Advances in Engineering Research, volume 150

4th International Conference on Machinery, Materials and Computer (MACMC 2017)

Research on Real-time Video Stitching Technology

Tao ZOU[1, a], Fang Deng[1]

[1]Beijing University of Posts and Telecommunications, Beijing, China

[a]email: 364583297@qq.com

Keywords: Image Stitching; Image Registration; Image Composition; Feature Correlation; Panorama

Abstract: Limited by the viewing angles and sizes of imaging instruments like cameras, large pictures can't be generated, and image stitching can solve this problem. image stitching is such a technique that seamlessly mosaics two or more adjacent images with partially overlapping and thus generates a high-definition picture with a relatively wide viewing angle or a panorama with a 360º viewing angle. This technique involves computer vision, computer graphics, image processing and some mathematical tools and so on. And it makes it possible for noise reduction, field of view (FOV) widening, removal of mobiles, blur removing, spatial resolution improvement and dynamic range enhancement. The paper gives a systemic introduction to the origin, actuality, field of application and mosaic methods and gives a respective and detailed introduction to the two main processes: image registration and image composition. Image registration is the core technique of image stitching. The paper also classifies and summarizes the methods of image registration and described their advantages and disadvantages. At last, it gives a detailed introduction to the common methods of image composition at present and puts forward a feasible method of real-time image stitching.

1. Introduction

In recent years, the image stitching-based wide-FOV imaging technique is much studied in digital imaging field and image registration is the core technique of it. Through earlier technological exploration, image registration method based on geometric features becomes the focus and mainstream in image stitching. Richard Szeliski proposed the movement-based panorama image stitching model in 1996, which used Levenberg-Marquardt iteration nonlinear minimization and conducted image registration by the transformation relation between images. The advantage is that it can better process the image stitching that contains various transformations such as translation, rotation and affine, making this a classic algorithm and making him the founder in the field of image stitching. Through years of technology upgrading, the interactive visual media research team of Microsoft developed the famous image panorama mosaic software known as Microsoft ICE (Image Composite Editor) and summarized the method into basic courses of image stitching. There is a wide application prospect for image stitching, such as on-board panorama image stitching and monitoring image stitching. Its key requirement is real-time performance, namely effective and quick completion of image stitching in a designated area. The earliest image stitching method was to select the overlapping area as the template and search similar corresponding modules in other images to match the corresponding position. However, this method could hardly realize the real-time performance due to the huge calculated amount. The precondition for image stitching is that adjacent images shall be logically identical in some parts, namely there must be certain overlapping. The main work of mosaic is to determine the overlapping degree of adjacent images in width and height, remove the overlapping and smoothly joint to the panorama. General procedures include: (1) shoot the scene image sequences by cameras; (2) digitize all the image frames for computer processing; (3) geometrically calibrate the images (optional) so that the same scene can have an identical shape and consistent relative space position in the overlapped image; (4) Determine the registration method for image registration and positioning.

2. Algorithm Process

Basic process for image composition: calibrate camera parameters, collect video images, calculate the distortion correction parameters, calculate homography matrix and composite the video images. The most important and time-consuming step is to calculate homography matrix. In this step, we should first extract the feature points of each image, calculate the affine or perspective transformation parameters of the relative reference plane according to the matched point coordinates, and then composite the overlapping area of the image on the reference plane to get the panorama. If video image stitching calculates the homography matrix of each image frame in the video streaming, there will be reduction in accuracy and failure in real-time mosaic. The solve this problem, this paper proposed the idea of extracting the transformation parameters between images by control frames, the method of which is to fix the relative position of the cameras, use the cameras to control the homography matrix parameters obtained from image frame calculation, establish a parameter list according to the transformation parameters obtained by control frames, then project each image frame in the video streaming to the panorama reference plane via the parameter list, so as to form the final panorama video streaming.

3. Camera Calibration

Camera calibration is a process to establish the correspondence between the position coordinate of camera image pixel and the position coordinate of spatial scene and refer to the appropriate camera imaging model to solve the parameters of camera model according to the imaging point and spatial point coordinates. By camera calibration, we can solve the parameters of camera model according to the known image coordinates and world coordinates of the feature points. The calibrated model parameters include intrinsic parameters and distortion parameters and parameters obtained by calibration are used for image distortion correction.

The process of camera imaging is the transformation of several coordinate systems in essence. First, a point in the space is transformed from the world coordinate system into the camera coordinate system, then projected to the imaging plane (image physics coordinate system) and at last transformed to the image plane (image pixel coordinate system). The distortion-free coordinate (U, V) in image pixel coordinate system (uOv coordinate system) falls on coordinate (Ud, Vd) on the uOv coordinate system after radial and tangential distortion. In other words, the relation between Real image imgR and distorted image ImgD is imgR (U, V) = imgD (Ud, Vd). All the imgR (U, V) can be found out by this relation. The coordinate (Ud, Vd) projected from (U, V) to (Ud, Vd) is usually not an integral number (U and V are integral numbers. Since it is the pixel coordinate position used to form the image, use the coordinate position of the normal image to solve the coordinate position in the distorted image and extract the corresponding pixel value, which is also the pixel value of the normal image).

4. Feature Description and Matching

Feature extraction and feature matching are the most time-consuming in the process, and the difficulty of realizing the real-time performance in the real-time video mosaic lies in this. Meanwhile, the accuracy of feature matching directly affects the output accuracy of the panoramic video.

Commonly used methods include the corner detection, ORB feature description detection, HOG feature description detection, sift algorithm, surf algorithm, etc.

(1) Corner detection

Corner is usually defined as the intersection of two sides. Strictly speaking, the local neighborhood of the corner should have two borders of different regions in different directions. In practical application, most of the so-called corner detection methods detect image points with specific features, rather than just "corners". Commonly used corner detection algorithms include Moravec corner detection algorithm, Harris corner detection algorithm, FAST corner detection

algorithm, LOCOCO corner detection method, etc.

(2) ORB feature detection algorithm

ORB uses FAST (features from accelerated segment test) algorithm to detect feature points. The core idea of FAST algorithm is to find out the unique point, that is, a point is compared with the points around it, and it can be considered as a feature point if it is different from most of the points.

(3) HOG feature detection method

HOG (Histogram of Oriented Gradient) feature is a feature descriptor used for object detection in computer vision and image processing. In this method, the feature is constructed through calculation and statistics of the gradient histogram of the local area of the image.

Its main idea is that in an image, the appearance and shape of a local target can be well described by the gradient or edge orientation density distribution. (The essence of this idea is to obtain the statistical information of the gradient, while the gradient mainly exists in the edge). The specific implementation of the method: First, the image is divided into small connected areas, which are called cell units. Then the gradient or edge orientation histograms of the pixels in the cell units are collected. Finally, these histograms are combined to form a feature descriptor.

(4) SIFT algorithm

SIFT (Scale-invariant feature transform) is an algorithm for detecting local features, in which the features and images are obtained and the image feature point matching is carried out by finding the interest points or corner points of an image and the descriptors of the scale and orientation.

The algorithm generally has five processes:

1. Build the scale space
2. Look for extreme points
3. Screen extreme points
4. Select the feature orientation
5. Construct the feature point description operator

SIFT feature not only has scale invariance, and a good detection effect can still be got even if the rotation angle, image brightness or shooting angle is changed.

(5) SURF algorithm

SIFT algorithm is relatively stable and can detect a large number of feature points, but its computational complexity is high, which is its largest disadvantage. Many scholars improved it and created new algorithms, in which a famous one is the SURF (speeded up robust feature) algorithm used in this paper. In the occasions requiring real-time operations, such as real-time target tracking system based on feature point matching, it is necessary to process images of 8-24 frames per second and complete the feature point search, feature vector generation, feature vector matching, target lock and so on. In this case, the SIFT algorithm is difficult to satisfy the requirement. SURF draws on the idea of simplified approximation in SIFT and simplifies the Gaussian second-order differential template in DoH, so that the image filtering of the template only requires a few simple addition and subtraction operations, and this operation is irrelevant to the scale of the filter. Experiments show that the SURF algorithm is about 3 times faster than the SIFT in operation speed.

The feature dimension of SIFT is generally 4 * 4 * 8 = 128D, and that of SURF is 4 * 4 * 4 = 64D. In the situation with high requirements for real-time performance like the real-time video mosaic, it is obvious that the SURF algorithm more conforms to the requirements.

5. Image Composition

In general, the image composition algorithm should meet two conditions:

1) The mosaic joints of two images should present smooth transition after the processing of image composition algorithm, and there should be no obvious mosaic traces.

2) After the image composition processing, the original image information should not be lost.

At present, the classical pixel-level image composition algorithm includes the averaging method and the weighted average image composition algorithm.

5.1. Averaging method

In general, the averaging method means that the pixel values of the matching image is directly superimposed in the overlapping area, and then the average value of the pixels after the superposition is taken as the final pixel value of this point. It is assumed that I1(x, y) is the pixel value of the image A to be mosaicked at the point (x, y); I2 (x, y) is the pixel value of the image B to be mosaicked at the point (x, y); and I(x, y) is the pixel value of point (x, y) after the composition. The averaging method is simple with fast operation speed, but the mosaicked image has a poor quality and is prone to have a relatively obvious mosaic trace.

5.2. Weighted average method

In general, the weighted average method is different from the averaging method. It is not a simple superposition of the pixel values in the overlapping area of images, but the average superposition after the introduction of certain weighting factors. It is now widely used. The cap weighted average algorithm is used in this paper. The principle of the cap weighted average algorithm is that the contribution of each pixel to the image pixel after composition is inversely proportional to the distance from the pixel to the center of the image within the overlap range of two images, so in this method, the weight value of the pixel close to the image center is large and the weight value of the pixel in the region away from the image center is small.

Gain Compensation:

Due to the difference in the position of the camera, the overall brightness between the images may be different. Generally, this deviation is handled by solving the overall brightness compensation coefficient of the image through the normalized gain intensity error *e* of the overlapping image pixels. The coefficient of the gain compensation is calculated by transforming the original RGB888 image into a single channel gray image, and then the gain compensation coefficient is used for the compensation of the same proportion in the original RGB channels, so that the brightness can be compensated without changing the color.

6. Experimental Results

In the experimental scene, the video recording and composition of the passing car on the overpass was performed. After two cameras were fixed, the video recording, composition and mosaic began. The algorithm described in this paper ran on the computer with the configuration of Intel Core i5 and memory of 4G in the VS2013 environment, and the images in the video capture card were the 1280 × 720RGB888 color images. The experimental results are shown in the figure. It can be seen that the mosaic has good real-time performance, the single-frame mosaic time is 70ms and 15 frames can be processed per second. Finally, there is no ghosting, the brightness difference between different cameras is well compensated and there is no obvious mosaic joint in the overlapping area of images. Thus, the purpose of video image stitching is achieved.

References

[1] Lv, Z., Halawani, A., Feng, S., Li, H., & Réhman, S. U. (2014). Multimodal hand and foot gesture interaction for handheld devices. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 11(1s), 10.

[2] Weisen Pan, Shizhan Chen, Zhiyong Feng. Automatic Clustering of Social Tag using Community Detection. Applied Mathematics & Information Sciences, 2013, 7(2): 675-681.

[3] Yingyue Zhang, Jennifer W. Chan, Alysha Moretti, and Kathryn E. Uhrich, Designing Polymers with Sugar-based Advantages for Bioactive Delivery Applications, Journal of Controlled Release, 2015, 219, 355-368.