Natural UAV tele-operation for agricultural application by using Kinect sensor

Xuanchun Yin¹, Yubin Lan¹,²,³*, Sheng Wen⁴, Jiantao Zhang⁵, Shifan Wu¹

(¹. College of Engineering, South China Agricultural University, Guangzhou 510642, China; 2. National Center for International Collaboration Research on Precision Agricultural Aviation Pesticides Spraying Technology, Guangzhou 510642, China; 3. International Laboratory of Agricultural Aviation Applied Technology, Guangzhou 510642, China; 4. Engineering fundamental Teaching and Trainer Center, South China Agricultural University, Guangzhou 510642, China; 5. College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China)

Abstract: Remote-controlled (RC) unmanned aerial vehicles (UAVs) have been extensively applied in agricultural areas, such as remote sensing, precise spraying pesticides for crop protection, agricultural situation inspection and so on, but these telemanipulated UAVs systems are operated entirely by a ground-based pilot with a need of eyes focus on the remote site UAV flight. The key issue existed in agricultural UAV teleoperation area is a longtime training needed. In this paper, a novel natural UAV teleoperation control system in agricultural application was proposed. In UAV teleoperation scenario, human operator gestures measured by using Kinect sensor can be used as control commands for UAV flight. Moreover, some UAV teleoperation control commands related to human hand gestures were defined, which is similar to the radio gymnastic exercises in China. Therefore, gesture recognition-based UAV teleoperation control is easy to learn and master. In addition, a new real time human hand gesture recognition algorithm was proposed. The stability of UAV flight dynamic of roll, pitch and yaw as well as attitude control were verified with the experiments based on the proposed method. At last, the usability and effectiveness of the proposed method has been verified by the experimental results.

Keywords: gesture recognition, human pose estimation, depth image, skeleton tracking, teleoperation, unmanned aerial vehicle (UAV)

DOI: 10.25165/j.ijabe.20181104.4096


1 Introduction

Unmanned aerial vehicle (UAV), also known as drones, are becoming a lot of popularity among an ever-growing community of amateurs and other service providers. Advent of LTE 4G/5G networks and communication technology, will widen the use case scenarios of UAVs. UAVs are used for providing diverse civilian, commercial, and governmental services. Especially in agricultural, it plays an important role in information acquisition of agricultural production and agricultural condition monitoring. Such as rice lodging monitoring, pest and disease monitoring, crop growth monitoring, crop nutrient diagnosis and so on. Moreover, a big event happened in the last year- agricultural drone finally cleared for take-off-according to new U.S. federal rules[1]. Tech-savvy farmers can use drones to check the conditions of their fields spraying pesticides for their crop protection or chemical fertilizer for the growing of the crops. It becomes easier for farmers in planting crops through drones (UAV) spraying[2-7] than traditional ground spraying machine as well as knapsack sprayer.

The application of pesticides and fertilizers in agricultural areas in China is of significant importance for improving crop yields. The use of UAV is becoming increasingly common in carry these tasks because of pesticide spraying efficiency and effectiveness of UAV spraying manner. However, some factors may reduce the crop yields or even cause damage[8] (e.g. crop areas are not covered during the UAV spraying process, or overlapping spraying of crop areas, or the misuse of pesticide dosage). Naturally, farming management issues has been considered. Precision agriculture (PA) concept is presented and viewed as one of the top ten revolutions in agriculture and widely accepted which dates back to the middle of the 1980’s. Moreover, unmanned aerial vehicles are proposed as a useful alternative to manned aircraft for some animal management.

Although remote monitoring of animal behavior offers great potential to improve livestock management with the information technology development, challenges are still existed in data collecting at high frequency, classification as well as evaluation model. Currently, UAV control has three formats: pre-programed, remote control, and self-thinking. Remote control is, the most common control system for UAVs, the operator (in local site) receives the flight data and other user desired information feedback from the remote site (UAV platform) and sends the flight commands and desired information acquirement instructions to the UAV platform. Moreover, in most cases, UAV remote control is...
realized with GPS based working model. During UAV application research and practice, the flight control and autonomous mission execution is one of the most active area. The significant crucial points are the localization of the vehicle and obstacle avoidance, which is especially difficult in unknown, GPS denied, unstructured environments. In order to solve this problem, the vision-based guidance and control method[9] and vision aided inertial navigation systems (V-INS)[10] have been deeply done in UAV flight control.

Continuous human action or gesture recognition is a crucial and challenging task in many research and application fields, such as Kinect-based method[11] allows a human operator to communicate his motions instructions to the dual robot manipulators for object manipulation task in unstructured environment. Human skeletal data can be extracted from RGB-D data used in computer vision reach area which are captured by a Kinect sensor[12,13]. Human action recognition may become a widespread application in our future life, partly owing to Kinect sensors with the characteristics of low-cost high resolution depth and visual (RGB) sensing. Zhu et al.[14] proposed an online continuous human action recognition algorithm based on the Kinect sensor. Kim et al.[15] proposed a method of classifying Korean pop (K-pop) dances based on human skeletal motion data obtained from a Kinect sensor in a motion-capture studio environment. Literature reviews show that most paper focus on the algorithm research for human hand gesture recognition acquired by Kinect sensor.

It is noted to point out that it is cost effective and complexity in realizing UAV control in GPS denied and unstructured environment. So, a human-in-the-loop control method has been proposed in this paper. In detail, a novel natural UAV teleoperation based hand gesture recognition method through untouched Microsoft Kinect sensor for agricultural inspection or management application has been proposed. The proposed natural teleoperation system has some advantages: 1) it is no need to long train the flight control through operation UAV controller; 2) the farmer only need to bear some basic body gesture to control UAV flight; 3) AR. Drone was chosen as agricultural application platform in GPS denied environment, which is used to add extra lightweight device for information acquisition, such as camera; 4) the designed body gesture for UAV control is similar to radio gymnastic exercise, which makes UAV flight control easy and acquires comfortable sensation during their work.

2 Materials and methods

2.1 System scheme

In this paper, we proposed a natural teleoperation system for animal monitoring by using AR.Drone shown in Figure 1. Human (administer of the cattle livestock farmers) was standing in the front of the Kinect sensor, human gesture captured by the Kinect sensor (operate site or called local site) was used to control the AR.Drone (remote site), the flight situation of AR.Drone and livestock situation are feedback with the image information from image sensor (camera) carried by the AR.Drone.

The whole system is consisted of the gesture recognition subsystem in the local site, and the UAV-based animal monitoring subsystem in the remote site, and the information communication channel, in which the control commands of the UAV flight were transmitted to the remote site from the local site and the feedback information with the flight situation and livestock information were transmitted back from the remote site.

![Figure 1](image1.png)

Figure 1  Proposed natural teleoperation system for animal monitoring by using UAV

2.2 Human hand gesture recognition subsystem

Human gesture recognition subsystem is made by the Kinect sensor 2.0, and human livestock administer and a computer. Kinect sensor is consisting of an infrared (IR) projector, an IR camera, and a color camera shown in Figure 2. The depth sensor comprises of the IR projector and the IR camera, which is a monochrome complementary metal-oxide semiconductor (CMOS) sensor. In this paper, we developed a human skeleton acquisition (especially hand gesture) from multiple Kinect sensors, the experimental scenario is shown in Figure 3.

![Figure 2](image2.png)

Figure 2  Inside structure of the Kinect sensor, the infrared (IR) projector, IR camera, and RGB camera.

![Figure 3](image3.png)

Figure 3  Hand gesture extraction image scenario in our laboratory

In order to acquisition the depth human administer skeleton for natural teleoperation UAV, the acquisition process is shown in Figure 4. The details as follows: firstly, human skeleton depth frame data should be written in a format of array in the computer while justify the one dimension valid. Once the data is valid, output of the skeleton joint array is ready for the articulation joint data process, then the results is used to draw human skeleton by invoking draw () function.

It is needed to point out that, each articulation joint has three state situations: Tracking State_Not Tracked, Tracking State_Tracked, and Tracking State inferred, which means human articulation joint detection failure, successful, and need to infer.
Here, the third situation illustrates the intelligence of the Microsoft Kinect sensor. In our experimental to ensure the safety of tele-operated UAV for livestock inspection, we proposed an easy method-a matching method based on the input human hand gesture database-to eliminate the third recognition situation, which shown in Figure 5.

The matching method consists of two steps. Firstly, according to the defined human gesture used for UAV teleoperation, all of the articulation joint depth data are expressed by a format of coordination and viewed as a standard data base; secondly, the real time acquired human articulation joint depth data matched with the standard data base. If they are matched, the UAV flight teleoperation control command is transmitted from the local site to the remote site by communication channel under Wi-Fi help while the red lamp is on in the local site shown in Figure 8.

Local parameter of human skeleton extraction method is also proposed in this paper to improve the response characteristics of the UAV teleoperation system through making simplification of the calculation during the human articulation joint point recognition process shown in Figure 6. In the method, the human skeleton was projected in 2-D (x-y plane) plane as shown in Figure 6. For example, at time \( t_0 \), we stand in the front of the Kinect sensor, after a sampling time \( t_{\text{sample}} \), at time \( t_0 + t_{\text{sample}} \), up our hands, the change of coordination in the y direction is only to be calculated to recognize the changed human hand gesture. The characteristics of hand gesture recognition are governed by the following equations.

\[
\begin{align*}
\text{hand}_{\text{gt}} & = \begin{pmatrix} x_{t_0} \\ y_{t_0} \end{pmatrix} \\
\text{hand}_{\text{gt}} & = \begin{pmatrix} x_{t_1} \\ y_{t_1} \end{pmatrix}
\end{align*}
\]

On condition that Equation (2) and (3)

\[
\text{hand}_{\text{gt}} = \begin{pmatrix} x_{t_0} - x_{t_0} > 0 \\ y_{t_1} - y_{t_0} > 0 \end{pmatrix}
\]

\[
\text{hand}_{\text{gt}} = \begin{pmatrix} x_{t_1} - x_{t_0} > 0 \\ y_{t_1} - y_{t_0} > 0 \end{pmatrix}
\]

where, \( \text{hand}_{\text{gt}} \) is the human hand gesture sampling at the time \( t_0 \) and \( t_1 \); \( \text{hand}_{\text{gt}} \) means the calibration data of hand gesture sampling at \( t_0 \) and \( t_1 \) in 2-D plane. \( \text{hand}_{\text{gt}} \) means the change of hand gesture, \( \varepsilon \) means the amount of calibration in the 2-D plane, \( \Delta \varepsilon \) is change amount in the experiment, where it is changed with the different human UAV operator. 0.10 is the basic constant according to the UAV natural teleoperation experiment experience. \( \text{hand}_{\text{gt}} \) is the results of hand gesture recognition, \( \text{quadrotorflying commands} \) is relative to the \( \text{hand}_{\text{gt}} \) based on the UAV flying control hand gesture define.

The abovementioned method is only used to simplify the calibration. However, it needs a lot of sampling to establish the standard human hand gesture data base to meet the control requirements. At the same time, during the UAV teleoperation, more high requirements and constraints has deteriorated natural teleoperation characteristics. To this end, an improved local parameter extraction method has proposed to improve the recognition accuracy and speed while keeping the validity of UAV teleoperation commands shown in Figure 7. The characteristics of improved local parameter extraction methods have:

1. Based on the original characteristics, some easy confused human hand gesture used as UAV teleoperation control commands is depended on an additional “triggered condition”. Only if the additional “triggered condition” has been detected, the different control commands have been established.

2. Each triggered condition is extra added “threshold value” to control trigger to ensure the validity of the hand gesture recognition. At the same time, the validity region (sector region III) and invalidity blind sector region (Sector region I and II) is used to justify the validity of recognition hand gesture shown in Figure 7. It is noted that the different threshold value is determined by the different hand gesture characteristic point extraction.
2.3 AR.Drone platform

One of the advantages of unmanned aerial vehicles (UAVs) is that they can be autonomous, being able to inspect the environment that they are in, while carrying out their designated mission. Quadrotors can be used both outdoor and indoor, where several localization techniques (mainly dependent on computer vision) are exploited to sense itself the position and orientation in the GPS denied environment.

For the quadrotors control, the human interface plays a key role. RC-transmitters and joysticks are the most common input devices used to control the quadrotors. The innovative natural user interface for the AR.Drone has been proposed in [16-18], where the hand gestures are translated into commands for the platform. The AR.Drone 2.0 (shown in Figure 8) was selected for the purposes of this experiment because it is relatively cheap compared to other drones and can be easily accessible for anybody willing to test it. The parrot AR.Drone is a quadrotor helicopter with Wi-Fi communications and two on-board cameras[19,20]: a wide angle front camera and a high speed vertical camera shown in Figure 4. Although the quality of the signal drops rapidly with the distance, according to test, 100 m range distance is not need to add extra signal amplification.

2.4 Establishment of natural-UAV teleoperation system

According to livestock inspection application, eleven UAV flight control commands based human hand gestures has been defined during in the natural UAV teleoperation process shown in Figure 9. During the design process, four key points should bear in the mind: (1) there is no conflict or overlap hand gesture existed in UAV control command; (2) pay attention to the logical and inclusion relationship existed in the control command, for example, some flight attitude control commands included the takeoff instruction; (3) safety flight is significant important during the design process, to this end, hovering of the UAV flight control command hand gesture design should keep completely different with other control command; (4) all of the hand gestures for flight control should keep appropriate blind section to avoid mis-trigger.

### Table 1 Relationship between the flying command and characteristics points needed to recognition

<table>
<thead>
<tr>
<th>FC</th>
<th>CP</th>
<th>FC</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Take off</td>
<td>CoRP, EoRA</td>
<td>Hovering</td>
<td>No CP</td>
</tr>
<tr>
<td>To the left</td>
<td>CoRP, RT</td>
<td>To the right</td>
<td>CoRP, RT</td>
</tr>
<tr>
<td>Forward</td>
<td>CoLP, EoLA</td>
<td>Backward</td>
<td>CoLP, EoLA</td>
</tr>
<tr>
<td>Ascend</td>
<td>CoLP, EoLA</td>
<td>Descend</td>
<td>CoLP, EoLA</td>
</tr>
<tr>
<td>Turn left</td>
<td>CoRP, EoRA</td>
<td>Turn right</td>
<td>CoRP, EoRA</td>
</tr>
<tr>
<td>Landing</td>
<td>CoLP, EoLA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Flying command (FC), characteristics points (CP), center of right palm (CoRP), elbow of right arm (EoRA), Right thumb (RT), center of left palm (CoLP), elbow of left arm (EoLA)

The advantage of the proposed hand gesture recognition algorithms is that the characteristics points (CP) needed to extract are few, such as center of right palm (CoRP), elbow of right arm (EoRA), Right thumb (RT), center of left palm (CoLP) and elbow of left arm (EoLA). Therefore, it is easy to establish data base for recognition of hand gesture. In addition, the human operator for natural UAV teleoperation must remember the hand gesture definition, as shown in Figure 9.

![Figure 8 AR.Drone flight attitude control by adjusting roll, pitch and yaw angle](image)

![Figure 9 Design of human hand gesture for UAV teleoperation](image)

3 Experiment and results

In the section, the effectiveness of human hand gesture recognition algorithm experiment has been done firstly.

(1) Human hand gesture recognition algorithm has been proposed in this paper, firstly, the local parameter of human hand gesture extraction and match with the designed standard gesture method experiments has been done, each UAV teleoperation instruction has been tested 100 turns, the times of success of recognition has been written down, the rate of success of hand gesture recognition has been calculated. The whole results are shown in Table 2. In order to improve the success recognition rate, the improved method has been proposed, four kinds of flight instruction has been tested 100 turns, the rate of recognition and the rate of improvement recognition have been calibrated and shown in Table 3.

### Table 2 Local parameter of human skeleton extraction and match method

<table>
<thead>
<tr>
<th>Instruc.N</th>
<th>ToRS</th>
<th>RoRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Take off</td>
<td>100</td>
<td>100%</td>
</tr>
<tr>
<td>To the left</td>
<td>84</td>
<td>84%</td>
</tr>
<tr>
<td>Forward</td>
<td>85</td>
<td>85%</td>
</tr>
<tr>
<td>Ascend</td>
<td>80</td>
<td>80%</td>
</tr>
<tr>
<td>Backward</td>
<td>82</td>
<td>82%</td>
</tr>
<tr>
<td>Landing</td>
<td>96</td>
<td>96%</td>
</tr>
</tbody>
</table>

Notes: Instruction names: Instruc.N; Time of recognition success: ToRS; Rate of recognition success: RoRS.
Table 3  Improved local parameter recognition method

<table>
<thead>
<tr>
<th>Instruction</th>
<th>ToRS</th>
<th>RoRS</th>
<th>RSoI</th>
</tr>
</thead>
<tbody>
<tr>
<td>To the left</td>
<td>93</td>
<td>93%</td>
<td>9%</td>
</tr>
<tr>
<td>To the right</td>
<td>96</td>
<td>96%</td>
<td>11%</td>
</tr>
<tr>
<td>backward</td>
<td>95</td>
<td>95%</td>
<td>13%</td>
</tr>
<tr>
<td>Forward</td>
<td>98</td>
<td>98%</td>
<td>18%</td>
</tr>
</tbody>
</table>

Notes: RSoI: rate success of improvement

Then natural UAV teleoperation experiments based on the proposed human hand gesture recognition algorithm have been done to verify the usability of natural UAV teleoperation concept.

During the natural UAV teleoperation experimental testing, in the local site, UAV flight control commands acquired and represent to the human operator through watching the Kinect based depth skeleton acquiring window, shown in Figure 10. View this image window, human operator can sense and understand the UAV flight control commands. At the same time, this image window can be used to train the human operator in the local site through correcting and standardizing the human actions to ensure safety UAV flight in the remote site. UAV flight situation through UAV flight situation feedback window provided to the human operator for inspect the UAV flight shown in Figure 11.

Then, natural UAV teleoperation experiments based on the proposed human hand gesture recognition algorithm have been done to verify the usability of natural UAV teleoperation concept.

During the natural UAV teleoperation experimental testing, in the local site, UAV flight control commands acquired and represent to the human operator through watching the Kinect based depth skeleton acquiring window, shown in Figure 10. View this image window, human operator can sense and understand the UAV flight control commands. At the same time, this image window can be used to train the human operator in the local site through correcting and standardizing the human actions to ensure safety UAV flight in the remote site. UAV flight situation through UAV flight situation feedback window provided to the human operator for inspect the UAV flight shown in Figure 11.

Then, the stability of UAV teleoperation experiments have been verified through UAV flight attitude control (roll, pitch, yaw, and attitude) based on human hand gesture recognition algorithm. The experimental results are shown in the following figures: Figure 12 shows roll experimental test, Figure 13 shows pitch experimental test, Figure 14 shows yaw experimental test, and Figure 15 shows the flight attitude experimental test respectively.

Figure 10  Local site: Kinect based depth human hand gesture acquire window

Figure 11  Remote site: UAV flight situation feedback monitoring window

Figure 12  UAV roll control results feedback

Figure 13  UAV pitch control results feedback

Figure 14  UAV yaw control results feedback

Form the experiments results, the stability of all flight situation can be obtained based on the proposed human hand gesture recognition algorithm. In Figure 12, the red part shows roll angle real measured situation when the UAV flight in the right direction, the blue means UAV flight in the left direction. In Figure 13, the yaw angle was measured in real time by the UAV tune to the right direction in red part, and turn left direction in the blue part. In Figure 14, the UAV yaw flight control experiments, the red shows the UAV flight in backward direction and the blue in forward situation. In Figure 15, the UAV flight attitude has been measured by ascend flight (red) and descend flight (blue).
From the above experiments, the effectiveness of the proposed human hand gesture recognition algorithm has been verified and the usability of natural UAV teleoperation by using Kinect sensor concept has been evaluated.

Figure 15  UAV flight altitude control results feedback

4 Conclusions

In this paper, a natural UAV teleoperation control system for agricultural application based on human gesture recognition has been established. A series UAV teleoperation control command related human hand gestures are defined, which is refined by the radio gymnastic exercises. It is easy to learn and master for the farmer in UAV agricultural application. In addition, a new human hand gesture recognition algorithm, which has an advantage of easy to build the data base for human hand gesture recognition, has been proposed and the effectiveness has been verified through the experiments test the defined human gesture. Moreover, in the local site human operator can sense the control command and monitoring the remote site UAV flight situation watching window has been designed. At last, the UAV roll, pitch, yaw and attitude control is experimental verified in the experiments.

In future, some experiments should be done with the local farmer in agricultural application. Such as, cows and sheep monitoring free-ranging flocks in Inner Mongolia prairie. At the same time, the proposed natural UAV teleoperation system can also be used to entertainment area.

Acknowledgments

The research was supported by the National Natural Science Foundation of China (Grant No.61773171), National Key Research and Development Program of China (No.2016YFD0200700); Science and Technology Planning Project of Guangdong Province (Grant No.2017B010117010), Guangdong Natural Science Foundation (Grant No.2017A03103083), Educational Commission of Guangdong Province of China for Platform Construction: International Cooperation on R&D of Key Technology of Precision Agricultural Aviation (Grant No.2015KJGHZ007).

|References|