Image Perspective Invariant Features Algorithm Based on Particle Swarm Optimization

Ge Lu
Department of School of Art & Design, Zhengzhou Institute of Aeronautical Industry Management, Zhengzhou 450016, China

Abstract—To solve the problems like affine sampling strategy and sampling point discrete settings etc. of traditional image matching techniques, this paper proposes the image perspective invariant features algorithm based on particle swarm optimization. The algorithm uses the features of the perspective transforming sampling methods to extract the algorithm and uses perspective transforming sampling to simulate scenery. Particle swarm algorithm is used as the optimization tool to optimize the parameters of the sampling, which aims at searching the optimal transformation model between images in continuous rotation parameter space. Experimental results show that the proposed algorithm can get more correctly matched points, search the better transformation model and reasonably reflect the correspondence between the matching images.

Index Terms—Fitness Function; Sampling; Discrete Set; Particle Encoding

I. INTRODUCTION

Image matching method can be divided as gray-based method and feature matching-based method. Wherein, gray-based method uses the grayscale statistics of the image itself to measure the degree of similarity between images. The methods mainly include related information, interactive information and so on. Such characteristic of algorithm is that the dense matching of pixels which can be realized and its t precision is high, it is sensitive to the factors like scale changes, rotation, uneven illumination and so on. Matching method based on image features typically includes feature extraction, feature matching, transformation model estimation and resampling. On the aspects of feature extraction, point feature is the most commonly used image features, generally including corners, extreme points, intersections etc [1-3]. The point feature extraction methods commonly include Harris operator, Forstner operator, Susan operator, FAST operator, ORB operator and so on, which mainly uses the comparison between the extreme of neighborhood gradient to detect the feature points. In recent years, Lowe has proposed the scale invariant feature transform (SIFT) operator. This algorithm uses the properties of the scale-space, takes the extreme points achieved from the scale domain and space domain as feature points and achieves the scale and rotation invariant through the distribution of the gradient direction.

After the SIFT algorithm is proposed, in the field of image processing it has been caroused great attention and spawned a series of improved algorithms, such as the use of principal component to analysis PCA-SIFT. Compared with SIFT, the matching efficiency of algorithm has been greatly improved in the transforms of rotation and scaling and noise jamming and illumination changes. On May 2006, Bay et al put forward the Speeded up robust features (SURF) algorithm, where its matching performs exceeds SIFT and it can get faster. Zeng Luan et al through anglicizing neighborhood information of SIFT feature points, presents a feature descriptor in circular area based on segmentation in sector region and reduces the SIFT feature to 64 dimensions while maintaining the performance at the same time [4-7].

Although the above algorithm achieves very good results on the aspects of feature description and matching, Mikolajczyk et al discussed that the extraction method based on SIFT feature is ineffective for multi-angle image matching, especially when there is a severe deformation between scenes, this algorithms is basically failure [8-9]. For solving this problem, Random ferns and Robust make affine transformation for the block, in which feature point is the center, and use the compared pixel information to structure the image features fast classifier. It fast the matching speed and improve the robustness of perspective changes at the same time. Lobaton et al proposed the robust topological features on homology theory, which get a good match results on the local deformation image matching [10-11]. Schmidt et al analyzed various symmetrical features of the nature and proposed the local symmetry characteristics. It will greatly improve the robustness of the multi-angle image matching when the building is affected by light and has texture change. In 2009, Morel et al analyzed the efficiency of SIFT in case of the affine variation and proposed the Affine scale invariant feature transform (ASIFT) algorithm based on the set of image transform, which simulate the transformation of the target at different angles through discrete affine sampling [12-13]. Compare with the SIFT, Harris-Affine algorithm and other algorithms, the experimental comparison shows that ASIFT can maintain invariance on rotation, translation, scaling zooming, but also it has a strong robustness to the distortion problem caused by perspective changes. After the presentation of ASIFT algorithm it has attracted wide attention, in which Podbreznik et al combined with the image segmentation method and pushed the AISFT to the matching with large viewing angle [14]. Cao et al used the background information and improved the matching
results of the improved ASIFT to repeated pattern. Similarly for the repeat mode problem, Brese et al used the graph transform matching method to delete ASIFT mismatch, which improve the robustness of feature extraction and matching [15]. Liu et al generalized the ASIFT algorithm to the feature points matching of behavior recognition, which achieved good recognition effect [16]. In our previous work, on the basis of obtained the transformation model by ASIFT algorithm, the simple method based on fuzzy control is proposed to search the transformation model around the optimal matching, which get a better matching results.

Although ASIFT algorithm uses the affine sampling, which is conducive to enhancing the robustness of multi-angle image matching, through the theoretical analysis, there are two problems of ASIFT: 1) ASIFT simulate the scene change brought by perspective distortion through affine transformation sampling. However, according to the camera perspective imaging principle, the correspondence relationship between the scene and image plane is the perspective transformation relationship, which can be described by the collinear; 2) the sampling of ASIFT to the transformation parameters (rotation angle) is discrete, that is, limited sampling points are used to "guess" the transformation model between images. Although through experimental comparison ASIFT gives the sampling point setting standards, the parameters of rotation angle are continuous, therefore the solution obtained by ASIFT algorithm is generally not the optimal transformation model.

For the first question, this thesis proposes a feature extraction algorithm based on perspective transform sampling method, which analog the variation of the scene shot at different angles by using perspective transformation sampling. For the second problem, the paper on the basis of perspective transform sampling methods optimizes the parameters of the sampling by using the particle swarm algorithm as the optimization tool. Its aim is searching the optimal transformation model between the images in the continuous rotation parameter space. Through three different image types experiments, it can be found that compared with the ASIFT, SIFT, Harris affine, and other algorithms, the deformation of ASIFT caused by the changes of perspective has strong robustness. For any defined affine transformation A, ASIFT verifies A can be decomposed by singular value (SVD) as followings:

\[
B = \varphi \begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix} = \varphi \begin{pmatrix} \cos \theta - \sin \theta \\ \sin \theta \cos \theta \end{pmatrix} \begin{pmatrix} \cos m \sin m \\ \sin m \cos m \end{pmatrix} \quad (1)
\]

In the sampling process, image needed to be matched is placed in the center of the world coordinate, and from the camera orientation and angle the sampling images can be deduced. Since SIFT operator is invariance for rotation and scale, ASIFT only samples from t and k in formula (1), so the affine transformation of sampling image can be obtained by formula (2):

\[
U_{r,k} = \begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix} T_r^k (U) \quad (2)
\]

Figure 1 describes the sampling settings of t and k recommended by the ASIFT, where the black dot indicates the position of the virtual camera. Through experimental analysis, Morel recommends that \( r \in \{1, \sqrt{3}, 3, \sqrt{3}, 6, 6 \sqrt{3}\} \) and \( k = \{0, b / r, \ldots, kb / r\} \) can ensure the balance between algorithm speed and matching accuracy, where \( b = 2\pi / 5, k = \pi r / b \).
After the sample image is generated, all images make feature extraction and matching with SIFT. In order to accelerate the algorithm computing speed, ASIFT uses a dual resolution strategy, that is, firstly the $3 \times 3$ down-sampling of image is taken to reduce the image resolution and accelerate the features extraction and matching of SIFT, which can store up the maximum five matching transformation model. Finally, the resamplings of corresponding models to the original high resolution image are taken to enhance the robustness of SIFT features to affine transformation.

![Figure 1. Sampling settings of $t$ and $\kappa$](image)

### B. Perspective Sampling

Although affine sampling-based ASIFT, Random ferns and others achieved very good results in multi-view matching, according to the camera imaging principle, the relationship between object point and image point is correspondingly perspective relationship. As shown in Figure 2, the frames are deformed in different shooting angles. Wherein, Figure 2 (a) is a front view image; Figure 2 (b) and Figure 2 (c) are side views. From Figure 2 (b) it can be seen that affine transformation (parallel rectangular frame) can not simulate the image distortion. Conversely, as shown in 2 (c), the perspective transformation can better estimate the perspective distortion from the multi-angle shooting. This paper uses the perspective sampling instead of affine sampling to improve the matching robustness of the algorithm from the multi-view images. The SIFT matching method based on perspective sampling is referred to Perspective scale invariant feature (PSIFT) algorithm. As shown in Figure 3, in the image sampling process, the reference image is placed in the center of the world coordinate (XOY plane center). Suppose $S = (x_s, y_s, z_s)$ is the analog camera orientation, according to the image principle of perspective transformation, the corresponding relationship between the point $R(x, y, z)$ on the reference image and the point $R(x, y)$ on the sampling image can be expressed by collinear equation:

$$
\begin{align*}
    x' &= -p \beta_1 (x - x_s) + z_s (y - y_s) + j_1 (z - z_s) \\
    y' &= -p \beta_2 (x - x_s) + z_s (y - y_s) + j_1 (z - z_s)
\end{align*}
$$

(3)

where in, $p$ represents the camera focal length; $(x_s, y_s, z_s) = (r \sin \varphi \sin \kappa, -r \sin \varphi \cos \kappa, r \cos \kappa)$ is the camera in the world coordinate; $r$ represents the distance between the projection center and origin. Since the reference image is placed on the XOY plane, it can be obtained $z = 0$. On the other hand, since in the sampling process main optical axis is always aligned to the center of the reference image (the origin of the world coordinates as shown in Figure 3). Thus, it is only needed to sampling $\kappa$ and $\varphi$, and the rotation matrix $R$ can be rewritten as:

$$
\begin{align*}
    p &= \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \cos(-\kappa) - \sin(-\kappa) 0 \\
                            0 \sin(-\varphi) \cos(-\varphi) \\
                            \end{pmatrix} \begin{pmatrix} \sin(-\kappa) \cos(-\kappa) 0 \\
                            0 0 1 \\
                            \end{pmatrix} \\
    &= \begin{pmatrix} 
    -\cos \varphi \sin \kappa \\
    \cos \varphi \cos k \varphi \sin \varphi \\
    \sin \varphi \sin k - \sin \varphi \cos k \cos \varphi \\
\end{pmatrix}
\end{align*}
$$

(4)

Substituting (4) into equation (3), singular transformation matrix between images can be obtained as following:

$$
T = \begin{pmatrix}
    -p \cos k & -p \sin k & 0 \\
    p \cos \varphi \sin k & -p \cos \kappa \cos \varphi \sin \varphi & 0 \\
    \sin \varphi \sin k & -\sin \varphi \cos k & -p \\
\end{pmatrix}
$$

(5)

![Figure 2. Perspective distortion caused by view changes](image)

![Figure 3. Camera model of the sampling perspective](image)

Figure 4 (a) and Figure 4 (b) are two low-altitude remote sensing image (white box is the outside outline of buildings) offered by Jimei University International Academic Exchange Center, in which it can be seen that from different perspectives the buildings image exists perspective distortion. Wherein, Figure 4 (c) is the sampling image of the reference image Figure 4 (a); sampling parameter $\varphi = \arccos 1/4$, $k = 4\pi/5$. From the sampling image it can be seen that the appearance of the building is basically the same with the input image (Figure 4 (b)).

### III. PSIFT Algorithm Based on Particle Swarm Optimization

Sample points of ASIFT algorithm is set by the discrete, which can only generate the limited and fixed
angle sample images, therefore, in a continuous parameter space it can not make sure the obtained image has maximum matching. As shown in Figure 5 (a) and Figure 5 (b), there is distortion between the two images. Through the affine sampling of $\phi$ and $t$ in the reference image, SIFT feature matching can be calculated. Thematching point function corresponding to the rotation angle is shown in Figure 5 (c). It can been seen that matching feature is a multimodal problem. The matching number obtained through the finite discrete sampling points can not guarantee the global optimum. Therefore, paper introduces the particle swarm optimization (PSO) based on perspective sampling to fast search the angle $\phi$ and angle $\kappa$ corresponding to the optimized images by using the particle swarm optimization algorithm. This section will briefly describe the PSO algorithm principle, PSIFT parameter space and fitness function settings.

A. Particle Swarm Optimization and Improvement

PSO algorithm is a swarm intelligence search algorithm proposed by Kennedy et al in 1995. The algorithm is initialized n random particles in the problem space, and then through the transfer of information and iteration between the particles to find the optimal solution. The particles update themselves by tracking the two extremes. Namely the particle itself finds the optimal solution $P_{best}$ and currently populations find the optimal solution $g_{best}$. After obtained the two optimal values, the particles update the velocity and position according to the following formula:

$$h_s(r+1) = \mu h_s(r) + p_r \cdot rand \left( t_s - z_s(r) \right) + p_r \cdot Rand \left( t_s - z_s(r) \right)$$

$$z_s = (r+1) = z_s(r) + z_s(r+1)$$

where in, $z_s(r)$ represents the location of the $i_{th}$ particle at time $t$; $z_s(r)$ represents the current location of the $d$-dimensional components; $h_s(r)$ indicates the speed of the particle at time $r$; $p_r$ represents the best location of the $i_{th}$ particle reached; $pg$ represents the best optimal point of the population; $p$ is the inertia weight; $c1$ and $c2$ are acceleration factor; $rand( )$ and $rand( )$ are random variation functions in the range of $[0,1]$. PSO repeatedly uses equation (6) and (7) to change the particle velocity and position, until the termination condition is reached.

B. Particle Encoding and Fitness Function

In the PSIFT optimization process, each particle represents the sampling angle $\phi$ and $\kappa$ in the image needed to be matched, so the particles coding can be expressed as: $\phi = (\phi_\alpha, j_\beta, \phi_\lambda, j_\gamma)$. Wherein, $\phi_\alpha \in [0, \pi/3], k_\beta \in [0, 2\pi]$, represents the transformation parameters of the $i_{th}$ particle to the $j_{th}$ image needed to be matched. Due to the purpose of the optimization is searching the optimal matching sampling image pairs, the SIFT correct matching value between images is the fitness function. The final output results combines correct matching features extracted in the iterative process.

C. Algorithm

Algorithm1. PSIFT algorithm based on particle swarm optimization.

Input: Images $U_1$ and $U_2$ needed to be matched $U_1$, $U_2$ Be matched.
Outputs: feature matching I1 and I2 and the matching characteristic coordinates.

**Step 1** Setting the initial parameters (including the population size n, the acceleration factor $p_1$ and $p_2$, inertia weight $z$, iterations $t_{max}$ and so on);

**Step 2** According to the rotation angle corresponds to each particle, perspective sampling the image; calculating the SIFT feature matching of the sampling image number as the particle's fitness value and recording the matching feature coordinates.

**Step 3** According to the fitness value of each particle updating the optimal $P_{best}$ and the global optimal $G_{best}$;

**Step 4** For each particle, according to formula (6) and (7) updating the velocity and position of particle;

**Step 5** If the number of iterations reaches the setting value, outputting the optimization results, otherwise go to step 2.

IV. EXPERIMENTAL SIMULATION AND ANALYSIS

The purpose of the experiment is to test the scale and stability of viewpoint changes of PSO + PSIFT, PSO+ASIFT, ASIFT, SIFT, Harris affine and MSER six algorithms. Due to the low-altitude remote sensing images generally have large differences in the depth of field, high resolution, fierce view changes and other features, the paper selects three groups of low altitude remote sensing images as the data set to evaluate the efficiency of various matching algorithm. The image is obtained by unmanned low-altitude remote sensing system in the Jimei University Engineering Technology Center. There sets of images respectively have the features like scale changes, weak angle changes, great angle changes and so on. Experimental evaluation index is the correct matching of the image feature points.

A. Experimental Environment and Setup

In the PSIFT parameters, let $r = 3k$, $p = 2k$, wherein, $k$ represents the length of the diagonal in the image. On the aspect of the swarm, since the particles code are only the four-dimension, the search space is small. In the experiments, it can be seen that the population generally converges in the fewer iterations. Thus, let the population size of PSO algorithm as $M = 16$ and the maximum number of iterations is $t_{max} = 100$ generations. Inertia weight $w$ linearly reduces with the number of iterations, where $z$ Initial = 1, $z$ Final = 0.3, $p_1 = p_2 = 2.2$. Search range of parameters is as $\phi = [0, \pi/2]$, $k \in [0, \pi/2]$. PSO + PSIFT, PSO + ASIFT, ASIFT and SIFT algorithm are using RANSAC algorithm to remove the mismatch.

B. Results Analysis

The first set of experiments mainly evaluates the robustness of various algorithms for the scale changes. Taken Figure 6 as an example, the image Corridor was taken in July 2011, and the building is the Jimei University corridor. It can be seen that the shooting height in Figure 6 (a) and Figure 6 (b) are different, resulting in the scale changes between scenes. Figure 6 (c) ~ Figure 6(e) are respectively the matching results of PSO + PSIFT, PSO + ASIFT and ASIFT algorithms and the matching results of SIFT, Harris affine and MSER algorithms are shown in Table 1. It can be seen that Harris affine and MSER are sensitive to scale changes.
while the SIFT algorithm finds 50 correct feature matching pairs, which explains that the Gaussian pyramid of the SIFT operator can effectively overcome the impact of feature extraction and description caused by scale changes. On the other hand, due to the SIFT features-based ASIFT, PSO + PSIFT and PSO + ASIFT generate a large number of sampling images, more matching features can be obtained compared to SIFT algorithm although the time increased.

The second set of experiments analyzes impact of the weak angle change to the algorithm. Taken Figure 7 as an example, the image (HQU) shot at Huaqiao University Xiamen campus. Due to the UAV shooting angle variation, there are rotations, scale changes, weak perspective deformation between the acquired images. It can be seen from the Table 1 and Figure 7 (c) ~ Figure 7 (e), SIFT features has a good matching effects to the rotation of the scene. On the other hand, due to the perspective distortion of the multi-angle scene, PSIFT algorithm is more robust of the characteristics obtained through the perspective sampling to the viewing angle, which can obtain more feature matching pairs than ASIFT. What’s more, due to the shooting height of these two images are different, the scale changes of the scene are not obvious, therefore Harris-affine and MSER also can find sufficient number of feature matching pairs.

In the third set of images, because altitude of UAV flight is low and the view angle shot rapidly changes, building perspective distortion phenomenon is obvious. Taken Figure 8 as an example, the group of pictures (Campus 1) was taken in September 2011; the buildings are the teaching podium in Jimei University new campus. Due to perspective changes large, buildings corresponding to two images severe deformed. On the other hand, the roof structures of buildings with Jiageng style are similar, so there are a large number of repeated modes and the extracted feature points are easily confusing. Figure 8 (c) to Figure 8 (e) are respectively the matching results of PSO+PSIFT, PSO+ASIFT and ASIFT algorithms. It can be seen from Table 1 that the ASIFT, SIFT, Harris affine and MSER algorithms are basic failure in the context of repeating pattern and large angle changes; ASIFT algorithm using PSO optimization extracts five pairs of correctly matched characteristics. Compared to the above methods, PSO + PSIFT strategy finds 85 pairs correct matching characteristics, which also shows the repeat mode will show different characteristics in different perspectives. As the previous analysis, perspective sampling strategy can estimate the perspective changes of repeating pattern, so the obtained characteristic has a stronger matching robustness under the circumstance of large angle changes and repeated modes.
TABLE I. MATCHING RESULTS OF SIX KINDS OF IMAGE MATCHING METHODS AT LOW ALTITUDE REMOTE SENSING DATA SETS

<table>
<thead>
<tr>
<th>Image class</th>
<th>Image</th>
<th>PSO+PSIFT</th>
<th>PSO+ASIFT</th>
<th>ASIFT</th>
<th>SIFT</th>
<th>Harris MSER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale changes</td>
<td>Park</td>
<td>384</td>
<td>487</td>
<td>432</td>
<td>68</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>People</td>
<td>835</td>
<td>488</td>
<td>266</td>
<td>75</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Tree</td>
<td>460</td>
<td>198</td>
<td>198</td>
<td>95</td>
<td>36</td>
</tr>
<tr>
<td>Weak perspective changes</td>
<td>Hat</td>
<td>1500</td>
<td>1456</td>
<td>1208</td>
<td>220</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Corridor</td>
<td>215</td>
<td>148</td>
<td>69</td>
<td>51</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Motorcycle</td>
<td>200</td>
<td>180</td>
<td>170</td>
<td>36</td>
<td>0</td>
</tr>
<tr>
<td>Large viewing angle changes</td>
<td>Lake</td>
<td>378</td>
<td>153</td>
<td>140</td>
<td>21</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Campus1</td>
<td>73</td>
<td>35</td>
<td>29</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Campus2</td>
<td>85</td>
<td>6</td>
<td>55</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Campus3</td>
<td>94</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 9 is the images of the rest of the three sets of experimental data at low altitude; the relevant matching results are shown in Table 1. It can be seen that SIFT-based algorithm maintain a good matching results on the scale and rotation changes of the scene, especially the matching effect of scaling images better than Harris affine and MSER. On the other hand, after the analysis of PSO + PSIFT, PSO + ASIFT and ASIFT matching results, it can be found that transforming the sampling strategy is able to simulate the deformation of the scene at various observation angles, therefore, it has a better perspective which will not be deformed.

C. Complexity Analysis

As expressed in the experimental setting, in the sampling image generation process, the code of each particle is four-dimension, so the search space is small. Therefore the population size of PSO algorithm is set as $M = 16$ ; the maximum number of iterations is $T_{max} = 100$ generations. Because each particle is corresponding to two sample images, in the PSO iterative process it requires 1500 times to extract and match the image feature to mismatch RANSAC. On the aspect of ASIFT algorithms, through experimental analysis, Morel recommends $r \in \{1, \sqrt{3}, 3, \sqrt{6}, 6, 6\sqrt{3}\}$ and $k = \{0, 1/2, ... \}$ as the set of discrete sampling points, so each piece of the reference image generates around $5 + 6 + 10 + 10 + 20 + 20 \approx 84$ sampling image. For the inputting two reference images, there is a need to extract features from $84 \times 84 = 168$ images, match features between images from $84 \times 84 = 7056$ and generate mismatch process of the RANSAC. Experiments show that image feature extraction and description of the time costs are related to the characteristics of the image itself, while the image matching and RANSAC mismatch are related to with the features of number. The statistics aiming at the low-altitude remote sensing data in this thesis show that the feature extraction, feature matching and RANSAC of the average time cost ratio of two images is about 56:13:1. So, the total time complexity ratio of PSO + PSIFT and ASIFT is about 1:04.

V. CONCLUSION

Due to there are the problems like affine transformation sampling point discrete settings of the ASIFT algorithm, so this paper presents a perspective invariant image matching PSIFT algorithm based on particle swarm optimization, which estimate the different shooting angle distortion through the perspective transformation sampling and uses the particle swarm algorithm as optimization tool to optimize the angle of rotation. The aim is searching the optimal transformation model between the images in the continuous parameter space. Through experimental results of the three different image types show that compared with the AS-FIT, SIFT, Harris affine and MSER algorithms the PSIFT algorithm adopting the practical swarm optimization can obtain more correct matching points to improve the robustness of the algorithm to perspective change and repeat mode.

ACKNOWLEDGMENT

This work was supported in part by the major program of Henan provincial education department Foundation of China under (Grant No. 12B430024); this work was supported in part by the scientific and technology project of Henan provincial Science and Technology Agency (No.122102310469).

REFERENCES


