Research on Property of Image Matching Method Based on the Deep Fusion of Harris and SIFT

Qingxi Zeng\textsuperscript{a,b},*, Shan Ma\textsuperscript{a}, Yupeng Feng\textsuperscript{a}, Xiao Fang\textsuperscript{b}

\textsuperscript{a}Department of Vehicle Engineering Nanjing University of Aeronautics and Astronautics, NO.29 Yudao Road Qinhuai Distric Nanjing Jiangsu Province, Nanjing and 210016, China
\textsuperscript{b}Chery Automobile Co., Ltd.,NO.35 Hengshan Road Economic & Technological Develop Zone Wuhu Anhui Province, Wuhu and 241009, China

*Qingxi Zeng: jslyzqx@nuaa.edu.cn

Abstract

In order to meet the requirements of real time and saving resources for the visual odometer in driverless cars, a method combining Harris and SIFT by taking advantage of their respective features is proposed. First, Harris operator is used to detect corner points which were regarded as the candidate feature points of an image. And then SIFT algorithm is used to extract features from those corners. Finally, Random Sample Consensus (RANSAC) is used to remove the wrong matching points. Results show that the proposed method has reduced the number of features to be detected, made the algorithm simpler and extracted features faster while good matching efficiency is maintained.

Keywords: Autonomous driving, image matching, SIFT, Harris.

1. Introduction

As an important branch in the field of image, the purpose of image matching is to find the same feature points in the same scene with overlapping images and to match the feature points. It is widely used in predicting the position of unmanned vehicles in an unknown environment. At present, two widely used matching methods are based on either area or feature [1]. The method based on feature matching has small amount of calculation and good robustness and some features are not sensitive to image rotation, scale changes, illumination and noise, making it a hot topic in today’s studies of image matching method.

The method based on feature matching consists of two steps: feature detection and feature matching. Harris corner detection [2] and SIFT algorithm [3, 4] are two feature detection and matching algorithms more effective and commonly used than other algorithms, but they have their own advantages and disadvantages. SIFT feature is the local feature of images, which is invariant to rotation, scale and illumination changes, and maintains stability to some degree despite perspective changes, affine transformation and noises. But the algorithm has high computational complexity and low efficiency because of a large amount of convolution operations and histogram statistics. Harris algorithm is an effective corner detection algorithm, which has less amount of calculation and extracts well-distributed and proper feature points. It is also not sensitive to rotation and noise. But it does not have scale invariance and the corner detected is not ideal for matching [7,8]. So this paper presents an algorithm integrating the Harris and the SIFT based on the advantages and disadvantages of the above two algorithms.

2. Harris-SIFT Algorithm

2.1 The flow diagram of algorithm

According to the analysis above, this paper presents an algorithm integrating the Harris and the SIFT. The flow diagram of algorithm is shown in Fig. 1.

2.2 Harris corner detection

Harris algorithm is based on Moravec algorithm. Corner points are detected by means of differential and auto correlation functions. Its principle is to move the window centered on the target pixel to any directions, calculate
changes of the gray scale before and after moving, and then decide whether the pixel is a corner. If the changes in both the horizontal and vertical directions are big, the pixel is considered as a corner.

The window function of Harris algorithm is show as (1), which is designed to eliminate the isolated points in the image, and avoid these points are mistaken for the corner. The closer the pixel is to the center, the larger its weight.

\[
w(x, y) = \exp\left[-\frac{(x^2 + y^2)}{2\sigma^2}\right]
\] (1)

The window centered on pixel \((x, y)\) have small mobile \(u\) and \(v\) in \(x\) direction and \(y\) direction, the amount of gray level transformation computation formula is as follows:

\[
E(u, v) = \sum_{x} \left[I(x + u, y + v) - I(x, y)\right] ^ 2
\]

\[
= \sum_{x} w(x, y) \left[I(x + u, y + v) - I(x, y)\right] ^ 2
\]

\[
= \sum_{x} w(x, y) \begin{bmatrix}
I_{xx} & I_{xy} \\
I_{xy} & I_{yy}
\end{bmatrix} \begin{bmatrix}
u \\
v
\end{bmatrix}^2
\] (2)

The autocorrelation function \(M\) of a pixel at the location \((x, y)\) is given as follows:

\[
M = \sum_{x} w(x, y) \begin{bmatrix}
I_{xx} & I_{xy} \\
I_{xy} & I_{yy}
\end{bmatrix}
\] (3)

The corner identification can be measured by the corner response function \(R\), which can be represented as:

\[
R = \det(M) - ktr^2(M)
\] (4)

Where \(\det\) and \(tr\) is the determinant and trace of autocorrelation matrix \(M\), \(k\) is a parameter whose values range is \([0.04, 0.06]\). If \(R\) of of a pixel at the location \((x, y)\) is local maximum within a 8 size window (non-maximum suppression), then the pixels \((x, y)\) can be considered as a corner.

2.3 Harris-SIFT feature point extract

The corner is an important feature in image processing. Harris corners are often have high stability and repeatability. The SIFT features, which are selected in Harris corner points, have the characteristics of both. Specific steps are as follows:

(a) Build local Gaussian pyramid and DoG pyramid

A local Gaussian pyramid and a DoG pyramid are established in the 16*16 neighborhood of Harris corners instead of in the whole picture, which can greatly reduce the number of pixels of the calculation and the complex computation algorithm. The algorithm is using the Gaussian kernel convolute with the neighborhood of Harris corners. Multi-scale space is obtained by the change of the parameter sigma in Gaussian kernels. The computation formulas of Gaussian pyramid and DoG pyramid are given as follow:

\[
L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)
\] (5)

\[
G(x, y, \sigma) = \frac{1}{2 \pi \sigma^2} e^{-\frac{(x^2 + y^2)}{2\sigma^2}}
\] (6)

\[
D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)
\] (7)

(b) Feature point detection in local scale space

In order to detect the local extreme point of DoG scale-space, each sample points is compared to its neighbors (see Fig.2). To ensure detect extreme point in scale space and 2D image space, the point to be detected should be compared to its eight neighbors in the current image and nine neighbors in the scale above and below. It can be considered as a feature point if it is larger than all of these neighbors or smaller than all of them. If the scale space has extreme value point, then keep the angular point;
Otherwise, delete the angular point.

c) Assign a major orientation to each key point

Before determined the major orientation of key point, the gradient magnitude and orientation, is precomputed as following:

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

(8)

\[
\theta(x, y) = \alpha \tan 2 \frac{L(x, y+1)-L(x, y-1)}{L(x+1, y)-L(x-1, y)}
\]

(9)

where \( L \) is the scale of the key point. In actual calculation, we sample in the neighborhood window centered of the key point and count the gradient direction. The range of gradient histogram is 0 ~ 360 degrees. There are either eight columns and each column is of 45 degrees, or there are 36 columns and each column is of 10 degrees. The peak of the histogram is the main direction of the key points in the neighborhood gradient, which is the direction of the key points.

d) Descriptor Representation

As Harris corners are rotation invariant, a 16*16 region around the key point is selected, which is shown in Fig. 3. The region is further divided by a size of 4*4 into 4*4 areas. The gradient histogram of the eight directions in each area is calculated to get 4*4*8=128 element feature vectors as the SIFT descriptor.

![Image](Image)  
Fig. 3. 16 x16 area around the key point.

e) Feature point matching

The best candidate match for each key point is found by comparing the similarity measure between key points from two images. However, many features from an image have more than one matching point. So in order to determine which one is the best match, the mutual consistency is used to check matching points. The Euclidean distance between points of the image is matched and the reference image computed, and then they are used to determine the similarity of the feature points in the two images. The pair features with the shortest Euclidean distance in the two images are considered the best candidate match. And finally, RANSAC is used to remove the wrong matching points.

3. Results

The experiments are conducted with 2012b version of MATLAB on an Intel-Core-i5 PC with 4G memory and Windows 7 system. The images to be matched are 640 *480 in resolution and taken with the same camera from different angles are shown as in Fig. 4.

![Images](a) (b)

Fig. 4. The images to be matched

First, experiments on MATLAB are conducted to verify efficiency of the proposed algorithm over SIFT algorithm. Features of the images in Fig. 4 are detected using SIFT algorithm and the proposed algorithm in succession. The number of features extracted and the running time of these algorithms are shown in Table 1. As can be seen in Table 1, the processing time of the proposed algorithm is significantly less than the SIFT algorithm. The reason is that the proposed algorithm detect features only in the neighborhood of Harris corners instead of in the whole picture.

<table>
<thead>
<tr>
<th>Image</th>
<th>Number of feature points</th>
<th>Time of feature points detection/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig.4(a)</td>
<td>SIFT: 372, Harris-SIFT: 115</td>
<td>SIFT: 5.638, Harris-SIFT: 2.524</td>
</tr>
<tr>
<td>Fig.4(b)</td>
<td>SIFT: 416, Harris-SIFT: 150</td>
<td>SIFT: 4.005, Harris-SIFT: 2.426</td>
</tr>
</tbody>
</table>

Then the experiment shows that the proposed algorithm have greatly enhanced the matching speed while matching enough features and maintaining high correct matching rate. The computation formula of the matching precision is as follow:

\[
\text{matching precision} = \frac{\text{number of matching after eliminate outliers}}{\text{number of matching}} 
\]

using RANSAC/number of matching
Feature points in Fig. 4 are matched using the SIFT algorithm and the proposed algorithm based on the experimental environment, and their respective number of feature points, the matching precision and time for features matching are shown in Table 2. As can be seen in Table 1, the number of features matching obtained by the SIFT algorithm is 184 pairs, the matching precision is 84.24%. The matching result is shown as Fig. 5. And the number of features matching obtained by the proposed algorithm is 52 pairs, the matching precision is 82.69%. The matching result is shown as Fig. 6. Although less features are matched, matching precision slightly lower than the SIFT algorithm by the proposed algorithm, they are enough to calculate the camera pose. Moreover, the proposed algorithm consumes 65% less time on average in matching images than the SIFT algorithm in 100 simulation experiments.

Table 2. Comparison of feature matching.

<table>
<thead>
<tr>
<th>Method</th>
<th>Matches</th>
<th>Correct matches</th>
<th>Percent of correct matches/%</th>
<th>Total time/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIFT</td>
<td>184</td>
<td>155</td>
<td>84.24</td>
<td>10.112</td>
</tr>
<tr>
<td>Harris-SIFT</td>
<td>52</td>
<td>44</td>
<td>82.69</td>
<td>5.856</td>
</tr>
</tbody>
</table>

Fig. 5. Matching image by SIFT algorithm.

Fig. 6. Matching image by the proposed algorithm.

4. Conclusions

In this paper, we present a fast and robust image matching method integrating Harris algorithm and SIFT algorithm. First, the Harris corners are extracted from the original scale image. Then SIFT features are detected in the neighborhood of the Harris corners before describing them with SIFT descriptors. Finally, image matching is realized with a small amount of features. The features extracted using the proposed algorithm have the advantages of both Harris and SIFT. It has not only improved extracting speed of feature points, but also achieved a good matching efficiency, and it is robust for image matching.

Acknowledgment

This research was financially supported by the Open Fund of Nanjing University of Aeronautics and Astronautics graduate innovation base (Laboratory) (kfjj20160216) and the National Natural Science Foundation of China (51505221).

References