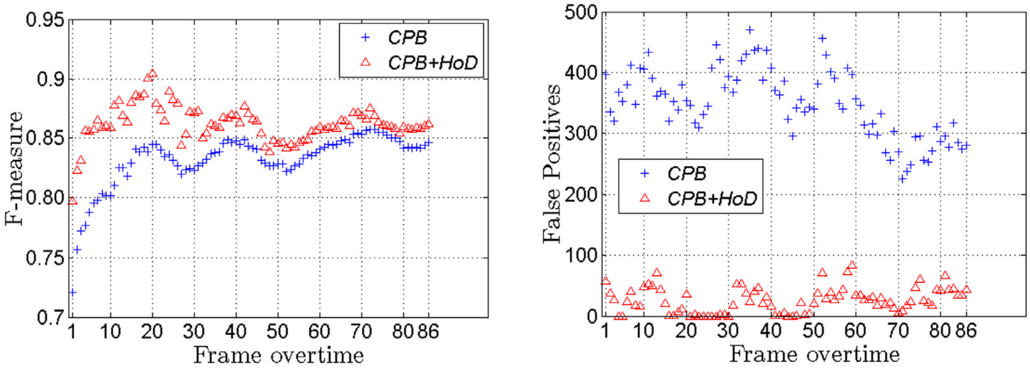


Fig. 9. Typical results for CPB and CPB+HoD in the sequence *canoe*.



(a) F-measure comparison

(b) False Positives comparison

Fig. 10. Comparison of CPB and CPB+HoD in the sequence *canoe* overtime.

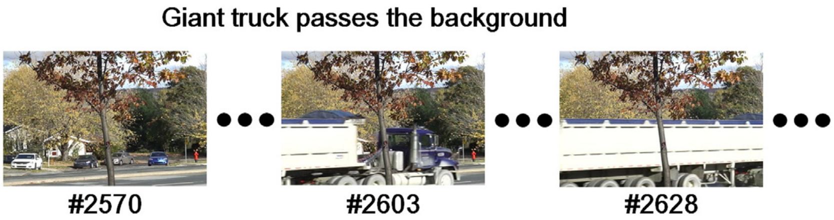


Fig. 11. A typical example of adversarial data: giant truck passes the background.

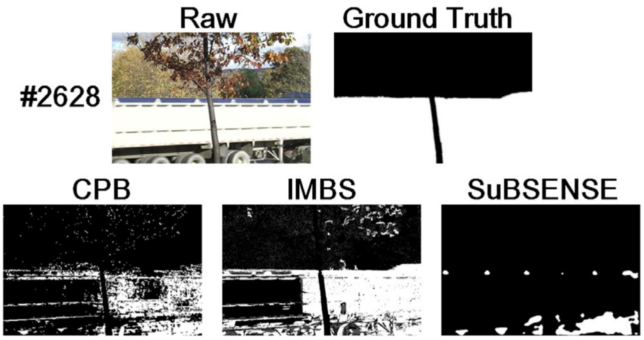


Fig. 12. Influence of adversarial data on the detection.

- training data includes a high-density crowd or large-scale object;
- foreground information is mixed with or even covers background.

Fig. 11 shows a typical example of adversarial data, which is from the sequence *fall* with swaying branches in the CDW-2012 dataset [25]. The giant truck passes the background and covers half of the background information. The typical results under adversarial training are shown in Fig. 12. In this case, the training data includes 150 frames (#2460-#2609, selected from the sequence *fall*): (a) 120 frames without any large-scale objects (#2460-#2579); (b) 30 frames with a giant truck (#2580-#2609) and the interference rate is 20 %, we define the interference rate as the percentage of adversarial frames to total frames.

As mentioned in Section 2.4, CPB is not good at the adversarial data. However, HoD can help CPB to resist the interference from the adversarial data and we design the experiments to verify the ability of HoD and the details are presented in Table. 2. Fig. 13 shows the typical results of CPB and CPB+HoD and Fig. 14 shows the *F-measure* comparison between CPB and CPB+HoD in the six different cases.

We note that CPB will lose the efficiency as the interference increases. However, HoD can help CPB to resist the interference, which is demonstrated in Figs. 13 and 14, because HoD can repair and stabilize the initial model structure of CPB by selecting new pixel-block pairs from the candidate supporting block set as described in Section 3. The results demonstrate the ability of HoD under the adversarial data.

Table 2
Experimental design under the adversarial training data from the sequence *fall*.

Case	Interference rate	Frames without any large-scale objects	Frames with a giant truck	Total number of frames
1	5%	#2010-#2579 (570 frames)	#2580-#2609 (30 frames)	600 frames
2	10%	#2310-#2579 (270 frames)		300 frames
3	15%	#2410-#2579 (170 frames)		200 frames
4	20%	#2460-#2579 (120 frames)		150 frames
5	25%	#2490-#2579 (90 frames)		120 frames
6	30%	#2510-#2579 (70 frames)		100 frames

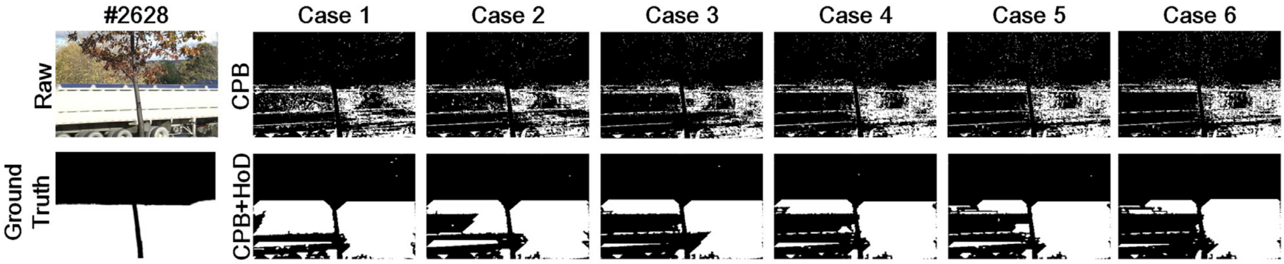


Fig. 13. Typical results of CPB and CPB+HoD in different interference cases.

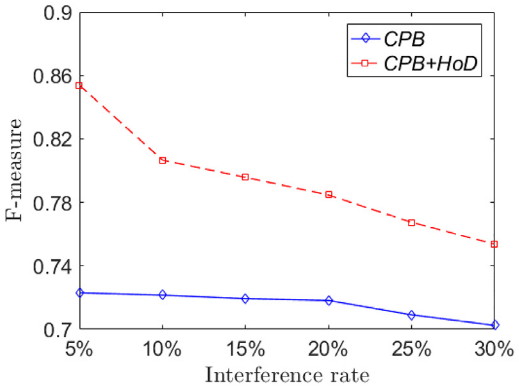


Fig. 14. Comparison of CPB and CPB+HoD in six interference cases.

4.3. Experimental setup

Considering the several challenges of video surveillance for background subtraction algorithm [32]. We consider the following datasets to evaluate the proposed methods:

- **PETS2001 dataset** [23]: one typical sequence of gradual illumination changes.
- **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.
- **SBMnet dataset** [24]: one sequence *advertisement Board* with strong background motion is selected from SBMnet dataset for testing, and this sequence contains an ever-changing advertising board in the scene.
- **CDW-2012 dataset** [25]: one typical sequence *canoe* with water rippling is selected from the CDW-2012 dataset.

We compare the proposed CPB and CPB+HoD with six different foreground detection techniques: GMM [11] and KDE [12], which are two well-known traditional algorithms, and four state of the art techniques IMBS [31], T2FMRF-UV [33], ViBe [14] and SuBSENSE [15].

Table 3
Parameters setting of CPB.

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

At first, GMM [11] and KDE [12] are two main basic standard techniques that are often used to make the basic comparison [14,34–36]. Second, the state of the art techniques IMBS [31] and T2FMRF-UV [33] are the foreground extraction techniques specifically for dynamic background. And then, ViBe [14] and SuBSENSE [15], which are two of the leading unsupervised techniques for foreground detection, especially SuBSENSE [15] is one of the top-ranked techniques in CDW-2012 dataset at present. Based on the above reasons, we select these six different techniques for comparative experiments.

In contrast to the methods with complex strategies [15,31,33], CPB is a low-complexity algorithm that is more easily realized. The parameters for GMM, KDE, IMBS, T2FMRF-UV, ViBe and SuBSENSE were set by using the tool bgslibrary [37]. In experiments, we set each block as 8×8 pixels for CPB, the parameters are shown in Table. 3 and have been discussed in [20] how to decide.

In order to evaluate the methods in pixel level, we utilize three common analysis measurements [38–40]: *Precision*, *Recall*, and *F-measure*. These metrics are widely used to estimate the quality of background subtraction methods [10,32],

$$Precision = \frac{TP}{TP + FP}, \tag{13}$$

and

$$Recall = \frac{TP}{TP + FN}, \tag{14}$$

where *TP*, *FP* and *FN* indicate the number of true positives, false positives and false negatives, respectively. Meanwhile, we use *F-measure* as the harmonic mean of *Precision* and *Recall*,

$$F - measure = \frac{2Precision \cdot Recall}{Precision + Recall}. \tag{15}$$

For further evaluating our CPB and CPB+HoD, we introduce the peak signal-to-noise ratio (PSNR) as our metric [41,42], which can be used o measure the quality of the estimated resulted compared with the background truth [43]. The definition of *PSNR* is

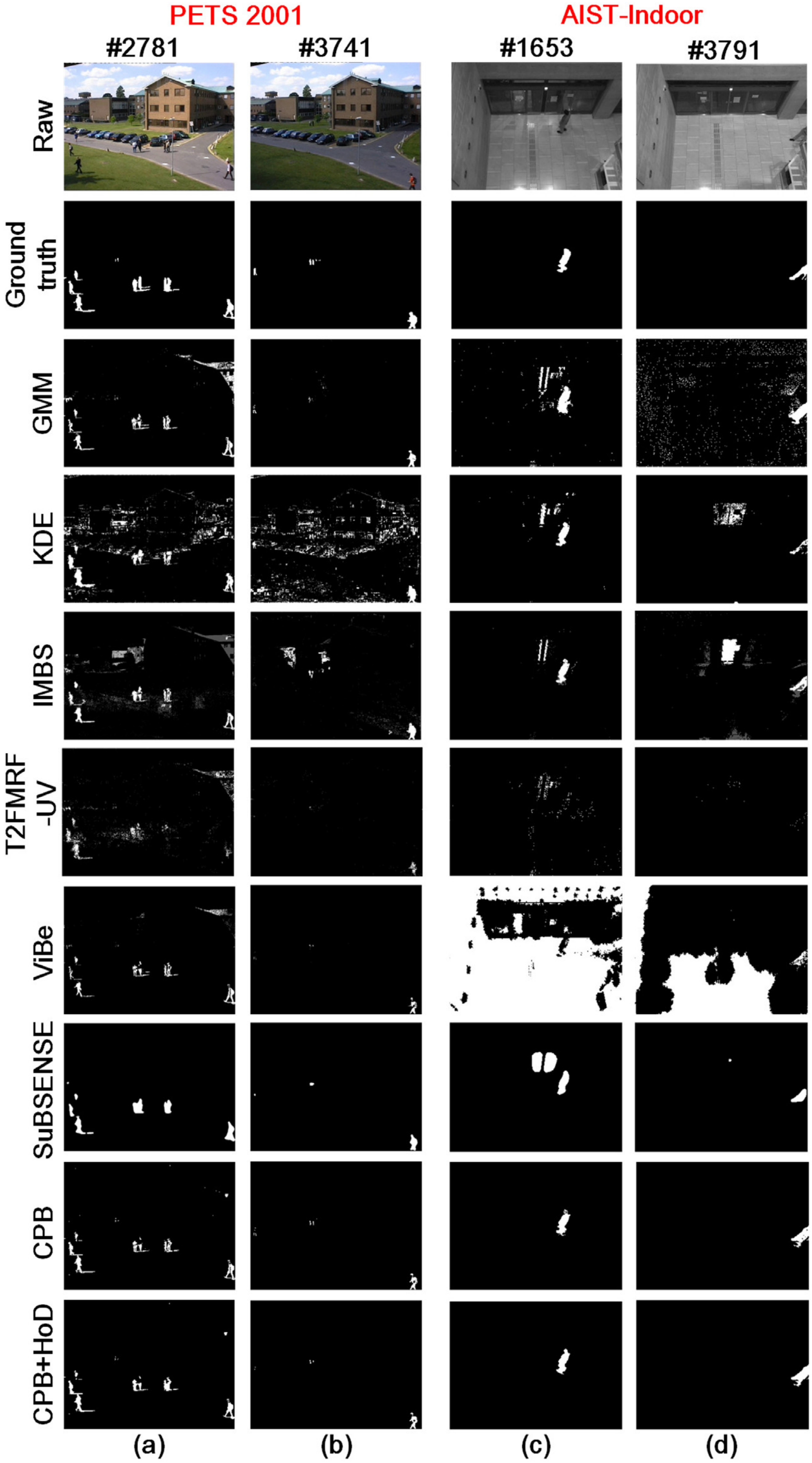


Fig. 15. Representative results from the illumination change challenges: (a) illumination becomes stronger in daylight; (b) illumination becomes lower in daylight; (c) automatic door suddenly opens and the light changes; (d) person suddenly enters the scene, and the light switches on automatically.

calculated as follows:,

$$PSNR = 10 \cdot \log_{10}\left(\frac{255^2}{MSE}\right), \tag{16}$$

where *MSE* is the mean square error.

4.4. Experimental comparison with other algorithms

Experimental results of the foreground detection are presented in Figs. 15 and 16. Tables 4 and 5 list 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.

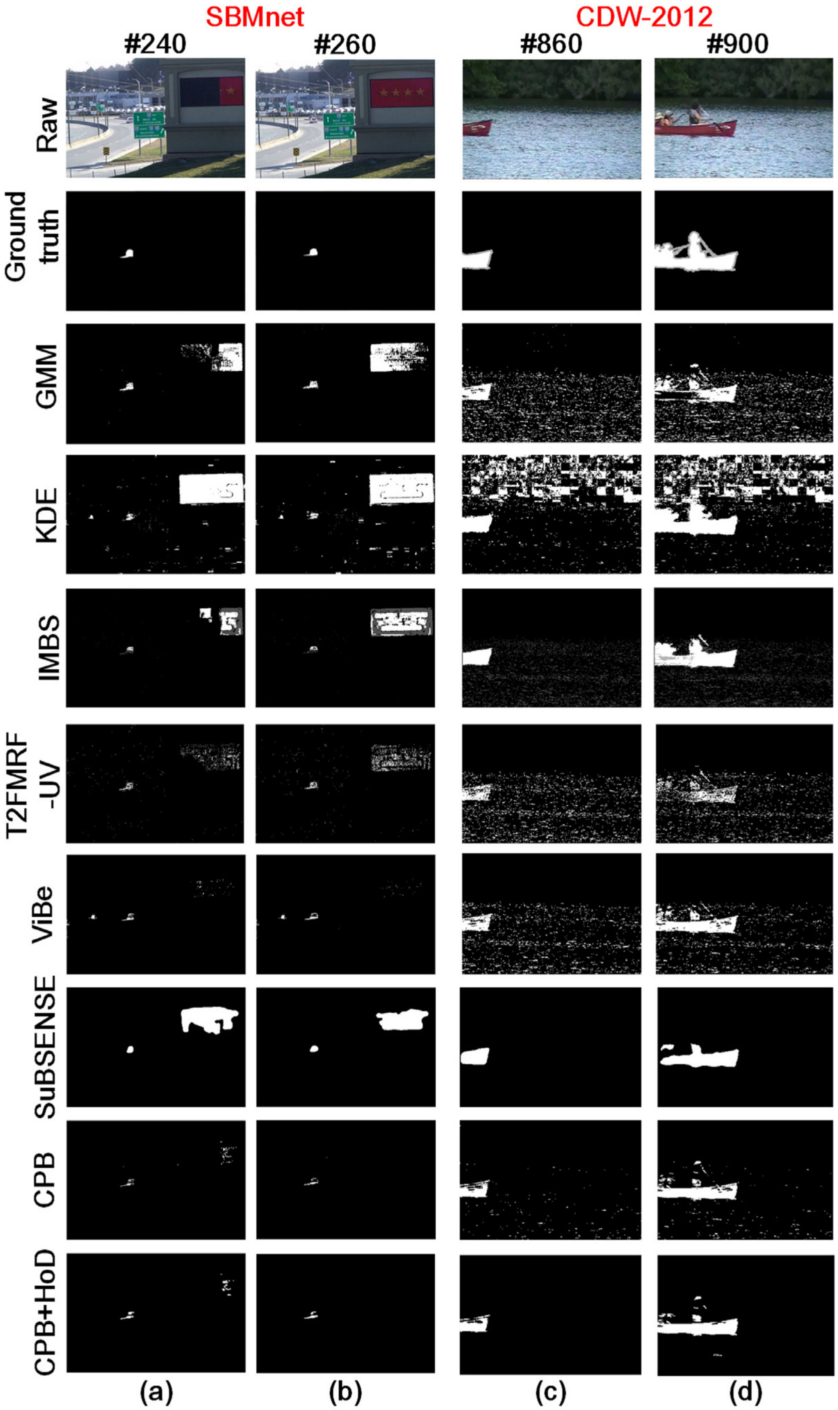


Fig. 16. Representative results from background motion challenges: (a) advertisement board starts to change; (b) advertisement board stops changing; (c) canoe enters the scene; (d) canoe continues to move.

- **Illumination changes:** Fig. 15 shows the illumination change challenges, which are gradual illumination changes (global illumination changes) and sudden illumination changes (local illumination changes). The results demonstrate that our methods work well during illumination changes, especially sudden illumination changes. Here, we explain the difference in our model. For example, ViBe [14] is based on an assumption that the correlation of pixels, that is depended on the distance in spatial between them (e.g. the LBP feature in SuBSENSE [15], where the target pixel has a high correlation with its neighboring pixels). However, this mechanism ignores the localized relation between each pixel, and the detection is insensitive and cannot adapt to local illumination changes as shown in Fig. 15. In CPB, due to the multiple supporting blocks for each target pixel, the co-occurrence pixel-block pairs build a multiple and spatial structure; thus, this structure maintains a stable statistical correlation more steadily for each target pixel and abandons the prior assumption of local correlation. This is why, CPB can extract the foreground sensitively under both global and local illumination changes as shown in Table 4.

Table 4
Performance evaluation for foreground detection during illumination changes.

Datasets	PETS 2001				AIST			
Methods	Precision	Recall	F-measure	PSNR	Precision	Recall	F-measure	PSNR
GMM [11]	0.6465	**0.9508**	0.7697	39.46	0.6523	**0.9207**	0.7636	40.57
KDE [12]	0.5181	0.8836	0.6531	17.77	0.5896	0.6944	0.6377	38.16
IMBS [31]	0.5162	*0.8841*	0.6518	16.20	0.5760	0.6923	0.6288	36.36
T2FMRF-UV [33]	0.5818	0.8365	0.6863	34.94	0.6382	0.5818	0.6087	45.65
ViBe [14]	0.7059	0.8821	0.7842	43.42	0.5005	0.5146	0.5074	9.11
SuBSENSE [15]	0.9008	0.8840	**0.8923**	54.11	0.5864	0.7047	0.6401	37.14
CPB	*0.9566*	0.7517	0.8418	*56.05*	*0.8651*	0.8181	*0.8409*	*53.14*
CPB+HoD	**0.9652**	0.7562	*0.8480*	**56.39**	**0.8668**	*0.8227*	**0.8442**	**53.31**

* Note that **red entries** indicate the best in measurement, and *blue entries* indicate the second best.

Table 5
Performance evaluation for foreground detection during background motion.

Datasets	SBMnet				CDW-2012			
Methods	Precision	Recall	F-measure	PSNR	Precision	Recall	F-measure	PSNR
GMM [11]	0.5151	0.5196	0.5174	26.92	0.6748	0.7024	0.6883	21.71
KDE [12]	0.4962	0.4856	0.4909	21.67	0.6584	*0.8630*	0.7468	17.22
IMBS [31]	0.5095	0.5118	0.5107	30.09	0.7315	**0.8911**	0.8035	21.60
T2FMRF-UV [33]	0.5508	0.5179	0.5338	35.38	0.6797	0.6114	0.6438	23.49
ViBe [14]	0.6427	**0.5368**	0.5850	35.16	0.8114	0.7821	0.7965	28.02
SuBSENSE [15]	0.5018	0.5033	0.5025	27.62	*0.9766*	0.7649	*0.8573*	30.80
CPB	*0.7653*	0.5118	*0.6133*	*36.64*	0.9283	0.7730	0.8436	*32.61*
CPB+HoD	**0.7973**	*0.5214*	**0.6350**	**37.39**	**0.9809**	0.7830	**0.8708**	**34.16**

* Note that **red entries** indicate the best in measurement, and *blue entries* indicate the second best.

- **Background motion:** Fig. 16 also shows two background motion challenges, which are sudden changes in background like a continuously changing advertising board in the scene and regular movement like rippling water. Video sequences contain the temporal context information and our CPB model can learn this information from the training data to avoid interference from background information such as background motion, during the detection process, and then accurately extract the current foreground information (*object*). This is different from the approaches based on local features (e.g., SuBSENSE [15] or ViBe [14]), which cannot adapt in non-ideal cases, for example, where textures are missing or there is a dynamic background. Based on this knowledge, our model can handle both of the changes well, and outperforms other methods significantly for sudden background motion as shown in Table 5.

4.5. Computational cost

This section, we compare the processing time of our proposed methods with others in terms of fps. We evaluate the time required in foreground detection with the tool in the MATLAB platform (Intel E3 3.5GHZ and 16G) and utilize the testing frames from *canoe* [25] (frame size: 320 × 240). From the above results in Table. 6, observing that our methods lead an intermediate level in the detection process. CPB does not dominate in detecting time as illustrated in the Table. 6. Because of the multiple “pixel to block” structure of CPB, it takes some time to estimate the current state of the target pixel p as discussed in [20]. For each pair (p, Q^B), we define all the pixels Q_{mn} of block Q^B follow the Eq. (7), then we can estimate the current state of the pair (p, Q^B) as described in Section 2.3. Based on this mechanism, it takes time on detection. To solve this problem, on the one hand, we can appropriately reduce the size of block to achieve the reduction in detecting time. On the other hand, we would like to introduce the parallel processing implement into our detection, which is also employed in [44]. For example, we can divide the current input frame into N non-overlapping regions based on the number of available CPU cores. Then, detecting the foreground pixels of each individual region instead of the full scene detection. We would like to optimize the program to further reduce the processing time on detection in the future.

Table 6
Processing time comparison in FPS.

Methods	Processing speed
GMM	81
KDE	69
IMBS	33
T2FMRF-UV	60
ViBe	149
SuBSENSE	14
CPB	**30**
CPB+HoD	**27**

5. Discussion

Compared with the classification algorithms based on ConvNets, CPB is a simple algorithm in training data preparation. CPB is a statistical model based on the extraction of background information to distinguish the outliers (i.e., foreground) from background. Quite different from the ConvNets based algorithms, it doesn’t need any labeled data (separate background and foreground data) for training. Hence, CPB is low cost in training data preparation. However, CPB has its own disadvantage in training data selection. For instance, CPB cannot deal well with the adversarial data if the training data includes a high-density crowd or large-scale object as described in Section 4.2. Because CPB needs to learn the background information to build the initial background model, which is a single Gaussian model based on the training data. If the foreground information severely interferes the background, that will lead a faulty background model for CPB. Therefore, we must be careful to select the training data for CPB.

In order to reinforce the robustness of CPB, we propose the Hypothesis on Degradation Modification (HoD) and the experimental results demonstrate the performance of HoD. One can introduce

HoD for this purpose at any timing when it is needed for repairing or fixing broken models. In general, one can estimate the timing by checking the amount of detected noises in the frames and comparing them with the normal frequency of the detection. Another problem of HoD is that it may lose the effectiveness in some specific applications, such like small object detection. For example, in the honeybees detecting and tracking [45], if we employ HoD to modify the initial result, that may remove the true target (honeybees) to lead a wrong modification. Because, small objects are too similar to the hypothesized noise as defined in Section 3, in this case, HoD may detect the foreground as the hypothesized noise and then lead a wrong modification. HoD may not so effective in such applications in reality.

6. Conclusions

In this paper, we developed a prospective background model with hypothesis on degradation modification (HoD) for foreground detection under dynamic scenes. It was designed to handle the problem of strong background changes in reality. With the help of HoD, we further improve the robustness of CPB and stabilize the effectiveness in the long-term use. And HoD also can help CPB to resist the interference under the adversarial data. Experimental results from different challenges show the interest of proposed method. For foreground detection in dynamic scenes, our method outperforms other methods significantly in the most challenging sequences. This background model performances a fairly good detection under extreme environments such as illumination changes and background motion. Furthermore, as discussed in the paper, HoD provides a new and natural thought: the structure of background model can be updated by the designed correlation weigh, which is a new strategy can be utilized in the pixel-correlation based algorithms for the background model update. Our future work would like to develop an on-line mode of CPB structure by using the hypothesis on degradation modification (HoD).

Conflict of interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled, “Foreground Detection based on Co-occurrence Background Model with Hypothesis on Degradation Modification in Dynamic Scenes” (ID: No. SIGPRO-D-18-00969).

References

[1] T. Bouwmans, Traditional and recent approaches in background modeling for foreground detection: an overview, Comput. Sci. Rev. 11 (2014) 31–66.

[2] T.B. Moeslund, A. Hilton, V. Krüger, A survey of advances in vision-based human motion capture and analysis, Comput. Vis. Image Understand. 104 (2) (2006) 90–126.

[3] S.-C.S. Cheung, C. Kamath, Robust background subtraction with foreground validation for urban traffic video, EURASIP J. Adv. Signal Process. 2005 (14) (2005). 726–261

[4] L. Maddalena, A. Petrosino, Towards benchmarking scene background initialization, in: International Conference on Image Analysis and Processing, Springer, 2015, pp. 469–476.

[5] T. Bouwmans, L. Maddalena, A. Petrosino, Scene background initialization: a taxonomy, Pattern Recognit. Lett. 96 (2017) 3–11.

[6] A. Yilmaz, O. Javed, M. Shah, Object tracking: a survey, Acm Comput.Surv. (CSUR) 38 (4) (2006) 13.

[7] V. Mahadevan, W. Li, V. Bhalodia, N. Vasconcelos, Anomaly detection in crowded scenes, IEEE, Comput. Vis.Pattern Recognit. (CVPR), 2010 IEEE Conf. on, 2010, pp. 1975–1981.

[8] M. Piccardi, Background subtraction techniques: a review, 4, IEEE, Systems, Man and Cybernetics, 2004 IEEE International Conference on, 2004, pp. 3099–3104.

[9] L. Li, W. Huang, I.Y.-H. Gu, Q. Tian, Statistical modeling of complex backgrounds for foreground object detection, IEEE Trans. Image Process. 13 (11) (2004) 1459–1472.

[10] A. Vacavant, T. Chateau, A. Wilhelm, L. Lequièvre, A benchmark dataset for outdoor foreground/background extraction, Asian Conf. Comput. Vis. (2012) 291–300.

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

[12] A. Elgammal, R. Duraiswami, D. Harwood, L.S. Davis, Background and foreground modeling using nonparametric kernel density estimation for visual surveillance, Proc. IEEE 90 (7) (2002) 1151–1163.

[13] P.-M. Jodoin, M. Mignotte, J. Konrad, Statistical background subtraction using spatial cues, IEEE Trans. Circuit Syst. Video Technol. 17 (12) (2007) 1758–1763.

[14] O. Barnich, M. Van Droogenbroeck, Vibe: a universal background subtraction algorithm for video sequences, IEEE Trans. Image Process. 20 (6) (2011) 1709–1724.

[15] P.-L. St-Charles, G.-A. Bilodeau, R. Bergevin, Subsense: a universal change detection method with local adaptive sensitivity, IEEE Trans. Image Process. 24 (1) (2015) 359–373.

[16] G.-A. Bilodeau, J.-P. Jodoin, N. Saunier, Change detection in feature space using local binary similarity patterns, in: International Conference on Computer and Robot Vision, IEEE, 2013. https://ieeexplore.ieee.org/abstract/document/6569191.

[17] Y. Zhang, X. Li, Z. Zhang, F. Wu, L. Zhao, Deep learning driven blockwise moving object detection with binary scene modeling, Neurocomputing 168 (2015) 454–463.

[18] M. Braham, M.V. Droogenbroeck, Deep background subtraction with scene-specific convolutional neural networks, in: international conference on systems, signals and image processing (IWSSIP), IEEE, 2016. https://ieeexplore.ieee.org/document/7502717.

[19] S. Javed, A. Mahmood, T. Bouwmans, S.K. Jung, Background–foreground modeling based on spatiotemporal sparse subspace clustering, IEEE Trans. Image Process. 26 (12) (2017) 5840–5854.

[20] W. Zhou, S. Kaneko, D. Liang, M. Hashimoto, Y. Satoh, Background subtraction based on co-occurrence pixel-block pairs for robust object detection in dynamic scenes, IIEEJ.Trans.Image ElectronicVisual Comput. 5 (2) (2017) 146–159.

[21] W. Zhou, S. Kaneko, M. Hashimoto, et al., Co-occurrence Background Model with Hypothesis on Degradation Modification for Robust Object Detection, 2018. https://pdfs.semanticscholar.org/5148/1b05f4d6f31c541770cbacd7c6f978b3eb04.pdf.

[22] I.J. Goodfellow, J. Shlens, C. Szegedy, Explaining and harnessing adversarial examples, 2015.

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

[24] P.-M. Jodoin, L. Maddalena, A. Petrosino, Y. Wang, Extensive benchmark and survey of modeling methods for scene background initialization, IEEE Trans. Image Process. 26 (11) (2017) 5244–5256.

[25] N. Goyette, et al., Changedetection. net: A new change detection benchmark dataset, in: IEEE computer society conference on computer vision and pattern recognition workshops, IEEE, 2012. https://ieeexplore.ieee.org/document/6238919.

[26] D. Liang, S. Kaneko, M. Hashimoto, K. Iwata, X. Zhao, Co-occurrence probability-based pixel pairs background model for robust object detection in dynamic scenes, Pattern Recognit. 48 (4) (2015) 1374–1390.

[27] X. Li, K. Liu, Y. Dong, Superpixel-based foreground extraction with fast adaptive trimaps, IEEE Trans. Cybern. 48 (9) (2018) 2609–2619.

[28] X. Li, K. Liu, Y. Dong, D. Tao, Patch alignment manifold matting, IEEE Trans. Neural Netw. Learn. Syst. 29 (7) (2018) 3214–3226.

[29] D. Stutz, A. Hermans, B. Leibe, Superpixels: an evaluation of the state-of-the-art, Comput. Vis. Image Understand. 166 (2018) 1–27.

[30] S.Y. Elhabian, K.M. El-Sayed, S.H. Ahmed, Moving object detection in spatial domain using background removal techniques-state-of-art, Recent PatentComput.Sci. 1 (1) (2008) 32–54.

[31] D. Bloisi, L. Iocchi, Independent multimodal background subtraction., CompIMAGE (2012) 39–44.

[32] S. Brutzer, B. Höferlin, G. Heidemann, Evaluation of background subtraction techniques for video surveillance, CVPR 2011, IEEE, 2011. https://ieeexplore.ieee.org/document/5995508.

[33] Z. Zhao, T. Bouwmans, X. Zhang, Y. Fang, A fuzzy background modeling approach for motion detection in dynamic backgrounds, MultimediaSignal Process. 346 (2012) 177–185.

[34] L. Maddalena, A. Petrosino, et al., A self-organizing approach to background subtraction for visual surveillance applications, IEEE Trans. Image Process. 17 (7) (2008) 1168.

[35] M. Hofmann, P. Tiefenbacher, G. Rigoll, Background segmentation with feedback: The pixel-based adaptive segmenter, in: Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on, IEEE, 2012, pp. 38–43.

[36] T. Elguebaly, N. Bouguila, Background subtraction using finite mixtures of asymmetric gaussian distributions and shadow detection, Mach. Vis. Appl. 25 (5) (2014) 1145–1162.

[37] A. Sobral, Bgslibrary: An opencv c++ background subtraction library, 7, IX Workshop de Visao Computacional (WVC 2013), 2013.

[38] T. Fawcett, An introduction to roc analysis, Pattern Recognit. Lett. 27 (8) (2006) 861–874.

[39] A. Sobral, A. Vacavant, A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos, Comput. Vision Image Understand. 122 (2014) 4–21.

[40] N. Lazarevic-McManus, J. Renno, G. Jones, Performance evaluation in visual surveillance using the f-measure, in: Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks, ACM, 2006, pp. 45–52.

[41] Q. Huynh-Thu, M. Ghanbari, Scope of validity of psnr in image/video quality assessment, Electron Lett 44 (13) (2008) 800–801.

[42] S. Winkler, P. Mohandas, The evolution of video quality measurement: from psnr to hybrid metrics, IEEE Trans. Broadcast. 54 (3) (2008) 660–668.

[43] T. Huynh-The, O. Banos, S. Lee, B.H. Kang, E.-S. Kim, T. Le-Tien, Nic: a robust background extraction algorithm for foreground detection in dynamic scenes, IEEE Trans. Circuit Syst. Video Technol. 27 (7) (2017) 1478–1490.

[44] D.D. Bloisi, A. Pennisi, L. Iocchi, Parallel multi-modal background modeling, Pattern Recognit. Lett. 96 (2017) 45–54.

[45] G. Chiron, P. Gomez-Krämer, M. Ménard, Detecting and tracking honeybees in 3d at the beehive entrance using stereo vision, EURASIP J. Image Video Process. 2013 (1) (2013) 59.