# An Improved SIFT Algorithm for Unmanned Aerial Vehicle Imagery



2014 IOP Conf. Ser.: Earth Environ. Sci. 17 012187

(http://iopscience.iop.org/1755-1315/17/1/012187)

IOP Conf. Series: Earth and Environmental Science **17** (2014) 012187 doi:10.1088/1755-1315/17/1/012187

# An Improved SIFT Algorithm for Unmanned Aerial Vehicle Imagery

**J M Li [1, 2], D M Yan[1], G Wang[1] and L Zhang[1]**

[1]Key Laboratory of Digital Earth Sciences, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China.

[2]China University of Geosciences (Beijing), Beijing 100083, China.

E-mail: lijianming24@gmail.com

**Abstract**. The Unmanned Aerial Vehicle (UAV) platform has the benefits of low cost and convenience compared with satellites. Recently, UAVs have shown a wide range of applications such as land use change, mineral resources management and local topographic mapping. Because of the instability of the UAV air gesture, an image matching method is necessary to match different images of an object or scene. Scale Invariant Feature Transform (SIFT) features are invariant to image scaling, rotation and translation. However, the main drawback of a SIFT algorithm is its significant memory consumption and low computational speed, particularly in the case of high-resolution imagery. In this study, in order to overcome these drawbacks, we have analysed the construction of the scale-space in the SIFT algorithm and selected new parameters to construct the SIFT scale-space to improve the memory consumption and computational speed for the processing of UAV imagery. Here, we propose a restriction on the number of octaves and levels for Gaussian image pyramids. Our experiment shows that the proposed algorithm effectively reduces memory consumption and significantly improves the operational efficiency of the feature point extraction and matching under the premise of maintaining the precision of the extracted feature points.

## 1. Introduction

An Unmanned Aerial Vehicle (UAV) remote sensing system, which is flexible and easy to operate, can effectively overcome the inaccessibility of high-resolution remote sensing data in cloudy areas. Currently, UAV has a wide range of applications in various fields of investigation, such as land use change, mineral resources management and local topographic mapping [1, 2]. In order to combine it with other technology and acquire timely and accurate information of the entire survey area, real-time data processing for UAV imagery is required. Image registration is the key role of the entire processing in a UAV remote sensing system. Because of the large amount of data and the requirement of real-time processing, it is difficult to directly apply the current image registration algorithm to a UAV remote sensing system.

The description of interest points is a critical aspect of point correspondence, which is vital in UAV image registration. Because SIFT features [3, 4] are highly distinctive and invariant to scale, rotation and illumination changes, the SIFT algorithm is currently the most widely used in computer vision applications. However, the main drawback of this algorithm is the problem of significant memory consumption and low computational speed, particularly in the case of high-resolution imagery. In order to overcome this drawback, various improved SIFT algorithms have been proposed for remote sensing imagery [5-7] and UAV imagery [8, 9] registration. Li et al. [5] proposed an adaptive method of selecting the size of a Gaussian kernel to solve the problem of low computational speed of the construction of scale-space. Zhu et al. [6] improved the original SIFT algorithm in four aspects, including formulating pyramids and partitioning strategy, constraining the number of pyramid octaves, filtering feature points and enabling parallelization. In spite of improving the memory consumption and the computational speed of the SIFT algorithm, these abovementioned methods do not consider the characteristics of UAV imagery. In order to achieve the goal of real-time processing for UAV imagery, Xi et al. [8] proposed the use of a simplified Forstner operator to improve the SIFT algorithm by reducing the computation of feature point recognition. Xiong et al. [9] introduced the preconditions into the SIFT algorithm and reduced the searching range in the matching stage.

In this study, we analysed the construction of scale-space in the SIFT algorithm and selected new parameters to construct SIFT scale-space to improve the memory consumption and computational speed of the processing of UAV imagery. The proposed algorithm could ensure the stability and accuracy of feature points.

## 2. Methods

### *2.1. Original SIFT Algorithm*

The original SIFT algorithm consists of four stages: scale-space extrema detection, accurate keypoint localization, orientation assignment calculation, and keypoint descriptor generation. In this paper, we focus on scale-space extrema detection.

*2.1.1. Scale-space extrema detection.* The scale space of an image is defined as a function, $L(x, y, \sigma)$ that is produced from the convolution of a variable-scale Gaussian, $G(x, y, \sigma)$, and an input image, $I(x, y)$:

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

where * denotes the convolution operation in $x$ and $y$, $\sigma$ represents the scale, and

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

To efficiently detect stable keypoint locations in the scale-space, SIFT use scale-space extreme in the difference-of-Gaussian function convolved with the image $D(x, y, \sigma)$, which can be computed from the difference of two nearby scales:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma), \tag{3}$$

where $k$ denotes a constant multiplicative factor in two nearby scales.

In fact, the difference-of-Gaussian function is an approximation of the Laplacian-of-Gaussian function. The approximation error will reach zero as $k$ goes to 1. SIFT chooses to divide each octave of the scale-space into an integer number, $s$, of intervals; hence, $k = 2^{1/s}$. The scale-space consists of $O + 3$ Gaussian pyramid image octaves. Each octave has $s + 3$ images in the stack of blurred images. Adjacent image scales are subtracted to produce the difference-of-Gaussian images. Once a complete octave has been processed, the next octave is constructed by resampling the Gaussian image by taking every two pixels in each row and column.

The SIFT algorithm uses $\sigma = 1.6$, $s = 3$, $k = 2^{1/s}$ to construct the scale-space. Because of the expansion and pre-smoothing of the input image, the initial parameter of scale is $\sqrt{\sigma^2 - 1^2}$. If the top of the octaves in the scale-space has only 4 pixels, we set

$$O = [\ln(\min(h, w)) / \ln(2)] - 2, \tag{4}$$

where $w$ denotes the width of the input image; and $h$, the height of the input image. The scale-space consists of $-1, 0, 1, ..., O - 1$ octaves that include 6 Gaussian pyramids with different scales.

### *2.2. Improved SIFT Algorithm*

Because of the complexity of the SIFT algorithm, particularly that of the construction of the scale-space from the Gaussian pyramids, it is difficult to UAV remote sense the data on computer memory and calculate the corresponding time. In practice, it is difficult to use this algorithm directly. The memory needs to consider the original SIFT algorithm from the point of view of constructing a scale-space. We set the parameters as follows: the width of the input image = w, the height of the input image = h and the size of the top octave image = $2 \times 2$. In order to make full use of the input, the original SIFT method expands the image. After the above sampling, the image doubling increases the number of stable keypoints by almost a factor of 4 [4]. Therefore, the initial value of o is -1. Each pixel in the pyramids and the difference pyramids takes 4 bytes. Each octave of the scale-space has 3 intervals ($s = 3$), and the memory size of the Gaussian pyramids of octave o is $\mathrm{Mo}$, which can be calculated as follows:

$$\mathrm{Mo} = h \times w \times (s + 3) \times 4^{-o} \times 4, o \in [-1, \mathrm{O} - 1], \tag{5}$$

If $\mathrm{O} > 4$, the total memory of the Gaussian pyramids and the Gaussian difference pyramids is approximately $235 \times h \times w$. For instance, when the size of the input image is $4272 \times 2848$ in the SIFT algorithm, the memory consumed for constructing the scale-space is approximately 2.7 GB. In another case related to the construction of the scale-space, Li et al. [5] found that the time expenditure of constructing a scale-space accounts for more than 30% of the total time required for executing the SIFT algorithm. In response to these issues, while constructing the scale-space, we choose to build a Gaussian pyramid image from the original size of the input image, but not from an image that is twice the size of the input image. Therefore, the initial value of o is 0. In this case, if $\mathrm{O} > 4$, the total memory of the Gaussian pyramids and the Gaussian difference pyramids is approximately $59 \times \mathrm{h} \times \mathrm{w}$.

In our experiment, we choose the UAV remote sensing imagery to test our hypothesis. Three different sizes, namely $2136 \times 1424, 1068 \times 712, 534 \times 356$, of the input imagery are considered. In the matching stage, we choose 0.49 as the threshold of the Best-Bin-First (BBF) algorithm.

**Table 1.** Size of the input image = $2136 \times 1424$

| Type of SIFT | Time expenditure (s) | Memory consumption (Mb) | Number of keypoints | Number of matching pairs | Match ratio |
|---|---|---|---|---|---|
| Original SIFT | 0.95 | 681.7 | 25632 | 6085 | 0.24 |
| Improved SIFT | 0.21 | 167.5 | 8966 | 2790 | 0.31 |

**Table 2.** Size of the input image = $1068 \times 712$

| Type of SIFT | Time expenditure (s) | Memory consumption (Mb) | Number of keypoints | Number of matching pairs | Match ratio |
|---|---|---|---|---|---|
| Original SIFT | 0.22 | 170.5 | 8664 | 2497 | 0.29 |
| Improved SIFT | 0.05 | 41.9 | 2817 | 945 | 0.34 |

**Table 3.** Size of the input image = $534 \times 356$

| Type of SIFT | Time expenditure (s) | Memory consumption (Mb) | Number of keypoints | Number of matching pairs | Match ratio |
|---|---|---|---|---|---|
| Original SIFT | 0.05 | 42.6 | 2539 | 777 | 0.31 |
| Improved SIFT | 0.01 | 10.5 | 762 | 260 | 0.34 |

We use the match ratio and repeatability of feature points to measure the stability of feature points. The match ratio is defined as follows:

$$\text{match ratio} = \frac{correspondences}{keypoints}, \tag{6}$$

where *correspondences* denotes the number of match pairs and *keypoints* represents the number of feature points featured with the SIFT algorithm. The repeatability of feature points is defined as follows:

$$\text{repeatability ratio} = \frac{number\ of\ same\ keypoints}{number\ of\ full\ keypoints}, \tag{7}$$

where *number of same keypoints* represents the number of keypoints that appear in the original SIFT and improved SIFT algorithms at the same time, and *number of full keypoints* denotes the number of keypoints that appear in the improved SIFT algorithm.

From Tables 1–3, we can see that when constructing a scale-space from the imagery of the original size, the number of detected feature points and that of the matching points seem to have reduced. Compared with the original SIFT algorithm, the improved SIFT algorithm does not exhibit a significant decrease. At the same time, the time and memory consumption of the construction of the scale-space also decreases. The improved SIFT algorithm maintains a certain stability of the match ratio. In order to evaluate calculating time and memory consumption efficiently, we count the number of keypoints, number of same keypoints and repeatability of feature points for three different sizes of input imagery. From Table 4, we can see that the ratio of the repeatability of feature points is almost 80% for the abovementioned three sizes of the imagery; this leads the proposed SIFT algorithm to have a similar match ratio as that of the original SIFT, as given in Tables 1–3.

**Table 4.** Keypoints in two SIFT algorithms

| Size of image | Number of keypoints (original SIFT) | Number of same keypoints (improved SIFT) | Repeatability ratio |
|---|---|---|---|
| 2136 × 1424 | 8966 | 7282 | 0.81 |
| 1068 × 712 | 2817 | 2220 | 0.79 |
| 534 × 356 | 762 | 592 | 0.78 |

## 3. Experimental Results

In this experiment, four different sizes, namely $4272 \times 2848$, $2136 \times 1424$, $1068 \times 712$, $534 \times 356$, of the input imagery have been considered. The main object types in the input images, as shown in Figures 1 and 2, are buildings and farmland. We count the number of keypoints per image, match ratio, calculation time and memory consumption for the original SIFT algorithm and the improved SIFT algorithm.

**Figure 1.** UAV images mostly contain building object types.

**Figure 2.** UAV images mostly contain farmland object types.

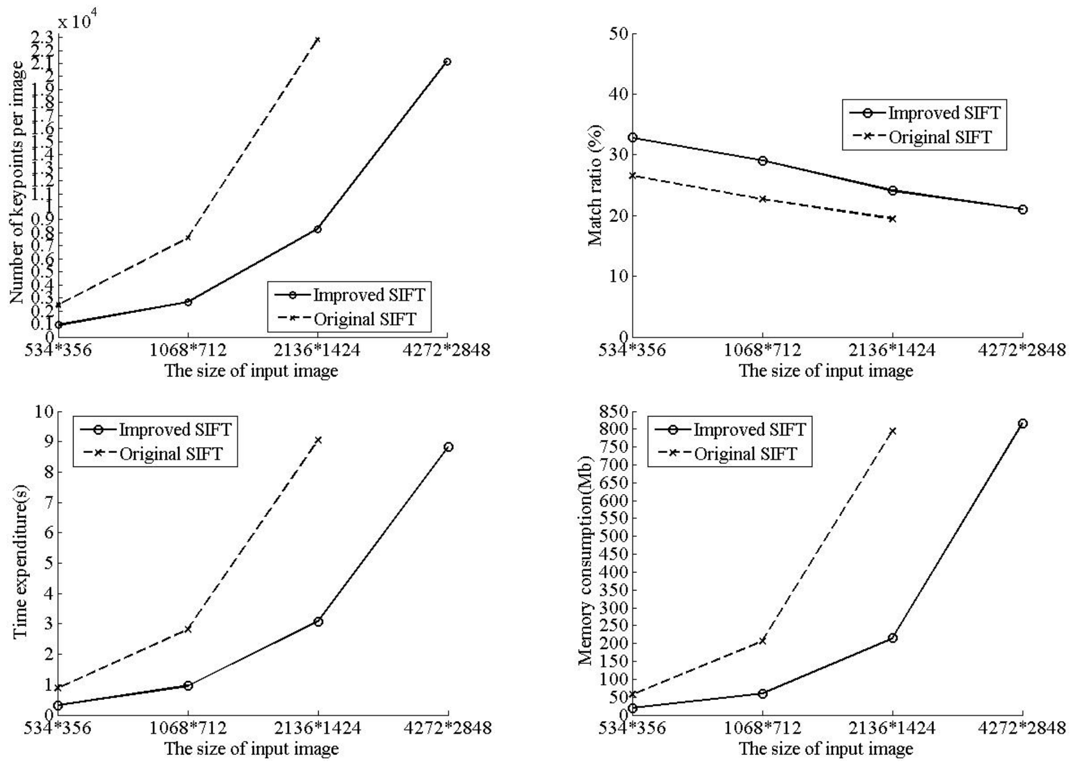


**Figure 3.** Comparison of the four indicators of the improved and original SIFT methods for the image shown in Figure 1.

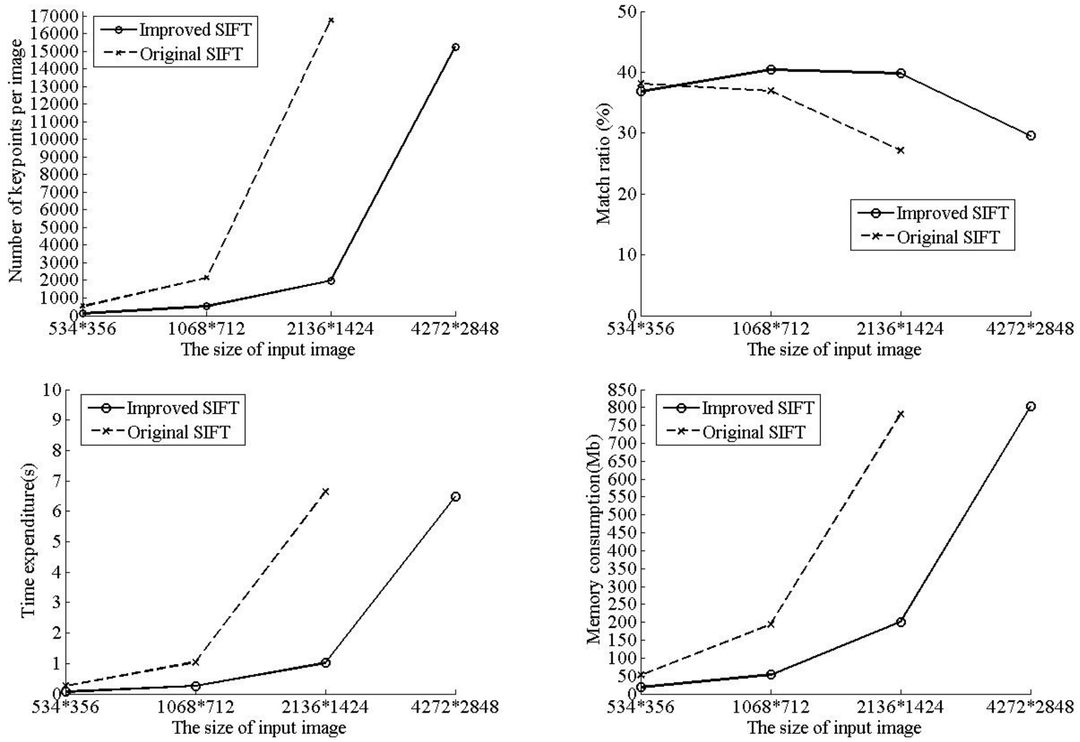


**Figure 4.** Comparison of four indicators of the improved and original SIFT methods for the image shown in Figure 2.

Because of the problem of the computer performance, the original SIFT algorithm is not suitable for the UAV imagery having a large size of $4272 \times 2848$. The abovementioned four indicators are not computed using the original SIFT algorithm.

From Figures 3 and 4, we can conclude that the time consumption of the improved SIFT method is less than $1/3$ of that of the original SIFT method, the memory consumption of the improved SIFT algorithm is $1/4$ of that of the original SIFT algorithm, and the match rate for the improved SIFT is comparable to that of the original SIFT. Meanwhile, the number of keypoints are decreased for the improved SIFT method. On the basis of the matching requirements for UAV remote sensing image, we can conclude that the number of feature points and that of the match points meet the requirements of solving a collinear equation.

## 4. Conclusion

In summary, when extracting feature points in UAV remote sensing imagery, we construct a scale-space from the original size of an image, not from an image that is twice the size of the original image. The improved SIFT algorithm is better than the original SIFT algorithm with respect to the time and memory consumption. However, the real-time requirement for UAV remote sensing systems is still a problem, particularly for the large UAV images (such as images that have a size of $4272 \times 2848$). Although, the proposed algorithm already reduces the number of keypoints, a number of keypoints are redundant. In order to further reduce the number of keypoints, we should consider the quality of keypoints. Therefore, our next step will focus on the filtration of the extrema points.

## Acknowledgements

The study was supported by the National Key Technology R&D Program (Grant No. 2012BAJ23B02).

## References

[1] Lin Z J 2008 Uav for mapping—low altitude photogrammetric survey *International Archives of Photogrammetry and Remote Sensing, Beijing, China* 1183-6

[2] Everaerts J 2008 The use of unmanned aerial vehicles (uavs) for remote sensing and mapping *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **37** 1187-92

[3] Lowe D G 1999 Object recognition from local scale-invariant features *The Seventh IEEE International Conference on Computer Vision* **2** 1150-7

[4] Lowe D G 2004 Distinctive image features from scale-invariant keypoints *International journal of computer vision* **60** 91-110

[5] Li F F, Xiao B L, Jia Y H and Mao X L 2009 Improved sift algorithm and its application in automatic registration of remotely-sensed imagery *Geomatics and Information Science of Wuhan University* **34** 1245-9

[6] Zhu Z W, F. S Z and Luo J C 2011 Parallel remote sensing image registration based on improved sift point feature *Journal of Remote Sensing* **15** 1024-39

[7] Alhwarin F, Ristić–Durrant D and Gräser A 2010 Vf-sift:Very fast sift feature matching *Pattern Recognition* 222-31

[8] Xi C X and Guo S X 2011 Image target identification of uav based on sift *Procedia Engineering* **15** 3205-9

[9] Xiong Z M, Wan G, Yan H and M. L 2012 Unmanned aerial vehicle serial aerial image automatic registration based on improved sift algorithm *Journal of Geomatics Science and Technology* **29** 153-6