Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 52-12 (2019) 61–66

Sense and Avoid using Hybrid Convolutional and Recurrent Neural Networks

Daniel Vidal Navarro * **Chang-Hun Lee** ** **Antonios Tsourdos** ***

* *School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, MK430AL, United Kingdom (e-mail: vidaldani8@gmail.com).*
** *Department of Aerospace Engineering, KAIST, Daejeon, 34141, Republic of Korea (e-mail: lckdgns@kaist.ac.kr)*
*** *School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, MK430AL, United Kingdom (e-mail: a.tsourdos@cranfield.ac.uk)*

Abstract: This work develops a Sense and Avoid strategy based on a deep learning approach to be used by UAVs using only one electro-optical camera to sense the environment. Hybrid Convolutional and Recurrent Neural Networks (CRNN) are used for object detection, classification and tracking whereas an Extended Kalman Filter (EKF) is considered for relative range estimation. Probabilistic conflict detection and geometric avoidance trajectory are considered for the last stage of this technique. The results show that the considered deep learning approach can work faster than other state-of-the-art computer vision methods. They also show that the collision can be successfully avoided considering design parameters that can be adjusted to adapt to different scenarios.

Keywords: Sense and Avoid, neural networks, deep learning, computer vision, Kalman filter, range estimation, UAV.

1. INTRODUCTION

Nowadays, Unmanned Aerial Vehicles (UAVs) have become a reality and they are being introduced into daily life. One of their most critical points is safety. Since there are no human pilots inside UAVs, it is very important that they are capable to ensure safe flights even in unexpected situations such as when an obstacle appears in the trajectory. UAV Sense and Avoid is the artificial system equivalent of a human pilot detecting and avoiding hazard situations. The operation of UAVs requires an Equivalent Level of Safety (ELOS) to existing manned air vehicles. Hence, Sense and Avoid systems must have at least the same See and Avoid capabilities.

Air vehicles have usually got many different sensors onboard in order to be aware of the surrounding environment. Passive sensors such as cameras have not been usually used just on their own because of the limiting information that they are able to extract directly. However, technology advancements are allowing more powerful techniques to process the information provided by cameras so that they can be used for collision avoidance applications. Electro-optical cameras have low size, weight, cost and power consumption as well as a fast scan rate, however, an additional infrared camera may be required to operate under low light conditions or difficult weather conditions such as rainy, cloudy or fog presence.

Several post-processing solutions have been used to detect and track objects when using only one electro-optical camera. Some of them have been morphological based methods with additional temporal filtering, dynamic programming or support vector machines (SVM); bio-inspired methods such as optical flow and saliency methods; and, most recently, deep learning solutions.

It is possible to calculate the relative angular measurements making use of camera information such as the focal length or the Field of View. However, relative range and velocity cannot be measured from a single 2D image so passive ranging techniques are needed. The most common technique has been to filter the measurements making use of Kalman filters (KF). Common KF used for this purpose have been Modified Spherical (MS-EKF), Range-Parametrised (RP-EKF), Log-Polar Coordinate (LPC-EKF) and Unscented (UKF) (Mcfadyen and Mejias (2016)). A known issue is the observability. Depending on the relative motion between vehicle and target, it is possible that the range cannot be estimated from the line of sight (LOS) measurement. It has been shown that the range and the range rate are only observable when the vehicle performs a manoeuvre with an acceleration component perpendicular to the LOS vector (Shakernia et al. (2005)).

With respect to the conflict decision, it can be taken deterministically (Bareiss et al. (2015)) or probabilistically,

10.1016/j.ifacol.2019.11.070

which can be performed by Monte-Carlo samplings (Nordlund and Gustafsson (2011)) or analytical approximations (Irvine (2001)).

Deep learning solutions have also been used to process the data gathered from electro-optical cameras. Convolutional Neural Networks (CNN) have been the most common approach to computer vision applications with networks such as GoogLeNet (Szegedy et al. (2015)) and ResNet (He et al. (2016)). Some extensions have also been designed in order to locate objects in addition to detect and classify them. Some of the best performances were reported with Faster R-CNN (Ren et al. (2017)) and You Only Look Once (YOLOv3) (Redmon and Farhadi (2018)). Recurrent Neural Networks (RNN) are a different type of neural networks that are able to learn temporal information (Lipton et al. (2015)). A combination of these two types of neural networks are hybrid Convolutional and Recurrent Neural Networks (CRNN) (Donahue et al. (2017) Yang and Chan (2018)).

In this study, a Sense and Avoid system considering an electro-optical camera to capture the environment and hybrid Convolutional and Recurrent Neural Networks as processing solution is designed. The complete system consists of the obstacle detection and tracking, a conflict evaluation procedure and an avoidance manoeuvre generation. The considered processing approach have only recently started to be used in computer vision applications. After several years studying CNN and RNN in depth, a solution that combine the advantages of these both techniques can be very useful in tracking applications. Hence, the main aims of this work are to analyse the feasibility of using this new kind of neural networks in Sense and Avoid applications for UAVs, to identify possible issues that might occur and to suggest future directions in its research.

In the next section, the designed object detection and tracking technique based on a hybrid CRNN is described. Then, the collision avoidance strategy is explained. This includes angle measurements, range estimation using and Extended Kalman Filter (EKF), conflict probability calculation and geometric avoidance manoeuvre. The next section contains the obtained results. The last section includes the main conclusions extracted from the project.

2. OBJECT DETECTION AND TRACKING

The main objective of the study is to analyse the performance of hybrid CRNN in Sense and Avoid applications. Therefore, no neural network training is carried out so the designed technique is based on state-of-the-art open-source solutions that has been previously trained. The hybrid CRNN used is Re^3 (Gordon et al. (2017)). However, that network is a tracker but not an object detector. Hence, YOLO9000 (Redmon and Farhadi (2017)) is selected as CNN with object localisation to generate the initial bounding box around the object. The new technique is named ReLO and it includes a combination of the previous networks with additional parameters.

2.1 YOLO9000: You Only Look Once

YOLO is capable of detecting and classifying objects using a single neural network. It generates a certain number of bounding boxes and assigns a detection score to each one of them so only the ones with the highest values are the output.

2.2 Re^3: Real-time Recurrent Regression Network

Re^3 can track any object with a bounding box around it. It works by comparing two consecutive frames. It consists of a first CNN stage to extract features from a crop of the image. The output of the CNN consists of the object features so it is possible to track any object, independently of the classes which it was trained on. Furthermore, it is capable of tracking objects that change their features over time such as when they are rotating or that are temporally occluded. That is possible thanks to a second stage consisting of a special kind of RNN called Long Short-Term Memory (LSTM).

2.3 ReLO

The working idea of the proposed approach consists of using YOLO to detect and classify objects generating a bounding box around them and then using Re^3 to keep tracking them. A design parameter called refresh parameter is introduced. With this new parameter, YOLO is applied every certain number of frames so the objects in the FOV are refreshed and tracked using Re^3 for the next frames until YOLO is run again. The refresh parameter is the number of frames between two YOLO runs and it can be tuned as a trade-off between processing load and detection performance. Furthermore, this parameter can be adjusted accordingly to the environment where the vehicle operates.

3. COLLISION AVOIDANCE

The designed collision avoidance strategy assumes an initial positive detection and consists on the following procedure. Once an object is detected, an EKF is used to estimate the relative range between the vehicle and the object, which is initially considered as a mass point. With that information, the conflict decision is made by a stochastic method. The probability of conflict is estimated by an analytical expression so an avoidance trajectory is performed when this probability is over a predefined threshold. The considered avoidance trajectory is generated by a geometrical approach. The required command is applied and the next object position measurement is taken so a new decision with a new trajectory are proposed according to the relative velocity of the object with respect to the vehicle.

3.1 Angle Measurements

The EKF uses the relative bearing and elevation angles to estimate the relative range. However, ReLO outputs the object position as a pixel location in the frame. Therefore, a conversion from pixels to bearing and elevation angles is performed as in Fig. 1 where p_h and p_v are respectively the horizontal and vertical pixel locations with respect to the centre of the frame, W and H are the width and height of the frame in pixels (i.e. the resolution), FOV_h and FOV_v are the horizontal and vertical FOVs of the camera and

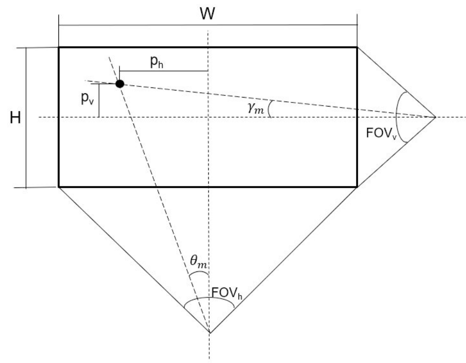

Fig. 1. Basic graphical scheme for angle measurements extraction.

θ_m and γ_m are the measured bearing and elevation angles respectively. Assuming that the camera characteristics are known (camera-specific adjustment required), the relative bearing and elevation angles are measured as:

$$\theta_m = \frac{p_h FOV_h}{W} \tag{1}$$

$$\gamma_m = \frac{p_v FOV_v}{H} \tag{2}$$

3.2 Range Estimation

The range estimator used is an EKF due to the non-linearity of the process. In Ahmadian and Pezeshk (2014), an EKF is designed to estimate the relative range in 2D from a bearing measurement. In this case, it is extended making use of the elevation angle to estimate the relative range in 3D. The considered definitions and equations can be found in the previous paper. To those equations, an extra z coordinate is added, defining an elevation angle γ as:

$$\gamma = \arctan\left(\frac{z}{\sqrt{x^2+y^2}}\right) \tag{3}$$

Initial UAV trajectory The UAV needs to perform a manoeuvre with an acceleration component perpendicular to the LOS vector to ensure range observability (Shakernia et al. (2005)). Otherwise, it may be impossible to estimate the range with a Kalman filter. However, the optimal manoeuvre depends on the relative velocity. In this work, a sinusoidal trajectory in the x-y direction is considered during the first seconds after an object has been detected in order to allow time for the range to converge.

3.3 Conflict Probability

The expression to calculate the conflict probability is NP-hard but it can be approximated by some analytical expressions, what requires much less computational power than Monte Carlo samplings. In Sense and Avoid applications, processing speed is critical since later a conflict is detected, lower the available time to avoid that conflict. Therefore, an analytical approximation is used so the conflict probability can be assessed quickly to allow enough time to execute an avoidance manoeuvre if needed.

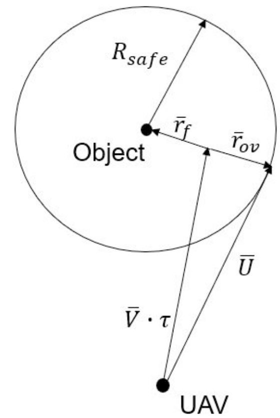

Fig. 2. Avoidance manoeuvre technique.

All the proposed approximations are usually very similar but with different noise modellings. In this case, a simple expression extracted from Krozel and Peters (1997) is used. Derivation details can be found in the mentioned paper, which approximates the probability of conflict as:

$$P(c) = \frac{1}{2}erf\left(\frac{R_{safe}+\bar{r}_f}{\sqrt{2}\sigma_{r_f}}\right) + \frac{1}{2}erf\left(\frac{R_{safe}-\bar{r}_f}{\sqrt{2}\sigma_{r_f}}\right) \tag{4}$$

where R_{safe} is the minimum safe distance between the UAV and the obstacle, $\bar{r}_f$ is the miss distance vector, σ_{r_f} is the variance of the miss distance and $erf(x)$ is the Gaussian error function:

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}du \tag{5}$$

The conflict is defined as the situation when the distance between the UAV and the obstacle is below R_{safe}. This parameter R_{safe} must be set accordingly to the required distance that should be kept in order to ensure safety. The bounding box size given by ReLO can also be considered to define R_{safe}.

3.4 Avoidance Manoeuvre

When the probability of conflict is over certain threshold, an avoidance manoeuvre should be executed in order to avoid a collision. A geometric approach is considered in this case. The proposed technique is based on the solution from Park et al. (2008). In that case, it was designed to be used with a deterministic conflict assessment and with two aircraft that could execute a cooperative avoidance trajectory. However, it is adjusted now to be used in this new application where the UAV is the unique responsible to avoid the collision. In this approach, it is assumed that the UAV is following certain trajectory and the command required to modify that trajectory in order to avoid the collision is the designed parameter.

If the conflict probability is high, it means that the UAV is likely to cross the R_{safe} distance with respect to the obstacle. Therefore, the proposed avoidance trajectory consists of modifying the UAV velocity direction so it goes through the surface of the safe sphere and, thus, going further from the obstacle, as can be visualized in Fig. 2.

The UAV flies with a velocity $\bar{V}$. Propagating that velocity until the time-to-closest-approach τ, the miss distance vector $\bar{r}_f$ is obtained as discussed before. The distance vector for obstacle avoidance $\bar{r}_{ov}$ can be obtained by simple geometry as:

$$\bar{r}_{ov} = (R_{safe} - r_f) \cdot \left(-\frac{\bar{r}_f}{\|\bar{r}_f\|}\right) \tag{6}$$

Hence, the unitary vector for the new velocity direction to point through the safe circumference can be expressed as:

$$\bar{U} = \frac{\bar{V} \cdot \tau + \bar{r}_{ov}}{\|\bar{V} \cdot \tau + \bar{r}_{ov}\|} \tag{7}$$

Once the desired velocity vector is obtained, it is only needed to generate the required control commands in order to execute this new velocity. It is assumed that changing only azimuth and/or elevation angles would generate the desired velocity direction so no thrust control is needed. Hence, the wished trajectory angles can be obtained from the director vector of the new goal velocity:

$$\alpha = \arctan\left(\frac{U_y}{U_x}\right) \tag{8}$$

$$\psi = \arctan\left(\frac{U_z}{\sqrt{U_x^2 + U_y^2}}\right) \tag{9}$$

Therefore, this trajectory variation is generated only when the conflict probability is over a threshold. Once the UAV gets further from the obstacle, the conflict probability decreases and the avoidance manoeuvre is not generated any more, returning the UAV to its previous trajectory.

4. RESULTS

ReLO and the collision avoidance technique performances are assessed separately and then an experimental test is carried out to merge both parts of the Sense and Avoid method.

4.1 Object Detection and Tracking

The object detection and tracking performance using ReLO is assessed using videos of objects that a UAV would likely find during its operation. A common CPU with Intel i7 and 8 GB RAM has been used for this purpose. First, YOLO9000 is compared to Re^3 in order to obtain orders of magnitude of the processing speed so ReLO performance can be analysed in relation to these values. A video of an aircraft in the sky as in Fig. 3 is considered. A processing speed of 1.3 fps (frames per second) is obtained using YOLO at every frame whereas the processing speed is 12.2 fps when YOLO is used only at the first frame and Re^3 in the rest of frames. Therefore, it is shown that hybrid CRNN trackers such as Re^3 can work almost ten times faster than one of the fastest CNN as YOLO is.

Fig. 3. Frame of a video used to analyse YOLO and Re^3.

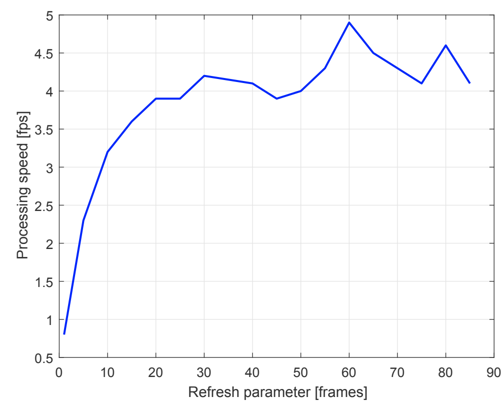

Fig. 4. Tracking speed for different refresh parameter values.

Furthermore, according to Gordon et al. (2017), Re^3 could work at 150 fps with an Intel Xeon CPU E5-2696 v4 @ 2.20GHz and an Nvidia Titan X (Pascal). Hence, it is assumed that the obtained processing speeds could be up to 12 times faster with more powerful processing units.

In order to have a safe performance assessment of ReLO, more complex scenarios are considered. In these new videos, several objects are constantly present in the FOV so high computational power is required and, furthermore, objects are continuously coming in and out from the FOV. Different values of the refresh parameter are tested obtaining the results of Fig. 4. It can be seen that lower the refresh parameter slower the speed since YOLO is run more frequently, what requires more computational load. However, in such complex environments, high refresh parameters affect the detection performance since it takes longer to start tracking new objects appearing in the FOV and bounding boxes remain for longer after the corresponding objects have disappeared. In this case, a value for the refresh parameter of 30 fps reports good detection and tracking performance and a mean tracking speed of 4.2 fps is achieved, what could be up to 50 fps with a more powerful processing unit. Assuming a speed of 30 fps to work in real time operations, ReLO is capable of working in demanding real time applications such as Sense and Avoid systems of UAVs. Moreover, this example shows a very complex scenario that a UAV would probably never find so the refresh parameter could be tuned larger in most of the cases and, thus, the speed would be even higher.

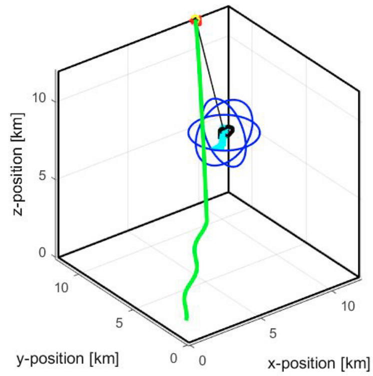

Fig. 5. 3D simulation scenario. The green line represents the UAV trajectory, the blue circumferences represent the safety distance and the cyan line represents the estimated object location obtained from the EKF.

4.2 Collision Avoidance Simulations

A MATLAB simulation is also designed to assess the performance of the collision avoidance strategy. The scenario can be seen in Fig. 5. Simulations are performed at several UAV and object initial locations, initial heading and elevation angles and linear speeds. Both dynamic and static obstacles are simulated. The design parameters are the safety distance and the conflict probability threshold. The safety distance is represented by three blue circumferences around the vehicle but it is actually a sphere. An example of one of the results is shown in Fig. 6. The noise parameters in the EKF are guessed in order to get accurate relative range estimations. An optimal probability of conflict of 50% is detected since very low values deviates the UAV further from the predefined desired trajectory whereas very high values may generate a command to avoid the collision too late, generating very aggressive manoeuvres that the UAV may be unable to perform. However, simulations with all kind of different parameters show that a UAV can always avoid the collision with an obstacle with this kind of strategy if the EKF does not diverge and detects correctly the obstacle location. Hence, the EKF plays a key role in the proposed collision avoidance technique.

4.3 Experimental Tests

Due to the criticality of the EKF performance on the overall Sense and Avoid technique, an additional experimental test is performed to combine both the object detection and tracking and the collision avoidance parts. The objective is to measure the noise of the object location measurement with ReLO and to use that real noise to check the collision avoidance behaviour since guessed values were used previously. A GoPro Hero 4 camera mounted on a remote controlled rover is selected for this purpose. Different trajectories are executed by the rover while the camera is recording a keyboard. The recorded videos are run through ReLO so the measured object location is obtained. This experiment is carried out in an experimental laboratory equipped with a VICON navigation system at Cranfield University so the processed results are compared to the

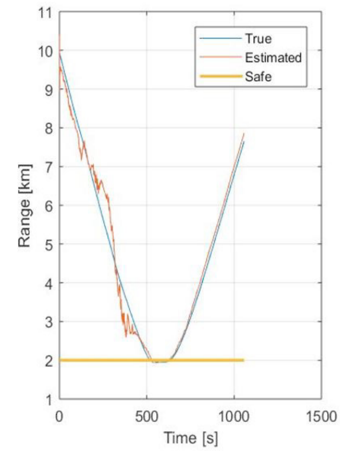

Fig. 6. True and EKF-estimated relative range over time with safety distance.

Fig. 7. Experiment set up in the flying lab at Cranfield University equipped with a VICON navigation system.

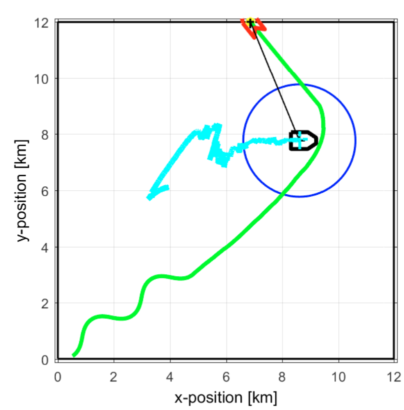

Fig. 8. Considered 2D scenario with dynamic object.The green line represents the UAV trajectory, the blue circumference represents the safety distance and the cyan line represents the estimated object location obtained from the EKF.

accurate VICON measurements and noise information can be extracted. The experiment set up is shown in Fig. 7.

The average standard deviation of the bearing angle after 12 tests is $\sigma_\theta = 4.97$ deg. Since the vehicle in the experimental test is a rover, a 2D version of the EKF with the measured noise is considered. 2D simulations with the resulting filter are carried out in a 2D scenario as the one

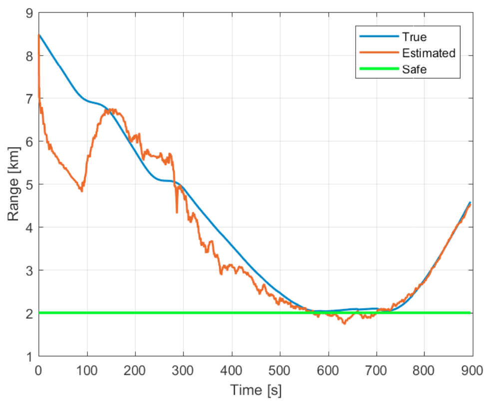

Fig. 9. True and EKF-estimated relative range and safety distance over time.

in Fig. 8. An example of the results is shown in Fig. 9 and it can be seen that the EKF can accurately estimate the true relative range so the collision is always avoided.

5. CONCLUSIONS

A Sense and Avoid technique for small UAVs using only an electro-optical camera as sensor has been designed. This technique uses a hybrid CRNN to detect and track objects named ReLO, an EKF to estimate the relative range, a probabilistic conflict decision and a geometrical avoidance manoeuvre. ReLO shows to work efficiently in complex scenarios, what ensures safety which is critical in Sense and Avoid applications, but the detection performance still relies in a CNN such as YOLO. The collision avoidance strategy shows to be quite reliable where fine EKF performance is critical.

In general terms, results show that hybrid CRNN can be a feasible solution for the considered Sense and Avoid problem. The processing speed is higher while the detection performance remains the same. Furthermore, the proposed approach is adaptable to different scenarios since some design parameters are considered so the technique behaviour can be adjusted according to the expected UAV operation. The main limitations are the detection performance and the relative range estimation so further research is suggested in those fields.

REFERENCES

Ahmadian, A. and Pezeshk, A.M. (2014). A Modified Extended Kalman Filter for Bearings- Only Tracking. *The 22nd Iranian Conference on Electrical Engineering (ICEE 2014)*, (Icee), 1712–1716.

Bareiss, D., Van Den Berg, J., and Leang, K.K. (2015). Stochastic automatic collision avoidance for tele-operated unmanned aerial vehicles. *IEEE International Conference on Intelligent Robots and Systems*, 2015-Decem, 4818–4825. doi:10.1109/IROS.2015.7354054.

Donahue, J., Hendricks, L.A., Rohrbach, M., Venugopalan, S., Guadarrama, S., Saenko, K., and Darrell, T. (2017). Long-Term Recurrent Convolutional Networks for Visual Recognition and Description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4), 677–691. doi:10.1109/TPAMI.2016.2599174.

Gordon, D., Farhadi, A., and Fox, D. (2017). Re3 : Real-Time Recurrent Regression Networks for Visual Tracking of Generic Objects. 1–8. doi:10.1109/LRA.2018.2792152. URL `http://arxiv.org/abs/1705.06368`.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. doi:10.1109/CVPR.2016.90. URL `http://ieeexplore.ieee.org/document/7780459/`.

Irvine, R. (2001). A geometrical approach to conflict probability estimation. *Encounter*, 1–15. doi:10.1109/DASC.2001.964214.

Krozel, J. and Peters, M. (1997). Strategic conflict detection and resolution for free flight. *, {Proceedings} of the 36th {IEEE} {Conference} on {Decision} and {Control}, 1997*, 2(December), 1822—-1828 vol.2. doi:10.1109/CDC.1997.657844.

Lipton, Z.C., Berkowitz, J., and Elkan, C. (2015). A Critical Review of Recurrent Neural Networks for Sequence Learning. 1–38. doi:10.1145/2647868.2654889. URL `http://arxiv.org/abs/1506.00019`.

Mcfadyen, A. and Mejias, L. (2016). A survey of autonomous vision-based See and Avoid for Unmanned Aircraft Systems. *Progress in Aerospace Sciences*, 80, 1–17. doi:10.1016/j.paerosci.2015.10.002.

Nordlund, P.J. and Gustafsson, F. (2011). Probabilistic noncooperative near mid-air collision avoidance. *IEEE Transactions on Aerospace and Electronic Systems*, 47(2), 1265–1276. doi:10.1109/TAES.2011.5751257.

Park, J.W., Oh, H.D., and Tahk, M.J. (2008). UAV collision avoidance based on geometric approach. *2008 SICE Annual Conference*, 2122–2126. doi:10.1109/SICE.2008.4655013. URL `http://ieeexplore.ieee.org/document/4655013/`.

Redmon, J. and Farhadi, A. (2017). YOLO9000: Better, faster, stronger. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-Janua, 6517–6525. doi:10.1109/CVPR.2017.690.

Redmon, J. and Farhadi, A. (2018). YOLOv3: An Incremental Improvement. doi:10.1109/CVPR.2017.690. URL `http://arxiv.org/abs/1804.02767`.

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149. doi:10.1109/TPAMI.2016.2577031.

Shakernia, O., Chen, W.z., and Raska, M.V.M. (2005). Passive Ranging for UAV Sense and Avoid Applications. *AIAA InfotechAerospace 2005*, (September), 1–10. doi:10.2514/6.2005-7179.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June, 1–9. doi:10.1109/CVPR.2015.7298594.

Yang, T. and Chan, A.B. (2018). Recurrent Filter Learning for Visual Tracking. *Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017*, 2018-Janua, 2010–2019. doi:10.1109/ICCVW.2017.235.

2019-11-25

Sense and avoid using hybrid convolutional and recurrent neural networks

Navarro, Daniel Vidal

Elsevier