ADAPTIVE LOCAL SPATIAL MODELING FOR ONLINE CHANGE DETECTION UNDER ABRUPT DYNAMIC BACKGROUND

Dong Liang[1], *Shun'ichi Kaneko*[2], *Han Sun*[1], *Bin Kang*[3]

[1]Collaborative Innovation Center of Novel Software Technology and Industrialization,
College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, China
[2]Graduate School of Information Science and Technology, Hokkaido University, Japan
[3]Nanjing University of Posts and Telecommunications, China

ABSTRACT

Change detection is an important theme in video processing. To provide reliable detection results in challenging scenes, traditional methods introduced sophisticated statistical distributions and hand-craft spatial features to build background models. In this paper, we develop an intuitive background model based on simple statistical distribution and adaptive spatial correlation among pixels: For each observed pixel, we select a group of supporting pixels with high correlation, and then employ a single Gaussian to model the intensity deviations of each pixel pair. To compensate camera motion and fast adapt to dynamic pattern that coming afterwards, a randomized multichannel on-line updating mechanism is introduced. This observation is robust to abrupt illumination variation and dynamic background. Experimental results using all the video sequences provided by three challenging benchmarks (CDW-2012, CDW-2014 and SABS) validate it outperforms many state-of-the-art methods under various situations.

Index Terms— Background model, change detection, illumination variation, dynamic scenarios

1. INTRODUCTION

Change detection in real world scenario should be free from illumination changes and dynamic backgrounds. However, the obstacle to introduce more sophisticated learning technique is that it is a scene-dependent and pixel-wise processing procedure [1]. Typically, it provides pixel-wise detection result via a specific scenario training stage using a relatively lightweight training and detection model to reduce the resources occupancy.

Previous studies focused on statistical distributions to describe challenging background: from the earliest single Gaussian model for each pixel [2], to the multi-modal background representation Mixture of Gaussian (MoG) [3], and the non-parametric models [4, 5, 6, 7]. However, this kind of approach is flawed, as shown in Fig.1 for example, to cover the variation of illumination change, the background model (MoG model here) occupies a large range of intensity, so that the detection is insensitive. Local features which can represent spatial characters have shown great potential [8, 9, 10, 11, 12, 13], however, such kind of handcraft feature cannot adapt to many non-ideal cases, such as texture-less and dynamic texture background.

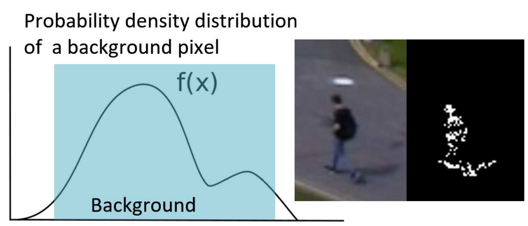

Fig. 1. The background model occupy a large range of the intensity so that the detection is insensitive.

This paper presents an intuitive and robust change detection method, which is based on our previous proposed background model: Co-occurrence Probability Pixel Pairs (CP3) [14, 15]. CP3 is originally designed for off-line object detection under sudden illumination changes, and it is proved to be robust in weak illumination, and dynamic texture. In this paper, we devote to: (1) compensate camera motion; (2) fast adapt to new coming dynamic pattern; (3) better distinguish objects from camouflage effects (photometric similarity of object and background). Section 2 gives a brief review of CP3 and its characteristics, Section 3 introduces a randomized multichannel on-line updating mechanism to solve the above issue, Section 4 describes the experiment procedure, parameter setting, and experimental results testing on Change Detection(CDW-2012 and CDW-2014) [16] and SABS [17] benchmark, which validate its comprehensive performance is better than or comparable with many other methods under various scenes.

2. FACING ILLUMINATION VARIATION AND DYNAMICS

2.1. Co-occurrence Probability Pixel Pairs

Fig.2 is a diagram of CP3 background model. The intensities of a pixel P have different variation from other arbitrary pixel Q as time goes by; nevertheless, it is also natural that P shows a co-occurrence characteristic with some other pixels, for example, has a stable difference value with its neighboring pixels. When such kind of co-occurring relation is relative stable, the deviation of the pixel pair could be a single Gaussian. To reduce the risk of individual error and enhance the robustness of detection, it is necessary to maintain a sufficient number of Q with scattered locations as supporting pixels, denoted as $\{Q_k^P\}_{k=1,2,\ldots,K}$, which provides a group of estimation

Thanks to National Key R&D Program 2017YFB0802300, NSFC 61601223, NSF of Jiangsu Province BK20150756, Fundamental Research Funds for the Central Universities NS2016091 and Postdoctoral Science Foundation of China 2015M580427 for funding.
liangdong@nuaa.edu.cn

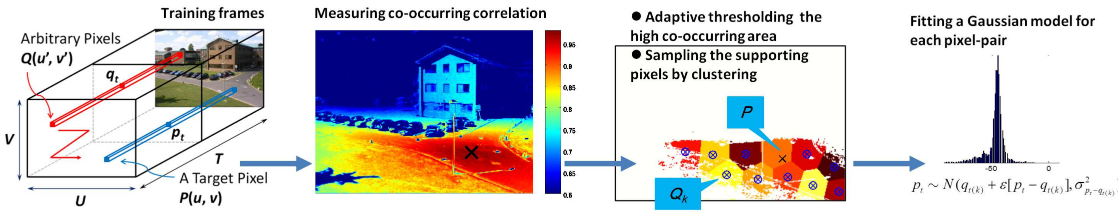

Fig. 2. Diagram of CP3 background model using PATS2001-dataset3-cam1 as a demonstration.

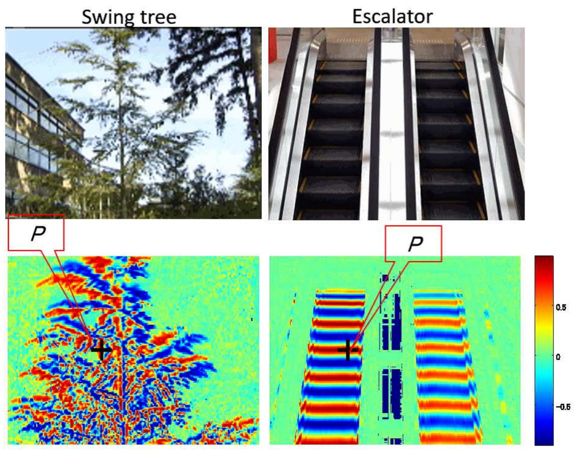

Fig. 3. Examples of dynamic scenarios: tree swinging and auto-induction escalators. The following heatmaps show the correlation distributions of pixel pairs.

for P via a unique Gaussian of each pixel pair. With this algorithm, only two parameters, the mean value $\mathcal{E}[p_t - q_{t(k)}]$ and the standard deviation $\sigma_{p_t - q_{t(k)}}$, are recorded for the following detection procedure. Once the intensity of P is an outlier of the background model, P would be regarded as an abnormal-status/foreground-element.

Computing the co-occurring relation of an arbitrary pixel pair uses a covariance-based correlation matrix. Each row and column of the symmetric matrix is an array of correlation coefficient $\gamma_{(P,\ Q)}$ for each $P(u, v)$. Then Q_n corresponds to the highest N components in the array $\gamma_{(P,\ Q(u',v'))}$ can be selected as the candidates of preferred supporting pixels, namely

$$\{Q_n\} = \{Q(u', v') | \gamma_{(P,\ Q)} > \check{\gamma}\},\ n = 1, 2, ..., N. \tag{1}$$

where $\check{\gamma}$ is the adaptive lower limit, which can be determined by a computable ratio of the signal variance $\sigma^2_{p_t}$ and the random noise variance σ^2_n [15].

2.2. Distribution characteristics of supporting pixel

For a pixel on a static background (this is the most common case), the supporting pixels randomly locate in the scene; For a pixel under illumination changes, high co-occurring supporting pixels generally distribute around the area that has identical illumination changes (For a color space observation, e.g. RGB, the distribution also relies on homogeneous color). Nevertheless, the distribution of the supporting pixels are also related to the geometrical characteristic (position, orientation, shape and relative distance), as shown in Fig.2; For a pixel passed by moving background, it could select the supporting pixels with simultaneous motion while locating along the vertical direction of motion. Fig.3 shows examples of the moving background: tree swinging and auto-induction escalators, both of which often appearing in outdoor/indoor surveillance scenes. The following heatmaps show the correlation distributions of pixel pairs.

Generally, a dense cluster of pixels will be selected as supporting pixels. Once such a cluster is covered by object motion at the same frame, large area of false detection would appear. CP3 avoids most of such cases by taking into account the supporting pixels' scattered distribution by sampling spatial clusters [15].

When the sudden illumination variation pattern and moving background patterns are learned by the background model, without any model updating, the deviation of pixel pair could be stable and the corresponding probability distribution model is explicit (only occupies a narrower intensity range than MoG approach for individual pixel) so that the detection could be sensitive.

3. RANDOMIZED ON-LINE UPDATING

Although CP3 is proved to be robust in sudden illumination changes, weak illumination and dynamic texture, it is originally designed for off-line object detection, so that it could not compensate camera motion. Also, without specific learning procedure, it is hard for CP3 to adapt to new coming dynamic patterns. Hereafter, we generalize it to adapt to on-line change detection.

3.1. Online model parameter updating

It is unreasonable to expect on-line training could outperforms off-line training because on-line training often has deficient prior. The consistency assumption of the pixel pairs would be also broken when detecting on an unlearned phase of a video . We expect to preserve the merits of CP3 at background initialization stage, and then introduce a mechanism to adapt new coming frames.

Firstly, to distinguish objects from camouflage background, we define $\mathbf{p_t}$ and $\mathbf{q_{t(k)}}$ to be the colour vector on a colour space, and $\mathbf{\Delta_{t(k)}}$ to be the mean of $(\mathbf{p_t} - \mathbf{q_{t(k)}})$, and $\mathbf{\Sigma_{t(k)}}$ to be the corresponding covariance matrix. In each recent frame, we update each pixel-pair's statistics recursively. For mean value,

$$\mathbf{\Delta_{t(k)}} = \alpha(\mathbf{p_t} - \mathbf{q_{t(k)}}) + (\mathbf{1} - \alpha)\mathbf{\Delta_{(t-1)(k)}}. \tag{2}$$

For covariance matrix,

$$\mathbf{\Sigma_{t(k)}} = \alpha(\mathbf{p_t} - \mathbf{q_{t(k)}} - \mathbf{\Delta_{t(k)}})(\mathbf{p_t} - \mathbf{q_{t(k)}} - \mathbf{\Delta_{t(k)}})^{\mathbf{T}} + (1 - \alpha)\mathbf{\Sigma_{(t-1)(k)}}. \tag{3}$$

Categories	Recall	Specificity	FPR	FNR	PWC	Precision	F-Measure
baseline	0.8500	0.9972	0.0027	0.1499	0.7724	0.9251	0.8856
dynamicBackground	0.6851	0.9876	0.0124	0.3149	1.5025	0.5159	0.5353
cameraJitter	0.6628	0.9518	0.0481	0.3371	5.9332	0.4561	0.5207
intermittentObjectMotion	0.7825	0.8745	0.1254	0.2174	11.5284	0.5630	0.6176
shadow	0.7839	0.9832	0.0167	0.2160	2.5175	0.6539	0.7036
thermal **CDW-2012**	0.8229	0.9893	0.0106	0.1770	1.6973	0.7663	0.7917
PTZ	0.5694	0.9744	0.0255	0.4306	2.9299	0.2173	0.2794
badWeather	0.8350	0.9954	0.0045	0.1649	0.7411	0.7423	0.7766
lowFramerate	0.6627	0.9965	0.0035	0.3373	1.3753	0.6699	0.5549
nightVideos	0.6308	0.9470	0.0530	0.3692	6.0131	0.2775	0.3483
turbulence **CDW-2014**	0.6885	0.9948	0.0051	0.3114	0.6646	0.3950	0.4724
Overall/MoG Method	0.723/0.660	0.971/0.973	0.030/0.028	0.278/0.340	3.432/4.000	0.556/0.597	0.581/0.557

Table 1. Experiment metrics of CP3+ on CDW-2012 and CDW-2014 benchmark with the reference value of MoG Method.

We use a blind updating policy that just adds the new sample to the model regardless whether it is classified as a background or foreground.

Generally speaking, in order to adapt to quick background change, we should better use a relatively large updating rate α. But blind updating with large updating rate would produce object's "tail". In the following experiments, we set α to be a small value (0.01 in Tab. 2). Although in most of other background models, small update rate could not adapt to fast background change, in CP3 method, the status of a target pixel P is represented jointly by a group of supporting pixels to offset fast change. The updating formulas Eq. (2) and Eq. (3) are introduced to adjust each pixel-pair's model parameters, rather than to deal with quick change. Another motivation to use blind updating is that it avoids a risk that an object is deadlocked as a foreground for a long time.

3.2. Randomized supporting pixel selection

The proposed method also takes into account randomization of the supporting pixel. As mentioned in Section 2.2 [15], the supporting pixels' scattered distribution is created by sampling K-means spatial distance clusters. We argue that this fixed distribution is unsuitable for on-line modelling in a nature environment, as random light change and background motion are unpredictable.

The opinion of randomization is reasonable for background subtraction, because no background model could guarantee an exact description of a current frame according to the processing of historic frames or spatial relations. Here we use a random sampled mechanism to select supporting pixels, which means even for two neighboring and homogeneous pixels, the locations of their supporting pixels can be extremely different. Such kind of scattered and random supporting pixels effectively reduce the risk of false positive, most of them are decomposed into sparse noise, which is facile to be removed by post-processing.

The idea of randomization is also used for background modelling by [4, 18], adopting random policy could be helpful to decompose large-area false-detection risk into sparse-distributed noise (which can be removed easily by a median filter).

3.3. Other details

Besides model updating, another signification modification is introducing the minimum $\hat{\mathbf{p}}_{\mathbf{t}}$ and maximum $\check{\mathbf{p}}_{\mathbf{t}}$ colour values to represent a dynamic range of a background pixel. This step is similar to [6]. The motivation of introducing such mechanism is to overcome a weakness of CP3: the Gaussian model is to model the relative relation of a pixel pair (pixel pair's deviation), whereas a pixel's intensity range has been ignored. When processing a short-distance surveillance video, the object can be so "huge" that target pixel and most of the supporting pixels are covered at the same frame. If the object's texture is similar to the background's (for example, both are smooth with poor texture), it would result a large number of false detection, because the passed object would not break the pixel pair's deviation relation, even if its intensity has obvious change. $\hat{\mathbf{p}}_{\mathbf{t}}$ and $\check{\mathbf{p}}_{\mathbf{t}}$ are also updated with other two parameters together, the details of their updating are same as [6]. We combine the upper and lower thresholds with the Gaussian constrain together, and use logic AND operation to integrate them.

4. EXPERIMENTS

4.1. Experiment setting

With the purpose of retaining the generality in real world applications and ensuring a fair comparison with other background subtraction methods on CDW-2012, CDW-2014 [16] and SABS [17] benchmarks, we just use the first 100 frames of every dataset for model initialization, and then use the on-line model mentioned above to detect object and update background model. All the parameters used are listed in Tab. 2, and a detailed discussion of parameters can be found in [15]. Note that, we use different pf and C rather than the values in [15] as we hope to reduce the false positive rate holistically. The changes of pf and C provide a higher tolerance to different dynamic background but would reduce model's sensitivity. But in this version, we have a minimum and maximum to restrain the dynamic range of a pixel, which helps to maintain sensitive detection. In addition, we have no parameter to deal with cast shadows, although it is possible to simply integrate other studies to suppress cast shadows [19, 5, 20], by using dual (or more) thresholds to restrain the range

Number of supporting pixels K	20
Probability function threshold pf	0.35
Gaussian model threshold C	3.0
Updating rate α	0.01

Table 2. Parameter setting.

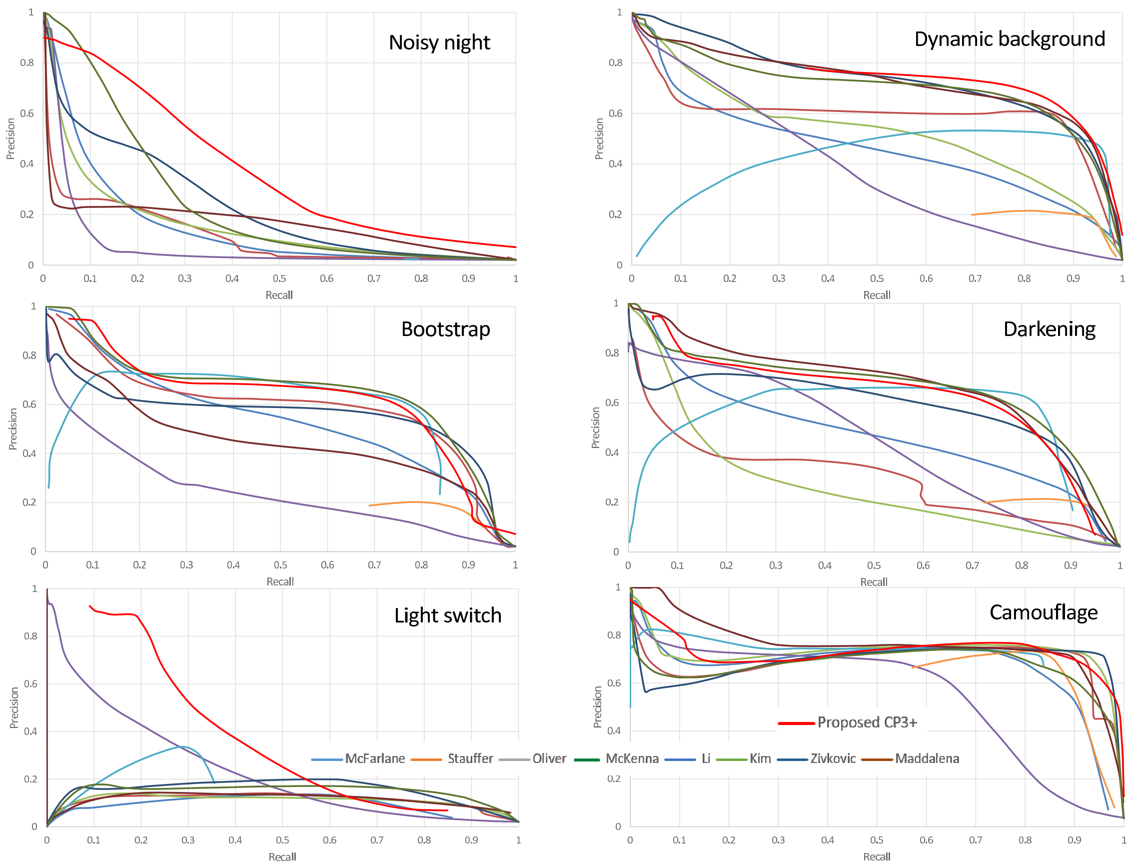

Fig. 4. Precision-Recall curve of CP3+ and other methods on SABS benchmark.

on an appropriate colour space. However, because the appearance of cast shadow in reality varies according to different illumination situations, imaging sensor and scenario, those threshoding-based methods are not so robust under different cases, for example, in the dark region of an image, colour analysis works poorly. More complicate methods involve the combination of colour, geometrical and temporal characteristics [21, 22] which are out of our scope.

Tab. 1 gives the category-wise quantitative analysis using the 7 metrics provided by CDW-2012 and CDW-2014 utilities based on True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). $Sp(Specificity) = TN/(TN + FP)$, $FPR(FalsePositiveRate) = FP/(FP + TN)$,$Precision = TP/(TP + FP)$. More detailed about specific metrics, the average $Recall = TP/(TP + FN) = 0.723$ and False Negative Rate $FNR = FN/(TP + FN) = 0.278$. Recall and FNR are just the inverse of each other to present the completeness of the foreground after detection, which means our method can preserve well completeness of the foreground blob. We also pay attention to Percentage of Wrong Classifications $PWC = 100*(FN+FP)/(TP+FN+FP+TN)$ and $F-Measure = 2*Precision*Recall)/(Precision+Recall)$. PWC should be as small as possible and $F - Measure$ just the reverse. PWC and $F - Measure$ of our method is better than most of other methods (PWC=3.43% and F-Measure=0.580).

Fig.4 shows the experiment results on SABS benchmark. We can see that in the six scenarios, CP3+ outperforms other methods significantly in three scenarios: Night noise, Dynamic background, Light switch, which indicates that it is robust to sudden illumination changes and dynamic background.

For computation time, a Matlab code for on-line updating and detection of the method reaches around 20 fps with size of 320 by 240 image on a Intel i7 3.0 GHz PC.

5. CONCLUSIONS

This paper reports CP3+ for on-line change detection and its performance using video sequences provided by CDW-2012, CDW-2014 and SABS benchmark. A randomized multichannel on-line updating mechanism is introduced to compensate camera motion and fast adapt to new coming dynamic pattern. The experimental results under different categories validate its comprehensive performance outperform many change detection methods. More details including source code could be found in the webpage:
`https://liangdongcv.cn/cp3.htm`

6. REFERENCES

[1] Mingliang Chen, Qingxiong Yang, Qing Li, Gang Wang, and Ming Hsuan Yang, "Spatiotemporal background subtraction using minimum spanning tree and optical flow," in *ECCV*, 2014, pp. 521–534.

[2] Christopher Richard Wren, Ali Azarbayejani, Trevor Darrell, and Alex Paul Pentland, "Pfinder: Real-time tracking of the human body," *TPAMI*, vol. 19, no. 7, pp. 780–785, 1997.

[3] Thierry Bouwmans, Fida El Baf, and Bertrand Vachon, "Background modeling using mixture of gaussians for foreground detection-a survey," *Recent Patents on Computer Science*, vol. 1, no. 3, pp. 219–237, 2008.

[4] 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, June 2011.

[5] Ahmed Elgammal, Ramani Duraiswami, David Harwood, and Larry 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.

[6] Kyungnam Kim, Thanarat H Chalidabhongse, David Harwood, and Larry Davis, "Real-time foreground–background segmentation using codebook model," *Real-time imaging*, vol. 11, no. 3, pp. 172–185, 2005.

[7] Hanzi Wang and David Suter, "Background subtraction based on a robust consensus method," in *ICPR*. IEEE, 2006, vol. 1, pp. 223–226.

[8] Marko Heikkila and Matti Pietikainen, "A texture-based method for modeling the background and detecting moving objects," *TPAMI*, vol. 28, no. 4, pp. 657–662, 2006.

[9] Erik G Learned-Miller, Manjunath Narayana, and Allen Hanson, "Background modeling using adaptive pixelwise kernel variances in a hybrid feature space," in *CVPR*, 2012.

[10] Shengcai Liao, Guoying Zhao, Vili Kellokumpu, Matti Pietikainen, and Stan Z Li, "Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes," in *CVPR*. IEEE, 2010, pp. 1301–1306.

[11] Yaser Sheikh and Mubarak Shah, "Bayesian modeling of dynamic scenes for object detection," *TPAMI*, vol. 27, no. 11, pp. 1778–1792, 2005.

[12] Xinyue Zhao, Yutaka Satoh, Hidenori Takauji, Shunichi Kaneko, Kenji Iwata, and Ryushi Ozaki, "Object detection based on a robust and accurate statistical multi-point-pair model," *Pattern Recognition*, vol. 44, no. 6, pp. 1296–1311, 2011.

[13] Thien Huynh-The, Oresti Banos, Sungyoung Lee, Byeong Ho Kang, Eun Soo Kim, and Thuong Le-Tien, "Nic: A robust background extraction algorithm for foreground detection in dynamic scenes," *IEEE Transactions on Circuits and Systems for Video Technology*, 2016.

[14] Dong Liang, Shun'ichi Kaneko, M Hashimoto, K Iwata, X Zhao, and Y Satoh, "Co-occurrence-based adaptive background model for robust object detection," in *AVSS*. IEEE, 2013, pp. 401–406.

[15] Dong Liang, Shun'ichi 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.

[16] Nil Goyette, Pierre-Marc Jodoin, Fatih Porikli, Janusz Konrad, and Prakash Ishwar, "Changedetection.net: A new change detection benchmark dataset," in *CVPR CDW*. IEEE, 2012, pp. 1–8.

[17] S. Brutzer, B. Hoferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *CVPR*. IEEE, 2011, pp. 1937–1944.

[18] Martin Hofmann, Philipp Tiefenbacher, and Gerhard Rigoll, "Background segmentation with feedback: The pixel-based adaptive segmenter," in *CVPR CDW*. IEEE, 2012, pp. 38–43.

[19] Rita Cucchiara, Costantino Grana, Massimo Piccardi, and Andrea Prati, "Detecting moving objects, ghosts, and shadows in video streams," *TPAMI*, vol. 25, no. 10, pp. 1337–1342, 2003.

[20] Andrea Prati, Ivana Mikic, Mohan M Trivedi, and Rita Cucchiara, "Detecting moving shadows: algorithms and evaluation," *TPAMI*, vol. 25, no. 7, pp. 918–923, 2003.

[21] Andres Sanin, Conrad Sanderson, and Brian C Lovell, "Shadow detection: A survey and comparative evaluation of recent methods," *Pattern recognition*, vol. 45, no. 4, pp. 1684–1695, 2012.

[22] S. H. Khan, M Bennamoun, F Sohel, and R Togneri, "Automatic shadow detection and removal from a single image," *TPAMI*, vol. 38, no. 3, pp. 431, 2016.