Optimizing Sector Ring Histogram of Oriented Gradients for human injured detection from drone images

Marzieh Ghasemi\textsuperscript{a,b}, Masood Varshosaz\textsuperscript{b}, Saied Pirasteh\textsuperscript{a} and Ghazal Shamsipour\textsuperscript{a}

\textsuperscript{a}Department of Surveying and Geoinformatics, Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong, Chengdu, China; \textsuperscript{b}Department of Photogrammetry, Faculty of Geomatics Engineering, K. N. Toosi University of Technology, Mirdamad Cross, Tehran, Iran

\textbf{ABSTRACT}
Developing a system for emergency response and rescue to find injured people within images has been an interest to many researchers. The key challenge is to define a proper feature to describe the human body’s appearance. Various features are often extracted from low-level data, such as texture and colour (Zhang et al., Sensors. 19(5):1005, 2020). One of the strong features is the Sector Ring Histogram of Oriented Gradients (SRHOG) that has been successfully applied to human detection tasks. Despite good accuracy in finding humans, an SRHOG detection method produces a large amount of false-negative labels. Locating an injured body after a disaster in drone images using an imaging camera remained a challenge. This study presents a new extension to SRHOG, so-called AdSRHOG, to reduce the number of false labels. In our approach, the gradient filters used by SRHOG defined adaptively depending on the corresponding pixel location. The proposed feature was used the Support Vector Machine (SVM) algorithm to detect humans on drone images. The experiments showed a significant improvement of up to 54.3\% in reducing the false labels. It was also found that the overall accuracy of the human detection process had a notable improvement of 13.1\% over a traditional SRHOG detection technique.

\textbf{ARTICLE HISTORY}
Received 12 May 2020
Accepted 28 January 2021

\textbf{KEYWORDS}
Human detection; drone; SRHOG; adaptive filters; Disaster; HOG; emergency response

1. Introduction
Today, natural disasters, including earthquakes, floods, landslides, and avalanches, take place in all parts of the world. In this regard, the most important task is the rescue of injured people. According to the Red Crescent (IRAN Red Crescent 2004), the first step of any search and rescue operation is to locate the injured people. Unfortunately, this can be a very difficult and time-consuming process where only a few extra minutes may be all needed to save a victim’s life. To this end, the use of drones has attracted the attention of many researchers (Blondel et al. 2013; Sun et al. 2016). Drones can be used...
to collect information from a remote distance without actually needing to be in a disaster environment (Doherty et al. 2007). Numerous rescue systems, including teleoperated drones, have been proposed to establish robotic assistance for urgent rescue tasks (Yeong et al. 2015; Jurecka and Niedzielski 2017; Zhang et al. 2019).

A multi-UAV system for unmanned search operations was implemented within the European project ICARUS framework, which uses a camera-based on a novel and very promising Quantum Cascade Detector (QCD) technology (De Cubber et al. 2013). This ultra-sensitive but commercial high-resolution lower-sensitivity microbolometer camera complements a relatively low-resolution QCD camera. AltiGator (AltiGator European Drone Manufacturer Company 2017) is another multi-copter aiming to aid the information acquisition with real-time imagery to make better decisions and gain extra time for the rescue teams in a search and rescue (SAR) operation (Silvagni et al. 2017; Jurecka and Niedzielski 2017). It carries an infrared (IR) thermal imaging camera that can detect human body heat. This capability dramatically increases the ability to find people or objects, even at night. Waharte and Trigoni. (2010) investigated a supporting search and rescue operation with UAVs. They addressed the fundamental parameters needed to locate and rescue a missing person using UAVs. AUXdron Lifeguard is another robot drone exclusively designed and implemented for urgent situations for swimmer rescue operations. This robot sends a video to the control station to localize the victims interactively.

The above examples show the importance and attention paid to the use of UAVs in SAR operations. Most of the technologies mentioned use thermal (Bravo et al. 2019; De Cubber et al. 2013) or Near Infrared (NIR) cameras (Lygoura et al. 2018; AltiGator European Drone Manufacturer Company 2017). However, drones carrying small RGB cameras are also of great interest. The images provide a rich source of information that could effectively locate injured people in post-disaster missions. In this regard, the view angle, scale and rotation of drone images cause humans to be viewed differently from terrestrial images. Besides, the injured person may be partially occluded by snow, rock and the like (Liu et al. 2017). Therefore, these issues make the detection of a human on UAV images a difficult and challenging task. As a result, defining a strong feature that can better describe the form of the human body in an image has been a central focus of different studies (Benenson et al. 2014; Malik et al. 2020).

Up until now, various features have been proposed that often are extracted from low-level data such as texture (Ojala et al. 2002; Leibe et al. 2005), colour (Ott and Everingham 2009; Wall et al. 2010) and edge (Olson and Huttenlocher 1997; Nguyen et al. 2009). For instance, Leibe et al. (2005) used texture information to identify pedestrians in a crowded scene. For this purpose, the so-called Haar Wavelets were utilized in which grey-level patterns determine considering the magnitude of the difference between nearby pixel intensities (Dollár et al. 2008). Another example is the Colour Self-Similarity (CSS) feature, which is defined by using the histogram of colour tones present in different parts of an image (Wall et al. 2010). In general, techniques that are based on texture or colour characteristics are highly dependent on the pixel values and, therefore, the image background may interfere with their performance. They are typically used along with the background subtraction or motion analysis techniques (Cutler and Davis 1998).
The edge features are useful to find the shape of an object because objects can represent edges. In contrast to colour and texture, the edge features describe objects mainly based on their geometry (Nguyen et al. 2011).

Ioffe and Forsyth (2001) used rectangular contours to model human body parts. The technique works best when other objects do not occlude the human body. To describe more complex shapes, template matching techniques have also been used (Gavrila and Philomin 1999). These techniques are often trained by using many sample images and, thus, depend mainly on the training data’s completeness.

Another alternative is the definition of body parts with curves and basic shapes, called Edgelets (Wu 2008). The Edgelet is a small part of a line or curve (Wu and Nevatia 2005). It includes information on the shape and brightness of an object and describes gradients. This approach may encounter problems if the object is so far away from the camera that its details cannot be properly recognized. A more complex gradient-based feature is called Shapelet (Sabzmeydani and Mori 2007). It uses a series of gradients with different directions and magnitudes to represent different parts of the human body (e.g. head, hand, and shoulders). Shapelets are used by unsupervised classifiers and are often prone to background complexities.

The Scale-Invariant Feature Transform (SIFT) is one of the most widely used features and applies the Histogram of Oriented Gradient (HOG) (Lowe 2004; Dalal and Triggs 2005). HOG was introduced in 2005 for pedestrian detection but has since been used in numerous applications (Blondel et al. 2013; Mihiçioğlu and Alkar 2019; Bahri et al. 2020). Nowadays, it has variations of its own. For instance, the Compressed Histogram of Oriented Gradient (CHOG) (Munder and Gavrila 2006) and SRHOG (Liu et al. 2017) are two examples of improvement of the HOG’s resistance to the brightness and rotation changes, respectively. Unlike SIFT, HOG uses the gradient of some keypoints and calculates the histograms of pixel gradients in a block. To calculate the HOG, first, the horizontal and vertical gradients have to be calculated; then, the magnitude and direction of the gradient vectors are determined to form a gradient image of which most redundant background detail has been omitted. This is achieved by generating a gradient vector orientation histogram (Dalal and Triggs 2005) (Figure 1).

Figure 1. An example of a gradient image created using the HOG descriptor. Source: Author.
In the studies above, the images used are mostly taken in front of the object or from a certain view angle (Gavrila 1999; Aggarwal and Cai 1999; Liu et al. 2017). However, the drone images use to identify the location of a person after a disaster that is prone to 3D irregular movements. As a result, an individual can appear at various poses, which in turn complicates the detection process.

To overcome the above problems, Liu et al. (2017) proposed SRHOG (an extension to HOG) that uses a dynamically defined local coordinate system r, t (Figure 2) instead of using the Cartesian coordinate system. SRHOG uses the deformed Sobel filter Approximated Radial Gradient Transform (ARGT) to calculate the gradients (Takacs et al. 2010; Takacs et al. 2013).

The gradient components (Gr, Gt) are defined in r and t directions shown in Figure 2. Therefore, the resulting vector does not change when this local coordinate system is rotated. This makes the feature vector resistant to the image rotations.

The experiments carried out by us (Ghasemi et al. 2020) have shown that drone images containing humans can be successfully and correctly labelled using a classifier that uses SRHOG as input. However, the results also include a lot of false image labels. This study presents a new SRHOG-based feature to overcome this problem that is called Adaptive SRHOG (AdSRHOG). The overall structure of this feature is the same as that of SRHOG. However, the shape of the gradient filter changes with respect to the location of the pixel. As will be shown in the Evaluations section, the proposed feature outperforms SRHOG by reducing the number of false labels. It can be used successfully and correctly for labelling drone images containing humans.

The following sections are structured into three parts. Section 2 explains how to calculate the new feature. The evaluation of the feature and experiments are described in the experimental exercises Section. The result discusses in Sect. 3, and finally, recommendations and suggestions are delivered in conclusion Sect. 4.

2. Methodology

Detecting humans on UAV images involves two main steps: forming image features and classifying the images using the formed features. In this study, a new feature, so-
called AdSRHOG, is developed to reduce false labels produced by the classification technique. In the following section, the methodology of computing the AdSRHOG feature is described. Concerning the classification process, different classifiers could be used. However, since HOG’s presentation, most researchers have used the Support Vector Machine (SVM) classifier. The same is true with SRHOG. Therefore, in this study, it is also decided to use SVM as the classifier.

SVM is a supervised Machine Learning (ML) algorithm used in many classification and regression problems. It works by finding an optimal separation line called ‘hyperplane’ to accurately separate two or more classes in the classification problem. The goal is to find the optimal hyperplane separation through a training process that
tries to separate the data linearly. In case the data cannot be separated linearly, SVM creates a hyperplane in the higher dimensional space. A good separation of classes is achieved by having a hyperplane with the largest distance (margin) to the nearest training data points (Hsu & Lin 2002).

The classification is done in two main steps: training and test. In the beginning, a pyramid is created for each image. The images are then scanned using a search window. For each window, the desired feature is calculated and used for training the classifier. The result of the training is a model, which is then used to label the test images. The results are the test images, and each labelled with one class.

In the following, first, it is shown how the AdSRHOG features are computed. Then, in the Implementation section, the way the SVM classifier uses them to detect humans on the UAV images is described.

2.1. Computing AdSRHOG feature

As already mentioned, this study aims to improve the SRHOG feature to reduce the number of false labels produced by the classifier. In this section, it is shown how the AdSRHOG features of an image are computed. Figure 3 shows the steps of how to calculate the proposed feature.

The stages in this diagram are mostly the same as those of SRHOG. However, the filters used to convolve the search window are uniquely defined depending on the pixel’s angular location within the search window. The overall process is divided into three main steps: Similar to SRHOG (Liu et al. 2017), first, the image is scanned by 128 × 128 search windows. The search window is convolved with two gradient filters at each location, whose shape changes depending on the polar angle θ (Figure 4).

θ is the anti-clockwise polar angle of the pixel. Once the gradients are calculated, the search window is divided into sector ring blocks to compute gradients’ histogram.
In each block, the collection of bin heights forms its feature vector. Putting these vectors together creates the final vector feature of the search window. Also, \(M_i\) and \(M_j\) are Magnitude vectors for \(i\) and \(j\) pixel position at XY coordinate system on the image.

The following algorithm shows how to calculate the ADSRHOG feature of a given search window.

**Algorithm 1: Calculate ADSRHOG**

**Input:** A \(128 \times 128\) window  
**Output:** ADSRHOG vector

Centrale pixel = \((C_i, C_j)\), \(W_p\) = A \(3 \times 3\) neighbor window for each pixel, Direction bins = \{\(z_1, z_2, \ldots, z_9\)\}, bins magnitude = \{\(m_1, m_2, \ldots, m_9\)\}

1. For each pixel \((i, j)\)
   \[\theta = \text{atan}\left(\frac{(j - C_j)}{(i - C_i)}\right)\]
   Determine the Gradient kernels \(g_r\) and \(g_t\) base on \(\theta\)
   \[G_r = W_p \times g_r\]
   \[G_t = W_p \times g_t\]
   Magnitude = \(\sqrt{G_r^2 + G_t^2}\), Direction = \(\text{atan}(G_t/G_r)\)

2. For each 4 neighbor sectors binning the histogram by weighted sum of the inner pixels

   If \(z_i < \theta < z_j\)
   \[M_i = m_i + (z_j - \theta)/20 \times \text{Magnitude}\]
   \[M_j = m_j + (\theta - z_i)/20 \times \text{Magnitude}\]

3. Join the vectors together

In summary, the main steps of computing the AdSRHOG feature of any search window are calculating the gradients, forming the histogram of the computed gradients, and forming the search window’s feature vector. These steps are described in detail as follows.

**2.1.1. Calculating the gradient of search window pixels**

To compute the AdSRHOG feature, the first step is to calculate the gradient of the search window pixels. For this, first, the search window is divided into eight \(45^\circ\) angular sections. Then, for each pixel, two filters, along with and perpendicular to the pixel’s radial bearing, are used to calculate its gradient values. The main contribution of this study is, in fact, here that to propose a new way of defining these filters. Two filters are proposed, one of which is radial, and the other is tangential. The radial filters are shaped according to \(\theta\), the pixel’s angular bearing with respect to the centre of the search window (see Figure 4).

If \(\theta\) is equal to \(0^\circ\) or \(180^\circ\), the filters are the same as those of the Sobel filter in the \(x\) and \(y\) directions. However, if \(0^\circ < \theta \leq 45^\circ\), \(45^\circ < \theta \leq 90^\circ\), and so on, the radial filter is calculated by the values shown in Figure 3. To calculate the tangential filter, the radial filter is rotated by \(90^\circ\). An important note is that, if \(\theta\) is not an integer
multiplication of 45°, the filter coefficients are computed by averaging the neighbouring filters.

Once the search window is convolved with the filters defined as explained, the magnitude and direction of the gradient vector can be computed by the following equations:

$$\text{Gradient magnitude} = \sqrt{G_r^2 + G_t^2}$$

$$\text{Gradient direction} = \arctan \left( \frac{G_t}{G_r} \right)$$

where $G_r$ and $G_t$ are the radials and tangential gradients, respectively.

### 2.1.2. Computing the histogram of gradients

Once the gradients of all search window pixels are computed, the next step is to compute the histogram of the gradients. As described, this part is similar to that of the SRHOG and is briefly explained in this section. In this step, the search window is divided into 15 sectors, each of which is also divided into 16 cells along the sector radius. The numbers, i.e. 15 and 16, are the optimum values for calculating the histogram of gradients (Liu et al. 2017). Each of the four adjacent cells forms a block (see the green part in Figure 5). To compute the feature vector of the search window, we need to calculate the histogram of gradients of these blocks. As can be seen, each block has some overlap with the preceding one (Figure 5a). This is done to make the feature more resistant against illumination changes.

As shown in Figure 5(c), the horizontal axis of the histogram corresponds to the gradient directions ranging from $0^\circ$ to $160^\circ$ (i.e. 9 parts per $20^\circ$). To compute the histogram of gradients for each block, we need to define some bins and assign each block pixel to one of the bins. The bins are considered at angles $0^\circ$, $20^\circ$, $40^\circ$, $60^\circ$ and so on until $160^\circ$. To fill in the bins, any block pixel whose gradient direction value is
in the first half of a bin is put in that bin; otherwise, it is assigned to the following bin (Takacs et al. 2010). The vertical axis contributes to the height of the histogram bar and is proportional to the weighted sum of the pixel gradient. The weight of each pixel gradient is calculated based on its distance to the centre of the search window and its angle to the x-axis.

### 2.1.3. Forming the feature vector

Once the histograms of all the search window blocks are computed, the final step is to define its feature vector. For this, the vector is generated by putting the block
It should be noted that the blocks have some overlaps. Therefore, the number of feature vectors that have been amended together is much higher than that of the window cells.

In the following sections, first, it is explained how the AdSRHOG feature is used by SVM to label UAV images (the Implementation section). Then, various tests carried out to evaluate the efficiency of the proposed feature for human detection in drone images are described.

### 2.2. Implementation

As stated in the introduction section, this study aims to develop a method for human detection utilizing drone images after natural disasters. Figure 6 illustrates the proposed process. As already mentioned, depending on their distance from the drone, humans appear in different shapes and dimensions. The detection process must, therefore, be carried out at different scales. For this purpose, an image pyramid is formed in four levels, each level of which is the product from the preceding level multiplied by 1.05.

In this research, the Matrix Laboratory (MATLAB) platform and Python is used to program the proposed algorithm. The supervised classification technique SVM was applied to detect humans. There are three main features of SVM: (i) use of an optimised hyper plane to distinguish the two classes, (ii) establishment of the maximum boundary margin between the classes and (iii) the minimal noise effects when the quality of the data is low (Cortes and Vapnik 1995). As a result, it has been traditionally used in most HOG-based classifications to detect humans within images (Dalal and Triggs 2005; Pang et al. 2011; Cao et al. 2011). SVM consists of two stages of
training and test. During the training stage, a classification model is built to define whether an image (i.e. the test image) has to be labelled positive or negative. Positive samples contain a human being (either in whole or in part), whereas a negative sample indicates no human presence in the image. The image here refers to the 128×128-pixel samples, which could be either positive or negative.

As shown in Figure 6, the whole process begins by extracting the AdSRHOG features of the training samples, which are then fed into the SVM classifier. The result is a training model that includes the classification parameters. At this point, we can evaluate any sample to see if it is positive or negative. For this, similar to the training stage, the test image is scanned using a sliding window. The dimensions of this window are equal to those of the training samples. To define the label of the sliding window, its AdSRHOG features are extracted and passed to the trained SVM. Therefore, if the test sample contains a body, either in whole or in part, a positive label is given; otherwise, a negative is given.

In the experiments carried out, unfortunately, the authors could not get hold of images taken after a real disaster. However, different image sets that almost entirely resemble real search and rescue task conditions have been used. The first dataset used is the INRIA dataset, which has been commonly used along with the HOG feature as a comprehensive benchmark for testing human and pedestrian detection algorithms in several studies (Ojala et al. 2002; Dalal and Triggs 2005; Benenson et al. 2014). The dataset includes ground images from human positions in standing, walking, and upright views only. In addition, since a person is viewed mostly from a top-down/oblique angle in a drone, many additional images taken by several drones at different angles were acquired and used. These include images taken by a DJI Tello from an oblique angle at low altitude, and those taken by Phantom 4 Pro, Inspire 2, and Parrot ArDrone 2.0, all taken nadir at different heights. Among these, in one of the sets, students were spread in a relatively large area. They took different poses, including standing, lying down, and sitting to closely resemble UAV images used in a

Figure 8. Some examples of drone images used in this study. Source: Author.
real search and rescue operation. A Phantom 4 Pro took these images at 60 m height. Details of these images are given in the following lines.

INRIA is a dataset (Figure 7) from which 2164 positive and 432 negative samples are used for training and 1126 positive and 453 negative samples for testing. The dimensions of the samples are all $64 \times 128$ pixels. However, as mentioned above, the image samples must be $128 \times 128$ pixels for our AdSRHOG-extraction feature. Therefore, the border parts of the samples were duplicated to produce $128 \times 128$ pixel samples (Figure 7a).

Also, the drone images acquired by the authors were 592 and 200 positive and negative samples, respectively. Figures 8 and 9 show some examples used in these evaluations.

The Tello images were taken at lower altitudes to enrich the training dataset with images having different view angles. Table 1 shows the number of samples taken from various sources to test the proposed feature.
In the following section, the tests carried out to evaluate the efficiency of AdSRHOG for classifying UAV images taken in SAR operations.

### 2.3. Evaluations methodology

For the evaluation process, three indices **Recall**, **Recall_neg**, and **Precision** (Wójcikowski 2016) are used. They are computed using True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) figures (Table 2).

The Recall index shows the ratio of the correctly identified positive samples over the total number of positive windows and is computed by:

\[
\text{Recall} = \frac{TP}{TP + FN} \tag{3}
\]

In effect, Recall shows how strong the proposed feature is in identifying the positive samples.

The second index is Recall value for negative samples, which is named Recall_neg refers to the ratio of correctly identified negative samples over the total number of all negative samples, and it is calculated by:

\[
\text{Recall}_\text{neg} = \frac{TN}{TN + FP} \tag{4}
\]

A bigger Recall_neg suggests that the procedure is stronger, and it makes fewer mistakes.

The precision refers to the ratio of correctly identified positive windows over all of the positive windows, and it is computed by:

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{5}
\]

| Type                  | Train | Test     |
|-----------------------|-------|----------|
|                       | Positive | Negative | Positive | Negative |
| INRIA                 | 2164   | 432      | 1126     | 453      |
| Parrot AR Drone 2.0   | 592    | 200      | 60       | 60       |
| DJI Inspire 2         |        |          |          |          |
| DJI Phantom4 Pro      |        |          |          |          |
| DJI Tello             | 162    | 92       | 2500 (different positions) | 0 |

#### Table 2. Meaning of indices.

| Index | Meaning                                      |
|-------|----------------------------------------------|
| TP    | Positive windows correctly identified        |
| FN    | Positive windows identified as negative      |
| TN    | Negative windows correctly identified        |
| FP    | Negative windows identified as positive      |

In the following section, the tests carried out to evaluate the efficiency of AdSRHOG for classifying UAV images taken in SAR operations.
3. Results and discussion

In order to evaluate the performance of the improved feature, five experiments are carried out.

Comparing the overall accuracy of the proposed feature to that of SRHOG: This was done twice; at first, the INRIA dataset is used to examine the strength of AdSRHOG in detecting humans in nadir images.

Evaluating the effect of the equivalent angle on the accuracy and time of computation: As indicated above, the gradient filter arrays utilizing the corresponding pixel bearing angle are computed. It means that a unique filter must be established for each pixel. It involves a lot of calculations and is not cost-effective. As an alternative, the search window can be divided into equal angular sections, all pixels of which use the same equivalent \( \theta \), so-called the equivalent \( \theta \), for filter computations. In other words, any pixels within a 45° or 22.5° section will use the same \( \theta \) to calculate its filter elements (see Figure 10).

This experiment aims to find the optimum number of sections.

Evaluating the efficiency of AdSRHOG against SRHOG in detecting humans appearing in different/more complex situations: As already mentioned, the primary purpose of this study is to detect injured humans in drone images. In general, humans can appear in different poses, like standing, sitting, lying down. Other objects can also occlude them. Besides, there is no guarantee that the images will be taken under proper lighting conditions. Thus, in a third test, the proposed feature’s performance is studied for human detection under such circumstances. Depending on the system configuration, these steps can be time-consuming. This issue more crucial in SAR operations as saving the life of a victim of an accident could depend on a few

Precision shows the overall accuracy of the method.
extra minutes (Silvagni et al. 2017; Jurecka and Niedzielski 2017). Thus, the calculation times are also studied in this section.

a. Test on real data: This test was performed in conditions close to the actual conditions of natural disasters. The test aimed to determine the potential of the proposed approach under real conditions.

b. Study the impact of height on human identification: In this experiment, the performance of AdSRHOG is compared against that of SRHOG using two different datasets taken at different heights. The aim is to get a better insight impact of the drone height on the human identification results.

The results of the experiments are presented in the following sections.

### 3.1. Comparing the overall accuracy of AdSRHOG against SRHOG

The purpose of this test was to determine whether the change made to the gradient calculation led to an improvement in the performance of the SRHOG method or not. Each algorithm was individually trained with INRIA dataset training samples. The algorithms are evaluated on the 1126 positive search windows and 453 negative test sample windows on the INRIA dataset. Then the evaluation indices were calculated for them. Table 3 illustrates the results of this study.

As shown in Table 3, the precision of SRHOG is 73.69%, while the precision of the proposed method is 94.21%. This difference is 20.52%. Although the precision index increased significantly, the Recall value that indicates the number of correctly identified humans has not improved much (1.1%). In other words, out of 1126 positive windows, 846 windows were correctly classified by using SRHOG, while only 16 more humans (862 in total) were correctly identified using AdSRHOG. However, as it can be seen, out of 453 negative samples, SRHOG classified 151 cases correctly, and the AdSRHOG correctly labelled 400 items. This means the Recall value of the negative data has significantly increased (from 33.33% to 88.30%). Therefore, the improvement in precision is the result of a reduction in miss-classifications (shown by Recall_neg in Table 3).

To evaluate the ability of the proposed feature and to detect humans from the top view, 60 positive and 60 negative samples by Parrot ArDrone were taken. Then, the drone images (Figure 8) were extracted and tested by applying SRHOG and AdSRHOG. Table 4 depicts the results of the test.

From the 60 positive samples, 52 cases were correctly classified by SRHOG, while AdSRHOG correctly classified only 48 cases. However, only 20 cases were correctly classified by SRHOG among 60 negative samples, while 51 cases were correctly classified by the improved method. Therefore, achieving a precision of 84.21%, AdSRHOG performed better than SRHOG (56.52%).

---

### Table 3. Comparing SRHOG with AdSRHOG.

| Method | TP   | FN   | Recall | TN   | FP   | Recall_neg | Precision (%) |
|--------|------|------|--------|------|------|------------|---------------|
| SRHOG  | 846  | 280  | 0.7513 | 151  | 302  | 0.3333     | 0.7369        |
| AdSRHOG| 862  | 264  | 0.7655 | 400  | 53   | 0.8830     | 0.9421        |

---
3.2. Comparison of different estimates of AdSRHOG

As mentioned above, dividing the search window into equal angular sections (Figure 10) can reduce the computation time. However, the question is the number of sections that, in addition to reducing the time, would have a minimal impact on the accuracy of the results. To do this, we need to see if the number of sections is changed and how the performance of the proposed feature is affected, from the speed and accuracy points of view. To define the optimal figure, the search window is divided into a different number of angular sections. The proposed feature was computed using the equivalent $h$ and used to classify the images in each case. Besides, the test was carried out by using actual $h$ values, i.e. without dividing the search window into angular sections. All tests are performed on the INRIA dataset, including 1126 and 453 positive and negative samples, respectively. The results are shown in Table 5, the feature vector computation speed of each scenario (last column), is calculated using 100 samples. The computer used in these experiments was an Intel® core™ i7-7700 HQ CPU and 16 GB RAM.

As can be seen, there is no significant difference between the time of calculation and also the value of Precision or Recall_neg. However, the reference value for 45° sections is 76.55%, whereas that for 22.5° or bigger is 90.76% or more. This means, although dividing the search window by 16 sections or more does not significantly impact the computation time, it leads to an improvement of 21.14% over the accuracy of detecting positive samples.

Overall, it was revealed that the algorithm’s performance is particularly improved by dividing the central angle into 16 sections. Therefore, the number of sections is used to perform the following experiment (see next section).

3.3. Evaluation of AdSRHOG in various human states

Injured people may appear in various positions, such as standing, sitting, lying down, and part of the body occluded or be in the shadows. There is also no guarantee that
Figure 11. Human in different situations. Source: Author.
a human being will be imaged under proper lighting conditions. In this section, the performance of SRHOG is compared to that of AdSRHOG using images in which humans appear in different shapes. In this experiment, the gradient sections are computed at 22.5° angles, as justified above. For each situation, 500 positive samples were used (Figure 11).

Only are used the Tello images in this experiment. This dataset could capture different standing, sitting, lying down, occluded, and being in imperfect lighting conditions. The experiment was carried out by using 500 images. Figure 11 depicts examples of the Tello images.

The images have been selected in such a way that humans appear in different angles of view, including sitting, standing and lying down poses. In some situations, the upper or lower half of the person’s body is occluded by other objects, such as trees or shadows. Also, in some images, the lighting conditions are not very good, as shadows partially cover the human body. The images were once again used to test the proposed feature for human detection. Table 6 illustrates the results of the third experiment.

As can be seen, both features are performing relatively well in standing and lying down situations. In all situations, however, the AdSRHOG-based classification has a higher recall value than that of SRHOG. In a lying down position, this value is, of course, only 3%. As shown, a human being has a better chance of being detected by both features when in a standing and lying down position. The Recall values for SRHOG in standing and sitting states are 73.42% and 51.63%, respectively. Also, in sitting cases, the accuracy is 21.79% less than that of standing poses.

In weak lighting conditions, the Recall value is 55.48%, a relatively low value concerning the standing cases. The Recall values for AdSRHOG in standing, sitting situations are respectively 90.60%, 77.20%. In weak lighting conditions, this value is equal to and 69.40%, which is a relatively low value. Perhaps improving the contrast could help the results.

The low percentage of the Recall value for occluded human detection is worth noting. Only 26.2% of images containing humans have been detected using the SRHOG algorithm, while this figure has been increased to 36.21% by implementing the proposed feature. Despite the 10% increase, this is not an acceptable figure for an image classification technique. Perhaps the low accuracy of the results is due to the lack of adequate training data set. In our experiments, algorithms have been trained mainly using images containing the entire human body. As a result, the classification gave poor results when trying to test data that only partially contained humans.

| Position                  | Improved method | SRHOG |
|---------------------------|-----------------|-------|
|                           | TP   | FN   | Recall | TP   | FN   | Recall |
| Standing                  | 453  | 47   | 90.6   | 367  | 133  | 73.42  |
| Sitting                   | 386  | 114  | 77.20  | 258  | 242  | 51.63  |
| Lying down                | 470  | 96   | 88     | 425  | 75   | 85.10  |
| Weak lighting condition   | 347  | 153  | 69.40  | 377  | 183  | 55.48  |
| Occluded body             | 181  | 319  | 36.21  | 131  | 369  | 26.20  |

Table 6. Comparison of SRHOG against AdSRHOG in different human poses.
3.4. Evaluation of AdSRHOG with images similar to those taken in real SAR operations

As mentioned before, the authors were not able to get access to images taken after a real disaster. To this end, nadir images taken by a Phantom 4 Pro at 60 m height from several people spread in a relatively large area are used. The test area is located around the city of Qazvin in Iran, which is located at longitude 49.85° and latitude 36.9°. Figure 12 illustrates examples of these images.

As in previous tests, 125 × 125 windows were extracted from 546 images at different pyramid levels of each image frame. Half of the windows did not contain a human, while the other half did. In this test, the people presented different postures.

Figure 12. An example of semi-realistic images; Pi: Positive images; Ni: Negative images. Source: Author.
of the human body, including standing, sitting and lying down. These images were taken at 60 meters’ height.

At this stage, 1400 positive windows, 1400 negative windows were used for training and 600 positive windows and 600 negative windows for testing. The results are given in Table 7.

Table 7 illustrates the accuracy of the ADSRHOG method, and it is 88.63%, while the accuracy of the SRHOG method is 56.12%. According to the tests relevant to Sect. 3.1, a gap of 32.51% is expected. Similar to the test in Sect. 3.1, this difference between the precision of both approaches results from SRHOG misclassification of negative samples. The number of negative samples correctly classified is 531 out of a total of 600 negative samples. Regarding the positive samples, the Recall value of AdSRHOG is 89.67%. It shows an improvement of 4.17% compared to SRHOG, which is 85.50%. Therefore, confirming the previous tests’ findings, this experiment also suggested that the proposed feature is an effective one, even in situations close to real SAR applications.

3.5. Study the impact of height on human identification

In this experiment, the performance of AdSRHOG is compared against that of SRHOG by using two different datasets taken at different heights. For this, images taken by Phanotom 4 Pro at 60 m and 100 m heights are used. In both datasets, 300 samples (200 positive and 100 negative samples) were used for training and 120 samples (i.e. 60 positive and 60 negative samples) for the test. In each case, both training and test samples were from the same image set, i.e. of either 60 m or 100 m height. Then, the images were classified once using AdSRHOG and next with SRHOG. The results are shown in Table 8.

As can be seen, at different heights, the Recall values of both AdSRHOG and SRHOG are almost identical. Also, in both cases, the Recall_neg of AdSRHOG is around 30% bigger than that of SRHOG. Therefore, it is seen, the precision obtained using AdSRHOG is nearly 30% better than that achieved using SRHOG features. This means changing the height has not had a notable effect on the relationship between SRHOG and AdSRHOG. In other words, even when the height is changed, similar to the previous tests, AdSRHOG is still the preferred feature for human detection tasks.

| Table 7. Evaluating AdSRHOG in a semi-realistic project. |
|----------------|-------|-------|-------|-------|-------|-------|-------|
| Method        | Dataset          | TP    | FN    | TN    | FP    | Recall | Recall_neg | Precision (%) |
| AdSRHOG       | Semi natural     | 538   | 62    | 531   | 69    | 0.8967 | 0.8850     | 0.8863       |
| SRHOG         | Semi natural     | 513   | 87    | 199   | 401   | 0.8550 | 0.3316     | 0.5612       |

| Table 8. The effect of drone height on human detection. |
|----------------|-------|-------|-------|-------|-------|-------|-------|
| Height (m)    | Feature used | TP    | FN    | TN    | FP    | Recall | Recall_neg | Precision |
| 60             | AdSRHOG      | 53    | 7     | 50    | 10    | 88.33  | 83.33     | 84.13      |
|                | SRHOG        | 49    | 11    | 21    | 39    | 81.67  | 35.00     | 55.68      |
| 100            | AdSRHOG      | 54    | 6     | 49    | 11    | 90.00  | 81.67     | 83.08      |
|                | SRHOG        | 47    | 13    | 19    | 41    | 78.33  | 31.67     | 53.41      |
when using UAV images. However, when moving from 60 m to 100 m images, the precision has slightly decreased by around 1.5% for both features. Overall, it does not seem that changing the height has a great effect on the accuracy of the classification, provided that both training and test images are taken at the same height.

4. Conclusions

We introduced a new feature in this paper. This study concluded that they could work well with the human detection classification techniques on drone images. Replacing the Sobel filter with a new one, the proposed filter (AdSRHOG) showed more accurate than SRHOG in reducing the number of false detections. To compare the performance of the proposed feature with that of SRHOG, several experiments were carried out in which humans appeared in drone images under different posing and lighting conditions. The experiments were carried out using different datasets. Table 9 illustrates a summary of the results discussed in the previous sections. As suggested by this table (Table 9), it can be concluded that despite some weakness in detecting humans appearing in some difficult cases, the proposed feature outperforms SRHOG in all scenarios. This study shows a 54.3% improvement in Recall_neg, which means the number of false labels is substantially lower than that of the figure obtained using SRHOG. Also, the Recall value that represents the overall accuracy of an algorithm is increased by 13.1%.

It was observed that some of the pose and lighting conditions are the main bottleneck. The results relating to the former may be enhanced by the extension of the training dataset, while contrast enhancement techniques could improve the latter. Both issues are proposed to be explored in future studies. Besides, in the experiments presented, the effect of drone view angle on the results was not studied. This is also an important and interesting issue to be looked at in future research activities.

Data and code availability

Data and code are available upon request

Disclosure statement

The authors declare no conflict of interest.

Table 9. Summary of the results.

| Dataset                        | AdSRHOG      | SRHOG        |
|-------------------------------|--------------|--------------|
|                               | Recall | Recall_neg | Precision (%) | Recall | Recall_neg | Precision (%) |
| INRIA                         | 0.9076 | 0.8808 | 0.9498 | 0.7513 | 0.3333 | 0.7369 |
| Drone                         | 0.8000 | 0.8500 | 0.8421 | 0.8667 | 0.3333 | 0.5652 |
| Data Containing different human poses | 0.7159 | 0.3914 | | |
| Overall                       | 0.7747 | 0.8771 | 0.9787 | 0.5072 | 0.3333 | 0.8474 |
Acknowledgement

The authors appreciate the GeoAI Smarter Map and LiDAR Lab of the Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University (SWJTU), for the technical research support. This work is the outcome of one of the joint research studies with international University collaborators and has not been published or submitted previously.

Funding

This research received funds for the publication of the research outcome from the start-up funds (A1920502051907-6), Faculty of Geosciences and Environmental Engineering (FGEE), Southwest Jiaotong University (SWJTU), China.

ORCID

Masood Varshosaz http://orcid.org/0000-0002-2703-1515
Saied Pirasteh http://orcid.org/0000-0002-3177-037X

References

Aggarwal JK, Cai Q. 1999. Human motion analysis: a review. Comput Vision Image Understanding. 73(3):428–440.
AltiGator European Drone Manufacturer Company. 2017. Drones for search & rescue missions; [accessed 19 November 2020]. http://altigator.com/drones-for-searchrescue.missions.
Bahri H., Chouchene M., Sayadi F.E., Arti M. 2020. Real-time moving human detection using HOG and Fourier descriptor based on CUDA implementation. J Real-Time Image Proc 17, 1841–1856 (2020). https://doi.org/10.1007/s11554-019-00935-1.
Benenson R, Omran M, Hosang J, Schiele B. 2014. Ten years of pedestrian detection, what have we learned?. European Conference on Computer Vision. Cham: Springer; p. 613–627.
Blondel P, Potelle A, Pegard C, Lozano R. 2013. How to improve the HOG detector in the UAV context. 2nd IFAC Workshop on Research, Education and Development of Unmanned Aerial Systems (RED UAS 2013), Nov 2013, Compiègne, France. 46(30) :46–51.
Bravo RZB, Leiras A, Cyrino Oliveira FL. 2019. The use of UAV s in humanitarian relief: an application of POMDP-based methodology for finding victims. Prod Oper Manag. 28(2): 421–440.
Cao X, Wu C, Yan P, Li X. 2011. Linear SVM classification using boosting HOG features for vehicle detection in low-altitude airborne videos. 2011 18th IEEE International Conference on Image Processing. IEEE; p. 2421–2424.
Cortes C, Vapnik V. 1995. Support-vector networks. Mach Learn. 20(3):273–297.
Cutler R, Davis L. 1998. View-based detection and analysis of periodic motion. Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No. 98EX170). Vol. 1. IEEE; p. 495–500.
Dalal N, Triggs B. 2005. Histograms of oriented gradients for human detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05). Vol. 1. IEEE; p. 886–893.
De Cubber G, Doroffei D, Serrano D, Chintamani K, Sabino R, Ourevitch S. 2013. The EU-ICARUS project: developing assistive robotic tools for search and rescue operations. 2013 IEEE international symposium on safety, security, and rescue robotics (SSRR). IEEE; p. 1–4.
Doherty P, Rudol P. 2007. A UAV search and rescue scenario with human body detection and geolocalization. In: Orgun M.A., Thornton J. (eds) AI 2007: Advances in Artificial
Intelligence. AI 2007. Lecture Notes in Computer Science, vol 4830. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-76928-6_1

Dollár P, Babenko B, Belongie S, Perona P, Tu Z. 2008. Multiple component learning for object detection. European Conference on Computer Vision. Berlin, Heidelberg: Springer; p. 211–224.

Gavrila DM. 1999. The visual analysis of human movement: a survey. Comput Vision Image Understanding. 73(1):82–98.

Gavrila DM, Philomin V. 1999. Real-time object detection for “smart” vehicles. Proceedings of the Seventh IEEE International Conference on Computer Vision. Vol. 1. IEEE; pp. 87–93.

Ghasemi M, Varshosaz M, Pirasteh S. 2020. Evaluating Sector Ring Histogram of Oriented Gradients filter in locating humans within UAV images. ISPRS The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Nice, France, 14–20 June, 43, p. 23–27.

Hsu CW, Lin CJ. 2002. A comparison of methods for multiclass support vector machines. IEEE Trans Neural Netw. 13(2):415–425. doi: 10.1109/72.991427

Ioffe S, Forsyth DA. 2001. Probabilistic methods for finding people. International Journal of Computer Vision. 43(1):45–68.

INRIA Person Dataset. 2020. Last accessed 20 November 2020. http://pascal.inrialpes.fr/data/human/.

IRAN Red Crescent. 2004., Rescuers’s website. Principles of search and rescue; [accessed 19 November 2020]. http://www.emdadgar.com/special/nejat/sar/osool.htm.

Jurecka M, Niedzielski T. 2017. A procedure for delineating a search region in the UAV-based SAR activities. Geomatics Nat Hazards Risk. 8(1):53–72.

Leibe B, Seemann E, Schiele B. 2005. Pedestrian detection in crowded scenes. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05). Vol. 1. IEEE; p. 878–885.

Liu B, Wu H, Su W, Sun J. 2017. Sector-ring HOG for rotation-invariant human detection. Signal Process Image Commun. 54:1–10. doi: 10.1016/j.image.2017.02.008

Lowe DG. 2004. Distinctive image features from scale-invariant keypoints. Int J Comput Vis. 60(2):91–110.

Lygouras E, Gasteratos A, Tarchanidis K, Mitropoulos A. 2018. ROLFER: a fully autonomous aerial rescue support system. Microprocess Microsyst. 61:32–42.

Malik A, He P, Wang H, Khan AN, Pirasteh S, Abdullahi SM. 2020. High-capacity reversible data hiding in encrypted images using multi-layer embedding. IEEE Access. 8: 148997–149010.

Mihcioglu ME, Alkar AZ. 2019. Improving pedestrian safety using combined HOG and Haar partial detection in mobile systems. Traffic Inj Prev. 20(6):619–623.

Munder S, Gavrila DM. 2006. An experimental study on pedestrian classification. IEEE Trans Pattern Anal Mach Intell. 28(11):1863–1868.

Nguyen DT, Li W, Ogunbona P. 2009. A part-based template matching method for multi-view human detection. 2009 24th International Conference Image and Vision Computing New Zealand. IEEE; p. 357–362.

Nguyen DT, Li W, Ogunbona PO. 2011. Local intensity distribution descriptor for object detection. Electron Lett. 47(5):321–322.

Ojala T, Pietikainen M, Maenpaa T. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Machine Intell. 24(7):971–987.

Olson CF, Huttenlocher DP. 1997. Automatic target recognition by matching oriented edge pixels. IEEE Trans Image Process. 6(1):103–113.

Ott P, Everingham M. 2009. Implicit color segmentation features for pedestrian and object detection. 2009 IEEE 12th International Conference on Computer Vision, Kyoto. IEEE; p. 723–730. doi: 10.1109/ICCV.2009.5459238

Pang Y, Yuan Y, Li X, Pan J. 2011. Efficient HOG human detection. Signal Process. 91(4):773–781.
Sabzmeydani P, Mori G. 2007. Detecting pedestrians by learning shapelet features. 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE; p. 1–8.

Silvagni M, Tonoli A, Zenerino E, Chiaberge M. 2017. Multipurpose UAV for search and rescue operations in mountain avalanche events. Geomatics Nat Hazards Risk. 8(1):18–33.

Sun J, Li B, Jiang Y, Wen CY. 2016. A camera-based target detection and positioning UAV system for search and rescue (SAR) purposes. Sensors. 16(11):1778.

Takacs G, Chandrasekhar V, Tsai S, Chen D, Grzeszczuk R, Girod B. 2010. Unified real-time tracking and recognition with rotation-invariant fast features. IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Takacs G, Chandrasekhar V, Tsai SS, Chen D, Grzeszczuk R, Girod B. 2013. Fast computation of rotation-invariant image features by an approximate radial gradient transform. IEEE Trans Image Process. 22(8):2970–2982. doi: 10.1109/TIP.2012.2230011

Waharte S, Trigoni N. 2010. September. Supporting search and rescue operations with UAVs. 2010 International Conference on Emerging Security Technologies. Canterbury. IEEE; p. 142–147. doi: 10.1109/EST.2010.31

Walk S, Majer N, Schindler K, Schiele B. 2010. New features and insights for pedestrian detection. 2010 IEEE Computer society conference on computer vision and pattern recognition. IEEE; p. 1030–1037.

Wójcikowski M. 2016. Histogram of oriented gradients with cell average brightness for human detection. Metrol Meas Syst. 23(1):27–36.

Wu B. 2008. Part based object detection, segmentation, and tracking by boosting simple shape feature based weak classifiers [Ph.D. thesis]. University of Southern California, ProQuest Dissertations Publishing; p. 192.

Wu B, Nevatia R. 2005. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. Tenth IEEE International Conference on Computer Vision (ICCV’05). Vol. 1. IEEE; p. 90–97.

Yeong SP, King LM, Dol SS. 2015. A review on marine search and rescue operations using unmanned aerial vehicles. Int J Mar Environ Sci. 9(2):396–399.

Zhang HB, Zhang YX, Zhong B, Lei Q, Yang L, Du JX, Chen DS. 2019. A comprehensive survey of vision-based human action recognition methods. Sensors. 19(5):1005.

Zhang J, Wu X, Hoi SC, Zhu J. 2020. Feature agglomeration networks for single stage face detection. Neurocomputing. 380:180–189.