2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

Real Time Continuous Image Stitching Algorithm Based on SIFT

YANG Rui-jun
School of Computer Science & Information Engineering
Shanghai Institute of Technology
Shanghai China
yangruijun@sit.edu.cn

ZHANG Chu
School of Computer Science & Information Engineering
Shanghai Institute of Technology
Shanghai China
sunnyzhang2020@163.com

CHENG Yan
School of Criminal Justice
East China University of Political Science and Law
Shanghai China
chengyan@ecupl.edu.cn

***Abstract*-Image stitching technology has application scenarios in many fields. This algorithm achieves the acquisition of simultaneously synthesized images, using the RGB module of the Intel Realsense D435 camera for image acquisition. Firstly, a raw image is created to store the final result, and one image is collected at 100ms intervals each time. Images with similarity less than 5/8 are taken and saved. Use the SIFT scale invariant feature detection algorithm to extract image feature points from the collected images, use RANSAC to extract feature points, use the random sample consistency algorithm to filter effective points, and calculate the homography transformation matrix. Synthesize every two collected images into one image and overlay it at the corresponding position in the resulting image. Through the experiment in this article, the average time for single acquisition and synthesis is 70ms, achieving the real-time goal. The similarity between the experimental group and the control group can reach 70%, and the resolution has been increased by 1.65 times, achieving the goal of continuous splicing.**

Keywords: image stitching; SIFT; RANSAC; real-time; continuous

I. INTRODUCTION

In recent years, image stitching technology has become a hot topic. Image stitching technology is the process of combining two or more images, based on one image, comparing other images with this image, and covering the same area to generate a new image. The process of image stitching is mainly divided into three steps: image matching [1], image transformation [2] and image synthesis [3].

The current commonly used technical methods usually use the method of collecting first and then processing. The synthesis results are not displayed together with the collection, so they are lacking in real-time and scalability.

This paper proposes a method that uses the SIFT algorithm for feature point extraction, uses the RANSAC algorithm to calculate the homography matrix transformation method for image transformation, and performs image synthesis after processing. On this basis, by creating a result map, the synthesized results are overlaid on the result map one by one. The result map is matched according to the size of the result of image synthesis. If it exceeds the range, it will be extended. In this way, continuous image synthesis is realized.

II. Correlation Algorithms

A. Feature Point Detection Algorithm

This paper uses the SIFT algorithm, which was published by Professor David Lowe in 1999 [4] and further improved in 2004 [5].

Using the SIFT algorithm, the scale space should first be constructed by using the Gaussian kernel function to blur the image, that is, to combine several local pixels into one pixel. The Gaussian function is:

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \qquad (1)$$

where σ is the standard deviation of the normal distribution, the calculation formula for the $m \times n$ convolution kernel is:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-0.5m)^2+(y-0.5n)^2}{2\sigma^2}} \qquad (2)$$

LOG is Laplacian of Gaussian, DOG is Difference of Gaussian, where DOG is used to approximate LOG, and DOG can provide more stable image features than LOG.

The SIFT algorithm solves the scale invariance through the scale space, and the expression of the scale space $L(x,y,\sigma)$ is:

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \qquad (3)$$

In the formula, $G(x,y,\sigma)$ is a Gaussian function; "*" is a convolution operation, and $I(x,y)$ is the original image.

The Gaussian pyramid is constructed by calculating the volume of the Gaussian kernel function $G(x,y,\sigma)$ and $I(x,y)$ under the constantly changing σ Product calculated.

$$\begin{aligned} D(x,y,\sigma) &= \big(G(x,y,k\sigma) - G(x,y,\sigma)\big) * I(x,y) \\ &= L(x,y,k\sigma) - L(x,y,\sigma) \end{aligned} \qquad (4)$$

The DOG pyramid is obtained by subtracting two adjacent images of the same group of images in the tower.

On the DOG pyramid, it is necessary to compare the point of 29 pixels need to be compared points, so as to ensure that extreme points can be detected in both image space and scale space. Only if a pixel in this layer is larger than all these adjacent pixels or smaller than all adjacent pixels, it will be selected as an extremum point.

B Image Stitching Method

This article uses a random sampling consensus algorithm proposed by Fischler and Bolls in 1981. The RANSAC algorithm randomly selects 4 matching pairs from the matching data set (these 4 matching pairs are required to be non-collinear), and obtains a homography matrix by calculating the matching data.

III. Algorithm implementation process

Collect two images successively, use the SIFT algorithm for key point matching, use the matched key points to calculate the transformation matrix through the RANSAC algorithm, perform image transformation through this transformation matrix, and synthesize the two images into one image.

A. Acquire two images in real time

In order to realize the real-time synthesis function, this paper closes the previous data stream every 100 milliseconds, stores the image *img1* obtained before closing, and then re-reads the camera data stream to obtain the image *img2*,

In practical applications, the camera may move slightly and should not continue to acquire new images at this time. This article uses the method of comparing image hash values to preserve images with a similarity of less than 5/8.

B. Synthesize two images into one image

Use the SIFT algorithm to obtain the key points *KP1* and *KP2* of two images. Using the *knnMatch* method, find the point m closest to the *img1* key point and the second closest point n in the *img2* key point, and match them together to form a match. Use the ratio of Euclidean distance to match the key points in m and n. When it is less than 0.5, the appropriate key points are retained, otherwise they will be discarded:

$$m.distance < minValue \times n.distance \quad (5)$$

Next, use the RANSAC method to calculate the transformation matrix *M*. Transform *img2* through *M* to obtain the synthesized *resultImg*.

C. Into the result image

When the camera moves to the left, it corresponds to the negative direction; when the camera moves to the right, corresponds to the positive direction. This movement value is superimposed in the positive direction. If it does not exceed the boundary, it is considered to be only moving to the right. If it exceeds the boundary, it is considered that the right side needs to increase the range.

IV. Experiment and Result Analysis

A. Exp data collection and processing

In this experiment, all images of the target are obtained by moving the camera horizontally. The composite target is a tiled image of 600mm*245mm.

This experiment uses a solid color background in which to place the target image. In the process of synthesizing the result image by means of translation, the key points collected by the SIFT algorithm are all on the target picture, and the captured or synthesized image is rectangularly cropped, and only the part of the target picture is kept for comparison.

Randomly select two of the images to match and synthesize the results as shown in Figure 1:

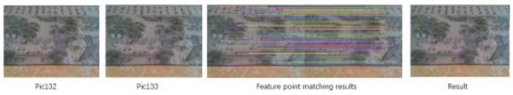

Fig.1 Experimental process

Directly shooting and collecting complete images as the result, therefore the results of this experiment are used as the control group. The experimental group completed the overall shooting through 55 images and performed the synthesis operation, as shown in Figure 2.

Fig.2 Control group(up) and experiment(down)

C. Analysis of results

The average duration of a single synthesis is 70ms, and the total duration of multiple synthesis varies linearly to achieve real-time synthesis goals.

For the post-experimental results, use 16*16 hash operations to calculate the similarity with the first set of results, that is, the original image, and compare the resolution change ratio.

The algorithm proposed in this paper has a similarity of 70% between the synthetic result and the original image. The resolution is also increased from the original 629*263 to the highest 1113*448, the length is increased to 1.76 times, and the width is increased to 1.70 times.

REFERENCES

[1] M. Brown, D.G. Lowe, Automatic panoramic image stitching using invariant features[J]. Int. J. Comput Vis. 2007, 74 (1):59–73.
[2] S.Peleg, J.Herman, Panoramic mosaics by manifold projection, Computer Vision and Pattern Recognition[C]// 1997. Proceedings., 1997 IEEE Computer Society Conference on. 1997:338–343.
[3] M.-S. Su, W.-L. Hwang, K.-Y. Cheng, Analysis on multiresolution mosaic images[C]// IEEE Trans. Image Process. 2004, 13 (7): 952–959. 2022,48(8):159-162
[4] Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.
[5] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981,24(6):381–395