Detection of Aircrafts on a Collision Course using Spatio-Temporal HOG*

Artem Rozantsev (artem.rozantsev@epfl.ch)

Mario Christoudias (mario.christoudias@epfl.ch)

Vincent Lepetit (vincent.lepetit@epfl.ch)

Pascal Fua (pascal.fua@epfl.ch)

School of Computer and Communication Sciences Swiss Federal Institute of Technology, Lausanne (EPFL)

Technical Report

March 5, 2013

*This work was supported by the EU funded myCopter project.

Abstract. We have developed a method for the detection of both generic flight neighbouring aircrafts and those on a collision course. Our approach employs a sliding window linear Support Vector Machine (SVM) classifier with a Histogram of Oriented Gradients (HOG) feature representation. An extension of this approach to the spatio-temporal domain is also considered and we demonstrate its advantage for the detection of aircrafts on a collision path. We evaluated our approach for the detection of both small rotorcraft and larger fixed-wing aircrafts in challenging video sequences. Our results show that aircrafts on a collision course can be detected more reliably than when assuming a generic flight path. This is very interesting in practice, since this case is of critical importance. We also show that our spatio-temporal approach improves the detection accuracy with respect to conventional single-frame approaches.

1 Introduction

The ability to detect and estimate the relative distance and bearing of neighboring aircrafts plays a crucial role in automated flight, central to such tasks as mid-air collision avoidance and formation flying. Vision-based relative positioning is of particular interest as cameras generally require less power consumption and are more lightweight than active sensor alternatives like radar and laser. It also has potential application in non-collaborative flight scenarios or in situations where GPS-based collision-warning systems are either unreliable or not commonly available on all aircrafts.

Vision-based relative positioning poses several challenges. Unlike other detection tasks a missed detection can be quite costly. A high detection accuracy is therefore required across a variety of operating conditions. Aircrafts travel at a high speeds and must be detected quickly particularly in collision avoidance scenarios. Also, neighbouring aircrafts can appear across a wide range of distances and can be difficult to detect when far away.

Of particular importance is the detection of aircrafts on a collision course for which highly accurate detection is a must. As Figure 1 shows, in such a case, the aircraft remains static in the video while increasing in apparent size. As demonstrated in our evaluations, this is a highly discriminative feature and we show how to exploit it to greatly increase detection accuracy.

In this work we developed a method for the detection of both generic flight neighboring aircrafts and those on a collision course. We adopt a

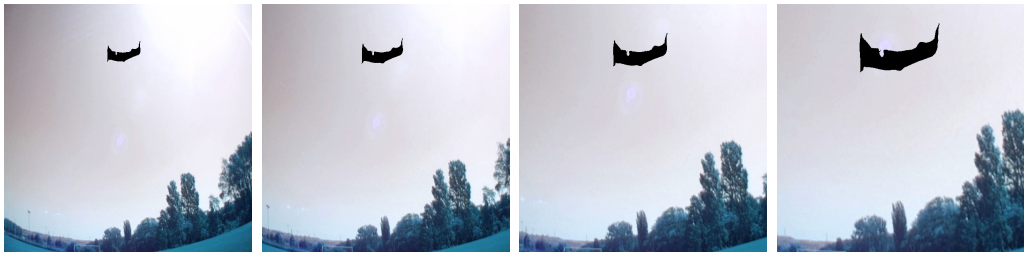


Figure 1: Image sequence of an aircraft on a collision course. Aircrafts that are on a collision path appear at a constant head and to be increasing in size.

detection framework similar to the one in [3] which has been demonstrated to perform well across a variety of detection tasks [4, 5, 6, 9]. An extension of this approach to function over the spatio-temporal domain, similar to [10], is then considered and we show its advantage for the detection of aircrafts on a collision course. We evaluated our approach for the detection of both small rotorcraft and larger fixed-wing aircrafts in challenging video sequences.

Our results show that:

- Aircrafts on a collision course can be detected more reliably than when assuming a generic flight path. This is very interesting in practice, since this case is of critical importance.
- Our spatio-temporal approach improves the detection accuracy with respect to conventional single-frame approaches.

The remainder of this report is organized as follows. Our aircraft detection algorithm and its spatio-temporal extension is first detailed in Section 2. We then present an evaluation of our approach in Section 3. The conclusion summarizes the method and discusses our quantitative results.

2 Aircraft Detection Algorithm

Our approach closely follows the detection framework of [3]. We first outline our classification method and then describe the image features we use. Finally, we discuss an extension for the detection of aircrafts on a collision course.

(a) (b)

Figure 2: Example images from the datasets used in our evaluation: (a) rotorcraft dataset and (b) fixed-wing dataset.

2.1 Support Vector Machine Classification

We employ a statistical learning algorithm for the detection of neighbouring aircraft. Provided a dataset of annotated images $\{\mathbf{x}_i, y_i\}_{i=1}^N$ as shown in Figure 2, with training image features $\mathbf{x}_i \in \mathcal{R}^d$ and binary labels $y_i \in \{-1, 1\}$, we seek to learn a linear classification function

$$f(\mathbf{x}) = \text{sgn}\left(\mathbf{w}^T\mathbf{x} + b\right) \tag{1}$$

where $\mathbf{w}$ and b define a separating hyperplane in $\mathcal{R}^d$ that segragates aircraft images from background images. In particular, an image $\mathbf{x}$ is classified as of an aircraft if $f(\mathbf{x})$ is positive and as background otherwise.

An optimal separating hyperplane can be found by minimizing the regularized hinge loss on the training data

$$\mathcal{L} = \min_{\mathbf{w},b} \sum_{i=1}^{N} \left[1 - (\mathbf{w}^T\mathbf{x}_i + b)y_i\right]_+ + \lambda\|\mathbf{w}\|_2^2 \tag{2}$$

where $[x]_+ = \max(0, x)$ is the hinge error function. It can be shown that minimizing Equation (2) yields a separating hyperplane of maximal *margin*, *i.e.*, the smallest distance between the hyperplane and any of the training examples. As shown in [1], Equation (2) can be expressed as a constrained

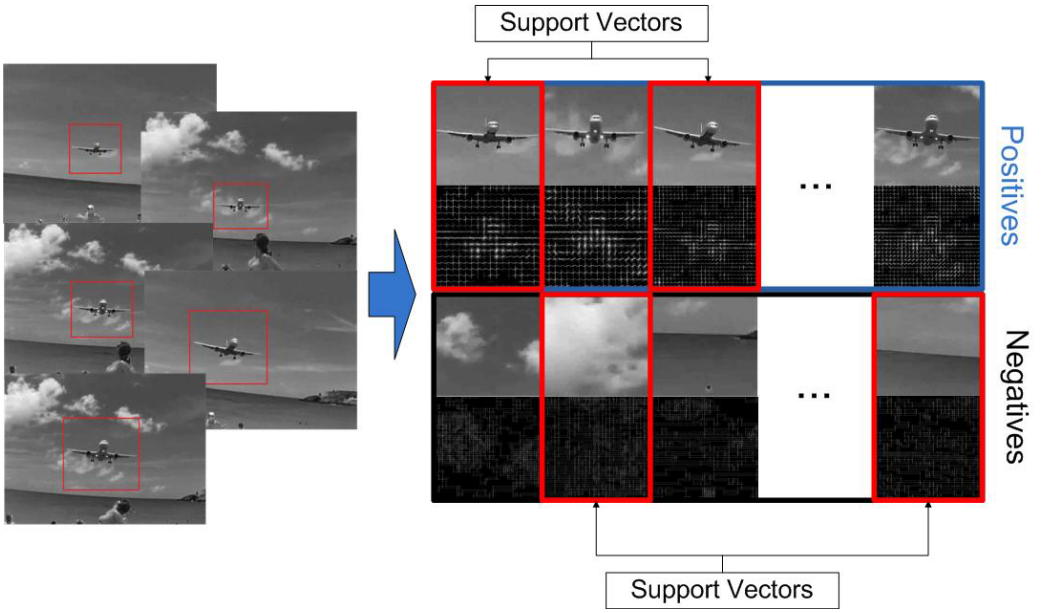


Figure 3: Learned SVM. Example aircraft and background training images used to learn the SVM are displayed along with their associated HOG features. The learned support vectors are highlighted in red.

minimization problem and solved in its dual form with respect to its Lagrange multipliers $\mathbf{a} \in \mathcal{R}^N$:

$$\tilde{\mathcal{L}} = \min_{\mathbf{a}} \sum_{i=1}^{N} a_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \tag{3}$$

subject to the constraints

$$0 \leq a_i \leq (2\lambda)^{-1}, \quad i = 1, \ldots, N \tag{4}$$

$$\sum_{i=1}^{N} a_i y_i = 0. \tag{5}$$

Given the optimal $\mathbf{a}$ that minimizes Equation (3), the linear classification function of Equation (1) can be re-expressed in terms of $\mathbf{a}$ as

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^{N} y_i a_i \mathbf{x}_i^T \mathbf{x} + b \right) \tag{6}$$

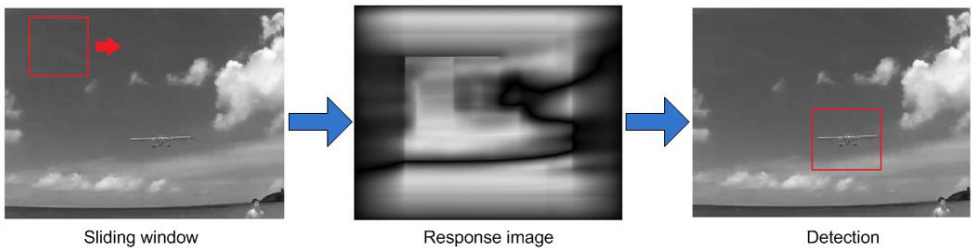


Figure 4: Sliding window detection. (left) The SVM is evaluated at different spatial locations and scales. (middle) At each scale a response image is computed from the sliding-window SVM. (right) Detections are found as local maxima in the response image whose value is above a detection threshold.

The training examples $\mathbf{x}_i$ that correspond to the non-zero a_i are referred to as the *support vectors* of the resulting maximum margin classifier and are those examples that lie along the margin of the learned classifier [1]. This classification algorithm called the *Support Vector Machine* (SVM) classifier was originally introduced by Vapnik [2] and has since found wide spread application.

Example images from our training dataset are shown in Figure 3 along with the learned support vectors. As illustrated in the figure, the support vectors represent distinctive features useful for differentiating aircraft from background images. The training examples were generated by annotating the extent of each aircraft using a bounding box and normalizing the resulting image to a canonical scale. The feature representation used to describe the normalized training patches is discussed in the following section.

Detection is performed by sliding a window across the image at multiple locations and scales, as illustrated in Figure 4. Detections are then found as those whose SVM score is above a threshold (b in Equation (1)) that is also a free parameter to our algorithm and whose value is typically cross-validated. In our evaluation, we show results obtained by varying this threshold.

2.2 Histogram of Oriented Gradients

We use the image gradient features of [3] as input to the SVM classifier. This representation bears close similarity to the widely used Scale Invariant Feature Transform descriptor [8] and has been further demonstrated to work well under challenging imaging conditions [3, 4, 5, 9].

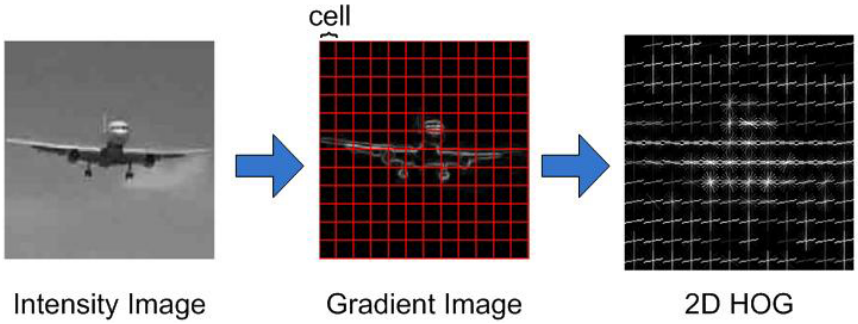


Figure 5: Histogram of Oriented Gradients (HOG) descriptor. Image gradients are binned according to their location and orientation about a regular grid of spatial cells. An illustration of the resulting description is shown on the right, gradient orientation responses are displayed as bold lines, with the thickness of the line indicating the strength of the response [3].

With our approach each image is represented using its local gradient statistics. This process is illustrated in Figure 5. A gradient feature histogram is computed by binning the image gradients according to their location and orientation over a predefined set of regularly spaced histogram *cells*. The cell size and count is a parameter to the representation, whose value is typically cross-validated. The resulting representation defines a Histogram of Oriented Gradients (HOG) descriptor [3] useful for characterizing the local shape and texture patterns of an image. It also can be shown to exhibit a fair degree of invariance to changes in illumination and viewpoint [8].

The sliding window SVM evaluation is fairly quick as it only involves an inner product about each location and scale. The computation of HOG forms the main bottleneck of our approach, however, this step can also be made fast using integral images [11].

2.3 Collision Avoidance

The accurate detection of aircrafts on a collision path is of particular importance as this is critical for achieving a safe flight. Unlike other aircrafts, aircrafts that are on a collision path exhibit a constant heading and are increasing in size as demonstrated in Figure 1. To detect this unique temporal pattern we extend our classification framework to additionally function over the time domain. To this end, we evaluate our SVM classifier over a spatio-

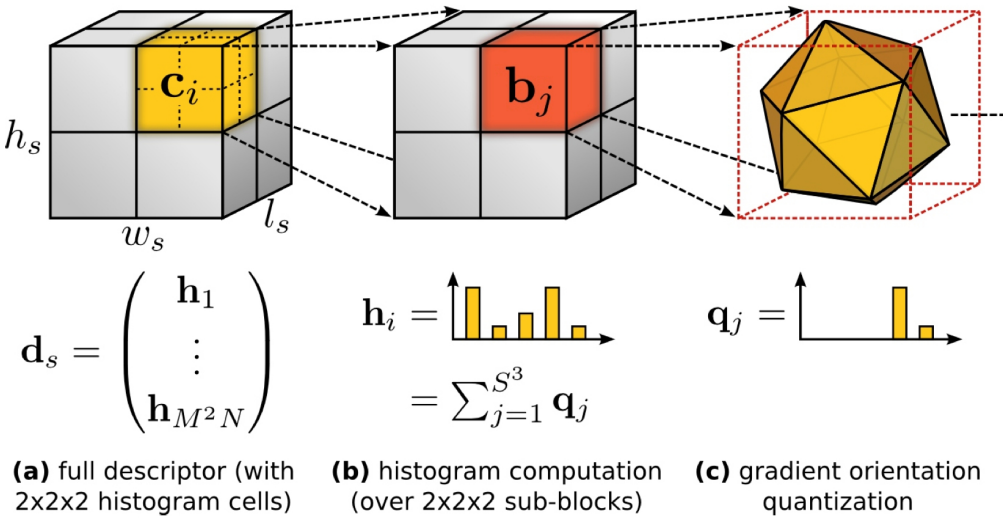


Figure 6: Illustration of Spatio-Temporal Histogram of Oriented Gradients (ST-HOG) composition for a block [7],[10]. (a) the support region around a point of interest is divided into a grid of gradient orientation histograms, $(\mathbf{h}_1, \ldots, \mathbf{h}_{M^2N})$; (b) each histogram $\mathbf{h}_i$ is computed over a grid of mean gradients, $\mathbf{q}_j$; (c) each gradient orientation is quantized using regular polyhedrons.

temporal sliding window, consisting of multiple input frames, and compute a spatio-temporal HOG descriptor [7],[10].

Spatio-temporal HOG (ST-HOG) defines a 3D descriptor as illustrated in Figure 6. In addition to the spatial gradient statistics, it also encodes the *temporal* edge statistics and therefore models the motion pattern of the objects that appear in the input sequence. As shown by our evaluation, ST-HOG offers a faithful representation for the detection of aircrafts on a collision path. The temporal resolution is an additional search parameter to our approach similar to spatial scale and location, and in our evaluation we experiment with different values of this parameter. Our complete detection framework is summarized in Figure 7.

Once the aircraft is detected the size and position of the detection can be used to compute additional information such as the speed, and time and distance to collision of the neighbouring aircraft, useful for performing mid-air collision avoidance. To estimate the time and distance to collision, we consider two basic models of aircraft movement: (1) both aircrafts are going

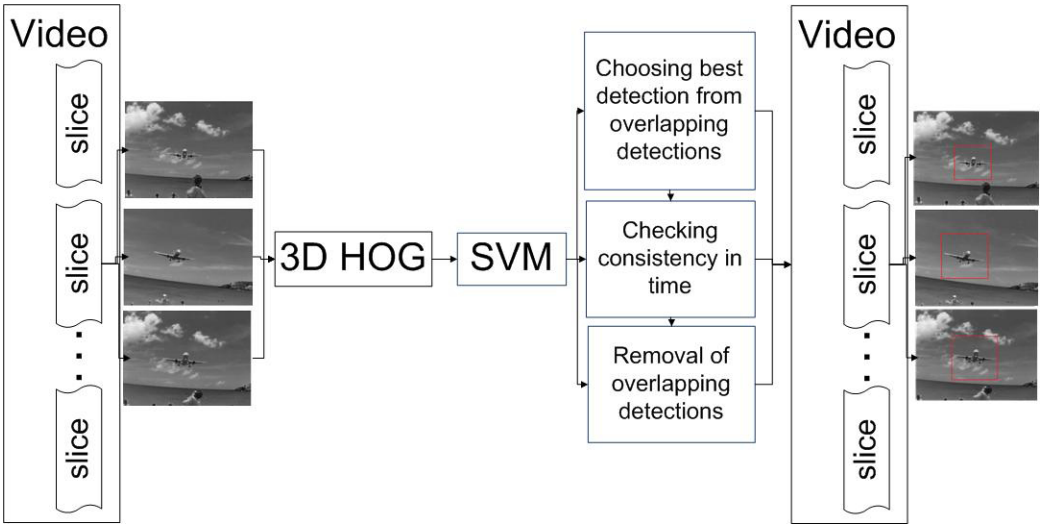
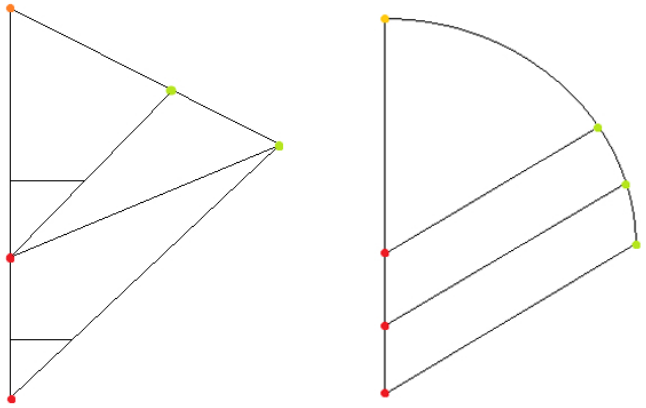


Figure 7: Our approach to detecting an aircraft on a collision course using spatio-temporal constraints. An ST-HOG descriptor is computed about each temporal window location and scale and is provided as input to the SVM classifier followed by a non-maximal supression to remove repeated detections.

straightforward to the collision point and (2) one aircraft is making a turn, and the other is moving straight forward to the collision point. These models are illustrated in Figure 8.

In the first case, it is enough to have two measurements to estimate the distance to the collision point. In the second case, three measurements are needed. If available, however, additional measurements at different times can increase the accuracy of the estimation. Due to the uncertainty in the size of the aircraft, the estimation process is intrinsically affected by errors, but the inaccuracy is not large and depends mainly on the physical parameters of the camera deployed such as focal length and field of view.

We have computed this error for typical camera features under the first motion model, in which both UAVs are moving straight forward to the collision point. The results are presented in Table 1 over various assumed aircraft sizes.



(a) (b)

Figure 8: Illustration of two basic models, that we used to estimate time and distance to collision. In this figure points of different colour represent different aircrafts. In (a) both aircrafts are going straightforward to the collision point and in (b) one aircraft is making a turn, and the other is moving straight forward to the collision point.

3 Evaluation

In this section we present an evaluation of our approach for the detection of both generic flight neighboring aircrafts and those on a collision path. We first discuss our experimental setup along with the datasets used in our evaluation. We then present the results of our approach. Finally, we provide an evaluation of our detection framework when considering aircrafts at various distances.

3.1 Experimental Setup and Datasets

The two settings of our approach with and without ST-HOG are evaluated. For aircrafts on a collision course, we also consider a temporal smoothing baseline that filters the 2D detections that do not appear in at least t consective frames.

As an evaluation metric we report the Receiver Operating Characteristic (ROC) curve of each detector and plot the true positive rate and corresponding false alarm rates obtained by varying the detection threshold b. The true positive rate is computed as the total number of correct detections divided

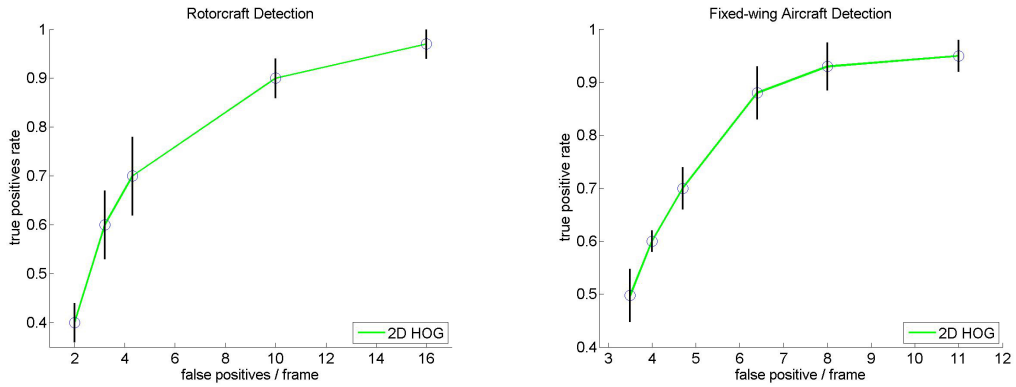
Camera Parameters	Real Size	Predicted Size	Maximum Error
focal length:	0.5 m	3 m	0.24 m
8 mm	1 m	0.5 m	0.13 m
field of view:	2 m	0.5 m	0.21 m
$35^\circ \times 26^\circ$	3 m	0.5 m	0.24 m
focal length:	0.5 m	3 m	0.28 m
4.2 mm	1 m	0.5 m	0.16 m
field of view:	2 m	0.5 m	0.25 m
$71^\circ \times 49^\circ$	3 m	0.5 m	0.28 m

Table 1: Error in collision distance estimate with respect to predicted aircraft size. The video frame size was set to 752×480 pixels, the size of the observed aircraft at the first measurement is 22 pixels, and the distance to the collision point is 24 m. The speed of the observing aircraft is 10 m/s and that of the observed one is approximately 9 m/s and its heading is calculated for collision.

by the number of positively labeled test examples. For the false alarm rate, we report the number of false detections per frame. We also report 95% error rates which is the false alarm rate achieved at 95% true positive rate. We perform leave-one-out experiments where one video is held out for testing and the remainder are used for training, and show average performance across all splits.

We consider two video datasets for the evaluation of our method, one of small-sized remote-controlled rotorcrafts and the other of full-size fixed-wing commercial and private aircrafts. Each dataset consists of 9 and 19 short video sequences respectively with image resolution of 640×480. Both datasets image neighboring aircrafts undergoing generic flight patterns from fixed and moving platforms. The fixed-wing dataset also consists of near-collision path instances taken of landing aircraft near an airport runway. Sample images from these datasets are illustrated in Figure 2.

For each video sequence we labeled a bounding box marking the extent of each aircraft present in each frame. Examples of the ground-truth were shown in Figure 3.



(a) (b)

Figure 9: Aircraft detection for (a) rotorcraft and (b) fixed-wing aircraft in general case. Average ROCs are shown across the leave-one-out evaluation. The error bars indicate ± 1 standard deviation. Our approach results in a reasonably accurate detection across both datasets.

3.2 Results

We first consider the detection of generic flight aircraft. Figure 9 displays the average leave-one-out accuracy of our approach on each dataset. We can faithfully detect the rotorcraft and fixed-wing aircrafts in these sequences achieving an average 95% error rate of 12 false detections per frame. The typical detections made by our system are also illustrated in Figure 10.

Next, we evaluate our approach for detecting aircrafts on a collision course. For this experiment, we used a subset of the fixed-wing dataset taken of landing aircraft near an airport runway. Example images from these sequences are displayed in Figure 11. The average leave-one-out accuracy of our approach is provided in Figure 12. We compare the ST-HOG approach with generic 2D detection with and without temporal smoothing with $t = 4$. We experimented with different values of t and here report the best baseline performance. By exploiting their unique temporal pattern our full approach with ST-HOG results in a significant boost in accuracy compared with a simple 2D detection of aircrafts on a collision course, giving a 95% error rate of only 2 false detections per frame. The accuracy of our approach for different temporal window sizes is also shown in Figure 12. When additional computational resources are available an even further increase in accuracy can be achieved using larger window sizes.

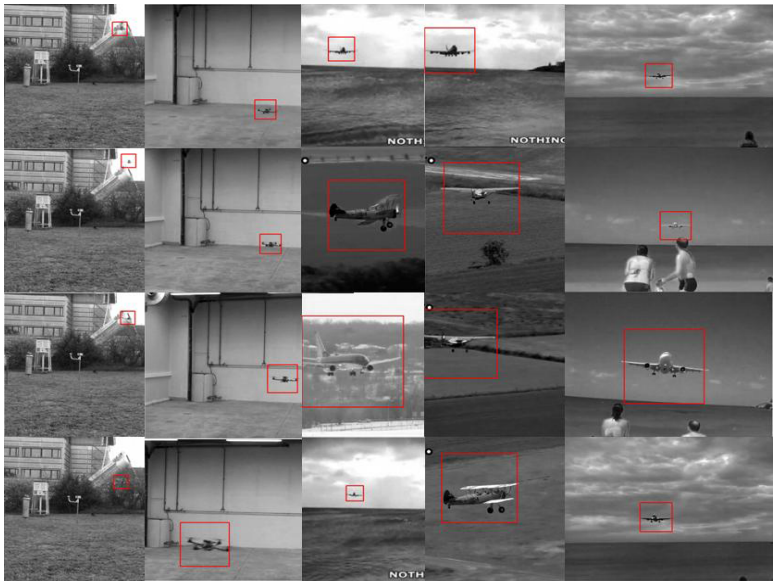


Figure 10: Illustration of detections made by our system on a set of example test images.

Finally, in Figure 13 we report the performance of our approach for aircrafts at various distances. This figure gives the detection accuracy with respect to the distance for different assumed wingspan sizes used to estimate its distance to the aircraft given its resolution in the image. The results assume an average false detection rate of 4.3 false detections per frame. Our approach is able to detect the aircrafts even when far away at distances up to 400 meters, and displays a reasonable degradation at increasing distances. There is also a degradation at closer distances where the aircrafts are seen under a different aspect than commonly available in the training data. In fact, the aircrafts are mostly seen from mid-range distances in these sequences, which accounts for the classifier's good performance in this range. Provided more training data of aircrafts at both nearby and distance locations would likely further improve our results and extend our range of operability.

4 Conclusion

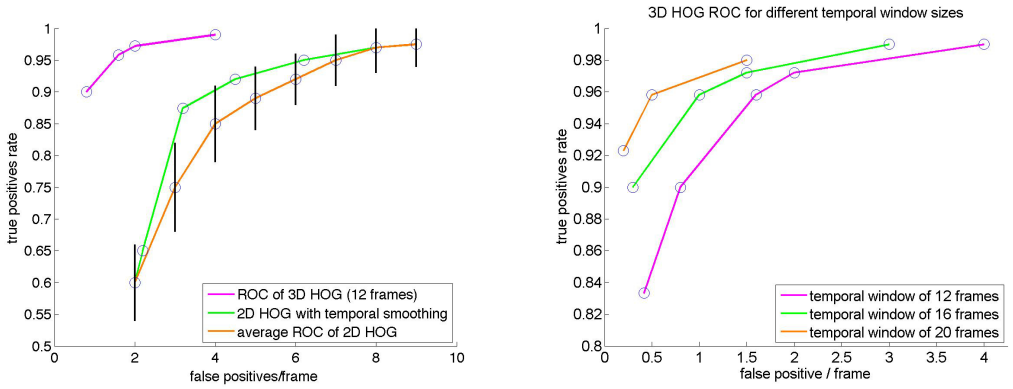
We presented an approach for the detection of generic flight neighboring aircrafts and those on a collision course. Our approach employs a sliding-window

Figure 11: Example video sequences from the fixed-wing dataset of landing aircraft taken near an aiport runway that are on a near-collision path with observer.

linear SVM classifier with a HOG feature representation. We also considered a spatio-temporal extension of HOG for the detection of aircrafts on a collision path. An evaluation was performed for the detection of both small-sized rotor and large fixed-wing aircrafts. Our approach gave an accurate detection with a 95% error rate of 12 false detections per frame. Achieving a high detection accuracy is crucial for mid-air collision avoidance and we demonstrated our system to achieve an even greater accuracy for the detection of aircrafts on a collision course with a 95% error rate of only 2 false detections per frame. Finally, we also demonstrated the performance of our approach when operating at different distances and saw a relatively good performance for the detection of distant aircrafts with the ability to accurately detect aircrafts with distances of up to 400 meters.

References

[1] C.M. Bishop. *Pattern Recognition and Machine Learning*. 2006.



(a) (b)

Figure 12: Detection results for an aircraft on a collision course. (a) Average ROCs are shown across the leave-one-out evaluation. The error bars indicate ± 1 standard deviation. Exploiting the temporal pattern of aircraft on a collision course results in a highly accurate performance, significantly outperforming 2D detection. A conventional temporal smoothing is able to increase the performance of the 2D detection, however, it does not compare to the spatio-temporal detector. (b) Performance of our spatio-temporal detector for different temporal window sizes. When additional computational resources are available an even further increase in accuracy can be achieved using larger window sizes.

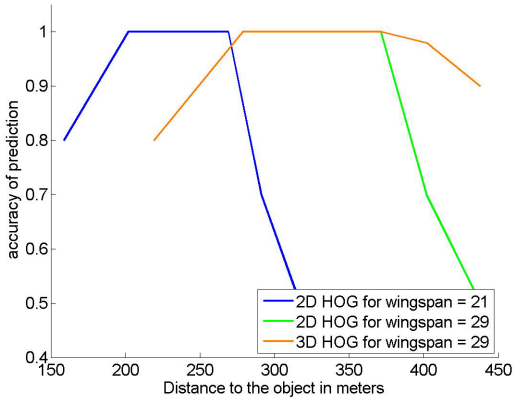


Figure 13: Evaluation of our approach for aircrafts at various distances. The plot shows the detection accuracy with respect to different distances and aircraft wingspans allowing for an average of 4.3 false detections per frame. Our approach is able to detect the aircrafts even when far away at distances up to 400 meters, and displays a reasonable degradation at increasing distances.

[2] C. Cortes and V. Vapnik. Support-Vector Networks. *ML*, 20(3):273–297, 1995.

[3] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *CVPR'05*.

[4] N. Dalal, B. Triggs, and C. Schmid. Human Detection Using Oriented Histograms of Flow and Appearance. In *ECCV'06*.

[5] M. Everingham, L. Van Gool, C. Williams, and A. Zisserman. The Pascal Visual Object Classes Challeng Results, 2005.

[6] P. Felzenszwalb, D. Mcallester, and D. Ramanan. A Discriminatively Trained, Multiscale, Deformable Part Model. In *CVPR'08*.

[7] A. Kläser, M. Marszałek, and C. Schmid. A Spatio-Temporal Descriptor Based on 3D-Gradients. In *BMVC'08*.

[8] D.G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *IJCV*, 20(2):91–110, 2004.

[9] T. Malisiewicz, A. Gupta, and A. Efros. Ensemble of Exemplar-SVMs for Object Detection and Beyond. In *ICCV'11*.

[10] D. Weinland, M. Ozuysal, and P. Fua. Making Action Recognition Robust to Occlusions and Viewpoint Changes. In *ECCV'10*.

[11] Q. Zhu, S. Avidan, M-C. Yeh, and K-Ting Cheng. Fast Human Detection Using a Cascade of Histograms of Oriented Gradients. In *CVPR'06*.