Method of Key-Frame Selection for Video Images

Xiaowei Han, Zhiyong Wang, Wenjun Yang and Yinghong Xie

ABSTRACT

In order to improve the quality and efficiency of image processing and 3D reconstruction, a key frame screening method for video image is proposed. Firstly, the image degradation is performed by the normalized image sharpness evaluation function based on the image degradation caused by camera motion and key frame segmentation. Secondly, in order to avoid the problem of low accuracy and motion degradation of triangulation, the geometric robust information criterion GRIC was used as the screening condition to screen the results of shallow screening. The experimental results show that the three-dimensional reconstruction of the selected image sequence as the source image reduces the number of feature points, shortens the time of feature matching, and improves the efficiency of 3D reconstruction by 3.6 times.

INTRODUCTION

UAV aerial video image key frame sequence screening is the video image stitching and three-dimensional reconstruction with high accuracy and real-time key[1]. If all the frames in the video image sequence are used for 3D reconstruction, it will take a lot of time and the improvement of the reconstruction effect will be minimal[2]. And if the baseline between the two adjacent frames of the video image sequence is too small, the accuracy of the triangulation will be reduced, and too large will increase the difficulty of feature matching[3].

Xiaowei Han, Zhiyong Wang, Yinghong Xie, Electronic Information Engineering of Shenyang University, Shenyang, China
Wenjun Yang, Xi'an Jiaotong University, Xi'an, China
In this paper, we summarize the shortcomings of several key frame extraction algorithms, and propose a hierarchical filtering algorithm for video image key frame, which divides the key frame into shallow screening and deep screening. Shallow screening factors are image clarity, the speed of the aircraft and the number of camera frames. Shallow screening of the key frame, and then according to the key frame overlap, the width of the baseline and other factors that can affect the three-dimensional reconstruction of the effect of deep screening.

**SHALLOW SCREENING**

Nowadays, the flying platform is stable and the speed control is accurate. The collected data has good effect in illumination and contrast[4]. Therefore, the shallow screening mainly considers the image degradation caused by camera movement and key frame segmentation. Since there is no a priori reference image in the process of filtering keyframes, this problem can be classified as a non-reference image clarity evaluation problem.

Shallow Screening. The image clarity evaluation function is used to filter the images of all the frames of the video. After normalizing the image clarity evaluation value, a peak is generated at a similar time interval, which represents the video image of the frame’s highest clarity. The frame image is output as a result of shallow screening. Image clarity evaluation flow chart shown in Figure 1.

![Image clarity evaluation flow chart](image)

**DEEP SCREENING**

After the shallow screening, in order to further improve the efficiency of the establishment of the three-dimensional model, to avoid the triangular measurement accuracy and motion degradation due to the small baseline, the feature matching degree M and the geometric robust information criterion GRIC (Geometric robust information criterion)[5] as the screening condition Screening results for deep screening.

The feature matching degree M is the ratio of the feature points matching the reference image pair matching the matching image to the reference image.
Where \( U \) is the number of feature points matching the reference image and the reference image, and \( T \) is the number of feature points matching the reference image pair. The closer the value of \( M \) is close to 1, the more the number of feature points matching the reference image to the reference image is closer to the number of matching feature points of the reference image, then the more overlapping area of the two images is, the smaller the camera motion. Due to the degradation of motion between the two key frames, the degradation of the structure will lead to a large computational error in the base matrix, and the base matrix is an important basis for camera estimation and 3D reconstruction. Therefore, GRIC is used to distinguish between general phenomena and motion degradation. The GRIC measure is defined as:

\[
GRIC = \sum_i \rho(e_i^2) + \lambda_1 dn + \lambda_2 k
\]

\[
\rho(e_i^2) = \min(\frac{e_i^2}{\sigma^2}, \lambda_3 (r - d))
\]

Where \( e_i \) is the retention number, \( \lambda_1 = \ln r \), \( \lambda_2 = \ln(m) \), \( \lambda_3 \) is the value of the limit residue. \( n \) denotes the number of two image matching points, \( k \) is the degree of freedom of the model, the single matrix model \( k = 7 \), the characteristic matrix model \( k = 8 \). \( \sigma^2 \) denotes the error variance, \( r \) denotes the dimension of the data, for the two-dimensional matching point \( r = 4 \), \( d \) denotes the model dimension, \( d = 3 \) for the plane model and \( d = 2 \) for the feature matrix model.

For the video frame key frames to be matched, the GRIC values of the base matrix \( F \) and the single matrix \( M \) are calculated separately. If \( GRIC_F < GRIC_M \), the two frame images are considered to have no degradation relation, so the frame is used as a candidate key frame.

In the deep screening of the video image, the video image frame is selected in consideration of the above two screening factors.

**EXPERIMENTAL RESULTS**

Experimental data for UAV aerial video images. Unmanned aerial vehicles to capture the experimental video image resolution of 3840 * 2160, frame rate of 30 frames per second, UAV flight speed of 6M / S. Select a length of 32s aerial video images to the key frame screening method to experiment.
After shallow screening and deep screening, a total of 16 images were obtained. The reference image adopts the key frame sampling and extraction algorithm. The sampling time interval is 0.5S, and 64 images are collected. The 3D reconstruction effect comparison chart is shown in figure 2, figure 3 and table 1.

![Image](image.png)

Table 2. Reference image splicing effect Figure 3 screened image mosaic effect map.

<table>
<thead>
<tr>
<th>Image</th>
<th>Number of images</th>
<th>Feature extraction time</th>
<th>Feature matching time</th>
<th>Dense point cloud number</th>
<th>Efficiency improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference image</td>
<td>64</td>
<td>367</td>
<td>58</td>
<td>186733</td>
<td>1</td>
</tr>
<tr>
<td>Filter the image</td>
<td>16</td>
<td>104</td>
<td>13</td>
<td>167322</td>
<td>3.63</td>
</tr>
</tbody>
</table>

**CONCLUSIONS**

Based on the problem of image sharpness degradation, low triangular measurement accuracy and image motion degradation, this paper makes a shallow screening and deep screening of video image sequences by normalized image sharpness algorithm and geometric robust information criterion GRIC, which not only improves the image Three-dimensional reconstruction of the quality, but also the image splicing efficiency increased by nearly 3.6 times. Lay the foundation for real-time stitching of 3D images.

**ACKNOWLEDGEMENTS**

This work was financially supported by Shenyang Program for Tackling Key Problems in Science and Technology (F15-126-9-00).

**REFERENCES**

