Research on the Application of UAV Wide-angle Shooting Technology in Multimedia Video Playing

Peitong Liu¹, Denghan Wang²,*

¹Department of Journalism and Communication, Nanjing University of Aeronautics and Astronautics
²Digital Media Digital, Nanjing University of Aeronautics and Astronautics

*Corresponding author: denghanwang@nuaa.edu.cn

Abstract. Unmanned aerial vehicles continue to increase the requirements for image resolution in the fields of reconnaissance, surveying, and mapping. The fusion of the video image system and the traditional remote control and telemetry system also needs to be considered in the design. Aiming at the above problems, the article designs and implements an airborne high-definition video image system based on FPGA+DSP. The system can store the video image on the local hard disk, and at the same time, the compressed video image is also convenient for real-time transmission to the ground control station through the wireless link equipment. The system detection found that the buffer occupancy can fluctuate smoothly in a small range, and the subjective image quality has also been improved, which better meets the low-latency application requirements of the airborne reconnaissance system.

Keywords: UAV; wide-angle shooting; multimedia video playback; video compression

1. Introduction

UAVs are widely used in military reconnaissance and civil surveying and mapping, among which the airborne video image system is one of the important links in the airborne electronic system. UAVs take pictures of ground scenes in high-altitude flight, and the target pixels in the obtained image frames are small and the number of targets is large. This requires increasing the image resolution to improve the recognition of the target object. Currently, the amount of data increases significantly, which is in contradiction with the limited wireless bandwidth resources. Therefore, it is required to apply efficient video coding and decoding technology. In special application fields such as reconnaissance, real-time video transmission is required. At this time, real-time video encoding and decoding and small image delay are required to be guaranteed. In the case of multi-channel video transmission and a fixed total data rate, it can be achieved by reducing the single-channel video resolution or increasing the single-channel compression ratio, so the system should be able to dynamically switch the resolution and compression ratio, with greater flexibility [1]. At the same time, with the continuous development of wireless broadband communication technology and video compression technology, drones can not only store the video information shot in the air locally for later analysis, but also transmit the compressed video images to the ground control station in real time. The operator controls the flight status and
reconnaissance information in real time. Here, an efficient visible light and infrared dual-channel video image compression and storage system is given to provide technical support for ground personnel to realize real-time remote control of UAVs and prevent collisions and obstacles.

2. System composition principle
The UAV multimedia video system mainly includes UAV system hardware integration module, data transmission link module, software processing module, etc. The overall design process is shown in Figure 1. The drone platform is equipped with high-definition video equipment, positioning and attitude module, and wireless transmission and transmission device [2]. The platform receives ground station instructions through the radio, and sends the acquired sensor position and attitude information to the multimedia video system, while the video signal is returned by the image transmission transmitter After being received by the image transmission antenna and processed by the video capture card, it is also input to the multimedia video system; due to the high overlap of adjacent video frames, the video frames need to be filtered at a certain frame interval. This part of the video frames is called key frames.

The multimedia video system uses image processing technology to extract video key frame images, combines digital camera distortion check files to correct the key frame distortion, and uses the survey area average elevation and external orientation element data to perform orthorectification on the corrected image to generate an orthophoto Image; the two-dimensional multimedia platform obtains the UAV's position data in real time and displays its track coordinates. At the same time, the orthophoto is matched with the two-dimensional multimedia data in real time according to the coordinates, and finally realizes the real-time display of the UAV track coordinates and video Functions such as positioning and distance area measurement.

3. System hardware and software

3.1. Hardware equipment

3.1.1. Airborne video capture. To ensure signal quality, high-definition analog video signals often use component output formats, such as the three-component output of luminance signal (Y) and color difference signals (Cr, CBS). At the same time, vertical synchronization and horizontal synchronization signals can adopt internal synchronization superimposed on the component signal. Mode output. The high-definition video signal in this system adopts 720p (1280×720) format, 25 frames per second, and selects TI's video ADC chip TVP70025I to complete the analog-to-digital conversion of the high-definition analog video. The sampling accuracy and method are selected as 8 bits, 4:2:2(Y: Cb: Cr)
sampling. The connection relation between TVP70025I and external device is shown as in Fig. 2. TVP70025I outputs digitized video components (Y, Cb and Cr), data clock (DATACLK), vertical synchronization (VSOUT) and line synchronization (HSOUT) signals to FPGA, DSP initializes TVP70025I internal registers through I2C bus, and starts modulo Conversion process.

![Fig 2. TVP70025I hardware connection](image)

3.1.2. Airborne DSP+FPGA processor. The airborne terminal processing part mainly completes the functions of digital video signal acquisition, video compression coding and remote-control telemetry interface. Considering the powerful digital signal processing capability of DSP and the large-scale logic design capability of FPGA, the system adopts the hardware architecture of FPGA+DSP. FPGA chooses Altera company StratixII series chip EP2S30F672C5, this chip has abundant logic resources and embedded memory, supports LVTTL, LVDS and other interface level standards. The main function design is shown in Figure 3.

![Fig 3. FPGA functional design](image)

3.2. Software design
Hi3531D is the main control SOC processor, which completes the functions of video encoding, storage, and communication management. FPGA realizes video and communication interface processing,
completes video and data format conversion, and docks with the main control SOC. The functional architecture of Hi3531D software is shown in Figure 5. The basic functions implemented are as follows:

1. Video encoding. Complete the coding of video data, the coding mode H264/265 is configurable, and the bit rate is configurable.

2. Video playback function. Support 1-channel video playback, that is, read recorded video files for playback.

3. OSD module. The aircraft attitude information, carrier pod information, etc. are superimposed into the video and encoded together to achieve local storage and transmission.

4. Video file storage management function. The encoded video can be stored in the electronic disk according to the corresponding format, and it can be automatically overwritten (when the disk is full).

5. RS422 interface control command analysis. Receive control instructions and analyze, and realize functions such as video recording, video playback, OSD overlay, and synchronous transmission of packaged data according to requirements.

6. BIT acquisition function. The storage system status information is reported periodically, including the remaining recording time, the number of recorded files, and the storage system status. The software is based on Linux "2.6.38SMP development kit, which has strong reliability and is easy to upgrade, expand and maintain.

4. System module design

4.1. Video display and storage module
This module is developed by Microsoft DirectShow under the Windows platform. First, it detects whether there is a video signal input [3]. If there is, it converts each frame of YUYV video data acquired into an RGB image and draws it on the display device; the video can be displayed while being displayed. Storage, storage is divided into two formats: one is the AVI format directly encapsulated without compression encoding; the other is via H. MP4 format encapsulated after 264 compression and encoding.

4.2. Positioning and attitude data receiving module
This module receives real-time positioning and attitude determination data from the radio. The GPS information is parsed according to the NMEA-0183 protocol. It mainly obtains fields 1, field 3, and field 5 in each packet of data $GPRMC sentence; the attitude data includes the yaw angle of the sensor, Pitch angle, roll angle, these three-angle data need to be processed into the external azimuth angle element used in photogrammetry, that is, the HPR coordinate system is converted to the OPK coordinate system.

4.3. Key frame extraction and processing module
The key frame extraction mainly adopts the method based on time interval. First, calculate the time interval of adjacent key frames according to the aerial photography design parameters, and then use the Set Timer() timer function in the MFC library to call the key frame extraction function, which mainly uses OpenCV The inwrite() function under the library; when the key frame is extracted, the time-stamp method is used to synchronize the time of the video data and multimedia, and the key frame image is corrected according to the camera calibration file, and finally the orthoimage is generated.

4.4. Two-dimensional multimedia platform application modules
The system uses Baidu online map as a two-dimensional multimedia platform, and the drone location coordinates are transferred to the map marker interface of Baidu Maps JavaScript API for track display in real time; the orthophotos are posted in real time by the map overlay interface of Baidu Maps JavaScript API according to the coordinates [4]. Map, with the help of Baidu map measurement tool to realize the distance and area measurement of the video frame image.
5. Software module related algorithms

5.1. Video codec algorithm
Currently, the most used video codec standards include the MPEG-4 standard established by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), and the H.264 standard established by the ISO and the International Telecommunication Union (ITU). Compared with the MPEG-4 standard, H.264 has high complexity, especially for high-resolution applications, the hardware implementation is very difficult, and MPEG-4 has low implementation complexity and is suitable for high and low bit rate videos. It has been widely used in embedded video coding and decoding systems. Considering the above points, the MPEG-4 standard is selected as the video codec algorithm in this system. The open-source MPEG-4 video codec XVID Simple Profile is selected as the source code in the system. The airborne terminal completes the video encoding function. After removing some redundant codes such as user interaction from the source code, the transplantation to the TMSC320C6455 platform is completed [5]. The XVID program for PC can't meet the requirements of real-time coding, the program needs to be optimized: improve the running speed. First, some algorithms are optimized, and motion estimation based on prediction is adopted. Secondly, code optimization: The first step is to use CCS software to analyze the time-consuming parts of each part of the code, and to observe the less efficient parts; the second step is to adopt C language-level optimization, such as using keywords and pseudo-instructions to provide optimization information to the compiler. Use inline functions, write software pipelines, etc.; In the third step, select the code segments that are not obvious in the second step, such as quantization/inverse quantization, 1/2 pixel interpolation, and motion compensation, and rewrite them with linear assembly. The upper computer at the ground station completes the video decoding function, and the source code can be tested to meet the real-time performance of video decoding.

5.2. Target tracking algorithm
We assume that \( \{x_i\}_{i=1}^n \) is the pixel points at different positions after the target area is normalized, and take out the points in the target area in turn. If the \( FLBP_{8,1}^{c} \) value of the point is not 0, then the Cb and Cr color components of the point are quantified as required to obtain the corresponding Eigenvectors. If the \( FLBP_{8,1}^{c} \) value of this point is 0, it means that the point is a smooth or noisy point, and no statistics are performed. In order to reduce the influence of noise on the peripheral sample points of the target, a weighted histogram is used, and different weights are given to the pixels according to the distance of different pixels from the target center. Let \( \{x_i^*\}_{i=1}^N \) be a pixel whose \( FLBP_{8,1}^{c} \) value is not 0 in the target area, then the weighted histogram of the target is calculated as follows:

\[
q_u = \sum_{i=1}^{N} k(\|x_i^*\|^2)\delta\left[ c\left(x_i^*\right) - u \right], \quad u = 1, ..., m
\]

Among them, \( c\left(x_i^*\right) \) means that the pixel point \( x_i^* \) whose \( FLBP_{8,1}^{c} \) value is not 0 in the target area is mapped to the index value in the corresponding histogram box, \( u \) is the index of the histogram box, \( m \) is the number of elements in the feature space, \( m = 8 \times 8 \times 5 \). \( k(\|x_i^*\|^2) \) in this article. It is a monotonically decreasing kernel function. Its function is to give a smaller weight to pixels far from the center of the window, where \( \|x_i^*\|^2 \) is the distance from point \( x_i^* \) to the center of the target window [6]. This article uses the most used kernel functions:
6. System detection

First, initialize a search window in the current frame. The center position of the search window in the previous frame is the center, and the window size is slightly larger than the target area. Take out each point in the search window in turn. If the value of the point is not zero, then the Cb and Cr color components of the point are quantified. By querying the target’s color-texture joint histogram model, you can get the pixel as the probability of the target pixel. If the value of the pixel point is zero, then the probability of the point is zero. All the pixels in the search window are replaced by the probability values of the corresponding pixels in the histogram, and the color-texture joint probability distribution map is obtained. Find the centroid position of the target in the probability map by Manshift, and iterate in this way to realize the tracking of the target in continuous video frames. Figure 4 shows the calculation results of the color-texture joint probability projection map. It can be seen from the figure that the feature values of the target are mainly concentrated in the five peak areas. This is because the histogram statistics are only performed on the pixels in the target area whose \( FLBP_{8,1} \) texture value is not 0. The color-texture joint probability projection map obtained by calculating the entire image [7]. At this time, the area of the moving object with similar characteristics to the target in the image is also retained, which will cause interference to the tracking target in the follow-up tracking. The color-texture joint probability projection map obtained by calculating only the area in the search window, where the size of the search window is set to increase by 20 pixels based on the actual target height and width. It can be seen from the figure that the probability value of the other regions except the search window is 0, which well suppresses the interference of the background with similar characteristics and other moving objects, and reduces the amount of calculation.

\[
k(r) = \begin{cases} 
1 - r & \text{if } r \leq 1 \\
0 & \text{else}
\end{cases}
\]  

(2)

Histogram model of normalized target:

\[
p_x = q_x / \sum_{i=1}^{N} k(||x_i||)
\]  

(3)

Fig 4. Color-texture joint probability projection map taken by drone
Figure 5 shows the tracking results of the Camshaft algorithm based on the color-texture joint feature. It can be seen from the figure that the improved Camshaft algorithm can accurately track the target because of the obvious difference between the target and the background color when tracking the player in white. Since the background is relatively smooth relative to the target, and the adopted texture features can effectively suppress the smooth background and only extract the useful texture of the target, the Camshaft algorithm realizes accurate tracking of the target based on the extracted effective target features. Experiments show that under strong background interference, the improved Camshaft algorithm can still track the target accurately, and the tracking performance is significantly improved compared to the traditional Camshaft algorithm based on color features.

![Fig 5. Tracking result of Camshift algorithm based on color-texture joint feature](image)

7. Conclusion
The paper designs and implements a high-definition video encoder based on the DSP platform. Through the optimization of the system structure and encoding algorithm, it effectively solves the problems of large data volume and high computational complexity of high-definition video encoding, and meets the requirements of real-time and reliability. Engineering application requirements. The results of the loading test and the flying test show that the encoder can encode high-definition video in real time, with good picture quality and a low bit rate control. The encoder has been successfully applied to a certain type of UAV ground monitoring system, and all indicators have passed the acceptance.

References
[1] Singh, V. P. RtC Drone: Implementation of intelligent autonomous patrolling using Round the Clock Drone. Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol.12, pp. 582-590, Thirteen 2021.
[2] Mototolea, D.& Mindroiu, A. Counter-Unmanned Aerial Systems. Scientific Bulletin" Mircea cel Batran" Naval Academy, vol. 22, pp. 1-5, January 2019.
[3] Lee, K. H. Jeong, J. K.& Ryu, G. S. Research on the Design of Drone System for Field Support Using AR Smart Glasses Technology. Journal of the Korea Convergence Society, vol. 11, pp. 27-32, April 2020.
[4] Kim, H. G.& Kim, Y. H. Design of Water Surface Hovering Drone for Underwater Stereo Photography. Journal of Convergence for Information Technology, vol. 9, pp. 7-12, June 2019.
[5] Li, Q. Wang, G. Sarrafi, A. Niu, X. Lu, W. & Mao, Z. Dynamic Characteristics Identification of an Arch Dam Model via the Phase-Based Video Processing. KSCE Journal of Civil Engineering, vol. 25, pp. 140-152, January 2021.

[6] Mademlis, I. Mygdalis, V. Nikolaidis, N. Montagnuolo, M. Negro, F. Messina, A. & Pitas, I. High-level multiple-UAV cinematography tools for covering outdoor events. IEEE Transactions on Broadcasting, vol. 65, pp. 627-635, March 2019.

[7] Cho, I. J. Kim, G. B. & Park, B. S. Development of a Forest Fire Tracking and GIS Mapping Base on Live Streaming. Journal of Convergence for Information Technology, vol. 10, pp. 123-127, October 2020.