MULTI-CHANNEL HUMAN-MACHINE INTERACTIVE SYSTEM FOR MULTIPLE UAVS' CONTROL

Peng-cheng Wen, Ya-di Zhang, Ya-hui Li
AVIC Xi'an Aeronautics Computing Technique Research Institute, P. R. China

Keywords: *Human-machine interaction, Multi-channel control, UAV*

Abstract

The UAVs in formation flight executing a same task coordinately have been more and more applied in the military and civil domains. In order to maximize the effectiveness of the multiple UAVs, meanwhile to reduce the workload of the operator, a new multi-channel human-machine interactive system is designed, in which the fusion of three different control signals (touch, gesture & voice) is used to carry out the control and the clear & concise visual feedback is used to monitor the control effect. Simulation experimental results demonstrated the feasibility and superiority in application of the designed system, where it had higher control efficiency and the operator could control more than four UAVs at one time to avoid obstacles with a higher success rate.

1 Introduction

Since the multiple unmanned aerial vehicles (UAVs) in formation flight executing a same task coordinately have the characteristics of multi-angle, omnibearing and large-depth view compared with a single UAV, they have been more and more applied in the military and civil domains such as multi-purpose ISR, close air support, wide-area seamless-coverage search, geodetic surveying, meterogical observation, environment monitoring, resource exploration, earthquake / fire rescue and city security [1, 2].

In a traditional control system, UAVs are frequently operated by multiple operators working in parallel. Different operations like navigation, flight control and communication & load management usually need different operators to take guidance at different levels. For example, the UAV of Predator and the UAV of Shadow need more than three operators to carry out the operations for a single platform. The UAV of Global Hawk and the UAV of Reaper also need operators more than the number of platforms. However, with the increasing UAVs' number and the increasing tasks' parallelism and complexity, this control mechanism would not be suitable. Because in this situation, the workload of the operator would increase and the coordination between the operators would be difficult to achieve optimal. Then the final task execution performance would be influenced. To avoid the above situation meanwhile to maximize the effectiveness of the multiple UAVs as much as possible, a new control system especially for multiple UAVs should be designed, in which the system should have the capability to operate a large number of UAVs with a small number of operators, ideally to operate a large number of UAVs with only one operator.

All the developed countries in the world, especially the United States, have not ceased to explore multiple UAVs' control [3].

The project of Mixed Initiative Control of Automa-teams (MICA) [4, 5], as a research program of DARPA, aims to develop a battlefield management system that can compensate for the lack of operator's ability through the study of human-machine efficiency, and achieve operating multiple UAVs with fewer operators. Compared with the current proportion (2:1 or 3:1) of the operators to the UAV, MICA plans to operate the UAVs, three times more than the number of operators in the first stage, and thirty times more in the second stage.

The project of Collaborative Operations in Denied Environment (CODE) [6], as another research program of DARPA, also advocates that a task commander can maintain good situation awareness and guide multiple UAVs through the technologies of open architecture, formation autonomy and human-machine interaction.

Based on the Software, Hardware, Environment and Liveware (SHEL) model describing human's ability and its limitation, human is the most valuable and flexible part of system. The design and the use of system must be centered on human. Mapping to the control interface, it should reduce the work load of the operator and help the operator execute attention allocation and decision making by some simple and practical technical means, so as to realize the task management and collaboration of multiple UAVs, in which the necessary intervention can be implemented and the extra burden cannot be brought.

With the growth of the intelligence level of UAVs, the UAVs' control mode changes from the low-level action control to the high-level supervisory control. According to the unmanned aircraft system roadmap (2005-2030) [7], there is a very urgent need to improve the supervisory control ability of the multiple UAVs, and the operator-UAVs interaction is a first key element. The Naval Studies Board and the DoD (Department of Defense, USA) hold the same views [8, 9]. They strongly consider that the future unmanned platforms' control system is a joint human-machine cognitive and decision-making system. Thus, when we design the multiple UAVs' control system, it is important to study the fusion relationship between the operator and the UAVs. In other words, a natural and efficient human-machine (operator-UAVs) interactive system should be paid much attention to.

The rest of the paper is organized as follows. Section 2 describes the framework of the multiple UAVs' control system. Section 3 discusses the consideration of the multi-channel. Section 4 expresses the detailed implementation and gives the technical models. Section 5 demonstrates the feasibility and superiority in application of the designed system through simulation experiments. Section 6 addresses the concluding remarks.

2 Framework of the multiple UAVs' control system

The designed framework of the multiple UAVs' control system, as shown in Fig. 1, consists of four parts: the operator, the UAVs, the human-machine interface and the function allocation module. The operator, as an intelligence element at the top level, is responsible for the mission assignment and the decision adjustment. It has the highest priority. The UAVs, as controlled objects with the capability to perform some autonomous tasks, are responsible for the situation awareness and the action execution. Their autonomy is constrained by the operator. The multi-channel human-machine interactive system, as a mutual interface connecting the operator and the UAVs, is responsible for the instruction analysis and the state feedback. The function allocation module, as a function regulating controller with a variable autonomous level, is responsible for the balance between the operator and the UAVs.

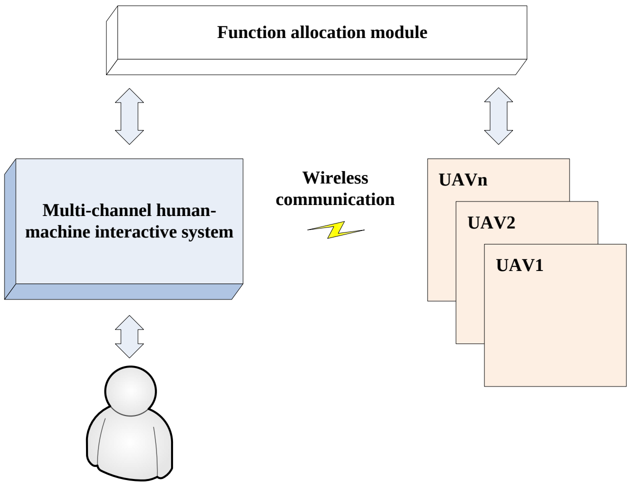

Fig. 1. Framework of the multiple UAVs' control system

This design mechanism has a lot of advantages.

Firstly, it is more suitable for the variety of tasks. If the UAVs perform routine tasks under predefined scenarios, they will have a high level of autonomy and their operator will give a rare intervention. If the UAVs perform safety critical tasks or perform a task under an unknown scenario, they will have a low level of autonomy and their operator will handle most operations.

Secondly, it is more suitable for the change of the task environment. In a complex electromagnetic condition like in a strong electromagnetic interference condition or in a wireless rejection condition, the UAVs may lose the connection with the ground control station. They need to improve the level of autonomy to ensure their tasks' success on the premise of their safety. In an opposite condition, i.e. in a condition of smooth communication, the operator may understand the task environment better than the UAVs by receiving multi external data. At this point, the operator needs to take back the control of UAVs.

Thirdly, it is more suitable for the different cognitive ability and working state of different operators. By monitoring the operator's physiological index and analyzing his physical status, meanwhile on the basis of the operator's operating reaction condition under a high pressure and time sensitive environment, it would make difference for different operators to set up the UAVs' autonomous level.

3 Consideration of the multi-channel

In view of the special application background of aviation, safety is the key problem that the UAVs must consider at all times. In order to ensure UAVs flight safety, the UAVs' control system must be stable. That is to say:

The human-machine interactive execution process and result must be determined, in which the meaning of the task instruction issued by the interactive system must be unique without any misleading. If the current task instruction cannot be identified exactly, it should not be used. Meanwhile, this information should be sent to the operator and other related systems, to make them take a safer way instead.

The analysis time of the task instruction must be fixed without any overtime. Beyond the specified time range, even if the task instruction is analyzed accurately, it will lose its meaning and lead to the failure of the task.

The system's working property must be consistent & reliable, no matter the operator is full of energy and concentrated or the operator's workload is huge and his attention is distracted. Even when the system works for a long time, it should maintain the same.

Compared to the single-channel human-machine interactive system which may have a lower interactive semantic recognition rate, a longer control reaction time and an unstable working state, the multi-channel human-machine interactive system is more close to practical. It could expand the information exchange bandwidth effectively to promote the efficiency of control and could express the different cognitive potential between human and machine to reduce the human's cognitive load. By fusing control signals from different channels, not only the number of instructions is increased but also the relatively redundant mutual information is used to cover the weak points of the single interactive mode.

Through a psychological research [10], the human is more willing to believe that the eyes (sense of seeing), the hands (sense of touching) and the mouth (sense of hearing) are the most reliable sensory and behavioral channels in the word. And about 83% of the information got by the human comes from seeing, 1.15% comes from touching, 11% comes from hearing, and the rest of the 4.85% comes from other channels [11, 12]. Therefore, in this paper, the technology of the fusion of three different control signals including the touch screen control, the gesture control and the voice control is used to carry out the control and the technology of the high-resolution integrated display is used to monitor the control effect.

4 Implementation

The technical reference model of the designed multi-channel human-machine interactive system, as shown in Fig. 2, adopts the dual closed loop format. At the beginning, the pressing positions and the sliding tracks of human fingers are recorded by the touching screen, the 3D positions and the gestures of human hands are extracted by the optical camera, and the human speeches are acquired by the microphone. Then these signals are fused and mapped into the preliminary task instructions in light of the predefined multi-channel interactive protocol. After finishing the consistency verification based on the context perception, the final confirmed task instructions can be obtained. Further, the execution instructions accepted by UAVs can be generated according to the operator-UAVs cooperation strategy. Here, the channel of visual information feedback is used to help monitoring the sending state of the instructions on one hand (i.e. the first closed loop) and monitoring the executing state of the tasks on the other hand (i.e. the second closed loop), so as to help making the necessary correction.

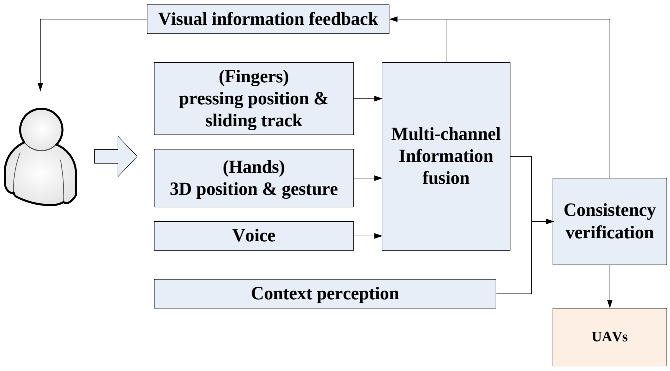

Fig. 2. Technical reference model of the designed multi-channel human-machine interactive system

4.1 Control part

In the control part, the key point is how to fused different control signals with a high accuracy. Different from the previous methods, which would lose a lot of internal associated information because the fusion always happened after the respective identification of each signal, a new end to end fusing method based on deep learning is adopted.

Firstly, the raw signals from different human-machine interactive channels are sent to two types of deep neural networks for pre-processing, where the touching signal and the gesture signal are sent to the Convolution Neural Network (CNN) and the voice signal is sent to the Long & Shot Time Memory Neural Network (LSTMNN) due to their different characteristics.

CNN is suitable for image processing, and both the touching signal and the gesture signal can be just converted to the images, so CNN is used for the processing of touching signal and gesture signal. CNN has three important properties, as shown in Fig. 3.

The local region perception completed in convolution layers (C1, C2, …) can find some local features of the images, which is the basis of visual recognition.

The pooling completed in sampling layers (S1, S2, …) can confuse the position of the feature, which helps the algorithm be strongly robust to the distortion of space and time.

The weight sharing completed in both convolution layers and sampling layers can reduce the number of parameters needed to be trained, which improves the model's performance of generalization.

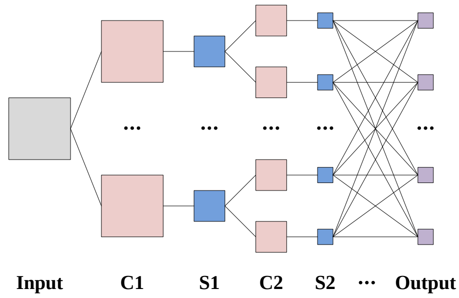

Fig. 3. Structure of CNN

LSTMNN, as an improved form of Recurrent Neural Network (RNN) to solve the problem of gradient disappearance, is suitable for modeling the long sequence like the voice signal, and improving the modeling accuracy as well. For many sequence modeling tasks, in addition to the historical information, the future information is also helpful. Although the single-directional LSTMNN can use all the historical information before the current moment, it cannot use the future information. Thus, the bi-directional LSTMNN is presented, as shown in Fig. 4. The basic idea is: each feature of forward and backward sequences will be sent to two independent auto-regression hidden layers respectively, and the two hidden layers connect with only one input layer and one output layer, which will make the modeling use the historical and future information completely and symmetrically.

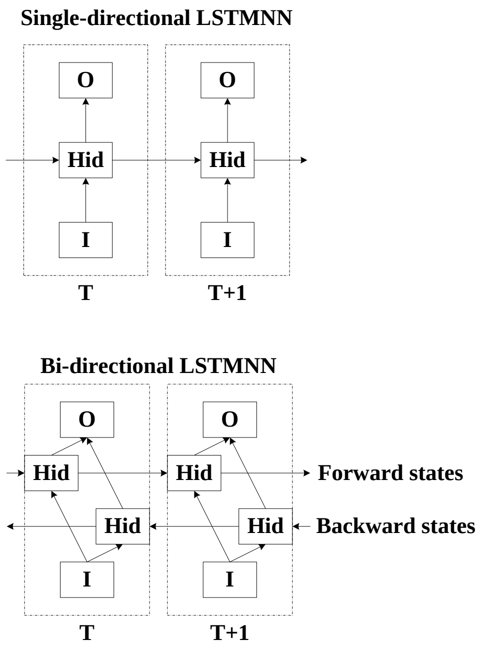

Fig. 4. LSTMNN: Single-directional VS Bi-directional

Secondly, the outputs are combined as a fully meshed matrix. After performing a number of connection operations and calculating the feature vector, the Soft regression is applied to get the whole system function.

$$h_\theta(\mathbf{p}^{(i)}) = \frac{1}{\sum_{i,j=1}^{k} e^{\boldsymbol{\theta}_j^T \mathbf{p}^{(i)}}} \begin{bmatrix} e^{\boldsymbol{\theta}_1^T \mathbf{p}^{(i)}} \\ e^{\boldsymbol{\theta}_2^T \mathbf{p}^{(i)}} \\ \vdots \\ e^{\boldsymbol{\theta}_k^T \mathbf{p}^{(i)}} \end{bmatrix} \quad (1)$$

Where $\boldsymbol{\theta}$ is regression vector, $\mathbf{p}$ is the input sample and $h(\mathbf{p})$ is the output sample. There are k interactive signals and $\mathbf{p}^{(i)}$ denotes the occurrence probability of each interactive signal.

As the generalization of Logistic regression, Soft regression extends the problem of two-classification to the problem of multi-classification. Then the maximum value and the second maximum value can be detected. If the ratio of the maximum value to the second maximum value is bigger than the predefined threshold, the integrated interactive instruction will be a certain result which is corresponding to the maximum value. If the ratio of the maximum value to the second maximum value is equal to or smaller than the predefined threshold, the integrated interactive instruction will be a fuzzy result which is either corresponding to the maximum value or corresponding to the second maximum value. In this case, there are two outputs. Each of them has a cue signal indicating the uncertainty.

At the same time, the consistency verification based on context perception is used to enhance the accuracy of control. Applying the perception and reasoning of a group of states or variables (i.e. context), the intentional task instruction will be generated in light of the current application. Then this instruction is compared to the instruction issued by the multi-channel interactive system. If they are same, the instruction will be directly used for further control. If they are different, the context query mechanism will be started and the most similar result will be suggested. Because the context data have different expressive forms, it is necessary to build a specialized agent to format and extract the basic elements. After shielding useless and eliminating redundancy, the clear and coincident context will be formed for the subsequent processing.

4.2 Monitoring part

In the monitoring part, the key point is how to get the visual information feedback clearly and concisely to help the operator make a further modification of control. After decoding and noise removal, the task related data like the environment data, the instruction data and the action executing data will be displayed via a combined vision system, in which the virtual graph of the indicated parameter is drawn to make a seamless overlap to the real video of the external scene and many different kinds of heterogeneous data are fused in a unified coordinate. This mechanism will provide an intuitive feeling according with human eye observation.

5 Experimental results

In order to test the feasibility and superiority in application of the designed system, the virtual functional control interface was developed based on the traditional interface of UAVs' control, as shown in Fig. 5. In addition to the flight task information and engine parameters, it supported map, navigation, multimedia, and other applications. All of them were imported simultaneously into Vega Prime to configure the virtual scenario using the Lynx Prime graphic design interface. The API functions were called for visual simulation driving as well as the display and the interaction of various functional components. According to the ergonomics, the display subsystem was determined based on functions, frequencies, properties, associations and densities of the displayed information. And the interaction subsystem was developed using Creator and then imported to GL Studio for further model design. Finally, the corresponding C++ class objects of the model were created to complete the internal logic simulation of the device.

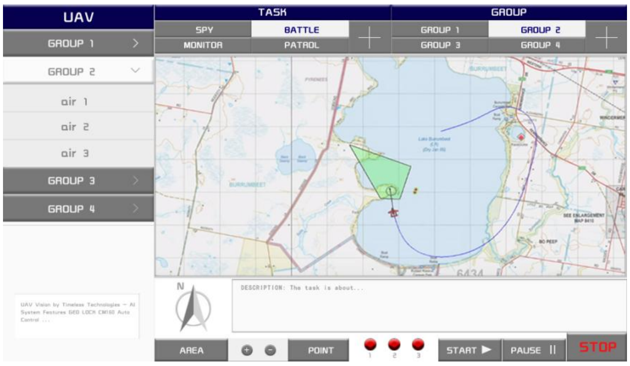

Fig. 5. Virtual functional control interface

Simulation experiments including single channel experiments and multi-channel experiments were performed under three tasks of obstacle avoidance like the air mass avoidance, no fly zone avoidance and moving object avoidance. Ten adults from 22 to 35 year old were chosen and they would be trained two hours before experiments. After they were familiar with the operation and the process, the basic experiments would be started.

The data of the single channel control and the multi-channel control was recorded in Tab. 1 and their comparison was shown in Fig. 6.

The results demonstrated that the multi-channel human-machine interactive system was very promising to improve the capability of multiple UAVs' control, in which it had higher control efficiency than the single-channel system and the operator could control more than four UAVs at one time to avoid obstacles with a higher success rate.

Tab.1. Data of the single channel control and the multi-channel control

A. Control reaction time (s)

Control method \ Numbers of UAVs	1	2	3	4
Touch screen control	1.05	2.2	3.6	4.5
Gesture control	1.28	2.32	3.7	4
Voice control	1.4	2.3	3.35	3.8
Multi-channel control	1.6	1.8	2.5	3

B. Control accuracy (%)

Control method \ Numbers of UAVs	1	2	3	4
Touch screen control	95.75	93.15	91.50	89.01
Gesture control	92.68	88.37	85.17	80.03
Voice control	93.05	90.10	88.70	85.19
Multi-channel control	93.20	91.86	90.24	89.85

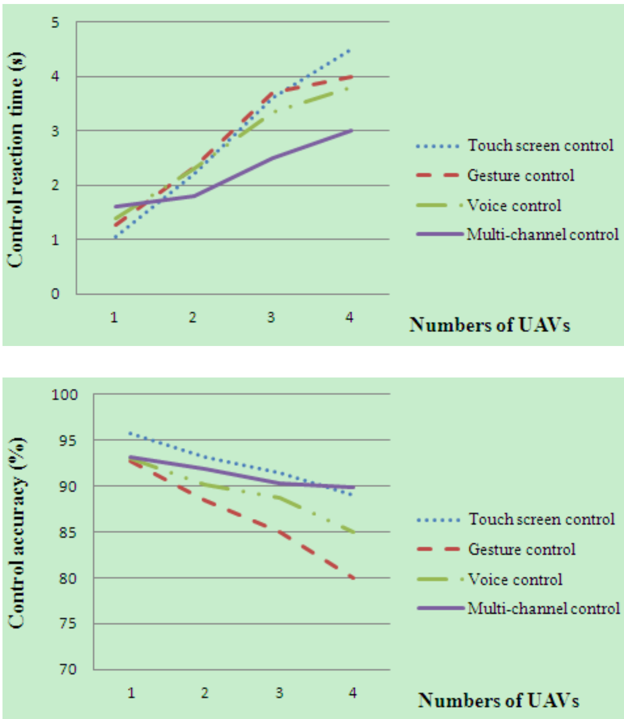

Fig. 6. Comparison of the single channel control and the multi-channel control

6 Conclusions

The designed multi-channel human-machine interactive system is promising for multiple UAVs' control. By discussing the system framework and its consideration, a detailed implementation of the system is presented, which is of good performance in the simulation experiments. Next, a prototype will be developed.

References

[1] Designing a Responsive UAS for Today's Operations and Tomorrow's Missions, Open Publication Reference Number: 88ABW-2009-2331, Air Force Research Laboratory, USA.

[2] Mini UAVs - the Next Small Thing, Jane's International Defense Review, No. 6, 2005, pp. 58-63.

[3] Zhang G, Shen L, Zhu H. The Current Situation and Enlightenment of Supervisory Control Technology for Multiple UAVs. National Defense Science & Technology, 2009, 30 (4): 5-10. (in Chinese)

[4] Ownby, Michael. Mixed Initiative Control of Automa-teams (MICA) - a Progress Report. DOI: 10.2514/6.2004-6483, September 2004.

[5] Corman D, Herm T, Paunicka J, etc. Mixed Initiative Control of Automa-teams Open Experimentation Platform - Final Technical Report. Boeing-STL 2004P0014, 2004.

[6] DARPA's CODE for Unmanned Aircraft to Fly as Collaborative Teams. http://www.uasvision.com/2016/06/07/darpa-establishes-the-code-for-unmanned-aircraft-to-fly-as-collaborative-teams/

[7] Unmanned Systems Integrated Roadmap FY2005-2030, Open Publication Reference Number: 14-S-0553, Department of Defense, USA, 2013.

[8] The Role of Autonomy in DoD Systems. Defense Science Board, Department of Defense, USA, July 2012.

[9] Naval Studies Board. Autonomous Vehicles in Support of Naval Operations. National Research Council, Washington DC, 2005.

[10] Margaret W M, Hugh J F. Seatation and Perception. Massachusetts: Allyn & Bacon, 1996.

[11] Salvendy G. Handbook of human factors and ergonomics. John Wiley & Sons, 2012.

[12] Hecht D, Reiner M, Halevy G. Multimodal virtual environments: response times, attention, and presence. Presence: Teleoperators and virtual environments, 2006, 15 (5): 515-523.

Contact Author Email Address

wpcheng@avic.com

Copyright Statement