Study on the traditional pattern retrieval method of minorities in Gansu province

Gang Zheng\textsuperscript{1,2,*}, Beizhan Wang\textsuperscript{2}, Yuchun Sun\textsuperscript{2}, Jin Xu\textsuperscript{1}

\textsuperscript{1}Lanzhou Institute of Technology, Lanzhou Gansu, 730050 China; \\
\textsuperscript{2}Xiamen University, Xiamen Fujian, 361005 China;

*Corresponding author e-mail: 32028095@qq.com

Abstract. The traditional patterns of ethnic minorities in Gansu province are ethnic arts with strong ethnic characteristics. It is the crystallization of the hard work and wisdom of minority nationalities in Gansu province. Unique traditional patterns of ethnic minorities in Gansu province with rich ethnic folk arts, is the crystallization of geographical environment in Gansu minority diligence and wisdom. By using the Surf feature point identification algorithm, the feature point extractor in OpenCV is used to extract the feature points. And the feature points are applied to compare the pattern features to find patterns similar to the artistic features. The application of this method can quickly or efficiently extract pattern information in a database.

1. Introduction

With the development of the digitization process of information technology and culture, the traditional data access and sharing methods are changed. In Gansu province, the traditional patterns of ethnic minorities are drawn by hand and cannot be combined with digital tools. In Gansu province, the traditional patterns of ethnic minorities are drawn by hand and cannot be combined with digital tools. These factors restrict the transmission and inheritance of traditional cultural resources of the unique ethnic minorities in Gansu province, which has influenced the wide application of traditional culture of minority nationalities. Therefore, it is very urgent and important to establish the traditional pattern information database and the digitization of cultural resources in Gansu province. An important task in constructing the digital platform of minority cultural resources in Gansu is the retrieval function of the pattern and the realization of the algorithm. This function can be more convenient for minority research enthusiasts to find relevant pattern information more easily, which increases the convenience of research and saves a lot of time. The introduction of graph sense algorithm provides a theoretical basis for the construction of digital retrieval technology of traditional patterns in the future. According to the minority pattern information database needs, in order to ensure the real-time and accuracy of the map in map function, image retrieval using surf feature retrieval algorithm in the system, and by comparing the feature points to retrieve similar images.

2. The key technology

2.1. SURF algorithm

SURF algorithm is a robust image recognition and description algorithm, first published in 2006 in the ECCV conference\cite{1}. SURF algorithm is proposed by Bay and other people, mainly for object recognition, image positioning and classification, 3D reconstruction and other tasks. It’s inspiration
mainly comes from the SIFT algorithm, and USES a more efficient way to complete the feature extraction and description. The SURF standard version algorithm is several times faster than the SIFT algorithm, and its author claims that it is more stable and powerful than SIFT in different image transformations[2].

2.2. OpenCV technology
OpenCV is a function library for digital image processing and computer vision, developed by visual interactive group of Intel Microprocessor Research laboratory. OpenCV consists of a series of C functions and a small number of C++ classes, and implements many general algorithms in image processing and computer vision[3]. OpenCV also provides a transparent interface for Intel's Integrated Performance Primitives (IPP). This means that OpenCV will automatically load these libraries at run time if there is a IPP library optimized for Intel processors. Just call the basic function library of OpenCV and add the program written by yourself, and then you can complete the complex and huge development task, which can achieve twice the result with half the effort.

3. Algorithm design and Implementation
According to the requirements of the project, in order to ensure the real-time and accuracy of the image drawing function, the SURF feature point retrieval algorithm is adopted in this project, and the similar patterns are searched through the comparison of feature points.

3.1. Data pretreatment
Data preprocessing consists of two parts: The first part is, to read program patterns in the zoom operation, because if each pattern is not the same size, the number of feature points extracted by then will be particularly big difference, sometimes even up to two orders of magnitude, it will have a very negative impact on the feature point matching algorithm, so the pattern all zoom to the same size of the operation can be a good solution to this problem. The scaling scale greatly affects the time of feature extraction algorithm. In order to give users better application experience and improve the practicability of the algorithm, all the patterns are uniformly scaled to 256 * 256 pixels.

The second part is the gray operation of all patterns, because the data used in this test are from the traditional Chinese patterns, and all of them are made up of black and white. But in actual application of this program, often inserted patterns are not all black and white, and sometimes colored patterns. But because of the use of matching feature points is based on the pattern shape, therefore the color information does not cause too much impact on the pattern, but in order to avoid the bias, the system will all the image to gray image, in order to increase the accuracy of the system.

3.2. Establishment of characteristic point matrix.
The feature extraction function, mainly used in the OpenCV SURFFeatureDetector and SURFDescriptorExtractor. When the system works, the SURFFeatureDetector method of detect is applied to each pattern of the data set, and the SURF feature points are searched for, and the detected feature points are shown in figure 1:

![Fig.1 Characteristic points in an image](image)

There are many circles in Figure 1. The center of the circle is the detected feature point. The length of the radius indicates the scale at which the feature points are located. The radius in the circle
represents the direction of the feature points. Applying SURFDescriptorExtractor's compute method, these feature points are transformed into a matrix form of 1 * 64 and stored in a large feature matrix called matTotalDesc.

The matTotalDesc is a collection of all the feature points of all the patterns in the dataset, a 64 dimensional matrix, whose rows are the total number of all feature points. To distinguish the feature points of each pattern from the position of the feature point matrix matTotalDesc, you need to define a structured variable, matDescToImgfile, which is shown in figure 2:

```
struct matDescToImgfile{
    int i_begin_index ;
    int i_end_index ;
    string s_img_file_name ;
};
```

**Fig.2** Code for structure variables

Which contains 3 variables, s_img_file_name is the name of the image, i_begin_index feature points of the image in the feature matrix in the starting position (line number), i_end_index is the final position of the feature points of the image in the feature points in the matrix (line number). Each pattern corresponds to such a structural variable, echoing the feature points of the pattern.

In the establishment of the structure of the above variables, you can through a cycle of every piece of patterns in the library all the feature point information stored in the matTotalDesc matrix, so that subsequent treatment cycle related code as shown in figure 3:

```
for (imageName = trainAllFileName.begin() ;
     imageName != trainAllFileName.end() ; imageName++) {
    vec<keypoints> vecKps;
    matDesc.release();
    Mat imgMat;
    imdecode(trainDirName+"\"+imageName,CV_LOAD_IMAGE_GRAYSCALE) ;
    imwrite(imgMat,Resize(256,256));
    cvcvtColor(imgMat,Mat,CV_BGR2GRAY);
    m_pExtractor->Detect(Mat,vecKps);
    if(vecKps.size() == 0) {
        return ;
    }
    for (size_t i = 0 ; i < vecKps.size() ; i++) {
        matDesc.push_back(matDescToImgfile{ i_begin_index , i_end_index , imageName });
    }
}
```

**Fig.3** Part of the code for the establishment of feature points matrix

Through such a series of processing, the matrix of all the feature points corresponding to all the patterns in the pattern library can be obtained, and the location of the feature points of each pattern in the feature point matrix can be obtained.

### 3.3. Pattern retrieval

FLANN library is a very perfect nearest neighbor open source library, which contains a variety of search algorithms, and this system is applied to the library in OpenCV. The index class in FLANN is used to establish an index for the feature point matrix, and the index needs two parameters. The first parameter is the index constructed matrix, and the second parameter is the algorithm of constructing index. At present, the commonly used algorithms are LinearIndexParams, KTreeIndexParams, KMeansIndexParams, CompositeIndexParams.
In this system, we select the feature matrix matTotalDesc of the data set to construct the KD tree index. KD tree index contains the knnSearch method used to implement KNN algorithm, the feature point sample pattern set is input to the knnSearch method. This method is to find the matching points in the feature matrix in matTotalDesc index stored in the variable Indices. The key code is shown in figure 4:

```cpp
m_pDetector->detect(QueryMac, QueryVecKF);
compute(QueryMat, QueryVecKF, QueryMatDesc);
num = QueryMatDesc.size();

for(int i = 0; i < num; i++)
{
    key = *it_indices[i];

    for (oneDescLog = Desc_Imgfile.begin(); oneDescLog != Desc_Imgfile.end(); oneDescLog++)
    {
        if (key == oneDescLog->p1.begin_index) // Key <= oneDescLog->p1.end_index
            flag_count[1] ++; //Flag ++
    }
}
```

**Fig.4** Feature point sample pattern extraction part of the code

In Figure 4, we can see that in the project, 10 feature points similar to the sample pattern are selected in the pattern library by using the KNN algorithm. The 10 feature points are the first 10 closest to the feature points of the exemplar pattern. Compared with retaining only the nearest feature points, selecting the closest 10 feature points increases the robustness of the system. For example, the user wants to display 10 pieces of similar patterns, and if a similar feature points only algorithm, then when the pattern in a library of two pictures and sample patterns are very similar, even contains all the feature points sample pattern, then the similar characteristics high number of these two patterns, and other patterns the similar number of feature points is zero, even though the other pattern also has the similar pattern, the program is not good to find out. The 10 points of similarity can be used to solve this problem. Then, according to the indices and the structure variable matDescToImgfile for each pattern, the number of feature points of each pattern in the dataset is similar to the sample pattern, and the related code is shown in figure 5.

```cpp
for (it_indices = Indices.begin(); it_indices != Indices.end(); it_indices++)
{
    key = *it_indices;

    for (oneDescLog = Desc_Imgfile.begin(); oneDescLog != Desc_Imgfile.end(); oneDescLog++)
    {
        if (key == oneDescLog->p1.begin_index)
            flag_count[1] ++; //Flag ++
    }
}
```

**Fig.5** Key code in the database pattern similar to feature point statistics

At the end of the pattern in the library is sorted according to the pattern similar to the number of feature points and feature points matching more, think of two picture more similar, and according to the needs of users, the highest level of N showed a similar pattern. The result of pattern feature matching is shown in figure 6. There are many lines in Figure 6, and the two ends of the line represent two matching feature points. There are a large number of matching feature points in the two images, which means that the two patterns are highly similar.

**Fig.6** Feature point matching results
4. Interface display
In order to increase the availability and practicability of the algorithm is convenient, minority patterns and patterns of fans and collectors, the project with the development of a Qt interface, the user can use the local computer to the application of the design pattern library retrieval algorithm, to achieve local pattern search function. The program interface is shown in Figure 7 and figure 8:

![Fig.7 program interface](image)

![Fig.8 search results](image)

The interface is all made of Qt drag and drop components, which are divided into two pattern display boxes, three buttons and a text display box. Through the three buttons set the corresponding click function to operate the program, and display the corresponding results. The user can click on the open button to select the desired pattern, the search box will display the selected sample pattern, the pattern of the left pattern display, text display box will display the name and path of the selected pattern, then click on the button to select the test set for a test set of directory, selected text box displays the path directory. Finally, click on the search button to start, start the search algorithm, completed in the bar will display the search and sample pattern the most similar pattern on the right side of the pattern display, and in the text below shows the path and name displayed similar degree before the 5 patterns in the column.

5. Conclusion
A pattern retrieval method based on SURF algorithm is proposed in this paper. Because the system needs a lot of pattern processing operations, the OpenCV image processing function library is used to improve the programming efficiency, and finally a set of pattern retrieval system is realized. The system obtains the pattern with higher similarity by comparing the patterns in the pattern library. The experimental results show that the recognition has good application of this algorithm in the field of image retrieval, it has low complexity, fast and good robustness, can further enhance the user experience and pattern search speed in the future. For example, the user can be only a pattern library, you can write a record of the pattern library in each image and its corresponding feature points, when the picture library content change, only slight modifications to the record can be, without the need for every time when you run the program to re read all the data and to find the feature points, save a lot of time.

Acknowledgement
The work was supported by Projects of the National Social Science Foundation of China under Grant No.15xmz035.National Natural Science Foundation of China under Grant No.61462057. Lanzhou Science and Technology Development Project under Grant No.2015-3-30. National Science Foundation project under grant NO.61462057. The Lanzhou special project of strategic new industry technology No.2013-4-23. China postdoctoral science foundation granted No:2015M582722.

References
[1] Bay H,Tuytelaars T,Van Gool L. SURF: Speeded up robust features[J]. Computer vision–ECCV 2006, 2006: 404-417.
[2] Bay Herbert, Tuytelaars Tinne, Gool Luc Van. SURF: Speeded Up Robust Features[J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.

[3] Adrian Kaehler & Dr. Gary Rost Bradski (2008), Learning OpenCV, New York: oreilly