Long Distance Visual Ground-Based Signaling for Unmanned Aerial Vehicles

Volker Grabe and Stephen T. Nuske

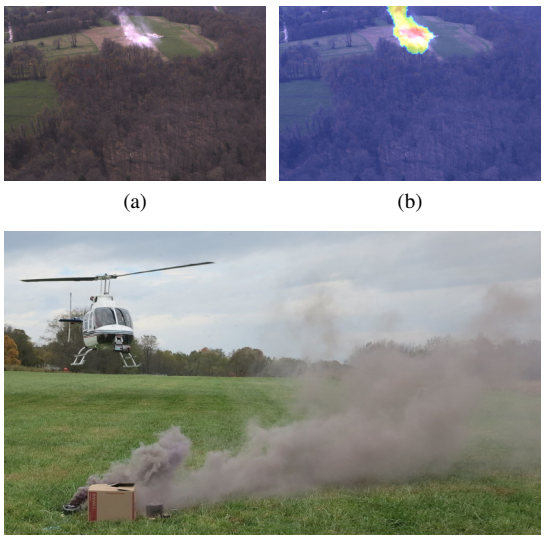
***Abstract*— We present a long-range visual signal detection system that is suitable for an unmanned aerial vehicle to find an optical signal released at a desired landing site for the purposes of cargo delivery or rescue situations where radio signals or other communication systems are not available or the wind conditions at the landing site need to be signaled. The challenge here is to have a signal and detection system that works from long range ($>$ 1000m) amongst ground clutter during various seasonal conditions on passive imagery. We use a smoke-grenade as a ground signal, which has the advantageous properties of being easy to carry by ground crews because of its light weight and small size, but when released has a long visual signaling range. We employ a camera system on the UAV with a visual texture feature extraction approach in a machine learning framework to classify image patches as 'signal' or 'background'. We study conventional approaches and develop a visual feature descriptor that can better differentiate the appearance of the visual signal under varying conditions and, when used to train a random-forest classifier, outperforms commonly used feature descriptors. The system was rigorously and quantitatively evaluated on data collected from a camera mounted on a helicopter and flown towards a plume of signal smoke over a variety of seasons, ground conditions, weather conditions, and environments. Our system was capable of detecting the smoke cloud with both precision and recall rates greater than 0.95 from ranges between 1000m and 1500m. Further, we develop a method to estimate wind orientation and approximate wind strength by assessing the shape of the smoke signal. We present a preliminary evaluation of the wind estimation in conditions with different wind intensities and orientations relative to the approach direction.**

I. Introduction

Unmanned Aerial Vehicles (UAVs) with vertical take-off and landing abilities have proven to be ideal search and rescue or cargo delivery vehicles that can operate in environments and conditions that are too dangerous for human rescue forces. Most prominently UAVs were used following the 2011 Fukushima Daiichi nuclear disaster as well as after several recent earthquakes all around the world, replacing manned helicopters. With the increasing availability of larger UAVs up to full-scale helicopters, even the autonomous delivery of goods to satisfy urgent needs in remote and isolated areas is within reach.

In these situations, ground personnel often does not have a reliable communication mechanism to guide the helicopter to a designated landing zone or drop-off point. Target coordinates are either approximate or not available at all. Under these conditions, visual signals are often the only alternative to indicate a desired landing zone at long range. However, searching for portable and therefore small visual landmarks can be difficult as the search area is large and often the vehicle is at a far distance from the objects it is trying to detect.

V. Grabe and Stephen T. Nuske are with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A vgrabe, snuske@andrew.cmu.edu



(a) (b)

(c)

Fig. 1: Visual homing signal detection from an aerial perspective at long distances on the example of signal smoke. (a) Aerial view from a distance of 1000m, (b) generated signal smoke segmentation heat map overlay from the proposed detection system (blue: low likelihood, red: high likelihood). (c) Helicopter approaching landing side marked by signal smoke.

To attract the attention of aircraft from the ground, a wide abundance of signaling devices has been proposed by rescue forces and the military [1]. Signal mirrors rely on direct sun light, flare guns provide only limited information on the source of the signal, and portable fiducial markers such as the bright orange 'VS 17' signal panels are either portable but relatively small and therefore difficult to detect at long range or large and impractical in the field.

Signal smoke from smoke grenades, on the other hand, provides a large visual target while not obscuring its origin. Additionally, signal smoke provides a visible clue of the local wind conditions in the landing zone, a very important factor in the planning of a safe final approach path. Different colors can be used to signal further information.

While large visual signals, when used as a visual homing beacon, can facilitate a quick and safe landing, its early detection and interpretation currently requires the trained eye of an on-board crew. In particular, the autonomous

detection of smoke signals on images from an unknown distance poses several additional challenges. Texture, color, and shape of a smoke plume are neither well defined nor static over time. Depending on the environmental conditions, the smoke texture can change dramatically from a large area of translucent haze to a thin stripe of thick clouds. Similarly, the appearance of the surrounding landscape varies with seasons and illumination conditions. The high speed of the camera on an aircraft and the therefore constantly changing background pose an additional challenge to the problem. To this end, we propose an autonomous visual smoke signal detection system that is robust to a large variety of environmental conditions, as visualized in Fig. 1.

In particular, we propose to train an autonomous classification system to effectively cover the various appearances of signal smoke plumes and backgrounds observed from an aerial perspective. To this end, we develop a supervised machine learning framework and compare a proposed radial feature descriptor to established descriptors commonly used for classification purposes using unbiased accuracy measures and cross-validation.

Based on the localization of the smoke plume, we will present a basic approach to also estimate the wind direction and intensity. The entire system is extensively evaluated based on data captured in a variety of seasonal, light, and environmental conditions using a helicopter-mounted camera flown at realistic flight speeds and trajectories.

The remainder of this paper is structured as follows. In Section I-A, we discuss existing work in relation to our proposed approach that is detailed in Sec. II. The system is experimentally evaluated and discussed in Sec. III. Finally, this work is concluded in Sec. IV.

A. Related Work

Homing and autonomous landing of UAVs has been accomplished using various fiducial marker such as an array of active LED-based beacons [2], concentric circles [3], two dimensional barcodes [4], and geometric patterns [5], [6], [7]. While the results of these approaches obtained on small aerial vehicles in constrained environments are very impressive, they do not scale well to the larger vehicles and distances addressed in this paper.

Long distance guidance of a helicopter towards a landing zone has been achieved using full-sized landing pad patterns [8]. Targets of this size and well defined structure are in use on almost all helicopter ports around the world since centuries, but, naturally, are not very suitable for quick deployment in the field.

The particular problem of smoke detection has been addressed mostly in the context of early fire warning systems. This prior work can be loosely grouped into two categories: (i) smoke detection in images from static cameras for, e.g., the purpose of forest fire detection from towers and (ii) smoke detection as a by-product of dynamic scene segmentation.

Work from the first category typically relies on the assumption of a static scene observed from a static camera and the availability of a smoke-free ground-truth image [9], [10], [11], [12]. Furthermore, the absence of other dynamic scene elements but smoke is assumed in most approaches.

Occasionally, the assumption of a known smoke color at given illumination conditions against a strong contrast gradient is exploited to segment smoke by color values [13]. While we had promising initial results using color segmentation, this approach did not generalize well with natural seasonal appearance changes or even the small color variations between different brands of smoke grenades.

The second category consists of work that is primarily dedicated to detect dynamic elements within a scene, such as, running water, waving leaves, or smoke. To this end, spatial temporal filters have been suggested [14], [15]. These approaches have been shown to even segment two overlaying dynamic scene elements from each other, but they do lack a specific categorization of the detected dynamic image components. Furthermore, the assumption of smoke as a highly dynamic scene element requires a close proximity of the camera to the scene.

None of these specific smoke detection approaches can be directly adopted for the detection of individual signal smoke plumes from a fast moving aircraft and over long distances as both concepts assume smoke as the predominant dynamic element in an otherwise mostly static scene.

In the last years, various machine learning techniques have become the standard in efficient real-time object detection and scene classification approaches. In the most basic case of a scene segmentation regression, the probability of an image region to belong to a certain class is estimated. To this end, the classifier is first trained using a training set of image patches with known class annotations.

In practice, segmentation of an entire image requires the evaluation of the trained classifier on many small patches of the image if not even every single pixel. To this end, it has been shown that a two staged classification system that first computes a low-bandwidth descriptor for each image patch which is then feed to the classifier can produces best results [16]. In this work, we have concentrated on the use of Random Forests [17] while Support Vector Machines (SVN) [18] and Convolutional Neuronal Networks (CNN) [19] are other established classification algorithms. Generic visual feature descriptors are SIFT [20] and SURF [21] while descriptors can be designed or trained for more specific purposes as well [16], [22].

In this work, we will rely on a random forest machine learning back-end for its ease of use and small number of system parameters. We will further use a novel feature descriptor that is inspired by both the SIFT and SURF descriptor which we found to be best suited for the given application as shown in the evaluation.

II. Approach

In the following, we describe the three major components of the proposed signal detection framework: (i) the machine learning classifier, (ii) the novel radial histogram feature descriptor, and (iii) a post processing step to threshold and

filter the output of the classifier obtained in (i) for entire images.

A. Machine Learning Framework

We used fully-supervised machine learning with rigorous cross-validation to train our system and achieve unbiased measures of accuracy; in particular we focus on precision and recall. We begin by first annotating images of the visual signal, e.g. smoke plumes in our case, obtained during several helicopter flights over the course of a year with the Sloth[1] labelling toolkit. Typically, the smoke density gradually decreases with increasing distance from the origin of the plume. Consequently, the plume has different levels of transparency and its appearance is therefore constantly changing. It is difficult to decide at what point to no longer consider the sparse regions of the plume as being a good representation of smoke appearance. We adopt the policy to only label regions of the smoke plume that are at least 50% opaque, which is to say that the smoke is more dominant than the background behind.

Once labelled, we parse this supervised training data and randomly select equally sized square image patches from both 'background' and 'signal' regions. We predefine this patch size and found that a size of 32×32 pixels works best for the scenario in consideration. It was ensured that a similar number of 'signal' and 'background' training patches was sampled from each image.

Once a sufficient number of patches has been obtained from both classes, the size of the positive and negative sample set is equalized. Equalization of the sets is done to not implicitly define an a priori probability of smoke observations during learning. This is particularly important since the probability of smoke observations naturally increases with decreasing distance towards the expected signal location. Entirely white or black patches caused by over or under illumination do not contain information and are discarded from both sets. Consequently, we extract a feature description for each patch that is a numerical description of texture (see Sec. II-B) and use these descriptors to train a random forest machine learning classifier which forms the appearance model of the signal. Assigning 'signal' patches a training label of 1 and 'background' patches a label of 0 allows us to create a regression that assigns each image patch a likelihood score in the interval $[0, 1]$. Using a 'leave one out' validation, the trained classifier is than evaluated on a test dataset that has not been used for training. The random forest will provide a confidence measure that gives a means to determine the optimal threshold on the precision/recall curve for detecting the visual signal.

B. Radial Histogram Feature Descriptor

Typically, the information contained in each image patch is compressed into a feature descriptor before being used to train a detector. This approach helps to avoid over-fitting and reduces the required memory as well as training and testing time of the detector. With the description of a patch being a lossy compression of the full information contained within a patch, the choice of a suitable descriptor that emphasizes the nature of the feature in question is critical.

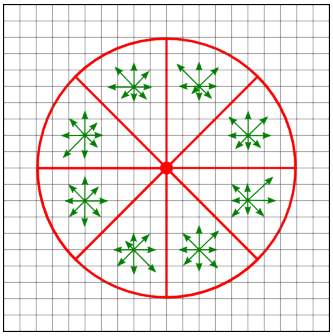


Fig. 2: Visualization of the Radial Histogram of oriented Gradients (RadHoG) descriptor. The image patch is divided into 8 circle segments (red). For each segment, an 8-bin gradient orientation histogram is computed (green arrows). The unity of all histogram bins forms the 64-byte descriptor.

Signal smoke in particular is characterized by smooth gradients while other features, such as color or shape, can vary significantly. In an effort to keep this work easily adaptable to other use cases, we avoided the use of manually engineered or trained classifiers that would have added another layer of specialization. However, a generic descriptor for the application in smoke detection should benefit from being based on gradients. Established examples of gradient-based descriptors are SIFT [20] and SURF [21].

We propose a generic Radial Histogram of Gradients (RadHoG) feature descriptor that combines the advantages from both SURF and SIFT into a new radial morphology. The purely circular design of the descriptor is intended to capture the gradient characteristics of the patch entirely with respect to its center while the traditional rectangular design of the SURF and SIFT descriptor covers the neighbourhood of the feature evenly. Furthermore, their rectangular design arbitrarily over-represents the diagonals of the sample patch.

To obtain the RadHoG descriptor, a gradient orientation and a gradient magnitude image are computed for each patch. Only the pixels within a circular region around the feature's origin are considered. The circle is divided into 8 equal segments and for each segment, a gradient orientation histogram with 8 bins is computed. The computation of the descriptor is visualized in Fig 2. Each pixel within a segment contributes to the corresponding gradient orientation bin with its gradient magnitude. All contributions to the gradient orientation histogram bins are further scaled by a distance dependant Gaussian weight to reduce the influence of pixels further from the center. This procedure results in a 64-byte descriptor.

Additional rotational invariance could be obtained by a representation relative to the circular segment with the strongest gradient magnitude or the predominant gradient of the entire patch. However, in this application, the diversity of the training set and the almost constant orientation of

[1] http://sloth.readthedocs.org

the camera with respect to the scene eliminates the need for rotational invariance.

C. Image Segmentation and Postprocessing

To segment entire images, we extract and classify descriptors for 32×32 pixel patches at every corner of a 10 pixel grid across the image. This process results in a initial signal likelihood map that is further filtered to refine the results. To this end, we first threshold the likelihood map to suppress low signal likelihood areas in the image and obtain a score map. The threshold is chosen to maximizes the precision at a desired recall level. Next, a uniform kernel is applied to compute the average signal score in the neighbourhood of each pixel. The resulting map is again threshold to obtain a final heat map. For visualization purposes, we scale the size of the heat map back to the original size of the image.

D. Wind Estimation

After localization of the smoke signal within an image, we perform analysis of the shape and orientation of the smoke plume to provide a coarse estimate of the wind direction. A binary classification of the velocity is established to asses the influence of the wind speed on safe flight operations. Estimating wind at the landing site is essential both because the wind on the ground may be different to the wind at altitude and also because some UAVs do not have a wind estimate from their onboard sensing systems at all.

The approach is to first identify the set of pixels which define the detected smoke region. Using the set of pixel locations in the smoke region, we perform Principal Components Analysis (PCA). The mean from the PCA provides the center coordinate (c_x, c_y) of the region. Then the orientation of the cloud is computed as the relationship between the x and y axis of the eigenvector e_x and e_y. The elongation of the cloud is computed as the ratio between the first and second eigenvalues v_x and v_y:

$$\theta = \arctan \frac{e_y}{e_x} \tag{1}$$

$$elongation = |\frac{v_x}{v_y}| \tag{2}$$

Using the central point of the region and the principal axis of the smoke plume, we compute the origin of the smoke. Our assumption is that the smoke plume is thinner at the origin and expands as it moves away from the grenade canister. Using this assumption, we transform all the points to a coordinate frame defined by the principal axis of the smoke plume:

$$\hat{p}_x = (p_x - c_x)\cos\theta - (p_y - c_y)\sin\theta \tag{3}$$

$$\hat{p}_y = (p_x - c_x)\sin\theta + (p_y - c_y)\cos\theta \tag{4}$$

Then, we collect all points to the left of the minor axis of the region $\hat{p}_x < 0$ and all the points to the right of the axis $\hat{p}_x >= 0$. We then again perform PCA on the left region of points and the right region separately. We compute the elongation of the left and right region as follows:

$$elongation_l = |\frac{v_x^l}{v_y^l}| \tag{5}$$

$$elongation_r = |\frac{v_x^r}{v_y^r}| \tag{6}$$

The origin is determined by the larger of $elongation_l$ and $elongation_r$, hence the thinner section of the region likely corresponds to the source of the smoke.

The velocity of the wind is approximated from the value of $elongation$ and discretized into two wind condition categories, strong and calm, which respectively correspond to wind speeds that do or do not affect flight operations. Values of the dimensional ratio $elongation$ greater than 10 are considered strong wind and were found to require special attention for save landings. Lower ratios are associated with calm conditions and landing is possible without consideration of the wind direction. Further work is required to map the separation between the two categories to actual flight requirements. See Fig.3 for an illustration of the entire method.

Once the wind vector in camera coordinates has been obtained, the wind vector can be projected onto the ground plane by usage of a GPS/INS state estimate and intersection of the ray towards the smoke with a Digital Elevation Map (DEM) of the environment where available. This enables us to calculate an approximation of the wind vector in world coordinates.

III. Experimental Evaluation

We devised a series of experiments to test the various aspects of the described smoke signal detection algorithm and to quantify its performance.

A. Experimental Equipment

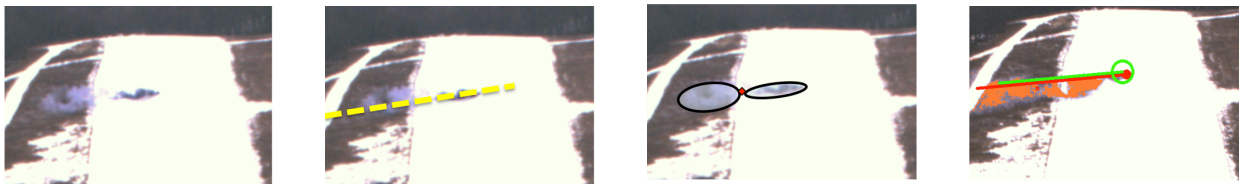
All datasets were collected using a sensor unit mounted to a Bell 206 LongRanger helicopter.

The sensor unit consisted of an Allied Vision Prosilica GT 3400 color camera for image collection with 10 megapixel resolution (3384×2704) using a Kowa LM16SC 16mm lens that provides a 43° field of view. Also integrated is a Novatel Span IMU/GPS system to obtain ground truth pose data for the helicopter. A similar multi-antenna GPS system was used to accurately measure the position of the smoke pot.

The smoke signals used were commercially available Enola Gaye EG-18 smoke pots. These units are typically used for theater, film, photography, and on paintball sites and produce a much lower volume of smoke than those used for aviation purposes. For legal reasons, we were unable to deploy aviation smoke grenades and therefore simultaneously deployed 7 EG-18 smoke pots to approximate the volume of smoke used in aviation smoke grenades.

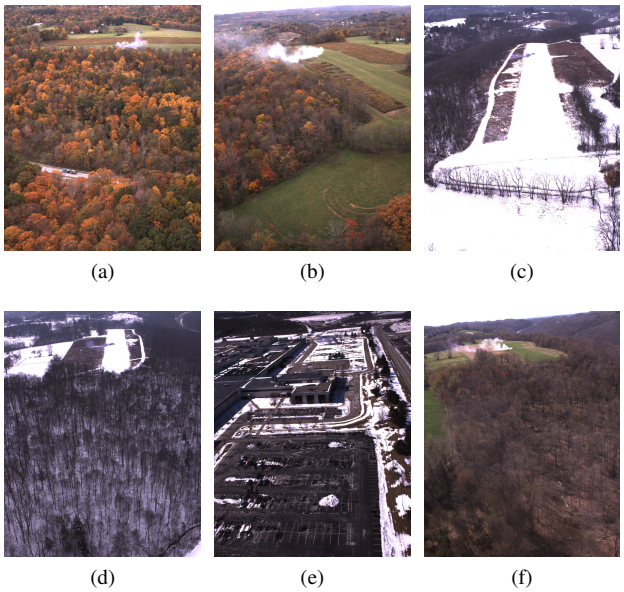
B. Experimental Data Collection

With this setup, we flew multiple approaches over the period of 12 months with the aim to cover a variety of conditions whilst collecting imagery and positioning data.



(a) (b) (c) (d)

Fig. 3: Example of the wind estimation algorithm. (a) Raw image. (b) Principal axis of wind region as estimated by PCA analysis. (c) Left and right regions of the smoke cloud analyzed to find the origin. Origin of cloud is the with more elongated side. (d) Origin highlighted with a circle, orientation shown as a line, and wind strength indicated by the length of the line (strong wind in this case).



(a) (b) (c)

(d) (e) (f)

Fig. 4: Sample images from different approaches towards a signal smoke plume. Each of the shown image samples was taken from a distance of 600m to the origin of the smoke. The conditions during each flight are documented in Table I. The dataset shown in (f) was used exclusively for testing.

In particular, we obtained a data collection that covers a substantial range of conditions that affect the visual appearance of both the environment and the smoke signal. Snow, foliage, and summer affect the background appearance while sunlight, cloud, and wind conditions affect the smoke plumes color and morphology. A representative subset of the database that we collected is shown in Fig. 4 and Table I. The example images shown in Fig. 4(a) to 4(e) stem from datasets that were used to train the machine learning framework. The dataset from which the image in Fig. 4(f) was taken was used exclusively for testing and evaluation. In particular, to test for seasonal invariance, no other dataset obtained under summer conditions was used to train the detector.

C. Comparison of Feature Descriptors

We compared the proposed radial histogram of gradients feature descriptor to several established descriptors, in particular the related SIFT and SURF descriptors. Furthermore, we validated our approach against two other descriptors: a 2-byte descriptor that holds the hue and saturation value

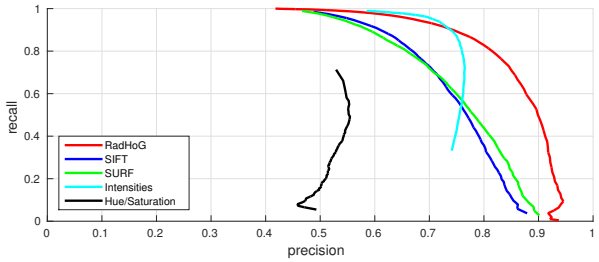


Fig. 5: Precision/recall rate averaged over 5 independent training/test cycles on identical training sets for the different feature descriptors in comparison.

averaged over the patch and a compression-free 32×32 byte linear concatenation of all pixel intensities of the patch. The hue/saturation descriptor is similar in design to descriptors that are typically used to train dense pixel-level scene segmentation systems, such as, skin detectors [23].

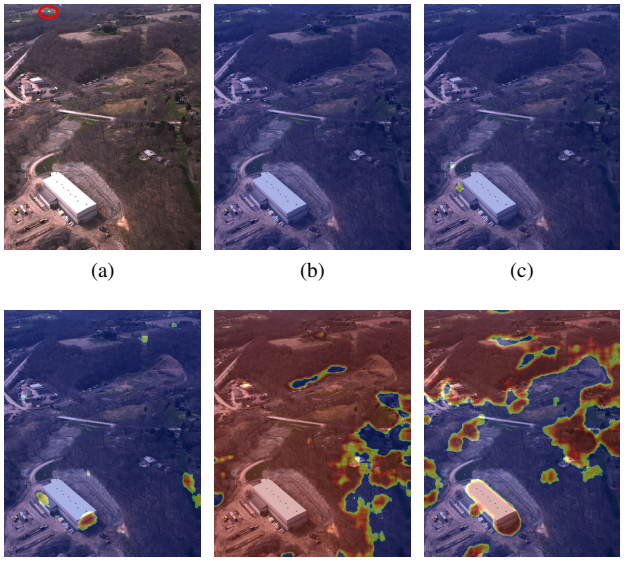
Precision/recall curves for all descriptors in comparison can be found in Fig. 5. These results were obtained by testing the labelled 32×32 sized patches of the balanced test set with the trained classifier. The curves show that a pure color/intensity based descriptor is not suitable to segment multi-colored smoke under a high variety of conditions and illuminations. Similarly, the uncompressed descriptor of all intensities cannot achieve the performance obtained with any of the more sophisticated descriptors. For this application, the proposed RadHoG descriptor outperforms the established SIFT and SURF descriptors significantly.

In the case of signal smoke detection and the assumption of only one such signal in the scene, one does not necessarily need to detect the smoke cloud in every image and as such a low recall rate can be acceptable. On the other hand, false detections have to be kept at a minimum to avoid confusion of the pilot in control or, in a fully automated system, a down-stream state estimation filter. Therefore, we decided to choose the classification threshold such that the precision rate is maximized at a recall rate of at least 15%. The quality of a descriptor is consequently judged by practical performance measures on the individual images of the test dataset.

While the quantitative results shown in Fig. 5 help to evaluate the performance of the different descriptors in the machine learning component of the proposed framework, they do not include the evaluation of the refining post-processing steps on the likelihood maps. Therefore, we

TABLE I: Selection of flight conditions that were used to train (datasets a to e) and test (dataset f) the smoke detection framework. The dataset ID matches the sample images shown in Figure 4.

ID	Season	Environment	Weather	Wind	Cloud Cover	Smoke Color
a	Autumn	Airfield	Forested / coloring foliage	Mild wind	Light cloud cover	purple
b	Autumn	Airfield	Forested / coloring foliage	Mild wind	Light cloud cover	white
c	Winter	Airfield	Snow cover, forested / no-leaves	Windy	Heavy cloud cover	purple
d	Winter	Airfield	Snow cover, forested / no-leaves	Windy	Heavy cloud cover	purple
e	Winter	Urban	Snow cover, industrial buildings, car park	Mild wind	Sunny	purple
f	Summer	Airfield	Forested	Moderate wind	Sunny	white



(a) (b) (c)

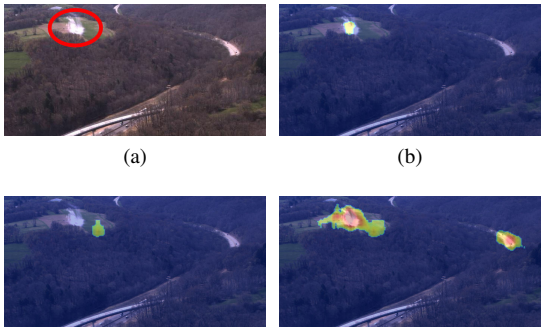
(d) (e) (f)

Fig. 6: Approach towards a smoke plume in 2440m distance with overlayed scene segmentation heat maps based on the five descriptors. Red represents a high, blue a low signal detection score. The smoke has not been detected at this distance using any of the descriptors. (a) Original image with smoke highlighted by red oval. (b) The proposed RadHoG descriptor does not show any false positive detections. (c) Using the SURF descriptor, two small false detections were found to the left of the warehouse. (d) The SIFT descriptor leads to several false high-confidence detections. (e, f) The results from the intensity and hue/saturation descriptor respectively do not provide any meaningful segmentation.

analyzed the smoke detection performance using the various classifiers also over the entire length of an approach towards the signal smoke.

Figure 6 shows the smoke segmentation results on a scene with a smoke plume in the far distance. The results illustrate the findings from the precision/recall curves in Fig. 5: the smoke was not detected using any of the tested detectors, yet did the classifier suggest the presence of two small and several larger smoke clouds with SURF and SIFT respectively as underlying descriptor, contributing to a low precision and recall rate for far distances. The segmentation obtained when using the two intensity based descriptors confirms the low performance indicated by the precision/recall curve. For the remainder of this paper, we will concentrate on a comparison of the gradient based descriptors RadHoG, SIFT, and SURF.

Figures 7 and 8 show a comparison of the smoke segmentation from a distance of 1270m and 288m respectively.



(a) (b)

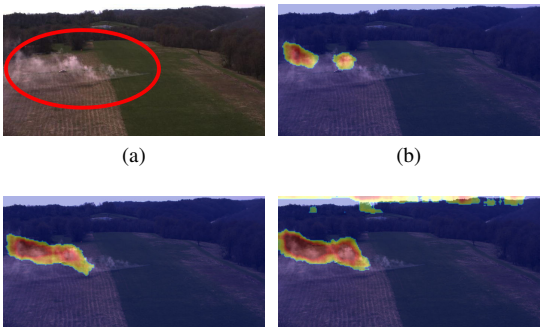
(c) (d)

Fig. 7: Smoke plume from a distance of 1270m. (a) Original image with highlighted smoke. (b) Using the RadHoG descriptor, the smoke cloud was optimally segmented. (c) The SURF detector leads to a false detection close to the real smoke. (d) SIFT produces a high rating on both the smoke and a road bend.

Figure 7 demonstrates that an unambiguous smoke signal detection during this important phase of the approach is only obtained with the suggested RadHoG descriptor. In particular, the SIFT descriptor fails to separate gradient patterns similar to smoke but with different magnitudes, such as, road bends, buildings, and trees alongside a bright field. Figure 8 shows the fading smoke plume from a distance of 288m just before landing in front of the smoke grenade. The RadHoG descriptor helps to segment the areas of dense smoke, as it was trained on 50% opaque smoke. Nevertheless, the SURF and SIFT descriptor allow for a dense labelling of almost the entire cloud. Again, several false detections along the horizon line occurred when the SIFT descriptor is used.

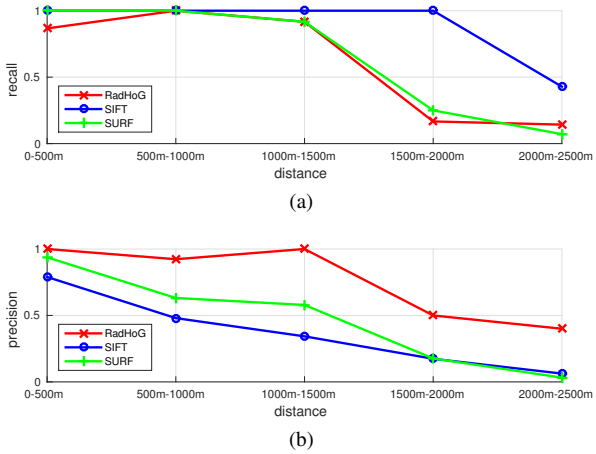
Our evaluations revealed that both precision and recall rates decrease with a growing distance from the smoke plume. These findings are visualized for a discrete distance histogram in Fig. 9. The recall rate with, in this case, only one true smoke cloud indicates the maximal distance at which the smoke has been detected using either of the descriptors. While the RadHoG and the SURF descriptor achieve similar results, the SIFT descriptor yielded a larger detection distance on this flight. However, the many false-positive detections using either the SURF or SIFT descriptor result in a poor precision rate compared to the proposed RadHoG descriptor.

Table II summarizes the quantitative results from all descriptors. The results show once more that, using the SIFT descriptor, we were able to detect the smoke cloud from a greater distance but at the expense of a very high false-



(a) (b) (c) (d)

Fig. 8: Fading smoke plume from a distance of 288m. (a) Original image with smoke circled. (b) Using the RadHog descriptor, only the more opaque parts of the smoke cloud are segmented. (c, d) SURF and SIFT descriptors lead to the segmentation of a larger portion of the cloud, but, in the case of SIFT, produce several outliers on the horizon line.



(a) (b)

Fig. 9: Detector performance over distance. Distances have been clustered into bins, representing 500m each. (a) Recall and (b) precision rate over distance from the smoke plume.

positive detection rate. On the other hand, the proposed RadHoG descriptor has been designed to provide very reliable estimates while accepting a higher false-negative rate.

D. Cross-Validation

Finally, we conducted a 'leave-one-out' cross-validation using each one of the autumn and winter datasets (ID (a) and (c) according to Table I) exclusively for testing while all other datasets were used for training. Table III lists the highest achievable precision rates at a recall rate above 15%. The results, when tested on the autumn dataset, conform with those found on the summer dataset, although all descriptors perform slightly better. However, all descriptors perform significantly worse on the winter dataset with the SURF descriptor leading the field. With snow covered ground, the strong wind makes the shape of the dark purple smoke plume indistinguishable from small rows of trees and bushes.

E. Timing

The time required for training the random forest with both 32,000 'signal' and 'background' sample patches as well as the time for the actual segmentation of an entire image can be found in Table II. All experiments were carried out on the CPU of common desktop hardware without any specific optimization. The training time varies with the complexity of the descriptor, but can be considered as a one time investment. With a given complexity of the machine learning framework, in this case a constant number and depth of trees in a random forest, the testing time is governed by the extraction of the descriptor. SURF has been reported to provide a faster feature extraction implementation than SIFT [21] which our results support. The complexity of the RadHoG descriptor is similar to the complexity of the SURF descriptor and we expect to achieve similar extraction times after optimization in future work.

TABLE III: Precision rates for a 'leave-one-out' cross-validation over different seasons.

Descriptor	Summer (ID: f)	Autumn (ID: a)	Winter (ID: c)
RadHoG	0.93	0.97	0.77
SURF	0.86	0.96	0.82
SIFT	0.84	0.94	0.71

F. Wind Estimation Results

We present preliminary evaluation results of the wind estimation system on approaches to smoke signals, assessing wind speed and approximate direction, see Fig. 10. Fig. 10a shows both the strength and orientation of a moderately-strong right-to-left crosswind being correctly estimated. Fig. 10b shows both the strength and orientation of a slight right-to-left crosswind being correctly estimated. Fig. 10c shows an example of an approach to a smoke signal in the presence of a light crosswind from right to left. In this example, the algorithm incorrectly estimated the conditions as a strong crosswind. Despite the incorrect wind velocity, the correct orientation was estimated as a crosswind from right to left. Fig. 10d shows an approach in calm conditions. The algorithm demonstrated that it could detect the conditions as being calm but the wind direction is often difficult to discern in calm conditions. However, in practice, the wind direction is less relevant in calm conditions as the wind does not require any special considerations in that case.

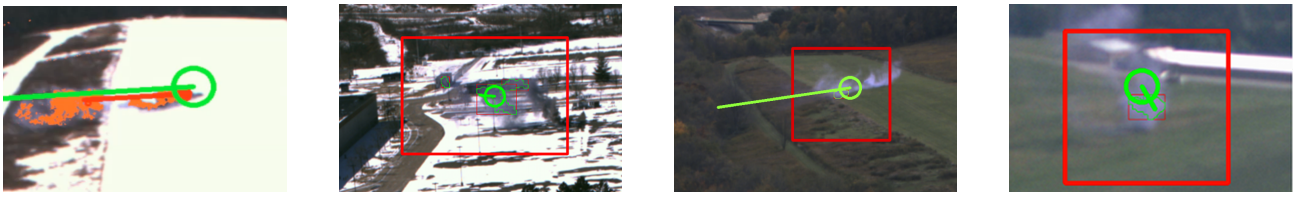
IV. Conclusions and Future Work

In this work, we developed an optical signal detection framework that is intended to allow UAVs to recognize and navigate towards visual distress signals in the absence of global localization capabilities. To this end, we trained a supervised regression system on a set of different feature descriptors. The proposed radial histogram of gradients descriptor was found to outperform the related traditional feature descriptors SIFT and SURF when cross-validated on an extensive dataset that spans various seasons, locations, and environmental conditions.

In future work, we plan to integrate the presented signal detector into the path planning system of an autonomous full-size helicopter. Therefore, we intend to continue to improve

TABLE II: Quantitative descriptor performance comparison. *The image segmentation obtained from the Hue/Saturation and the intensity descriptor did not allow for a meaningful analysis, see Figure 6.

Descriptor	Complexity	Maximal Precision	Initial Detection Distance	Consistent Detection Distance	False Detections	Training Time	Segmentation Time
RadHoG	64 byte	0.93	2160m	1650m	7	48s	2.8s
SURF	64 byte	0.86	2090m	1650m	46	43s	1.0s
SIFT	128 byte	0.84	2190m	2190m	180	74s	2.5s
Hue/Sat	2 byte	0.56	–*	–*	–*	3s	0.5s
Intensity	1024 byte	0.77	–*	–*	–*	462s	1.5s



(a) (b) (c) (d)

Fig. 10: Examples of wind estimation. The direction that the green line is pointing relative to the green circle indicates the orientation estimate, while the length of the line indicates approximate strength of the wind in two categories, calm and strong. (a) Successful estimation of both direction and wind strength. (b) The wind was correctly estimated as being a light crosswind. (c) The wind velocity was calm but the algorithm incorrectly estimated it as a strong crosswind. (d) The system correctly estimated calm wind conditions.

our texture feature descriptor and machine learning algorithms in terms of both accuracy and efficiency in particular under winter conditions. The collection of additional datasets with ground truth wind conditions will further allow for a quantitative evaluation of the wind estimation system.

V. Acknowledgments

This work was funded under U.S. Government Prime Award No. N00014-12-C-0671.

References

[1] United States Department of the Army, *U.S. Army Survival Manual*, 2003.

[2] K. E. Wenzel, A. Masselli, and A. Zell, "Automatic Take Off , Tracking and Landing of a Miniature UAV on a Moving Carrier Vehicle," *Journal of Intelligent & Robotic Systems*, vol. 61, no. 1, pp. 221–238, 2010.

[3] S. Lange, N. Sünderhauf, and P. Protzel, "Autonomous Landing for a Multirotor UAV Using Vision," in *Workshop Proceedings of the International Conference on Simulation, Modeling and Programming for Autonomous Robots*, Venice, Italy, 2008, pp. 482–491.

[4] P. Benavidez, J. Lambert, A. Jaimes, and M. Jamshidi, "Landing of an Ardrone 2.0 quadcopter on a mobile base using fuzzy logic," *World Automation Congress Proceedings*, vol. 1, no. 1-2, pp. 5–25, 2013.

[5] S. Saripalli, J. F. Montgomery, and G. Sukhatme, "Vision-based autonomous landing of an unmanned aerial vehicle," in *Proceedings of the International Conference on Robotics and Automation*, Washington, DC, USA, 2002, pp. 2799–2804.

[6] T. Merz, S. Duranti, and G. Conte, "Autonomous Landing of an Unmanned Helicopter based on Vision and Inertial Sensing," *Experimental Robotics IX*, vol. 21, no. c, pp. 343–352, 2006.

[7] S. Lin, M. Garratt, and A. Lambert, "Real-Time 6DoF Deck Pose Estimation and Target Tracking for Landing an UAV in a Cluttered Shipboard Environment using On-board Vision," in *Proceedings of the International Conference on Mechatronics and Automation*, Beijing, China, 2015.

[8] S. Arora, S. Jain, S. Scherer, S. Nuske, L. Chamberlain, and S. Singh, "Infrastructure-free shipdeck tracking for autonomous landing," in *Proceedings of the International Conference on Robotics and Automation*, Karlsruhe, Germany, 2013, pp. 323–330.

[9] D. Kim and Y. F. Wang, "Smoke detection in video," in *Proceedings of the World Congress on Computer Science and Information Engineering*, vol. 5, 2009, pp. 759–763.

[10] H. Tian, W. Li, L. Wang, and P. Ogunbona, "Smoke detection in video: An image separation approach," *International Journal of Computer Vision*, vol. 106, no. 2, pp. 192–209, 2014.

[11] B. U. Toreyin, Y. Dedeoglu, and A. E. Cetin, "Contour based smoke detection in video using wavelets," in *Proceedings of the European Signal Processings Conference*, Florence, Italy, 2006, pp. 6–10.

[12] T. Çelik, H. Özkaramanh, and H. Demirel, "Fire and Smoke Detection without Sensors: Image Processing," in *Proceedings of the European Signal Processing Conference*, Poznan, Poland, 2007, pp. 1794–1798.

[13] B. U. Töreyin, Yigithan Dedeoglu, and A. E. Çetin, "Wavelet Based Real-time Smoke Detection in Video," in *Proceedings of the European Signal Processings Conference*, 2005, p. 4.

[14] G. Doretto, D. Cremers, P. Favaro, and S. Soatto, "Dynamic Texture Segmentation," in *Proceedings of the International Conference on Computer Vision*, Los Angeles, CA, USA, 2003, pp. 1236–1242.

[15] D. Teney, M. Brown, D. Kit, and P. Hall, "Learning Similarity Metrics for Dynamic Scene Segmentation," in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2084–2093.

[16] A. S. Razavian, A. Hossein, S. Josephine, and C. Stefan, "CNN Features off-the-shelf : an Astounding Baseline for Recognition," in *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops*, 2014.

[17] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5–32, 2001.

[18] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[19] T. Homma, L. E. Atlas, and R. J. Marks II, "An Artificial Neural Network for Spatio-Temporal Bipolar Patters: Application to Phoneme Classification," *Advances in Neural Information Processing Systems*, vol. 1, pp. 31–40, 1988.

[20] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, nov 2004.

[21] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[22] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks," in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1717–1724.

[23] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," *Pattern Recognition*, vol. 40, no. 3, pp. 1106–1122, 2007.