Artificial Neural Networks for Airport Runway Safety Systems

Denis Bekasov, Bauman Moscow State Technical University

Address for correspondence: Denis Bekasov, Bauman Moscow State Technical University, Moscow, Russian Federation, e-mail: bekasov@bmstu.ru

Abstract

This paper presents the analysis of the existing approaches to ensuring the safety of aircraft`s takeoff and landing at airport runways using video surveillance systems. The subject area is formalized, and security threats and measures to prevent them are assessed. Optional architecture of the system designed for detection and classification of moving objects in the airport runway area is presented. The architecture is based on Neural Networks with AI elements. Also the original method of runway objects' trajectory tracking is proposed. And finally, the research results of the applicability of the proposed architecture are presented.

Keywords

Airport Runway Safety, Artificial neural networks, Video tracking, Object detection, Object classification

1. Introduction

Airport runways are the areas with high risks of occurrence of emergency situations. Largely they include various incidents during aircraft taxiing, take-off. These incidents may pose a threat to both a crew and passengers of an aircraft and people on a runway surface. Incidents can be of technical nature as well, caused by failure and damage of aircraft, ground vehicles, and ground infrastructure. Most of these threats are related to either unexpected appearance of static or dynamic foreign objects, including various items (garbage, vehicles, or people) or their untimely or unauthorized appearance on the runway.

The International Civil Aviation Organization (ICAO) and other aviation organizations regulate use of various systems aimed at mitigation of various runway-related risks (MLS, ILS, TLS, ADS–B, RAAS). Most of such systems use radars and other radioelectronic facilities in their operation. At the same time, there is no system that would solve comprehensive issues of runway safety in terms of monitoring of objects present on the runway. One possible solution is to employ artificial intelligence systems based on neural networks to analyze the information provided by the specialized CCTV system.

2. Problem analysis

According to ICAO's definition in (ICAO, 2006), a runway is a "defined rectangular area on a land aerodrome prepared for the landing and takeoff of aircraft. In fact, the following objects may be located on the runway for some reason or other at various times:

- aircraft;
- follow-me-cars;
- ground support vehicles;
- odd vehicles;
- bulky waste, technical waste;
- domestic waste;
- people;
- animals and birds.

In addition, both the runway itself (pits, cracks, worn-out marking) and the airport infrastructure can be damaged.

All these objects, their possible interactions, specific runway features, weather conditions, and human factor may cause runway incidents. The ICAO in total specifies 10 possible incident types listed below (ICAO, 2017).

- Category Description Abnormal Runway Contact (ARC);
- Bird Strike (Bird);
- Ground Collision (GCOL);
- Ground Handling (RAMP);
- Runway Excursion (RE);
- Runway Incursion (RI);
- Loss of Control on the Ground (LOC-G);
- Collision with Obstacle(s) (CTOL);
- Undershoot / Overshoot (USOS);
- Aerodrome (ADRM).

This paper presents the problems related to events on the ground and interaction with ground objects, which is: RAMP (occurrences during (or as a result of) ground handling operations), RE (an event in which an aircraft veers off or overruns off the runway surface during either take-off or landing), RI (any occurrence at an aerodrome involving the incorrect presence of an aircraft, vehicle or person on the protected area of a surface designated for the landing and take-off of aircraft), CTOL (collision with obstacle(s), during take-off or landing whilst airborne), ADRM (occurrences involving aerodrome design, service, or functionality issues).

Harm and severe consequences of the most of the listed incidents may be significantly reduced if air traffic control officers and operators of air traffic control stations:

- possess complete, unbiased, and formalized real-time information on runway-related events,
- receive analytical reports on the current situation and warnings on possible future incidents.

Radars and specialized vehicle- or personnel-mounted radio beacons and GPS-trackers are insufficient to collect all necessary information on the runway environment. Visual monitoring is required. However, even at a small airport, the amount of visual information is so large that a special team of operators working in shifts is required to do continuous monitoring of all cameras. At the same time, there is no guarantee to avoid a human error, lack of attention, accumulated fatigue. In addition, an operator may not always be able to notice, recognize, and predict a potentially dangerous incident promptly. Due to this fact, the problem of building up an automated runway video surveillance system, including AI-based video analysis and prediction modules, is becoming increasingly relevant.

Based on the incoming information received from some video camera network, such system shall:

- detect all objects within visual range,
- classify the detected objects,
- track the classified objects and generate their probable trajectories,
- make predictions on incident possibility.

When possibility of certain incident is identified, an instant notice with calculated hazard level shall be sent to the operator for him to check it and decide on a response action.

3. Optional system architecture

This paper proposes the following architecture of the video surveillance system. As the airport runway area may be large, it will be not enough to have only one camera to receive information on all runway events; a camera system is required. There are static and pan-tilt-zoom (PTZ) cameras. Static cameras are cheaper and easy in maintenance, allowing them to simplify a number of video analysis problems. However, their angle and zoom are fixed, so they cannot adjust themselves to object movement and have a little use in observation of distant objects. Static cameras may be united into optical clusters to cover a larger observation angle. It should be noted that runways may have a length of almost 12 km, (Matthews, 1995) and there may be a more such runways at the airport, so a number of optical clusters may be required. In order to compensate the disadvantages of static cameras, the optical clusters may be supplemented with rotary cameras that may be pointed at distant object with required zoom in either automatic or manual control mode.

The system proposed here includes k sets of m subsets of static CCTV cameras with fixed positions and fields of view, along with 2*k sets of PTZ-cameras. These cameras should cover the whole airport runway space to be monitored. A subset of m cameras is required to create a single panoramic view of a large runway area. This panoramic coverage is essential for detecting any objects in the runway zone. Each such subset is augmented with two PTZ-cameras: one for automatic object annotation and one for object tracking. The automatic object annotation camera is controlled by the system. It annotates and classifies runway objects. The object tracking camera, also controlled by the system, serves to support tracking of classified objects and automatic pointing at a detected incident site.

Depending on system configuration and availability of high-speed broadband airport communication channels, the computing units may be either installed immediately in the camera modules or implemented as a separate video stream analysis server processing all incoming information.

From the point of view of software architecture, the system consists of four modules:

1. Moving objects detection;
2. Object annotation;
3. Object tracking;
4. Incident prevention.

According to section 2, the system must detect and classify objects in the incoming video stream. Annotating convolutional neural networks are currently widely used in object detection and classification problems (Zhao, Zheng, Xu, & Wu, 2019). The problem is that such networks may efficiently classify objects having minimum 25 pixels in one linear dimension, and only foreground objects or aircraft meet this requirement for static cameras. Therefore, it is proposed to scan static camera streams for any dynamic objects, point the annotating PTZ-camera on them and then annotate all objects from static camera streams and separate frames of the PTZ-camera by means of the annotating convolutional neural network.

4. Proposed solution

4.1 Moving objects detection

The main purpose of the object detection module is to detect presence of dynamic objects and their aggregations in the frames of the video stream from the static cameras. This problem quite regular for computer vision is usually solved by using background subtraction algorithms.

The solution proposed in this paper involves the classical exponential smoothing algorithm (Oppenheim & Schafer, 1975) with formulas (1) and (2) for calculation of the expected value and variance.

$$bm_t = \alpha F_t + (1-\alpha)bm_{t-1} \qquad\qquad (1)$$

$$bs_t = \alpha(F_t - bm_{t-1})^2 + (1-\alpha)bs_{t-1} \qquad\qquad (2)$$

Where bm_t is the background expected value, bs_t is the background variance, F_t is the video frame image at moment t. Depending on system configuration and detection range requirements, objects may be very small-sized (a few pixels), which makes noise filtration difficult. However, the fact that the camera angles are known and always fixed provides unambiguous information on permissible dimensions of objects on different image locations. This allows cutting off most noise by quite simple techniques.

The background containing dynamic objects is defined by expression (3).

$$f_t = (I_t - bm_{t-1})^2 > \mu_{t-1}bs_{t-1} \qquad\qquad (3)$$

After that, connected components are extracted from the image by standard operations of mathematical morphology. These connected components are candidates for recognition as dynamic objects. It is worth noting individually that objects may be located in close proximity to each other and, after the set of morphological operations, united in one connected component. This does not affect detection quality as the final decision on presence of objects of specific types is taken at the next stage. On the contrary, this will help reduce the processing time by avoiding individual positioning of a camera on each object belonging to the derived connected component.

In order to make sequential annotation of the objects in the found connected components possible, ID numbers must be given to the objects. These ID numbers must be used by the detection module for tracking of the connected components: the system must ascertain that it spots the same object. The standard technique quite widely used in solving multi-object tracking problems is used for this purpose: the Kalman filter for trajectory analysis and solution of the assignment problem for search of similarities between the objects identified in different frames (Bewley, Ge, Ott, Ramos, & Upcroft, 2016). The position and speed of the connected component are the observable variables of the Kalman filter. The position and speed of the real objects (and their aggregations) are the latent variables. The connected components are matched to the objects by the maximum likelihood estimation method formalized as an assignment problem (Hungarian Algorithm as an example).

4.2 Object annotation

The purpose of the object annotation module is to create a relevant list of all objects at a specific time. Frames from static cameras and IDs of the objects detected at the previous stage are the input data of the module.

For object annotation, this paper proposes using a convolutional neural network based on modified YOLOv3 architecture (Redmon & Farhadi, 2018) that, in turn, is based on GoogLeNet(Szegedy, et al., 2015) and R-CNN architectures (Girshick, Donahue, Darrell, & Malik, 2014). The network is trained in a standard way on the collected data set of runway objects (people, vehicles, support equipment, infrastructure facilities, and aircraft). The dataset is formed on the basis of open accessible datasets of the urban environment and enriched with marked-up frames from airports. The chosen network architecture provides the best accuracy to performance ratio in annotation of random objects from CCTV images (see Figure 1).

Figure 1. Examples of operation of the annotation module based on the convolutional neural network

This module operates in the system asynchronously. It receives the latest relevant data from the moving objects detection module and static cameras and processes only this data. The speed of processing of all frames and objects may vary depending on a specific system configuration but it does not cause issues. General algorithm of module operation:

- annotation of all static camera frames (parallelized operation),
- annotation of all detected groups of moving objects by using the PTZ-camera, with the exception of those already processed at the first stage (serial operation, may be parallelized if a number of PTZ-cameras is available).

The first step allows annotating within a single iteration the frames from the static surveillance cameras, detect and classify both dynamic foreground objects and large static object that cannot be defined at the previous stage. The second step implies serial positioning of the PTZ-camera at each moving objects group identified at the moving objects detection stage. The purpose of this step is to quickly annotate conceivable runway objects. During this process, the dynamic clusters are divided into separate objects with outlined boundaries. All positions and speeds of movement of object boundaries are brought to a single coordinate system to combine results obtained from the static and PTZ-cameras. A few frames with good focus are sufficient for the purpose of annotation, so, while a queue of unannotated moving areas is present, the camera quickly switches between these areas, focusing on the objects for a period required for the first successful annotation. If the objects cannot be classified, an object of unknown type is designated within the boundaries outlined by the detector. Once all objects in the queue are annotated, the camera switches to the mode of annotation correction and distant runway area surveillance. In the distant area surveillance mode, the system annotates the frames from the PTZ-camera and searches static or small low-speed moving objects that cannot be detected by the static cameras.

4.3 Object trajectory tracking

The main purpose of the object trajectory tracking module is to track objects, define and predict their trajectories.

The proposed method is based on significantly modified classical Tracking-Learning-Detection (TLD) approach (Kalal, Mikolajczyk, & Matas, 2010). In this approach, the whole tracking problem is viewed as an aggregation of solutions of two separate problems: tracking (following the object movement) and detection (searching for the presence of the object in the frame). Tracking is performed by using the object movement model based on optical flows while detection is implemented via the trained model of an object, originally based on the random forest of object appearance examples by using the sliding window method and image cross-correlation for the object search itself.

In the case under study, scanning of the whole image is not required for object search; it will be enough to check only the most probable positions. The input data of the method is represented by both raw detections and annotated objects. For all objects, the module contains the descriptors forming the object models and trajectories.

The detection stage produces results within method in question only with sets of known possible object positions. Similarity of object appearance models is evaluated by using the modular approach. DDIS (Itamar, Mechrez, & Zelnik-Manor, 2017) is used as a basic mechanism to define similarity between the objects. In addition, various specific criteria improving matching quality (color of special clothing of the airport personnel, faces at short distances, vehicle and flight numbers) may be used depending on object class and video survey parameters.

The method also uses the output of the moving objects detection module that is in fact a tracker based on background subtraction and Kalman filter in addition to a classical optical flow-based tracker.

As a trajectory model, the recurrent neural network of LSTM (Hochreiter & Schmidhuber, 1997) type is used, it is learning, when tracking a specific object, and memorizing the character of its trajectory. The idea of use of similar neural network for trajectory prediction is inspired by work (Ning, et al., 2017).

At this stage, the assignment problem formulation is formed again but this time taking into account all acquired information. The tracked objects and output data of the object annotation module are assigned. Moreover, the problem formulation is supplemented by the information on object similarity, tracker suppositions, and predictions of the trajectory model. If the number of tracked objects is less than the number of annotation module's results, new object descriptors are created. The resulting optimization problem is solved by the linear programming methods.

The final step is clarification of the object boundaries on the basis of all input data. After the boundaries are clarified, the object model and object trajectory model start to learn, and the coordinates of the object on the runway are calculated.

The module also controls the tracking camera. The camera switches from one object to another in rotation, surveying the most important object types with specified time slices allocated for each object. Camera pointing is based on clarified borders of a tracked object. The information obtained from the camera is used for clarification of trajectories of specific tracking objects.

4.4 Incident prevention

The operation logic of the incident prevention module is based on the information on trajectories and current coordinates of the runway objects received from the object tracking module, runway areas information (taxiing area, take-off and landing area, etc.), information on runway events at the time of aircraft entry, take-off and landing received from the air traffic control station.

For each tracked dynamic object with a trajectory model, k spots of probable positions are generated with a specific sampling rate for the related time points. For each time point, a spatial map is generated. This map contains the data on spots of probable positions of objects of various classes, positions of static objects, different runway areas, and object positions known in advance on the basis on the information received from the air traffic control stations.

This information is provided in the form of a set of closed spatial areas and based on, among others, the calculated object dimensions. Each area belongs to a specific class (vehicle, aircraft, human, unknown object). At the last step, areas of intersection is identified in accordance with the set of rules for prohibited intersections of areas of different classes. An R*-tree is used for optimization of the process of search of intersection areas (Beckmann, Kriegel, Schneider, & Seeger, 1990). Depending on rule type and time point of the map where the intersection is found, a warning message is generated with designation of a specific hazard class, hazard level, incident probability and estimated time.

5. Experimental results

The test prototype of the proposed system is composed of 3 static detecting cameras with a resolution of 20 MP and view angle of 30 x 20 degrees and 2 PTZ-cameras with a resolution of 1,920 x 1,200 pixels and 30x optical zoom. The computing unit is a multi-processor server with 4 specialized graphic cards with high parallel performance.

The dependence of accuracy of object detection, annotation, and tracking on distance to an object (in meters) was measured for three major object types: people, vehicles, and aircraft (see Figure 2). The final annotated sample contains 324,500 frames. As an accuracy measure, ratio (4), standard for similar problems, was selected.

$$\frac{TP}{TP+FP} \qquad\qquad\qquad\qquad (4)$$

where *TP* - true positive of the ratio of areas of the real and experimental object restriction frames, *FP* - false positive of the same ratio.
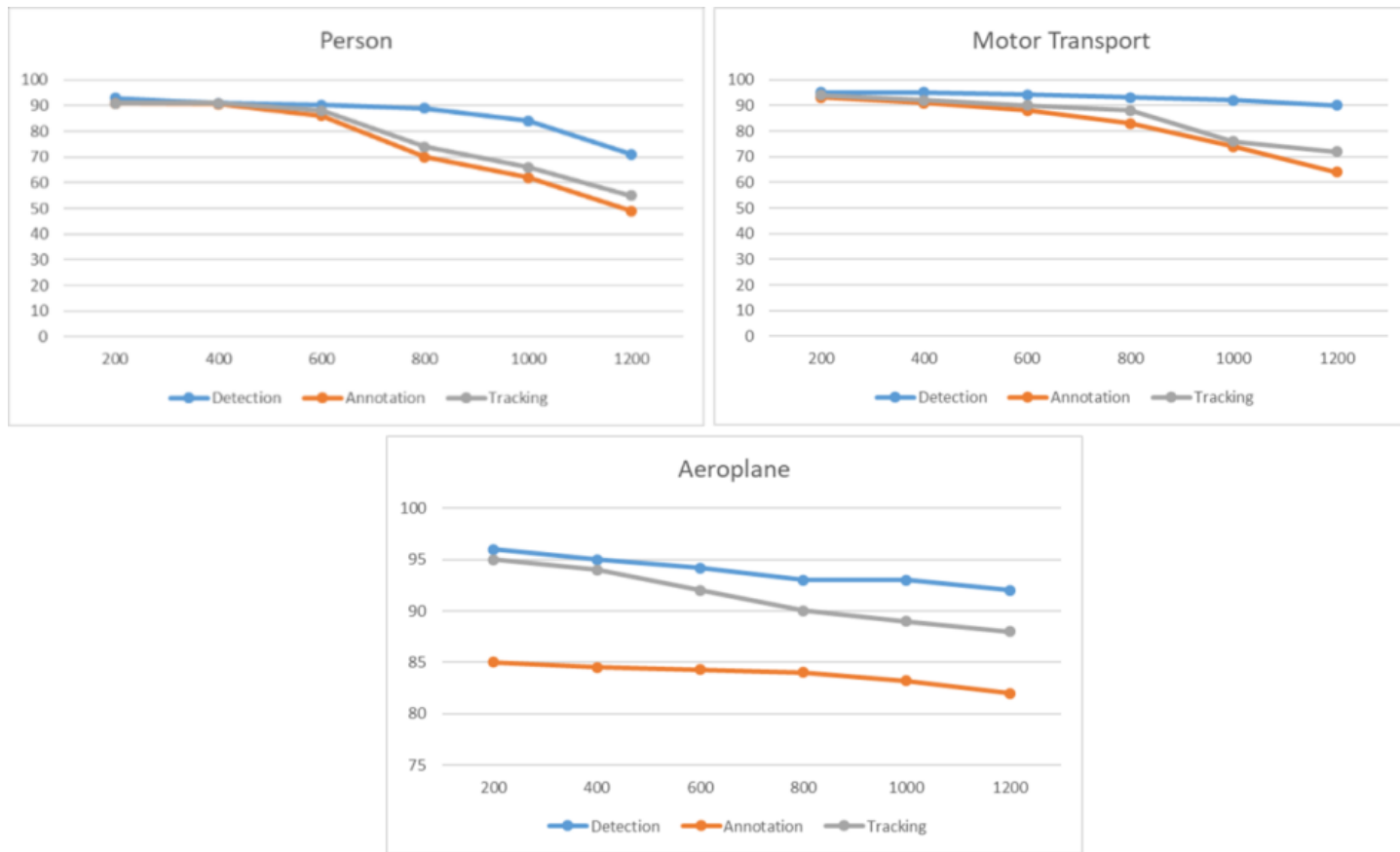

Figure 2. Relation between accuracy of the method stages (P) and distance (S) to an object

The research results on the test prototype show the acceptable accuracy of the proposed implementations of the method stages at a distance of 600 meters for people, 800 meters for motor transport and 1200 meters for airplanes. The lower accuracy of aircraft annotations at a close range is due to the fact that the available dataset of aircraft objects is less powerful than the dataset of people and motor transport.

6. Conclusion

The article addresses a number of runway safety problems related to incidents with ground objects. Due to these problems addressed and threats, an architecture of the specialized video surveillance system is proposed and runway incident probability prediction method by using AI technologies based on artificial neural networks is developed. The architecture of the system consists of four modules: moving objects detection, object annotation, object tracking and incident prevention. A technical solution is proposed for each of them. The applicability of the proposed technical solutions for solving real problems has been demonstrated on a test prototype. Suitability of the proposed approach is demonstrated as a solution for monitoring of movements of aircraft, vehicles, airport personnel and unexpected passengers on the runway, if a system configuration suitable for a specific airport is selected.

References

Beckmann, N., Kriegel, H., Schneider, R., & Seeger, B. (1990). The R*-tree: an efficient and robust access method for points and rectangles. *Proceedings of the 1990 ACM SIGMOD international conference on Management of data — SIGMOD '90*, (pp. 322-331).

Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking. *In Proceedings of the International Conference on Image Processing (ICIP). Phoenix, AZ, USA.*

Girshick, R. B., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Computer Vision and Pattern, 2014. CVPR 2014. IEEE Conference.*

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation, 9*(8), 1735–1780.

ICAO. (2006). *Aerodrome design manual. Part 1 Runways (Doc 9157).* ICAO.

ICAO. (2017). *Runway Safety Programme - Global Runway Safety Action Plan.* ICAO.

Itamar, T., Mechrez, R., & Zelnik-Manor, L. (2017). Template Matching with Deformable Diversity Similarity. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)*, (pp. 1311-1319).

Kalal, Z., Mikolajczyk, K., & Matas, J. (2010). Tracking-Learning-Detection. *IEEE transactions on pattern analysis and machine intelligence, 6*(1), pp. 1-14.

Matthews, P. (1995). Airports. In P. Matthews, *The Guinness Book of Records* (p. 128). Guinness Superlatives.

Ning, G., Zhang, Z., Huang, C., He, Z., Ren, X., & Wang, H. (2017). Spatially Supervised Recurrent Convolutional Neural Networks for Visual Object Tracking. *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, (pp. 1-4). Baltimore, MD.

Oppenheim, A. V., & Schafer, R. W. (1975). *Digital Signal Processing.* Prentice Hall.

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv:1804.02767.* Washington: University of Washington.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., . . . Rabinovich, A. (2015). Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (pp. 1-9). Boston, MA.

Zhao, Z., Zheng, P., Xu, S., & Wu, X. (2019, November). Object Detection with Deep Learning: A Review. *IEEE Transactions on Neural Networks and Learning Systems, 30*(11), 3212-3232.