

# **Anomaly Detection in Surveillance Video of Natural Environment**

Silas Santiago L. Pereira  
Federal Institute of Education, Science  
and Technology of Ceará - IFCE  
62800-000 Aracati-CE, Brazil

José E.B. Maia  
Centro de Ciências e Tecnologia - CCT  
Universidade Estadual do Ceará - UECE  
60714-903 Fortaleza-CE Brasil

## **ABSTRACT**

This work demonstrates the effectiveness of the median filter combined with morphological operators in the detection of anomalies in video surveillance of scenes of natural environment. Natural environment is characterized by backgrounds that are not static but whose dynamics are limited and do not include the appearance or disappearance of background objects in the scene. Examples include background images with seawater or river surfaces, or landscapes with trees, in which the wind produces waves and other movements of limited amplitude. The performance on four publicly available benchmark videos is compared to that of other published state-of-the-art works. The results obtained are promising.

## **Keywords**

Anomaly detection, background extraction, median filter, morphology operator, intelligent video surveillance

## **1. INTRODUCTION**

Video anomaly detection is a challenging problem in the computer vision area. Since it involves subjectivity, depends on localization and context, and can vary in content and duration, the definition of an anomaly could become a complex task. Beyond that, there is also the difficulty in the acquisition of anomaly samples. Perform frame and pixel-level annotations could be a laborious and expensive task, which leads to the creation of frequently unbalanced datasets and the disseminated use of unary classification approaches, although the existence of works on literature about binary classification [20].

Anomaly detection algorithms are typically effective in specific contexts. This study demonstrates the effectiveness of the combination of median filter and morphological operators for video surveillance anomaly detection in the context of natural environments. The environments considered are characterized by non-static backgrounds in which the dynamics do not include the entry or exit of objects in the background scene. Examples of these scenarios include backgrounds with water surfaces in seas and rivers, and natural landscapes with trees, in which the wind produces ripples and other movements with limited amplitude.

An anomaly can be defined as a pattern whose characteristics deviate from normal patterns [16] in a way that anomalous events occurs with low frequency among normal ones. An anomalous event

is also considered a rare situation, so they have a low probability to occur among the majority of normal events. In this sense, an anomalous event can be detected based on its low probability of occurrence in a given context. The context is fundamental to the decision-making process, once an anomalous activity in a given context can be considered normal in another context [16]. The low probability of occurrence of anomalous events is the concept used in the construction of the decision boundary for one class classifiers.

Background modeling is an essential step in the general procedure for anomaly detection. The literature presents different approaches for background estimation from an initial sequence of training frames. Frequent approaches include Kalman filter approaches [17], Gaussian mixtures [5], kernel density estimation [12], frame difference [11], median filtering [4], among others. In addition to these methods, subspace learning-based approaches have obtained promising results in background extraction problems [7]. Representatives of these methods are approaches based on Principal Component Analysis (PCA). Specifically, RPCA (Robust PCA) models have become a benchmark approach in the literature for the task of background extraction [10].

The performance scores of many published approaches in video anomaly detection are usually high or satisfactory, even in applications with completely dynamic and confusing scenarios. The hypothesis here is these methods present excessive complexity when built and applied in lesser confusing scenarios. In this sense, more simple strategies could perform well in the context of the natural environment, which is the focus in this work. Representatives of complex approaches are those which involve the definition and processing of spatiotemporal video volumes defined by partially overlapped frame-blocks, a necessary technique for challenging backgrounds [18]. In this way, the main contribution of this work is to demonstrate the effectiveness of the combination of the median filter approach and morphological operators in the context of video surveillance anomaly detection in natural environments. Additionally, three manifold-based approaches are evaluated and compared for the background modeling phase.

This work is organized as follows: Section 2 reviews some works directly related to this research, and Section 3 describes the methodology of the work. Section 4 presents and discusses the results and Section 5 concludes the article.

## 2. RELATED WORK

This section is a brief review of some works directly related to this one, in order to contextualize the present research in the literature. Recent surveys that expand the coverage of this section can be found in [21] e [6].

The study of [10] performs a comparison of Robust PCA and median filter approaches for static backgrounds. The study takes into account the robustness of compared models to noise and blur. Real and synthetic datasets are considered in evaluation performance experiments. The described approaches presuppose no alteration of background along with the frames in the scene so that the matrix representation of the static background consists of a set of identical columns, which allows the use of a median filtering approach for background representation. The authors also emphasize that due to the high dimensionality in video surveillance videos, the matrix operations necessary for singular value decompositions in RPCA models could involve high computational costs. From the results, the median filter performs better than RPCA and is also computationally faster.

The authors in [18] present a non-supervised real-time surveillance framework based on spatiotemporal volumes. Normal behavior statistical models are built for partially overlapped frames subsequences named spatiotemporal volumes. Anomalous events are described as configurations with low-occurrence frequency. The considered experiments involve distinct datasets with different anomaly patterns. The obtained experiments indicate the proposed approach is comparable to state of art in accuracy terms.

On the other hand, in [9], the proposed non-supervised approach is based on autoencoders trained over normal datasets. Five benchmark datasets were considered in the experiments (KDDCUP, MNIST, CIFAR-10, CatVsDog e UCF-Crime). The normal and anomalous sub-datasets are obtained by clustering techniques so that normal patterns are used to learn representations with the autoencoder technique. The proposed approach outperforms existing unsupervised techniques, in terms of accuracy and robustness. However, the authors argue that autoencoders may not enough to deal with complex patterns.

In [20], the authors explore video anomaly detection as a binary classification problem. The goal is to investigate an anomaly detection method that maximizes the distance among examples of different classes and minimizes distance among examples of the same class. The authors also explore the weakly supervised learning paradigm from Multiple-Instance Learning (MIL) approach to allow segment-level classification instead of frame-level. In this sense, only video-level labels are considered during the training phase. The authors propose, evaluate and compare the framework AR-NET (Anomaly Regression Net) for video anomaly detection. Moreover, the authors propose and evaluate two new cost functions applied to learn discriminants for anomaly detection, called Dynamic Multiple-Instance Learning Loss and Center Loss, respectively. The feature extraction is performed by using the pre-trained neural network model Inception-v1 I3D (Inflated 3D), which uses appearance (RGB) and motion information (optical flow). The experiments considered the challenge benchmark dataset ShanghaiTech. The authors evaluated the proposed approach with state-of-art literature. From the obtained results, the proposed model overcomes in performance the compared approaches for the experimented dataset. Moreover, from a subjective analysis, the authors also observe that anomaly detection for some scenarios is still a challenge for state-of-art models.

In [8], two approaches are applied to perform the detection of abrupt behavior changes and anomalous events: change points detection and topic modeling. The topic modeling approach is completely non-supervised and allows comprehensive analysis of video sequences. The detection of change points is used to identify global anomalies to detect sudden changes in global behaviors. This approach can be applied for unforeseen situations and a video could be represented as a temporal series.

In [3], the authors evaluate the application of spatiotemporal descriptors DTF (Dense Trajectory Based Features) e SOF (Silhouettes and Optic Flow-based Features) for the video anomaly detection problem. Normal sequence representations based on these descriptors are used to build models with the one-class SVM algorithm. The authors evaluate the performance on five datasets, and the obtained results are competitive with state-of-art. The DTF approach overcomes the literature performance for crowded scenarios, and the SOF approach obtained considerable performance for some datasets.

In [16], the authors present three adaptative inference mechanisms for the detection of anomalous patterns in video surveillance scenarios. The first mechanism is to detect an anomalous event with a short-term duration. The second one is for previously defined anomalies, and the last is intended to object detection involving contexts. The modeling takes into account the construction of non-overlapped volumes and ensembles over different scales. These volumes are used to generate representations named codewords, which are used for probabilistic modeling. The authors evaluate the proposed approach for seven distinct datasets for video anomaly detection. From the obtained results, the proposed approach is promising and presents a detection rate of 91.35%.

The results of these two last studies [3, 16] were selected for performance comparison in this work. This choice is because two of the used datasets are also used in this study and are their performances are comparable to other state-of-art works.

## 3. METHODOLOGY

This section describes the related steps for data preparation, modeling, and performance evaluation applied in the proposed approach for anomaly detection in video surveillance scenarios. The general procedure is composed of conjunction of steps, which are detailed in the following subsections and is described in the algorithm 1.

### 3.1 Background Modeling

Let  $M$  be a  $m \times p$  matrix whose columns are formed by vectors obtained by flattening a sequence of  $p$  frames with  $m$  pixels each. Usually  $m \gg p$ . The matrix is also centralized by rewriting it as:

$$M = [M_1 - \bar{M}, \dots, M_p - \bar{M}] \quad (1)$$

where  $\bar{M} = \frac{1}{p} \sum_i^p M_i$  and  $M_i$  is the  $i$ -th column of  $M$ . The matrix  $M$  is obtained from a sequence of frame images  $[I_1, I_2, \dots, I_p]$ , where each instance is originally an image  $I_i$  with dimension  $width \times height$  pixels.

In all the *background* modeling methods evaluated in this work, they look for the following decomposition [19]:

$$M = L + S \quad (2)$$

where  $L$  is a low rank matrix (*low rank*) in relation to  $p$ , and  $S$  is a sparse matrix.

The  $L$  matrix will be a background model and the  $S$  matrix will contain a representation of foreground in addition to various noises

**Algorithm 1** Macro-steps of modeling and evaluation in the anomaly detection procedure.

- 1: Load a grayscale frame sequence  $F = [f_1, f_2, \dots, f_N]$
- 2: Resize all frames in  $F$
- 3: Load the respective set of discrete labels  $Y = [y_1, \dots, y_N]$ , where  $y \in \{0, 1\}$ ;
- 4: Select the training sequence  $(F_{train}, Y_{train})$  and test sequence  $(F_{test}, Y_{test})$  from  $(F, Y)$ , so that  $F_{train}$  is composed only of normal frames;
- 5: Extract background model using  $k$  initial frames in  $F_{train}$ ;
- 6: Compute  $F'_{train}$  and  $F'_{test}$ , from the background subtraction followed by binarization by an automatic threshold;
- 7: Search the appropriate erosion and dilation kernels used in the morphological operators;
- 8: Apply morphological processing in sequences  $F'_{train}$  and  $F'_{test}$  to better distinguish the foreground from noise;
- 9: Train the unary classification algorithm OCSVM with  $F'_{train}$
- 10: Evaluate the predictive built model with  $F'_{test}$

and residues from the background modeling. In the video surveillance applications analyzed here, the presence of foreground will be treated as an anomaly to be detected, and the algorithm must be able to distinguish the presence of foreground from the different types of noise that may be present, in addition to of the residuals of possible imperfections in the representation of the background. Note that after building a model for  $L$ ,  $S$  can be obtained by the relationship  $S = M - L$ . In the case of a set of frames is available at design time, which are known to consist only of background, the strategy is to build  $L$  from this data and obtain  $S$  for test frames by the relation  $S = M - L$ . Several techniques can be used to obtain a background model  $L$ .

**3.1.1 Median Filtering.** The median filtering approach approximates the foreground detection problem with dynamic background by a static background model. In this case, the background is dynamic but with limited movements present in natural landscapes. By this approximation, the matrix decomposition for  $M$  can be written as:

$$M = \mathbf{u}\mathbf{1}^T + S \quad (3)$$

where  $\mathbf{u}\mathbf{1}^T$  corresponds to the background matrix  $\mathbf{u} \in \mathbf{R}^m$  and  $\mathbf{1} \in \mathbf{R}^n$ . Thus, given a  $M$  project matrix, the background model is obtained by solving the following optimization problem:

$$\min_{\mathbf{u} \in \mathbf{R}^m} \|\mathbf{u}\mathbf{1}^T - M\|_1 \quad (4)$$

where  $\|\cdot\|_1$  is a  $l_1$  norm. According to [10], this problem has a closed form solution, which is given by the following expression:

$$\text{Median}\{M_{i,1}, M_{i,2}, \dots, M_{i,m}\} \quad (5)$$

where  $M_{i,j}$  is a given  $(i, j)$  position in the  $M$  matrix.

**3.1.2 PCA/MCA - Principal/Minor Component Analysis.** [13] PCA is an unsupervised technique that aims to rotate the axes of the representation space of a data matrix, maintaining the orthogonality. When the data is projected on the new axes, the explained variance is maximized. PCA can be calculated from different structures such as covariance or correlation matrices. In this work, was used SVD (Singular Value Decomposition) technique. SVD performs the following decomposition of the data centered matrix:

$$M_{m \times p} = U_{m \times m} \cdot D_{m \times p} \cdot V_{p \times p}^T \quad (6)$$

where  $U \in \mathbf{R}^{m \times m}$  are  $V \in \mathbf{R}^{p \times p}$  are orthogonal matrices for which their columns correspond to the eigenvectors of  $M \cdot M^T$  e  $M^T \cdot M$ , respectively. The diagonal matrix  $D$  is formed by singular elements, which are the square roots of the  $\lambda_i$  of  $M \cdot M^T$  (or  $M^T \cdot M$ ). These eigenvalues are usually ordered so that  $\lambda_i \geq \lambda_{i+1}$ , for  $i = 1, 2, \dots, p - 1$ . For data with  $p$  columns, the PCA technique will project the original data into a subspace of dimension  $d$ , where the axes correspond to the  $d$  largest singular values. In MCA, these data are projected in the dimension subspace  $p - d$ , of the axes corresponding to the  $p - d$  lowest single values. Intuitively, PCA will preserve more details of the background, in addition, to also preserve the noise present in data. MCA will filter the noise, but with the cost of losing details present in the background. Once the SVD decomposition is obtained, the data projection in a component subspace composed of  $d \ll p$  eigenvectors (the first  $d$  columns of  $U_{m \times m}$  in the case of PCA, and the last  $p - d$  columns in the case of MCA) is finally obtained from the following expression:

$$M' = \hat{M} + U_{m \times d} \cdot U_{d \times m}^T \cdot (M - \hat{M})_{m \times p} \quad (7)$$

**3.1.3 RPCA - Robust Principal Component Analysis.** The RPCA (Robust Principal Component Analysis) technique has been considered a benchmark approach in background extraction tasks. It allows a matrix decomposition by the following expression:

$$M = L + S \quad (8)$$

where the background component is expressed by a low rank matrix  $L$ , while the foreground objects will be present in a sparse outlier matrix  $S$  [10]. According to [2], the matrices  $L$  and  $S$  can be obtained from the resolution of the convex optimization model expressed by the following equation:

$$\min_{L, S \in \mathbf{R}^{m \times n}} \{\|L\|_* + \tau \|S\|_1 \mid L + S = M\} \quad (9)$$

where  $\|\cdot\|_*$  is a nuclear norm and  $\|\cdot\|_1$  is a  $l_1$  norm [10].

## 3.2 Background subtraction, Binarization and Morphological Processing

Having the background model, and given a new gray-scale frame, the background subtraction followed by the binarization of resulting image are performed in a single operation. Let  $Bg$  a background representation,  $v$  the current frame, and  $T$  a given constant threshold. The background subtraction and binarization can be described by the following expression:

$$Fg = [|v - Bg| < T] \quad (10)$$

which limits the difference between this current frame  $v$  and the reconstructed background image [7]. The determination of  $T$  is automatic and by using the average pixel intensity in the background image.

For instance, the Figure 1 describes the a pareto histogram for the background image in BOAT RIVER dataset. The dashed vertical line corresponds to the average pixel intensity value in this frame of the 127.14, which corresponds to 49.67% of pixel intensity.

The erosion and dilation operations are fundamental in the morphological processing step. The binarized image is then submitted to the morphological operations erosion and dilation. This step intends to smooth and improve the obtained region boundaries. The morphological processing allows the extraction of image components that can be useful in the shape description of a given region [14].

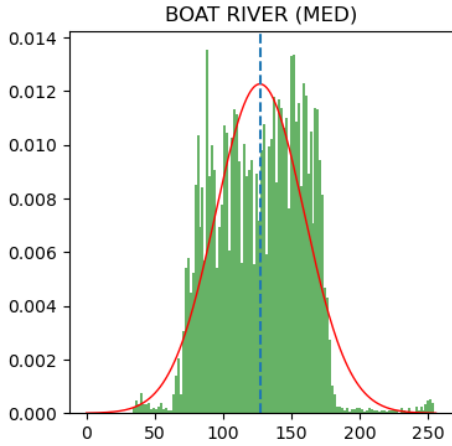


Fig. 1: Determination of the image binarization threshold.

Erosion is a diminishing or thinning operation, and it is used to remove noise that may impact in the final classifier performance. The erosion of  $A$  by  $B$  is given by the following expression [14]:

$$A \ominus B = \{z | (B)_z \subseteq A\} \quad (11)$$

where  $A, B$  are sets in  $\mathbf{Z}^2$ , and  $A \ominus B$  expresses the set of points  $z$  such that  $B$ , when translated by  $z$ , is contained in  $A$ .

Dilation is a magnifying operation, and is used to emphasize the presence of an object in a image. This is given by the following expression:

$$A \oplus B = \{z | (\hat{B})_z \cap A \neq \emptyset\} \quad (12)$$

which expresses the reflection of  $B$  around its origin, followed by the translation of this reflection by  $z$ .

In Figure 2 illustrates the application of morphological operators erosion and dilation in the binarized images for a frame in BOAT RIVER video:

The kernels were optimized for each experimented video from a iterative search procedure that considered different filter dimensions in order to maximize the balanced accuracy of the final classifier. The search for better kernel parameters proved to be a computationally expensive procedure at runtime ( $i = 1 \dots 10$ ), given the high dimensionality of the image attributes space. For this reason, the PCA technique is used specifically in this step to data dimensionality reduction in the successive evaluations of generated models from data processed with different kernel dimensions. The kernel selection search can be summarized by the following optimization rule:

$$kernel \leftarrow \arg \max_k bACC_{OCSVM} | kernel = k \quad (13)$$

Were found filters with dimension  $3 \times 3$ ,  $4 \times 4$ ,  $5 \times 5$  and  $5 \times 5$  for datasets BOAT RIVER, BOAT SEA, CANOE and BOAT CHARLES RIVER, respectively.

### 3.3 One-class Classification

One-class Classification (OCC) is used in a context in which there is a target class, called normal class, and you desire to distinguish future patterns in relation to this class, as is the case of the detection of outliers, novelties, or anomalies. In model design time, there are available data only from normal class. OCC has also proved to

be an adequate approach when you need to deal with unbalanced classes [15]. The general idea is to build a data description (DD) in the form of a decision boundary, combining the normal class data and classifying any pattern which lies from this boundary as out of class. This procedure can be accomplished via several techniques such as K-Means, Gaussian Mixture Model, or Support Vector Machine (SVM). The One-class SVM (OCSVM) unary classification approach was utilized in this study. OCSVM solves the following optimization problem, in the primal formulation:

$$\min_{w, \xi, \rho} \frac{1}{2} \|w\|^2 + \frac{1}{vN} \sum_i^N \xi_i - \rho \quad (14)$$

where  $R$  is the hypersphere radius around normal class data centered on the average point  $c$ ,  $\xi_i > 0$  are slack variables to allow some training data points to be considered as outliers, and  $C$  is a free parameter which allows adjusting the weight relative to the extreme points in the optimization process.

In the model application phase, the ocsvm predictive model will classify as anomalies the new points whose distances from center  $c$  are beyond the hypersphere radius  $R$ . In the application of this study, this center  $c$  is represented by background model  $Bg$ . Once known previously that the training data do not have outliers, the hyperparameter  $C$  should be made small:

$$\langle w, \Phi(x_i) \rangle \geq b - \xi_i, \quad \xi_i > 0 \quad (15)$$

where  $\xi_i$  are slack variables,  $\Phi(x)$  is a kernel mapping of the input space to a feature space,  $N$  is the number of training samples,  $v$  is a weight parameter and  $b$  is a bias, which corresponds to the center of the sample. After solving this optimization problem for  $w$ , a new pattern  $x$  is classified by testing the condition  $sgn(\langle w, \Phi(x) \rangle - b)$  as negative for anomaly.

## 4. RESULTS AND DISCUSSION

### 4.1 Background Similarities per Dataset

To verify the similarity level among the produced background models, the metric Single Scale Structural Similarity (SSIM), which allows measuring the similarity between two images  $x$  e  $y$  de  $N$  pixels [22]. This computation is given by the expression

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c1)(2\sigma_{xy} + c2)}{(\mu_x^2 + \mu_y^2 + c1)(\sigma_x^2 + \sigma_y^2 + c2)} \quad (16)$$

where  $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2$  correspond to the means and variances of  $x$  and  $y$ , respectively, and  $c_1$  e  $c_2$  are computed constant values included to avoid instability when  $\mu_x^2 + \mu_y^2$  tend to zero.

Table 1 shows the computed similarities among the four generated backgrounds. Notice the different levels of similarities among the images.

### 4.2 Performance Evaluation Results

The experimental results are displayed in Table 2 and describe the achieved results from performance evaluation of the OCSVM classifier from training and evaluation post-processing data that considered different background models. It was consider the balanced accuracy ( $BACC = \frac{TPR+TNR}{2}$ ), the precision of anomalies ( $PREC = \frac{TP}{TP+FP}$ ), false-positive rate ( $FPR = \frac{FP}{FP+TN}$ ), true-positive rate ( $recall = \frac{TP}{TP+FN}$ ) and the harmonic mean of

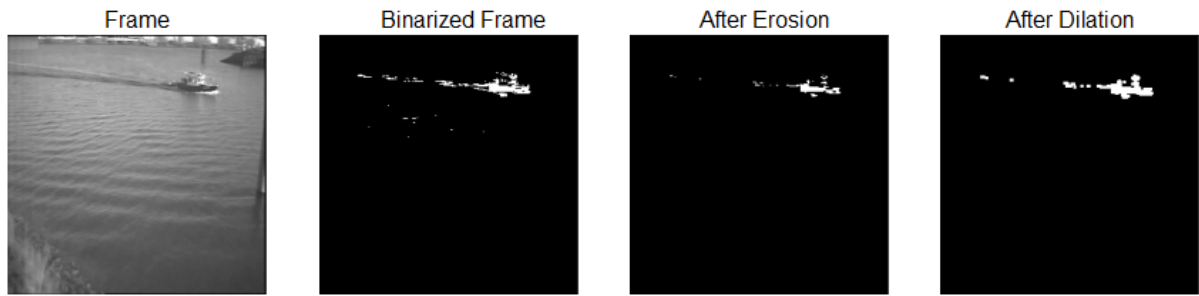


Fig. 2: An example of binarization followed by morphological processing of an image.

Table 1. : Structural similarity ( $SSIM()$ ) between the background models found by the various methods, for each dataset.

	BOAT RIVER	BOAT SEA	CANOE	BOAT CHARLES RIVER
MED x MCA	0.9412	0.9165	0.9696	0.9662
MED x PCA	0.3963	0.6939	0.9504	0.8064
MED x RPCA	0.3647	0.6508	0.6676	0.7807
MCA x PCA	0.4000	0.7070	0.9727	0.8080
MCA x RPCA	0.3711	0.6620	0.6696	0.7867
PCA x RPCA	0.2143	0.5731	0.6957	0.6733

precision and recall F1-Score  $F1 = \frac{precision \cdot recall}{precision + recall}$  as evaluation metrics in each experimented video. It can be observed the constructed OCSVM models achieve expressive results for the 4 considered datasets. It can be also verify that the predictive models achieve approximate results among the different background approaches in each experimented dataset. Beyond that, the median filter approach is computationally more efficient, and this one was enough to achieve comparable results to the subspace learning-based techniques.

In order to verify the number of positive and negative examples predicted by the classifier, the confusion matrix for the median filtering approach is described in Figure 3.

The achieved performance results were also compared with the studies of [3] and [16], according to Table 4. The authors in [16] evaluate an approach for unusual event detection, where it is considered that a test video contains an unusual event. The authors measure the unusual event detection rate (true positives among predicted examples as unusual events). The authors in [3] consider the recognition rate of anomalous activities.

It can be verify that the proposed approach achieves comparable results to the state-of-art literature. The compared state-of-art models are possible with high complexity and therefore capable of more complex anomaly detection tasks. Nonetheless, although the proposed approach has less capacity, can be verified the expressive achieved results, possibly due to the approach be adequate to the considered detection task.

An analogy can be made to Occam's knife [1] as to the results obtained here. This presumption states that the simplest hypothesis is the one that best fits a dataset.

## 5. CONCLUSION

This work presented an approach to detect anomalies in video surveillance of natural environments. The performance of the general procedure was evaluated when the background modeling technique was varied and compared with that of some works published.

Four different techniques were used to model the background: PCA, MCA, RPCA and MEDIANA filtering. The tests used four data sets of publicly available natural videos that are benchmarks used in other published works.

There are two main conclusions of this research: that the performance of the proposed general procedure is superior or equal to that of the published state of the art and that, in the scenarios of the tested data sets, the four methods evaluated for modeling of the background showed similar performance. Since the median filter is the lightest computationally, this is the recommendation of this work.

This research also opens up some lines for future investigation. A first is that the tests need to be expanded to videos whose anomaly events are more challenging. A second step is to quantify the effects of morphological processing, the type of OCC classifier used and the image reduction factor on performance. These last two have an impact on real-time processing.

## 6. REFERENCES

- [1] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K Warmuth. Occam's razor. *Information processing letters*, 24(6):377–380, 1987.
- [2] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):1–37, 2011.
- [3] KG Manosha Chathuramali, Sameera Ramasinghe, and Ranga Rodrigo. Abnormal activity recognition using spatio-temporal features. In *7th International Conference on Information and Automation for Sustainability*, pages 1–5. IEEE, 2014.
- [4] Rita Cucchiara, Costantino Grana, Massimo Piccardi, and Andrea Prati. Detecting moving objects, ghosts, and shadows in video streams. *IEEE transactions on pattern analysis and machine intelligence*, 25(10):1337–1342, 2003.

Table 2. : Performance evaluation of the anomaly detection procedure for each dataset.

	BOAT RIVER					BOAT SEA				
	bACC	PREC	FPR	REC	F1	bACC	PREC	FPR	REC	F1
MED	0.9309	0.9912	0.1000	0.9617	0.9762	0.9764	0.9953	0.0154	0.9681	0.9815
RPCA	0.8973	0.9867	0.1500	0.9447	0.9652	0.9771	0.9953	0.0154	0.9696	0.9823
PCA	0.9117	0.9909	0.1000	0.9234	0.9559	0.9802	0.9977	0.0077	0.9681	0.9827
MCA	0.9181	0.9910	0.1000	0.9362	0.9628	0.9764	0.9953	0.0154	0.9681	0.9815
	CANOE					BOAT CHARLES RIVER				
	bACC	PREC	FPR	REC	F1	bACC	PREC	FPR	REC	F1
MED	0.9979	1.0000	0.0000	0.9958	0.9979	0.9738	0.9984	0.0028	0.9504	0.9738
RPCA	0.8929	0.9875	0.2143	1.0000	0.9937	0.9716	0.9984	0.0028	0.9459	0.9714
PCA	0.9265	0.9916	0.1429	0.9958	0.9937	0.9729	1.0000	0.0000	0.9459	0.9722
MCA	0.9601	0.9958	0.0714	0.9916	0.9937	0.9744	1.0000	0.0000	0.9489	0.9738

Table 3. : Confusion Matrices for MEDIAN Filtering + OCSVM Approach (N = normal, A = anomaly)

	BOAT RIVER			BOAT SEA			CANOE			BOAT C. RIVER	
	N	A		N	A		N	A		N	A
N	18	2	N	384	6	N	14	0	N	361	1
A	9	226	A	42	1275	A	1	236	A	33	632

Table 4. : Performance comparison with other published works.

	BOAT RIVER		BOAT SEA		CANOE	
	PREC	REC	PREC	REC	PREC	REC
<b>Rao et al (2017)</b>	0.6666	-	-	-	1.0000	-
<b>Chaturamali et al. (2014)</b>	-	-	-	1.0000	-	-
<b>Proposed (MED)</b>	0.9912	0.9617	0.9953	0.9681	1.0000	0.9958

- [5] Nir Friedman and Stuart Russell. Image segmentation in video sequences: A probabilistic approach. *arXiv preprint arXiv:1302.1539*, 2013.
- [6] Belmar Garcia-Garcia, Thierry Bouwmans, and Alberto Jorge Rosales Silva. Background subtraction in real applications: Challenges, current models and future directions. *Computer Science Review*, 35:100204, 2020.
- [7] Charles Guyon, Thierry Bouwmans, El-hadi Zahzah, et al. Robust principal component analysis for background subtraction: Systematic evaluation and comparative analysis. *Principal component analysis*, 10:223–238, 2012.
- [8] Olga Isupova. *Machine learning methods for behaviour analysis and anomaly detection in video*. Springer, Cham, Switzerland, 2018.
- [9] Tangqing Li, Zheng Wang, Siying Liu, and Wen-Yan Lin. Deep unsupervised anomaly detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3636–3645, 2021.
- [10] Xinxin Li, Michael K Ng, and Xiaoming Yuan. Median filtering-based methods for static background extraction from surveillance video. *Numerical Linear Algebra with Applications*, 22(5):845–865, 2015.
- [11] Davide A Migliore, Matteo Matteucci, and Matteo Naccari. A reevaluation of frame difference in fast and robust motion detection. In *Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks*, pages 215–218, 2006.
- [12] Anurag Mittal and Nikos Paragios. Motion-based background subtraction using adaptive kernel density estimation. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 2, pages II–II. Ieee, 2004.
- [13] Erkki Oja. Principal components, minor components, and linear neural networks. *Neural networks*, 5(6):927–935, 1992.
- [14] Hélio Pedrini and William Robson Schwartz. *Análise de imagens digitais: Princípios. Algoritmos e Aplicações*, Thomsom, 2008.
- [15] Pramuditha Perera, Poojan Oza, and Vishal M Patel. One-class classification: A survey. *arXiv preprint arXiv:2101.03064*, 2021.
- [16] TJ Narendra Rao, GN Girish, and Jeny Rajan. An improved contextual information based approach for anomaly detection via adaptive inference for surveillance application. In *Proceedings of International Conference on Computer Vision and Image Processing*, pages 133–147. Springer, 2017.
- [17] Christof Ridder, Olaf Munkelt, and Harald Kirchner. Adaptive background estimation and foreground detection using kalman-filtering. In *Proceedings of International Conference on recent Advances in Mechatronics*, pages 193–199. Cite-seer, 1995.
- [18] Mehrrsan Javan Roshtkhari and Martin D Levine. An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions. *Computer vision and image understanding*, 117(10):1436–1452, 2013.

- [19] Baidya Nath Saha, Nilanjan Ray, and Hong Zhang. Snake validation: A pca-based outlier detection method. *IEEE Signal Processing Letters*, 16(6):549–552, 2009.
- [20] Boyang Wan, Yuming Fang, Xue Xia, and Jiajie Mei. Weakly supervised video anomaly detection via center-guided discriminative learning. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2020.
- [21] Guojun Yin, Bin Liu, Huihui Zhu, Tao Gong, and Nenghai Yu. A large scale urban surveillance video dataset for multiple-object tracking and behavior analysis. *arXiv preprint arXiv:1904.11784*, 2019.
- [22] Wang Zhou. Image quality assessment: from error measurement to structural similarity. *IEEE transactions on image processing*, 13:600–613, 2004.