Building Recognition System Based on Improved SIFT Algorithm and Positioning Information

Tao Liu[1], Feng Jiang[1], Yu Gao[2]

[1] College of Electronic and Information, Southwest Minzu University, Sichuan Chengdu, China
[2] The Australian National University, Canberra, Australia
Email: 2268979210@qq.com

Abstract. In order to solve the general problem, that is, the accurate recognition rate is low in a small extent or when the image resources are few and scattered. This article puts forward a building recognition system that combines GPS positioning information with an improved SIFT algorithm, and adds a pre-processing mechanism to predict the possibility of the building existence in the system, which further reduces mismatch and improves response speed. The final verification shows that this research is actually effective.

Keywords: building recognition; SIFT feature; straight line detection; Canny edge detection; feature matching

1 Introduction

Image processing and computer vision began to rise rapidly, and the recognition of image information is one of the research hotspots. The way users get information is no longer just staying in words. People prefer to get information directly from the image itself [1]. Building recognition is also one of the most population research directions. The traditional recognition method only recognizes the characteristics of the building itself or uses GPS for positioning. However, when only one processing technique is used in a small area or a small area with data, its accuracy and rapidity are difficult to be guaranteed. In response to these issues, this research will add a pre-processing mechanism to predict the existence of buildings. This article puts forward a building recognition system that combines GPS positioning information with an improved SIFT algorithm.

For decades, massive eminent researchers have promoted the development of this field and have put forward plenty of effective methods. In 2012, Songlin Li proposed a city building recognition system based on feature line matching, which can well adapt to the building conditions to quickly identify the requirements of the mobile devices, but its accuracy rate is not high[2]. In 2015, Xingquan Cai proposed a building recognition system based on the combination of GPS matching and SIFT feature matching to cope with the difficulties of traditional recognition methods in both response speed and recognition efficiency. There are also good results on mobile devices [3-4]. In 2019, Xintong Liu et al. used variance to generate grayscale images and reduce the influence of background areas. But it did not reduce the impact of mismatch points, and the effect was universal [5].

2 Building Identification System

In order to solve the general problem, that is, in a small extent or with few and scattered image resources, it is unable to efficiently and accurately identify. The innovative point of the building recognition system proposed in this paper is to add the pre-screening operation of the buildings to further reduce the waste of computing resources. As shown in Figure 1, this system is mainly parted into four modules, namely: pre-processing module, feature extraction module, database creation module, and feature matching module.

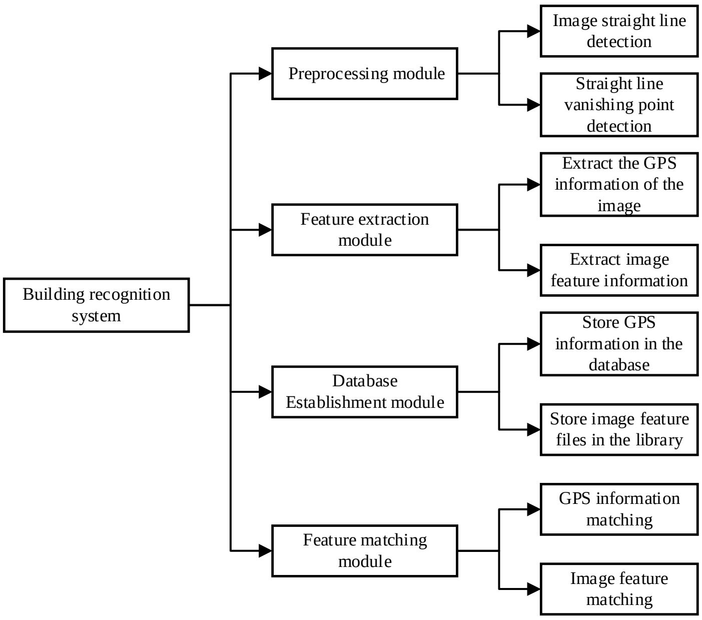

Building recognition system
Preprocessing module
Image straight line detection
Straight line vanishing point detection
Feature extraction module
Extract the GPS information of the image
Extract image feature information
Database Establishment module
Store GPS information in the database
Store image feature files in the library
Feature matching module
GPS information matching
Image feature matching

Figure 1. Structure Diagram of Building Recognition System

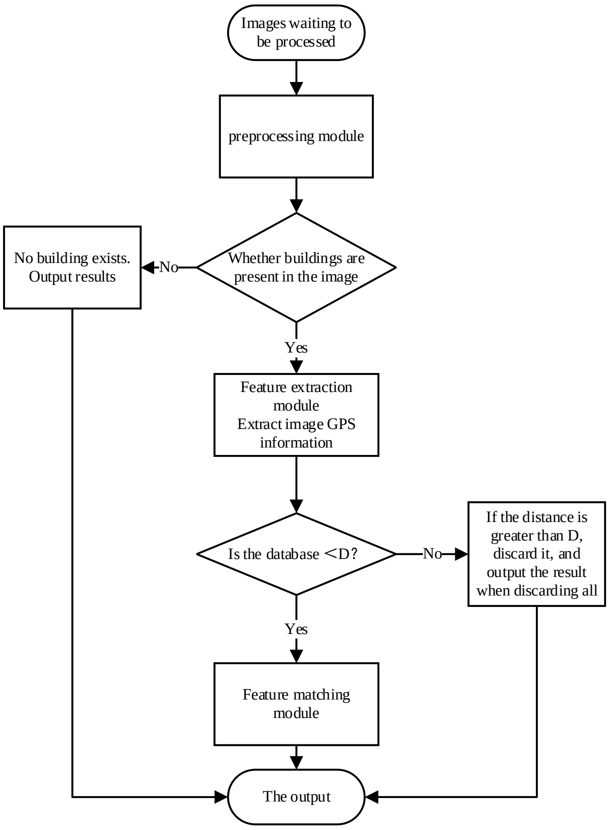

Images waiting to be processed
preprocessing module
Whether buildings are present in the image
No
No building exists. Output results
Yes
Feature extraction module
Extract image GPS information
Is the database <D?
No
If the distance is greater than D, discard it, and output the result when discarding all
Yes
Feature matching module
The output

Figure 2. System Logic Flow Chart

2.1 Preprocessing Module

At this stage, many building recognition systems are mainly based on image retrieval (CBIR) [6], and there is no pre-judgment whether there are buildings in the image to be recognized. Compared with natural objects, the structure of buildings is more regular, and there may be traces to follow in related images, such as the intersection of multiple straight lines, specific structural patterns, and so on. The SIFT algorithm is used in this study to detect the artificial structures in the image to be identified.

Due to the effect of noise on the image, smoothing processing is used first, and followed by edge detection. After the detection, the Hough transform is performed to get the line parameters and determine the quantity. The points on the same straight line in space coordinates are mapped to another parameter space to define the curve, and they all intersect at one point. But usually this intersection will be around a certain point (x, y). Because of those outliers, there may be some error lines, which will definitely affect the vanishing point detection in the next step. This paper finds the vanishing point clusters and calculates the distance D from these points to one of them. Set a comparison value T. When $D \leq T$, they are adjacent-points, if the number exceeds a certain set value, it is regarded as a vanishing point.

On top of this, a pre-judgment operation is performed on whether there are buildings in the entire image, instead of directly matching features, which will reduce operations of irrelevant images, thereby improving the accuracy and efficiency of the system.

2.2 Feature Extraction Module

It will be affected by external factors when acquiring images. Not every corresponding image is a standard size. Therefore, the size of the advanced image to be recognized is reduced. This article uses bilinear interpolation to reduce response time. By extracting GPS information, narrowing the matching range of the image to be recognized; then the subsequent feature extraction operation uses the optimized SIFT algorithm. Saving the extracted description operator to a feature file for subsequent processing.

The SIFT algorithm is mainly to extract the regional feature points of the object to be recognized. It has nothing to do with angle, size and other elements, and has a high tolerance for noise [7]. However, traditional SIFT has the problem of feature points mismatch in building recognition. The increase in the existence of irrelevant points will affect the action time and system exactness, and the 128-dimensional description operator will make it have higher time and space complexity.

The purpose of this research is to solve the above mentioned problems, improve algorithm performance and shorten the response time without affecting the robustness of the original algorithm.

2.2.1 Component Gaussian difference pyramid

$$L(x, y, \sigma) = G(x, y, \sigma)^{*} I(x, y) \tag{1}$$

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

Among them, $\boldsymbol{L(x, y, \sigma)}$ is the Gaussian scale-space model, $\boldsymbol{G(x, y, \sigma)}$ is the scale change Gaussian function, $\boldsymbol{I(x, y)}$ is the pixel gray value, and $*$ is the convolution operation.

2.2.2 Positioning of key points

The preliminary screening of key points refers to the detection of spatial extreme points, because key points are composed of local extreme points in DOG space. Each detection point will be compared with 26 points existing in 3 layers (including this layer). This layer has 8 pixels and the adjacent layer has 9 pixels. The purpose of this operation is to ensure that extreme points can be obtained. At this time, the detected extreme points are in discrete space, because some extreme points have a weak response, not all feature points are stable. Therefore, it is necessary to perform further determinations by fitting a three-dimensional quadratic function.

2.2.3 Calculation of the main direction of feature points

Here, for the sake of eliminating the impact of image rotation and other operations on the matching, it is necessary to calculate the gradient value and gradient direction. Construct a histogram of neighborhood gradient directions, in the end assign a reference direction to each feature point. The purpose is to ensure the rotation invariance of the descriptor.

$$m(x, y) = \sqrt{\left(L(x+1, y) - L(x-1, y)\right)^2 + \left(L(x, y+1) - L(x, y-1)\right)^2} \tag{3}$$

$$\theta(x, y) = tan^{-1}((L(x+1, y) - L(x-1, y)/L(x, y+1) - L(x, y-1))) \tag{4}$$

The above formulas (3) and (4) are used to calculate the gradient value and gradient direction, where L represents the value of the scale space where the key point is located.

2.2.4 Generate feature descriptor

After completing the above three parts, the next is to use the information that has been obtained to generate description operators. Based on experiments, this research divides the feature point area into 3×3 sub-regions, and a seed point is a sub-region. Each feature point will have $3 \times 3 \times 8 = 72$ data, which constitutes a 72-dimensional SIFT feature vector after dimensionality reduction.

The original SIFT algorithm directly performs grayscale processing on graphics in RGB mode. This article uses the laboratory mode that defines the most colors, which has nothing to do with lighting and equipment, and is as fast as the RGB mode in terms of processing speed. This mode is different from the traditional RGB mode, it is composed of the bright channel L and the other two color channels. Among them, a means the channel that the color gradient from dark green with a low brightness value to pink with a high brightness value; b means the channel that the color changes from dark blue to yellow. Compared with the RGB mode, the color range represented by this mode is wider.

After that, calculate the expectation and variance of each channel. For images with inconspicuous color contrast, this paper introduces the parameter R to adjust the grayscale image. The formula (6) is the gray scale calculation method.

$$S^2(j) = (P_j(L) - E(L))^2 + (P_j(a) - E(a))^2 + (P_j(b) - E(b))^2 \tag{5}$$

$$Gray(j) = \{S^2(j)\}^R \tag{6}$$

Among them, $Gray(j)$ represents the gray value of the pixel, and R is the impact factor. Use R to change the gray value to adjust the degree of discrimination so that the retained effective building information is relatively optimal. After actual testing, the effect of R is significant when the value of $R = 1.5$, and it can also be adjusted between 0.8 and 1.8 as needed. As the degree of discrimination increases, the corresponding buildings in the image will become more obvious.

After that, the original SIFT algorithm using Euclidean distance has been further improved. The original author's idea is that each feature description operator has 128-dimensional information. Then obtain the Euclidean distance between the vectors, and complete the matching through comparison [3]. This article uses Mahalanobis distance as a metric to judge whether the feature points are close to each other. The Mahalanobis distance considers the relationship between various characteristics and has nothing to do with the measurement scale. It further improves the comparison speed.

If the vector samples is $X_1 \sim X_n$, S is the corresponding covariance matrix, and u is the mean value. The Mahalanobis distance between the vector X and u is obtained by formula (7).

$$\boldsymbol{D(X) = \sqrt{(X-u)^T S^{-1}(X-u)}} \tag{7}$$

If there are any two samples $X_i \sim X_j$, as shown in formula (8):

$$D(\boldsymbol{X_i}, \boldsymbol{X_j}) = \sqrt{(\boldsymbol{X_i} - \boldsymbol{X_j})^T \boldsymbol{S}^{-1} (\boldsymbol{X_i} - \boldsymbol{X_j})} \tag{8}$$

Compared with the original algorithm, the feature extraction of building area has been improved, but there will still be some mismatch points when there are a large number of similar structures. On this basis, the RANSAC algorithm is added to improve performance. This algorithm can train the optimal parameter model from a set of observation data containing “irrelevant points" through iterative training. The models that do not meet the optimal parameters are defined as "outliers".

In OpenCV, RANSAC gets the best 3×3 homography matrix to filter out false matches. RANSAC needs to obtain the optimal parameter matrix to make the most data points that satisfy the matrix [8].

$$S\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{9}$$

In the above formula (9), (x, y) (x', y') are the corner points of the comparison image and input image respectively, and S is the scale parameter. The algorithm randomly selects four non-collinear samples from the matched data set, then calculates its homography matrix, then uses this model to test all the data, and calculates the number of data points and the projection error under this model (that is, the cost function). If this model is the optimal model, the corresponding value is the minimum value. The calculation cost function is formula (10).

$$\sum_{i=1}^{n}\left(x'_i\frac{h_{11}x_i+h_{12}y_i+h_{13}}{h_{31}x_i+h_{32}y_i+h_{33}}\right)^2+\left(y'_i\frac{h_{21}x_i+h_{22}y_i+h_{23}}{h_{31}x_i+h_{32}y_i+h_{33}}\right)^2 \tag{10}$$

2.3 Database Establishment Module

During the recognition operation, the corresponding image feature database must be created first. This article is designed for a small area. Areas with poor field of view and signal, and fewer image resources. Therefore, the establishment of a database is to search for the target image resources. Taking the Wuhou campus of Southwest Minzu University as an example, the author took pictures of the school buildings with GPS information from different angles, different times, and different lighting equipment to create a database. The GPS information of the above building photos obtained under different conditions and saved in the MySQL database is extracted, then change the image size and obtain the feature information, and then save it to the feature file.

2.4 Feature Matching Module

After judging by the preprocessing module, if there are buildings in the image, it will enter the feature matching module for final comparison. This module includes GPS information matching and feature matching.

First, the GPS information of the image to be recognized is matched with the information in the database established above. In the WGS-1984 coordinate system commonly used in my country, longitude $1'' \approx 23.6$ meters, latitude $1'' \approx 30.9$ meters. Calculate the distance between the position information of the image to be recognized and all the position information in the database. If the distance is less than 30 meters, that is, when it changes in seconds, the corresponding picture in the library will be used as a candidate image. Otherwise, it will be discarded. This step will ultimately improve the efficiency of feature matching. Finally, the to-be-recognized image is matched with the candidate image formed after the previous operation. The improved SIFT algorithm is adopted here to complete the matching and output the final matching result.

Theoretically, the influence of the actual topography is ignored, and the Earth can be seen as a theoretical sphere. In this case, the average radius of this point is $R = 6371.004$ km, and the distance between two points can be obtained by latitude and longitude. Suppose A and B are $(LonA, LatA)$, $(LonB, LatB)$ respectively. At this time, the 0 degree longitude is the benchmark, the east longitude is positive and the west longitude is negative.

In practice, latitude does not take into account positive or negative values. Formula (11) is the calculation formula for the distance between A and B.

$$C = sin(LatA)\,sin(LatB) + cos(LatA)\,cos(LatB)\,cos(NLonA - NLonB) \tag{11}$$

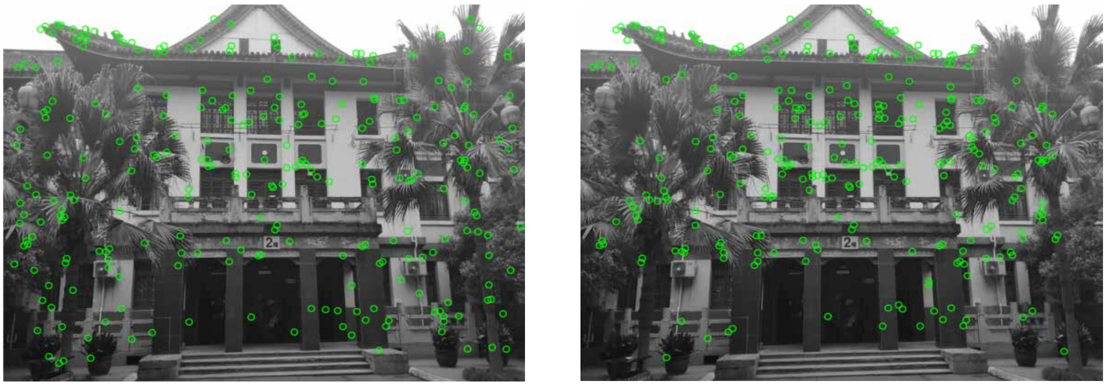
$$Distance = \frac{R \times arccos(C)Pi}{180} \tag{12}$$

3 Experiment and Summary

From the comparison of Figure 3, it can be seen that the original algorithm is relatively scattered in the extraction of feature points and has a weak processing capacity for environmental interference in the image. The improved algorithm has a better recognition effect on the building area，and feature points are more concentrated in the target area.

From the experimental results in Figure 4, it can be shown that the original algorithm has a lot of mismatches in the matching process, and it does not deal with the occlusion phenomenon well. But the improved algorithm shows more ideal results, and the matching effect on partial images is even better.

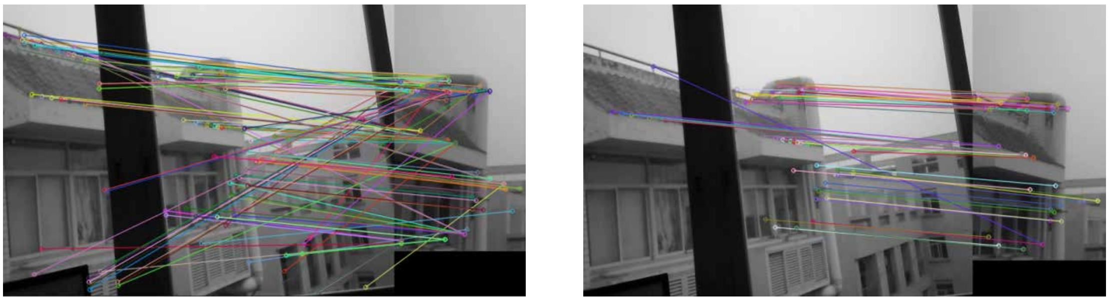
Aiming at the problem of insufficient accuracy of traditional building recognition methods in small areas, areas where the field of view and signal are poor. After the above optimization, the matching efficiency can be effectively improved and reduce the existence of mismatch points. This research greatly improves the performance of the system after further optimizing the operation. Through the actual tests, the building identification system proposed in this study has a better effect, but it needs to be improved in preprocessing to avoid the influence of irrelevant objects.



Original Algorithm Feature Points

Improved Feature Points

Figure 3. Comparison of Extracted Feature Points



Original Algorithm Matching Result

Improved Matching Results

Figure 4. Algorithm Improvement Comparison

References

1. Zehong Wang,Houquan Liu. Building recognition based on migration learning and adaptive feature fusion[J]. Computer Technology and Development, 2019, 29(12): 40-43.
2. Songlin Li,Haisheng Fan,Xiuwan Chen.Research on Urban Building Recognition Method Based on Feature Line Matching[J].Remote Sensing Technology and Application,2012, 27(002):190-196.
3. Jinghua Liu. Research on building image recognition technology for mobile applications [D]. North China University of Technology, 2015.
4. Xingquan Cai, Jinghua Liu. Design and implementation of building image recognition system[J]. Modern Computer (Professional Edition), 2015(14):18-20.
5. Xintong Liu,Hui Zhang,Zhiqiang Deng. Building image recognition based on SIFT algorithm. [J]. Science and Fortune, 2019; 350.
6. Tingfa Yan. Specific building recognition algorithm based on image sequence[J]. Journal of Taishan University, 2018,v.40;No.205(03): 68-72.
7. Lowe D G .Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.
8. Xu Xiaobin et al. LiDAR–camera calibration method based on ranging statistical characteristics and improved RANSAC algorithm[J]. Robotics and Autonomous Systems, 2021, 141(prepublish): 103776-.