Comparisons of feature extraction algorithm based on unmanned aerial vehicle image

Wenfei Xi, Zhengtao Shi*, and Dongsheng Li

Abstract: Feature point extraction technology has become a research hotspot in the photogrammetry and computer vision. The commonly used point feature extraction operators are SIFT operator, Förstner operator, Harris operator and Moravec operator, etc. With the high spatial resolution characteristics, UAV image is different from the traditional aviation image. Based on these characteristics of the unmanned aerial vehicle (UAV), this paper uses several operators referred above to extract feature points from the building images, grassland images, shrubbery images, and vegetable greenhouses images. Through the practical case analysis, the performance, advantages, disadvantages and adaptability of each algorithm are compared and analyzed by considering their speed and accuracy. Finally, the suggestions of how to adapt different algorithms in diverse environment are proposed.

Keywords: feature point extraction, SIFT operator, Förstner operator, Harris operator, Moravec operator

PACS: 89.20.Bb, 89.20.Ff

1 Introduction

Unmanned aerial vehicles (UAV) have become a fast, efficient, low-cost and flexible remote sensing data acquisition systems to get the images with high-resolution [1-3]. If UAV images are used for aerial triangulation, we can carry out the geometric reconstruction, and obtain the digital surface model and the point cloud data [4], which can be widely used in fields like archaeology, vegetation monitoring, traffic monitoring, disaster management and 3D reconstruction, etc [5-11]. In the process of unmanned aerial vehicle (UAV) model geometry reconstruction, it is a very critical step to extract and optimize the matching points in aerial triangulation [12]. In traditional aerial photography, the commonly used matching feature point extraction methods can be roughly divided into two categories: the matching based on pixel gray value, and the matching based on feature [13]. In traditional and standard aerial triangulation, it’s very convenient to obtain the matching points by using the method based on pixel gray value. However, there are a lot of problems in unmanned aerial vehicle (UAV) imaging, such as irregular image overlapping, or large rotation deviation angle, so some difficulties will be encountered when the common method and software are used for the UAV images. In other words, a great deal of labor force and material resources are needed [14-16].

In computer vision and photogrammetry field, there are lots of feature point extraction operators are proposed. For example, Dicksheld and Förstner compared and analyzed image orientation, performance of these feature extraction operators. However, they only used the low resolution close range images and analog images in the experiment [17]. Lowe applied SIFT algorithm to machine vision, image matching, image retrieval fields, etc. [18]. Wang, Zhang compared and evaluated the performance of several commonly used operators [19, 20]. However, they all only compare the image data of building images, without referring to other objects.

Based on the previous research, this paper uses SIFT operator, Förstner operator, Harris operator and Moravec operator to extract the feature image data from different objects and analyzes the data. Some mathematical formula used in this article can refer to [21-23].
# 2 Feature operator extraction

## 2.1 SIFT algorithm

The SIFT algorithm, proposed by David G. Lowe, is a feature-based matching algorithm based on scale space, mainly using Gaussian differential pyramid model [24]. The Gaussian pyramid is essentially a multi-scale representation of the signal, Gaussian blurring the same signal or picture many times, and down-sampling to produce multiple sets of signals or pictures at different scales for subsequent processing.

The application of SIFT matching algorithm in the photo grammetry has drawn more and more attention. Its superior performance has greatly enhanced the degree of automatic orientation in close range photography and unconventional aerial photography (like UAV images). In this way, it still can carry out automatic aerial triangulation in the lack of GPS/IMU data support [25].

### 2.1.1 The establishment of Gaussian differential pyramid

On the basis of a series of reasonable assumptions, Koenderink and Lindeberg proved that the Gaussian function is probably the only spatial kernel scale to realize image scale conversion. One tile of image spatial scale can be defined as:

\[
L(x, y, s) = G(x, y, s) * I(x, y), \tag{1}
\]

where * represents the sum of convolution operation, \((x, y)\) denotes the space coordinates of pixel, \(s\) stands for scale coordinates of pixel, and \(G(x, y, s)\) is scale variable Gaussian kernel function, which is defined as:

\[
G(x, y, s) = \frac{1}{2ps^2} e^{-(x^2+y^2)/2s^2}. \tag{2}
\]

David G. Lowe raised that it uses Gaussian difference operator to establish image Gaussian difference scale space. It is able to obtain Gaussian difference images \(D(x, y, s)\) with different scales by using the Gaussian difference kernels and image convolution on different scales. All Gaussian difference images together make up the Gaussian difference pyramid of the image.

\[
D(x, y, s) = (G(x, y, ks) - G(x, y, s)) * I(x, y) = L(x, y, ks) - L(x, y, s). \tag{3}
\]

Construction process of image Gaussian difference pyramid is shown in the left part of Figure 1. In the group, the images from the upper layer are convoluted with the Gaussian functions of different scales to obtain the spatial images in which the constant scale factor is increased \(k\) times. In other words, the scale of first layer is \(s\), the scale of the second layer is \(ks\), and the next layer follows this rule. The image of the first layer of each group is obtained from the first layer sample of the previous set of images. The size of the image after sampling is \(1/4\) of the size of the images of the previous group, but scale size is twice of the original one. In this way, the Gaussian difference pyramid is constructed. The right part of Figure 1 describes the process how to construct Gaussian pyramid. In the group, Gaussian difference image is obtained after the subtraction between adjacent images. Then, after repeating the previous operations on the full set of images, we get the Gaussian difference pyramid.

![Figure 1: The generation of Gaussian difference pyramid](image)

### 2.1.2 Generation of feature point

After the key point SIFT feature area is generated, firstly the axis direction is rotated to the main direction of the key points, which ensures the rotation invariance. Then we select the neighboring window with the key as the center, as shown in the left part of Figure 2. The central pixel is the key point to be described. The surrounding little grids represent the pixels of the neighboring key points; the length of the arrow represents the gradient value of the pixel and the direction of the arrow represents the gradient direction of the pixel. The circles in the figure represent the range of Gaussian weighting of pixel values. As shown in right part of Figure 2, in every sub block, if we calculate the gradient direction histogram in eight directions, we can get a seed point. When all the seed points are obtained, we get a descriptor.
2.2 Forstner operator [26]

First, the Robert gradient and the gray covariance matrix of each pixel are calculated. Then the feature points, whose error ellipse is the smallest and closest to the round, are searched as the characteristic point.

The process is described as follows:
1. Calculate the Robert gradient for each pixel
   \[ g_u = \frac{\partial g}{\partial u} = g_{i+1,j} - g_{i,j} \]
   \[ g_v = \frac{\partial g}{\partial v} = g_{i,j+1} - g_{i,j} \]
   (4)

2. Calculate |x| covariance matrix of the gray value in the window
   \[ Q = N^{-1} = \left( \frac{\sum g_u^2}{\sum g_v^2} \frac{\sum g_u g_v}{\sum g_v^2} \right)^{-1} \]
   where
   \[ \sum g_u^2 = \sum_{i=-k}^{c+k-1} \sum_{j=-k}^{c+k-1} (g_{i+1,j} - g_{i,j})^2 \]
   \[ \sum g_v^2 = \sum_{i=-k}^{c+k-1} \sum_{j=-k}^{c+k-1} (g_{i,j+1} - g_{i,j})^2 \]
   \[ \sum g_u g_v = \sum_{i=-c+k}^{c-k} \sum_{j=-c+k}^{c+k-1} (g_{i+1,j+1} - g_{i,j})(g_{i,j+1} - g_{i+1,j}) \]
   (5)

3. Calculate the interest values of q and w:
   \[ w = \frac{1}{\text{tr}Q} = \frac{\text{Det}N}{\text{tr}N}, \quad q = 4\text{Det}N \left( \frac{\text{tr}N}{2} \right)^2 \]
   where \( \text{Det}N \) represents the determinant of \( N \), and \( \text{tr}N \) represents the trace of matrix \( N \).
   (6)

4. Determine candidate points
   \[ T_q = 0.5 \sim 0.75 \]
   \[ T_w = \begin{cases} \frac{fW(f)}{0.5 \sim 1.5} \\ cW(c = 5) \end{cases} \]
   (7)

When \( q > T_q \) and \( w > T_w \) at the same time, it confirms to be the candidate point.

(5) Select extremum points
The extreme points are selected based on the weights, that is, the largest candidate points are selected in an appropriate window, and the rest of the remaining points are eliminated.

2.3 Harris operator

Harris operator, presented by Harris and Stephens, is a signal-based feature extraction operator [27].

The process is stated as follows:
\[ M = G(s) \bowtie \begin{pmatrix} g_x & g_x g_y \\ g_x g_y & g_y \end{pmatrix}, \]
\[ R = \text{Det}(M) - k \cdot \text{tr}^2(M), \quad k = 0.04 \sim 0.06, \]
where \( g_x \) is the gradient of \( x \) axis direction, \( g_y \) is the gradient of \( y \) axis direction, \( G(s) \) is Gaussian template, \( \text{det} \) is matrix determinant, \( \text{tr} \) is matrix trace, \( k \) is a constant, \( R \) represents the interest value of the corresponding pixels in the figure.

2.4 Moravec operator

Moravec operator, proposed by Moravec, uses the characteristics of gray variance to extract feature point operator [28].

1. Calculate the pixel IV (Interest Value). In the image window with \( w \times w \) as the center of the pixel \((m, n)\), the sum of the squares of the grayscale of the adjacent four directions is:
   \[ V_1 = \sum_{i=-k}^{k-1} (g_{m+i,n} - g_{m+i+1,n})^2, \]
   \[ V_2 = \sum_{i=-k}^{k-1} (g_{m+i,n+i} - g_{m+i,1,n+i})^2, \]
   \[ V_3 = \sum_{i=-k}^{k-1} (g_{m,n+i} - g_{m,n+i+1})^2, \]
   \[ V_4 = \sum_{i=-k}^{k-1} (g_{m+i,n-i} - g_{m+i,1,n-i})^2. \]
   (9)

In the function: \( k = \text{int}(w/2) \). Take the smallest value as interest value of the pixel, that is
\[ IV_{m,n} = \min\{V_1, V_2, V_3, V_4\}. \]

2. Give an empirical threshold value, the interest value, which is bigger than the threshold, is chosen as the candidate point. The threshold value should be selected to
include required feature points in the candidate points, without excessive non-feature points.

3. Select extreme value pointing the candidate point as the feature point, in a certain size of the window (which can be different from the interest value calculation window), then eliminate the candidate value of interest which is not the largest, leaving the only one largest interest value at last. The pixel of the remaining one is a feature point.

3 The experimental case

3.1 UAV data acquisition

This paper takes Yunnan Normal University campus as an example, where unmanned aerial vehicle (UAV) is chosen as a platform to collect image data. In order to verify the effect of four operators to extract the feature points, building image, grassland image, shrubbery image, and vegetable greenhouses images are obtained. In this UAV flight, we choose the dajiang PHANTOM3 unmanned aerial vehicles (UAV) and use GPS/GLONASS positioning system with 12.4 million pixels. Its take-off weight is 1280g. The main idea of this experiment is designed as follows:

1. Use unmanned aerial vehicle (UAV) to obtain the original image data. In this process, the overlap degree of the obtained images should not be less than 60%. The uniform source of light and the same flight heights should be ensured in the field data collection.
2. The quality of the acquired image data must be checked. The software is used to check whether the image meets the quality requirements of modeling. If not, the UAV must fly again to collect enough data.
3. Use all kinds of operators to extract feature points.

The single image data from unmanned aerial vehicle (UAV) is shown in Figure 3.

3.1.1 The operator feature extraction

We use matlab2014 for programming. The running environment requires Windows 7 system, memory 4.00 GB, 64-bit operating system. Then the features are extracted by using SIFT operator, Forstner operator, Harris operator, and Moravec operator, respectively. The Forstner operator has a threshold of 0.65, the k value of the Harris operator is 0.06. The interest threshold of Moravec operator is 10000. And the results of the extraction are shown in Figures 4, 5, 6, and 7. The resolution of the images are: building (638 × 535), grassland (686 × 524), shrubbery (858 × 583), vegetable greenhouses (701 × 522).
Speed and accuracy analysis

In the same experimental conditions, the extraction speed of each operator is compared, which is shown in Table 1.

It can be seen from Table 1, in the process of building feature points extraction, Harris operator is slower; in the process of grassland feature points extraction process, Forstner operator is slower; in the process of bushes feature points extraction, SIFT operator is slower; in the process of vegetable greenhouses feature points extraction, Harris operator is slower.

In this experiment, in order to compare the accuracy of feature point extraction, the corner of the building is selected for our analysis. We measure 45 corner points coordinates. After that, we compare and analyzes the measured value and the coordinate value extracted from the feature points. The result is shown in Table 2.

As we can see from Table 2, the extraction accuracy of SIFT operator for building is the lowest, while the extraction accuracy of Forstner operator is the highest. SIFT operator extracts round feature points better than the corner point, so SIFT operator is not suitable for corner point feature extraction.

4 Conclusion

In this paper, the feature extraction algorithm is studied, and the main algorithms of photogrammetry and computer vision are analyzed and compared systematically in terms of speed and accuracy. The result shows that each algorithm has its own advantages and disadvantages, so the appropriate algorithm should be selected according to the specific task requirements.

Conflict of interests: The authors declare that there is no conflict of interests regarding the publication of this paper.
Comparisons of feature extraction algorithm based on unmanned aerial vehicle image

| Table 1: Time comparison of each operator's extraction (unit: second) |
|---------------------------------|----------------|----------------|----------------|
|                                 | building      | Grassland     | shrubbery      | vegetable greenhouses |
| SIFT operator                   | 4.6           | 2.8           | 2.1            | 2.7                   |
| Forstner operator               | 2.1           | 1.6           | 2.0            | 1.5                   |
| Harris operator                 | 5.1           | 2.6           | 3.6            | 2.9                   |
| Moravec operator                | 1.2           | 1.5           | 1.6            | 1.2                   |

| Table 2: The RMS of each algorithm |
|-----------------------------------|----------------|----------------|----------------|
| algorithm                         | RMSE \ pits   | RMSE\x \ pits | RMSE\y \ pits |
| SIFT operator                     | 4.849         | 2.718          | 4.016          |
| Forstner operator                 | 1.212         | 0.857          | 0.857          |
| Harris operator                   | 1.887         | 1.358          | 1.31           |
| Moravec operator                  | 2.276         | 1.359          | 1.826          |

Acknowledgement: We thank the reviewers for their constructive comments on improving the quality of this paper. This work research was supported in part by the Yunnan Provincial Department of education research foundation (2016ZZX067).

References


