MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

GODS: Generalized One-class Discriminative Subspaces for Anomaly Detection

Wang, J.; Cherian, A.

TR2019-121 October 31, 2019

Abstract

One-class learning is the classic problem of fitting a model to data for which annotations are available only for a single class. In this paper, we propose a novel objective for one-class learning. Our key idea is to use a pair of orthonormal frames – as subspaces – to "sandwich" the labeled data via optimizing for two objectives jointly: i) minimize the distance between the origins of the two subspaces, and ii) to maximize the margin between the hyperplanes and the data, either subspace demanding the data to be in its positive and negative orthant respectively. Our proposed objective however leads to a non-convex optimization problem, to which we resort to Riemannian optimization schemes and derive an efficient conjugate gradient scheme on the Stiefel manifold. To study the effectiveness of our scheme, we propose a new dataset Dash-Cam-Pose, consisting of clips with skeleton poses of humans seated in a car, the task being to classify the clips as normal or abnormal; the latter is when any human pose is out-of-position with regard to say an airbag deployment. Our experiments on the proposed Dash-Cam-Pose dataset, as well as several other standard anomaly/novelty detection benchmarks demonstrate the benefits of our scheme, achieving state-of-the-art oneclass accuracy.

IEEE International Conference on Computer Vision (ICCV)

GODS: Generalized One-class Discriminative Subspaces for Anomaly Detection

Jue Wang*
Australian National University, Canberra
jue.wang@anu.edu.au

Anoop Cherian
Mitsubishi Electric Research Labs, Cambridge, MA
cherian@merl.com

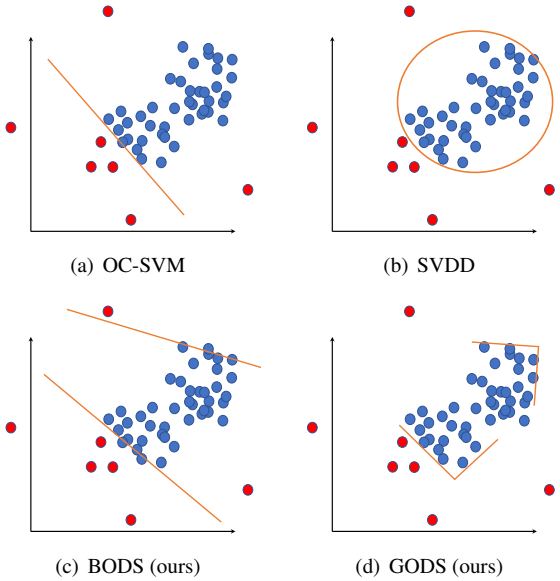
Abstract

One-class learning is the classic problem of fitting a model to data for which annotations are available only for a single class. In this paper, we propose a novel objective for one-class learning. Our key idea is to use a pair of orthonormal frames – as subspaces – to "sandwich" the labeled data via optimizing for two objectives jointly: i) minimize the distance between the origins of the two subspaces, and ii) to maximize the margin between the hyperplanes and the data, either subspace demanding the data to be in its positive and negative orthant respectively. Our proposed objective however leads to a non-convex optimization problem, to which we resort to Riemannian optimization schemes and derive an efficient conjugate gradient scheme on the Stiefel manifold.

To study the effectiveness of our scheme, we propose a new dataset Dash-Cam-Pose*, consisting of clips with skeleton poses of humans seated in a car, the task being to classify the clips as normal or abnormal; the latter is when any human pose is out-of-position with regard to say an airbag deployment. Our experiments on the proposed Dash-Cam-Pose dataset, as well as several other standard anomaly/novelty detection benchmarks demonstrate the benefits of our scheme, achieving state-of-the-art one-class accuracy.*

1. Introduction

There are several real-world problems in which it may be easy to characterize the normal operating behavior of a system or collect data for it, however may be difficult or sometimes even impossible to have data when a system is at fault or is improperly used. Examples include but not limited to an air conditioner making an unwanted vibration, a network attacked by an intruder, abnormal patient conditions such as heart rates, an accident captured in a video surveillance camera, or a car engine firing at irregular intervals, among others [11]. In machine learning literature, such problems



(a) OC-SVM (b) SVDD (c) BODS (ours) (d) GODS (ours)

Figure 1. A graphical illustration of classical OC-SVM and SVDD in relation to our proposed BODS and GODS schemes. The blue points show the given one-class data, the red-points are outliers, and decision boundary of each method is shown by orange curves/lines.

are usually called one-class problems [4, 44], signifying the fact that we may be able to have unlimited supply of labeled training data for one-class (corresponding to the normal operation of the system), but do not have any labels or training data for situations corresponding to abnormalities. The main goal of such problems is thus to learn a model that fits to the normal set, such that abnormalities can be characterized as outliers of this model.

Classical solutions for one-class problems are mainly extensions to support vector machines (SVMs), such as the one-class SVM (OC-SVM) that maximizes the margin of the discriminative hyperplane from the origin [46]. There are extensions of this scheme, such as the least-squares one-class SVM (LS-OSVM) [14] or its online variants [58] that learn to find a tube of minimal diameter that includes all the labeled data. Another popular approach is the support-vector data description (SVDD) that finds a hy-

*Work done while interning at MERL.

persphere of minimum radius that encapsulates the training data [51]. There have also been kernelized extensions of these schemes that use the kernel trick to embed the data points in a reproducible kernel Hilbert space, potentially enclosing the 'normal' data with arbitrarily-shaped boundaries.

While, these approaches have shown benefits and have been widely adopted in several applications [11], they have drawbacks that motivate us to look beyond prior solutions. For example, the OC-SVM uses only a single hyperplane, however using multiple hyperplanes may be beneficial and provide a richer characterization of the labeled set, as also recently advocated in [56]. The SVDD scheme makes a strong assumption on the spherical nature of the data distribution, which may be seldom true in practice. Further, using kernel methods may impact scalability. Motivated by these observations, we propose a novel one-class classification objective that: (i) learns a set of discriminative and orthogonal hyperplanes, as a subspace, to model a multi-linear classifier, (ii) learns a pair of such subspaces, one bounding the data from below and the other one from above, and (iii) minimizes the distances between these subspaces such that the data is captured within a region of minimal volume (as in SVDD). Our framework generates a piecewise linear decision boundary and operates in the input space.

Albeit these benefits, our objective is non-convex due to the orthogonality constraints. However, such non-convexity fortunately is not a significant concern as the orthogonality constraints naturally place the optimization objective on the Stiefel manifold [17]. This is a well-studied Riemannian manifold [6] for which there exist efficient non-linear optimization methods at our disposal. We use one such optimization scheme, dubbed Riemannian conjugate gradient [1], which is fast and efficient.

To evaluate the usefulness of our proposed scheme, we apply it to the concrete setting of detecting abnormal or 'out-of-position' human poses [54, 53] in cars; specifically, our goal is to detect if the passengers or the driver are seated "out-of-position" (OOP) as captured by an inward looking dashboard camera. This problem is of at most importance in vehicle passenger safety as humans seated OOP may be subject to fatal injuries if the airbags are deployed [43, 36, 25]. The problem is even more serious in autonomous cars, which may not (in the future) have any drivers at all to monitor the safety of the passengers. Such OOP human poses include abnormal positions of the face (such as turning back), legs on the dashboard, etc., to name a few. As it may be easy to define what normal seating poses are, while it may be far too difficult to model abnormal ones, we cast this problem in the one-class setting. As there are no public datasets available to study this problem, we propose a novel dataset, *Dash-Cam-Pose* consisting of nearly 5K short video clips and comprising of nearly a million human poses (extracted using OpenPose [8]). Each clip is collected from long Internet videos or Hollywood road movies and weakly-annotated with a binary label signifying if passengers are seated correctly or out-of-position for the entire duration of the clip.

We showcase the effectiveness of our approach on the Dash-Cam-Pose dataset, as well as several other popular benchmarks such as UCF-Crime [50], action recognition datasets such as JHMDB [23], and two standard UCI anomaly datasets. Our experiments demonstrate that our proposed scheme leads to more than 10% improvement in performance over classical and recent approaches on all the datasets we evaluate on.

Before moving ahead detailing our method, we summarize below the main contributions of this paper:

1. We first introduce a one-class discriminative subspace (BODS) classifier that uses a pair of hyperplanes.
2. We generalize BODS to use multiple hyperplanes, termed generalized one-class discriminative subspaces (GODS).
3. We propose a new Dash-Cam-Pose dataset for anomalous pose detection of passengers in cars, and
4. We provide experiments on the Dash-Cam-Pose dataset, as well as four other public datasets, demonstrating state-of-the-art performance.

2. Background and Related Works

Let $\mathcal{D} \subset \mathbb{R}^d$ denote the data set consisting of our one class-of-interest and everything outside it, denoted $\overline{\mathcal{D}}$, be the anomaly set. Suppose we are given n data instances $\mathcal{D}_o = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\} \subset \mathcal{D}$. The goal of one-class classifiers is to use $\mathcal{D}_o$ to learn a functional f which is positive on $\mathcal{D}$ and negative on $\overline{\mathcal{D}}$. Typically, the label of $\mathcal{D}$ is assumed $+1$ and that of $\overline{\mathcal{D}}$ as -1.

In One-Class Support Vector Machine (OC-SVM) [46], f is modeled as an extension of an SVM objective by learning a max-margin hyperplane that separates the origin from the data points in $\mathcal{D}_o$. Mathematically, f has the form $\text{sgn}(\mathbf{w}^T\mathbf{x} + b)$, where $(\mathbf{w}, b) \in \mathbb{R}^d \times \mathbb{R}^1$ and is learned by minimizing the following objective:

$$\min_{\mathbf{w},b,\xi\geq 0} \frac{1}{2}\|\mathbf{w}\|_2^2 - b + C\sum_{i=1}^{n}\xi_i, \text{ s.t. } \mathbf{w}^T\mathbf{x}_i \geq b - \xi_i, \forall \mathbf{x}_i \in \mathcal{D}_o,$$

where ξ_i's are non-negative slacks, b is the hyperplane intercept, and C is the slack penalty. As a single hyperplane might be insufficient to capture all the non-linearities associated with the one-class, there are extensions using non-linear kernels via the kernel-trick [46]. However, as is common with kernelized SVM, such a formulation is difficult

to scale with the number of data points. Another popular variant of one-class classifiers is the support vector data description (SVDD) [51] that instead of modeling data to belong to an open half-space of $\mathbb{R}^d$ (as in OC-SVM), assumes the labeled data inhabits a bounded set; specifically, the optimization seeks the centroid $\mathbf{c} \in \mathbb{R}^d$ of a hypersphere of minimum radius $R > 0$ that contains all points in $\mathcal{D}_o$. Mathematically, the objective reads:

$$\min_{\mathbf{c},R,\xi\geq 0} \frac{1}{2}R^2 + C\sum_{i=1}^{n}\xi_i, \text{ s.t. } \|\mathbf{x}_i - \mathbf{c}\|_2^2 \leq R^2 - \xi_i, \forall \mathbf{x}_i \in \mathcal{D}_o,$$

where, as in OC-SVM, the ξ's model the slack. There have been extensions of this scheme, such as the mSVDD that uses a mixture of such hyperspheres [29], density-induced SVDD [30], using kernelized variants [52], and more recently, to use subspaces for data description [49]. A major drawback of SVDD in general is the strong assumption it makes on the isotropic nature of the underlying data distribution. Such a demand is ameliorated by combining OC-SVM with the idea of SVDD in least-squares one-class SVM (LS-OSVM) [14] that learns a tube around the discriminative hyperplane that contains the input; however, this scheme also makes strong assumptions on the data distribution (such as being cylindrical). In Figures 1(a) and 1(b), we graphically illustrate OC-SVM and SVDD schemes.

Unlike OC-SVM that learns a compact data model to enclose as many training samples as possible, a different approach is to use principal component analysis (PCA) (and its kernelized counterpart, such as Kernel PCA and Roboust PCA[7, 15, 20, 38, 60]) to summarize the data by using its principal subspaces. However, such an approach is usually unfavorable due to its high computational cost, especially when the dataset is large. Similar in motivation to the proposed technique, Bodesheim et al. [5] use null space transform for novelty detection and while Liu et al. [34] optimize a kernel-based max-margin objective for outlier removal and soft label assignment. However, their problem setups are different from ours in that [5] requires multi-class labels in the training data and [34] is proposed for unsupervised learning.

In contrast to these prior methods, in this paper, we explore the one-class objective from a very unique perspective; specifically, to use subspaces as in PCA, however instead of approximating the one-class data, these subspaces are aligned in such a way as to bound the data in a piecewise linear manner, via solving a discriminative objective. We first present a simplified variant of this objective by using two different (sets of) hyperplanes, dubbed Basic One-class Discriminative Subspaces (BODS), that can sandwich the labeled data by bounding from different sides; these hyperplanes are independently parameterized and thus can be oriented differently to better fit to the labeled data. Note that there is a similar prior work, termed Slab-SVM [18], that learns two hyperplanes for one-class classification. However, these hyperplanes are constrained to have the same slope, which we do not impose in our BODS model, as a result, our model is more general than Slab-SVM. We extend the BODS formulation by using multiple hyperplanes, as a discriminative subspace, which we call Generalized One-class Discriminative Subspaces (GODS); these subspaces provide better support for the one-class data, while also circumscribing the data distribution. The use of such discriminative subspaces has been recently explored in the context of representation learning on videos in Wang and Cherian [56] and Wang et al. [57], however demands a surrogate negative bag of features found via adversarial means.

Anomaly Detection: In computer vision, anomaly detection has been explored from several facets and we refer interested readers to excellent surveys provided in [11, 42] on this topic. Here we pickout a few prior works that are related to the experiments we present. To this end, Adam et al., [2] and Kim et al. [26] use optical flow to capture motion dynamics, characterizing anomalies. A Gaussian mixture modelling of people and object trajectories is used in [32, 48] for identifying anomalies in video sequences. Saliency is used in [22, 24] and detecting out-of-context objects is explored in [13, 40] using support graph and generative models for characterizing normal and abnormal data. We are also aware of recent deep learning methods for one-class problems. Feature embeddings (via a CNN) is explored in [31, 33] minimizing the "in-distribution" sample distances, so that "out-of-distribution" samples can be found via suitable distance measures. Differently, we attempt at finding a suitable "in-distribution" data boundary which is agnostic to the data embedding. A deep variant of SVDD is proposed in [45], however assumes the one-class data is unimodal. There are extensions of OC-SVM to a deep setting in [59, 10, 41]. Due to the immense capacity of modern CNN models, it is often found that the learned parameters overfit quickly to the one-class; requiring heuristic workarounds for regularization or avoiding model collapse. Thus, deep methods so far have been primarily used as feature extractors, these features are then used in a traditional one-class formulation, such as in [10]. We follow this trend.

3. Proposed Method

Using the notation above, in this section, we formally introduce our schemes. First, we present our basic idea using a pair of hyperplanes, which we generalize using a pair of discriminative subspaces for one-class classification.

3.1. Basic One-class Discriminative Subspaces

Suppose $(\mathbf{w}_1, b_1)$ and $(\mathbf{w}_2, b_2)$ define the parameters of a pair of hyperplanes respectively; our goal in the basic variant of one-class discriminative subspace (BODS) classifiers

is to minimize an objective such that all data points $\mathbf{x}_i$ be classified to the positive half-space of $(\mathbf{w}_1, b_1)$ and to the negative half-space of $(\mathbf{w}_2, b_2)$, while also minimizing a suitable distance between the two hyperplanes. Mathematically, BODS can be formulated as solving:

$$\min_{\substack{(\mathbf{w}_1,b_1),(\mathbf{w}_2,b_2),\\ \xi_1,\xi_2,\beta>0}} \frac{1}{2}\|\mathbf{w}_1\|_2^2+\frac{1}{2}\|\mathbf{w}_2\|_2^2-b_1-b_2+\boldsymbol{\Omega}(\xi_{1i},\xi_{2i}) \tag{1}$$

$$\text{s.t. } \left(\mathbf{w}_1^T\mathbf{x}_i-b_1\right)\geq\eta-\xi_{1i} \tag{2}$$

$$\left(\mathbf{w}_2^T\mathbf{x}_i-b_2\right)\leq-\eta+\xi_{2i} \tag{3}$$

$$\text{dist}^2((\mathbf{w}_i,b_1),(\mathbf{w}_2,b_2))\leq\beta,\forall i=1,2,\cdots,n, \tag{4}$$

where (2) constraints the points such that they belong to the positive half-space of $(\mathbf{w}_1, b_1)$, while (3) constraints the points to belong to the negative half-space of $(\mathbf{w}_2, b_2)$. We use the notation $\boldsymbol{\Omega}(\xi_{1i},\xi_{2i}) = C\sum_{i=1}^{n}(\xi_{1i}+\xi_{2i})$ for the slack regularization and $\eta > 0$ specifies a (given) classification margin. The two hyperplanes have their own parameters, however are constrained together by (4), which aims to minimize the distance dist between them (by β). One possibility is to assume dist to be the Euclidean distance, i.e., $\text{dist}^2((\mathbf{w}_1,b_1),(\mathbf{w}_2,b_2)) = \|\mathbf{w}_1-\mathbf{w}_2\|_2^2+(b_1-b_2)^2$.

It is often found empirically, especially in a one-class setting, that allowing the weights $\mathbf{w}_i$'s to be unconstrained leads to overfitting to the labeled data; a practical idea is to explicitly regularize them to have unit norm (and so are the data point $\mathbf{x}_i$'s), i.e., $\|\mathbf{w}_1\|_2 = \|\mathbf{w}_2\|_2 = 1$. In this case, these weights belong to a unit hypersphere U^{d-1}, which is a sub-manifold of the Euclidean manifold $\mathbb{R}^d$. Using such manifold constraints, the optimization in (1) can be rewritten (using a hinge loss variant for other constraints) as follows, which we term as our *basic one-class discriminative subspace* (BODS) classifier.

$$P1 := \min_{\substack{\mathbf{w}_1,\mathbf{w}_2\in U^{d-1}\\ \xi_1,\xi_2\geq0,b_1,b_2}} \boldsymbol{\alpha}(b_1,b_2)-2\mathbf{w}_1^T\mathbf{w}_2+\boldsymbol{\Omega}(\xi_{1i},\xi_{2i}) \tag{5}$$
$$+\sum_i\left[\eta-\left(\mathbf{w}_1^T\mathbf{x}_i+b_1\right)-\xi_{1i}\right]_+ +\left[\eta+\left(\mathbf{w}_2^T\mathbf{x}_i+b_2\right)+\xi_{2i}\right]_+,$$

where using the unit-norm constraints dist^2 simplifies to $-2\mathbf{w}_1^T\mathbf{w}_2+(b_1-b_2)^2$, and $\boldsymbol{\alpha}(b_1,b_2) = (b_1-b_2)^2-b_1-b_2$. The notation $[\ ]_+$ stands for the hinge loss. In Figure 1(c), we illustrate the decision boundaries of BODS model.

3.2. Generalized One-class Discriminative Subspaces

To set the stage, let us first see what happens if we introduce subspaces instead of hyperplanes in BODS. To this end, let $\mathbf{W}_1, \mathbf{W}_2 \in \mathcal{S}_d^K$ be subspace frames – that is, matrices of dimensions $d \times K$, each with K columns where each column is orthonormal to the rest; i.e., $\mathbf{W}_1^T\mathbf{W}_1 = \mathbf{W}_2^T\mathbf{W}_2 = \mathbf{I}_K$, where $\mathbf{I}_K$ is the $K \times K$ identity matrix. Such frames belong to the so-called Stiefel manifold, denoted $\mathcal{S}_d^K$, with K d-dimensional subspaces. Note that the orthogonality assumption on the $\mathbf{W}_i$'s is to ensure they capture diverse discriminative directions, leading to better regularization; further also improving their characterization of the data distribution. A direct extension of P1 leads to:

$$P2 := \min_{\mathbf{W}\in\mathcal{S}_d^K,\xi\geq0,\mathbf{b}} \text{dist}_\text{W}^2(\mathbf{W}_1,\mathbf{W}_2)+\boldsymbol{\alpha}(\mathbf{b}_1,\mathbf{b}_2)+\boldsymbol{\Omega}(\xi_{1i},\xi_{2i})$$
$$+\sum_i\left[\eta-\min(\mathbf{W}_1^T\mathbf{x}_i+\mathbf{b}_1)-\xi_{1i}\right]_+^2 \tag{6}$$
$$+\sum_i\left[\eta+\max(\mathbf{W}_2^T\mathbf{x}_i+\mathbf{b}_2)+\xi_{2i}\right]_+^2, \tag{7}$$

where dist_W is a suitable distance between subspaces, and $\mathbf{b} \in \mathbb{R}^K$ is a vector of biases, one for each hyperplane. Note that in (6) and (7), unlike BODS, $\mathbf{W}^T\mathbf{x}_i+\mathbf{b}$ is a K-dimensional vector. Thus, (6) says that the minimum value of this vector should be greater than η and (7) says that the maximum value of it is less than $-\eta$.

Now, let us take a closer look at the $\text{dist}_\text{W}(\mathbf{W}_1,\mathbf{W}_2)$. Given that $\mathbf{W}_1, \mathbf{W}_2$ are subspaces, one standard possibility for a distance is the *Procrustes distance* [12, 55] defined as $\min_{\Pi\in\mathcal{P}_K}\|\mathbf{W}_1-\mathbf{W}_2\Pi\|_F$, where $\mathcal{P}_K$ is the set of $K \times K$ permutation matrices. However, including such a distance in Problem P2 makes it computationally expensive. To this end, we propose a slightly different variant of this distance which is much cheaper. Recall that the main motivation to define the distance between the subspaces is so that they sandwich the (one-class) data points to the best possible manner while also catering to the data distribution. Thus, rather than defining a distance between such subspaces, one could also use a measure that minimizes the Euclidean distance of each data point from both the hyperplanes; thereby achieving the same effect. That is, we redefine dist_W^2 as:

$$\text{dist}_\text{W}^2(\mathbf{W}_1,\mathbf{W}_2,\mathbf{b}_1,\mathbf{b}_2|\mathbf{x}) = \sum_{j=1}^{2}\left\|\mathbf{W}_j^T\mathbf{x}+\mathbf{b}_j\right\|_2^2, \tag{8}$$

where now we minimize the sum of the lengths of each $\mathbf{x}$ after projecting on to the respective subspaces; thereby pulling both the subspaces closer to the data point. Using this definition of dist_W^2, we formulate our *generalized one-class discriminative subspace* (GODS) classifier as:

$$P3 := \min_{\substack{\mathbf{W}\in\mathcal{S}_d^K\\ \xi\geq0,b}} F = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{2}\left\|\mathbf{W}_j^T\mathbf{x}_i+\mathbf{b}_j\right\|_2^2+\boldsymbol{\alpha}(\mathbf{b}_1,\mathbf{b}_2)$$
$$+\boldsymbol{\Omega}(\xi_{1i},\xi_{2i})+\frac{\nu}{n}\sum_i\left[\eta-\min(\mathbf{W}_1^T\mathbf{x}_i+\mathbf{b}_1)-\xi_{1i}\right]_+^2$$
$$+\frac{1}{2n}\sum_i\left[\eta+\max(\mathbf{W}_2^T\mathbf{x}_i+\mathbf{b}_2)+\xi_{2i}\right]_+^2. \tag{9}$$

Figure 1(d) depicts the subspaces in GODS model in relation to other methods. As is intuitively clear, using multiple hyperplanes allows richer characterization of the one-class, which is difficult in other schemes.

4. Efficient Optimization

In contrast to OC-SVM and SVDD, the problem P3 is non-convex due to the orthogonality constraints on $\mathbf{W}_1$ and $\mathbf{W}_2$.[1] However, these constraints naturally impose a geometry to the solution space and in our case, puts the $\mathbf{W}$'s on the well-known Stiefel manifold [37] – a Riemannian manifold characterizing the space of all orthogonal frames. There exist several schemes for geometric optimization over Riemannian manifolds (see [1] for a detailed survey) from which we use the Riemannian conjugate gradient (RCG) scheme in this paper, due to its stable and fast convergence. In the following, we review some essential components of the RCG scheme and provide the necessary formulae for using it to solve our objective.

4.1. Riemannian Conjugate Gradient

Recall that the standard (Euclidean) conjugate gradient (CG) method [1][Sec.8.3] is a variant of the steepest descent method, however chooses its descent along directions conjugate to previous descent directions with respect to the parameters of the objective. Formally, suppose $F(\mathbf{W})$ represents our objective. Then, the CG method uses the following recurrence at the k-th iteration:

$$\mathbf{W}^k = \mathbf{W}^{k-1} + \lambda^{k-1}\alpha^{k-1}, \tag{10}$$

where λ is a suitable step-size (found using line-search) and $\alpha^{k-1} = -\operatorname{grad} F(\mathbf{W}^{k-1}) + \mu^{k-1}\alpha^{k-2}$, where $\operatorname{grad} F(\mathbf{W}^{k-1})$ defines the gradient of F at $\mathbf{W}^{k-1}$ and α^{k-1} is a direction built over the current residual and conjugate to previous descent directions (see [1][pp.182])).

When $\mathbf{W}$ belongs to a curved Riemannian manifold, we may use the same recurrence, however there are a few important differences from the Euclidean CG case, namely (i) we need to ensure that the updated point $\mathbf{W}^k$ belongs to the manifold, (ii) there exists efficient vector transports[2] for computing α^{k-1}, and (iii) the gradient grad is along tangent spaces to the manifold. For (i) and (ii), we may resort to computationally efficient retractions (using QR factorizations; see [1][Ex.4.1.2]) and vector transports [1][pp.182], respectively. For (iii), there exist standard ways that take as input a Euclidean gradient of the objective (i.e., assuming no manifold constraints exist), and maps them to the Riemannian gradients [1][Chap.3]. Specifically, for the Stiefel manifold, let $\nabla_{\mathbf{W}} F(\mathbf{W})$ define the Euclidean gradient of F (without the manifold constraints), then the Riemannian gradient is given by:

$$\operatorname{grad} F(\mathbf{W}) = (\mathbf{I} - \mathbf{W}\mathbf{W}^T)\nabla_{\mathbf{W}} F(\mathbf{W}). \tag{11}$$

The direction $\operatorname{grad} F(\mathbf{W})$ corresponds to a curve along the manifold, descending along which ensures the optimization objective is decreased (atleast locally).

Now, getting back to our one-class objective, all we need to derive to use the RCG, is compute the Euclidean gradients $\nabla_{\mathbf{W}} F(\mathbf{W})$ of our objective in P3 with regard to the variables $\mathbf{W}_j$'s; the other variables (such as the biases) are Euclidean and their gradients are straightforward, and the joint objective can be solved via RCG on the product manifold comprising the Cartesian product of the Stiefel and the Euclidean manifolds. Thus, the only non-trivial part is the expression for the Euclidean gradient of our objective with respect to the $\mathbf{W}$'s, which is given by:

$$\frac{\partial F}{\partial \mathbf{W}_j} = \sum_{i=1}^{n} \mathbf{x}_i \left(\mathbf{W}_j^T \mathbf{x}_i + \mathbf{b}_1\right)^T - \mathbf{Z}_{i^*}\left[\eta - \mathbf{W}_j^T \mathbf{x}_i - b_j - \xi_{ji}\right]_+, \tag{12}$$

where $i^* = h(\mathbf{W}_j^T \mathbf{x}_i + \mathbf{b}_j)$, h abstracts $\arg\min_k$ and $-\arg\max_k$ for $\mathbf{W}_1$ and $\mathbf{W}_2$ respectively, i^* denotes the selected hyperplane index (out of K) and $\mathbf{Z}_{i^*}$ is a $d \times K$ matrix with all zeros, except i^*-th column which is $\mathbf{x}_i$.

4.2. Initialization

Due to the non-convexity of our objective, there could be multiple local solutions. To this end, we resort to the following initialization of our optimization variables, which we found to be empirically beneficial. Specifically, we first sort all the data points based on their Euclidean distances from the origin. Next, we gather a suitable number (depending on the number of subspaces) of such sorted points near and far from the origin, compute a singular value decomposition (SVD) of these points, and initialize the GODS subspaces using these orthonormal matrices from the SVD.

5. One-class Classification

At test time, suppose we are given m data points, and our task is to classify each of them as belonging to either $\mathcal{D}$ or $\overline{\mathcal{D}}$. To this end, we use the learned parameters of our problem P3 as above, and compute the score for each point (using (9)). Next, we use K-means clustering (we could also use graph-cut) on these scores with $K = 2$. Those points belonging to the cluster with smaller scores are deemed to belong to $\mathcal{D}$ and the rest to $\overline{\mathcal{D}}$.

6. Experiments

In this section, we provide experiments demonstrating the performance of our proposed schemes on several one-

[1] Note that the function $\max(0, \min(z))$ for z in some convex set is also a non-convex function.

[2] This is required for computing α_{k-1} that involves the sum of two terms in potentially different tangent spaces, which would need vector transport for moving between them (see [1][pp.182].

class tasks, namely (i) out-of-position human pose detection using the Dash-Cam-Pose dataset, (ii) human action recognition in videos using the popular JHMDB dataset, (iii) UCF-Crime dataset to find anomalous video events, (iv) discriminating sonar signals from a metal cylinder and a roughly cylindrical rock using the Sonar dataset[3], and (v) abnormality detection in a submersible pump using the Delft pump dataset[4]. Before proceeding, we first introduce our new Dash-Cam-Pose dataset.

6.1. Dash-Cam-Pose: Data Collection

Out-of-position (OOP) human pose detection is an important problem with regard to the safety of passengers in a car. While, there are very large public datasets for human pose estimation – such as the Pose Track [21] and MPII Pose [3] datasets, among others – these datasets are for generic pose estimation tasks, and neither they contain any in-vehicle poses as captured by a dashboard camera, nor are they annotated for pose anomalies. To this end, we collected about 104 videos, each 20-30 min long from the Internet (including Youtube, ShutterStock, and Hollywood road movies). As these videos were originally recorded for diverse reasons, there are significant shifts in camera angles, perspectives, locations of the camera, scene changes, etc.

To extract as many clips as possible from these videos, we segmented them to three second clips at 30fps, which resulted in approximately 7000 clips. Next, we selected only those clips where the camera is approximately placed on the dashboard looking inwards, which amounted to 4,875 clips. We annotated each clip with a weak binary label based on the poses of humans in the front seat (the back seat passengers often underwent severe occlusions, as a result, was harder to estimate their poses). Specifically, if all the front-seat humans (passengers and the driver) are seated in-position, the clip was given a positive label, while if any human is seated OOP for the entire 3s, the clip was labeled as negative. We do not give annotations for which human is seated in OOP. The in-position and out-of-position criteria are defined loosely based on the case studies in [39, 16], the primary goal being to avoid passenger fatality due to an OOP if airbags are deployed.

After annotating the clips with binary labels, we applied Open Pose [8] on each clip extracting a sequence of poses for every person. These sequences are filtered for poses belonging to only the front seat humans. Figure 2 shows a few frames from various clips. As is clear from the examples, the OOP poses could be quite arbitrary and difficult to model; which is the primary motivation to seek a one-class solution for this task. In the following section, we detail our data preparation and evaluation scheme. Some statistics of the dataset are provided in Table 1.

[3] https://www.kaggle.com/adx891/sonar-data-set

[4] http://homepage.tudelft.nl/n9d04/occ/547/oc_547.html

Dash-Cam-Pose Dataset	
Total # clips	4875
% of clips with OOP poses	28.5%
Total # poses	1.06M
Total # OOP poses	310,996

Table 1. Attributes of the proposed Dash-Cam-Pose dataset.

6.2. Dash-Cam-Pose: Preparation and Evaluation

Suitable representation of the poses is important for using them in the one-class task. To this end, we explore two representations, namely (i) a simple bag-of-words (BoW) model of poses learned from the training set, and (ii) using a Temporal Convolutional Network (TCN) [27] which uses residual units with 1D convolutional layers, capturing both local and global information via convolutions for each joint across time. For the former, we 1024 pose centroids, while for the latter the poses from each person in each frame are vectorized and stacked over the temporal dimension. The TCN model we use has been pre-trained on the larger NTU-RGBD dataset [47] on 3D-skeletons for the task of human action recognition. For each pose thus passed through TCN, we extract features from the last pooling layer, which are 256-D vectors for each clip.

We use a four-fold cross-validation for evaluating on Dash-Cam-Pose. Specifically, we divide the entire dataset into four non-overlapping splits, each split consisting of approximately 1/4-th the dataset, of which roughly 2/3rd's are the labeled positive and the rest are OOP. *We use only the positive data in each split to train our one-class models.* Once the models are trained, we evaluate on the held out split. For every embedded-pose feature, we use the binary classification accuracy against the annotated ground truth for measuring performance. The evaluation is repeated on all the four splits and the performance averaged.

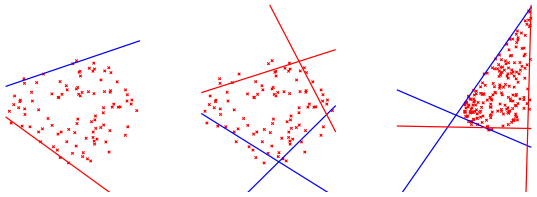
6.3. Public Datasets

JHMDB dataset: is a video action recognition dataset [23] consisting of 968 clips with 21 classes (illustrative frames are provided in Figure 3). To adapt the dataset for a one-class evaluation, we use a one-versus-rest strategy by choosing sequences from an action class as "normal" while those from the rest 20 classes are treated as "abnormal". To evaluate the performance over the entire dataset, we cycle over the 21 classes, and the scores are averaged. For representing the frames, we use an image-net [28] pre-trained VGG-16 model and extract frame-level features from the 'fc-6' layer (4096-D).

UCF-Crime dataset: is the largest publicly available real-world anomaly detection dataset [50], consisting of 1900 surveillance videos and 13 categories such as *fighting*, *robbery*, as well as several "normal" activities. Illustrative video frames from this dataset and their class labels are shown in Figure 3. To encode the videos, we use the state-

Figure 2. Frames from our proposed Dash-Cam-Pose dataset. The leftmost frame has poses in-position (one-class), while the rest of the frames are from videos labeled out-of-position.

Figure 3. Some examples from JHMDB (left-two) and UCF-Crime (right-two) datasets, with respective categories.



(a) BODS-Gaussian (b) GODS-Gaussian (c) GODS-Arbitrary

Figure 4. Visualizations of subspaces found by BODS (leftmost) and GODS on various data distributions.

of-the-art Inflated-3D (I3D) neural network [9]. Specifically, video frames from non-overlapping sliding windows (8 frames each) is passed through the I3D network; features are extracted from the 'Mix_5c' network layer, that are then reshaped to 2048-D vectors. For anomaly detections on the test set, we first map back the features classified as anomalies by our scheme to the frame-level and apply the official evaluation metrics [50].

Sonar and Delft pump dataset: are two UCI datasets, having 208 and 1500 data points respectively, and two classes. We directly adopt the raw feature (60-D and 64-D) without any feature embedding. We keep the train/test ratio as 7/3 while keeping the original proportion of each class in each set. We randomly pick train/test splits and the evaluation is repeated 5 times and performances averaged.

6.4. Evaluation Metrics

On the UCF-Crime dataset, we follow the official evaluation protocol, reporting AUC as well as the false alarm rate. For other datasets, we use the F1 score to reflect the sensitivity and accuracy of our classification models. As the datasets we use - especially the Dash-Cam-Pose – are unbalanced across the two classes, having a single performance metric over the entire dataset may fail to characterize the quality of the discrimination for each class separately, which is of primary importance for the one-class task. To this end, we also report True Negative Rate $TNR = \frac{TN}{N}$, Negative Predictive Value $NPV = \frac{TN}{TN+FN}$, and $\overline{F1} = \frac{2\times TNR\times NPV}{TNR+NPV}$ alongside standard F1 scores.

6.5. Ablative Studies

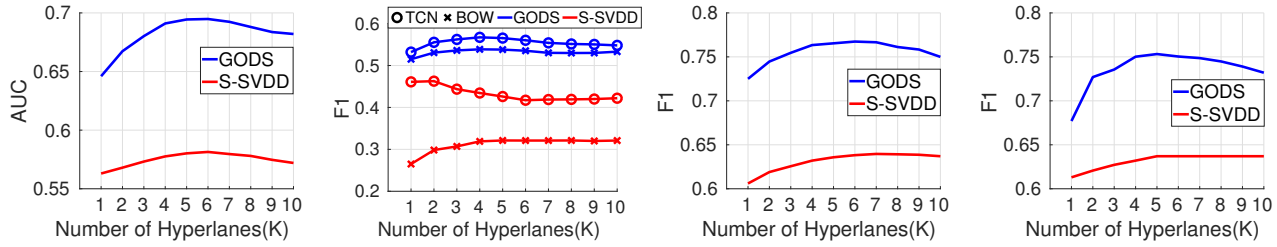
Synthetic Experiments: To gain insights into the inner workings of our schemes, we present results on several 2D synthetic toy datasets. In Figure 4, we show four plots with 100 points distributed as (i) Gaussian and (ii) some arbitrary distribution[5]. We show the BODS hyperplanes in the first plot, and the rest two plots show the GODS 2D subspaces with the hyperplanes belonging to each subspace shown in same color. As the plots show, our models are able to orient the subspaces such that they confine the data within a minimal volume. More results are provided in the supplementary materials.

Parameter Study: In Figure 5, we plot the influence of increasing number of hyperplanes on four of the datasets. We find that after a certain number of hyperplanes, the performance saturates, which is expected, and suggests that more hyperplanes might lead to overfitting to the positive class. We also find that the TCN embedding is significantly better than the BoW model (by nearly 3%) on the Dash-Cam-Pose dataset when using our proposed methods. Surprisingly, S-SVDD is found to perform quite inferior against ours; note that this scheme learns a low-dimensional subspace to project the data to (as in PCA), and applies SVDD on this subspace. We believe, these subspaces perhaps are common to the negative points as well that it cannot be suitably discriminated, leading to poor performance. We make a similar observation on the other datasets as well.

6.6. State-of-the-Art Comparisons

In Tables 2, we compare our variants to the state-of-the-art methods. As alluded to earlier, for our Dash-Cam-Pose dataset, as its positive and negative classes are unbalanced, we resort to reporting the $\overline{F1}$ score on the negative set. As is clear from the table, our variants outperform prior meth-

[5]The data follows the formula $f(x) = \sqrt{x} * (x + sign(randn) * rand)$, where randn and rand are standard MATLAB functions.



(a) UCF-Crime (b) Dash-Cam-Pose (c) JHMDB (d) Sonar

Figure 5. Performance of our method on various datasets for an increasing number of subspaces.

Table 2. Average performances on the four datasets, where Dash-Cam-Pose use the $\overline{F1}$ score while the rest use $F1$ score as evaluation metric (classification accuracy is shown in the brackets). K-OC-SVM and K-SVDD are the RBF kernelized variants.

Method	Dash-Cam-Pose_BOW	Dash-Cam-Pose_TCN	JHMDB	Sonar	Pump
OC-SVM [46]	0.167 (0.517)	0.279(0.527)	0.301 (0.568)	0.578 (0.459)	0.623(0.482)
SVDD [51]	0.448 (0.489)	0.477(0.482)	0.407 (0.566)	0.605 (0.479)	0.813 (0.516)
K-OC-SVM [46]	0.327 (0.495)	0.361(0.491)	0.562 (0.412)	0.565 (0.429)	0.601 (0.499)
K-SVDD [51]	0.476 (0.477)	0.489 (0.505)	0.209 (0.441)	0.585 (0.474)	0.809 (0.529)
K-PCA [20]	0.145 (0.502)	0.258 (0.492)	0.245 (0.557)	0.530 (0.426)	0.611 (0.416)
Slab-SVM [18]	0.468 (0.568)	0.498 (0.577)	0.643 (0.637)	0.600 (0.619)	0.809 (0.621)
LS-OSVM [14]	0.234 (0.440)	0.246(0.460)	0.663(0.582)	0.643 (0.466)	0.831 (0.448)
S-SVDD [49]	0.325 (0.490)	0.464 (0.500)	0.642 (0.498)	0.637 (0.500)	0.865 (0.500)
BODS	0.523 (0.582)	0.532 (0.579)	0.725 (0.714)	0.677 (0.662)	0.823 (0.714)
GODS	**0.553 (0.629)**	**0.584 (0.601)**	**0.777 (0.752)**	**0.762 (0.775)**	**0.892 (0.755)**

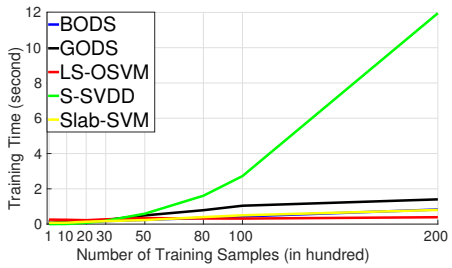


Figure 6. Training time of each method with increasing number of training samples.

Method	AUC	False alarm rate
Random	50.00	-
Hasan et al. [19]	50.60	27.2
Lu et al. [35]	65.51	3.1
*Waqas et al. [50]	75.41	1.9
Sohrab et al. [49]	58.50	10.5
BODS	68.26	2.7
GODS	**70.46**	**2.1**

Table 3. Performances on UCF-Crime dataset. *Setup is different.

ods by a considerable margin. For example, using TCN, GODS is over 30% better than OC-SVM; even we outperform the kernelized variants by about 20%. Similarly, on the JHMDB and the other two datasets, GODS is better than the next best method by about 3-13%, associated with a significant improvement for the classification accuracy (by over 10%). As the classes used in the test set for these datasets are balanced, we report the F1 scores. Overall, the experiments clearly substantiate the performance benefits afforded by our method on the one-class task. In the Figure 6, we demonstrate the time consumption for training different models. It can be seen that the GODS & BODS algorithm are not computationally expensive than other methods, while being is empirically superior (Table 2).

In Table 3, we present results against the state of the art on the UCF-Crime dataset using the AUC metric and false alarm rates (we use the standard threshold of 50%). While, our results are lower than [50], their problem setup is completely different from ours in that they use weakly labeled abnormal videos as well in their training, which we do not use and which as per definition is not a one-class problem. Thus, our results are incomparable to theirs. On other methods for this dataset, our methods are about 5-20% better.

7. Conclusions

In this paper, we presented a novel one-class learning formulation using subspaces in a discriminative setup, these subspaces are oriented in such a way as to sandwich the data. Due to the non-linear constraints optimization problem that ensues, we cast the objective in Riemannian context however, for which we derived efficient numerical solutions. Experiments on a diverse collection of five datasets, including our new Dash-Cam-Pose dataset, demonstrated the usefulness of our approach achieving state-of-the-art performances.

References

[1] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009. 2, 5

[2] Amit Adam, Ehud Rivlin, Ilan Shimshoni, and Daviv Reinitz. Robust real-time unusual event detection using multiple fixed-location monitors. *IEEE transactions on pattern analysis and machine intelligence*, 30(3):555–560, 2008. 3

[3] Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, and Bernt Schiele. 2d human pose estimation: New benchmark and state of the art analysis. In *ICCV*, 2014. 6

[4] Christopher M Bishop. Novelty detection and neural network validation. *IEEE Proceedings-Vision, Image and Signal processing*, 141(4):217–222, 1994. 1

[5] Paul Bodesheim, Alexander Freytag, Erik Rodner, Michael Kemmler, and Joachim Denzler. Kernel null space methods for novelty detection. In *CVPR*, pages 3374–3381, 2013. 3

[6] William M Boothby. *An introduction to differentiable manifolds and Riemannian geometry*, volume 120. Academic press, 1986. 2

[7] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011. 3

[8] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. *arXiv preprint arXiv:1611.08050*, 2016. 2, 6

[9] Joao Carreira and Andrew Zisserman. Quo vadis, action recognition? a new model and the kinetics dataset. In *CVPR*, pages 4724–4733. IEEE, 2017. 7

[10] Raghavendra Chalapathy, Aditya Krishna Menon, and Sanjay Chawla. Anomaly detection using one-class neural networks. *arXiv preprint arXiv:1802.06360*, 2018. 3

[11] V Chandala, A Banerjee, and V Kumar. Anomaly detection: A survey, ACM computing surveys. *University of Minnesota*, 2009. 1, 2, 3

[12] Yasuko Chikuse. *Statistics on special manifolds*, volume 174. Springer Science & Business Media, 2012. 4

[13] Myung Jin Choi, Antonio Torralba, and Alan S Willsky. Context models and out-of-context objects. *Pattern Recognition Letters*, 33(7):853–862, 2012. 3

[14] Young-Sik Choi. Least squares one-class support vector machine. *Pattern Recognition Letters*, 30(13):1236–1240, 2009. 1, 3, 8

[15] Fernando De La Torre and Michael J Black. A framework for robust subspace learning. *IJCV*, 54(1-3):117–142, 2003. 3

[16] Stefan M Duma, Tyler A Kress, David J Porta, Charles D Woods, John N Snider, Peter M Fuller, and Rod J Simmons. Airbag-induced eye injuries: a report of 25 cases. *Journal of Trauma and Acute Care Surgery*, 41(1):114–119, 1996. 6

[17] Alan Edelman, Tomás A Arias, and Steven T Smith. The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications*, 20(2):303–353, 1998. 2

[18] Victor Fragoso, Walter Scheirer, Joao Hespanha, and Matthew Turk. One-class slab support vector machine. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 420–425. IEEE, 2016. 3, 8

[19] Mahmudul Hasan, Jonghyun Choi, Jan Neumann, Amit K Roy-Chowdhury, and Larry S Davis. Learning temporal regularity in video sequences. In *CVPR*, pages 733–742, 2016. 8

[20] Heiko Hoffmann. Kernel pca for novelty detection. *Pattern recognition*, 40(3):863–874, 2007. 3, 8

[21] Umar Iqbal, Anton Milan, and Juergen Gall. Posetrack: Joint multi-person pose estimation and tracking. *arXiv preprint arXiv:1611.07727*, 2016. 6

[22] Laurent Itti and Christof Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, 40(10-12):1489–1506, 2000. 3

[23] Hueihan Jhuang, Juergen Gall, Silvia Zuffi, Cordelia Schmid, and Michael J Black. Towards understanding action recognition. In *ICCV*, 2013. 2, 6

[24] Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba. Learning to predict where humans look. In *ICCV*, 2009. 3

[25] MU Khan, M Moatamedi, Mhamed Souli, and Tayeb Zeguer. Multiphysics out of position airbag simulation. *International journal of crashworthiness*, 13(2):159–166, 2008. 2

[26] Jaechul Kim and Kristen Grauman. Observe locally, infer globally: a space-time mrf for detecting abnormal activities with incremental updates. In *CVPR*. IEEE, 2009. 3

[27] Tae Soo Kim and Austin Reiter. Interpretable 3d human action analysis with temporal convolutional networks. In *CVPRW*, 2017. 6

[28] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012. 6

[29] Vinh Lai, Duy Nguyen, Khanh Nguyen, and Trung Le. Mixture of support vector data descriptions. In *Conference on Information and Computer Science*. IEEE, 2015. 3

[30] KiYoung Lee, Dae-Won Kim, Kwang H Lee, and Doheon Lee. Density-induced support vector data description. *IEEE Transactions on Neural Networks*, 18(1):284–289, 2007. 3

[31] Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. In *NIPS*, pages 7167–7177, 2018. 3

[32] Ce Li, Zhenjun Han, Qixiang Ye, and Jianbin Jiao. Visual abnormal behavior detection based on trajectory sparse reconstruction analysis. *Neurocomputing*, 119:94–100, 2013. 3

[33] Shiyu Liang, Yixuan Li, and R. Srikant. Principled detection of out-of-distribution examples in neural networks. In *International Conference on Learning Representations*, 2018. 3

[34] Wei Liu, Gang Hua, and John R Smith. Unsupervised one-class learning for automatic outlier removal. In *CVPR*, pages 3826–3833, 2014. 3

[35] Cewu Lu, Jianping Shi, and Jiaya Jia. Abnormal event detection at 150 fps in matlab. In *CVPR*, pages 2720–2727, 2013. 8

[36] P-O Marklund and Larsgunnar Nilsson. Optimization of airbag inflation parameters for the minimization of out of position occupant injury. *Computational Mechanics*, 31(6):496–504, 2003. 2

[37] Robb J Muirhead. *Aspects of multivariate statistical theory*, volume 197. John Wiley & Sons, 2009. 5

[38] Minh H Nguyen and Fernando Torre. Robust kernel principal component analysis. In *NIPS*, pages 1185–1192, 2009. 3

[39] Larry S Nordhoff. *Motor vehicle collision injuries: biomechanics, diagnosis, and management*. Jones & Bartlett Learning, 2005. 6

[40] Sangdon Park, Wonsik Kim, and Kyoung Mu Lee. Abnormal object detection by canonical scene-based contextual model. In *ECCV*, 2012. 3

[41] Pramuditha Perera and Vishal M Patel. Learning deep features for one-class classification. *arXiv preprint arXiv:1801.05365*, 2018. 3

[42] Marco AF Pimentel, David A Clifton, Lei Clifton, and Lionel Tarassenko. A review of novelty detection. *Signal Processing*, 99:215–249, 2014. 3

[43] Gordon R Plank, Michael Kleinberger, and Rolf H Eppinger. Analytical investigation of driver thoracic response to out of position airbag deployment. Technical report, SAE Technical Paper, 1998. 2

[44] Gunter Ritter and María Teresa Gallegos. Outliers in statistical pattern recognition and an application to automatic chromosome classification. *Pattern Recognition Letters*, 18(6):525–539, 1997. 1

[45] Lukas Ruff, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Robert Vandermeulen, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *ICML*, pages 4390–4399, 2018. 3

[46] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001. 1, 2, 8

[47] Amir Shahroudy, Jun Liu, Tian-Tsong Ng, and Gang Wang. NTU RGB+ D: A large scale dataset for 3d human activity analysis. In *CVPR*, 2016. 6

[48] Rowland R Sillito and Robert B Fisher. Semi-supervised learning for anomalous trajectory detection. In *BMVC*, 2008. 3

[49] Fahad Sohrab, Jenni Raitoharju, Moncef Gabbouj, et al. Subspace support vector data description. *arXiv preprint arXiv:1802.03989*, 2018. 3, 8

[50] Waqas Sultani, Chen Chen, and Mubarak Shah. Real-world anomaly detection in surveillance videos. *Center for Research in Computer Vision (CRCV), University of Central Florida (UCF)*, 2018. 2, 6, 7, 8

[51] David MJ Tax and Robert PW Duin. Support vector data description. *Machine learning*, 54(1):45–66, 2004. 2, 3, 8

[52] David Martinus Johannes Tax. One-class classification: concept-learning in the absence of counter-examples [ph. d. thesis]. *Delft University of Technology*, 2001. 3

[53] Mohan M Trivedi, Shinko Yuanhsien Cheng, Edwin MC Childers, and Stephen J Krotosky. Occupant posture analysis with stereo and thermal infrared video: Algorithms and experimental evaluation. *IEEE transactions on vehicular technology*, 53(6):1698–1712, 2004. 2

[54] Mohan Manubhai Trivedi, Tarak Gandhi, and Joel McCall. Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety. *IEEE Transactions on Intelligent Transportation Systems*, 8(1):108–120, 2007. 2

[55] Pavan Turaga, Ashok Veeraraghavan, and Rama Chellappa. Statistical analysis on stiefel and grassmann manifolds with applications in computer vision. In *CVPR*, 2008. 4

[56] Jue Wang and Anoop Cherian. Learning discriminative video representations using adversarial perturbations. In *ECCV*, 2018. 2, 3

[57] Jue Wang, Anoop Cherian, Fatih Porikli, and Stephen Gould. Video representation learning using discriminative pooling. In *CVPR*, 2018. 3

[58] Tian Wang, Jie Chen, Yi Zhou, and Hichem Snoussi. Online least squares one-class support vector machines-based abnormal visual event detection. *Sensors*, 13(12):17130–17155, 2013. 1

[59] Dan Xu, Yan Yan, Elisa Ricci, and Nicu Sebe. Detecting anomalous events in videos by learning deep representations of appearance and motion. *CVIU*, 156:117–127, 2017. 3

[60] Huan Xu, Constantine Caramanis, and Shie Mannor. Outlier-robust pca: the high-dimensional case. *IEEE transactions on information theory*, 59(1):546–572, 2013. 3