Delft University of Technology

Reconstruction of Pilot Behaviour from Cockpit Image Recorder

Tsuda, Hiroka; Stroosma, Olaf; Mulder, Max

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Reconstruction of Pilot Behaviour
from Cockpit Image Recorder

Hiroka TSUDA*

Japan Aerospace Exploration Agency, Mitaka, Tokyo, 181-0015, Japan

Olaf Stroosma† and Max Mulder‡

Delft University of Technology, Delft, 2629HS, The Netherlands

A method to automatically identify pilot actions from cockpit camera footage is reported in this paper. Although they have long been considered for the enhancement of flight safety, cockpit image recorders have not yet been standard equipment in aircraft cockpits. The rules on Flight Data Recorders have been changed, however, to include a cockpit image recorder as one of the safety devices, and it is recommended to be installed in small aircraft as a substitute for a Flight Data Recorder. With cockpit images becoming available, it would surely be useful for accident analysis as well as for daily flight analysis. Especially for the latter purpose, pilot behavior should be automatically analyzed and classified into specific actions, or procedures. The authors conducted a study to assess the feasibility of automatic detection of pilot actions in the cockpit by a machine learning process. Results show that even with a small amount of training data, the resulting algorithm could identify some typical actions, such as manipulation of the switches on the glare shield, with 80% accuracy. Even in cases with a button and a switch positioned very close to each other, the actions 'pushing the switch' and 'pushing the button' could be distinguished by the algorithm. The action estimation accuracy improves up to 90% when using the training data focused on the pilot's body parts, rather than the data focused on the whole body.

I. Introduction

Traditional safety activity in aviation has been conducted in a "reactive" manner; once an accident occurred, the goal has been to analyze it and develop countermeasures to prevent re-occurrence of the same type of accident. As both required and achieved safety levels improved, reactive safety activity became less effective because the number of accidents simply decreased. Then, proactive safety activity, to extract precursors of accidents from data of daily operation, has been focused on. FDM (Flight Data Monitoring) or FOQA (Flight Operations Quality Assurance) are some of those proactive activities, to find out off-nominal flight parameters from daily flight data. So far, only numerical flight data, such as air data, inertial data, and status of systems have been used for FDM, while audio and video recordings in the cockpit were not used, for several reasons. Although cockpit video is considered to be very useful, not only for FDM but also for accident investigation, the installation of cockpit cameras to record pilot behavior has been considered as problematic.

Recently, this situation has begun to change. The installation of flight recording devices, sometimes substituted by commercial action cameras, is recommended by safety authorities [1–3]. Although installation in transport category aircraft has not been done so far, using cockpit image in FDM turned out to be a realistic approach.

A natural use of cockpit images is to extract a crew member's physical action, that is not recorded in flight data nor in audio data. "OPSAMS" (Operational Procedure Safety Analysis and Monitoring System) is a flight data analysis system developed by JAXA [4–6], to reconstruct pilot behavior, only by using flight data and a human model. Although OPSAMS shows good reconstruction accuracy, it could not tell whether an erroneous operation was conducted by one crew member or another, because the cockpit image is not available. For accident investigation purposes, it might be easy to tell which crew member does what action from the cockpit image at a glance, however, it is not realistic for FDM, where automation is necessary to analyze large amounts of video data.

*Associate Senior Researcher, Flight Research Unit., tsuda.hiroka@jaxa.jp.

†Researcher, section Control and Simulation,Faculty of Aerospace Engineering, P.O Box 5058, 2600GB Delft, The Netherlands; O.Stroosma@tudelft.nl, Senior Member AIAA.

‡Professor, section Control and Simulation, Faculty of Aerospace Engineering, P.O Box 5058,2600GB Delft, The Netherlands; M.Mulder@tudelft.nl, Associate Fellow AIAA.

This paper discusses the feasibility of identifying types of pilot action from cockpit images and video, using various types of machine-learning processes. The basic concept, method of experiment and the results are reported.

II. Concept Design

The goal of the proposed technology is to identify whether a procedure is properly conducted, or not. The procedure is comprised of some actions by each pilot, Pilot Flying (PF) and Pilot Monitoring (PM). Fig. 1 illustrates an example of the relationship between procedure and actions. As described, some actions can be detected in voice recording, and the result of some actions is recorded as system status in flight parameter recording. Obvious actions such as stretching an arm to a switch can be detected from the cockpit image; on the other hand, actions with a more subtle motion of (parts of) the body, such as listening, seeing and other mental actions are hardly observed from the cockpit image. At this stage of the research, we defined the objective of the system as to identify the obvious "visible actions" from the cockpit image, while the reconstruction of the procedure from the visible and non-visible actions will be conducted by other systems, e.g., [4–6].

The functional structure of the proposed system is described in Fig. 2. First, body motion of the crew in the cockpit is captured by a camera and recorded. As discussed later on, the captured image data could be either of 2D or 3D nature. Although real-time analysis is in future scope, post-flight analysis is considered as appropriate for the initial stage of the study. The expected output of the system are the visible actions, as already discussed.

PF	PM
Order Gear Down	Response
	Down Gear Lever
Acknowledge Gear is Down	Check Gear is Down

PF Action	PM Action	Record		
		Image	Voice	System
Order Gear Down				
	Hear PF's order			
	Respond to order		✓	
	Move Head	✓		
	See Gear Lever			
	Stretch Arm	✓		
	Down Gear Lever	✓		✓
Move Head		✓		
See Gear Lever status	Hand on Knee	✓		

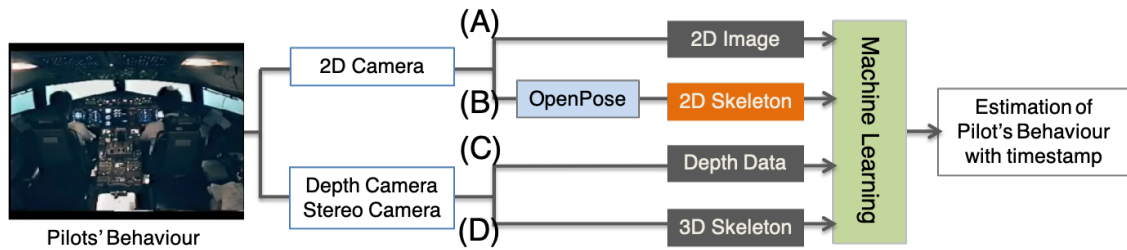**Fig. 1   Procedure and actions.**

Fig. 2   Functional structure of the proposed system.

Taking into account emerging technologies such as marker-less motion capture, several methods are considered to be candidates to extract Pilot Behaviour from the camera. These candidates are categorized in Table 1. There are two types of camera, *conventional (2D) cameras* and *depth (3D) cameras*. Depth cameras produce the image data with depth information by using stereo cameras, depth sensors like LiDAR (Light Detection and Ranging), or other technologies. For both 2D and 3D cameras, there is an option to extract a human body skeleton from the 2D or depth image or not. When implementation in the real cockpit would be considered, using a 2D camera seems to be a more practical solution than using a depth camera. Once a 3D skeleton would be extracted, interaction with hands or fingers with control surfaces in the flight deck can be identified in a deterministic way, if locations of those control surfaces in

Table 1   Variation of the system structure.

Type of Camera	Skeleton extraction		Action Analysis	Calculation Cost
2D camera	None	2D Image	Stochastic	High
	e.g. "OpenPose"	2D Skeleton	Stochastic	Low
Depth Camera,	None	Depth Data	Stochastic	High
Stereo Camera	e.g. "Kinect"	3D Skeleton	Stochastic	Low
			Deterministic	Low

the cockpit are exactly known. In this approach, however, the greatest difficulties lie on how to define the criteria of each human action. Taking those factors and available resources into account, the authors selected to use a 2D camera and extract the skeleton by using OpenPose [7] before the process of behavioral analysis.

The function of the core system is then defined as shown in Fig. 2 (B). The input of the system is the 2D position of joints of the subject's body, as the output of OpenPose, and the outputs are the subject's actions.

III. Procedure

A. Recording Pilot's Behaviours.

Some sets of the movie are taken in the cockpits of the SIMONA Research Simulator (SRS)[8], and the JAXA Flight Simulator (FSCAT-A)[9]. One subject sat on the left-hand seat as a pilot, and conducted the following actions:

1) Push a switch on FMS/ CDU,
2) Push a switch on Mode Control Panel,
3) Rotary a Knob on Mode Control Panel,
4) Push a switch on Overhead Panel,
5) Grip the control Wheel,
6) Put their hands on their knee, and
7) Look outside the window.

In the SRS, six types of above actions except for No. 3 were recorded. In the FSCAT-A, No. 2, 3 and 6 actions were recorded. In addition, pilots had to put their hands on their lap before starting those actions and after finishing them.

Although some cameras were installed and used in the cockpit, only one camera which was installed to take from the right rear of the subject was utilized in the analysis. The resolution of the camera was 1920x1080 pixels and its frequency was 30Hz.

The simulator systems recorded all of the control manipulations and produced logs with timestamps. The same timestamp was presented on one of cockpit displays and was used to manually add events of pilot actions which were not recorded by the simulator system, such as looking out the window.

B. Generating the data-sets for Machine-Learning.

As described in Section II, the movies taken in the SRS's and FSCAT-A cockpit were once analyzed by OpenPose to extract a series of data-sets, the 2D positions of joints of the body. Fig.3 shows a snapshot of the resultant image of OpenPose. Then a label of pilot action was attached to the data-set by the following procedure.

First, by synchronizing the beginning time and the ending time of both movie recording and the simulators' logfile, the duration was divided by the number of the data-set and the timestamp was attached to each data-set. Second, a label was attached to the data-set by comparing its timestamp with the time of simulators' logfiles automatically. In the FSCAT-A cockpit, there is a push switch on the pilot's lap and he pushed it on every time before he started a new action and after he finished the action. The time when he pushed it was recorded into FSCAT-A's logfiles. If the actions were not recorded by the simulator system, a label was attached to the data-set manually.

Table 2 contains a list of labels. The contents of the data-set are described in Eq. 1.

3

Table 2   Type of Actions and Labels.

Simulator	No.	Type of Action	Label
SRS	s1	None.	Knee
	s2	Right and left hands are Moving.	Moving
	s3	Right and left hands grip the control wheel.	WHL
	s4	Right hand is on CDU. Left hand grips the control wheel.	R=CDU+L=WHL
	s5	Right hand is on CDU. Left hand in on the lap.	R=CDU+L=Knee
	s6	Right hand is on MCP. Left hand is on the lap.	R=MCP+L=Knee
	s7	Right hand is on OHP. Left hand is on the lap.	R=OHP+L=Knee
	s8	Right hand is moving. Left hand grips the control wheel.	R=Moving+L=WHL
	s9	Right hand is moving. Left hand is on the lap.	R=Moving+L=Knee
	s10	Right hand is on the lap. Left hand grips the control wheel.	R=Knee+L=WHL
	s11	Right hand is on the lap. Left hand is moving.	R=Knee+L=Moving
	s12	Pilot sees the forward.	F=FWD
	s13	Pilot looks down.	F=Down
	s14	Pilot looks up.	F=Up
	s15	Pilot looks back.	F=Back
	s16	Pilot looks to the center direction.	F=CTR
	s17	Pilot looks at CDU.	F=CDU
	s18	Pilot looks at MCP.	F=MCP
	s19	Pilot looks at OHP.	F=OHP
FSCAT-A	f1	None.	Knee
	f2	Right hand is moving.	Moving
	f3	Right hand is pushing the button.	Button
	f4	Right hand is pushing the knob.	Knob
	f5	Right hand is rotating the knob.	Rotation

$$data-set = \begin{cases} time, label, \\ Body-x1, Body-y1, \cdots, Body-x18, Body-y18, \\ HandL-x1, HandL-y1, \cdots, HandL-x21, HandL-y21, \\ HandR-x1, HandR-y1, \cdots, HandR-x21, HandR-y21 \end{cases} \tag{1}$$

C. Machine-Learning process.

As a preliminary setting, a total of three input files were prepared for input to Machine-Learning. An input file is a series of data-sets that contain one or more cases of each of the seven actions described in section A of this chapter. The number of data frames means the number of data-sets included in the input file and each input file is composed of 2769, 3133 or 4140 data-sets. The contents and the number of data frames of each input files are shown in Table 3.

An input file was divided into training data and test data in a ratio of 8:2. Scikit-learn [10] was utilized as a Machine-Learning algorithm. A variety of algorithms in Scikit-learn, as listed in Table 4 was tested for accuracy.

IV. Experiments and Results

We conducted three types of experiments.

The first is an attempt to find the algorithm with the highest accuracy rate through Machine-Learning using several kinds of algorithms.

4
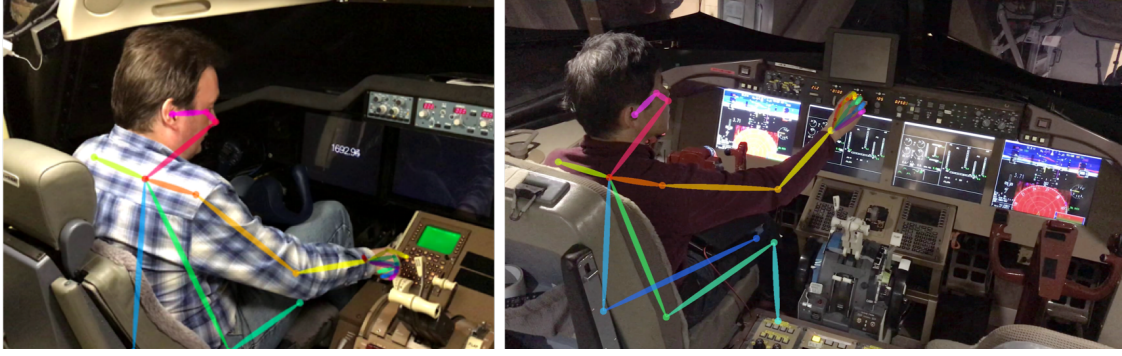
Fig. 3 OpenPose extraction of pilot's joints of body and hand.

Table 3 The contents and the number of data frames of each input files.

Input file	Contents	The number of data frames.
1	Action No.s1-s19. The pilot acted from s1 to s19 continuously. Then, the pilot repeated this series of actions four times.	3133
2	Action No.f1-f5. The pilot acted from f1 to f5 continuously. Then, the pilot repeated this series of actions ten times.	2769
3	Action No.f1-f5. The pilot performed each action ten times repeatedly. The pilot acted each action ten times repeatedly. For example, when the pilot finished repeating f2, he started the next action f3.	4140

The second is to verify whether the pilot's action can be classified using this new approach combining OpenPose and Machine-Learning.

Third and final, any differences in the accuracy of Machine-Learning when using a different body action categorization policy were investigated.

A. Comparison of Machine-Learning algorithms.

Table 4 shows the results of the trial of Machine-Learning with several types of algorithms. For this series of trials the data from SIMONA (file 1 from Table 3) were used. Among the variety of algorithms, LogisticRegression shows the best results of 80% accuracy; this algorithm was chosen for analyzing the remainder of the data.

Table 4 Results of each Machine-Learning algorithm.

Kind of Machine-Learning Algorithm	Average Accuracy %
StochasticGradientDescent(SGD)	54.72
SupportVectorMachine Classification(SVC), LinearSVC	80.09
LogisticRegression	83.77
RamdomForestRegressor	58.32
NeuralNetwork	68.63

### B. Validation of the applicability.

Table 5 shows the average accuracy of estimation through Machine-Learning with the LogisticRegression algorithm. Here, the input files were No.1, 2 and 3. All of the results show above 75%.

Fig.4 shows the plot of the result of the input file No.1 and No.2. The x-axis shows the number of the data frame,

5

that is, time. The y-axis shows the type of pilot's actions. Blue lines show the actual actions; amber lines indicate the estimated actions by Machine-Learning.

Table 5   The average accuracy of estimation by LogisticRegression algorithm.

Input file	Average Accuracy %
1	83.25
2	76.80
3	75.20

(A) The result of input file No.1.

(B) The result of input file No.2.
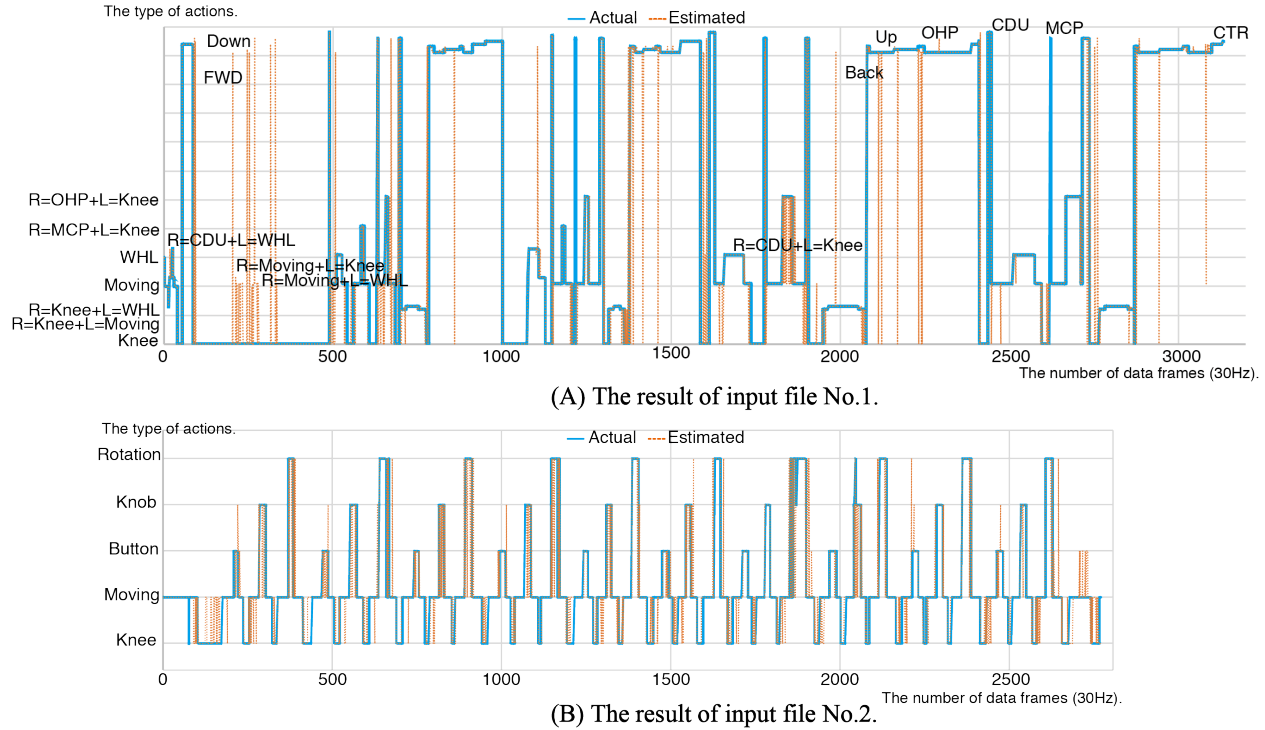
Fig. 4   The timeline plots of the results by Machine-Learning using input file No.1 and No.2.

In some cases, the Machine-Learning estimate was wrong, and we counted where failures occurred. Failures in estimation are more often observed when either of the actual action or the estimated action is 'Knee' or 'Moving' than when both of the actual action and the estimated action are 'Button', 'Knob' or 'Rotation.'

Fig. 5 shows the percentage of the errors in input file No.2 and No.3. As shown in this figure, the failure occurs rarely when both of the actual and estimated actions are the cockpit devices. For example, in the case of input file No.2, the confusion of 'Button' and 'Knob' happened four times and the confusion of 'Knob' and 'Rotation' happened three times. These results suggest that this procedure combining OpenPose and Machine-Learning can distinguish the pilot's one action and the other action.

One of the reasons for success that this procedure combing OpenPose and Machine-Learning can distinguish a pilot's action from other actions, is considered the resolution of the video. The angle-of-view of the camera also seems to be important. In the input file No.2 and No.3, the button and the knob are near each other in the cockpit of FSCAT-A as shown in Fis.6. It's thought that the resolution and the angle-of-view are enough.

C. The effect of action labeling with a different policy.

The label, indicating the pilots's action, can take two different forms. One is to describe *"Overall"* action of the body, such as "Looking forward and both hands on the knee", "Looking at FMS-CDU and manipulating CDU by right

6

Left: Actual, Right: Estimated action.
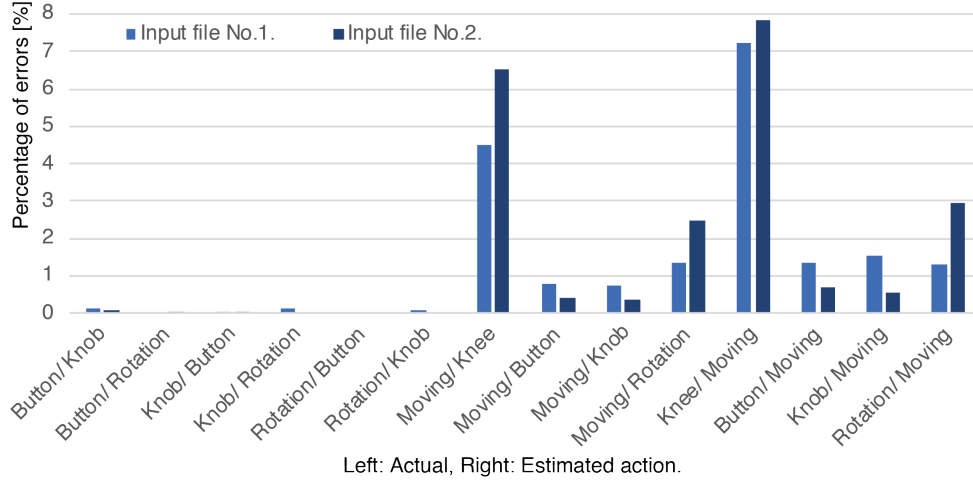
Fig. 5   The percentage % of the failures in input file No.2 and No.3.
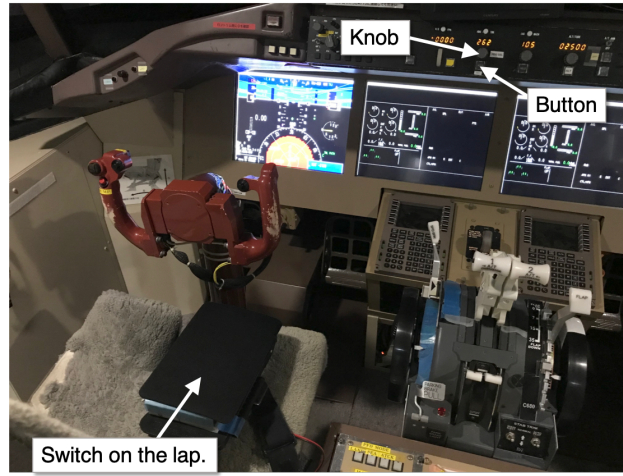
Fig. 6   Button and Knob on the MCP and Switch on the lap in the cockpit of FSCAT-A.

hand", and, the other is to focus on single *"Element"* of body, such as "right hand on knee", "right hand on CDU", "right hand moving toward control column". Both types of labelling as shown in Table 6 were compared.

Table 7 shows the results for "Element" labelling focusing on the subject's right hand in the case of input file No.1. Slightly better results were obtained for "Element" labelling than "Overall". Here:

$n = 1$ The hand or head starts moving from the original position to the target device, or the target direction;

$n = 2$ The hand or head is on the target device, or moving toward the target direction;

$n = 3$ The hand or head starts coming back from the target device to the original position.

V. Conclusion

The accuracies of using various Machine-Learning algorithms were compared, to investigate what would be the most suitable algorithm to distinguish the pilots' actions. We conclude that the LogisticRegression algorithm performs best, and will be used in future work.

The chosen algorithm could distinguish some typical pilot actions even with a relatively small number of data-sets, and also when the target buttons were positioned very close to each other.

We also found that the accuracy of the estimates improves, up to 90%, in case of using the data-sets with labels focusing on the pilot's body parts rather than on the whole body.

7

Table 6   List of two types of labels.

Overall	Element		
	Right Hand (RH)	Left Hand (LH)	Face
Knee	R=Knee	L=Knee	F=FWD
WHL	R=WHLn*	L=WHLn	F=FWD
R=CDU+L=WHL	R=CDUn	L=WHLn	F=CDU
R=MCP+L=Knee	R=MCPn	L=Knee	F=MCP
R=OHP+L=Knee	R=OHPn	L=Knee	F=OHP
R=Knee+L=WHL	R=Knee	L= WHLn	F=WHL
F=MCP	R=Knee	L=Knee	F=MCP
F=Up	R=Knee	L=Knee	F=Up
F=Down	R=Knee	L=Knee	F=Down
F=CTR	R=Knee	L=Knee	F=CTR

Table 7   Comparison with the average of accuracy % of "Overall" and that of "Element" using LogisticRegression algorithm.

Overall	Element		
	Right Hand (RH)	Left Hand (LH)	Head
83.78	93.10	94.39	91.72

Acknowledgments

References

[1] "Helicopter Air Ambulance Operations," *Advisory Circular 135-14B*, Federal Aviation Administration, U.S.Department of Transportation, 2015.

[2] "State Helicopter Flight Recorder Standards," *Specification No.23 Issue:2*, United Kingdom Civil Aviation Authority, 2018.

[3] "In-flight recording for light aircraft," *Notice of Proposed Amendment 2017-03*, RMT.0271 (MDM.073(a)) and RMT.0272 (MDM.073(b)), European Aviation Safety Agency, 2017.

[4] Muraoka, K., and Tsuda, H., "Flight Crew Task Reconstruction for Flight Data Analysis," *Proceedings of the 50th Annual Meeting of Human Factors and Ergonomic Society*, 2006, pp. 1194–1198.

[5] Muraoka, K., and Tsuda, H., "Flight crew task reconstruction for Flight Data Analysis program," *JAXA Research and Development Report*, JAXA-RR-07-039, 2008, pp. 175–212.

[6] Funabiki, K., Noda, F., Kato, M., Tsuda, H., and Muraoka, K., "OPSAMS: Flight Crew Behavior Reconstruction from Flight Data," *Proceedings of the 49th JSASS Annual Meeting*, JSASS-2018-1007, 2018. (in Japanese).

[7] Zhe, C., Tomas, S., Shih-En, W., and Yaser, S., "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," *CVPR 2017*, 2017.

[8] Stroosma, O., Van Paassen, M. M., and Mulder, M., "Using The SIMONA Research Simulator for Human-Machine Interaction Research," *AIAA Modeling and Simulation Technologies Conference*, AIAA2003-5525, 2003.

[9] Wakairo, K., Noda, F., Muraoka, K., Iijima, T., Funabiki, K., and Nojima, T., "Development of flight simulator for research and development," *JAXA Research and Development Memorandum*, JAXA-RM-04-015, 2005. (in Japanese).

[10] Pedregosa, F., Varoquaux, G., Michel, V., Thirion, B., Grisel, O., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, Vol. 12, 2011, pp. 2825–2830.