Co-occurrence Background Model with Hypothesis on Degradation Modification for Robust Object Detection

Wenjun Zhou[1], Shun'ichi Kaneko[1], Manabu Hashimoto[2], Yutaka Satoh[3] and Dong Liang[4]
[1]*Graduate School of Information Science and Technology, Hokkaido University, Japan*
[2]*Chukyo University, Japan*
[3]*National Institute of Advanced Industrial Science and Technology, Japan*
[4]*Nanjing University of Aeronautics and Astronautics, China*

Keywords: Background Model, Co-occurrence Pixel-Block Pairs (CPB), Object Detection, Correlation Depended Decision Function, Severe Scenes, Hypothesis on Degradation (HoD).

Abstract: This paper presents a prospective background model for robust object detection in severe scenes. This background model using a novel algorithm, Co-occurrence Pixel-block Pairs (CPB), that extracts the spatiotemporal information of pixels from background and identifies the state of pixels at current frame. First, CPB realizes a robust background model for each pixel with spatiotemporal information based on a "pixel to block" structure. And then, CPB employs an efficient evaluation strategy to detect foreground sensitively, which is named as correlation dependent decision function. On the basis of this, a Hypothesis on Degradation Modification (HoD) for CPB is introduced to adapt dynamic changes in scenes and reinforce robustness of CPB to against "noise" in real conditions. This proposed model is robust to extract foreground against changes, such as illumination changes and background motion. Experimental results in different challenging datasets prove that our model has good effect for object detection.

1 INTRODUCTION

Object detection is one active area of research in the field of visual surveillance(Hu et al., 2004), where background subtraction has been widely used in various problems(Yilmaz et al., 2006; Moeslund et al., 2006; Cheung and Kamath, 2005). However, implementing background subtraction for real scenes with severe backgrounds is beset with challenges(Vacavant et al., 2012), not least of which are those related to ***illumination changes***, e.g. variable sunlight outdoors or lights being switched on and off indoors, and then ***background motions***, e.g. swaying trees or moving waves on the water.

To overcome such challenges, two types of design schema have been proposed. First is the *pixel-wise* model, in which the intensity of each pixel is independently analyzed in the temporal domain and then the current frame is subtracted. An example of this approach is Pfinder(Wren et al., 1997), a real-time method for analyzing the color information (Y/U/V) of each pixel and then building a pixel-wise model by the Gaussian mixture model (GMM)(Stauffer and Grimson, 1999), which is a well-known way to deal with multiple background objects. Elgammal et al.(Elgammal et al., 2002) proposed a non-parametric method that can be used to detect object in severe scenes by the using of kernel density estimation (KDE).

The second scheme is the *spatial-based* model, in which a background model is built by making decisions regarding the spatial correlations between pixels or blocks. Seki(Seki et al., 2003) proposed such a method that involved estimating the co-occurrence correlation between neighboring blocks. Subsense(St-Charles et al., 2015), a recently algorithm following ViBe's strategy(Barnich and Van Droogenbroeck, 2011) that presented one pixel-level segmentation method that relies on spatiotemporal binary features combined with color information to detect foreground.

The first design schema can not deal well with illumination changes in the absence of contextual spatial information and most of the second design schema pay much attention to the local spatial information of neighboring pixels or blocks and ignore the global spatial information.

To counter these overlooked issues, we propose an effective method of co-occurrence pixel-block pairs

Zhou, W., Kaneko, S., Hashimoto, M., Satoh, Y. and Liang, D.
Co-occurrence Background Model with Hypothesis on Degradation Modification for Robust Object Detection.
DOI: 10.5220/0006613202660273
In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018) - Volume 5: VISAPP*, pages 266-273
ISBN: 978-989-758-290-5

(CPB) for detecting objects robustly in severe scenes. This is based on our earlier works(Iwata et al., 2009; Zhao et al., 2011; Liang et al., 2015) with the innovations: 1) a “pixel to block” structure that can provide a fast statistical training solution, thereby allowing an on-line approach; 2) a novel evaluation strategy named correlation depended decision function for accurate object detection. Based on CPB, we propose a Hypothesis on Degradation Modification (HoD) for CPB to adapt dynamic changes in scenes and reinforce robustness of CPB to against “noise” in real conditions. More details are described in following sections and we also compare the proposed method with other advanced techniques to prove the efficiency of our method under various challenging datasets. This paper is organized as follows. Sections 2 introduces how CPB method works in details. Section 3 gives one introduction of Hypothesis on Degradation Modification (HoD) for CPB. Section 4 reports experimental results to compare the performance of the proposed method with other advanced approaches. Section 5 concludes the paper along with future work.

2 CPB BACKGROUND MODEL

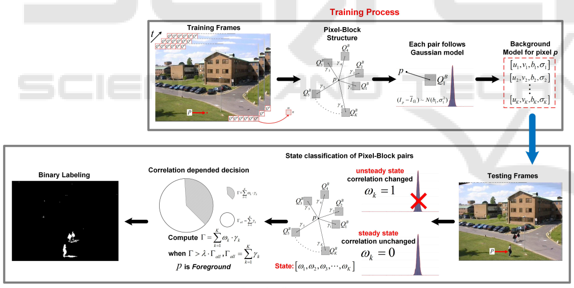


Figure 1: Overview of working mechanism of CPB.

2.1 Overview

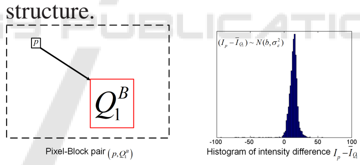
In general, the proposed CPB includes two processes: training process and detecting process. Fig. 1 shows the overview of working mechanism of CPB.

2.2 Model Building

As an extension from the “pixel to pixel ” structure that works SRF(Iwata et al., 2009), GAP(Zhao et al., 2011) and CP3(Liang et al., 2015) 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 $\{Q_k^B\}_{k=1,2,\dots,K} = \{Q_1^B, Q_2^B, \dots, Q_K^B\}$ to denote a supporting block set for the target pixel p. As an instance, we first divide each frame (the size is $U \times V$) into the blocks $\{Q^B\}$, the size of each block is $m \times n$ and the number of blocks is $M \times N$ ($\frac{U}{m} = M, \frac{V}{n} = N$). 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(Liang et al., 2015).

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.2, where $\bar{I}_Q$ is the average intensity of block Q^B. The relation shown in Fig.2(b) is called as “Co-occurrence between intensity,” and we can utilize this knowledge to design the statistical model for the characteristics in background pixels. Since the main purpose of this study is to design a robust detector of any foreground events, such as walking peoples, animals or cars on the roads or grasses without any detection of the meaningless events, such as moving clouds or shaking grasses, we utilize multiple relationship of the co-occurrence mentioned above to build the background model.

(a) Co-occurrence pixel-block pair structure.



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

Figure 2: Basic structure of co-occurrence pixel to block pair.

2.2.1 Selection of Supporting Blocks

A set of supporting blocks $\{Q_k^B\}$ is defined for each pixel p in scene by utilizing Pearson's product-moment correlation coefficient:

$$\gamma(p, Q_k^B) = \frac{C_{p,\bar{Q}_k}}{\sigma_p \cdot \sigma_{\bar{Q}_k}}, \tag{1}$$

where $C_{p,\bar{Q}_k}$ 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.

In general, we can expect that if the pixel-block pair (p, Q_k^B) keeps a high correlation coefficient, then

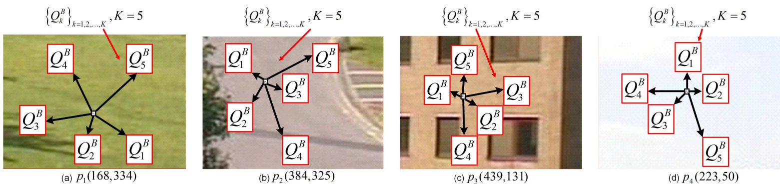


Figure 3: Example layouts of pixel-block pairs for different position pixels $p_1(168,334)$, $p_2(384,325)$, $p_3(439,131)$ and $p_4(223,50)$, respectively, where $K=5$ and the size of each block is 5×5.

the supporting block Q_k can provide some reliability to estimate the current sate of the target pixel p. According this approach, we propose a set of supporting blocks$\{Q_k^B\}_{k=1,2,...,K} = \{Q^B|\gamma(p,Q^B)$ is the K highest$\}$ for each pixel p. Fig.3 shows example layouts of the supporting blocks using $PETS2001-dataset3$ and the target pixels are selected from the four representative regions: “Grass,” “Road,” “Building,” “Sky,” respectively.

2.2.2 Statistical Modeling of Pairwise Intensity Co-occurrence

For the selected K pixel-block pairs, we build a statistical model using the single Gaussian distribution as defined in the following expression:

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

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. We assume that the difference in intensities between any co-occurrence pairs follows a normalized distributions $N(b,\sigma^2)$(Liang et al., 2015), and then we use the single Gaussian model to build the background model for each co-occurrence pair. The variance estimation σ_k^2 cis defined as follows:

$$\sigma_k^2 = \frac{1}{T}\sum_{t=1}^{T}(\Delta_k - b_k)^2, \tag{3}$$

and b_k is the differential increment

$$b_k = \frac{1}{T}\sum_{t=1}^{T}\Delta_k, \tag{4}$$

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 $\{Q_k^B\}_{k=1,2,...,K}$, where (u_k, v_k) is the coordinate of supporting block.

2.3 Object Detection

We contain a competitive binary classification process for the object detection(Elhabian et al., 2008) in our CPB by evaluating each pair (p, Q_k^B) of every pixel in turn. It includes two procedures: 1) to estimate the steady or unsteady state of each pair, and then 2) to distinguish a target pixel is belongs to foreground or background.

2.3.1 State Classification of Pixel-Block Pairs

To identify whether a pixel p is belongs to foreground or background , it is necessary to design a framework which can distinguish the difference between these two states at the detecting process. 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. This idea can be realized as the following expression for identifying pixel p is foreground or background as shown in Fig. 4.

In CPB, we can define this statistical structure in each relation between any pixel and its supporting block set as the collection of Gaussian distributions learned in the training process. In the detecting process, we utilize these knowledges to find any foreground pixels which may violate the knowledge due to a different intensity from its background pixel. For each pair (p, Q_k^B), a binary function for identifying its steady or unsteady state can be defined as follows:

$$\omega_k = \begin{cases} 1 & if \ \left|(p - Q_k^B) - b_k\right| \geq \eta \cdot \sigma_k \\ 0 & \text{otherwise} \end{cases}, \tag{5}$$

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 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.

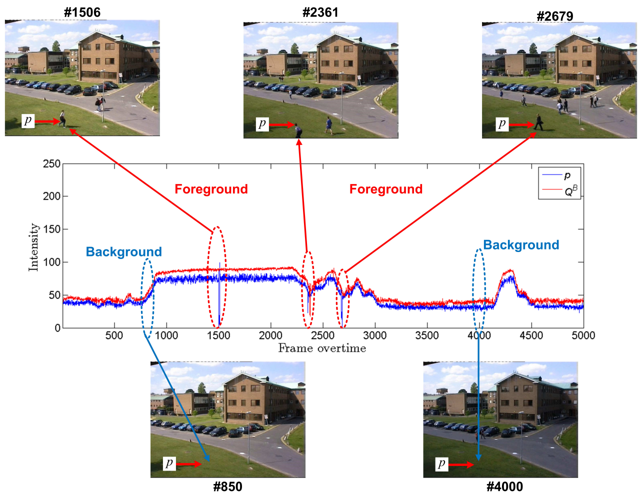


Figure 4: Relationship in the intensity changes between target pixel p and supporting block Q^B.

In this function, ω_k corresponds a logical value to represent **the steady state with 0** or **the unsteady state with 1** for each pair, respectively.

2.3.2 Correlation Depended Decision Function

In order to define an efficient decision function for target pixel, with considering the K supporting blocks around it, 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 (Liang et al., 2015; Elhabian et al., 2008). The lager γ_k may be stronger or more reliable on the state decision of target pixel p. The definition is realized as Γ as follows.

$$\Gamma = \sum_{k=1}^{K} \omega_k \cdot \gamma_k. \tag{6}$$

Γ has the following two significances: first, Γ can count up the unsteady pairs, second, Γ has its own ideal value, the maximum value of Γ is 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.

$$\Gamma_{all} = \sum_{k=1}^{K} \gamma_k. \tag{7}$$

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*
else
  p is *background*.

λ is a threshold parameter. As shown in Fig. 5, It is natural to evaluate the state of pixel p through a comparative analysis between Γ and Γ_{all}, if the value of Γ is high, it is highly likely that pixel p is a foreground pixel.

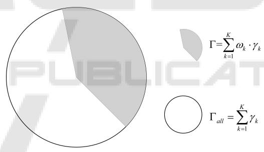


Figure 5: Relationship between Γ and Γ_{all}.

3 HoD MODIFICATION

We have introduced the basic algorithms for robust background subtraction so far, however, in the real world we have only a set of limited data for training model over some limited time range. We may have some mechanism to modify the model to fix some errors which may be observed in any new target frames out of the training set. In this section, we intend to introduce a simple mechanism named Hypothesis on Degradation Modification (HoD) extended from CPB to adapt dynamic changes in scenes and reinforce robustness of CPB to against "noise" in real conditions.

3.1 Hypothesis on Degradation

By use of the basic algorithm in learning and detecting structure of CPB, we can extract particular

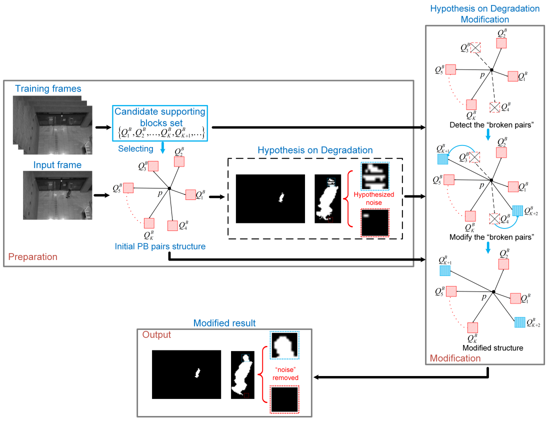


Figure 6: Overview of HoD Modification.

events, such as pedestrians or vehicles in scene. After a long utilization of initial CPB background model for real data, we may encounter some strange or unknown situations which do not belong to the initial training dataset, and then our initial CPB model may gradually or suddenly decrease its performance by reason of the change of observation in scene.

In practice, we propose an assumption that some "noise" may arise in detecting process due to some trouble in CPB structure over time. However, we could not know the true of these troubles without any ground truth data, and it is not possible to realize ground truth of future frames in real condition. Hence, we need an effective modification to adapt the possible changes in real condition and consolidate the performance of proposed CPB over time. In this study, we call this above assumption as "Hypothesis on Degradation" and name the "noise " in detecting process as "hypothetical noise."

Based on mentioned above, we propose a Hypothesis on Degradation Modification (HoD) for CPB to against the hypothetical noise by modifying the initial structure to a new one. Fig. 6 describes an overview of the proposed HoD. Here in Fig. 6, it is clear 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 is also a feasible on-line mode for CPB.

To estimate which Pixel-Block structure should be modified, we first define two types of 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 hypothetical noise using $AIST-Indoor$-dataset provided by the National Institute of Advanced Industrial Science and Technology in Japan.

For such pixels as mentioned above, we detect them as noise and do modification for the Pixel-Block structure of them.

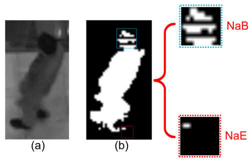


Figure 7: Example of hypothesized noise. (a) Raw data. (b) Description of hypothesized noise.

3.2 Detection of Possible Wrong Pixel-Block Pairs

For any detected noise pixels, we need to define wrong or broken elemental pairs in the Pixel-Block structure . As introduction in Section 2.3, we adopt the strategy that any Pixel-Block pair, which has the lager γ must hold the higher weight in the trained structures and such pair is more likely to affect the state of NaE and NaB. We propose a weight-based decision rule to detect the wrong pair:

$$if\ \gamma_m \geq \bar{\gamma},\ \ then\ (p, Q_m^B)\ is\ wrong \tag{8}$$

where (p, Q_m^B) is the 'wrong' pair, which is in unsteady state of NaE or steady state of NaB. Depending on the noise is NaE or NaB, the threshold $\bar{\gamma}$ has the 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:

$$\bar{\gamma} = \frac{1}{M} \sum_{k=1}^{K} \gamma_k \cdot \omega_k = \frac{1}{M} \Gamma. \tag{9}$$

While in the case of NaB, it is defined as follows:

$$\bar{\gamma} = \frac{1}{K-M} \sum_{k=1}^{K} \gamma_k \cdot (1-\omega_k) = \frac{1}{K-M} (\Gamma_{all} - \Gamma). \tag{10}$$

We can see a slight difference in the above definitions. The calculations contain the elemental correlation coefficient, the supporting block set and the total number of Pixel-Block pairs $\{(p, Q_m^B)\}$, and then we record these "broken pairs" as shown in Fig. 8.

3.3 Removal of Wrong Structure

We try to exchange the wrong pair by new one which is kept as a spare pair in the training process. Fig. 8 shows its schema for exchange to keep K pairs in any supporting block sets.

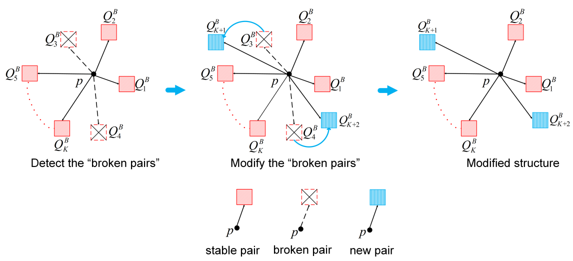


Figure 8: Modification process.

4 EXPERIMENTS

4.1 Experimental Setup

At first, considering the several challenges of video surveillance for background subtraction algorithm(Brutzer et al., 2011). We consider the following challenges for evaluation:

- **Gradual Illumination Changes:** the light intensity typically varies during day. We consider $PETS2001 - camera1$ as the testing data for evaluation with the illumination change during day.
- **Sudden Illumination Changes:** for example the sudden switch of light, strongly affects the observation of object to lead a fault for detection. We consider the dataset $AIST - Indoor$ with strong sudden light changes when the auto-door opening, in such moment it is difficult to detect true foreground from the scene.
- **Dynamic Background:** some movement in scene should be regarded as background e.g. swaying tree, waving water. We select one challenging sequence $advertisementBoard$ from $SceneBackgroundModeling.NET$(SBMnet) dataset for testing, and this sequence contains an ever-changing advertising board in the scene.
- **Intermittent Object Motion:** this category is one difficult challenge for object detection with background objects moving away, abandoned objects and objects stopping for a short while and then moving away. In this category, it is difficult to detect correct foreground objects. The $sofa$ sequence from $Change - detection\ dataset$(Goyette et al., 2012) is selected for testing.
- **Camera Jitter:** in video surveillance, camera jitter is one issue that need to be solved for background subtraction. In our experiment, we consider $sidewalk$ from $Change - detection\ dataset$(Goyette et al., 2012) to test the performance of proposed CPB and CPB+HoD in such extreme category.

4.2 Evaluation Measurement

To analyze the quality of our method, we utilize three common analysis measurements: *Precision*, *Recall*, and *F-measure*. These metrics are widely used to estimate the quality of background subtraction methods (Brutzer et al., 2011; Vacavant et al., 2012). For further evaluating our CPB and CPB+HoD, we introduce the peak signal-to-noise ratio (PSNR) as our metric(Huynh-Thu and Ghanbari, 2008), which can be used to measure the quality of the estimated resulted compared with the background truth(Huynh-The et al., 2016). The definition of $PSNR$ is calculated as follows:

$$PSNR = 10 \cdot \log_{10}\left(\frac{255^2}{MSE}\right), \tag{11}$$

where MSE is the mean square error.

4.3 Result Evaluation

In this section, we compare the proposed CPB and CPB+HoD with four different foreground detection methods: GMM(Stauffer and Grimson, 1999) and KDE(Elgammal et al., 2002), which are two well-known traditional algorithms, and two state of the art techniques IMBS(Bloisi and Iocchi, 2012) and SuBSENSE(St-Charles et al., 2015), especially SuBSENSE is one of the top-ranked methods in $Change - detection\ dataset$ at present. In contrast to methods with complex strategies(Bloisi and Iocchi, 2012; St-Charles et al., 2015), CPB is a low-complexity algorithm that is more easily realized. The parameters for GMM, KDE, IMBS and SuBSENSE were set by using the tool bgslibrary(Sobral, 2013). In experiments, we set each block as 8×8 pixels, $\lambda = 0.5$ and $\eta = 2.5$ for CPB.

Fig. 9 shows examples of foreground detection for a typical frame from each dataset sequence. Table 1 lists the results of the performance measurements of CPB and CPB+HoD with other methods from all the categories, respectively. Compared with above foreground detection results, the proposed algorithms outperform the methods GMM, KDE , IMBS and SuBSENSE in most testing sequences. Meanwhile, CPB+HoD is quite efficient in extracting foreground from sequences that suffers from sudden illumination changes and dynamic background. Furthermore, 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 1, that means our algorithm is robust against noise for detecting foreground in severe scenes.

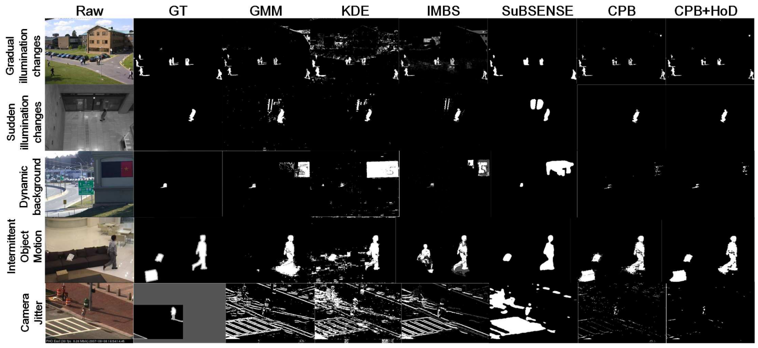


Figure 9: Foreground detection results in different challenging sequences.

Table 1: Comparison in different challenging categories.

Method	Measure	Category				
		Gradual illumination changes	Sudden illumination changes	Dynamic background	Intermittent object motion	Camera jitter
GMM	Precision	0.6465	0.6523	0.5151	0.8462	0.4150
	Recall	0.9508	0.9207	0.5196	0.6876	0.4811
	F-measure	0.7697	0.7636	0.5174	0.7587	0.4456
	PSNR	39.46	40.57	26.92	36.11	31.79
KDE	Precision	0.5181	0.5896	0.4962	0.7361	0.5640
	Recall	0.8836	0.6944	0.4856	0.7820	0.5167
	F-measure	0.6532	0.6377	0.4909	0.7583	0.5393
	PSNR	17.77	38.16	21.67	25.99	33.68
IMBS	Precision	0.5162	0.5760	0.5095	0.8353	0.4457
	Recall	0.8841	0.6923	0.5118	0.7298	0.4879
	F-measure	0.6518	0.6288	0.5107	0.7790	0.4658
	PSNR	16.20	36.36	30.09	28.21	32.10
SuBSENSE	Precision	0.9008	0.5864	0.5018	0.9556	0.5966
	Recall	0.8840	0.7047	0.5033	0.7803	0.5079
	F-measure	0.8923	0.6401	0.5025	0.8591	0.5487
	PSNR	54.11	37.14	27.62	32.29	34.92
CPB	Precision	0.9566	0.8651	0.7653	0.8928	0.6365
	Recall	0.7517	0.8181	0.5118	0.8691	0.5051
	F-measure	0.8418	0.8409	0.6133	0.8808	0.5633
	PSNR	56.05	53.14	36.64	32.04	34.22
CPB+HoD	Precision	0.9652	0.8668	0.7973	0.9079	0.6384
	Recall	0.7562	0.8227	0.5214	0.8750	0.5055
	F-measure	0.8480	0.8442	0.6305	0.8912	0.5642
	PSNR	56.39	53.31	37.39	32.69	34.28

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

Based on co-occurrence pixel-block pairs, CPB can build one prospective background model from a scene, such background model contains spatial and temporal information of each pixel in sequence, and then CPB can analyze the current sate of each pixel effectively with these information. In other words, at training process, CPB can learn the information of scene, whether the scene is dynamic or static, our model can acquire the regularity of scene. Then, at detecting process, when any object enters into the scene and the information of this object is out of range of our model, so we can extract the object from the scene efficiently.

For that reason, CPB does well in above scenes. On the basis of this, we introduce a HoD into CPB to adapt dynamic changes in scenes and reinforce robustness in real conditions. Through the results of above experiments, CPB+HoD leads a good performance in various scenes.

5 CONCLUSIONS

We have proposed a robust and efficient object detection approach named CPB in severe scenes. It was designed to reduce the computing cost in training process and also to keep the robustness against scene changes in reality. Furthermore, we realized a novel modification approach named hypothesis-on-degradation modification (HoD) for CPB to defend the possible degradation in practice and it is also a feasible on-line mode for the proposed CPB. The experimental results show the good performance of the proposed approach. In future, we would like to improve our CPB to be an on-line approach by hypothesis on degradation modification (HoD).

REFERENCES

Barnich, O. and Van Droogenbroeck, M. (2011). Vibe: A universal background subtraction algorithm for video sequences. *IEEE Transactions on Image processing*, 20(6):1709–1724.

Bloisi, D. and Iocchi, L. (2012). Independent multimodal background subtraction. In *CompIMAGE*, pages 39–44.

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

Cheung, S.-C. S. and Kamath, C. (2005). Robust background subtraction with foreground validation for urban traffic video. *EURASIP Journal on Advances in Signal Processing*, 2005(14):726261.

Elgammal, A., Duraiswami, R., Harwood, D., and Davis, L. S. (2002). Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. *Proceedings of the IEEE*, 90(7):1151–1163.

Elhabian, S. Y., El-Sayed, K. M., and Ahmed, S. H. (2008). Moving object detection in spatial domain using background removal techniques-state-of-art. *Recent patents on computer science*, 1(1):32–54.

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

Hu, W., Tan, T., Wang, L., and Maybank, S. (2004). A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 34(3):334–352.

Huynh-The, T., Banos, O., Lee, S., Kang, B. H., Kim, E.-S., and Le-Tien, T. (2016). Nic: a robust background extraction algorithm for foreground detection in dynamic scenes. *IEEE transactions on circuits and systems for video technology*.

Huynh-Thu, Q. and Ghanbari, M. (2008). Scope of validity of psnr in image/video quality assessment. *Electronics letters*, 44(13):800–801.

Iwata, K., Satoh, Y., Ozaki, R., and Sakaue, K. (2009). Robust background subtraction based on statistical reach feature method. *IEICE Trans. on Information and Systems*, 92:1251–1259.

Liang, Dong Kaneko, S., Hashimoto, M., Iwata, K., and Zhao, X. (2015). Co-occurrence probability-based pixel pairs background model for robust object detection in dynamic scenes. *Pattern Recognition*, 48(4):1374–1390.

Moeslund, T. B., Hilton, A., and Krüger, V. (2006). A survey of advances in vision-based human motion capture and analysis. *Computer vision and image understanding*, 104(2):90–126.

Seki, M., Wada, T., Fujiwara, H., and Sumi, K. (2003). Background subtraction based on cooccurrence of image variations. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–II. IEEE.

Sobral, A. (2013). Bgslibrary: An opencv c++ background subtraction library. In *IX Workshop de Visao Computacional (WVC2013)*, volume 7.

St-Charles, P.-L., Bilodeau, G.-A., and Bergevin, R. (2015). Subsense: A universal change detection method with local adaptive sensitivity. *IEEE Transactions on Image Processing*, 24(1):359–373.

Stauffer, C. and Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2, pages 246–252. IEEE.

Vacavant, A., Chateau, T., Wilhelm, A., and Lequièvre, L. (2012). A benchmark dataset for outdoor foreground/background extraction. In *Asian Conference on Computer Vision*, pages 291–300. Springer.

Wren, C. R., Azarbayejani, A., Darrell, T., and Pentland, A. P. (1997). Pfinder: Real-time tracking of the human body. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):780–785.

Yilmaz, A., Javed, O., and Shah, M. (2006). Object tracking: A survey. *Acm computing surveys (CSUR)*, 38(4):13.

Zhao, X., Satoh, Y., Takauji, H., Kaneko, S., Iwata, K., and Ozaki, R. (2011). Object detection based on a robust and accurate statistical multi-point-pair model. *Pattern Recognition*, 44(6):1296–1311.