Architectural Scene Construction Model Based on Adaptive Error Correction RANSAC Algorithm

Zhaofeng Wang, Guoxin Lu
College of Art, Agricultural University of Hebei, Baoding071001, Hebei, China

Abstract:
This paper proposes an architectural scene construction model based on adaptive error correction RANSAC (random sample consensus) algorithm. Firstly, SIFT algorithm is used to perform initial matching on the feature point of two frames in architecture scene. Secondly, extreme value detection on DoG space is conducted to locate the position of the feature point and its location scale. Thirdly, the principal curvature is calculated by using Hessian matrix to remove the unstable marginal answering point. Finally, RANSAC algorithm is used to remove the mismatching point, then, the adaptive error is corrected through polar geometric principle so that the accuracy of RANSAC algorithm can be improved. Simulations show the improved model proposed in this paper has better anti-noise performance and higher point-matching accuracy.

Key words: Adaptive Error Correction, RANSAC Algorithm, Mismatching Point Removal, Polar Geometric Principle, Marginal Answering Point, Architectural Scene Reconstruction

1. INTRODUCTION

In recent years, with the development of Internet technology and computer vision theory, threedimensional(3D) reconstruction has become an important research topic in the field of computer vision. Because of the deficiency of manually construction of 3D model, 3D reconstruction technology (Furukawa and Ponce, 2010) has emerged. In the meantime, virtual reality technology, computer graphics, engineering modeling and other related fields have also become the hot research topics (Shotton, Winn, Rother, et al., 2012). The 3D reconstruction technology is a 3D model reconstruction process by using the existed modeling related data material or graphics (Schnabel, Degener and Klein, 2009).

In the past decades, experts have devoted more and more efforts into the field of 3D construction, and many advanced application systems and softwares have appeared. Debevec and others (Debevec,Taylor and MaKK, 2016) presented a facade 3D reconstruction system, which is mainly applicable to architectural scenes. Before the system is operated, simplex geometric model should be established and the camera parameters should be given in advance. If condition is satisfied, we compare the anti-projection model with the real scene, and reduce the error of anti-projection to obtain the precise 3D model. Faugeras and others (Faugeras, Robert, Laveau, et al., 2008) combined self-calibration method and hierarchical thinking to reconstruct model from sequence images. This system presents the final reconstructing model with polyhedron. Then, combining the architectural styles, the authors realized the calibration of the camera by using the physical information such as 3D point coordinates and angles of the model. Pollefeys and others (Pollefeys, Koch and Van Gool, 2009) proposed an automatic reconstruction system of the surface, which can combine camera self-calibration techniques with the continuous-shooting sequential images of the objects to accomplish the 3D reconstruction. Furukawa and others (Furukawa and Ponce, 2010) put forward a dense 3D modeling reconstruction algorithm, which can present the surface of the object through a set of the facets. It is simple and effective and do not need any initial information. It is a classical multiple-perspective dimensional reconstructing algorithm on the basis of surface patch. Jiang and others (Jiang and Peng, 2008) developed a real-time rendering system based on visual shell. The system mainly combines the modeling method based on the outline with marching cubes method based on voxel to extract object surface model and scene reflection parameters. Then, we can obtain the diffuse reflection of the 3D point of the model. Wang Wenhai (Wang, 2014) proposed a dynamic 3D reconstruction system, which mainly makes use of fringe projection and Fourier transformation to realize the measure of 3D surface. The dynamic reconstruction of the object can be accomplished precisely only by one image. Liu and others (Liu, Peng and Bao, 2015) discovered a 3D modeling system based on scene interaction. The input of this system are multiple images shot by one moving camera. By means of interaction tool, the user can simply interact with the scene, and finally the 3D model with texture can be obtained.

In order to satisfy the requirement of the 3D architectural scene reconstruction, this paper proposes an architectural scene reconstruction model based on adaptive error correction RANSAC algorithm.
2. INITIAL FEATURE MATCHING BASED ON ANTI-NOISE OPTIMIZATION SIFT ALGORITHM

Generally, the flexible and efficient self-calibration method is adopted in the sparse reconstruction phase of facade modeling. Therefore, image matching is performed to find the view correspondence in the sequence. SIFT algorithm (Fu, Wang, and Gao, 2016) is used to perform initial matching of the feature points on the two frames of the architectural scene. The detailed processes are as follows:

(1) Extreme value detection of the scale space

Extreme value detection of the scale space is to find the potential extreme point, which will remain unchanged with the change of scale in the whole scale and figure. In order to improve the speed of extreme value detection, we choose to perform extreme value detection on DoG space to locate the position of the feature point and its scale.

\[
L(x, y, \delta) = G(x, y, \delta) \cdot I(x, y)
\]

(1)

where \( G(x, y, \delta) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+\delta^2+y^2)}{2\sigma^2}} \) is Gaussian convolution kernel, \( I(x, y) \) is gray level image, \( \sigma \) is scale factor. And the larger the value, the smoother the image is, and vice versa. \( L(x, y, \delta) \) is Gaussian image, and we call a series of Gaussian images produced by different multiples \( \sigma \) as Gaussian pyramid (Mu, Zhao, and Hui, 2015).

\[
D(x, y, \delta) = L(x, y, k\delta) - L(x, y, \delta)
\]

(2)

DoG pyramid is formed through the subtraction of the adjacent scale Gaussian images in the Gaussian pyramid, which is shown in figure 1.

\[ \text{Figure 1 DoG pyramid model} \]

As shown in figure 2, the extreme value detection process is: we compare the point with total 26 pixels, which are the adjacent 8 pixels in the same layer, the adjacent 9 pixels in the upper layer, and the adjacent 9 pixels in the lower layer. If the point is bigger or smaller than the other 26 points, then we consider this point as local extreme value, which are called feature candidate point and should be marked down its location and scale.

\[ \text{Figure 2 Extreme point detection} \]
(2) Precise extreme point

The marginal point is unstable and sensitive to noise. And the point with low contrast is also sensitive to noise. Therefore, the extreme point which was previously detected needs to be further tested to be precisely located as feature point. Then, anti-noise can be achieved.

1) Low-contrast point removal

The local extreme value is performed 3D quadratic function fitting to locate the position of the feature point and its scale. Taylor expansion of scaling space function $D(x, y, \delta)$ at local minimums $(x_0, y_0, \delta_0)$ is shown as follows:

$$D(x, y, \delta) = D(x_0, y_0, \delta_0) + \frac{\partial D^T}{\partial x} X + \frac{1}{2} X^T \frac{\partial^2 D^T}{\partial x^2} X$$

where $X = (x, y, \delta)^T$ is offset

$$\frac{\partial^2 D}{\partial \sigma^2} = \left(\frac{D_{xx}^{\alpha} - D_{xy}^{\alpha}}{4}\right) - \left(\frac{D_{yx}^{\alpha} - D_{yy}^{\alpha}}{4}\right)$$

$$\frac{\partial^2 D}{\partial x \sigma} = \left(\frac{D_{xx}^{\alpha} - D_{xy}^{\alpha}}{4}\right) - \left(\frac{D_{yx}^{\alpha} - D_{yy}^{\alpha}}{4}\right)$$

And so on, the first and second order derivative of equation(3) is obtained by difference approximation in nearby space. The derivative of equation(3) is carried out and let it be 0, then, we can obtain the precise extreme point $\hat{X}$, which is shown in equation(6):

$$\hat{X} = -\left(\frac{\partial^2 D}{\partial x^2}\right)^{-1} \frac{\partial D}{\partial x}$$

If the value of $\hat{X}$ is bigger than 0.5 at any direction, this point is not the extreme point. When the extreme point changes, we can use interpolation to replace sample point.

Equation (6) is inserted into equation (3) and we take the first two terms, then, the following is obtained:

$$D(\hat{X}) = D + \frac{1}{2} \frac{\partial D^T}{\partial x} X$$

If $|D(\hat{X})| \geq 0.03$, the point remains, otherwise it can be discarded.

2) The unstable marginal answering point removal

In order to ensure the stability of the feature point, it is not enough to just remove low-contrast points. As the marginal extreme points are very sensitive to noise, these points also should be removed. The peak value of Gaussian difference have a large principal curvature along the margin, but in the vertical margin direction, the curvature is small. We calculate the principal curvature by using the 2x2 Hessian matrix, which is shown in equation(8). The partial derivative in this equation is the partial derivative in the feature point determined previously, which is also approximately estimated by the difference in the nearby space.

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix}$$

The principal curvature of $D$ and the feature value of $H$ are in direct proportion relationship. Suppose $\alpha$ is the biggest feature value, $\beta$ is the second biggest feature value. What we concern is the ratio of the two, their values do not need to be calculated.

Suppose $\gamma = \frac{\alpha}{\beta}$, then the following equations hold:

$$\text{Det}(H) = D_{xx} D_{yy} - D_{xy} D_{yx} = \alpha \beta$$

$$\text{ratio} = \frac{\text{Tr}(H)^2}{\text{Det}(H)} = \frac{(\alpha + \beta)^2}{\alpha \beta} = \frac{(\gamma + 1)^2}{\gamma}$$
Set $\gamma = 10$. If $\text{ratio} \geq \left(\frac{\gamma + 1}{\gamma}\right)^2$, we consider this point as a marginal point, which should be removed. Otherwise it is remained.

3) Principle distribution direction

The gradient direction distribution characteristic of the pixel is considered as principle direction of each feature point.

$$m(x, y) = \sqrt{((L(x+1, y) - L(x-1, y)) + (L(x, y+1) - L(x, y-1))^2}}$$

$$\theta(x, y) = \tan^{-1}\left(\frac{L(x+1, y) - L(x-1, y)}{L(x, y+1) - L(x, y-1)}\right)$$

Equation (11) and equation (12) are gradient value and gradient direction in $(x, y)$, respectively. $L$ is Gaussian image which is closest to the scale of the feature point. Gradient direction histogram is adopted to express each gradient direction of the adjacent different pixels. The range of gradient direction histogram is $0^\circ$~$360^\circ$, there is a pillar every $10^\circ$, and total is 36 pillars. We regard the peak of histogram in gradient direction as the main direction of the feature point in neighborhood gradient. In gradient direction histogram, when there is a local peak of $80\%$ energy of main peak, this direction is treated as the auxiliary direction of the feature point. Only $15\%$ feature points have multiple directions. Obviously, the stability of matching is improved.

3. ARCHITECTURAL SCENE RECONSTRUCTION BASED ON ADAPTIVE ERROR CORRECTION RANSAC ALGORITHM

3.1. Mismatching Removal Based on Adaptive Error Correction RANSAC

The improvement of self-calibration accuracy of the camera is one of the key points for the quality of sparse reconstruction. Removal of the mismatching point in the initial matching can greatly improve the accuracy of self-calibration. This paper combines RANSAC method with polar geometric principle to remove mismatching point.

RANSAC is a method belongs to robust principle. The method estimates the model with the least observation points. And the observation distance from observation point to the estimated model can be calculated. Then, a distance threshold is set to judge whether the observation distance is smaller than the threshold. If it is, then we call it as interior point, otherwise it is exterior point. The program is performed Loop execution, and finally we adopt the estimation of observation points with the most interior points as the final result.

We can see from figure 3 that, splattering robust is estimated as a straight line, two points are randomly selected to fit the initial straight line $l$, and the distance between all the points and $l$ is calculated. Set distance threshold, and let interior point set be $s(l)$. Then, randomly cycles are carried out a couple of times, we can obtain straight lines $l_1, l_2, ..., l_n$ and their corresponding interior points $S(l_1), S(l_2), ..., S(l_n)$. The fitting of biggest interior points set is selected as the final result.

Figure 3 RANSAC estimation straight line
According to the feature of the adjacent view polar geometry, the primary matrix of the two views are estimated by RANSAC method to remove the mismatching point. This process has six steps:

1. 8 groups of corresponding points of the two views are randomly selected, and these points are used to estimate primary matrix $F$.

2. Denote one group of corresponding point as $m <\rightarrow m'$, $m$ and $m'$ are on image $I_1$ and $I_2$, respectively. According to the feature of the primary matrix, we use $F$ to calculate epipolar $l'$ which is $m$ relative to $I_1$ and epipolar $l$ which is $m'$ relative to $I_1$.

3. The distances between $m$, $m'$ and epipolars $l$, $l'$ are calculated, which are $d(m,F^T m')$, $d(m',Fm')$, respectively.

4. Let the threshold be $T$, if the matching point $m <\rightarrow m'$ satisfies equation(13), we mark it as interior point:

$$\begin{cases} d(m,F^T m') < T \\ d(m',Fm') < T \end{cases}$$

(13)

Of all the matching points, we obtain the $i$-th estimated interior point set $S_i$.

Loop executions from (1) to (4) are performed until iteration times are bigger than sampling numbers $K$ in equation (14). And we use the biggest interior point to re-estimate the primary matrix $F'$.

$$K = E[K] + 3SD[K]$$

$$= \frac{1 + 3\sqrt{1 - w^n}}{w^n}$$

(14)

where $n$ is the sampling point, $w$ is interior point proportion.

Let the threshold be $T'$, $F'$ is used to calculate the distance between all the matching points and the corresponding epipolars. The interior point set is selected out and retained, the others should be removed as mismatching point.

### 3.2. 3D Architectural Scene Reconstruction

First of all, the image sequences are verified and numbered. Then we use the optimal SIFT algorithm to detect the feature point of every view. These feature points may lie in location where the colors and outlines could change severely, such as the ridges and corners. We restore the 3D coordinates of 3D feature points corresponding to 2D feature points. It can be realized through the following 6 steps:

1. Sequence $I$ is divided into several successive views. It is guaranteed that the first and last view in each group have more than 50% overlapped scenes. Pair matching of the 2D feature points of every view in the group are performed and the matching points presenting the same space are obtained.

2. We establish the correspondence with the matching point of every three neighboring image $I_i, I_{i+1}, I_{i+2}$, and the matching-point pair that presents the same space are obtained.

3. The camera parameters $P_i, P_{i+1}, P_{i+2}$ of $I_i, I_{i+1}, I_{i+2}$ based on the corresponding points and the 3D feature points $S(I_i, I_{i+1}, I_{i+2})$ are calculated.

4. Views $I_{i+3}, I_{i+4}$ are added. Step (2)-(3) are repeated to $I_{i+1}, I_{i+3}, I_{i+4}$ to calculate 3D feature points $S(I_{i+2}, I_{i+3}, I_{i+4})$.

5. Repeat step(4) until the sequence finishes, then we can obtain the initial 3D feature point set $S(I)$ of $I$.

6. The coordinate optimization is carried out by using the adaptive error corrective RANSAC algorithm. And the abnormal point and remote point with large deviations are removed, which can make the coordinates concentrate on the architectural vertical area.

### 4. SIMULATIONS OF THE ALGORITHM

In this paper, simulations and analysis are performed on the feature matching of scene image sequences.

The operating system is Windows 7, processor is Intel Core i7-3630QM@2, CPU is 40GHz, memory is 4G RAM. And it also has NVIDIA GeForce GT 645M discrete video card. The experiment is carried out under VS2010 with VC++. The source figure are two pictures of architectural scenes waiting for matching, which is shown in figure 4. The original size of the figure is 1024×768.
Figure 4  Scenes waiting for matching

(1) Extraction of figure feature points and experiment of matching algorithm

Figure 5 and figure 6 are the matching results of the standard SIFT algorithm and the improved SIFT algorithm, respectively. Table 1 is the comparison of the average time of 50 times executions between the two algorithms when the two algorithms deal with different-multiples source figures.

Figure 5  Matching results of the standard SIFT Algorithm

Figure 6  Matching results of the improved SIFT algorithm

From the matching process of the above two sets of pictures, we can find that, both the standard SIFT algorithm and the improved SIFT algorithm can extract abundant and stable feature points. After comparing the matching results of figure 5 and figure 6, we can find that the improved SIFT algorithm can obtain more feature
points than those of the standard SIFT algorithm. Because the improved algorithm uses different-scale figures and Gaussian-function as convolution and also has many scanning directions, it can better calculate the scales and the shifting angles, and has larger matching scale.

<table>
<thead>
<tr>
<th>Image size</th>
<th>SIFT</th>
<th>Improved SIFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full-scale</td>
<td>15.42</td>
<td>7.92</td>
</tr>
<tr>
<td>Zoom 1 times</td>
<td>31.57</td>
<td>16.81</td>
</tr>
<tr>
<td>Zoom 0.5 times</td>
<td>21.24</td>
<td>12.74</td>
</tr>
<tr>
<td>Narrow 1 times</td>
<td>10.19</td>
<td>7.29</td>
</tr>
<tr>
<td>Narrow 0.5 times</td>
<td>7.63</td>
<td>3.26</td>
</tr>
</tbody>
</table>

From table 1, we can find that, the speed of the improved algorithm is 2 times faster than the standard one. As the improved SIFT uses integral image techniques and Hessian matrix, which can reduce the sampling times. And its feature vector dimension is less than standard SIFT, the speed of the matching can be also improved.

(2) Experiment of mismatching removal

The matching results of figure 6 through the standard RANSAC algorithm and the improved RANSAC algorithm are used to perform mismatching removal. The experiments repeat 20 times, and the average is used to obtain the results, which are shown in figure 7 and 8. The comparisons of the average matching number and the average accurate matching rate before and after the experiment is shown in table2.

![Figure 7](image1.png)  
**Figure 7** Error elimination result of Standard RANSAC algorithm

![Figure 8](image2.png)  
**Figure 8** Error elimination result of the Improved RANSAC algorithm

<table>
<thead>
<tr>
<th></th>
<th>RANSAC Before</th>
<th>RANSAC After</th>
<th>Improved RANSAC Before</th>
<th>Improved RANSAC After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Match logarithmic</td>
<td>51</td>
<td>48</td>
<td>51</td>
<td>42</td>
</tr>
<tr>
<td>Correct number/ %</td>
<td>74.2</td>
<td>81.3</td>
<td>74.2</td>
<td>93.6</td>
</tr>
</tbody>
</table>

The experiments above show that, the matching point obtained based on the matching algorithm can satisfy the requirement of polar geometry, and the related RANSAC algorithm is reliable to eliminate the mismatching.

5. CONCLUSIONS

3D reconstruction technique is to extract the spatial information of the scene or scene objects from 2D images. According to the obtained information, 3D structure model can be restored. Currently, it has become the hot research topic in the field of computer graphics, computer animation, virtual reality, scientific calculation and digital media. This paper proposes the architectural scene reconstruction model based on adaptive error
correction RANSAC algorithm. Simulations show that the improved model has better anti-noise performance and higher matching accuracy compared with the standard algorithm.

REFERENCES


