

Received July 14, 2019, accepted July 25, 2019, date of publication July 30, 2019, date of current version August 15, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2932137

# Random Finite Set-Based Anomaly Detection for Safety Monitoring in Construction Sites

**AMMAR MANSOOR KAMOONA<sup>1,3</sup>, (Student Member, IEEE), AMIRALI KHODADADIAN GOSTAR<sup>1</sup>, RUWAN TENNAKOON<sup>1</sup>, ALIREZA BAB-HADIASHAR<sup>1</sup>, DAVID ACCADIA<sup>2</sup>, JOSHUA THORPE<sup>2</sup>, AND REZA HOSEINNEZHAD<sup>1</sup>**

<sup>1</sup>School of Engineering, RMIT University, Melbourne, VIC 3000, Australia

<sup>2</sup>Cornerstone Solutions Pty., Ltd., Hawthorn, VIC 3122, Australia

<sup>3</sup>Department of Electrical Engineering, Faculty of Engineering, University of Kufa, Najaf, Iraq

Corresponding author: Reza Hoseinnezhad (rezah@rmit.edu.au)

This work was supported by the Australian Research Council through the ARC Linkage Project under Grant LP160101081.

**ABSTRACT** Low visibility hazard detection in construction sites is a crucial task for prevention of fatal accidents. Manual monitoring of construction workers to ensure they follow the safety rules (e.g., wear high-visibility vests) is a cumbersome task and practically infeasible in many applications. Therefore, an automated monitoring system is of both fundamental and practical interest. This paper proposes an intelligent solution that uses live camera images to detect workers who breach safety rules by not wearing high-visibility vests. The proposed solution is formulated in the form of an anomaly detection algorithm developed in the random finite set (RFS) framework. The proposed system is comprised of three steps: 1) applying a deep neural network to extract people in the image; 2) extracting particularly engineered features from each blob returned by the deep neural network; and 3) applying the RFS-based anomaly detection algorithm to each set of detected features. The experimental results demonstrate that in terms of F1-score, the proposed solution (as the combination of the newly engineered features and RFS-based anomaly detection algorithm) significantly outperforms various combinations of common and the state-of-the-art features and anomaly detection algorithms employed in machine vision applications.

**INDEX TERMS** Random finite sets, construction safety, safety monitoring, Poisson point patterns, IID clusters, PHD filter, anomaly detection.

## I. INTRODUCTION

Safety in construction sites is an extremely important priority which safety monitoring and hazard detection tasks are necessary parts of operational practices in the construction industry to prevent fatal accidents and injuries. The current practice involves manual safety monitoring which is labor intensive and prone to human error. In addition, the complex environment of construction sites makes the safety monitoring highly challenging. Therefore, development of an automated and yet reliable system for safety monitoring is of fundamental and practical interest, and has attracted significant attention from the intelligent systems community [1].

One of the common causes of hazards in complex environments such as construction sites is the *low visibility* to which the workers who are not wearing appropriate apparel

The associate editor coordinating the review of this manuscript and approving it for publication was Baoping Cai.

are exposed. In order to ensure the workers' safety, they require to wear *high-visibility safety apparel* (HVSA). There is an abundance of work in the computer vision literature in which machine vision techniques are proposed to automatically detect and identify workers with no HVSA [2]–[5]. With the recent advancements in computer vision, it is possible to detect and track workers, material, and equipment. However, one of the most challenging tasks is to analyze the content of the detected objects in order to distinguish non-desirable from desirable features.

The most common approach in recent works for low-visibility hazard detection is based on classifying observations into two or more classes [1], [6]. Classification-based methods commonly require a large amount of annotated/labeled data for their training. An alternative approach is *anomaly detection* which does not require the negative (anomalous) data to be annotated. In [1], the HSV color space is used to detect workers who wear safety vest in which they

used the histogram of H and S component in construction of their feature. In [6], different color space models were used to train different methods. However, they divide the training data based on the color of safety vest (yellow, orange) and train two different networks. Anomaly detection is at its core a classification method, however there are some important differences. In anomaly detection-based methods, the aim is to distinguish between *normal/positive* and *anomalous/negative* observations. Anomalous observations do not conform to the expected pattern of other observations in the dataset. To the best of our knowledge, no anomaly detection-based solution has been reported for safety monitoring in construction sites.

In a wide range of applications that involve information sources (data) other than images and videos, numerous anomaly detection techniques have also been developed. There are classification-based [7], clustering-based [8], statistical model-based [9], nearest neighbors-based [10], spectral-based [11], [12] and information theoretic [13] solutions.

Anomaly detection using images and video data has been also explored in different practical applications. Ramezani *et al.* [14] proposed a method for novelty detection in video streams for surveillance applications. In his solution, a kernel density estimator is used in conjunction with an evolving clustering approach suitable for real-time applications in fully autonomous and unsupervised systems. Also, Diehl and Hampshire [15] proposed a real-time anomaly detection mechanism for video surveillance based on extracting monochromatic spatial features in image sequences to represent moving objects. Anomaly detection has also been used to monitor the wildlife. Yong *et al.* [16], [17] proposed to solve the anomaly detection problem in wildlife scenes by classifying an image in terms of its scene class; if the image does not belong to any scene class, it is considered as an anomaly. Singh and Markou [18], [19] proposed a framework for anomaly detection in terms of environmental conditions. In their solution, they employ neural networks as adaptive classifiers that are capable of anomaly detection. Sultani *et al.* [20] proposed an anomaly detection framework for real world surveillance videos by using pre-trained deep learning network for extracting spatial-temporal features followed by deep multiple instance learning for anomaly detection.

The focus of this paper is on formulating a parametric statistical solution (a *statistical model*, a branch of model based solutions [21] for anomaly detection. Such a solution, due to its statistical nature, would incorporate the existing or learned knowledge around uncertainties within the sensing mechanisms involved in the application. In a statistical model-based method, the assumption is that anomalies are distributed outside the positive data distribution [22] and have significantly lower likelihoods compared to positive data. In machine learning, many data sources are in the form of *point pattern* data. A point pattern is a set or multi-set of unordered points, where each point is a vector representing the state or features of an object of interest. Commonly, for

point data, Gaussian mixture models are used to represent positive data [23], [24] in which the parameters of the model could be learned by expectation maximization (EM).

Vo *et al.* [25] proposed the use of point processes to model point pattern data for anomaly detection via *multiple instance learning*. Their proposed approach is based on treating each point pattern as a Random Finite Set (RFS), and derive likelihood functions based on common RFS density assumptions (e.g. Poisson or IID Clusters). Employing the RFS framework allows the formulated likelihood function to incorporate both *cardinality* (the number of extracted features) information and individual feature information.

This paper proposes a novel anomaly detection solution for safety monitoring in construction sites. The proposed method uses live camera images to detect workers who breach safety rules by not wearing high-visibility vests. The entire solution is implemented in three steps. First the Faster R-CNN neural network is used to detect people in the image. Then a novel feature, particularly engineered for this application, is extracted from each blob returned by the Faster R-CNN. Finally, the RFS-based anomaly detection algorithm is applied to each set of extracted features.

The major points of contribution and novelty in this work are: (i) applying anomaly detection for the first time for the detection of safety vests not being worn by workers in construction sites, (ii) a novel solution that combines the power of deep learning for object detection (Faster R-CNN) with engineered features to achieve high detection performance, (iii) using the YCbCr color space for the first time, for feature extraction in anomaly detection as our investigations demonstrated that those features work best in differentiating anomalous events of safety vests not being worn from the normal/positive situation where all workers are abiding with the rules, and (iv) development of an RFS-based anomaly detection solution that unlike the state-of-the-art [6], different networks for the detection of different safety vests of different colors. Experimental results demonstrate that in terms of F1-score, the proposed solution (as the combination of the newly engineered features and RFS-based anomaly detection algorithm) significantly outperforms various combinations of common and state-of-the-art features and anomaly detection algorithms employed in machine vision applications.

The rest of this paper is organized as follows: Section II presents the design problem and a background on existing solutions in the literature. Section III covers the proposed algorithm. Section IV presents experimental results, and Section V concludes the paper.

## II. BACKGROUND

The main goal of this work is to detect persons who do not wear high-visibility vests in digital images taken from long distances. In such images, each person may cover a very small area and appear in low quality/resolution. Fig. 1 shows a sample of such images taken in a construction site. The workers cover a very small area and the containing blob is



**FIGURE 1.** A sample image taken in a construction site. The workers cover a small area and the image information related to each person in the blob containing them, is of low quality.

low resolution in contrast to the whole image that appears in very high quality.

This problem can be tackled by applying a classification solution. However, as mentioned before, classification methods usually require a large annotated dataset for all classes, which may not be available in many applications. Solving the problem via anomaly detection relaxes the need for a large training dataset with the class of interest (anomaly) being annotated, as anomaly detection only needs training data for the *normal/positive* situation which is usually available in abundance.

In this section, a brief background on anomaly detection methods using *naïve Bayes* and *Random Finite set* (RFS) approaches is introduced.

#### A. ANOMALY DETECTION: NAÏVE BAYES APPROACH

The naïve Bayes approach is commonly used to solve classification problems. It is formulated based on assuming independence between the measured or detected features used for classification.

Consider a data point denoted as a *measurement set*  $Z$  which is comprised of  $n$  presumably independent feature vectors  $z_1, \dots, z_n$ . A classifier is to assign a class label  $c \in \{c_1, \dots, c_m\}$  to the data point. A statistical classifier would generally assign the most probable class label to the data point:

$$\hat{c} = \underset{c_i}{\operatorname{argmax}} \Pr(c = c_i | Z).$$

From Bayes' rule, assuming prior class probabilities denoted by  $\{\Pr(c = c_i)\}_{i=1}^m$ , the classifier can be formulated as:

$$\hat{c} = \underset{c_i}{\operatorname{argmax}} \mathcal{L}(z_1, \dots, z_m | c = c_i) \Pr(c = c_i), \quad (1)$$

where  $\mathcal{L}(z_1, \dots, z_m | c = c_i)$  is the likelihood of the measured features given the class label. Assuming independent features, the above Bayes' classifier reduces to the following Naïve Bayes classifier [26]:

$$\hat{c} = \underset{c_i}{\operatorname{argmax}} \left( \prod_{i=1}^m \mathcal{L}(z_i | c = c_i) \right) \Pr(c = c_i). \quad (2)$$

When this approach is used for anomaly detection, there are only two possible classes: normal/positive and anomalous/negative. In the absence of sufficient data to train the likelihood for anomalous class, we can only compute the likelihood for the measurement representing the positive class. In this case, the approach turns into ranking the data points according to their likelihood values for the positive class, and considering those less than a user-defined threshold as anomalous [11].

Vo *et al.* [25] have shown that the naïve Bayes model cannot be used for ranking point pattern data due to unit measurement inconsistencies. They also suggested that the naïve Bayes model does not make use of the information contained in number of elements in point patterns (cardinality). Vo *et al.* [25] introduced a new method based on the Random Finite Set framework for data point patterns which will be introduced in following section.

#### B. RANDOM FINITE SET-BASED ANOMALY DETECTION

According to Chandola *et al.* [11], an effective solution for an anomaly detection problem should be designed based on the nature of the information acquired from sources/sensors and used for detection. When such information are in the form of data points, the natural approach is to treat them as point pattern data.

In the RFS-based anomaly detection methods, the data point is modeled as an RFS. The rationale is mainly the observation that in the set  $Z$ , not only the members  $z_i$  ( $i = 1, \dots, n$ ) vary randomly with time, but also the cardinality of the set  $|Z| = n$  may vary randomly with time. Hence, each measured set of features  $Z = \{z_i\}_{i=1}^{|Z|}$  is treated as an RFS, and its density with respect to measure  $\mu$  is given by [27]:

$$p(Z) = p_c(|Z|) (|Z|)! U^{|Z|} p_{|Z|}(z_1, \dots, z_{|Z|}), \quad (3)$$

where  $p_c(n) = \Pr(|Z| = n)$  is the cardinality distribution,  $U$  is the unit hyper-volume, and  $p_n(z_1, \dots, z_n)$  is a symmetric joint feature density for the given cardinality  $|Z| = n$  [28].

As it was mentioned earlier, for the purpose of anomaly detection, we only need to compute the likelihood of the measured set of features  $Z$  given that it represents the positive case (a worker wearing the high-visibility vest in our application). In that case, the likelihood is given by (3) in which the functions  $p_c(\cdot)$  and  $p_n(z_1, \dots, z_n)$  need to be learned. In practice, we can make some particular assumptions on the mathematical form of those functions, and estimate the parameters involved in those models through a training phase.

Examples of parametric RFS densities include the Poisson RFS [29], the Bernoulli RFS [27], the multi-Bernoulli RFS [30], the IID-cluster RFS [31], the labeled multi-Bernoulli RFS [32] and the generalized labeled multi-Bernoulli RFS [33]. Our investigations have shown that assuming Poisson densities is suitable for anomaly detection particularly in the application of interest (safety in construction sites). Most of the previously mentioned RFS densities haven been used for multi-object tracking. Their approach is

to treat the multi-object entity as a Random Finite Set and apply the prediction and update steps in a similar fashion applied in a traditional single-object tracking algorithm such as adaptive combination kernel filter [34].

An IID-cluster RFS  $Z = \{z_i\}_{i=1}^n$  satisfies the basic IID assumption for its feature elements. Hence, its density is given by:

$$p(Z) = p_c(|Z|)|Z|!U^{|Z|} [p_f]^Z, \quad (4)$$

where,  $p_f(\cdot)$  is the feature probability density and

$$[p_f]^Z \triangleq \prod_{z \in Z} p_f(z)$$

is a finite set exponential. With IID-cluster assumption, we need to assume a special form for the  $p_c(\cdot)$  and  $p_f(\cdot)$  functions and estimate their parameters through the training step. In a special case that turns an IID-cluster RFS into a Poisson RFS, the cardinality density is Poisson:

$$p(Z) = \rho^{|Z|} \exp(-\rho) [Up_f]^Z, \quad (5)$$

where  $\rho$  is a non-negative intensity of the Poisson RFS density, which is also the statistical mean of cardinality.

### C. TRAINING PHASE

In the training phase of the parametric statistical learning, the goal is to estimate parameters of the likelihood function that best fits the training data (associated with positive cases) using the maximum likelihood estimator (MLE). In the RFS framework, assuming that the cardinality distribution and multi-feature joint density are parameterized as  $p_c(n; \theta_c)$  and  $p_n(z_1, \dots, z_n; \theta_f)$ , respectively, then the training phase turns into the estimation of the parameters  $\theta_c$  and  $\theta_f$  as follows:

$$\begin{aligned} (\hat{\theta}_c, \hat{\theta}_f) &= \operatorname{argmax}_{\theta_c, \theta_f} \prod_{\text{positive } Z^*} \mathcal{L}(\theta_c, \theta_f | Z^*), \\ &= \operatorname{argmax}_{\theta_c, \theta_f} \prod_{\text{positive } Z^*} \mathcal{L}(Z^* | \theta_c, \theta_f) p(\theta_c, \theta_f), \end{aligned} \quad (6)$$

where  $Z^*$  is a single set of measurements in the training dataset. In the absence of any prior information on the parameters, prior density  $p(\theta_c, \theta_f)$  can be assumed constant. Hence,

$$(\hat{\theta}_c, \hat{\theta}_f) = \operatorname{argmax}_{\theta_c, \theta_f} \prod_{\text{positive } Z^*} \mathcal{L}(Z^* | \theta_c, \theta_f). \quad (7)$$

With Poisson assumption for the density of feature set given the parameters,  $\mathcal{L}(Z | \theta_c, \theta_f)$ , equation (7) is simplified to:

$$\begin{aligned} (\hat{\theta}_c, \hat{\theta}_f) &= \operatorname{argmax}_{\theta_c, \theta_f} \prod_{\text{positive } Z^*} p_c(|Z^*|; \theta_c) |Z^*|! U^{|Z^*|} [p_f(\cdot; \theta_f)]^{Z^*}. \end{aligned} \quad (8)$$

Vo et al. [25] have proven that the above optimization can be decomposed into separate optimizations for the parameters of cardinality and feature distributions:

$$\hat{\theta}_c = \operatorname{argmax}_{\theta_c} \prod_{\text{positive } Z^*} p_c(|Z^*|; \theta_c) \quad (9)$$

$$\hat{\theta}_f = \operatorname{argmax}_{\theta_f} \prod_{\text{positive } Z^*} [p_f(\cdot; \theta_f)]^{Z^*}. \quad (10)$$

Note that in equation (9), the terms  $|Z^*|!U^{|Z^*|}$  have been removed because they do not depend on  $\theta_c$ . Let us assume that cardinality is Poisson distributed and the parameter  $\theta_c$ , is the mean  $\rho$ . Equation (9) turns into:

$$\begin{aligned} \hat{\rho} &= \operatorname{argmax}_{\rho} \prod_{\text{positive } Z^*} \rho^{|Z^*|} \exp(-\rho), \\ &= \operatorname{argmax}_{\rho} \rho^{\left(\sum_{\text{positive } Z^*} |Z^*|\right)} \exp(-N\rho), \end{aligned} \quad (11)$$

where  $N$  is the number of feature set measurements in the training dataset. After differentiation with respect to  $\rho$  and equating it to zero, we derive:

$$\hat{\rho} = \frac{\sum_{\text{positive } Z^*} |Z^*|}{N}. \quad (12)$$

Assuming that each feature  $z^* \in Z^*$  is IID with Gaussian distribution with parameters  $\theta_f = (\mu, \Sigma)$ , equation (10) turns into:

$$(\hat{\mu}, \hat{\Sigma}) = \operatorname{argmax}_{(\mu, \Sigma)} \prod_{\text{positive } Z^*} \left[ \prod_{z^* \in Z^*} \mathcal{N}(z^*; \mu, \Sigma) \right], \quad (13)$$

where  $\mathcal{N}(\cdot; \mu, \Sigma)$  denotes a multi-variate Gaussian density with mean  $\mu$  and covariance  $\Sigma$ . After algebraic manipulation and differentiation, the following MLE results are obtained:

$$\hat{\mu} = \sum_{\text{positive } Z^*} \sum_{z^* \in Z^*} z^* / \sum_{\text{positive } Z^*} |Z^*| \quad (14)$$

$$\hat{\Sigma} = \sum_{\text{positive } Z^*} \sum_{z^* \in Z^*} (z^* - \hat{\mu})(z^* - \hat{\mu})^\top / \sum_{\text{positive } Z^*} |Z^*|. \quad (15)$$

In addition, the RFS likelihood function in (3) as claimed in [25] does not provide a consistent ranking measure because of the unit inconsistency of feature density  $p_f$ . Therefore, a new RFS ranking function is introduced as follows [25]:

$$R(Z) \propto p_c(|Z|) \left[ \frac{p_f}{\|p_f\|_2^2} \right]^Z \quad (16)$$

where  $R(\cdot)$  is a ranking function,  $\|p_f\|_2^2 = \int p_f^2(z) dz$ , is the squared  $L^2$ -norm of  $p_f$ . The common approach to make a decision on occurrence of an anomaly, as used by almost all statistical methods such as naïve Bayes [35], is to compute and threshold the joint probability density function. Alternatively, Vo et al. [25] suggest to compute and threshold the ranking function (16). This paper presents the results of both approaches in the RFS framework with Poisson assumption for the feature set distribution.

### III. PROPOSED ALGORITHM

In this section, we present the results of our investigations on how the RFS framework can be applied in machine vision for anomaly detection in visual data. We formulate our proposed algorithm for the application of low visibility safety

monitoring in construction sites, but with little tweaking of parameters, the proposed algorithm can be widely used for anomaly detection in images in other safety monitoring and autonomy applications.

Let us consider a construction site where workers are expected to be wearing their high visibility vests. Those entities are modeled as positive data; while any other workers who are not wearing their high visibility vest are considered as anomalous. The proposed solution is comprised of three steps: (i) using a deep neural network for people detection, (ii) extracting features from each blob that is returned by the neural network, (iii) applying anomaly detection on the extracted features that form a set of measurements.

**Note:** Vo *et al.* [25] presents some results for anomaly detection in textures images. They employ SIFT or Harris features which cannot be used on blobs returned from object detection modules, which are in considerably low resolution. In addition, SIFT or Harris features are based on edges and corners and do not take into account the color information which are key when high visibility vests are to be detected.

Algorithm 1 shows an overview of our proposed method. The above mentioned three steps explained in the following sections.

#### A. PEOPLE DETECTION

In the context of detection of workers in construction sites, who do not wear their safety vests, the OBJECT\_DETECTOR(·) function in Algorithm 1 is implemented via a people detection module. In this work, we found the Faster R-CNN [36] to be an effective tool for this purpose. It is a deep learning object detector based on the solution proposed by Girshick [37] but with improved training and testing speed while increasing the detection accuracy. It consists of two components. The first component is a fully convolutional Region Proposal Network(RPN), followed by a Faster R-CNN detector. The Faster R-CNN detector is a purely CNN-based method without using hand-crafted features.

#### B. FEATURE EXTRACTION

In order to extract features, the object detections (blobs) returned by the OBJECT\_DETECTOR(·) function are first resized to 200 pixels  $\times$  180 pixels. This ensures consistency in feature detection, as well as preserves the aspect ratio of objects. RGB color space is well known to be not robust to illumination changes. We first transform the image into the YCbCr color space. This color space is selected due to observations that: (i) in our particular application, the shiny colors of safety vests stand out in the Cb component, more specifically the negative Cb ( $\bar{C}_b$ ), for most color variations, (ii) the Y component holds the gray-scale information (intensity) and can be used to detect shape-related information (see Figure 2).

Using the information embedded in the YCbCr image, we first detect the shiny color of the vest (by focusing on the Cb component) then check shape information (by focusing on the Y component). Checking shape information is crucial because in many practical circumstances, there might be

---

**Algorithm 1** Anomaly Detection for Objects in Image Data

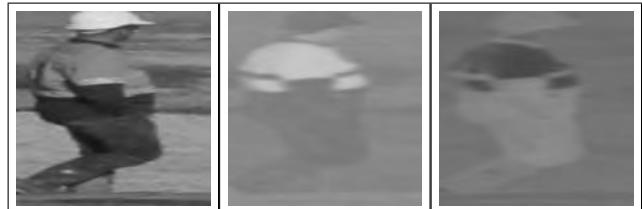
---

```

1: procedure OBJECT_ANOMALY_DETECTION( $I; \rho, \mu, \Sigma, L_{th}$ )
   ▷  $I$  : RGB image of the construction site.
   ▷  $\rho, \mu, \Sigma$  : RFS density parameters returned by
      training. ▷  $L_{th}$  : Likelihood threshold.
2:    $D \leftarrow \emptyset$ 
3:    $B \leftarrow \text{OBJECT\_DETECTOR}(I)$  ▷  $B$  is a set of detected
   blobs in the image. ▷ Extract a set of features from the contents of each
   blob in  $B$ .
4:   for  $b \in B$  do
5:      $[Y, Cb, Cr] \leftarrow \text{rgb2ycbcr}(b)$  ▷ Apply Hessian method to extract key points.
6:      $x \leftarrow \text{HESSIAN}(\bar{C}_b)$ 
7:      $Z \leftarrow \emptyset$ 
8:     for  $x \in x$  do ▷ For each key point, computer LIOP descriptor on  $Cb$ 
   and  $Y$ .
9:        $z_1 \leftarrow \text{LIOP}(x, \bar{C}_b)$ 
10:       $z_2 \leftarrow \text{LIOP}(x, Y)$ 
11:       $z \leftarrow [z_1 z_2]^T$ 
12:       $Z \leftarrow Z \cup \{z\}$ 
13:    end for
14:     $L \leftarrow [\rho U]^{|\mathcal{Z}|} \exp(-\rho) \prod_{z \in Z} \mathcal{N}(z; \mu, \Sigma)$ 
15:    if  $L < L_{th}$  then
16:       $\eta \leftarrow 1$ 
17:    else
18:       $\eta \leftarrow 0$ 
19:    end if
20:     $D = D \cup \{\eta\}$  ▷ A set of decisions, 1 for anomaly,
   0 otherwise.
21:  end for ▷ Report the anomaly detection output as a set of binary
   detections.
22:  return  $D$ 
23: end procedure

```

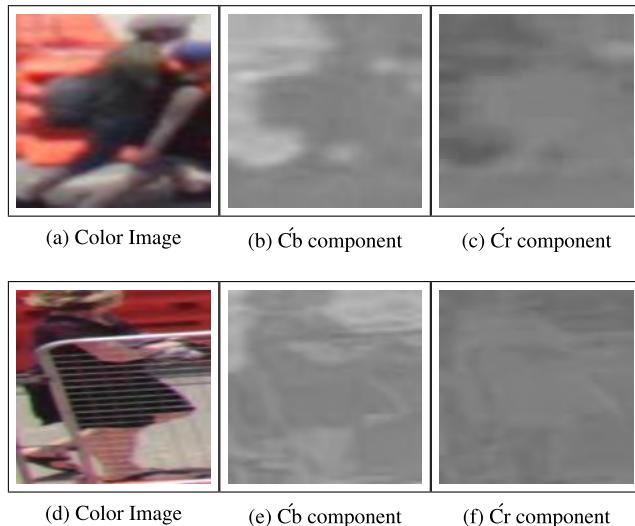
---



**FIGURE 2.** The shiny colors of safety vests stand out in the negative Cb component of YCbCr color space, and the Y component is same as the gray-scale version of the image which holds shape information.

parts of the image in shiny color but not representing a vest (See Figure 3).

Details of our proposed algorithm for feature extraction is presented in Algorithm 1, lines 4–13. The contents of the input blob  $b$  are first converted to YCbCr colors values.



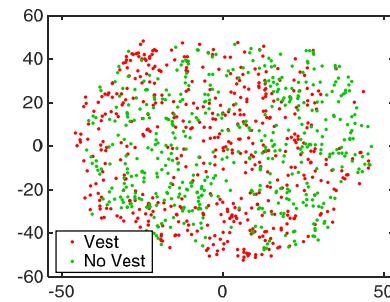
**FIGURE 3.** Objects with color similar to safety vest color and their negative  $C_b$  and  $C_r$  components,  $\bar{C}_b$  and  $\bar{C}_r$  respectively.

The Hessian feature detection method [38] is then utilized to extract key points. For each key point, a 144-D descriptor vector called local intensity order pattern (LIOP) [39] is then computed, once using the contents of the  $C_b$  component, and once using the Y component. The two vectors are then concatenated into a 288-D descriptor vector and saved as a member of the final feature set which gradually grows as more key points are detected and processed.

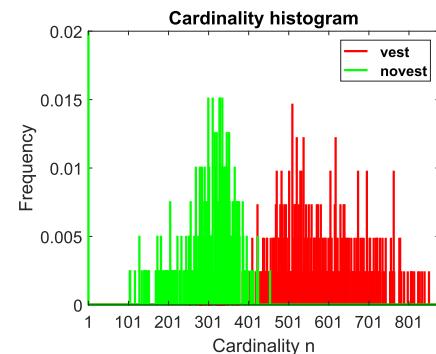
The main reason why we compute the LIOP descriptors for each key point is its invariance to rotation and monotonic intensity changes. As it was previously mentioned, we compute the descriptor for each point, once to capture color information (when applying the contents of  $C_b$  component) and once to capture shape information (when applying the contents of the Y component). Capturing both color and affine shape information enhances the robustness of our method to changes in resolution, illumination and scale.

It is important to note that the number of key points detected by Hessian method can vary from blob to blob. In addition, the order of key points does not matter. Hence, the 288-D descriptors computed for key points are accumulated in a set (not a matrix), which is then treated as a Random Finite set (mainly due to random variations in its cardinality).

Figures 4 and 5 show the features scattering in a dimension-reduced 2-D space, and their cardinality distribution for training blobs that contain positive and negative (anomalous) events in terms of the worker wearing a high-visibility vest. For the purpose of visualization, each 288-D feature vector has been reduced into a 2-D vector using the t-Distributed Stochastic Neighbor Embedding (t-SNE) technique which is well-known to be particularly well suited for the visualization of high-dimensional datasets [40]. The presented graphics in Figures 4 and 5 demonstrate that although individual feature densities for positive and negative cases are distributed with significant overlap, the likely number of those feature for positive and negative blobs are distributed with little overlap.



**FIGURE 4.** Visualization of the features extracted from normal (positive) images and anomalous images, with feature dimension reduced using the t-SNE technique.



**FIGURE 5.** Cardinality distribution of the features from normal (positive) images and anomalous images.

Similarity of the feature densities justify the use of Poisson (and an IID) model for the distribution of feature sets as a Random Finite set. Differences in cardinality distribution will lead to pretty small likelihood values associated with negative (anomalous) sets of features. Using RFS-framework method both the cardinality information and the probability density will be used to detect the anomalous data.

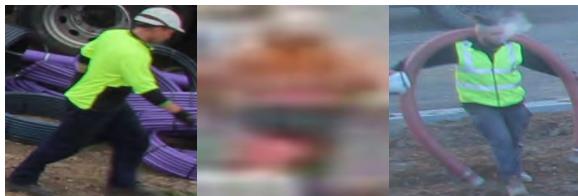
### C. ANOMALY DETECTION

Our proposed RFS-based anomaly detection routine is implemented in lines 14–20 in algorithm 1. First, the Poisson density equation 4 is used to calculate the likelihood of the measurement set  $Z$  that is comprised of all the features extracted from the contents of a given blob  $b$ . Then, if computed likelihood is less than a user-defined threshold, it is detected as an anomaly. Note that the trained parameters  $\rho$ ,  $\mu$  and  $\Sigma$  are used here, assuming that they have been already calculated using a training dataset of positive samples (blobs containing no anomaly), based on equations (12), (14) and (15).

*Remark:* In an alternative approach, we can simply compute and analyze (compare with a threshold) the ranking function given in equation (16). In that case, the only change to Algorithm 1 would be in its lines 15 and 16 in which  $L$  would be substituted by  $R$  given by the expression in equation 16.

### IV. EXPERIMENTAL RESULTS

The proposed algorithm was evaluated based on detections resulted from Faster R-CNN algorithm in the form of bounding boxes. The original images were acquired by the authors



**FIGURE 6.** Samples of images with workers wearing safety vest.

in construction sites. Overall, 400 blobs of workers wearing high visibility vests (positive samples) were collected as well as another 400 containing anomalies (workers not wearing their safety vests). Samples of the detected blobs (with  $200 \times 180$  pixel resolution) are shown in Figure 6.

In our experiments we used 80% of the 400 positive samples for training, and the union of the rest of the positive samples and the negative samples for validation. For anomaly detection evaluations, the most common metrics are the  $F_1$  score and the Area Under the Curve (AUC). We have employed the  $F_1$  score for performance evaluation since it has been used most widely in the literature [25], [41], [42], including the major reference where RFS anomaly detection is presented [25]. Also, we used four fold cross validation and computed the well-known  $F_1$  score which is defined as follows:

$$F_1 \triangleq 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (17)$$

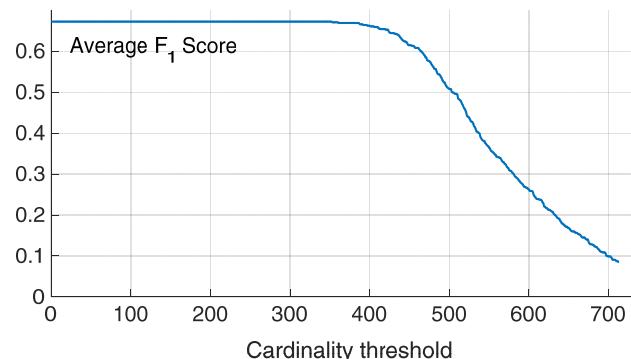
where precision and recall can be defined as follows:

$$\text{precision} = \frac{TP}{TP + FP}, \quad \text{recall} = \frac{TP}{TP + FN} \quad (18)$$

in which  $TP$ ,  $FP$  and  $FN$  stand for the true positive, false positive and false negative. In our particular application scenario,  $TP$  represents the number of blobs in the validation dataset that have been correctly identified as positive, while  $FP$  represents the number of blobs in the validation dataset that have been incorrectly identified as anomalous. Similarly,  $FN$  represents the number of blobs incorrectly identified as negative (anomalous).

For the sake of comparative analysis, we implemented and tested the following methods:

- **RFS likelihood density:** This is the method proposed in Algorithm 1.
- **RFS Ranking function:** This is the variation of Algorithm 1 in which the ranking function (16) is computed instead of density function, for decision making.
- **NB likelihood [26]:** Naïve Bayes classifier.
- **CNN [43]:** One-class convolutional neural network (CNN) classifier.
- **Fisher-OCSVM [44]:** The one-class support vector machine (SVM) [45] in which a Fisher vector representation is used to convert the features sets into vectors.
- **BOVW-SVM [46]:** The one-class SVM [45] in which the feature sets are turned into bags of visual words.



**FIGURE 7.** The  $F_1$  scores (averaged over four-fold validations) of the Cardinality Threshold method, returned for cardinality values chosen as the threshold.



**FIGURE 8.** Randomly chosen 50 samples detected as anomalous by the RFS likelihood density method.

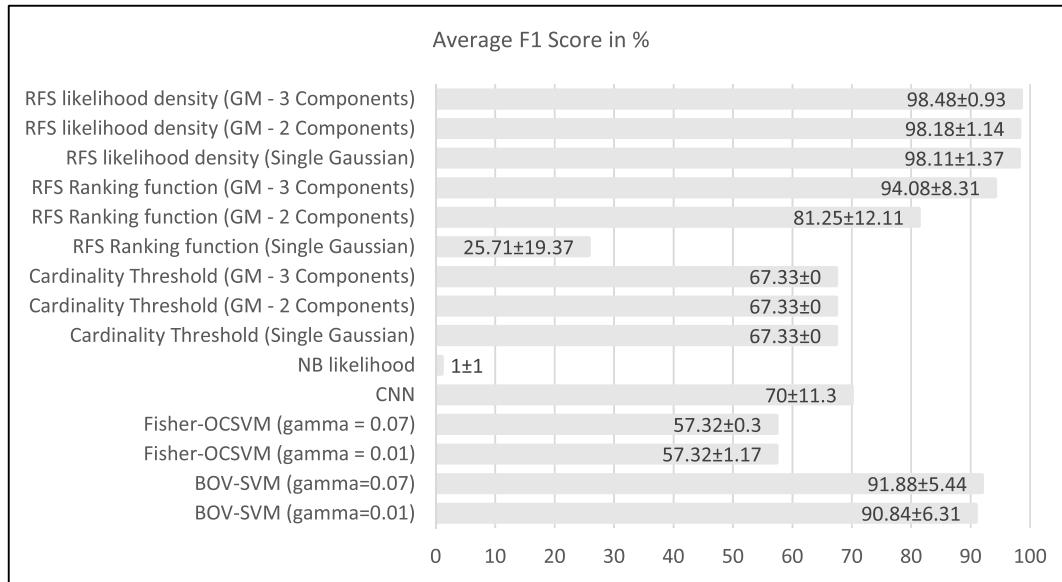
- **Cardinality Threshold:** A heuristic detection method based on merely comparing the cardinality of features with a user-defined threshold.

We have used MATLAB 2018b for implementing all the anomaly detection methods. For Fisher-OCSVM method, the number of Gaussian clusters is set to 30. In the BOVW-SVM method, the size of codebook is set to 100000, the number of iteration of k-means clustering is set to 50 and the size of visual word is set to 500. We investigated the performance the two SVM-based methods both with the gamma parameter set to 0.07 and set to 0.01, respectively.

The **Cardinality Threshold** is an intuitive alternative method inspired by the following observation: From Figures 4 and 5, it appears that the major factor that discriminates between positive and negative sets of features is the cardinality. We have tested this hypothesis and demonstrated that it is not accurate, by indicating that our method (which uses both cardinality and feature density information all together) performs significantly better.

To determine the best-performing threshold for cardinality, we tested numerous candidates ranging from 0 to the maximum value of 718 and applied it for anomaly detection. Figure 7 shows the  $F_1$  score averaged over all validation samples, plotted versus different choices of cardinality thresholds. It demonstrates that for any threshold less than 342, the best score of 0.67 would be achieved.

Figure 8 shows 50 samples randomly chosen from those detected as anomalous by our proposed algorithm



**FIGURE 9.** The average and standard deviations of  $F_1$  scores in % for various methods benchmarked in this study.

(RFS likelihood density), demonstrating its effectiveness in detecting the absence of safety vest. All of the randomly chosen samples are indeed representative of anomalies, and our proposed method has returned the right decision.

We first examined the performance of the **RFS likelihood density**, the **RFS ranking**, the **Cardinality Threshold**, and the **naïve Bayesian** in which the single-object densities are modeled each as a single Gaussian. We also investigated whether any significant improvement in anomaly detection accuracy of those methods could be achieved by considering more than one Gaussian components in our presumed model for feature densities. Note that with a mixture of Gaussian components, equations (14) and (15) are no longer valid, and training for parameter estimation should be conducted using a recursive algorithm such as the Expectation Maximization (EM) algorithm.

The overall results for different numbers Gaussians, and different gamma parameter values for the SVM-based methods are visualized as bar plots in Figure 9. Inside each bar, we specify the average and standard deviation of the  $F_1$  score in %. We observe that (i) the **RFS likelihood density** method outperforms the benchmarked solutions in all cases with returning above 98% accuracy which is not significantly improved by increasing the number of Gaussian components, (ii) the **NB likelihood** completely fails due to the huge overlap between the feature densities as significant difference in cardinalities that are not picked up by the method, (iii) the **Cardinality Threshold** method performs the same in all cases despite large disparities in cardinality of positive and negative samples, which is still weaker than the accuracy returned by the **RFS likelihood density** method because it utilizes both the cardinality and feature information in an efficient manner, and (iv) the performance of SVM-based methods does not change with varying the gamma parameter

**TABLE 1.** The  $F_1$  score, precision, and recall in % for different anomaly detection algorithms with 3 Gaussian components for all statistical-based methods.

Methods	$F_1$ score	Precision	Recall
NB likelihood	1.0	0.8	1.5
RFS likelihood density	98.7	99.0	98.5
RFS Ranking Function	94.5	93.5	95.0
BOVW-SVM	90.4	83.0	99.7
Fisher-OCSVM	58.2	44.8	85.0
CNN	70.0	80.0	63.0

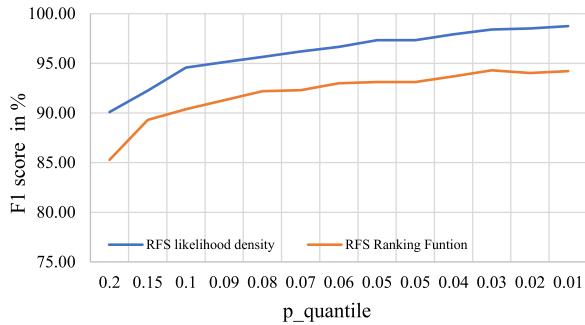
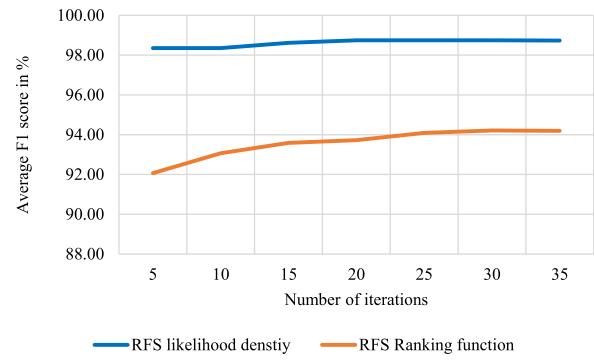
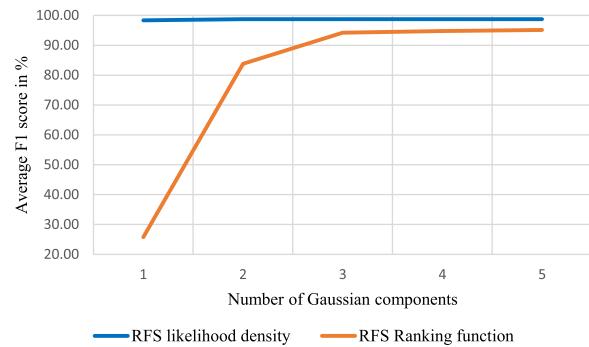
from 0.01 to 0.07, and both methods return a lower accuracy than the proposed RFS-based method.

For more quantitative analysis, we recorded all of the  $F_1$  score, precision and recall for all methods with the same settings reported before as shown in Table 1. It is clear from Table 1 that the proposed solution not only achieves the highest results in  $F_1$  score but also in the precision and recall metrics as well.

Besides proposing the application of RFS-based anomaly detection in visual analytics, an important contribution of this paper is the introduction of the particularly crafted features (with both color and shape-related information properly embedded in them) for the purpose of anomaly detection in visual data. To ensure we are running a fair comparison, we investigated whether the benchmarked methods would outperform our method (i.e. would return an accuracy of more than %98) using other features that are widely used for various detection and identification purposes in the computer vision literature. Those features include the well-known SIFT [47] and DSP-SIFT [48], the color-based features in different domains such as CSIFT [49], HSV-SIFT [50], OPP-SIFT [51] and Hue-SIFT [52], and the recent deep learning local feature LF-Net [53]. The comparative results are presented in Table (2), demonstrating that our proposed

**TABLE 2.** Average  $F_1$  score for different feature detection methods using different anomaly detection methods.

Method\Features	SIFT	CSIFT [50]	DSP-SIFT [49]	HSV-SIFT [51]	Hue-SIFT [53]	OPP-SIFT [52]	LF-Net [54]	Proposed
Fisher-SVM	0.584	0.649	0.000	0.675	0.588	0.637	0.610	0.568
BOVW-SVM	0.644	0.746	0.000	0.761	0.686	0.723	0.580	0.919
naïve	0.076	0.004	0.816	0.004	0.004	0.004	0.423	0.000
RFS-likelihood	0.076	0.004	0.816	0.004	0.004	0.004	0.917	<b>0.985</b>
RFS-Ranking	0.710	0.860	0.816	0.850	0.829	0.880	0.057	<b>0.941</b>
Cardinality threshold	0.672	0.673	0.672	0.673	0.673	0.673	0.648	0.673

**FIGURE 10.** The average  $F_1$  scores in % for different p\_quantile for 3 Gaussian components with 30 iterations.**FIGURE 12.** The average  $F_1$  scores in % for different number of iterations with 3 Gaussian components.**FIGURE 11.** The average  $F_1$  scores in % for different Gaussian components with number of iteration set to 30.

RFS-based anomaly detection methods return the highest accuracy (in terms of the average  $F_1$  score) with the proposed feature extracted and used as measurement.

#### A. ANALYSIS OF THE PROPOSED SOLUTION

In this section, we study the effect of different parameters on the performance of the proposed solution. The proposed solution is based on the assumption that normal and anomalous data are separated by a user-defined threshold value ( $p_{quantile}$  of the RFS density). Therefore, we studied the effect of different threshold values ( $p_{quantile}$ ) on the performance of the proposed solution as shown in Figure 10. It is clear that the proposed solution performance varies according to threshold value and in our application (non-vest detection), the best performance (the highest  $F_1$  score) is achieved with the  $p_{quantile}$  set to 0.01 of its density.

In addition, we studied the effect of choosing different number of Gaussian components for modeling the single-object densities, on the performance of the proposed solution (reported as the average  $F_1$  score) as shown in Figure 11.

It is clear that **RFS likelihood density** is more robust than **RFS Ranking function**, as its performance is less affected by the number of Gaussian components. Note that the more Gaussian components required during the training means more computational cost.

Finally, since the parameters of the Gaussian mixture components are determined through a recursive operation by EM algorithm, we studied the effect of the number of algorithm iterations on the performance of the proposed solutions and reported it in Figure 12. We can see that the performance of both algorithms increases with the number of iterations but the effect is less significant for the **RFS likelihood density** method.

#### V. CONCLUSIONS

In this paper, monitoring of low visibility hazards in construction sites is investigated with a focus on detection of workers who are not wearing any safety vest. The automated intelligent solution takes live images from construction sites and detects the existence of workers who do not abide with safety rule as an anomalous event. The proposed solution is formulated as an anomaly detection in RFS framework. The resulting algorithm consists of three steps: (i) people detection using deep learning (Faster R-CNN), (ii) extracting a set of particularly engineered features from the blobs returned by the Faster R-CNN, and finally (iii) anomaly detection using RFS framework on each set of the extracted features. The engineered features take into account both the shiny color of safety vest (using YCbCr color space) and shape information during the extraction process. The experimental results show that the proposed **RFS likelihood density** outperforms the other methods with 98%  $F_1$  score. In addition, we showed

that anomaly detection can be performed using cardinality information (**Cardinality Threshold**) but with lower performance compared with **RFS likelihood density**. The proposed solution depends on the accuracy of the blobs returned by the deep learning tool, and on weather conditions such as fog. Investigating of the effects of those parameters on the detection performance, and robustifying the solution against them are excellent candidates for our future research.

## REFERENCES

- [1] M.-W. Park and I. Brilakis, “Continuous localization of construction workers via integration of detection and tracking,” *Autom. Construct.*, vol. 72, pp. 129–142, Dec. 2016.
- [2] R. Stewart, M. Andriluka, and A. Y. Ng, “End-to-end people detection in crowded scenes,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2325–2333.
- [3] Q. Fang, H. Li, X. Luo, L. Ding, H. Luo, T. M. Rose, and W. An, “Detecting non-hardhat-use by a deep learning method from far-field surveillance videos,” *Automat. Construct.*, vol. 85, pp. 1–9, Jan. 2018.
- [4] Q. Fang, H. Li, X. Luo, L. Ding, T. M. Rose, W. An, and Y. Yu, “A deep learning-based method for detecting non-certified work on construction sites,” *Adv. Eng. Inform.*, vol. 35, pp. 56–68, Jan. 2018.
- [5] J. Seo, S. Han, S. Lee, and H. Kim, “Computer vision techniques for construction safety and health monitoring,” *Adv. Eng. Inform.*, vol. 29, no. 2, pp. 239–251, Apr. 2015.
- [6] H. Seong, H. Son, and C. Kim, “A comparative study of machine learning classification for color-based safety vest detection on construction-site images,” *KSCE J. Civil Eng.*, vol. 22, no. 11, pp. 4254–4262, Nov. 2018.
- [7] S. Garg, K. Kaur, N. Kumar, and J. J. P. C. Rodrigues, “Hybrid deep-learning-based anomaly detection scheme for suspicious flow detection in SDN: A social multimedia perspective,” *IEEE Trans. Multimedia*, vol. 21, no. 3, pp. 566–578, Mar. 2019.
- [8] F. Jiang, G. Liu, J. Du, and Y. Sui, “Initialization of K-modes clustering using outlier detection techniques,” *Inf. Sci.*, vol. 332, pp. 167–183, Mar. 2016.
- [9] G. Qin, Z. Huang, Y. Xiang, and J. Sun, “ProbDetect: A choice probability-based taxi trip anomaly detection model considering traffic variability,” *Transp. Res. C, Emerg. Technol.*, vol. 98, pp. 221–238, Jan. 2019.
- [10] W.-C. Lin, S.-W. Ke, and C.-F. Tsai, “CANN: An intrusion detection system based on combining cluster centers and nearest neighbors,” *Knowl.-Based Syst.*, vol. 78, pp. 13–21, Apr. 2015.
- [11] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey,” *ACM Comput. Surv.*, vol. 41, no. 3, pp. 15:1–15:58, Jul. 2009.
- [12] Y. Wang, L. Han, W. Liu, S. Yang, and Y. Gao, “Study on wavelet neural network based anomaly detection in ocean observing data series,” *Ocean Eng.*, vol. 186, Aug. 2019, Art. no. 106129.
- [13] A. Taha and A. S. Hadi, “Anomaly detection methods for categorical data: A review,” *ACM Comput. Surv.*, vol. 52, no. 2, pp. 38:1–38:35, May 2019. [Online]. Available: <http://doi.acm.org/10.1145/3312739>
- [14] R. Ramezani, P. Angelov, and X. Zhou, “A fast approach to novelty detection in video streams using recursive density estimation,” in *Proc. 4th Int. IEEE Conf. Intell. Syst.*, vol. 2, Sep. 2008, pp. 14-2–14-7.
- [15] C. P. Diehl and J. B. Hampshire, “Real-time object classification and novelty detection for collaborative video surveillance,” in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, vol. 3, May 2002, pp. 2620–2625.
- [16] S.-P. Yong, J. D. Deng, and M. K. Purvis, “Novelty detection in wildlife scenes through semantic context modelling,” *Pattern Recognit.*, vol. 45, no. 9, pp. 3439–3450, 2012.
- [17] S.-P. Yong, J. D. Deng, and M. K. Purvis, “Wildlife video key-frame extraction based on novelty detection in semantic context,” *Multimedia Tools Appl.*, vol. 62, no. 2, pp. 359–376, Jan. 2013.
- [18] S. Singh and M. Markou, “An approach to novelty detection applied to the classification of image regions,” *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 4, pp. 396–407, Apr. 2004.
- [19] M. Markou and S. Singh, “A neural network-based novelty detector for image sequence analysis,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1664–1677, Oct. 2006.
- [20] W. Sultani, C. Chen, and M. Shah, “Real-world anomaly detection in surveillance videos,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 6479–6488.
- [21] C. Yin *et al.*, “Improved collaborative filtering recommendation algorithm based on differential privacy protection,” *J. Supercomput.*, 2019. doi: [10.1007/s11227-019-02751-7](https://doi.org/10.1007/s11227-019-02751-7).
- [22] N. Moustafa, K.-K. R. Choo, I. Radwan, and S. Camtepe, “Outlier Dirichlet mixture mechanism: Adversarial statistical learning for anomaly detection in the fog,” *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 8, pp. 1975–1987, Aug. 2019.
- [23] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, “A review of novelty detection,” *Signal Process.*, vol. 99, pp. 215–249, Jun. 2014.
- [24] L. Li, J. Hansman, R. Palacios, and R. Welsch, “Anomaly detection via a Gaussian mixture model for flight operation and safety monitoring,” *Transp. Res. C, Emerg. Technol.*, vol. 64, pp. 45–57, Mar. 2016.
- [25] B.-N. Vo, N. Dam, D. Phung, Q. N. Tran, and B.-T. Vo, “Model-based learning for point pattern data,” *Pattern Recognit.*, vol. 84, pp. 136–151, Dec. 2018.
- [26] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2010.
- [27] R. P. S. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Norwood, MA, USA: Artech House, 2007.
- [28] M. Markou and S. Singh, “Novelty detection: A review—Part 1: Statistical approaches,” *Signal Process.*, vol. 83, no. 12, pp. 2481–2497, Dec. 2003.
- [29] R. P. S. Mahler, “Multitarget Bayes filtering via first-order multitarget moments,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1152–1178, Oct. 2003.
- [30] B.-T. Vo, B.-N. Vo, and A. Cantoni, “The cardinality balanced multi-target multi-Bernoulli filter and its implementations,” *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 409–423, Feb. 2009.
- [31] R. P. S. Mahler, *Advances in Statistical Multisource-Multitarget Information* (Electronic Warfare). Norwood, MA, USA: Artech House, 2014. [Online]. Available: <https://books.google.com.au/books?id=zPbAABAAQBAJ>
- [32] S. Reuter, B. T. Vo, B. N. Vo, and K. Dietmayer, “The labeled multi-Bernoulli filter,” *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3246–3260, Dec. 2014.
- [33] B.-T. Vo and B.-N. Vo, “Labeled random finite sets and multi-object conjugate priors,” *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3460–3475, Jul. 2013.
- [34] Y. Chen, J. Wang, R. Xia, Q. Zhang, Z. Cao, and K. Yang, “The visual object tracking algorithm research based on adaptive combination kernel,” *J. Ambient Intell. Humanized Comput.*, pp. 1–13, Jan. 2019. doi: [10.1007/s12652-018-01171-4](https://doi.org/10.1007/s12652-018-01171-4).
- [35] N. F. Lepora, M. J. Pearson, B. Mitchinson, M. Evans, C. Fox, A. Pipe, K. Gurney, and T. J. Prescott, “Naive Bayes novelty detection for a moving robot with whiskers,” in *Proc. IEEE Int. Conf. Robot. Biomimetics*, Dec. 2010, pp. 131–136.
- [36] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” in *Proc. 28th Int. Conf. Neural Inf. Process. Syst. (NIPS)*. Cambridge, MA, USA: MIT Press, vol. 1, 2015, pp. 91–99.
- [37] R. Girshick, “Fast R-CNN,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Washington, DC, USA, Dec. 2015, pp. 1440–1448.
- [38] K. Mikolajczyk and C. Schmid, “An affine invariant interest point detector,” in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2002, pp. 128–142.
- [39] Z. Wang, B. Fan, and F. Wu, “Local intensity order pattern for feature description,” in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 603–610.
- [40] L. J. P. van der Maaten and G. E. Hinton, “Visualizing high-dimensional data using t-SNE,” *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
- [41] H. Schütze, C. D. Manning, and P. Raghavan, *Introduction to Information Retrieval*, vol. 39. Cambridge, U.K.: Cambridge Univ. Press, 2008.
- [42] R. Domingos, P. Michiardi, J. Zouaoui, and M. Filippone, “Deep Gaussian Process autoencoders for novelty detection,” *Mach. Learn.*, vol. 107, nos. 8–10, pp. 1363–1383, 2018.
- [43] R. Chalapathy, A. K. Menon, and S. Chawla, “Anomaly detection using one-class neural networks,” 2018, *arXiv:1802.06360*. [Online]. Available: <https://arxiv.org/abs/1802.06360>
- [44] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Fisher vector faces in the wild,” in *Proc. BMVC*, Sep. 2013, vol. 2, no. 3, p. 4.
- [45] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011.
- [46] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, “Visual categorization with bags of keypoints,” in *Proc. Workshop Stat. Learn. Comput. Vis. ECCV*, Prague, Czech Republic, vol. 1, nos. 1–22, 2004, pp. 1–16.

- [47] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [48] J. Dong and S. Soatto, "Domain-size pooling in local descriptors: DSP-SIFT," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5097–5106.
- [49] A. E. Abdel-Hakim and A. A. Farag, "CSIFT: A SIFT descriptor with color invariant characteristics," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2006, pp. 1978–1983.
- [50] A. Bosch, A. Zisserman, and X. Muñoz, "Scene classification using a hybrid generative/discriminative approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 4, pp. 712–727, Apr. 2008.
- [51] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1582–1596, Sep. 2010.
- [52] J. van de Weijer, T. Gevers, and A. D. Bagdanov, "Boosting color saliency in image feature detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 150–156, Jan. 2006.
- [53] Y. Ono, E. Trulls, P. Fua, and K. M. Yi, "LF-Net: learning local features from images," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 6234–6244.



**AMMAR MANSOOR KAMOONA** (S'13) received the master's degree in electronic and electrical engineering from the Swinburne University of Technology, Melbourne, Australia, in 2016. He is currently pursuing the Ph.D. degree with the RMIT University of Technology, Melbourne. He was an Assistant Lecturer with the Department of Electrical Engineering, University of Kufa, Iraq, from 2017 to 2018. His current research interests include computer vision, RFS filters, robotics and optimization, and FPGA applications. He was a recipient of two certificates of excellence in RF circuit design and Stochastic and Survival analysis from Swinburne University, as well as a Golden Key Certificate for being one of the top achiever students.



**AMIRALI KHODADADIAN GOSTAR** received the B.Sc. degree in electrical engineering, the M.Sc. degree in philosophy of science, and the Ph.D. degree in mechatronics engineering from RMIT University, where he is currently a Postdoctoral Research Fellow with the School of Engineering. His research interests include sensor management, data fusion, and multitarget tracking.



**RUWAN TENNAKON** received the B.Sc. degree (Hons.) in electrical and electronics engineering from the University of Peradeniya, Sri Lanka, in 2007, and the Ph.D. degree in computer vision from the Swinburne University of Technology, Australia, in 2015. He was a Research Scientist with IBM-Research Australia. Since 2015, he has been a Research Fellow with the RMIT School of Engineering, developing computer vision-based driver assist technologies for industrial vehicles. His main research interests include computer vision, machine learning, and medical image analysis.



**ALIREZA BAB-HADIASHAR** received the B.Sc. and M.Eng. degrees in mechanical engineering and the Ph.D. degree in robotics from Monash University. He has held various positions in Monash University, the Swinburne University of Technology, and RMIT University, where he is currently a Professor of mechatronics and leads the Intelligent Automation Research Group. His main research interests include intelligent automation in general, robust data fitting in machine vision, deep learning for detection and identification, and robust data segmentation, in particular.



**DAVID ACCADIA** received the bachelor's degree in computer science and the bachelor's degree in electronic engineering (Hons.) from the La Trobe University, Melbourne, Australia, in 2003. He is currently the Chief Technical Officer with Cornerstone Solutions and photoSentinel. He has held a lead position in the Research and Development team for over ten years.



**JOSHUA THORPE** received the B.Eng. degree (Hons.) in advanced manufacturing and mechatronic engineering with RMIT University, Melbourne. He is currently a Software Engineer with Cornerstone Solutions, researching and developing products in the field of deep learning and image analysis.



**REZA HOSEINNEZHAD** received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the University of Tehran, Iran, in 1994, 1996, and 2002, respectively. He has held various positions at the University of Tehran, the Swinburne University of Technology, The University of Melbourne, and RMIT University, where he has worked, since 2010, and is currently a Professor, and a Research Development Lead, as well as the Discipline Leader (Manufacturing and Mechatronics) with the School of Engineering. His main research interests include statistical information fusion, random finite sets, multi-object tracking, deep learning, and robust multi-structure data fitting in computer vision.