Introduction to the Project

Research on Real-Time Panoramic Video Stitching Technology

Yuan Yuan, Li Na, Bai Yiyun, Sun Dengyu, Hu Huiting
Supervised by
Prof. Yao Li
School of Computer Science and Engineering, Southeast University
yyuan1222@163.com

I BACKGROUND

With the popularity of consumer electronics, more and more people come to use digital cameras or other devices to obtain videos. A single device acquires a relatively small field of vision and even though the field of vision can be expended by adjusting the focal length, the resolution will be reduced as a consequence. Many times, people need to obtain videos of large vision and the usually adopted methods are as follows:

(1) Move or rotate video capture devices like pan-and-tilt camera to get an overview of the surroundings. But only one local vision can be obtained at a time.

(2) Use multiple video capture devices as cases in supermarket and other occasions that need surveillance. But the output is a number of separate interfaces, lacking the feeling of integration.

(3) Use fisheye camera. However, such devices are often expensive and may exhibit varying degrees of distortions.

Video stitching can be a good solution to all those problems. Video stitching refers to the technology of stitching small-vision videos collected by multiple devices into a video of large vision. Its advantage lies in that it can reduce the required number of video capture devices, while the method is simple and the resolution will not be reduced when expending the field of vision. Video stitching technology can greatly improve people's ability in distinguishing, perceiving and monitoring things and scenes. Video stitching technology can be applied to fields like military and defense, security monitoring, having great significance in reducing risks and improving security. This technology is also widely used in fields like resource exploration, satellite remote sensing, video conferencing, medical and health, automotive rear-view mirror systems, greatly facilitating people's production and life.

The research background of this project is a virtual-reality live system, aiming to collect remote events (sports events, stage performances, etc.) in real time and transmit them to the client. Users can use 3D helmets to watch the events and rotate their perspectives freely, having a strong sense of immersion, as if they are on the scene. Virtual-reality live system consists of three parts: video collecting, panoramic video generation and transmission, live broadcast and helmet interaction. The main research content of our project includes video acquiring, video stitching and panoramic video generation and transmission, preparing for the subsequent interaction with the helmet.

II IMPLEMENTATION

Video stitching is a kind of technology based on image stitching to acquire large-vision video. Our project technically realized the function of stitching prerecorded videos taken by digital cameras and real-time videos captured by USB cameras. And the generated panoramic video can be saved on local machine or livestreamed on RTMP server.

In order to meet the requirement of timeliness, it is necessary to optimize each step of the video stitching process. Video stitching can be divided into video processing and image stitching. The main steps of image stitching include image preprocessing, feature detection, feature description, feature matching and image fusion. Due to the restriction on the number of devices, we purchased cameras with wide-angle lens. Though the number of the required cameras has been reduced, image distortion occurs.

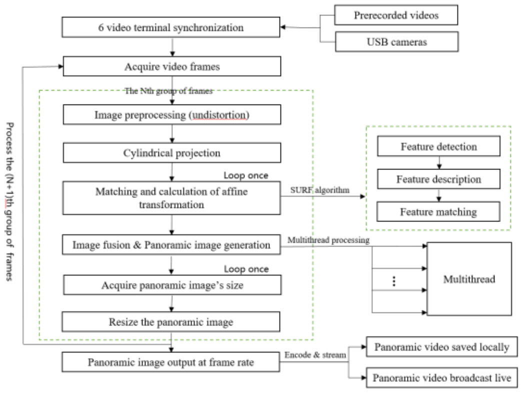

Figure 2.1 Overview of the system

The detailed panoramic video stitching steps are as follows:

(1) Calculate the intrinsic matrix and distortion coefficients using Zhang's camera calibration algorithm.

(2) As for prerecorded videos, extract image frames from each video using FFmpeg. As for USB cameras, grab a frame from each camera using OpenCV.

(3) Undistort frames with the distortion coefficients obtained previously and project the undistorted frames on to a cylinder.

(4) Detect feature points using SURF algorithm and match features using KNN and RANSAC.

(5) Combine 6 frames into a panoramic image using multithreading.

(6) Encode each panoramic image into the panoramic video file using FFmpeg.

(7) Save the video file locally or stream each panoramic frame to your RTMP server for live broadcast.

Since the configuration required for the devices are relatively high and we are underfunded, the digital cameras we purchased does not support remote controlling multiple cameras at the same time; that is to say we cannot start recording simultaneously. To solve this problem, we use special sounds like applauses as synchronizing signals, use BASS library to obtain volume corresponding to every single frame, and use self-designed algorithm to acquire the serial number of the first synchronized frame and the last synchronized frame. By doing so, the videos can be synchronized.

III RESULTS

Figure 3.1 Generated panoramic image

Figure 3.2 Generated panoramic video

IV FUTURE WORK

(1) Reduce the dependence on hardware. The current system demands high consistency between each camera's projection radius (the focal length of the camera + the distance from the camera to the center of the positioning board). Besides, our DIY chessboard and the positioning board have also produced some errors.

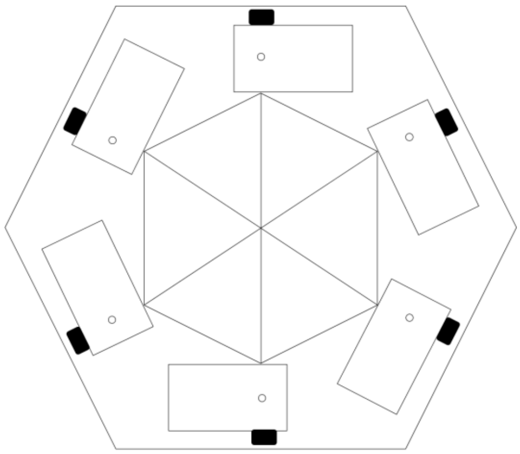

Figure 4.1 Positioning board

(2) Speed up the system. Our current system only run parallel processing in the image fusion part. With future improvements, we probably can conduct image preprocessing, image registration, cylindrical projection and etc. in parallel. Moreover, we can also hand over part of the image processing job to GPU for execution.