The Flight Guardian: Autonomous Flight Safety Improvement by Monitoring Aircraft Cockpit Instruments

Wasiq Khan[1], Darren Ansell[2], Kaya Kuru[3], Muhammad Bilal[4]

[1]Manchester Metropolitan University, M15 6BH, UK {W.Khan@mmu.ac.uk}

[2,3]University of Central Lancashire, PR1 2HE, UK {[2]DAnsell@UCLan.ac.uk; [3]KKuru@UCLan.ac.uk}

[4]University of California, Los Angeles, CA 90095-7227, USA {M.Bilal@UCLA.edu}

During in-flight emergencies, a pilot's workload increases significantly and it is often during this period of increased stress that human errors occur that consequently diminish the flight safety. Research studies indicate that many plane crashes can be attributed to ineffective cockpit instrument monitoring by the pilot. This manuscript entails the development of Flight Guardian[5] (FG) system being first of its kind that aims to provide efficient flight-deck awareness to improve flight safety while assisting the pilot in abnormal situations. The system is intended to be used in older aircrafts that cannot easily or cost effectively be modified with modern digital avionic systems. One of the important feature of FG system being not physically connected to the aircraft, avoids any impact on airworthiness or the need for re-certification. For the first time, a composite of techniques including video analysis, knowledge representation, and machine belief representations are combined to build a novel flight-deck warning system. The prototype system is tested in both; simulation based Lab and real flight environments under the guidance of expert pilots. The overall system performance is evaluated using statistical analysis of experimental results that proved the robustness of proposed methodology in terms of automated warning generation in hazardous situations.

Keywords. **Flight Safety; Cockpit Monitoring; Intelligent Warning System; Decision Support System; Theory of Evidence; Automated Dial Reading.**

---

[1] Research Associate, School of Mathematics and Digital Technologies, JD Building, MMU, Manchester, M156BH.
[2] Aerospace Lead, School of Engineering, C & T Building, UCLan, Preston, PR1 2HE.
[3] Research Associate, School of Engineering, C & T Building, UCLan, Preston, PR1 2HE.
[4] Assistant Researcher, UC Centre for Environmental Implications of Nanotechnology, University of California, LA.
[5] Flight Guardian is a first of a generation disruptive cockpit technology part funded by NATEP (Grant Ref: NWC2-S006) to improve the safety of aircraft. Further details are available at: http://www.aerospace.co.uk/news/140915-natep-projects

I. Introduction

THE current generation of civil air transport aircraft are usually equipped with advanced Intelligent Warning Systems (IWS). One of the important aspects of these systems is the centralized alerting and monitoring system that displays information directly to the pilot. Typical aircraft warning systems monitor specific environmental properties that are difficult to observe or not observable by the pilot or flight crew. These systems generate warnings to alert the pilot about hazardous situation while keeping track of a number of parameters. When warnings occur in the cockpit, a pilot has procedures and actions to perform to mitigate the problem. During airborne emergencies this requires a great deal of attention directed towards resolving the detected hazard that degrades the flight safety by increasing the pilot workload significantly. Research studies [1-4] address the consequences of multitasking on human brain and ability to simultaneously deal with several parameters [5] as needed for cockpit monitoring by the human pilot.

The National Transportation Safety Board (NTSB) has long identified ineffective pilot monitoring of flight data to be problematic and therefore leading to crashes [6]. A safety study of crew-involved air carrier accidents between 1978 and 1990 [7] indicated that 31 out of 37 (i.e. 84%) reviewed accidents were caused by ineffective monitoring of the cockpit instruments. A detailed study is presented by Iwadare and Oyama [8] about recent plane crashes that constituted a huge number of fatal disasters in the aviation history. Seventy five percent of these recent crashes were caused by the pilot errors such as complacency, loss of control, lack of knowledge, distraction, lack of situational awareness, and lack of assertiveness [8]. These errors have marked some of the most fatal aviation disasters in history and have served as great learning examples on what to avoid in the future [9]. Likewise, statistics of aircraft crashes with respect to type of service, airlines, and number of fatalities per flight presented by Oster et al. [10] indicate a high number of accidents and fatalities caused by the pilot error. Following the recent breakdown investigation reports, the NTSB made a number of recommendations to the Federal Aviation Administration (FAA) for the flight safety improvement [11] that can be achieved by facilitating the pilots with automated operational and safety related information [12]. The NTSB also reported that the effect of technology and safety of general aviation operations can be improved by involving aircraft with glass cockpit display [13]. The aforementioned statistics and reports manifest that a reliable automated monitoring of the cockpit instrument will improve the flight safety and would also decrease the pilot workload to recognize the nature of problem and eye-scanning during emergencies.

Background

In general, a cockpit warning system performs tasks of hazard detection, attention-getting, display of resolution status and commands, and resolution guidance [14]. Current cockpit warning indicators give indications in hazardous situations but require a considerable amount of attention from the pilot. In addition, it takes significant amount of time to recognize the nature of problems and choose the right procedure [15]. To the best of our knowledge, there are no third party systems in use to enable the external monitoring of cockpit instruments. The most commonly used existing warning systems are Ground Proximity Warning System (GPWS) and Traffic Alert and Collision Avoidance System (TCAS) which are standalone applications that have improved aircraft safety [16]. Recent improvements in cockpit warning systems include enhanced- GPWS, predictive wind-shear warning systems, and the anticipated future airborne separation assurance system. An agent based cockpit safety system is presented by Thatcher [17] by using three autonomous agents for aircraft path monitoring and cockpit behavior. The first agent is situated on the ground in the air traffic controller center of each airport. This agent uses an artificial neural network for training on a dataset of successful landings at this airport. The dataset consists of beliefs centered on the current landing, the aircraft's altitude, horizontal and lateral distance from landing strip, speed, heading, type of aircraft, and weather conditions that affect the landing performance. In case of unusual events during the landing, a warning alarm is generated to air traffic controller. The second and third agents are situated on-board the aircraft to monitor parameters such as aircraft position, heading, speed and the pilot's behavior. These agents use harmonic topographic map and topographic products of experts for the prediction of hazardous situation.

Despite of the technological advancements which have made IWS very sensitive to hazardous situations, tracing the exact problem and resolving it using a correct procedure is still a challenge that needs a considerable attention [17]. Due to the increasing number of cockpit warning systems in modern aircraft, the need to prioritize warning messages to avoid nuisance alerts and provide a more intuitive human interface becomes more apparent [16]. As an example, one of the challenging factors in the development of an effective IWS is the potential for false alarms. Alarms could be erroneously triggered during real time operation due to dynamics such as air turbulence, sunlight reflections, night time camera performance, and limitations of desired sensor/camera technology that could misidentify objects in the cockpit. These false alarms can result in pilot's general distrust of the system. The problems associated with false and nuisance alarms were demonstrated in the design of the first generation of TCAS system for commercial aircraft. These systems had such a high nuisance alarm rate that pilots failed to trust the system's validity as an alerting device. This situation could be analogous to what may happen in the automotive crash-warning domain [18].

Aforementioned studies indicate the significance of a reliable, intelligent, easy to expand, and supportive warning system for pilot assistance. The FG prototype presented in this manuscript provides such a platform using composite of techniques to improve the flight safety and pilot support. Image processing techniques are employed for Automated Dial Reading (ADR) to produce instantaneous state information of the cockpit instruments. For the first time, domain knowledge with respect to the appropriate range of cockpit instrument readings is presented which is gathered from domain experts to prove the proposed concept of FG system. The theory of evidence is used to build the Decision Support System (DSS) that combines beliefs from two cockpit instruments for the intelligent reasoning. In abnormal conditions, the FG system generates spoken and visual alerts and resolution advice after diagnosing the problem that helps pilot to choose the corresponding procedure for mitigating the hazardous situations.

II. Materials and Methods

A composite of techniques was sequentially combined to build a FG system as shown in Fig. 1. The inputs to the system consist of two video streams of cockpit instruments (analogue) acquired by the camera devices which were then processed by a sequence of ADR, data representation, transformation, and probabilistic modelling approaches to inform a final decision making component to decide if flight parameters need to be adjusted in order to return the aircraft to a safe state.

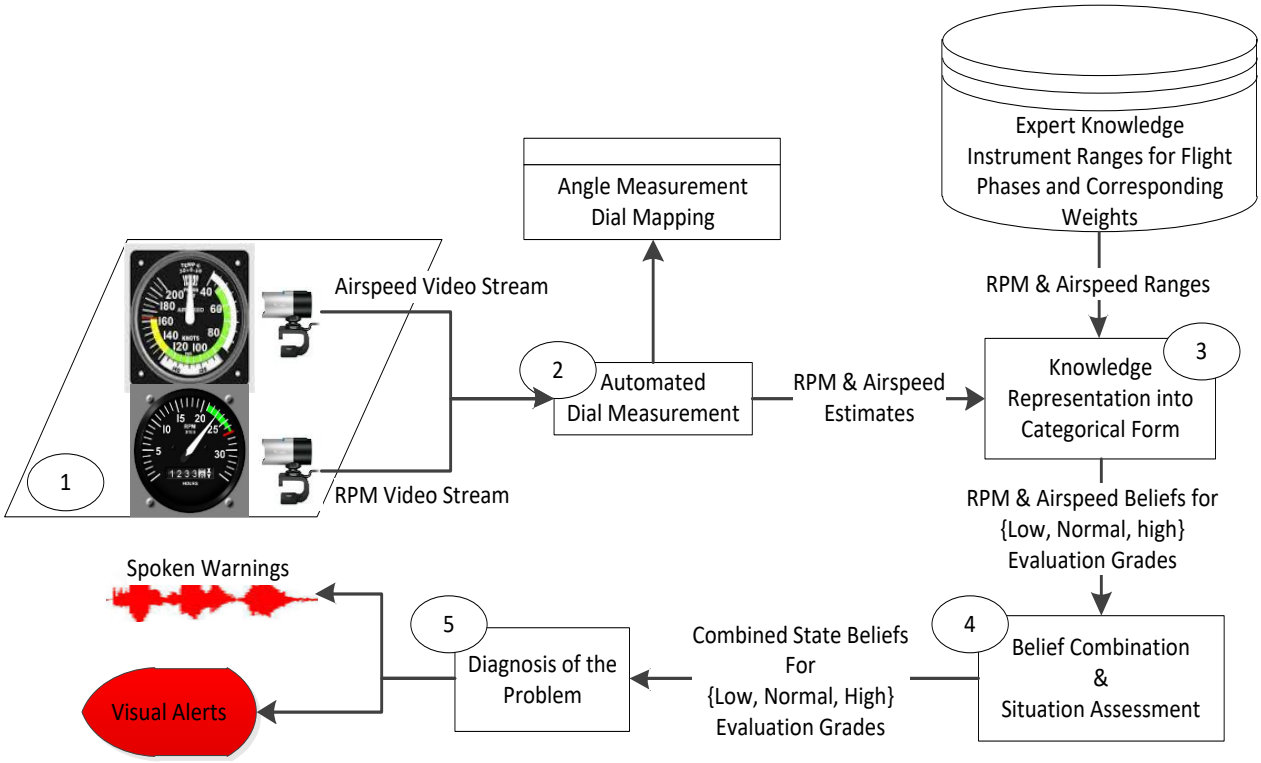


Fig. 1 Sequential Processing of the Deployed Methods for the Flight Guardian System Composition.

A. Data Acquisition and Automated Dial Reading

In the first step, two fixed position camera devices are placed in front of the Airspeed (ASp) and Revolutions per Minute (RPM) instruments to acquire the corresponding video streams as shown in Fig. 1. Microsoft Lifecam Cinema devices are used that produce 30 frames per second. The video streams are then processed to generate instantaneous snapshots (i.e. images) with a frequency of 0.5 Hz and forwarded to ADR component. The primary objective of the ADR is to provide instantaneous state evidences (i.e. current readings) for the cockpit instruments to the DSS. In regards to ADR, different approaches are explored which have been used in the literature. For instance, [19] presented a background subtraction based approach that subtracts the current image frame from a reference frame to identify the target object's movements in the scene. Similarly, pattern matching [20] provides the object segmentation based on the features extracted from the input image. Likewise, Gellaboina [21] presented a polar representation of the dial gauge image to identify the needle. These approaches are generic and simple to implement but exhibit poor performance in the presence of specular highlights [21] that may occur due to sunlight reflections on the cockpit instruments during flight.

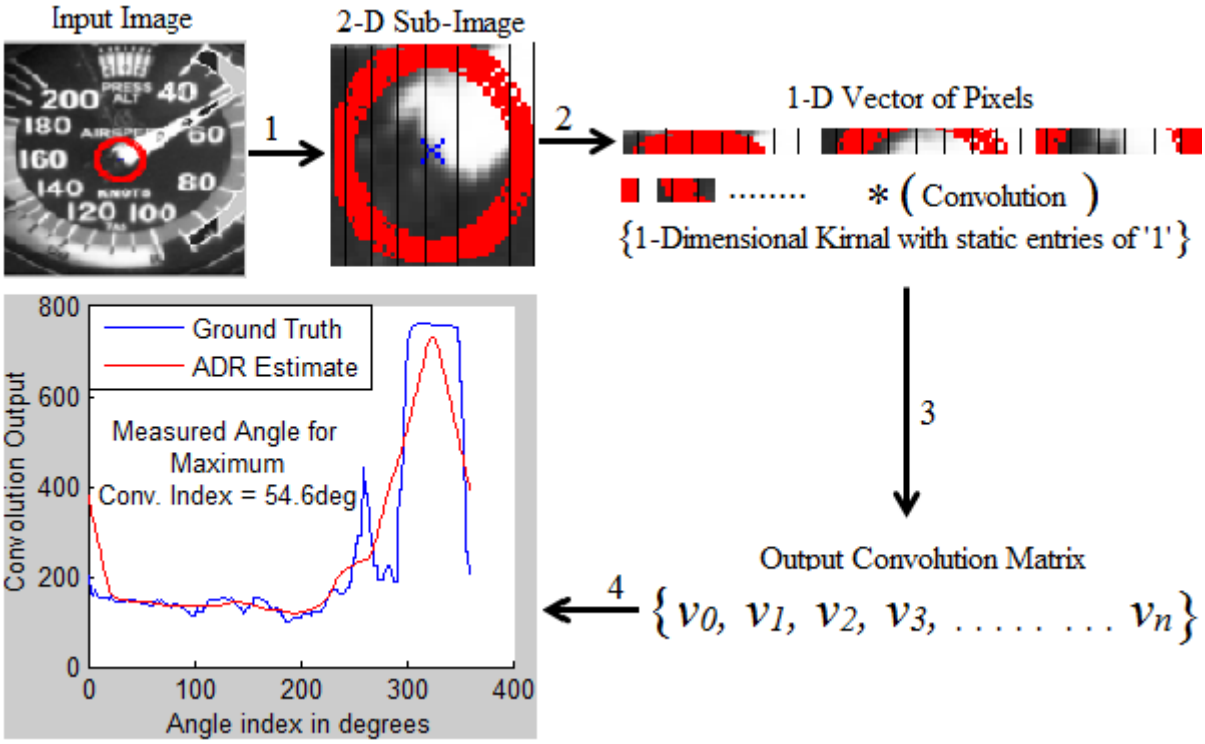


Fig. 2 Estimation of the Gauge Angular Position within the Airspeed Dial using Convolution

To deal with this issue, a convolution based ADR approach is used to detect the needle within the ASp and RPM dials as presented in our previous work [22]. Because of the circular shape and needle locations in both dials, focus of the camera devices is fixed at center coordinates of the cockpit instruments as shown in Fig. 2. It can be observed

that all pixels are black within a limited circular range (red circle in Fig. 2) around the image center except the white pixels occupied by the needle. Therefore, only a sub-image (within the red circle) is needed to identify the needle position in the cockpit instruments. In the second step, a convolution function with a static one-dimensional kernel is used to identify the needle and hence its position in the dial. As the entire pixels occupied by the target needle are white, the entries of the kernel vector are set to '1' such that it produces maximum convolution output for only white pixels within the cropped sub-image. The entire convolution process between selected sub-image pixels '*x*'and a kernel '*h*' over a pre-defined interval is described as a pseudo code below.

Inputs:

1- Dimensional vector $'x'$ for target **sub-image** pixels

1- Dimensional **Kernel** vector $h = \{1,1,1,\ldots. \ length\ (x)\}$

Output:

Angle $'\theta'$; maximum convolution index in output vector $'v'$

for all indexes $'i'$ in $'x'$ **do**

$v[i] = 0;$

for all indexes $'j'$ in $'h'$ **do**

$v[i] = v[i] + x[i-j] * h[j];$

end for

end for

$\max\_index$ = index of maximum value in the output vector $'v'$

$\theta = 360 / (length - of - x)\ *\ \max\_index$

B. Knowledge Representation

In the next step, expert knowledge is acquired from domain experts at The Great Circle[6] and related information booklets [23, 24] that demonstrate the ideal ranges for cockpit instruments (i.e. airspeed and engine revolution/minute) of Cessna C172SP aircraft for different flight phases as presented in Table 1. Based on the selected model for evidence combination and situation assessment, the numeric data in the form of cockpit instruments ranges and ADR measurements is transformed into the categorical form (i.e. *low*, *normal*, *high*). The $'Gaussian'$, $'Z'$and$'S'$Membership Functions (MFs) are used to interpret the numeric data into a categorical form with a degree of belief. Table 1 presents detailed information about the instrument ranges and the parameter values for these MFs with respect to different flight phases. These parameters describe the MFs in terms of their ranges within the data.

[6] The Great Circle Ltd and the University of Central Lancashire are working in partnership on the Flight Guardian Project. For further information about GC, follow the link: http://www.thegreatcircle.co.uk.

Table 1 Cockpit Instruments Ranges for Different Flight Phases and Parameter Setting for Gaussian, Z, and S Member Functions

FLIGHT PHASES	Engine Noise (RPM) Ranges	Airspeed (ASp) Ranges (Knots)	MF Parameters for Air-Speed Representation						MF Parameters for RPM Representation					
			Z		Gaussian		S		Z		Gaussian		S	
			a	*b*	*c*	*σ*	*a*	*b*	*a*	*b*	*c*	*σ*	*a*	*b*
Take-Off	**Low:** RPM<2000 **Normal:** 2000<RPM<2600 **High:** RPM>2600	**Low**: ASp<5 **Normal:** 5<ASp <60 **High:** ASp>65	0	3	30	30	60	70	1800	2000	2200	400	2400	2800
Climb	**Low:** RPM<2000 **Normal:** 2000<RPM<2600 **High:** RPM>2600	**Low:** ASp<55 **Normal:** 60<ASp<80 **High:** ASp>90	45	60	72	20	75	95	1800	2000	2200	400	2400	2800
Cruise	**Low:** RPM<1750 **Normal:** 1800<RPM<2500 **High:** RPM>2600	**Low:** ASp<60 **Normal:** 70<ASp<100 **High:** ASp>110	50	85	85	25	85	115	1700	1900	2150	450	2350	2700
Descent	**Low:** RPM<1600 **Normal:** 1800<RPM<2000 **High:** RPM>2300	**Low:** ASp<60 **Normal:** 70<ASp<90 **High:** ASp>100	50	70	80	20	75	115	1500	1900	2000	350	2150	2350

The parameter $'a'$ and $'b'$ locate the lower and upper bounds within the input numerical data respectively, for $'Z'$ and $'S'$ MFs. Similarly, $'c'$ and $'\sigma'$ represent the center and width parameters respectively for the $'Gaussian'$ MF. A detailed discussion about these member functions and their representation is presented in [25]. These functions take the ASp and RPM reading estimates along with the expert knowledge and transform the data into the degree of beliefs for multiple evaluation grades corresponding to each flight phase. These degrees of beliefs are then forwarded to an evidence combination stage for the flight situation assessment. Figure 3 demonstrates the categorical form of data representation for the ASp ranges in a cruise flight phase presented in Table 1 using the aforementioned member functions and the corresponding parameters.

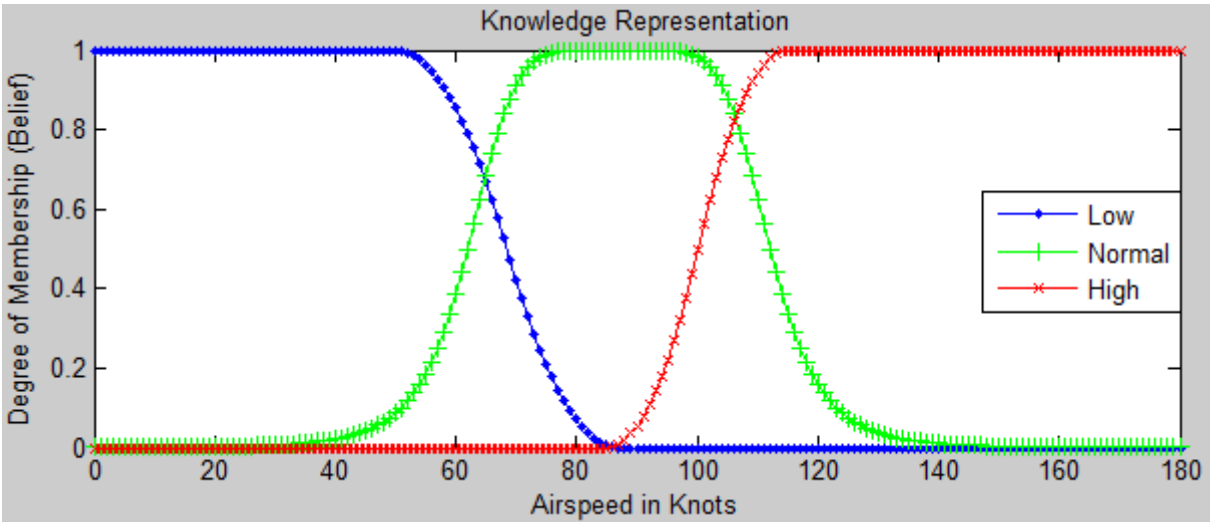


Fig. 3 Categorical Representation for the Airspeed Ranges in *Cruise* Phase using *Gaussian*, *S*, and *Z* Membership Functions

C. Evidence Combination

The categorical state information from the previous step is then forwarded to the evidence combination process that uses the Dempster-Shafer (DS) theory of evidence also considered as a generalization to the Bayesian theory in such a way that it can handle the degree of ignorance [26]. The DS provides the best state estimate by combination of evidences from multiple information resources while taking into account the corresponding weights. One of the interesting advantages of the DS is the model simplicity for the complex multi-layered situations where the system can be decomposed into many layers of simpler states and then the beliefs can be propagated upwards, combined with sibling layer states to get overall belief [27]. A detailed study on DS advantages, disadvantages, and its application areas is presented in [26, 27]. In the proposed study, the DS performs the flight situation assessment task using beliefs from two cockpit instruments (i.e. ASp and RPM) in different flight phases. Mathematical formulation of the DS for the FG system is presented in the following steps.

Let $E = \{AS_P, RPM\}$ represents the set of basic attributes representing the evidences produced by ASp and engine revolution speed respectively. The relative weights for the basic attributes are equally distributed according to the pilot instructions such that $0 \le \omega_i \le 1$ and that the weights sum to 1. The distinctive evaluation grades for each attribute are defined as set of three entities, i.e.

$$H = \{low, normal, high\} \tag{1}$$

For each attribute in 'E' and evaluation grade 'H', a degree of belief β_n is assigned by the member functions described earlier which denotes the source's level of confidence when assessing the level of fulfilment of a certain property.

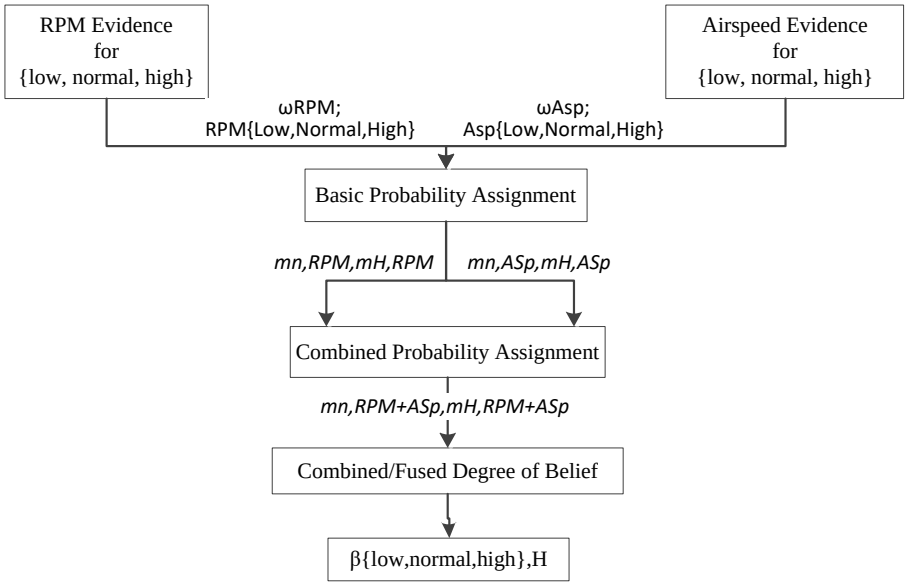


Fig. 4 Combining the Beliefs from Multiple Cockpit Instruments for being in "*Low, Medium,* and *High*" States

i. *Basic Probability Assignments for Each Basic Attribute*

Let $m_{n,i}$ be a basic probability mass representing the degree to which the i^{th} basic attribute is assessed. A hypothesis that the general attribute is assessed to the n^{th} evaluation grade H_n can be presented as:

$$m_{n,i} = \omega_i \beta_{n,i} \tag{2}$$

Where '$n = 3$' are the number of evaluation grades (*i.e.* $low, normal, high$). The remaining probability mass $m_{H,i}$ unassigned to each basic attribute is calculated as:

$$m_{H,i} = 1 - \sum_{n=1}^{N} m_{n,i} = 1 - \omega_i \sum_{n=1}^{3} \beta_{n,i} \tag{3}$$

The remaining probability mass is further decomposed into $\bar{m}_{H,i}$ and $\tilde{m}_{H,i}$ as:

$$\bar{m}_{H,i} = 1 - \omega_i \tag{4}$$

$$\tilde{m}_{H,i} = \omega_i \left(1 - \sum_{n=1}^{3} \beta_{n,i} \right) \tag{5}$$

Eq. (4) measures the degree to which final attributes have not yet been assessed to individual grades due to the relative importance of basic attributes after their aggregation. Eq. (5) measures the degree to which final attributes cannot be assessed to individual grades due to the incomplete assessments for basic attributes.

ii. *Combined Probability Assignments*

In next step, the probability mass of the basic attributes are aggregated to form a single assessment for the *low/normal/high* grades. The combined probability masses can be generated using the following set of recursive evidence reasoning equations:

$$\{H_n\}:$$
$$m_{n,i+1} = K_{i+1}[m_{n,i}.m_{n,i+1} + m_{H,i}.m_{n,i+1} + m_{n,i}.m_{H,i+1}] \tag{6}$$
$$n = 1, ..., N$$

Where $i = \{1, 2\}$ represents the number of basic attributes. In the above equation, $m_{n,1}.m_{n,2}$ measures the degree of both attributes $\{AS_p, RPM\}$ supporting the general attribute of decision to be assessed to H_n. The term $m_{n,1}.m_{H,2}$ measures the degree of only 1st attribute $\{AS_p\}$ supporting decision to be assessed to H_n. The term $m_{H,1}.m_{n,2}$ measures the degree of only 2nd attribute $\{RPM\}$ supporting final belief to be assessed to H_n.

$$m_{H,i} = \bar{m}_{H,i} + \tilde{m}_{H,i} \tag{7}$$

$$\tilde{m}_{H,i+1} = K_{i+1}[\tilde{m}_{H,i}.\tilde{m}_{H,i+1} + \bar{m}_{H,i}.\tilde{m}_{H,i+1} + \bar{m}_{H,i+1}.\tilde{m}_{H,i}] \quad (8)$$

$$\bar{m}_{H,i+1} = K_{i+1}[\bar{m}_{H,i}.\bar{m}_{H,i+1}] \quad (9)$$

$$K_{i+1} = \left[1 - \sum_{t=1}^{N=3} \sum_{\substack{j=1 \\ j \neq t}}^{N=3} m_{t,i}.\, m_{j,i+1}\right]^{-1} \quad (10)$$

In Eq. (8), $\tilde{m}_{H,1}.\tilde{m}_{H,2}$ measures the degree to which the final attribute cannot be assessed to any individual grades $low, normal, high$ due to the incomplete assessments for both attributes $\{AS_p, RPM\}$. Term $\bar{m}_{H,1}.\tilde{m}_{H,2}$ measures the degree to which final attributes cannot be assessed due to the incomplete assessments for $\{AS_p\}$ only. In Eq. (9), $\bar{m}_{H,1}.\bar{m}_{H,2}$ measures the degree to which final attributes have not been assessed yet to individual grades due to the relative importance of $\{AS_p\}$ and $\{RPM\}$ after $\{AS_p\}$ and $\{RPM\}$ have been aggregated. The normalization factor 'K' is used to normalize m_n, m_H such that $\sum_{n=1}^{2} m_n + m_H = 1$.

iii. *Calculation of the Combined Degrees of Belief*

Finally, let $c\beta_n$ denote the combined degree of belief that the situation is assessed to the grade H_n which is generated by combining the assessments for all the associated basic attributes $E = \{AS_p, RPM\}$, then $c\beta_n$ is calculated by:

$$\{H_n\} : c\beta_n = \frac{m_{n,i}}{1 - \bar{m}_{H,i}} \qquad n = 1,.., N \quad (11)$$

$$\{H\} : c\beta_H = \frac{\tilde{m}_{H,i}}{1 - \bar{m}_{H,i}} \quad (12)$$

Eq. (12) for $c\beta_H$ measures the belief that is left unassigned during the assessments. The set of equations: Eq. (1)-(12) provides the combined degree of beliefs for the $low, normal,$ and $high$ grades that are further used for the decision making about a *normal* or *abnormal* flight situation.

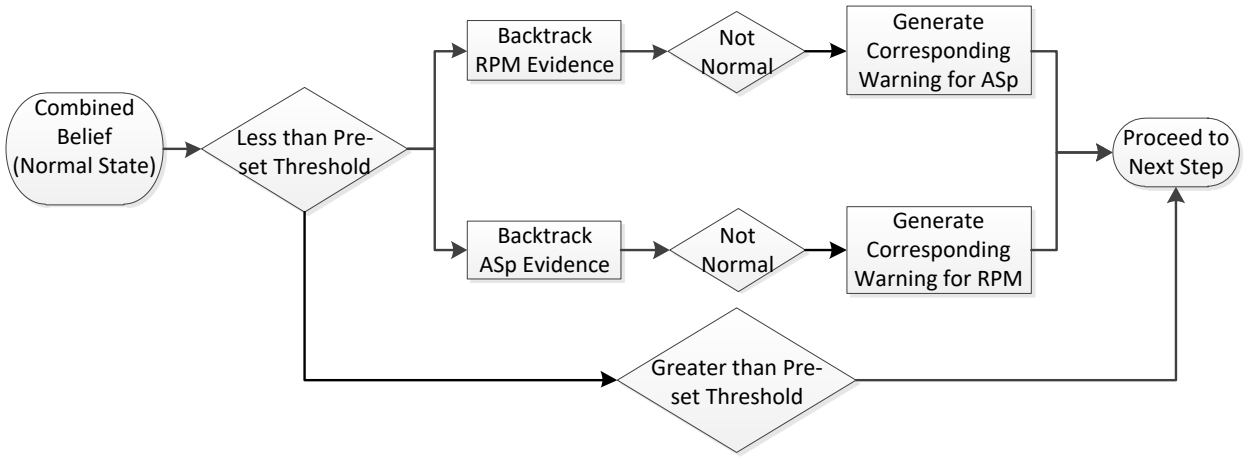


Fig. 5 Diagnosis of Hazard Cause by Back-tracking the Individual Beliefs from RPM and Airspeed, Corresponding ADR Measurements, and Domain Knowledge in Case of Abnormal Situation

A flight state is considered as '*abnormal'* if the combined belief for '*normal*' state is less than a pre-defined decision threshold as shown in Fig. 5. In case of abnormal situation, the problem is identified by automated analysis of the corresponding instrument state information (i.e. backtracking individual beliefs) and corresponding measurements retrieved from the ADR to identify the required adjustments with regards to the domain knowledge presented in Table 1. Finally, a warning message is generated to the pilot that provides precise information about the diagnosed problem and instructions to return the aircraft to a safe state.

D. Experimental Setup and System Evaluation

Performance of the proposed system is evaluated using different experimental setup that includes:

1. Lab Environment Using a Flight Simulator

In a Lab environment, experiments are conducted in a simulated flight using the Flight Simulator (FSX) platform. Two fixed position camera devices are placed directly in front of the ASp and RPM instruments that export continuous video streams to the processing device. The video streams are processed and converted into instantaneous snapshots (i.e. images) with a frequency of 0.5 Hz and forwarded for further processing. Microsoft Lifecam Cinema devices are used that produces 30 fps. In addition, virtual software camera (i.e. manyCam) is used to provide the zoom function to physical camera devices. The domain knowledge along with the simulated flight data are used to know the ground truth (i.e. true state) of whether the system was in a *normal* or *abnormal* situation which is then compared to the DSS output for that particular instance to evaluate the performance. As the output warnings generated by FG system are in the binary form of '*normal'* and '*abnormal'* situations, the statistical metrics for a binary classifier are used for the DSS performance evaluation that includes: (1) Sensitivity, (2) Specificity, and (3) Accuracy.

a) Lab Environment **b)** Real Flight Environment

Fig. 6 Experimental Setup for the Flight Guardian System in Simulated and Real Flight Environments

The sensitivity indicates the probability an instance is classified as a warning with high confidence in case of actual hazard. Specificity shows the probability an instance is classified as a normal situation with high confidence when it is indeed a normal situation and therefore no warning is generated by the DSS. Accuracy is a percentage of all the instances that were correctly classified. Detailed experiments are conducted for various conditions in terms of different flight phases (i.e. take-off, climb, cruise, and descent), decision boundaries, varying light intensity, and environmental dynamics to evaluate the ADR performance. Fig. 6(a) shows the experimental setup in the Lab environment for a simulated flight.

2. Real Aircraft Flight Environment Under the Guidance of Experienced Pilots

Despite of the high expenses associated with the real flight experimental setup, the proposed FG prototype is tested twice using Cessna C172SP aircraft in the presence of two expert pilots (one pilot in each flight). The selected method is evolved according to the test outcomes from each flight test to mitigate the real flight challenges. There is no significant change in the experimental setup for the real flight environment except the camera position is placed such that it does not interfere with the pilot's internal or external view as shown in Fig. 6(b). However, the automated ground truth data extraction from the cockpit instruments in real flight environment is not possible specifically, in older aircrafts. In this scene, a video data is recorded while testing the FG system in real flight by creating different hazardous scenarios by the pilot. The recorded video is then converted into image instances with a frequency of 0.5 Hz (i.e. execution time) and the ground truth (i.e. true states) are generated manually to retrieve the statistical results using the aforementioned binary classification metrics. In addition, the overall system performance is evaluated for multiple ADR approaches to identify the best one in terms of accuracy. Finally, a combined

statement is given by the pilots at the end of flight tests which empowers the significance of the proposed system for the flight safety and pilot assistance.

III. Results and Discussions

The experimental results are achieved to evaluate the FG system performance under the experimental setup described earlier. These statistical results are highly dependent upon the weight assigned to ASp and RPM measurements. However, both instruments are weighted equally according to the pilot instructions. In addition, the threshold value for the decision boundary for an instance to be considered as *normal* or *hazard* is defined using the Receiver of Characteristics (ROC) curve analysis as shown in Fig. 7.

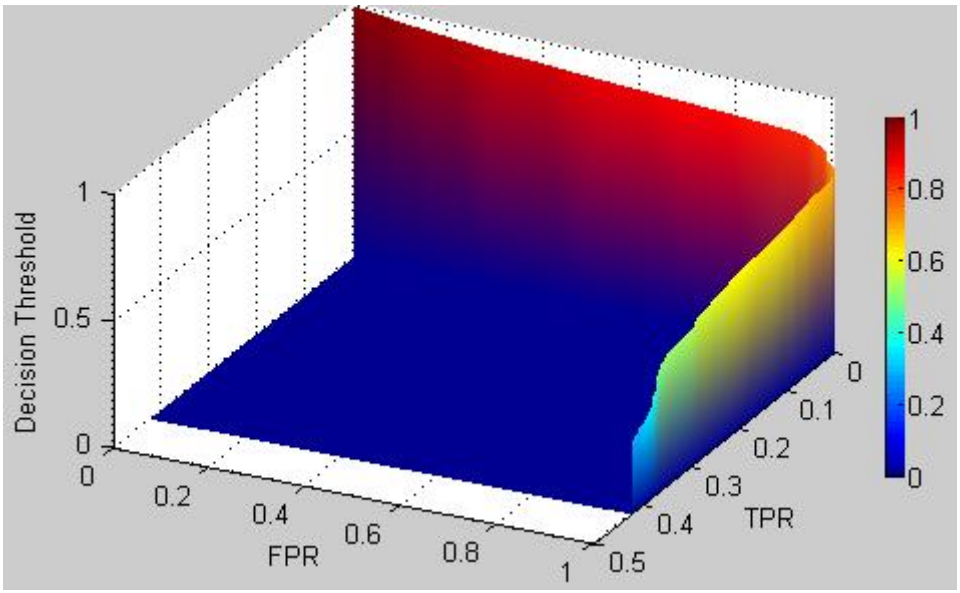


Fig. 7 ROC Curve for the Optimal Decision Boundary Selection using True Hazard Detection and False Alarms Rate

The threshold value is varied from 0 to 100 with a lag of 1 unit and the DSS performance is evaluated in terms of True Positive Rate (TPR), and False Positive Rate (FPR). An ROC curve for the decision boundary selection is achieved which indicated the optimal compromise between TPR and FPR for the overall flight at 0.72 (i.e. 72%) decision threshold value. Except the Take-off flight phase where the decision threshold slightly decreases (i.e. 70 %), it remains almost stable for the rest of flight phases. This is because in Take-off phase, an additional turbulence is generated in the Aircraft resulting the small vibration in the fixed positioned camera devices. Statistical results in Table 2 demonstrate the impacts of different ADR approaches on the performance of the FG system. It can be observed that the best performance in terms of TPR, TNR, and accuracy is obtained using the convolution based ADR that was introduced in our previous work [22]. Because of the rapid state variations in real flight, the

computation time is also an important factor to be considered. Table 2 indicates that the optimal computation time of 0.41 seconds is achieved for the background-subtraction based ADR approach [19] however; it sacrifices the main aim by producing 59.25% accuracy rate. Comparatively, the convolution based ADR produces the best accuracy rate of 96% and 100% in real flight and Lab environments respectively along-with an efficient computation time (i.e. 0.5 seconds).

Table 2 FG Performance for Different Automated Dial Reading Approaches in a Real and Simulated Flight Environments

	ADR approach	Instances /Scenarios					True Positive Rate (%)	True Negative Rate (%)	Accuracy (%)	Execution Time (Sec)	Frequency (Hz)
		Total	TP	TN	FP	FN					
Live Flight Test for (Take-off, Climb, Cruise, and Descent Phases)	Convolution	1350	450	863	37	0	100	95.9	97.20	0.52	0.5
	Background Subtraction	1350	200	600	300	250	44.40	66.66	59.25	0.41	0.5
	Pattern Matching	675	160	355	95	65	71.11	78.80	76.29	2.12	0.25
Lab Test for (Take-off, Climb, Cruise, and Descent Phases)	Convolution	1805	400	1405	0	0	100	100	100	0.52	0.5
	Background Subtraction	1805	300	1004	401	100	75.00	71.42	72.02	0.41	0.5
	Pattern Matching	902	177	650	52	23	88.75	92.52	91.68	2.12	0.25

Existing cockpit warning indicators give indications in hazardous situations but needs a considerable amount of attention from the pilot. In addition, it takes a lot of time to recognize the nature of problem and choosing the right procedure [15, 28], that consequently influences the flight safety in emergency situations. The proposed FG system resolves these issues by providing robust hazard detection in emergency situations that reduces the pilot workload in terms of recognizing the nature of problem. For instance, providing direct information related to hazard type, its cause, and instructions to mitigate the abnormal situation will significantly reduce the time spent by the pilot to identify the hazard cause and eye-scanning for sub-tasks. In comparison to Lab environment, the system generated 4% extra false alarms in real flight and hence reducing the system accuracy by 2.8%. This reduction occurs due to the dynamics associated with real flight environment such as light reflections and vibrations in the cameras due to air turbulence producing blurry images. Consequently, identification of the needle/gauge (i.e. ADR) becomes more challenging resulting in 3.9% false alarms. Despite of these challenges in real flight, the 100% rate of hazardous situation detection proves the FG system significance in terms of flight safety improvements and pilot assistance as discussed in [11, 12].

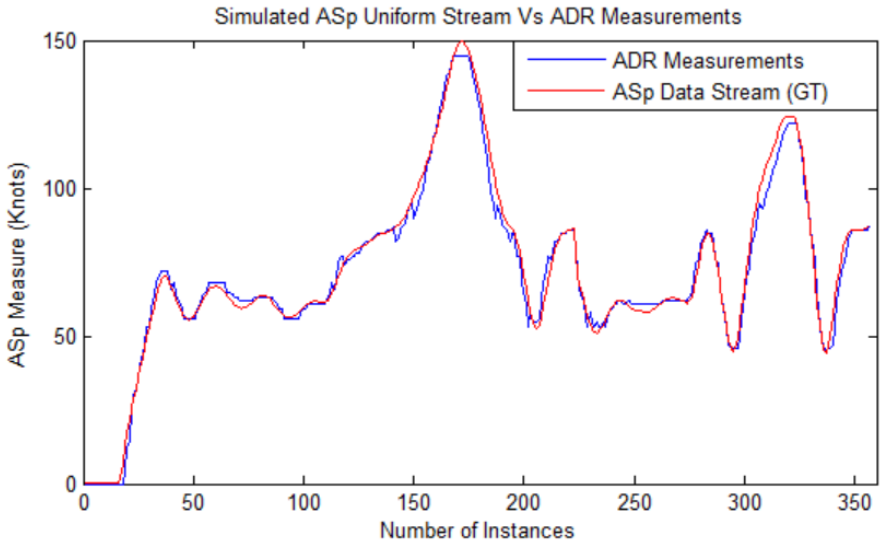


Fig. 8 Comparison between Simulated Uniform Data Stream Vs ADR Measurements

In addition to the aforementioned statistical results, the dial reading functionality is tested in the presence of specular highlights and varying light intensity while acquiring a uniform data stream using a flight simulator (FSX). A comparison between the uniform simulated data stream for Asp measurements (i.e. ground truth) acquired from FSX and predicted measurements by convolution based ADR system in the Lab environment is presented in Fig. 8. It can be analyzed that the proposed ADR approach performed robustly in terms of precise overlapping of the ground-truth stream that produces a negligible root mean squire error (1.63 in Knots). As the needle color in ASp and RPM instruments is white, it is challenging to differentiate the light reflections from actual needle as discussed by Vega et al. [20] and Gellaboina et al. [21]. However, the proposed convolution based ADR along with the circular constraints makes it more accurate by minimizing the probability of existence of light reflections in such a tiny area of the cockpit.

Statistical results achieved for varying lighting conditions in a Lab simulated environment are presented in Table 3. The intensity threshold for output images is adjusted by varying the contrast limits between 0 and 1 for two parameters (α, β). First parameter increases the brightness level from default (0) to complete bright image by increasing its values whereas; the second parameter reduces the brightness level from default (1) to complete dark image by decreasing its values. The p-Values are calculated using t-Test for the ground truth (GT) stream and ADR measurements while varying the ‘α’ and ‘β’ parameters in the simulated environment. It can be observed that the convolution based ADR produced consistency in output p-Value with 0.01 (i.e. 1% significance level) in dynamic lighting conditions that indicated the strong confidence in support of the hypothesis that there is no significance difference between the GT and ADR measurements data streams. A slightly higher p-Value (i.e. 0.02) is observed in the presence of direct specular light within the convolutional circular area around the needle that caused the miss-

identification of needle. On the other hand, p-Value is significantly increased when calculated for pattern matching (p=0.19) and background subtraction (p=0.49) based ADR approaches that indicate the significance difference between the GT data streams and the ADR measurements for these approaches. Likewise the root mean square error (RMSE) and false alarm rate are negligible for the convolution based method that is not the case for other approaches. The hypothetical test and statistical measures are the evident of the proposed convolution based ADR approach over the existing background subtraction and pattern matching based approaches.

Table 3 Performance Evaluation with Varying Lighting Conditions Using Different Automated Dial Reading Approaches in the Simulated Flight Environment

Lighting Conditions		*Convolution* based ADR			*Pattern matching* based ADR			*Background Subtraction* based ADR		
		False alarm (%)	p Value	RMSE	*False alarm* (%)	p Value	RMSE	*False alarm* (%)	p Value	RMSE
Room Light		0	0.01	0.25	5.80	0.09	4.50	22.40	0.29	13.84
α= 0	β= 1									
Lower Ambient Light		0	0.01	0.31	6.89	0.09	5.81	27.01	0.31	15.32
α= 0	β= 0.5									
Dark Ambient Light		0	0.01	0.35	8.09	0.11	5.95	29.20	0.33	17.44
α= 0	β= 0.3									
Higher Ambient Light		0	0.01	0.14	7.91	0.10	5.71	27.33	0.32	14.88
α= 0.5	β= 1									
Bright Ambient Light		0	0.01	0.22	8.87	0.12	6.23	28.15	0.34	16.01
α= 0.7	β= 1									
Direct Specular Light		1.7	0.02	0.64	17.03	0.19	8.13	42.00	0.49	31.15
Indirect Specular Light		0	0.01	0.17	6.23	0.09	4.81	34.92	0.43	24.92

Figure 9 shows multiple scenarios for the FG performance in the presence of specular highlights in the cockpit instruments and varying lighting conditions in the simulated flight environment. It is analyzed that the proposed ADR performed robustly in dynamic light intensity conditions by differentiating the pixels occupied by specular highlights and target object (needle) efficiently that consequently increases the performance of correct hazard identification and situation assessment.

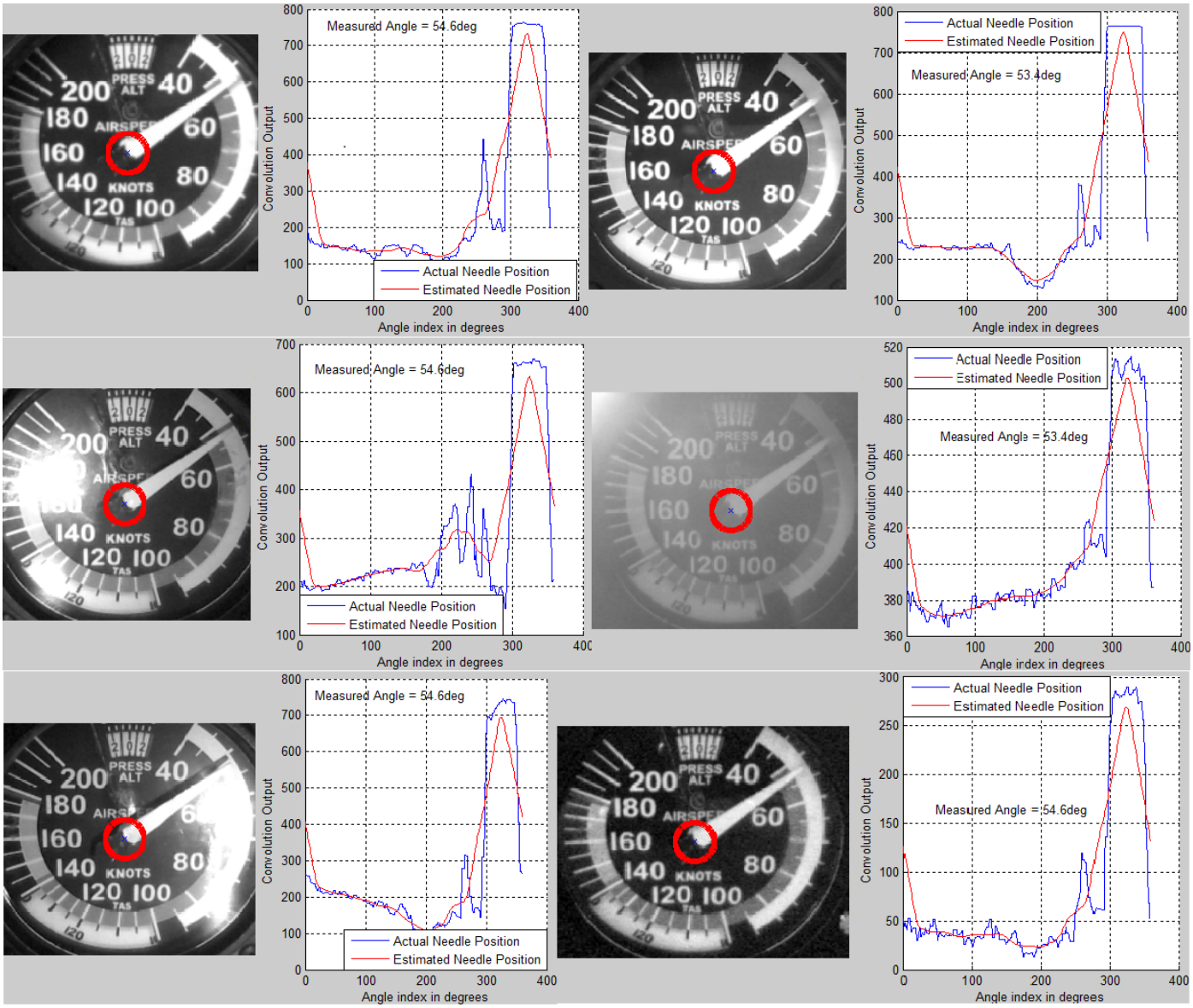
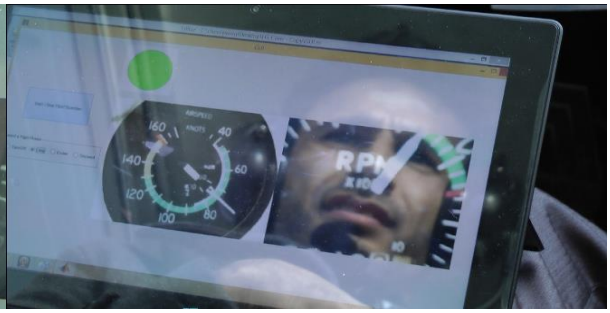


Fig. 9 Convolution based automated dial reading for varying light intensities and existence of specular highlights

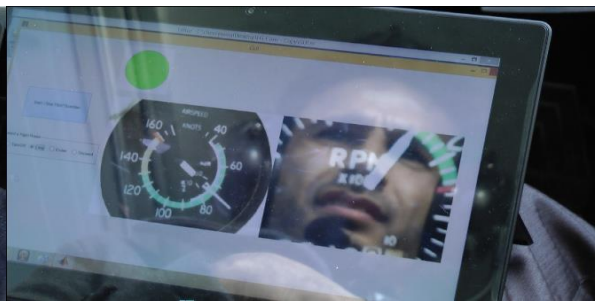
In addition to earlier described statistical analysis of the results, multiple hazardous scenarios are created in real flight environment using ASp and RPM instruments control to evaluate the FG system responses in emergency situations. Fig. 10 demonstrates the visual snapshots taken for different hazardous situations created by the pilot during real flight while exceeding and dropping-down the ASp and RPM values in climb, cruise, and descent flight phases. Despite of the real flight dynamics (e.g. air turbulence and sunlight reflections); the proposed methodology produced a robust hazard diagnosis and provided corresponding visual and spoken instructions to the pilot to mitigate the abnormal situation.

(i) Climb Normal Airspeed



(ii) Climb High Airspeed

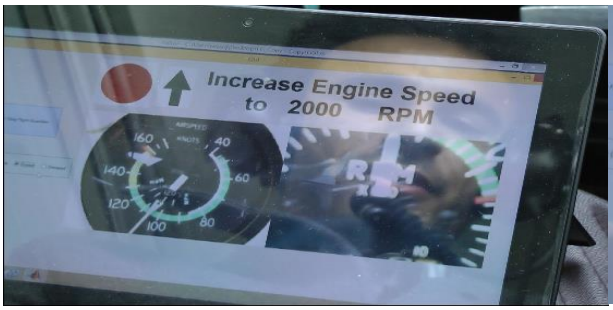
(iii) Cruise Low Airspeed



(iv) Climb Normal RPM

(v) Cruise Normal Airspeed & RPM

(vi) Cruise High Airspeed

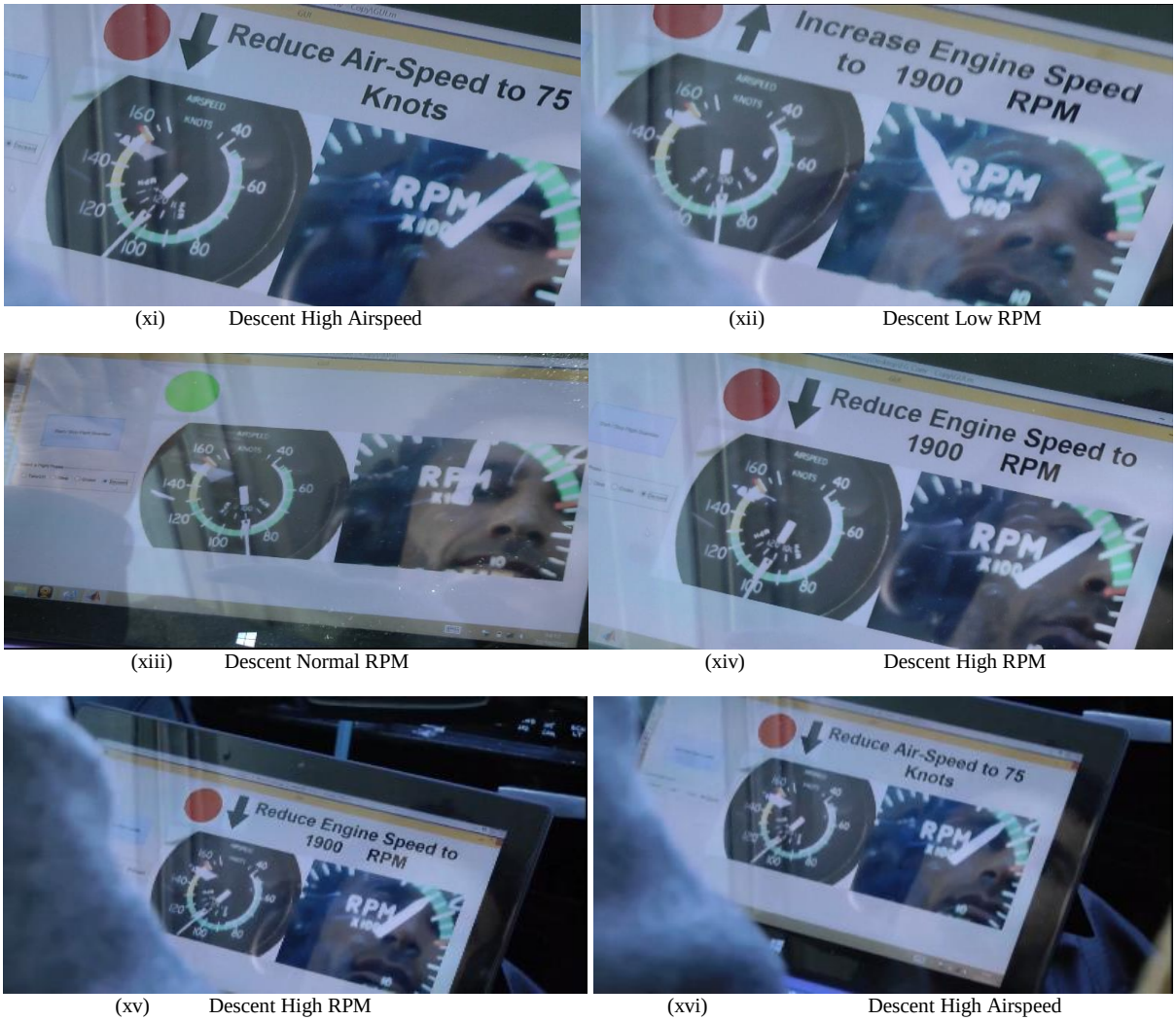


(vii) Cruise Low RPM

(viii) Cruise High RPM

(ix) Descent Low Airspeed

(x) Descent Normal Airspeed & RPM



(xi) Descent High Airspeed (xii) Descent Low RPM

(xiii) Descent Normal RPM (xiv) Descent High RPM

(xv) Descent High RPM (xvi) Descent High Airspeed

Fig. 10 Test Cases for Real Flight Evaluation of FG Prototype in Various Hazrdous Situations in Different Flight Phases

Finally, a summary of the real flight test is produced by the both pilots stated that: "*During the live flight test of Flight Guardian the benefits to pilots of all abilities became clear. It is like having a second pilot issue timely observations and warnings about potentially hazardous situations – it has real potential to reduce pilot error in the cockpit and ultimately save lives*". The pilots report and the aforementioned detailed statistical results prove the proposed system efficiency and significance in terms of flight safety improvements and pilot assistance. These outcomes also serve as an enabler for the extension of this research in terms of its applicability to the variety of aircrafts, fighting jets, commercial aircrafts, and helicopters. Despite of the future open ended research interests, the collection of expert knowledge for each application domain, multiple flight phases, and additional cockpit instruments would be a challenging task that is needed to be considered in relation to FG system applicability to further domains.

IV. Conclusions and Future Directions

The hypothesis underlying the principle objective as motivated for this research project was based on the development a regular monitoring platform for an aircraft that provides robust flight-deck awareness. The proposed system is able to intelligently generate voice activated warnings and visual indications in hazardous situations. Experimental setup was built up comprising data acquisition, automated dial reading using video analysis, expert knowledge representation, belief combination, and decision making. Detailed statistical results were achieved that validated the proposed methodology in terms of flight safety improvements and its significance for the pilot assistance. For the first time, domain knowledge is presented for the cockpit instrument ranges with respect to different flight phases that might be of great interest for the related research. In spite of the robust outputs, there are some limitations associated with the selected methodology. For instance, the ADR task works well only for fixed position camera devices focusing at the center position of the cockpit instrument. Similarly, the propose system uses expert knowledge for two cockpit instruments for specific aircraft (i.e. Cessna C1772SP) that may be extended for the additional cockpit instruments and variety of aircrafts including commercial aircrafts, helicopters, and fighting jets. These outcomes also serve as an enabler for the future research on different aspects. A generic dial reading approach may be developed that would automatically identify the desired cockpit instrument and rotate its image with respect to a fixed reference point (e.g. deep learning approaches). Likewise, the object detection can be further improved in a way that it handles the light reflections and vibrations more efficiently. In addition, the domain knowledge (i.e. dataset for instrument ranges) might be extended to include more cockpit instruments to make the proposed system more generic in terms of automated flight phase detection and applicability to a wider range of civil and military aircrafts.

Acknowledgement

Special thanks to The Great Circle Ltd who were the industrial partner collaborating with UCLan on the development of FG. The FG project was funded (Grant Ref: NATEP NWC2-S006) by the National Aerospace Technology Exploitation Programme (NATEP). The NATEP is managed by the UK's national aerospace body ADS, supported by the UK regional aerospace alliances – FAC, NWAA, MAA & WEAF.

References

[1] Wicken, C., Santamaria. A., and Sebok. A., "A Computational Model of Task Overload Management and Task Switching," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 57, 2013, pp. 763 – 767, DOI: https://doi.org/10.1177/1541931213571167.

[2] National Safety Council., "Understanding the Distracted Brain: Why Driving While Using Hands-Free Cell Phones is Risky Behaviour," 2012. Available at:

http://www.nsc.org/safety_road/Distracted_Driving/Documents/Cognitive%20Distraction%20White%20Paper.pdf (accessed August 20, 2015).

[3] Loukopoulos, L., Dismukes. K., and Barshi. E., "The Multitasking Myth: Handling Complexity in Real-World Operations," *Ashgate*, Surrey, England, 2015, pp. 13-14.

[4] Nikolic M., Orr. J., and Sarter. N., "Why Pilots Miss The Green Box: How Display Context Undermines Attention Capture," *International Journal of Aviation Psychology*. Vol. 14, No. 01, 2004, pp. 39 -52, DOI: https://doi.org/10.1207/s15327108ijap1401_3.

[5] Miller, G. A., "The Magical Number Seven, Plus or Minus Two Some Limits on Our Capacity for Processing Information," *Psychological Review by the American Psychological Association*. Vol. 63, No. 02, 1956, pp. 81-97, DOI: http://dx.doi.org/10.1037/h0043158.

[6] Sumwalt, R., Cross, D., and Lessard. D., "Examining How Breakdowns in Pilot Monitoring of the Aircraft Flight Path," *International Journal of Aviation, Aeronautics, and Aerospace, ERAU Scholarly Commons*. Vol. 02, No. 03, 2015, pp. 01-25, DOI: https://doi.org/10.15394/ijaaa.2015.1063.

[7] National Transportation Safety Board., "Safety Study: A Review Of Flight Crew-Involved, Major Accidents Of U.S. Air Carriers, 1978 Through 1990," NTSB Report No. NTSB/SS-94/01, Washington DC, 1994, Available at: http://libraryonline.erau.edu/online-full-text/ntsb/safety- (accessed June 03, 2015).

[8] Iwadare, K., Oyama, T., "Statistical Data Analyses on Aircraft Accidents in Japan: Occurrences, Causes and Countermeasures," *American Journal of Operations Research*, Vol. 5, 2015, pp. 222-245. DOI: http://dx.doi.org/10.4236/ajor.2015.53018.

[9] Boone, P., "Pilot Errors from a Pilot Perspective," June, 2016, Available at: http://www.b737mrg.net/downloads/fly_the_dog.pdf (accessed Aug 01, 2016).

[10] Oster, C. V., Strong, J. S., and Zorn, C. K., "Analysing aviation safety: Problems, challenges, opportunities," *Research in Transportation Economics, Elsevier*, Vol. 43, 2013, pp. 148-164. DOI: http://dx.doi.org/10.1016/j.retrec.2012.12.001.

[11] Grady, M., "NTSB Issues Glass Cockpit Safety Recommendations," 2010, Available at: http://www.avweb.com/avwebflash/news/NTSBIssuesGlassCockpitSafetyRecommendations_202250-1.html (accessed September 13, 2015).

[12] Sorrill, S. L., "What Safety Initiatives have Been Introduced in Response To," OpenSIUC, Southern Illinois University Carbondale, No. 316, 2012, DOI: 6. http://opensiuc.lib.siu.edu/gs_rp/316.

[13] National Transportation and Safety Board., "Introduction of Glass Cockpit Avionics into Light Aircraft," Washington DC: NTSB, 2012, Available at: http://www.ntsb.gov/safety/safetystudies/SS1001.html.

[14] Kuchar, K. J., "White Paper on Multiple Independent Alerting Systems," Written at the request of the NASA Ames Research Centre and the Cargo Airline Association, 1998, Available at: http://www.mit.edu/~jkkuchar/white/whitepaper.html.

[15] Mulder, M., Pedroso. P., Mulder. M., and Passen. M.M. V., "Integrating Aircraft Warning Systems. Delft University of Technology, Faculty of Aerospace Engineering," 2000, pp. 179- 183. Online Available at: https://scholar.google.com/citations?user=1nvtktYAAAAJ&hl=nl.

[16] Abeloos, A. L. M., Mulder, M., and Paassen, M. M. V., "The Applicability of an Adaptive Human-Machine Interface in the Cockpit," *19th European Annual Conference on Human Decision Making and Manual Control*, Italy, 2000, pp. 193-198.

[17] Thatcher, S., "An Artificial Intelligent Paradigm for Systems Safety in the Cockpit," AERO Lab, University of South Australia, 2008, Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.102.7992&rep=rep1&type=pdf.

[18] Marie, S. L., "An Investigation of the Effectiveness of a Strobe Light as an Imminent Rear Warning Signal," Master's Thesis, Digital Library and Archive, VirginiaTech University, 2000, URN: etd-12012000-084911. Available at: http://scholar.lib.vt.edu/theses/available/etd-12012000-084911/unrestricted/Literature.pdf (accessed March 12, 2015).

[19] Feng, H., and Zhao. J., "Application Research of Computer Vision in the Auto-Calibration of Dial Gauges," *International Conference on Computer Science and Software Engineering*, Vol. 02, 2008, pp. 845-848, DOI: 10.1109/CSSE.2008.350.

[20] Vega, R. O., Ante. G. S., Morales. L. E. F., and Sossa. H., "Automatic Reading of Electro-mechanical Utility Meters," *12th Mexican International Conference on Artificial Intelligence*, 2013, pp. 164-170, DOI: 10.1109/MICAI.2013.28.

[21] Gellaboina, M. K., Swaminathan. G., and Venkoparao. V., "Analogue Dial Gauge Reader for Handheld Devices," *IEEE 8th Conference on Industrial Electronics and Applications (ICIEA*), Melbourne, VIC, 2013, DOI: 10.1109/ICIEA.2013.6566539.

[22] Khan, W., Ansell. D., Kuru. K., and Amina. M., "Automated Aircraft Instrument Reading Using Real Time Video Analysis," *IEEE Conference on Intelligent Systems*, Sofia, Bulgaria, 6th Sep, 2016, DOI: 10.1109/IS.2016.7737454.

[23] Aviation, A., "Aircraft Information Booklet, Cessna 172S," 2013, Available at: http://www.airborne-aviation.com.au/resources/aircraft/booklet_cpq.pdf (accessed December 02, 2014).

[24] Skyhawk., "Pilot's Operating Handbook, Cessna 172N," 1978, Available at: http://www.skywarriorinc.com/downloads/POH%20BOOKS/172N%20POH.pdf. Accessed: 17th Feb, 2015.

[25] Mandal, S. N., Choudhury. J. P., and Chaudhuri. S. R., "In Search of Suitable Fuzzy Membership Function in Prediction of Time Series Data," *International Journal of Computer Science Issues*, Vol. 09, No, 03, 2012, pp. 293-302.

[26] Foley, B. G., "A Dempster-Shafer Method for Multi-Sensor Fusion," MSc Thesis, Department of Mathematics and Statistics, Air Force Institute of Technology, 2012, Available at: http://www.dtic.mil/dtic/tr/fulltext/u2/a557749.pdf (accessed April 07, 2015).

[27] Chen, Q., Whitbrook. A., Aickelin. U., and Roadknight. C., "Data Classification Using the Dempster-Shafer Method," *Journal of Experimental & Theoretical Artificial Intelligence, Taylor & Francis Online*, Vol. 4, No. 26, 2014, pp. 493-517. DOI: 10.1080/0952813X.2014.886301.

[28] Dismukes, K., Young. G., and Sumwalt. R., "Cockpit Interruptions and Distractions," *ASRS Directline*, Vol. 10, 1998, pp. 4-9, Available at: https://asrs.arc.nasa.gov/docs/dl/DL10.pdf.

Biography

Dr. Wasiq Khan **received his BSc in Math & Physics and MSc in Computer Science from COMSATS IIT, Pakistan. At Bradford University (UK), he completed MSc in Artificial Intelligence and Ph.D in Speech Processing research. His current research interests include Applied Computational Intelligence in diverse areas. Currently, Khan is working as a Research Project Manager in Intelligent Systems Research Group at Manchester Metropolitan University, UK. His recent publications cover the pattern matching, time warped speech similarity measurement, and spoken term detection**.

Dr. Darren Ansell **is the engineering lead for Space and Aerospace in the School of Computing, Engineering and Physical Sciences, University of Central Lancashire. He specializes in applied autonomous and intelligent systems research. He previously worked in industry at BAE Systems in research management and research and development roles, specializing in Mission Systems and Autonomy. Darren is research active within the area of digital engineering and is a member of Applied Digital Signal and Image Processing Research Centre (ADSIP).**

Dr. Kaya Kuru **is a Research Associate in the School of Engineering, University of Central Lancashire (UCLan). Before UCLan, he has studied and worked in several top ranked universities for 15 years such as Middle East Technical University (METU), Southampton University and Gulhane Military Medical Academy (GATA). He is interested in developing autonomous intelligent systems and decision support systems based on machine learning and image processing algorithms. He is keen to develop these kinds of applications on Android systems using wireless technologies. Kaya contributes to several scholarly journals as a reviewer.**

Dr. Muhammad Bilal **is an Assistant Researcher at Centre for Environmental Implications of Nanotechnology, University of California, LA. His research focuses data-driven solutions for the environmental and health impact assessment of engineered nanomaterials (ENMs) using advanced machine learning/data mining and simulation techniques. Bilal also has extensive experience in designing, implementing, and maintaining high-performance computational cluster using advanced techniques (such as Rocks cluster, Sun Grid Engine, Hadoop) as client-server based applications.**