

# Feature-based Aerial Image Discovery System with Multi-Scale City Image Database

Ma, bingjie

---

(出版者 / Publisher)

法政大学大学院情報科学研究科

(雑誌名 / Journal or Publication Title)

法政大学大学院紀要. 情報科学研究科編 / 法政大学大学院紀要. 情報科学研究科編

(巻 / Volume)

12

(開始ページ / Start Page)

1

(終了ページ / End Page)

6

(発行年 / Year)

2017-03-31

(URL)

<https://doi.org/10.15002/00014027>

# Feature-based Aerial Image Discovery System with Multi-Scale City Image Database

Bingjie Ma

Graduate School of Computer and Information Sciences

Hosei University

Tokyo 184-8584, Japan

bingjie.ma.34@stu.hosei.ac.jp

**Abstract**—In this paper, we propose a new method for achieving city aerial image discovery based on texture feature calculated using GLCM combined with SIFT algorithm. This algorithm uses gray-level co-occurrence matrix for pre-selecting search candidates, and LSH forest to process the pre-selected result. Then the scale invariant feature transform algorithm, homography and the random sample consensus algorithm are used to obtain the final matching result. Experimental results demonstrate that these sets of algorithms can greatly reduce the discovery time while meeting the requirements of precision and robustness.

**Keywords**—aerial image discovery; city image; GLCM; SIFT; LSH forest

## I. INTRODUCTION

Aerial image discovery is one of the key technologies in various services using geographic information system (GIS), in which a match is to be identified in a large image database with a given query image with unknown location, orientation, and scale. The task is more challenging in city areas because the views have higher degree of homogeneity compared to non-urban areas. It will be great help if we can extract the useful information from complicated aerial images and apply it to disaster relief and other aspects.

To achieve aerial image discovery, especially in city area, there are three main requirements need to meet: accuracy, efficiency and robustness. Although there have been various research attempts made in this area, the performance of automated discovery is not sufficiently high.

There is a high similarity among the features of city area in remote sensing images. A large number of duplicate streets and residential areas determine that the usual content-based image retrieval (CBIR) system [1] is inapplicable to city sub area localization. The currently CBIR systems are mostly built to help users find similar or same images without multi-scale or rotation. And the global feature is used to achieve image retrieval, most of which changes with the change of image's size, angle or brightness.

Scale invariant feature transform (SIFT) [2] algorithm has shown to be one of the best approaches of aerial image match and registration because of its high accuracy and robust. But few of researchers use it to do image retrieval cause it need pretty big calculation and takes too long time to get the result with large database.

To solve this complex problem, I propose a new method combining the global feature and local feature, which is robust and can greatly reduce the cost time while has high precision.

The remainder of this paper is organized as follows. The second section describes related work about content-based image retrieval, texture feature and local feature. The third section makes introduction of the algorithms that will be used in this system, including gray-level co-occurrence matrix, local sensitive hash forest and sift. The fourth section describes details of system implementation, such as system overview, working flow, database structure and so on. Last section analyses the experimental results and the last section concludes this paper and outline future research.

## II. RELATED WORK

Research related to image discovery so far mostly focuses on content-based approaches using color, texture and shape features. Liu Zhen and Zhou Shuqiu have proposed a method using color information and texture features obtained from Fourier transform in image retrieval by using colors and texture features in images [3]. The system can output several results similar to the input images.

In aerial image discovery field, Antonios Gasteratos et al. proposed a fuzzy aerial image retrieval system based on texture energy combined with color in an intelligent system for aerial image retrieval and classification [4]. Their experimental results show color information plays dominant role in the image retrieval process.

But aerial image match is easily affected by scale, rotation, noise and illumination. Color and texture features lack sufficient content expression ability for aerial images. The streets and buildings in different areas may have same color or texture histogram. In aerial image discovery, there are so many similarity images, if we want to find which area is the input image belonged to, color and texture feature is not enough to discover result accurately. Yan Ke et al. who proposed PCA-SIFT to reduce the calculation, apply PCA-SIFT to build a sub-image retrieval system [5]. They acclaimed that this system can satisfy the requirements of high recall and precision in efficient near-duplicate detection and sub-image retrieval. According to Image localization in satellite imagery with feature-based indexing, Wu Changchang et al. presented a method to do query image localization which has unknown scales and rotations [6]. The paper applied visual word combined SIFT features to make localization and tried GPU-based SIFT implementation to speed up. But the

recall of their experiments result using Google color maps is only 13%.

Gray-level co-occurrence matrix (GLCM) is put forward by Mryka [7]. Although it is not published by a professional journal, it shows good performance in texture analysis. In GLCM Textural Features for Brain Tumor Classification, Nitish Zulpe and Vrushesen Pawar are using GLCM textural features to do brain tumor classification, which has achieved 100% accuracy [8].

Local sensitive hash (LSH) can reduce the complexity of time and space while ensure a degree of accurate and it an support high-dimensional data retrieval. In 1998, P.Indy and R.Motwani proposed the theoretical basis of the LSH algorithm. In 2005, Mayank Bawa etc. put forward the LSH forest algorithm, which uses tree structure to replace of hash table and has the ability of self-parameters correcting [9].

In addition, most of the previous works have emphasis on images in non-urban areas. Thus, we need to establish a method to attain match accuracy, high degree of robustness, and computational speed, especially for city image database.

### III. PROPOSED APPROACH

#### A. Overview of the proposed approach

To achieve the target of aerial image discovery, I propose a robust approach can meet both accuracy and efficient, using the GLCM tutorial, LSH forest to do coarse matching [10] and SIFT algorithm comprehensively. The procedure of the algorithm is described in figure 1.

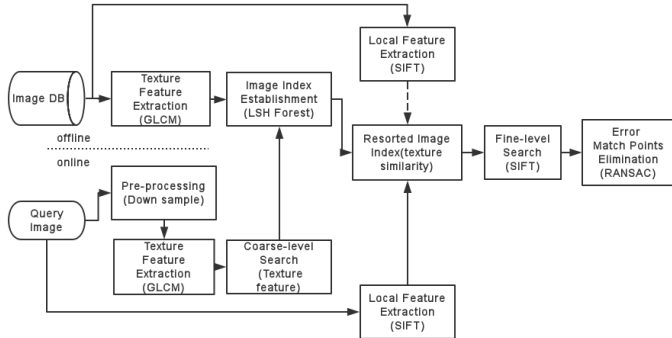


Fig. 1. Procedure of Proposed Approach

#### B. Texture Feature

##### 1) Grey-level Co-occurrence Matrix

Texture is the similar elements or graph structure with regularity, generally understood as the gray scale changes and repetition in space or the local mode recurred (texture units) and their regular arrangement.

Gray-level Co-occurrence Matrix (GLCM) is a tabulation of how often different combinations of pixel brightness values (gray levels) occur in an image. GLCM texture considers the relation between two pixels at a time, called the reference and the neighbor pixel.

GLCM provides the information of image gray scale direction, interval and magnitude, but cannot be used to do

texture classification. GLCM properties are extracted to describe the texture feature based on GLCM.

##### 2) GLCM Properties

Haralick defines 14 types of texture features. Commonly used to extract texture information in remote sensing images are: mean, variance, homogeneity, contrast, dissimilarity, entropy, angular second moment and correlation.

Homogeneity is also called inverse difference moment and it is a measure of local gray scale uniformity of the image. The bigger homogeneity means the more uniform gray scale of texture.

$$Homogeneity = \sum_{i=0}^{quant_k} \sum_{j=0}^{quant_k} p(i, j) * \frac{1}{1 + (i - j)^2} \quad (1)$$

Contrast reflects the clarity and the depth of texture striation. The greater the contrast, the deeper texture striation.

$$Contrast = \sum_{i=0}^{quant_k} \sum_{j=0}^{quant_k} p(i, j) * (i - j)^2 \quad (2)$$

Dissimilarity is similar to contrast but it is a linear increase.

$$Dissimilarity = \sum_{i=0}^{quant_k} \sum_{j=0}^{quant_k} p(i, j) * |i - j| \quad (3)$$

When the distribution of GLCM elements is concentrated around the main diagonal, the distribution of local region's gray scale is more uniform, the ASM value is bigger. On the contrary, if the values of co-occurrence matrix are all equal, ASM value is smaller.

$$ASM = \sum_{i=0}^{quant_k} \sum_{j=0}^{quant_k} p(i, j)^2 \quad (4)$$

Correlation describes the degree of similarity between rows elements or columns elements in GLCM, which can reflect the extension of a gray value along a direction.

$$Correlation = \sum_{i=0}^{quant_k} \sum_{j=0}^{quant_k} \frac{(i - Mean) * (j - Mean) * p(i, j)^2}{Variance} \quad (5)$$

#### C. LSH Forest

Locality sensitive hash (LSH) techniques have been used in many areas especially in near-duplicate detection, image retrieval and cluster. LSH hashes input items so that similar items map to the same "buckets" with high probability (the number of buckets being much smaller than the universe of possible input items).

In 2005, Mayank Bawa etc. put forward the LSH forest algorithm, which uses tree structure to replace of hash table and has the ability of self-parameters correcting. Figure 2 shows an LSH Tree that contains four points, with each hash function producing one bit as output.

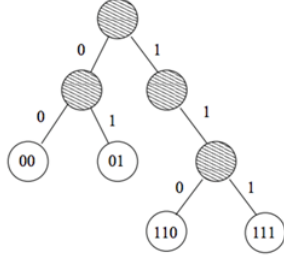


Fig. 2. A prefix tree on the set of LSH labels

#### D. Scale Invariant Feature Transform

Scale invariant feature transform (SIFT) is a local feature matching algorithm, the sift descriptor is scale-invariance, and has great performance when the perspective and brightness changes.

The function of scale is the convolution of the Gaussian function and the image function.  $(x,y)$  is the pixel localization;  $G(x,y,\sigma)$  is Gaussian function;  $\sigma$  determines the degree of image smoothing;  $I(x,y)$  is image function.

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \quad (6)$$

$$G(x_i, y_i, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(x-x_i)^2 + (y-y_i)^2}{2\sigma^2}\right] \quad (7)$$

The processing of SIFT algorithm as shown in figure 3.

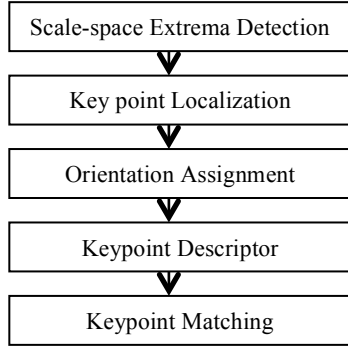


Fig. 3. SIFT algorithm process

### IV. IMPLEMENTATION

#### A. System Overview

##### 1) System Architecture

Aerial image discovery is one of the key technologies in various services using GIS, in which a match is to be identified in a large image database with a given query image with unknown location, orientation, and scale. The task is more challenging in city areas because the views have higher degree of homogeneity compared to non-urban areas.

Our proposed system is composed of three modules: offline processing module, online processing module and online search module. Figure 4 shows the system architecture.

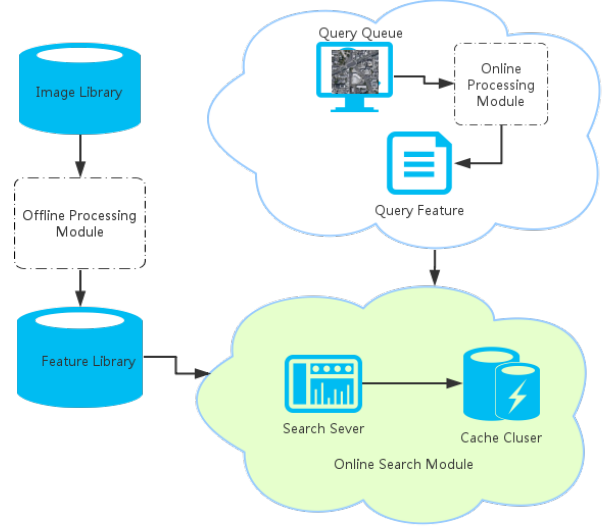


Fig. 4. System architecture

##### 2) System Working Flow

The system working flow is as shown in figure 5.

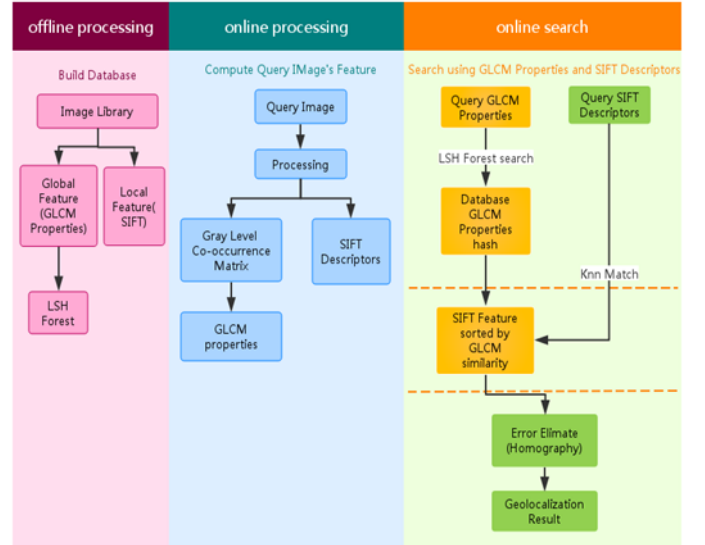


Fig. 5. System working flow

First of all, we build the database based on Kyoto maps downloaded using Google API. After image segmentation, texture features of images using GLCM properties are extracted. And then the data index is built using LSH forest. SIFT descriptors are also extracted and saved in feature database.

In online processing module, once a city aerial image input, Its texture feature and sift feature are both extracted in preparation for discovery.

The search module can be basically divided into two parts, preliminary matching and fine matching. First, we use texture feature to find the localization of similar texture image. After we get this pre-matching result, sift matching algorithm is applied to find exact matching result. Homography [11] and

RANSAC [12] is used here to do error matching points elimination.

### B. Database Structure

We construct our database covering the area as multi-layered structure, each layer containing images of the same magnitude. Figure 6 shows a conceptual illustration of the aerial image database consisting of multi-scale layers.

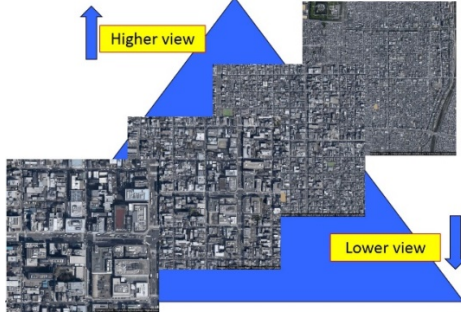


Fig. 6. Conceptual illustration of aerial image database consisting of multi-scale layers

The number of images in each level of Kyoto and west Tokyo is as shown in Table 1.

TABLE I. NUMBER OF IMAGES IN EACH LAYER

	L 1 (high view)	L 2	L 3	L 4 (low view)
<b>Kyoto</b>	120	360	918	2100
<b>West Tokyo</b>	450	1000	2700	4800

The system database is composed of file table, texture feature table and sift feature table, which shown as the figure below.

ID	FileName	Position(latitude&longitude)	Size
1	XXX	(X,Y)	
2	XXX	(X2,Y2)	

ID	GLCM Properties
1	
2	

ID	Local Descriptors
1	
2	

Fig. 7. Format of database structures

### C. Feature Extraction

#### 1) GLCM Properties Extraction

The analysis of texture feature based on GLCM should consider the image gray scale, window size, moving step and moving direction.

We first segment an image into 11\*11 sub-block, then create a 7\*7 window and set the value of pixel pair distance offsets to 5, moving direction to 0, moving step to 1.

#### 2) SIFT Descriptors Extraction

Firstly, we obtain the scale space of the image and do extrema detection in this space to preliminarily get the position and scale of feature points. Then calculating the 128 dimensional feature vectors using gradient information of the neighborhood pixels of the feature points.

### D. Image Retrieval

The matching algorithm performs better when matching pairs have similar size. We resize the input's size to the similar value of images' size in L1. Then extract the texture feature and sift feature of input to prepare for matching.

After feature extraction, we get the index list sorted by texture similarity and do fine matching based this result. The fine matching results will contain many error matching pairs, we choose homography and RANSAC to eliminate the error matching pairs. The image retrieval processing is as the sequence chart shows.

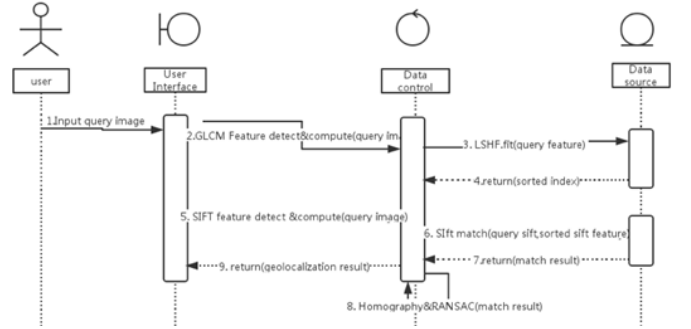


Fig. 8. Image Retrieval sequence chart

## V. RESULTS EVALUATION

### A. Experimental Setup

We prepared 100 query images by clipping randomly selected L4 (low-view) layer aerial images from Kyoto area. Then each is further modified to have random scale variation from 0.5 to 1.5 times the original scale, rotation at a random angle, and finally cropped to a square image. In the evaluation test, these 100 query test images were matched with all the file images in a L1 (high-view) layer using SIFT as interest point detector.

In this experiment, we used OpenCV 3.0 on MacBook Pro OS X EI Captain on a personal computer with 2.7 GHz Intel Core i5 CPU and 8 GB memory.

### B. Experimental Results

#### 1) Preliminary Test on GLCM Tutorial Performance

The result of different regions GLCM properties is shown in figure 9 The different regions like residential area, forest, river and so on have different properties that can be used to do

aerial image retrieval. LSH forest is implemented on these vectors in order to speed up the preliminary matching. Using these properties, the index of aerial images could be resorted according to the similarity of texture feature preparing for the next fine matching based on SIFT algorithm.

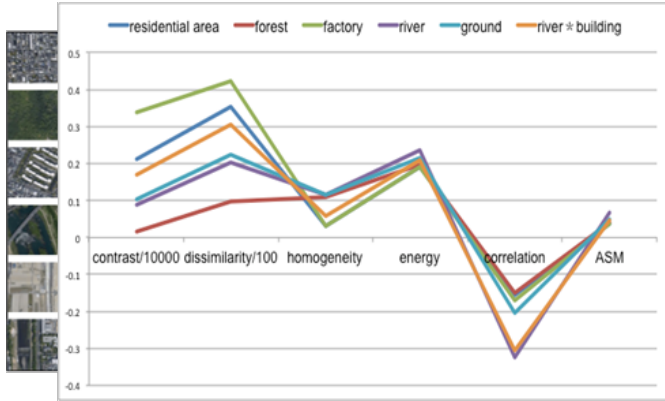


Fig. 9. GLCM properties of different regions

## 2) Retrieval Results of Different Query Image Sizes

In order to measure the experimental results, we use recall and precision to describe the performance of system's accuracy.

Recall and precision are defined as:

$$\text{Recall} = \frac{\text{Relevant Images retrieved}}{\text{Relevant Images in database}} \quad (8)$$

and

$$\text{Precision} = \frac{\text{Relevant Images retrieved}}{\text{Total Images retrived}} \quad (9)$$

Figure 10 shows the retrieval result of my algorithm and Amina's algorithm in Content based Image Retrieval using Texture Contents. My algorithm can find the exact localization result which integrate the GLCM tutorial with sift algorithm, while Amina's algorithm using GLCM tutorial [13] doesn't work well in the situation of aerial image discovery. This result shows the most currently image retrieval system can only find the similar images of the query instead of the exact one because they use global feature like color feature and texture feature which is not robust to do image retrieval. Global feature vectors change with the change of image's brightness, size, rotation and so on. SIFT algorithm can find the location of the query image precisely, so my algorithm integrating GLCM tutorial with SIFT algorithm can find the result correctly and accurately.

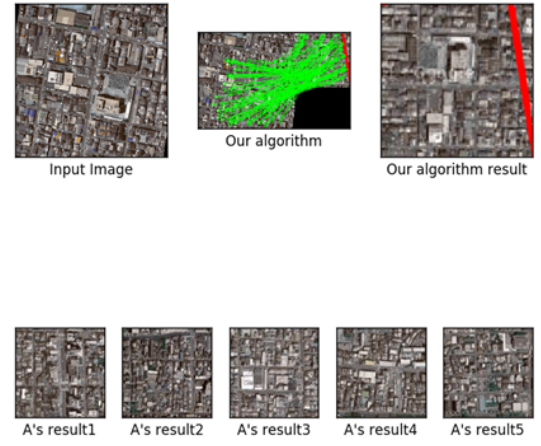


Fig. 10. An example of my algorithm result and Amina's algorithm result comparison (The sub image 1 is the original query image; the sub image 2&3 presents the matching points and the correct localization found using our algorithm; the images in second column are the results found using Amina's algorithm and there is no correct result among the most five similar images.)

As we can see, the algorithm successfully finds out a correct position from a considerably large, busy-looking, high-view image consisting of structures of similar appearances, using only a small, low quality image clip. This demonstrates the robustness of my approach against rotation and scale changes.

Table 2 summarizes the results of our matching trial with different query image sizes (one side in pixels). The results illustrate that our algorithm can meet the requirement of 100% recall while the query image processed to smaller size in order to reduce the matching time.

TABLE II. MATCHING PERFORMANCE OF DIFFERENT QUERY IMAGES SIZES

Query image size	Recall	Precision
400	90%	100%
200	90%	100%
150	90%	100%
100	85%	100%
75	56%	100%

## 3) Matching Time Results

Here the matching time means the average time one query data costs got from the experiments using 100 query data sets versus L1 image library.

The time using my algorithm1 is preliminary matching based on GLCM properties plus fine matching time and error elimination time. And the time using our algorithm2, which based on down sample and sift algorithm, means fine matching time plus error elimination time without preliminary matching. The matching time of original case is fine matching time plus error elimination time using original query data.

In table 3 and figure 11, the matching time results are compared between our algorithms and conversional algorithm while their precision is all 100%. In figure 11, the red line



represents the matching time only using conventional algorithm without coarse-level search and pre-processing which is much more than our algorithm represented by green line.

TABLE III. MATCHING TIME RESULTS COMPARISON

Query Image Size \ Average Time	Our Algorithm	Convectional Algorithm
400	13.51	30.59
200	5.41	7.17

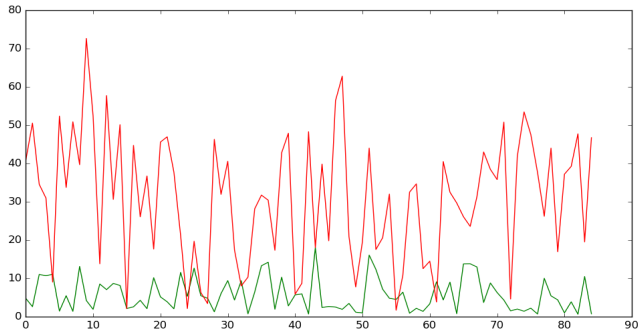


Fig. 11. Matching time results comparison

In figure 11, the red line represents the match time of conventional algorithm only using SIFT algorithm without coarse-level search and pre-processing while the green line represents the match time using our algorithm. It is observed that the algorithm in this paper can greatly reduce the matching processing time by simple down sampling combined texture feature based on GLCM and LSH forest, without any penalty of accuracy degradation.

## VI. CONCLUSIONS

We have addressed the task of aerial city image discovery, the difficulty is meet the requirements of accuracy, efficiency and robustness, because the global feature is easy to change with good efficiency, bad accuracy while the local feature is robust, accurate but very slow. To realize reasonably efficient query, I propose an approach with preliminary matching and

fine matching. The preliminary matching based on GLCM and LSH forest works to select the high similarity items with query data. In fine matching, we pre-process the query data into similar size with high view images in database cause the performance of sift algorithm is better between similar matching pairs.

In this paper, I outlined the proposed approach, the system configuration and how the discovery proceeds. The experimental results show this algorithm using down sample, GLCM, LSH forest and sift algorithm can greatly reduce the matching time while has high accuracy and robustness..

My future work will focus on ground to aerial image match using machine learning, which can expand the scope of application and realize the geo-localization through retrieval with a building image.

## REFERENCES

- [1] Smeulders A W M, Worring M, Santini S, et al. Content-based image retrieval at the end of the early years[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2000, 22(12):1349-1380.
- [2] Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(60):91-110.
- [3] Liu Z, Zhou S Q. Image retrieval by using of colors and texture features in images[J]. Computer Engineering & Design, 2007.
- [4] Gasteratos A, Zafeiridis P, Andreadis I. An Intelligent System for Aerial Image Retrieval and Classification[M]. Methods and Applications of Artificial Intelligence. Springer Berlin Heidelberg, 2004:63-71.
- [5] Ke Y, Sukthankar R, Huston L. Efficient nearduplicate detection and subimage retrieval[J]. Acm Multimedia, 2010:869--876.
- [6] Wu C, Fraundorfer F, Frahm J M, et al. Image Localization in satellite imagery with feature-based indexing[J]. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2008, XXXVII (B3a).
- [7] The GLCM Tutorial Home Page <http://www.fp.ucalgary.ca/mhallbey/tutorial.htm>
- [8] Zulpe N S, Pawar V P. GLCM textural features for Brain Tumor Classification[J]. International Journal of Computer Science Issues, 2012, 9(3).
- [9] Bawa M, Condie T, Ganesan P. LSH forest: self-tuning indexes for similarity search [C]. International Conference on World Wide Web. ACM, 2005:651-660.
- [10] McAfee R. Coarse matching. Econometrica 70(5), 2025–2034 (2002).
- [11] Kriegman D. Homography Estimation, Lecture Computer Vision I, CSE 252A, Winter 2007.
- [12] Fischler M A, Bolles R C. 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.
- [13] Khatra A. Content-based Image Retrieval using Texture Contents[J]. International Journal of Scientific Engineering & Research, 2013.