

Article

Low-Altitude Aerial Video Surveillance via One-Class SVM Anomaly Detection from Textural Features in UAV Images

Danilo Avola ^{1,*}, Luigi Cinque ¹, Angelo Di Mambro ¹, Anxhelo Diko ¹, Alessio Fagioli ^{1,*},
Gian Luca Foresti ², Marco Raoul Marini ¹, Alessio Mecca ¹ and Daniele Pannone ^{1,*}

¹ Department of Computer Science, Sapienza University, 00198 Rome, Italy; cinque@di.uniroma1.it (L.C.); dimambro@di.uniroma1.it (A.D.M.); diko@di.uniroma1.it (A.D.); marini@di.uniroma1.it (M.R.M.); mecca@di.uniroma1.it (A.M.)

² Department of Mathematics, Computer Science and Physics, University of Udine, 33100 Udine, Italy; gianluca.foresti@uniud.it

* Correspondence: avola@di.uniroma1.it (D.A.); fagioli@di.uniroma1.it (A.F.); pannone@di.uniroma1.it (D.P.)

Abstract: In recent years, small-scale Unmanned Aerial Vehicles (UAVs) have been used in many video surveillance applications, such as vehicle tracking, border control, dangerous object detection, and many others. Anomaly detection can represent a prerequisite of many of these applications thanks to its ability to identify areas and/or objects of interest without knowing them a priori. In this paper, a One-Class Support Vector Machine (OC-SVM) anomaly detector based on customized Haralick textural features for aerial video surveillance at low-altitude is presented. The use of a One-Class SVM, which is notoriously a lightweight and fast classifier, enables the implementation of real-time systems even when these are embedded in low-computational small-scale UAVs. At the same time, the use of textural features allows a vision-based system to detect micro and macro structures of an analyzed surface, thus allowing the identification of small and large anomalies, respectively. The latter aspect plays a key role in aerial video surveillance at low-altitude, i.e., 6 to 15 m, where the detection of common items, e.g., cars, is as important as the detection of little and undefined objects, e.g., Improvised Explosive Devices (IEDs). Experiments obtained on the UAV Mosaicking and Change Detection (UMCD) dataset show the effectiveness of the proposed system in terms of accuracy, precision, recall, and F₁-score, where the model achieves a 100% precision, i.e., never misses an anomaly, but at the expense of a reasonable trade-off in its recall, which still manages to reach up to a 71.23% score. Moreover, when compared to classical Haralick textural features, the model obtains significantly higher performances, i.e., ≈20% on all metrics, further demonstrating the approach effectiveness.

Keywords: anomaly detection; small-scale unmanned aerial vehicles; low-altitude flights; texture analysis; feature extraction; real-time applications; support vector machines



Citation: Avola, D.; Cinque, L.; Di Mambro, A.; Diko, A.; Fagioli, A.; Foresti, G.L.; Marini, M.R.; Mecca, A.; Pannone, D. Low-Altitude Aerial Video Surveillance via One-Class SVM Anomaly Detection from Textural Features in UAV Images. *Information* **2022**, *13*, 2. <https://doi.org/10.3390/info13010002>

Academic Editor: Gholamreza Anbarjafari (Shahab)

Received: 24 October 2021

Accepted: 18 December 2021

Published: 22 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the last decades, vision-based systems are becoming increasingly important in supporting a wide range of application areas, including environment modelling for moving cameras [1–5], human action and event recognition [6–9], target and object detection [10–16]; even in areas such as medical image analysis [17–23], emotion or deception recognition [24–29], and immersive rehabilitation by serious games [30–35], these systems are, now, of daily use. At the same time, the last 10 years have seen substantial improvements of small-scale Unmanned Aerial Vehicles (UAVs), hereinafter UAVs, in terms of flight time, automatic control, embedded processing, and remote transmission. All these aspects have enabled an increasing development of vision systems based on UAVs [36,37] that more and more support activities in civilian and military missions, thus making a new era in human-drone collaboration.

Regarding civilian-centered applications, the last few years have been characterized by the development of extraordinary UAV-based vision systems for research and commercial

purposes. This was possible thanks to the greater payload capacity of the current UAVs, together with advances in HD cameras and gimbal mounts that have made these devices ever more performing and at an increasingly limited cost. In this context, industries, research centers, and public authorities have started a massive usage of vision systems based on UAVs beginning from daily problems such as agriculture monitoring [38], traffic surveillance [39], and road pavement analysis [40]. In [41], for example, the authors propose a video system based on drones to map terrains and plan the construction of efficient drainage networks. Since the images provided by satellites are not detailed enough to reach the target reported above, the authors use a combination of aerial images captured by UAVs equipped with high-resolution cameras and multiple Differential Global Positioning System (DGPS) stations, thus providing a consistent amount of information for generating quality Digital Terrain Models (DTMs). The latter are exploited to propose multiple drainage networks and identify the optimal ones. The work just introduced highlights another essential topic of civilian-centered applications, i.e., the reconstruction of virtual models. In fact, numerous proposals [42–44] show how three-dimensional (3D) digitization of structures and objects can significantly improve the accomplishment of a wide range of aerial tasks. In [45], for example, the authors provide a method to reconstruct an entire scenario in 3D and allow the planning of new building structures. The method shows how a drone equipped with an HD camera performing a circular flight can capture all the required information about the area of interest to implement the related 3D digital model. In particular, the authors propose a technique based on a dense point cloud [46] to reconstruct a surface of an object starting from multi-view images.

Another civilian application area in which reconstruction techniques are widely used regards the change detection task. The work presented in [47], for instance, shows how 3D reconstructed models can support the analysis of volumetric changes in generic environments. The combination of Unmanned Aerial Systems (UASs), Structure-from-Motion (SfM) multi-view stereophotogrammetry [48], and Terrestrial Laser Scanning (TLS) allows the authors to retrieve precise information about the morphology of structures. As for the previous work, the algorithmic approach exploits point cloud techniques. The results obtained from the application on Bedrock Coastal Cliffs underline the efficiency of the proposed technique, even if, in very small areas (between 2.7 and 31.5 mm, with 20.5 mm of standard deviation), the changes may not be detected. In [49], instead, the UAVs are exploited for detecting changes in a specific area. The proposed method is designed to monitor the progress of road construction. Two orthoimages of the same area are generated with a drone equipped with an HD camera, and successively, these are used as input of a Convolutional Siamese Metric Network (CosimNet). The latter is a Convolutional Neural Network (CNN)-based structure that contains multiple identical sub-networks and allows to compare their outputs. Different features are detected and localized into the scene, thus allowing to segment the region. In this application area, Siamese networks are used for change detection missions [50] but also for tracking tasks [51] by UAVs. Regardless, especially in civilian applications, UAVs equipped with vision systems are suitable for all those cases in which a quick intervention might be necessary while safeguarding the life of human beings [52], e.g., interventions in quarantine zones or critical environments.

Moving on to the military usage of UAVs, the latter are used in different application areas, including mines detection [53], combat efficiency [54,55], battlefield mapping [56], and many others. A relevant context in which the UAVs are utilized regards land monitoring. Technical features of these devices, e.g., high silence and reduced size, make them especially suitable for critical missions [57,58]. In [59], for example, the authors propose a deep learning-based algorithm for tracking objects within a monitoring environment. Their work exploits a Siamese neural network, where features are extracted from both a reference and a probe image and processed through a multi-level prediction module. Using residual feature fusion blocks in the feature extractor and layer attention fusion blocks in the predictor drastically improves the obtained results. Some recent works [60–62] show how to provide monitoring of wide areas by exploiting geo-referenced maps built

by mosaicking techniques. In these cases, flight altitude is a crucial parameter that has to be carefully considered. In fact, tall elements (e.g., towers, high lamp lights, pylons, natural obstacles) acquired from low altitudes can produce occlusions, parallax errors, and other noise that can introduce artifacts during the mosaic construction, which will compromise the application of other subsequent operations, e.g., detection, classification, and localization. Moreover, another challenging parameter for mosaicking applications is the perspective. Improvements in searching for the best geometrical transformation (e.g., Homography vs. similarity) to build the best geo-referenced map at low-altitude is another main target of these surveillance vision systems. Notice that images acquired from low-altitude flights are generally the most interesting in terms of case studies due to their high detail level. Indeed, images acquired from satellite or common air-crafts cannot reach some parts of the monitored areas, and, in some cases, they can have a low resolution. Furthermore, UAVs can be used several times a day and do not require dedicated structures for take-off and landing.

Following the direction shown above, the last years have seen the design of increasingly advanced automatic systems for aerial video surveillance. In the [63], for instance, the authors propose an automatic estimation method to define optimal UAV flight parameters for real-time monitoring of wide areas. The authors of [64], instead, propose a visual cryptography approach to detect hidden targets, thus enabling the design of a new paradigm for the localization and communication of sensitive military objectives. In [65], the authors propose a faster Region-based CNN (R-CNN) for object detection and tracking. Finally, in [66], the authors present a feature-based Simultaneous Localization And Mapping (SLAM) algorithm for small-scale UAVs with nadir view. Although other relevant studies are in progress on the topics just introduced, e.g., object detection in [67], a particular argument, i.e., anomaly detection by UAVs, can be considered relatively new. Indeed, as shown in Section 2, anomaly detection is mainly applied to non-aerial images and presents several open issues such as the definition of anomaly or false positives/negatives during inference. Concluding, the use of UAVs flying at low altitudes to accomplish different military missions is encouraging the development of a new scientific area regarding the design of countermeasures for these vehicles [68]. In addition, recent studies on anomaly detection show that these techniques can be used in critical contexts to detect dangerous events or objects, thus playing a fundamental role in video surveillance.

In this paper, a set of customized textural features based on the original idea of R.M. Haralick [69,70], and a One-Class Support Vector Machine (OC-SVM) [71], are used to design a novel anomaly detection system for aerial video surveillance at low altitudes, i.e., between 6 and 15 m. The use of textural features to catch micro and macro structures of pixel patterns on the ground has been encouraged by both old [72–74] and recent [75,76] literature, where texture analysis is successfully used to detect pixel patterns of different sizes in several application areas. In addition, in this work, we have developed a customized set of textural features inspired by our previous experiences in the use of texture-based classifiers [17,19,77–80]; by defining (i) a discretized circumference-based spatial relationship to build a gray level co-occurrence matrix (GLCM), and (ii) generalized Haralick equations, accounting for this particular displacement. Moreover, using an OC-SVM as a classifier allows us to obtain different key advantages. First, as well-known, this model is lightweight and fast, thus enabling the possible embedding of real-time classifiers even in UAVs with low computational capacity. Second, when adequately trained, an OC-SVM classifier can be highly robust and reliable, achieving high-performance rates, especially in terms of precision. In the anomaly detection context, the latter highlights how many anomalies are correctly detected among those present on the ground during a mission at a low-altitude. Furthermore, this aspect plays a key role in video surveillance of critical environments where each anomaly can represent a dangerous event from an intrusion to an Improvised Explosive Device (IED).

To evaluate the proposed method, extensive experiments using standard classification metrics, such as accuracy, precision, recall, and F₁-score, were performed on the

UAV Mosaicking and Change Detection (UMCD) dataset [81]. A collection comprising 50 challenging aerial video sequences acquired at low altitudes in different environments (i.e., urban, dirt, and countryside) with and without the presence of vehicles, persons, and objects was used. Observe that no other aerial datasets could be used to test the proposed approach. Indeed, to the best of our knowledge, only the UMCD dataset presents the following key characteristics: (i) acquisitions at very low altitudes, i.e., between 6 and 15 m; (ii), different environments with diverse backgrounds; (iii), acquisitions along the same paths with and without anomalies on the ground. For these reasons, although an exhaustive model evaluation is reported in this paper, describing an impressive precision, i.e., 100%, at the expense of a reasonable recall trade-off, i.e., 71.23%, the only feasible and significant literature comparison could be carried out with respect to classical Haralick textural features since this is the first work addressing anomaly detection on the UMCD dataset. Regardless, concerning these baseline features, the proposed method achieves a $\approx 20\%$ gain across all metrics, highlighting the presented approach effectiveness.

Concluding, the main contributions of this paper can be summarized as follows:

- Designing a customized spatial relationship, i.e., a discretized circumference, and generalized Haralick equations to build meaningful GLCMs and textural features;
- Describing an approach that achieves real-time capabilities by exploiting the notoriously lightweight OC-SVM algorithm;
- Presenting quantitative and qualitative experiments of a novel method setting a new baseline for the anomaly detection task on the UMCD dataset.

The rest of this paper is structured as follows. Section 2 offers an overview about anomaly detection based on UAVs for civilian and military applications. Section 3 introduces the customized Haralick textural features by means of a discretized circumference and generalized Haralick equations. Section 4 reports a discussion on quantitative and qualitative results obtained on the UMCD dataset, as well as a comparison with classical Haralick textural features. Finally, Section 5 draws some conclusions on the presented work with some ideas on possible future improvements.

2. Related Work

Anomaly detection is a relatively new research field. It aims at identifying all possible items (e.g., objects, persons, animals) that are not expected to be found in a specific context. This means, for example, that an airplane can be considered a normal item in a specific context, e.g., a hangar of an airport, but not in another context, e.g., the seabed. However, at the same time, a group of airplanes in the sky could be considered an anomaly since it is a not common event, although the context, i.e., the sky, is correct for those items. In other words, anomaly detection aims at monitoring an area of interest, generally through video sequences, without knowing a priori which are all the possible items that can be considered an anomaly. The only knowledge of the system is about the normal condition for the monitored area. The latter is a typical real-life scenario where an automatic system needs to detect possible problems during their monitoring tasks of restricted or dangerous areas. A complete overview of the topic is presented in [82] for interested readers. Moreover, further details concerning methods using novel deep learning approaches in the anomaly detection field can be found in [83], while [84] offers a review on both shallow and deep architectures employed on this task. The following sections report recent state-of-the-art approaches addressing the anomaly detection task in different operative fields such as industrial and private applications as well as surveillance from standard cameras and UAVs.

2.1. Industrial and Private Applications

Industrial applications for anomaly detection are becoming widespread solutions enabling, among others, automatic supply chains management or infrastructural damage prevention. For instance, the proposal in [85] presents a strategy for anomaly detection in the former scenario. The goal is to balance supply and demand to avoid understocking

or overstocking. To this aim, anomaly detection is used to determine unexpected patterns for making more effective decisions. The authors suggest using an Autoencoder Long Short-Term Memory (LSTM) network for forecasting and an OC-SVM classifier for anomaly detection. In this scenario, employing an OC-SVM is very effective since the Autoencoder LSTM network eliminates the multivariate time series dependency and enables to achieve satisfactory performances. Another industrial work is presented in [86], where the authors devise a strategy to detect anomalies in large pipeline infrastructures (e.g., sewers and waterworks) to guarantee their correct functionality and to prevent incidents. Usually, these checks are performed by human operators, implying possible issues in feasibility and causing long processing times. To automate this activity, the proposed approach firstly reconstructs the images retrieved from a hemispherical camera and exploits mosaicking to create the section to analyze. Then, it extracts features by using the well-known SURF algorithm [87] and matches them by Brute Force Matcher (BFM) [88]. Moreover, since water flows can maintain a fixed level over time, the top and bottom of a pipe can result in different textures. For this reason, to help detect structural pipe anomalies, mosaic images are divided into horizontal stripes to analyze portions at the same height and with similar textures. Then, each stripe is further divided into patches, and each patch is processed with the Local Binary Pattern (LBP) [89] algorithm to extract textural features. Finally, the presence of structural anomalies is detected by an OC-SVM classifier.

Moving to the smart home field, the recent proposal in [90] presents a strategy for activity recognition and anomaly detection. The first part relies on the extraction of features from the pre-segmented activities. The authors use a Probabilistic Neural Network (PNN) on the extracted features and then identify anomalies by using a H2O autoencoder [91] applied on the network results. The approach has also been tested on two public datasets (<http://ailab.wsu.edu/casas/datasets>, accessed on 23 October 2021) with promising results. Similarly, the authors of [92] identify anomalies by defining sequences of everyday actions undertaken by house inhabitants. Such sequences are defined via activated IoT items, e.g., smart air purifier, coffee maker, microwave, refrigerator, television, or robot vacuum, among others, interconnected through the house Wi-Fi; and are organized into a tree structure of event sequences. Subsequently, by navigating this structure, the authors can detect anomalous behaviors possibly associated with intruders, both physical or virtual, due to the nature of interconnected smart devices. In particular, in the proposed approach, the authors leverage human habits and use as a root the tree node associated with the first action taken by a user. Then, if all the following actions are present in the sub-tree defined by the chosen root, the action can be considered normal; otherwise, a warning is raised due to an anomalous sequence of events.

2.2. Surveillance

Another relevant field where anomaly detection is fundamental and more correlated to our proposal is surveillance. For instance, the authors of [93] describe a strategy to improve surveillance networks based on RGB cameras. Specifically, the authors implement a pre-trained ResNet-50 and exploit the transfer learning paradigm by fine-tuning the network and have a robust feature extractor. In detail, given a sequence of 15-frames, the extracted features are collected from the last fully connected layer. Then, they are used as input for the proposed Multi-Layer Bi-Directional LSTM network. The choice of the Multi-Layer variation of the LSTM is due to the inability of a single LSTM cell to classify large portions of training data. The Multi-Layer is created by stacking several LSTM cells to learn long-term dependencies. The strategy has been tested on two publicly available datasets, namely the UCF-Crime [94] and the UCFCrime2Local [95], providing improved performance with respect to other state-of-the-art approaches. A second method leveraging deep learning algorithms for anomaly detection in video surveillance is presented in [96]. In particular, the authors introduce an Incremental Spatio-Temporal Learner (ISTL) that enables the detection and localization of anomalies in a video in real-time. In more detail, ISTL is implemented as an active learning method based on fuzzy aggregation. Through

this unsupervised deep learning approach, the authors continuously update the definition of normal frames, thus enabling a dynamic anomaly distinction that can evolve over time.

Differently from the approaches mentioned above, the work in [97] presents a novel strategy for online anomaly detection based on particle filtering [98]. The first contribution of this work is a new way to compute the likelihood of observed events, allowing a better assignment of the particle weights in particle filtering. The second is a novel strategy that aims at identifying anomalous activities based on a posteriori probability. The proposed algorithm splits each frame into cells with variable sizes. The first step consists of the prediction of possible activities occurring in the analyzed frame. Then, the initial predictions are refined by an update step analyzing motion, location, and size features. A clustering algorithm is then applied to separate the different activities. This updating step also evaluates the a posteriori distribution of the activities by using particle filtering to allow a better classification into the two classes: normal and anomalous. The strategy has been tested on two publicly available datasets, namely, UCSD [99], and LIVE [100], showing an overall improvement in performance compared to the other state-of-the-art proposals. An additional work using a clustering technique for anomaly detection is also presented in [101]. Specifically, the authors analyze crowd motions to detect anomalous behaviors through the spatial vicinity of pixels, which enables to describe and understand the dominant motion direction. What is more, since anomaly detection in crowd control is generally computational intensive, the authors exploit a K-means classifier with Univariate Gaussian Discriminant Analysis (KUGDA) to have a hardware-friendly model. Moreover, they also provide a field-programmable gate array (FPGA) implementation, showing that their approach is highly energy-efficient and can outperform methods based on both deep learning and handcrafted features in surveillance applications.

2.3. UAV Surveillance

With the widespread use of drones (generally of the UAV family), especially in the last years, surveillance with UAVs is increasing in its popularity, and it is becoming a hot research field. The proposal in [102] presents a novel end-to-end strategy based on unsupervised generative learning applied on deep one-class classification. The system has two main goals. The former is to guarantee characteristic compactness of normal events (described by optical flow and original images). The latter is to generate optical flow images directly from UAV videos during the testing phase to speed up the detection of anomalous events and satisfy a real-time constraint. Specifically, the proposed network is a deep CNN-based optical flow generator that produces new optical flows from the original images. In this case, however, the authors do not exploit the classical computation of optical flows but, instead, replace it with a convolution/deconvolution-based neural network to speed up the process. At the same time, the network also extracts compact features from both original and optical flow generated images. The architecture is also trained with a custom loss function, computed as the sum of three loss functions, i.e., reconstruction, generation, and compactness losses, to ensure a more efficient classification of events. The architecture has been tested on two publicly available datasets [99,103] and a novel in-house dataset composed of 1000 samples, proving its effectiveness with the collected results.

The work in [104] presents another strategy for anomaly detection through UAVs. The main contribution of the proposal is the comparison of four sets of features. The first set is composed of the deep features extracted from GoogLeNet [105]. The second is made up of local shape information extracted from the regions of each frame by the well-known Histogram of Oriented Gradients (HOG) [106]. The third set is obtained by applying the well-known Principal Component Analysis (PCA) [107] on the HOG features to reduce their space. Finally, the fourth set includes spatio-temporal features extracted with HOG3D [108]. In the case of PCA-HOG and GoogLeNet features, they are normalized using Min-Max normalization and subsequently scaled into the interval [0, 1]. Finally, the anomaly detection step relies on an OC-SVM trained on each feature set, resulting in four classifiers. The work shows that the HOG and the PCA-HOG sets perform significantly

worse compared to the other two. Moreover, the use of PCA barely impacts the accuracy, but it fastens the computation. As for the presented results, the GoogLeNet features seem to be robust enough for obtaining higher accuracy in anomaly detection tasks.

Moving to the anomaly detection by low-altitude UAVs, to the best of our knowledge, there are no other proposals similar to the one shown in this work. The latter aspect is crucial because, as is well-known, detecting anomalies at different altitudes presents very different challenges in terms of visual features, analysis strategies, targets, and so on due to varying object sizes. Moreover, the dataset used in this work, i.e., UMCD dataset [81], is the first that collects low-altitude video acquisitions with different items in different environments. In addition, it is the first dataset studied for change detection missions, i.e., recording over the same path with and without items on the ground; a strategy that can be suitably adapted to verify the anomaly detection task. Finally, it is important to observe that, unlike other anomaly detection works, the proposed one does not use deep learning techniques, thus enabling advantages due to a more straightforward training stage and lower computational aspects.

3. Materials and Methods

In this section, the proposed solution for the anomaly detection in aerial images is described, providing details on the customized Haralick feature extraction step and OC-SVM classifier. An overview of the presented approach is outlined in Figure 1. In more detail, starting from an UAV RGB video stream, the pipeline designed to analyze each frame can be summarized as follows:

1. Patch Generation: the input image is converted to grayscale and split into $n \times m$ patches, where n and m correspond to the width and height of a given patch, respectively. Notice that these smaller sub-regions enable the proposed algorithm to both detect and localize anomalies in the input;
2. Features Extraction: from each generated patch P , a gray level co-occurrence matrix G_P , representing the joint probability distributions of pixel pairs in a given sub-region, is computed using a customized geometric shape, i.e., by selecting pixels on a discretized r -radius circumference. Subsequently, Haralick textural features for patch P are extracted by computing several statistics on G_P ;
3. Anomaly Detection: using Haralick textural features of a given patch P , anomalies are detected exploiting the OC-SVM algorithm. Specifically, this classifier is trained by providing G_P statistics of anomaly-free patches. A hyperplane encompassing this single class is then calculated and used to detect anomalies in new patches.

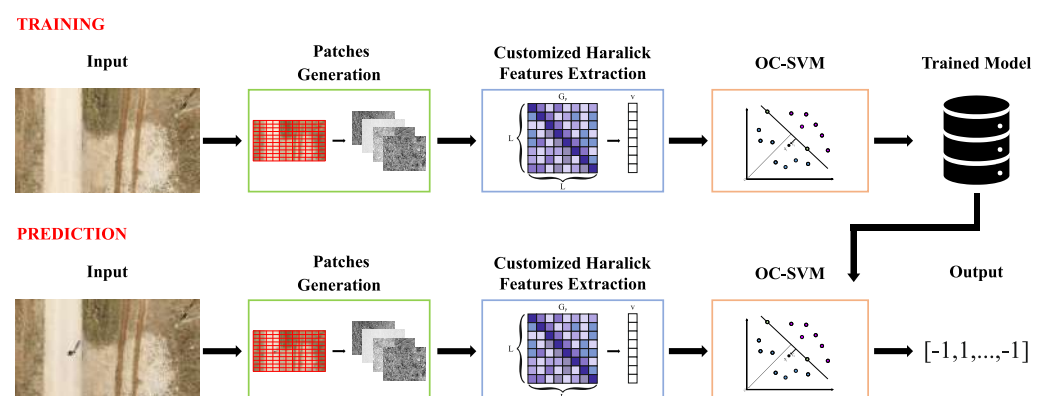


Figure 1. Overview of the proposed architecture. Starting from an RGB image of a video sequence, patches are generated on the grayscale transformed input. Subsequently, customized Haralick textural features based on a discretized circumference are generated and used to train an OC-SVM classifier. Finally, at inference time, the hyperplane found by the OC-SVM is used to detect anomalies in the patches extracted from the analyzed image.

3.1. Customized Haralick Feature Extraction

Haralick features [70] are descriptors defined to capture texture characteristics inside an image. In particular, given a grayscale input image I with shape $w \times h$, Haralick textural features are extracted by computing several statistics on the corresponding gray level co-occurrence matrix (GLCM). The latter is a square matrix with an $L \times L$ size, where L indicates the gray level values in I , and is used, intuitively, to count the co-occurrences of neighboring pixels that satisfy a relation defined by a specific offset. Formally, a GLCM G is computed as follows:

$$G_{\Delta_x, \Delta_y}(i, j) = \sum_{x=1}^w \sum_{y=1}^h \begin{cases} 1, & \text{if } I(x, y) = i \wedge I(x + \Delta_x, y + \Delta_y) = j; \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where Δ_x and Δ_y represent an offset associated to the spacial relation to be satisfied, and can be applied to any pixel in the image; i and j indicate the pixel values being counted; x and y correspond to the coordinates of pixels inside I ; while $I(x, y)$ indicates the pixel value, i.e., gray level, in position (x, y) . From this matrix G , textural features are computed in the form of statistics such as Angular Second Moment, Contrast, Correlation, Sum of Squares: Variance, Inverse Difference Moment, Sum Average, Sum Variance, Sum Entropy, Entropy, Difference Variance, Difference Entropy, Information Measure of Correlation 1, Information Measure of Correlation 2, Maximum Correlation Coefficient. The mathematical formulations for all of these features are fully described in the original paper [70].

A fundamental aspect of classical Haralick textural features lies in the parametrization of spatial relations during the GLCM construction. Specifically, a relationship between two neighboring pixels can be defined through a distance d and an axis orientation $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$, corresponding, respectively, to neighbors along a horizontal, right diagonal, vertical, and left diagonal axis, with respect to a central pixel, as depicted in Figure 2. Observe that, since a distinct co-occurrence matrix is generated for pixels along a single orientation axis, these features are not rotation invariant. Differently from this classic approach, in this work we design a spatial relation based on a discretized circumference with radius r . Through this pixel pattern, it is possible to build a single GLCM that (i) simultaneously accounts for satisfied relationships along all directions; (ii) achieves invariance with respect to rotations; (iii) examines more gray level co-occurrences by analyzing several neighbors for each pixel, therefore capturing textural characteristics of extended areas such as, for instance, uniform color distributions in objects. In more detail, starting from a grayscale patch P , a sliding window SW with size 1×1 is used to apply the desired pattern to all pixels. Specifically, for a given center pixel c , selected via the SW , the GLCM is computed by feeding neighboring pixels located along the r circumference to Equation (1). For instance, assuming a circumference pattern with radius $r = 2$, shown in Figure 3c, the spatial relations (Δ_x, Δ_y) with respect to c , would be defined via the displacement set $D = \{(0, \pm 2), (\pm 1, \pm 2), (\pm 2, \pm 1), (\pm 2, 0)\}$. Visual examples of neighboring pixels displacements along discretized circumferences with radius $r = 2, 3, 4, 5$, are shown in Figure 3.

Similarly to classical Haralick, the last step employed to extract features from the GLCM requires the computation of several statistics. However, in this work, starting from [70], we implement equations with generalized parameters that are adapted to account for the proposed textural pattern, namely, N-Order Momentum, N-Order Central Moment, Homogeneity, Contrast, Inverse Difference, Entropy, Correlation, and Difference Entropy. After the generation of these textural features, the eight resulting values associated to statistical aspects of the input patch P are concatenated into a single vector v , which is finally classified via the OC-SVM to detect possible anomalies inside P . Following, a brief introduction comprising equations complete with constraints for each generalized measure utilized as a textural feature. Such equations, despite similar to the originals presented by Haralick, are the formalization of the proposed modified features.

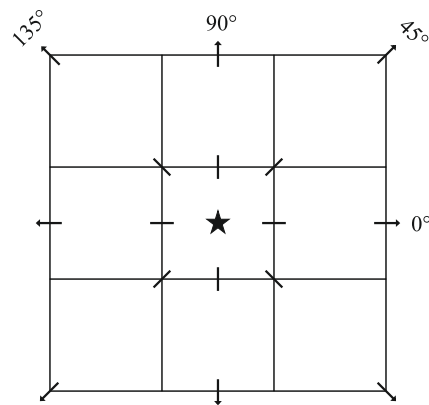


Figure 2. Classical spatial relationship to compute a GLCM for Haralick textural features using a 3×3 window. A relationship is defined through a single displacement using a distance d along one of the shown axes orientations, i.e., $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$, for the horizontal, right diagonal, vertical, and left diagonal neighbors.

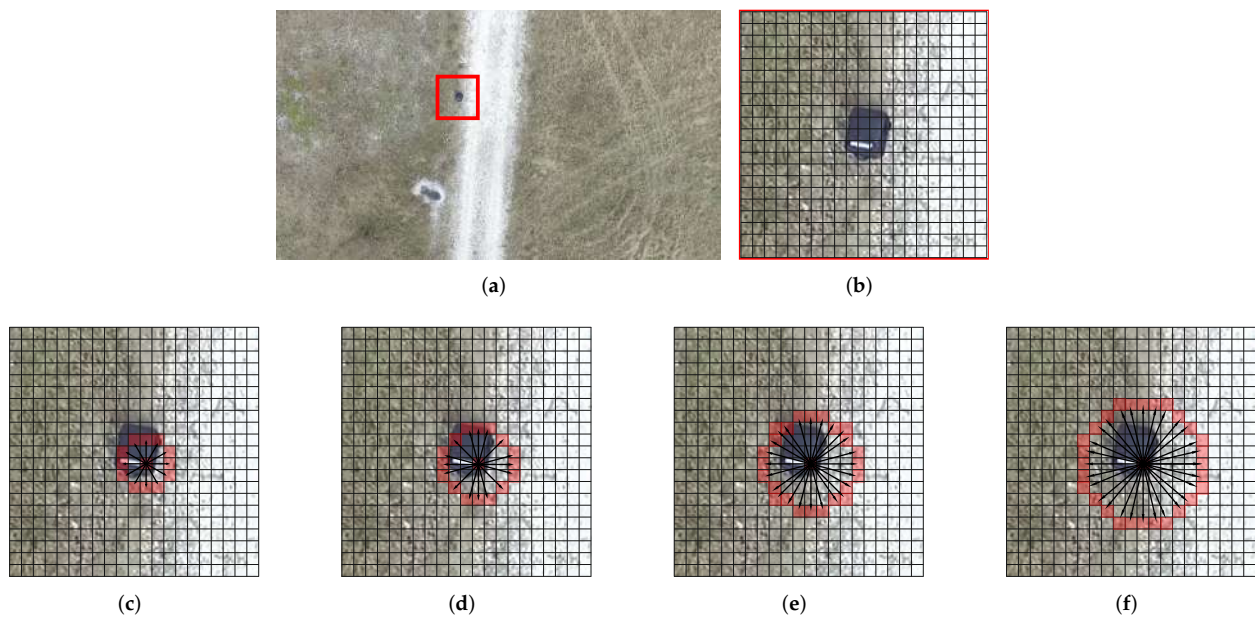


Figure 3. Proposed spatial relation of pixels used to build the GLCM. A sample frame containing an anomaly and a pixel level zoom are reported in (a,b). Examples of circumferences relations for $r = 2, 3, 4, 5$ are shown in (c–f), respectively. For each pixel in the patch P , discretized circles are built using it as a center. (a) Input image showing an anomaly, highlighted by the red bounding box. (b) Anomaly at the pixel level. (c) Circumference with $r = 2$. (d) Circumference with $r = 3$. (e) Circumference with $r = 4$. (f) Circumference with $r = 5$.

3.1.1. N-Order Momentum and N-Order Central Moment

Based on first order statistics, these operators evaluate the amount of information of pixels analyzed by the sliding window SW , i.e., neighboring pixels located along a given circumference. Intuitively, the N-Order Momentum M_{n_1} computes the average gray level for the SW , while the N-Order Central Moment C_{n_2} is concerned with the amplitude dispersion of pixels contained inside the SW with respect to their average, i.e., M_{n_1} . Formally, these two measures are computed as follows:

$$M_{n_1} = \sum_{i=0}^{L-1} i^{n_1} \cdot p(i), \quad (2)$$

$$C_{n_2} = \sum_{i=0}^{L-1} (i - M_{n_1})^{n_2} \cdot p(i), \quad (3)$$

which are constrained to:

$$\forall i \in [0, \dots, L-1] \subset \mathbb{N}, 0 \leq p(i) \leq 1, \sum_{i=0}^{L-1} p(i) = 1, n_1, n_2 \in \mathbb{N}, \quad (4)$$

where $p(i)$ represents the gray level probability $i \in [0, \dots, L-1]$; L corresponds to the gray level values in the input patch; while n_1 and n_2 indicate the N -th order of M_{n_1} and C_{n_2} , respectively.

3.1.2. Homogeneity and Contrast

Associated to second order statistics, these two operators can account for the circumference size r to analyze textural variations in a given neighborhood. In particular, Homogeneity $HG(r)_{n_3}$ examines the uniformity degree of a given area to detect textural structures; while Contrast $CT(r)_{n_4, n_5}$ can detect gray level variations in a zone by showing a high response in case of areas with extremely diverse intensities. Formally, they are defined as:

$$HG(r)_{n_3} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} [p_r(i, j)]^{n_3}, \quad (5)$$

$$CT(r)_{n_4, n_5} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} |i - j|^{n_4} \cdot [p_r(i, j)]^{n_5}, \quad (6)$$

which are constrained to:

$$\forall (i, j) \in [0, \dots, L-1] \times [0, \dots, L-1] \subset \mathbb{N}^2; 0 \leq p_r(i, j) \leq 1, \quad (7)$$

$$\sum_{i=0}^{L-1} p_r(i, j) = 1; n_3, n_4, n_5 \in \mathbb{N}, \quad (8)$$

where $p_r(i, j)$ represents the probability that two pixels with distance r , i.e., on the circumference, have $i, j \in [0, \dots, L-1]$ gray level values; L indicates the gray level values in the input patch; while n_3, n_4, n_5 correspond to the order of $HG(r)_{n_3}$ and $CT(r)_{n_4, n_5}$, respectively.

3.1.3. Inverse Difference and Entropy

Belonging to second order statistics, these operators can identify possible pattern sizes inside a patch. Specifically, the Inverse Difference $ID(r)_{n_6, n_7}$ observes local pixel distributions and can detect specific configurations or repetitions; while the Entropy $ET(r)_{n_8, n_9}$ captures the disorder degree inside patch P since its value will increase proportionally to the detected gray level randomness. Formally, these measure are calculated via:

$$ID(r)_{n_6, n_7} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{[p_r(i, j)]^{n_6}}{1 + (i, j)^{n_7}}, \quad (9)$$

$$ET(r)_{n_8, n_9} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} [p_r(i, j)]^{n_8} \cdot [\log_{k_1}(p_r(i, j))]^{n_9}, \quad (10)$$

where, again, $p_r(i, j)$ represents the probability that two pixels with distance r , i.e., on the circumference, have $i, j \in [0, \dots, L-1]$ gray level values; L indicates the gray level values in the input patch; while $n_6, n_7, n_8, n_9, k_1 \in \mathbb{N}$ correspond to the generalized $ID(r)_{n_6, n_7}$ and $ET(r)_{n_8, n_9}$ parameters.

3.1.4. Correlation and Difference Entropy

The last two operators have been defined to increase the generated textural features detail and reliability by recognizing possible relationships between near circumferences, i.e., associated to close proximity SWs. In particular, the Correlation $CR(r)_{n_{10},n_{11}}$ identifies spatial constraints for circumferences presenting similar patterns; while the Difference Entropy $DE(r)_{n_{12},n_{13}}$ is actually concerned with representing spatial constraints of dissimilar patterns. Formally, these measures are computed as follows:

$$CR(r)_{n_{10},n_{11}} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{(i - \mu_x) \cdot (j - \mu_y) \cdot [p_r(i, j)]^{n_{10}}}{[\sigma_x \cdot \sigma_y]_{11}^{n_{10}}}, \quad (11)$$

$$DE(r)_{n_{12},n_{13}} = - \sum_{i=0}^{L-1} [p_{x-y}(i)]^{n_{12}} \cdot [\log_{k_2}(p_{x-y}(i))]^{n_{13}}, \quad (12)$$

which are constrained to Equations (7) and (8) and where $n_{10}, n_{11}, n_{12}, n_{13}, k_2 \in \mathbb{N}$ represent the parameters of the generalized $CR(r)_{n_{10},n_{11}}$ and $DE(r)_{n_{12},n_{13}}$. Finally, the mean, standard deviation, and gray level pixel probability presented in Equations (11) and (12) are computed via the following equations:

$$\mu_x = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} [i \cdot p_r(i, j)]; \sigma_x = \sqrt{\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} [(i - \mu_x)^2 \cdot p_r(i, j)]}, \quad (13)$$

$$\mu_y = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} [j \cdot p_r(i, j)]; \sigma_y = \sqrt{\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} [(j - \mu_y)^2 \cdot p_r(i, j)]}, \quad (14)$$

$$p_{x-y}(k) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} [p_r(i, j)]^q; \text{ where } |i - j| = k. \quad (15)$$

3.2. OC-SVM Classifier

To classify the customized Haralick textural features contained in vector v , we employ an OC-SVM since it is an effective method for anomaly detection [109,110], where rare occurrences, i.e., anomalies, need to be identified. Intuitively, to achieve this objective, the OC-SVM finds a hyperplane describing a target class represented in the input dataset. This hyperplane has a maximum margin separation with respect to the origin, which is used as an outlier with respect to the target class, and allows to detect anomalies when a new input instance is given to the OC-SVM. Formally, starting from a collection describing the target class $\{x_i\}_{i=1}^s$, where s corresponds to the number of samples in the dataset, the hyperplane is computed as follows:

$$\min_{\omega, \rho, \xi} \left\{ \frac{1}{2} \|\omega\|^2 - \rho + \frac{1}{\nu s} \sum_i \xi_i \right\} \quad \forall i = 1, \dots, s, \quad (16)$$

which is constrained to:

$$\langle \omega, \phi(x_i) \rangle \geq \rho - \xi_i, \quad (17)$$

where ω is a vector perpendicular to the hyperplane, which is computed in a high-dimensional Hilbert feature space \mathcal{H} ; ρ indicates the hyperplane distance from the origin; $\xi_i \geq 0$ corresponds to a set of slack variables used to handle possible outliers in the input data; $\nu \in (0, 1]$ is a trade-off parameter that manages the number of samples mapped as positive in the training set, according to the decision function $f(x) = \text{sgn}(\langle \omega, \phi(x_i) \rangle - \rho)$; while $\phi(\cdot)$ identifies a non-linear transformation to map dataset samples into the \mathcal{H} space, exploited to manage non-linear problems. Moreover, through Lagrange multipliers α_i , Equation (17) constraints can be applied to Equation (16). Then, by deriving with respect to

ω , the primal and dual weights relation can be defined via a linear combination of samples mapped into \mathcal{H} with $\alpha_i \neq 0$, i.e., $\omega = \sum_i \alpha_i \phi(x_i)$. In addition, it is possible to avoid computing the non-linear mapping $\phi(\cdot)$ directly by defining a kernel $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$, through which the dual problem is defined as follows:

$$\min_{\alpha} \left\{ \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j) \right\}, \quad (18)$$

which is constrained to:

$$0 \leq \alpha_i \leq \frac{1}{\nu s} \quad \wedge \quad \sum_i \alpha_i = 1. \quad (19)$$

Finally after solving the dual problem shown in Equation (18) and obtaining model weights α_i , they can be used to classify new input instances and detect anomalies through the following decision function:

$$f(x_*) = \text{sgn} \left(\sum_i \alpha_i K(x_i, x_*) - \rho \right), \quad (20)$$

where x_* corresponds to any given test vector containing the customized Haralick textural features described in Section 3.1.

4. Experimental Results and Discussion

In this section, the experimental results for the OC-SVM anomaly detection are presented. In particular, the dataset utilized to evaluate the system is first introduced. Implementation details are then provided, and a discussion on the obtained results highlighting strong points as well as system limitations is finally reported.

4.1. Dataset

The collection used to evaluate the proposed anomaly detection system is the UMCD dataset (<http://www.umcd-dataset.net/>, accessed on 23 October 2021) [81], which was chosen for several reasons. First, to the best of our knowledge, this is the only dataset that contains low-altitude UAV videos, with a flight altitude lower than 15 m. At this height, it is possible to exploit the full camera resolution when trying to detect anomalies, contrary, for example, to data derived from higher altitudes or even from satellites, where it is much more challenging to identify fine-grained details inside an image due to the device distance from the ground. Second, this dataset, originally collected to address mosaicking and change detection tasks, contains several video sequences of given areas with and without anomalies to simulate real-life scenarios where unexpected objects might appear in the scene.

Concerning the samples themselves, the UCMD dataset contains 50 aerial video sequences collected in different environments, i.e., urban, dirt, and countryside. The videos were recorded using two small-scale UAVs, namely a DJI Phantom 3 Advanced with its built-in camera, and a custom home-made hexacopter employing cameras with different spatial resolutions, ranging from 720×540 (4:3, Standard Definition) to 1920×1080 (16:9, High Definition) pixels per frame. Moreover, all videos were acquired at a low-altitude, ranging from 6 to 15 m, with speeds ranging from 2 up to 12 m/s, at different daily times, i.e., morning and afternoon, and include metadata in the form of telemetry data such as flight height, speed, and GPS coordinates. From these 50 videos, only 20 were devised for change detection and therefore selected for this work. In particular, this selection contains pairs of videos in the three environments mentioned above, where each pair shows a given path with or without anomalies, as can be observed in Figure 4. In particular, different kinds of anomalies are identified through various objects properties, including size, shape, and color, i.e., tire, gas bottle, person, car, small box, big box, metal suitcase, suitcase, and bag. What is more, multiple anomalies can appear in a given video or frame of a

recording. Finally, notice that although the remaining 30 videos could be employed to train the OC-SVM, we purposely excluded them to avoid presenting unbalanced data since first, those recordings do not present anomalies, and second, they are exclusively focused on dirt and countryside backgrounds.



Figure 4. Image samples for different environments in the UMCD Dataset. In the first row, images associated to the normal state. In the second row, the same images presenting anomalies, highlighted in the red bounding box.

4.2. Implementation Details

All experiments were performed on the 20 video pairs devised for change detection. In particular, the 10 sequences without anomalies were employed to train the OC-SVM to recognize normal patches of urban, dirt, and countryside environments. The remaining 10 recordings containing anomalies were instead used to evaluate the model. Notice that since the OC-SVM classifies each patch generated from a given frame of a video stream, the sequences used to evaluate the model contain both normal and anomalous patches, enabling to test the system correctly.

Regarding the model configuration, the OC-SVM was trained using the customized Haralick textural features presented in Section 3.1, which are computed for each patch extracted from the input image. Notice that the patch size can be changed, producing a GLCM covering different input areas; therefore, the optimal patch size was found empirically, as shown in the evaluation section. Moreover, for the OC-SVM we employed the following parameters: *kernel* = radial basis function, *stop_tolerance* = 0.001, *nu* = 0.5.

Concerning the evaluation, common classification metrics were employed to determine the anomaly detection effectiveness of the system, i.e., *Accuracy*, *Precision*, *Recall*, and *F₁-score*. These metrics are computed analyzing all frames in a video as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}; \quad (21)$$

$$Precision = \frac{TP}{TP + FP}; \quad (22)$$

$$Recall = \frac{TP}{TP + FN}; \quad (23)$$

$$F_1\text{-score} = 2 * \frac{Precision * Recall}{Precision + Recall}; \quad (24)$$

where *TP* and *TN*, i.e., true positive and true negative, indicate normal and anomalous patches correctly classified as such; *FP*, i.e., false positive, identifies anomalous patches classified as normal; while *FN*, i.e., false negative, corresponds to normal patches that are classified as anomalous. Observe that, from these definitions, it is clear that *Precision*

enables to evaluate the ability of the system to detect all anomalies, while *Recall* allows to assess the model resistance to raise false alarms in normal situations.

Lastly, the system was implemented using the Python language as well as OpenCV and Scikit-learn libraries to create the custom Haralick textural features and OC-SVM model. Moreover, the experiments, which achieved real-time capabilities, were performed using an AMD Ryzen 7 CPU, 16 GB DDR4 RAM, and an NVidia RTX 2080 GPU. Notice, however, that the proposed method is extremely lightweight and, once trained, can be embedded directly in UAVs with low computational capacity without losing its real-time execution.

4.3. Performance Evaluation

Extensive experiments were performed to evaluate the proposed system. In particular, we explored how the selection of patch P size, and circumference radius r , affect the model performances in relation to the customized Haralick textural feature generation. Moreover, since no other work performs anomaly detection on this dataset, we compare the model behavior to classical Haralick spatial relations, where the latter can be considered as a baseline for the UCMD dataset anomaly detection via textural analysis.

Concerning the patch P size effects, performances are reported in Table 1. As shown, using medium-sized patches, i.e., 80×100 , results in the highest scores for this specific dataset, an outcome attributable to the co-occurrence matrix generation. As a matter of fact, a GLCM accounts for co-occurring gray levels inside the input patch; therefore, the resulting matrix is highly affected by the covered area dimensions. In detail, when smaller areas are selected, i.e., patches 40×50 and 50×75 , gray level values might be too close, thus resulting in the computation of similar statistics for different patches through the customized Haralick features presented in Section 3.1. Contrarily, when choosing wide areas, i.e., patches 120×150 and 160×200 , there might be a pixel distribution inside the GLCM that is too heterogeneous due to the amount of textural information. While this could also be beneficial for noisy environments such as those present in the dataset, uninformative statistics might be generated and, consequently, the system suffers a massive performance degradation as reported in Table 1.

Table 1. Performance evaluation on the UMCD dataset for Patch P size. All rows generate a co-occurrence matrix using a circumference radius $r = 3$.

Patch Size	Accuracy	Precision	Recall	F ₁ -Score
40×50	42.75%	100.00%	42.07%	59.22%
50×75	56.08%	100.00%	53.73%	69.94%
80×100	72.15%	100.00%	71.23%	83.19%
120×150	24.65%	65.37%	19.48%	30.01%
160×200	4.12%	33.05%	1.74%	3.30%

On a different note, small to medium patches can achieve a 100% precision on the UMCD dataset, meaning that for such sizes, the extracted textural features enable the system to never miss an anomaly inside an image. This outcome is of particular interest for security applications; however, it comes with a significant trade-off on the recall metric, which describes how many false negatives, i.e., normal patches classified as anomalous, there are at inference time. In fact, as can be seen in Table 1, even for the best performing patch size, i.e., 80×100 , recall only reaches a 71.23% score, indicating that there are many misclassifications. While raising false alarms might be a sub-optimal solution, for surveillance scenarios it is of paramount importance to never miss anomalies to avoid dangerous situations, e.g., a person violating a restricted area perimeter. To this end, wider patches result in being completely unusable since they can also miss anomalies, while smaller ones lead to too many normal patches being classified as anomalous. To have a better intuition of this behavior, classifications for the same input image using different patch sizes are reported in Figure 5. As shown, small and medium patches, i.e., Figure 5b,c, do not miss anomalies in the input. Nevertheless, they present many false positives,

especially on vehicle tracks and grass where there is a substantial contrast with the terrain texture, for the former, and illumination variation causing darker areas, for the latter. Contrarily, employing a wider patch, i.e., Figure 5d, causes classification errors for both normal and anomalous patches since uninformative statistics are extracted from the GLCM, resulting in a high number of false positives and negatives. Due to this issue, smaller patches are preferable for surveillance scenarios where anomalies must always be detected. Thus, in this work, the medium-sized patch 80×100 was selected as the optimal dimension for the UMCD dataset since it maximizes the system precision while showing a smaller recall trade-off with respect to the other sizes.

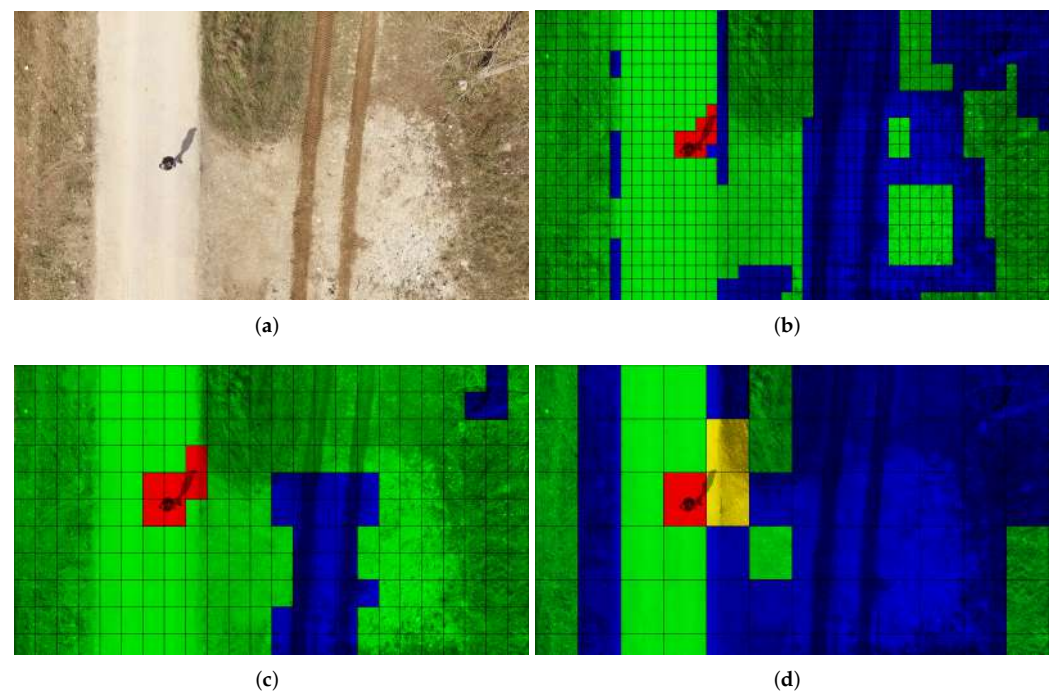


Figure 5. Output generated using different patch sizes. The red and green patches indicate, respectively, anomalous and normal patches correctly classified. Blue patches correspond to normal patches classified as anomalies. Yellow patches represent anomalous patches classified as normal. (a) Input image with an anomaly, i.e., a man and his shadow. (b) Anomaly detection output using a 40×50 patch size. (c) Anomaly detection output using an 80×100 patch size. (d) Anomaly detection output using a 160×200 patch size.

Regarding the circumference radius r size impact, performances are outlined in Table 2. As can be observed, the various dimensions show similar behaviors since they can detect roughly the same amount of information from a given patch, indicating that the proposed spatial relationship is effective on the UMCD dataset. Regardless, by increasing the radius, it is possible to notice that performances start to degrade across all metrics, with a particular emphasis on the recall measure. This outcome has a similar rationale to the patch size effects, where wider areas resulted in the extraction of uninformative statistics. Specifically, by increasing the radius size, fine-grained textural details and patterns associated with smaller objects might be missed altogether, even though a higher number of pixels is analyzed in the circumference-defined neighborhood. Therefore, normal patches presenting specific patterns, e.g., tracks shown in Figure 5a, would be considered anomalous at inference time. Such an effect can be easily explained by observing Figure 3, where the anomaly, i.e., a suitcase, is entirely skipped by the biggest radius size $r = 5$ employed in this work. Indeed, this issue highlights how the spatial relationship defined to build a GLCM is relevant to the final system performances since it directly affects the Haralick textural features extraction. Moreover, since the UMCD dataset is specifically designed

to contain aerial images, objects on the scene appear smaller than their real dimension, further supporting the improved performances for smaller radiuses such as $r = 2$, and $r = 3$, reported in Table 2.

Table 2. Performance evaluation on the UMCD dataset for Radius r size. All rows are computed using a Patch size $P = 80 \times 100$.

Radius r	Accuracy	Precision	Recall	F ₁ -Score
2	72.08%	100.00%	70.14%	82.44%
3	72.15%	100.00%	71.23%	83.19%
4	70.58%	99.85%	68.77%	81.44%
5	69.23%	99.23%	66.10%	79.34%

Finally, regarding the classical Haralick spatial relationships comparison, the obtained results are reported in Table 3. Notice that the chosen spatial offsets for classical Haralick textural features were taken along the circumference of radius $r = 3$ to better appreciate the effectiveness of using an entire circumference-defined neighborhood instead of a single pixel. Moreover, independently from the selected classical relationship, performances remain in line with the shown cases. Furthermore, the same statistics presented in Section 3.1 were employed for all reported relationships to ensure a fair comparison between the various extracted features. Concerning the results summarized in Table 3, the proposed relationship, based on a circumference, significantly outperforms classical ones, based on single offsets, on all metrics. This outcome indicates that classical relationships consistently struggle on the UMCD dataset, most likely due to the single spatial displacements being too simple to capture complex textural patterns inside the chosen noisy environments, i.e., urban, dirt, and countryside. Indeed, since in classical Haralick features the GLCM is computed by analyzing gray level co-occurrences of pixels with a single displacement, information from the input patch is inevitably lost, and anomalies might not be recognized due, for example, to missing rotation invariance of such features. Contrarily, the proposed circumference pattern can account for such rotations by analyzing an entire neighborhood along all directions with respect to a central pixel, as explained in Section 3.1, thus capturing as much information as possible from a given patch. As a matter of fact, the obtained performances demonstrate that the proposed spatial relationship, i.e., discretized circumference, is effective on the UMCD dataset, especially for security applications, since it has a perfect precision with a moderately low recall trade-off. To conclude, some final remarks on the execution time are reported. As already shown, the size of the patches strongly influences the accuracy of the model. This is not true for the execution time. In fact, the OC-SVM is able to predict the class of all the patches belonging to an image in just 0.01 seconds, independently from the patch size, allowing the use of the model for real-time applications.

Table 3. Performance evaluation on the UMCD dataset comparing classical Haralick textural features with single displacement to the proposed discretized circumference.

GLCM Spatial Relationship	Accuracy	Precision	Recall	F ₁ -Score
Single Offset $(\Delta_x, \Delta_y) = (0, 3)$	35.24%	87.42%	32.17%	47.03%
Single Offset $(\Delta_x, \Delta_y) = (-2, -2)$	36.05%	88.09%	32.27%	47.23%
Single Offset $(\Delta_x, \Delta_y) = (3, 0)$	33.98%	86.80%	31.67%	46.40%
Single Offset $(\Delta_x, \Delta_y) = (2, -2)$	34.72%	87.13%	32.05%	48.85%
Circumference Radius $r = 3$	72.15%	100.00%	71.23%	83.19%

5. Conclusions

This work presented a novel lightweight method with real-time capabilities for anomaly detection based on textural features and One-Class SVM in low-altitude aerial images. In particular, input frames are turned to grayscale and split into patches to extract GLCMs and, subsequently, textural statistics representing these patches that can be exploited to detect anomalies in the input via an OC-SVM. In detail, starting from classical Haralick textural features based on single offset displacements, we designed a new spatial relationship in the form of a discretized circumference. The latter simultaneously accounts for displacements along all directions by analyzing pixels in the circumference-defined neighborhood, improving the patch representation, and guaranteeing a rotation invariance property intrinsically via the chosen pattern. Moreover, generalized equations to compute Haralick textural features were also presented to handle the proposed spatial relationship correctly and ultimately extract meaningful patch characteristics. Experiments were performed on the public UMCD dataset to evaluate the system by assessing different patch sizes and circumferences with varying radiuses. Furthermore, to the best of our knowledge, there are currently no other methods addressing the anomaly detection via textures on this dataset; therefore, a comparison with baseline Haralick textural features employing single displacements was also provided to highlight the proposed method effectiveness. As reported in the experimental section, patch size and circumference radius are key components to achieve satisfactory performances for security applications, where anomalies must always be captured. Specifically, the system can reach a 100% precision at the expense of a reasonable recall trade-off, which obtained up to a 71.23% score. In addition, the chosen spatial relation and statistics significantly outperformed classical Haralick textural features based on single offset displacements, demonstrating the presented approach effectiveness.

Although encouraging performances were obtained on the UMCD dataset, enabling the proposed method to be considered a baseline on this collection for the anomaly detection task, there are still margins of improvement. Specifically, as per the obtained recall, the model currently produces several false negatives, i.e., it identifies normal patches as anomalous. Thus, as future work, we plan to introduce other textural statistics to extend a patch description and different spatial relationships to be used in conjunction with the presented circumference. Moreover, we are also considering exploring deep learning solutions to further improve the system performances while still trying to retain a lightweight model that can be implemented, in the future, directly by UAVs with low computational capabilities.

Author Contributions: Conceptualization, D.A. and D.P.; methodology, D.A., A.F. and D.P.; software, A.D.M., A.D. and D.P.; validation, M.R.M. and A.M.; writing—original draft preparation, D.A., A.F., M.R.M. and D.P.; writing—review and editing, D.A., L.C., A.F., G.L.F. and D.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This work was supported by the MIUR under grant “Departments of Excellence 2018–2022” of the Sapienza University Computer Science Department and the ERC Starting Grant no. 802554 (SPECGEO).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yang, M.H.; Huang, C.R.; Liu, W.C.; Lin, S.Z.; Chuang, K.T. Binary Descriptor Based Nonparametric Background Modeling for Foreground Extraction by Using Detection Theory. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *25*, 595–608. [\[CrossRef\]](#)
2. Avola, D.; Cinque, L.; Foresti, G.L.; Massaroni, C.; Pannone, D. A keypoint-based method for background modeling and foreground detection using a PTZ camera. *Pattern Recognit. Lett.* **2017**, *96*, 96–105. [\[CrossRef\]](#)
3. Javed, S.; Mahmood, A.; Bouwmans, T.; Jung, S.K. Background–Foreground Modeling Based on Spatiotemporal Sparse Subspace Clustering. *IEEE Trans. Image Process.* **2017**, *26*, 5840–5854. [\[CrossRef\]](#)
4. Avola, D.; Bernardi, M.; Cinque, L.; Foresti, G.L.; Massaroni, C. Adaptive bootstrapping management by keypoint clustering for background initialization. *Pattern Recognit. Lett.* **2017**, *100*, 110–116. [\[CrossRef\]](#)

5. Mohanty, S.K.; Rup, S.; Swamy, M. An improved scheme for multifeature-based foreground detection using challenging conditions. *Digit. Signal Process.* **2021**, *113*, 103030. [[CrossRef](#)]
6. Bakheet, S.; Al-Hamadi, A. A Discriminative Framework for Action Recognition Using f-HOL Features. *Information* **2016**, *7*, 68. [[CrossRef](#)]
7. Avola, D.; Bernardi, M.; Foresti, G.L. Fusing Depth and Colour Information for Human Action Recognition. *Multimed. Tools Appl.* **2019**, *78*, 5919–5939. [[CrossRef](#)]
8. Tu, Z.; Li, H.; Zhang, D.; Dauwels, J.; Li, B.; Yuan, J. Action-Stage Emphasized Spatiotemporal VLAD for Video Action Recognition. *IEEE Trans. Image Process.* **2019**, *28*, 2799–2812. [[CrossRef](#)] [[PubMed](#)]
9. Avola, D.; Cascio, M.; Cinque, L.; Foresti, G.L.; Massaroni, C.; Rodolà, E. 2-D Skeleton-Based Action Recognition via Two-Branch Stacked LSTM-RNNs. *IEEE Trans. Multimed.* **2020**, *22*, 2481–2496. [[CrossRef](#)]
10. Du, S.; Chen, S. Salient Object Detection via Random Forest. *IEEE Signal Process. Lett.* **2014**, *21*, 51–54. [[CrossRef](#)]
11. Avola, D.; Cinque, L.; Foresti, G.L.; Mercuri, C.; Pannone, D. A Practical Framework for the Development of Augmented Reality Applications by Using ARUCO Markers. In Proceedings of the International Conference on Pattern Recognition Applications and Methods (ICPRAM), Rome, Italy, 24–26 February 2016; pp. 645–654.
12. Hu, Q.; Paisitkriangkrai, S.; Shen, C.; van den Hengel, A.; Porikli, F. Fast Detection of Multiple Objects in Traffic Scenes with a Common Detection Framework. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 1002–1014. [[CrossRef](#)]
13. Avola, D.; Foresti, G.L.; Cinque, L.; Massaroni, C.; Vitale, G.; Lombardi, L. A multipurpose autonomous robot for target recognition in unknown environments. In Proceedings of the IEEE International Conference on Industrial Informatics (INDIN), Poitiers, France, 19–21 July 2016; pp. 766–771.
14. Cao, X.; Yang, L.; Guo, X. Total Variation Regularized RPCA for Irregularly Moving Object Detection Under Dynamic Background. *IEEE Trans. Cybern.* **2016**, *46*, 1014–1027. [[CrossRef](#)] [[PubMed](#)]
15. Avola, D.; Cinque, L.; Foresti, G.L.; Marini, M.R.; Pannone, D. A Rover-based System for Searching Encrypted Targets in Unknown Environments. In Proceedings of the International Conference on Pattern Recognition Applications and Methods (ICPRAM), Funchal, Madeira, 16–18 January 2018; pp. 254–261.
16. Lei, S.; Zhang, B.; Wang, Y.; Dong, B.; Li, X.; Xiao, F. Object recognition using non-negative matrix factorization with sparseness constraint and neural network. *Information* **2019**, *10*, 37. [[CrossRef](#)]
17. Avola, D.; Cinque, L.; Di Girolamo, M. A novel T-CAD framework to support medical image analysis and reconstruction. In Proceedings of the International Conference on Image Analysis and Processing, Ravenna, Italy, 14–16 September 2011; pp. 414–423.
18. Islam, A.; Reza, S.M.S.; Iftekharuddin, K.M. Multifractal Texture Estimation for Detection and Segmentation of Brain Tumors. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 3204–3215. [[CrossRef](#)] [[PubMed](#)]
19. Avola, D.; Cinque, L.; Placidi, G. Customized first and second order statistics based operators to support advanced texture analysis of MRI images. *Comput. Math. Methods Med.* **2013**, *2013*, 1–13. [[CrossRef](#)] [[PubMed](#)]
20. Song, W. A New Method for Refined Recognition for Heart Disease Diagnosis Based on Deep Learning. *Information* **2020**, *11*, 556. [[CrossRef](#)]
21. Avola, D.; Cinque, L.; Fagioli, A.; Foresti, G.; Mecca, A. Ultrasound Medical Imaging Techniques: A Survey. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–38. [[CrossRef](#)]
22. Yu, B.; Zhou, L.; Wang, L.; Yang, W.; Yang, M.; Bourgeat, P.; Fripp, J. SA-LuT-Nets: Learning Sample-Adaptive Intensity Lookup Tables for Brain Tumor Segmentation. *IEEE Trans. Med. Imaging* **2021**, *40*, 1417–1427. [[CrossRef](#)] [[PubMed](#)]
23. Avola, D.; Cinque, L.; Fagioli, A.; Filetti, S.; Grani, G.; Rodolà, E. Multimodal Feature Fusion and Knowledge-Driven Learning via Experts Consult for Thyroid Nodule Classification. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *1*. [[CrossRef](#)]
24. Mumenthaler, C.; Sander, D.; Manstead, A.S. Emotion recognition in simulated social interactions. *IEEE Trans. Affect. Comput.* **2018**, *11*, 308–312. [[CrossRef](#)]
25. Avola, D.; Cinque, L.; Fagioli, A.; Foresti, G.L.; Massaroni, C. Deep temporal analysis for non-acted body affect recognition. *IEEE Trans. Affect. Comput.* **2020**, *1*. [[CrossRef](#)]
26. Wu, Z.; Singh, B.; Davis, L.; Subrahmanian, V. Deception detection in videos. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), New Orleans, LA, USA, 2–7 February 2018; pp. 1–8.
27. Avola, D.; Cascio, M.; Cinque, L.; Fagioli, A.; Foresti, G.L. LieToMe: An Ensemble Approach for Deception Detection from Facial Cues. *Int. J. Neural Syst.* **2021**, *31*, 2050068. [[CrossRef](#)] [[PubMed](#)]
28. Khan, W.; Crockett, K.; O’Shea, J.; Hussain, A.; Khan, B.M. Deception in the eyes of deceiver: A computer vision and machine learning based automated deception detection. *Expert Syst. Appl.* **2021**, *169*, 114341. [[CrossRef](#)]
29. Avola, D.; Cinque, L.; De Marsico, M.; Fagioli, A.; Foresti, G.L. LieToMe: Preliminary study on hand gestures for deception detection via Fisher-LSTM. *Pattern Recognit. Lett.* **2020**, *138*, 455–461. [[CrossRef](#)]
30. Petracca, A.; Carrieri, M.; Avola, D.; Basso Moro, S.; Brigadoi, S.; Lancia, S.; Spezialetti, M.; Ferrari, M.; Quaresima, V.; Placidi, G. A virtual ball task driven by forearm movements for neuro-rehabilitation. In Proceedings of the International Conference on Virtual Rehabilitation (ICVR), Valencia, Spain, 9–12 June 2015; pp. 162–163.
31. Palestra, G.; Rebiai, M.; Courtial, E.; Koutsouris, D. Evaluation of a rehabilitation system for the elderly in a day care center. *Information* **2019**, *10*, 3. [[CrossRef](#)]

32. Avola, D.; Cinque, L.; Foresti, G.L.; Marini, M.R. An interactive and low-cost full body rehabilitation framework based on 3D immersive serious games. *J. Biomed. Inform.* **2019**, *89*, 81–100. [[CrossRef](#)] [[PubMed](#)]
33. Shum, L.C.; Valdés, B.A.; Hodges, N.J.; Van der Loos, H.F.M. Error Augmentation in Immersive Virtual Reality for Bimanual Upper-Limb Rehabilitation in Individuals with and without Hemiplegic Cerebral Palsy. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *28*, 541–549. [[CrossRef](#)]
34. Avola, D.; Cinque, L.; Pannone, D. Design of a 3D Platform for Immersive Neurocognitive Rehabilitation. *Information* **2020**, *11*, 134. [[CrossRef](#)]
35. Almasi, S.; Shahmoradi, L.; Ansari, N.N.; Honarpishe, R.; Ahmadi, h. Kinect-based Virtual Rehabilitation for Upper Extremity Motor Recovery in Chronic Stroke. In Proceedings of the International Serious Games Symposium (ISGS), Tehran, Iran, 23–25 December 2020; pp. 51–60.
36. Demirhan, M.; Premachandra, C. Development of an automated camera-based drone landing system. *IEEE Access* **2020**, *8*, 202111–202121. [[CrossRef](#)]
37. Premachandra, C.; Thanh, D.N.H.; Kimura, T.; Kawanaka, H. A study on hovering control of small aerial robot by sensing existing floor features. *IEEE/CAA J. Autom. Sin.* **2020**, *7*, 1016–1025. [[CrossRef](#)]
38. Tsouros, D.C.; Bibi, S.; Sarigiannidis, P.G. A review on UAV-based applications for precision agriculture. *Information* **2019**, *10*, 349. [[CrossRef](#)]
39. Gupta, H.; Verma, O.P. Monitoring and surveillance of urban road traffic using low altitude drone images: A deep learning approach. *Multimed. Tools Appl.* **2021**, 1–21. [[CrossRef](#)]
40. Roberts, R.; Inzerillo, L.; Di Mino, G. Using UAV based 3D modelling to provide smart monitoring of road pavement conditions. *Information* **2020**, *11*, 568. [[CrossRef](#)]
41. Pijl, A.; Tosoni, M.; Roder, G.; Sofia, G.; Tarolli, P. Design of terrace drainage networks using UAV-based high-resolution topographic data. *Water* **2019**, *11*, 814. [[CrossRef](#)]
42. Ahmed, W.; Shi, W.; Wenbin, X. Modeling complex building structure (LoD2) using image-based point cloud. In Proceedings of the IEEE International Conference on Image Processing, Applications and Systems (IPAS), Sophia Antipolis, France, 12–14 December 2018; pp. 110–114.
43. Zheng, X.; Wang, F.; Xia, J.; Gong, X. The methodology of UAV route planning for efficient 3D reconstruction of building model. In Proceedings of the International Conference on Geoinformatics (ICG), Buffalo, NY, USA, 2–4 August 2017; pp. 1–4.
44. Barrile, V.; Candela, G.; Fotia, A.; Bernardo, E. UAV survey of bridges and viaduct: Workflow and application. In Proceedings of the International Conference on Computational Science and Its Applications (ICCSA), Saint Petersburg, Russia, 1–4 July 2019; pp. 269–284.
45. Jiang, W.; Zhou, Y.; Ding, L.; Zhou, C.; Ning, X. UAV-based 3D reconstruction for hoist site mapping and layout planning in petrochemical construction. *Autom. Constr.* **2020**, *113*, 103137. [[CrossRef](#)]
46. Shao, Z.; Yang, N.; Xiao, X.; Zhang, L.; Peng, Z. A multi-view dense point cloud generation algorithm based on low-altitude remote sensing images. *Remote Sens.* **2016**, *8*, 381. [[CrossRef](#)]
47. Hayakawa, Y.S.; Obanawa, H. Volumetric change detection in bedrock coastal cliffs using terrestrial laser scanning and uas-based SFM. *Sensors* **2020**, *20*, 3403. [[CrossRef](#)] [[PubMed](#)]
48. Westoby, M.J.; Brasington, J.; Glasser, N.F.; Hambrey, M.J.; Reynolds, J.M. ‘Structure-from-Motion’ photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology* **2012**, *179*, 300–314. [[CrossRef](#)]
49. Han, D.; Lee, S.B.; Song, M.; Cho, J.S. Change Detection in Unmanned Aerial Vehicle Images for Progress Monitoring of Road Construction. *Buildings* **2021**, *11*, 150. [[CrossRef](#)]
50. Mesquita, D.B.; dos Santos, R.F.; Macharet, D.G.; Campos, M.F.; Nascimento, E.R. Fully convolutional siamese autoencoder for change detection in UAV aerial images. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 1455–1459. [[CrossRef](#)]
51. Zhao, X.; Zhou, S.; Lei, L.; Deng, Z. Siamese network for object tracking in aerial video. In Proceedings of the International Conference on Image, Vision and Computing (ICIVC), Chongqing, China, 27–29 June 2018; pp. 519–523.
52. Almehsal, A.M.; Alenezi, M.R.; Alshatti, A.K. Accuracy assessment of small unmanned aerial vehicle for traffic accident photogrammetry in the extreme operating conditions of Kuwait. *Information* **2020**, *11*, 442. [[CrossRef](#)]
53. Yoo, L.S.; Lee, J.H.; Lee, Y.K.; Jung, S.K.; Choi, Y. Application of a Drone Magnetometer System to Military Mine Detection in the Demilitarized Zone. *Sensors* **2021**, *21*, 3175. [[CrossRef](#)]
54. Jia, N.; Yang, Z.; Yang, K. Operational effectiveness evaluation of the swarming UAVs combat system based on a system dynamics model. *IEEE Access* **2019**, *7*, 25209–25224. [[CrossRef](#)]
55. Suresh, M.; Ghose, D. UAV grouping and coordination tactics for ground attack missions. *IEEE Trans. Aerosp. Electron. Syst.* **2012**, *48*, 673–692. [[CrossRef](#)]
56. Kerr, C.; Jaradat, R.; Hossain, N.U.I. Battlefield mapping by an unmanned aerial vehicle swarm: Applied systems engineering processes and architectural considerations from system of systems. *IEEE Access* **2020**, *8*, 20892–20903. [[CrossRef](#)]
57. Çintaş, E.; Özyer, B.; Şimşek, E. Vision-based moving UAV tracking by another UAV on low-cost hardware and a new ground control station. *IEEE Access* **2020**, *8*, 194601–194611. [[CrossRef](#)]
58. Bianchi, M.; Barfoot, T.D. UAV Localization Using Autoencoded Satellite Images. *IEEE Robot. Autom. Lett.* **2021**, *6*, 1761–1768. [[CrossRef](#)]

59. Zhu, M.; Zhang, H.; Zhang, J.; Zhuo, L. Multi-level prediction Siamese network for real-time UAV visual tracking. *Image Vis. Comput.* **2020**, *103*, 104002. [[CrossRef](#)]
60. Avola, D.; Foresti, G.L.; Martinel, N.; Micheloni, C.; Pannone, D.; Piciarelli, C. Aerial video surveillance system for small-scale UAV environment monitoring. In Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Lecce, Italy, 29 August–1 September 2017; pp. 1–6.
61. Avola, D.; Foresti, G.L.; Martinel, N.; Micheloni, C.; Pannone, D.; Piciarelli, C. Real-time incremental and geo-referenced mosaicking by small-scale UAVs. In Proceedings of the International Conference on Image Analysis and Processing (ICIAP), Catania, Italy, 11–15 September 2017; pp. 694–705.
62. Avola, D.; Cinque, L.; Foresti, G.L.; Pannone, D. Homography vs similarity transformation in aerial mosaicking: Which is the best at different altitudes? *Multimed. Tools Appl.* **2020**, *79*, 18387–18404. [[CrossRef](#)]
63. Avola, D.; Cinque, L.; Fagioli, A.; Foresti, G.L.; Pannone, D.; Piciarelli, C. Automatic estimation of optimal UAV flight parameters for real-time wide areas monitoring. *Multimed. Tools Appl.* **2021**, *80*, 25009–25031. [[CrossRef](#)]
64. Avola, D.; Cinque, L.; Foresti, G.L.; Pannone, D. Visual cryptography for detecting hidden targets by small-scale robots. In Proceedings of the International Conference on Pattern Recognition Applications and Methods (ICPRAM), Funchal, Madeira, 16–18 January 2018; pp. 186–201.
65. Avola, D.; Cinque, L.; Diko, A.; Fagioli, A.; Foresti, G.L.; Mecca, A.; Pannone, D.; Piciarelli, C. MS-Faster R-CNN: Multi-Stream Backbone for Improved Faster R-CNN Object Detection and Aerial Tracking from UAV Images. *Remote Sens.* **2021**, *13*, 1670. [[CrossRef](#)]
66. Avola, D.; Cinque, L.; Fagioli, A.; Foresti, G.L.; Massaroni, C.; Pannone, D. Feature-based SLAM algorithm for small scale UAV with nadir view. In Proceedings of the International Conference on Image Analysis and Processing (ICIAP), Trento, Italy, 9–13 September 2019; pp. 457–467.
67. Mittal, P.; Singh, R.; Sharma, A. Deep learning-based object detection in low-altitude UAV datasets: A survey. *Image Vis. Comput.* **2020**, *104*, 104046. [[CrossRef](#)]
68. Kang, H.; Joung, J.; Kim, J.; Kang, J.; Cho, Y.S. Protect your sky: A survey of counter unmanned aerial vehicle systems. *IEEE Access* **2020**, *8*, 168671–168710. [[CrossRef](#)]
69. Haralick, R.M. Statistical and structural approaches to texture. *Proc. IEEE* **1979**, *67*, 786–804. [[CrossRef](#)]
70. Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural features for image classification. *IEEE Trans. Syst. Man, Cybern.* **1973**, *3*, 610–621. [[CrossRef](#)]
71. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A training algorithm for optimal margin classifiers. In Proceedings of the Annual Workshop on Computational Learning Theory (AWCLT), Pittsburgh, PA, USA, 27–29 July 1992; pp. 144–152.
72. Peleg, S.; Naor, J.; Hartley, R.; Avnir, D. Multiple resolution texture analysis and classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **1984**, *6*, 518–523. [[CrossRef](#)] [[PubMed](#)]
73. Zhang, J.; Tan, T. Brief review of invariant texture analysis methods. *Pattern Recognit.* **2002**, *35*, 735–747. [[CrossRef](#)]
74. Davis, L.S.; Johns, S.A.; Aggarwal, J. Texture analysis using generalized co-occurrence matrices. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *1*, 251–259. [[CrossRef](#)] [[PubMed](#)]
75. Agarwal, R.; Jalal, A.S.; Arya, K. A multimodal liveness detection using statistical texture features and spatial analysis. *Multimed. Tools Appl.* **2020**, *79*, 13621–13645. [[CrossRef](#)]
76. Humeau-Heurtier, A. Texture feature extraction methods: A survey. *IEEE Access* **2019**, *7*, 8975–9000. [[CrossRef](#)]
77. Avola, D.; Cinque, L. Encephalic NMR image analysis by textural interpretation. In Proceedings of the ACM Symposium on Applied Computing (SAC), Fortaleza, Brazil, 16–20 March 2008; pp. 1338–1342.
78. Avola, D.; Ferri, F.; Grifoni, P. Genetic algorithms and other approaches in image feature extraction and representation. In *Artificial Intelligence for Maximizing Content Based Image Retrieval*; IGI Global: Hershey, PA, USA, 2009; pp. 1–19.
79. Avola, D.; Cinque, L. Encephalic NMR Tumor Diversification by Textural Interpretation. In Proceedings of the International Conference on Image Analysis and Processing (ICIAP), Vietri sul Mare, Italy, 8–11 September 2009; pp. 394–403.
80. Avola, D.; Cinque, L.; Foresti, G.; Martinel, N.; Pannone, D.; Piciarelli, C. Low-level feature detectors and descriptors for smart image and video analysis: A comparative study. In *Bridging the Semantic Gap in Image and Video Analysis*; Springer: Cham, Switzerland, 2018; pp. 7–29.
81. Avola, D.; Cinque, L.; Foresti, G.L.; Martinel, N.; Pannone, D.; Piciarelli, C. A UAV Video Dataset for Mosaicking and Change Detection From Low-Altitude Flights. *IEEE Trans. Syst. Man Cybern.* **2020**, *50*, 2139–2149. [[CrossRef](#)]
82. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection: A survey. *ACM Comput. Surv. (CSUR)* **2009**, *41*, 1–58. [[CrossRef](#)]
83. Pang, G.; Shen, C.; Cao, L.; Hengel, A.V.D. Deep learning for anomaly detection: A review. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–38. [[CrossRef](#)]
84. Ruff, L.; Kauffmann, J.R.; Vandermeulen, R.A.; Montavon, G.; Samek, W.; Kloft, M.; Dietterich, T.G.; Müller, K.R. A unifying review of deep and shallow anomaly detection. *Proc. IEEE* **2021**, *109*, 1–40. [[CrossRef](#)]
85. Nguyen, H.; Tran, K.P.; Thomassey, S.; Hamad, M. Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management. *Int. J. Inf. Manag.* **2021**, *57*, 102282. [[CrossRef](#)]
86. Piciarelli, C.; Avola, D.; Pannone, D.; Foresti, G.L. A vision-based system for internal pipeline inspection. *IEEE Trans. Ind. Inform.* **2018**, *15*, 3289–3299. [[CrossRef](#)]

87. Bay, H.; Tuytelaars, T.; Van Gool, L. Surf: Speeded up robust features. In Proceedings of the European Conference on Computer Vision (ECCV), Graz, Austria, 7–13 May 2006; pp. 404–417.
88. Song, B.C.; Kim, M.J.; Ra, J.B. A fast multiresolution feature matching algorithm for exhaustive search in large image databases. *IEEE Trans. Circuits Syst. Video Technol.* **2001**, *11*, 673–678. [\[CrossRef\]](#)
89. Ojala, T.; Pietikainen, M.; Harwood, D. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. In Proceedings of the International Conference on Pattern Recognition (ICPR), Jerusalem, Israel, 9–13 October 1994; Volume 1, pp. 582–585.
90. Fahad, L.G.; Tahir, S.F. Activity recognition and anomaly detection in smart homes. *Neurocomputing* **2021**, *423*, 362–372. [\[CrossRef\]](#)
91. Candel, A.; Parmar, V.; LeDell, E.; Arora, A. *Deep learning with H2O*; H2O.ai, Inc.: Mountain View, CA, USA, 2016; pp. 1–21.
92. Yamauchi, M.; Ohsita, Y.; Murata, M.; Ueda, K.; Kato, Y. Anomaly detection in smart home operation from user behaviors and home conditions. *IEEE Trans. Consum. Electron.* **2020**, *66*, 183–192. [\[CrossRef\]](#)
93. Ullah, W.; Ullah, A.; Haq, I.U.; Muhammad, K.; Sajjad, M.; Baik, S.W. CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks. *Multimed. Tools Appl.* **2021**, *80*, 16979–16995. [\[CrossRef\]](#)
94. Sultani, W.; Chen, C.; Shah, M. Real-world anomaly detection in surveillance videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 6479–6488.
95. Landi, F.; Snoek, C.G.; Cucchiara, R. Anomaly locality in video surveillance. *arXiv* **2019**, arXiv:1901.10364.
96. Nawaratne, R.; Alahakoon, D.; De Silva, D.; Yu, X. Spatiotemporal anomaly detection using deep learning for real-time video surveillance. *IEEE Trans. Ind. Inform.* **2019**, *16*, 393–402. [\[CrossRef\]](#)
97. Ata-Ur-Rehman; Tariq, S.; Farooq, H.; Jaleel, A.; Wasif, S.M. Anomaly detection with particle filtering for online video surveillance. *IEEE Access* **2021**, *9*, 19457–19468. [\[CrossRef\]](#)
98. Del Moral, P. Nonlinear filtering: Interacting particle resolution. *C. R. De L'Académie Des Sci.-Ser. I-Math.* **1997**, *325*, 653–658. [\[CrossRef\]](#)
99. Chan, A.; Vasconcelos, N. Ucsd pedestrian dataset. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 909–926. [\[CrossRef\]](#) [\[PubMed\]](#)
100. Sheikh, H. LIVE Image Quality Assessment Database Release 2. 2005; p. 1. Available online: <http://live.ece.utexas.edu/research/quality> (accessed on 23 October 2021).
101. Khan, M.U.K.; Park, H.S.; Kyung, C.M. Rejecting motion outliers for efficient crowd anomaly detection. *IEEE Trans. Inf. Forensics Secur.* **2018**, *14*, 541–556. [\[CrossRef\]](#)
102. Hamdi, S.; Bouindour, S.; Snoussi, H.; Wang, T.; Abid, M. End-to-End Deep One-Class Learning for Anomaly Detection in UAV Video Stream. *J. Imaging* **2021**, *7*, 90. [\[CrossRef\]](#)
103. Bonetto, M.; Korshunov, P.; Ramponi, G.; Ebrahimi, T. Privacy in mini-drone based video surveillance. In Proceedings of the IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 4–8 May 2015; pp. 1–6.
104. Chriki, A.; Touati, H.; Snoussi, H.; Kamoun, F. Deep learning and handcrafted features for one-class anomaly detection in UAV video. *Multimed. Tools Appl.* **2021**, *80*, 2599–2620. [\[CrossRef\]](#)
105. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–15 June 2015; pp. 1–9.
106. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, 20–26 June 2005; Volume 1, pp. 886–893.
107. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *Lond. Edinburgh Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559–572. [\[CrossRef\]](#)
108. Klaser, A.; Marszałek, M.; Schmid, C. A spatio-temporal descriptor based on 3d-gradients. In Proceedings of the British Machine Vision Conference (BMVC), Leeds, UK, 1–4 September 2008; p. 275.
109. Qiao, Y.; Wu, K.; Jin, P. Efficient Anomaly Detection for High-Dimensional Sensing Data with One-Class Support Vector Machine. *IEEE Trans. Knowl. Data Eng.* **2021**, *1*. [\[CrossRef\]](#)
110. Rasheed, W.; Tang, T.B. Anomaly detection of moderate traumatic brain injury using auto-regularized multi-instance one-class SVM. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *28*, 83–93. [\[CrossRef\]](#) [\[PubMed\]](#)