2012 Fifth International Conference on Intelligent Networks and Intelligent Systems

Characteristic Point Match Algorithm Based on the SURF in Binocular Stereo Vision

Wang Xiaoli[1] Yang Lei[2]
1.College of Electronic Information and Engineering
Changchun University
Changchun, China
e-mail: 445444726@qq.com

Wang Lirong[1*] Xu Jing[1]
2. College of Electronic Information and Engineering
Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences
Changchun, China
Correspondingauthor:wlr10012003@yahoo.com.cn
（Wang Lirong）

***Abstract*—One of the main components of the computer vision is 3D reconstruction. Extracts and the match based on two-dimensional picture characteristic point is the 3D reconstruction technology core. Taking binocular stereoscopic vision theory as afoundation, through extracts the characteristics of SURF based on the multi-scale analysis, this feature has specific scale reproducibility, put the epipolar constraints and disparity constraints as conditions judgement to specific screening, it is greatly reduced the search range. The experiment use the matching method for three different actual scenes, the results shows that the method can enhance the image matching speed and precision, and can get more precise dense parallax, used for reconstruction complete scene.**

Keywords- binocular stereo vision; 3D reconstruction;SURF; epipolar constraints; disparity constraints

I. INTRODUCTION

In computer vision navigation system[1], the collected visual signals are two-dimensional, In reality, the visual signals are three dimension ones which can not be get through the vision sensor, and the real world is three dimensional, we can't obtain three-dimensional spatial coordinates of the real world directly through the vision sensor. Therefore, the 3D space information can be reconstructed by 2D signals collected by visual sensor, It is the necessary step of scence analysis in stereo vision environment. This step usually consists of six modules: image acquisition, camera calibration, feature extraction, stereo matching, three-dimensional recovery and analysis of video.

Based on the theory of binocular stereo vision[2], the fantastic binocular stereo vision platform is established first to aquiring image and calibrating camera. shown in Figure 1:

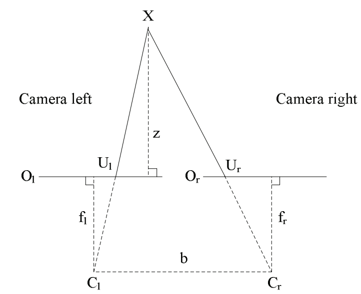

Figure 1. The fantastic binocular stereo vision platform

In Figure 1, the target point is X, the subscript l means left, r for right, U for imaging abscissa, f for the focal length, C for the optical center, b for the baseline. The rectifiable depth z can be get according to the similarity triangle:

$$z = \frac{bf}{|u_l - u_r|} \tag{1}$$

Where, $|u_l - u_r|$ is often referred to as disparity, indicated by d.

Feature extraction and matching are the most important and difficult problems in stereo vision and also are prerequisites for recovery three-dimensional information. It has been the focus of the study of stereoscopic vision field all the time. This extraction SURF characteristics based on multi-scale analysis, this feature has a specific scale reproducebility, and the epipolar constraint and disparity constraint as conditions to determine specific screening, reduces the search range greatly, improve the speed of image matching and precision, and thus get a more accurate dense disparity used to reconstruct a complete scene.

II. THE IDEAS AND STEPS OF SURF MATCHING ALGORITHM

SURF[3](Speed Up Robust Features) is proposed by Herbert Bay in 2006 which is based on fast robust feature matching algorithm. The algorithm mainly includes two parts of the feature point detection and feature points description. Not only the calculation speed has been greatly improved but the anti-scale changes, rotation, illumination, affine and geometric transformations have better performance by the algorithm. SURF feature extraction

978-0-7695-4855-5/12 $26.00 © 2012 IEEE

DOI 10.1109/ICINIS.2012.72

process are divided into two parts, namely feature point detection and description.

A. Feature Point Detection

The procedure of SURF feature detection is as follows:

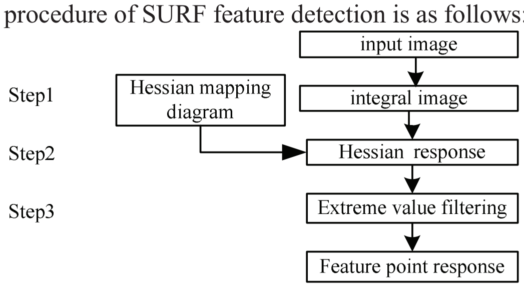

Figure 2. The procedure of SURF features' detection

Step 1: Calculate the integral of the input image image[4] and Hessian map.The mathematical form of the integral image is the double integral of the original image, the expression is:

$$I_{\Sigma}(x,y)=\sum_{i=0}^{i\le x}\sum_{j=0}^{j\le y}I(x,y) \tag{2}$$

From equation (2), any rectangular area can be quickly get on the integral image, shown in Figure 3:

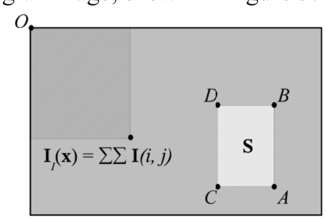

Figure 3. Using integral image to acquire the measure of area

So, with the area S=A+D-B-C, the property can be used to quickly gain the Hessian response. Using the same stereo vision platform and the constant input image size, the map can be calculated in advance. The calculation process can be speeded up through this way. The equation is as follows:

$$H(x,y)=\begin{bmatrix}\dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x\partial y}\\ \dfrac{\partial^2 f}{\partial x\partial y} & \dfrac{\partial^2 f}{\partial y^2}\end{bmatrix} \tag{3}$$

$\det(H(x,y))=\dfrac{\partial^2 f}{\partial x^2}\dfrac{\partial^2 f}{\partial y^2}-\left(\dfrac{\partial^2 f}{\partial x\partial y}\right)^2$, the four elements of the Hessian response related to the second derivative and mixed partial derivatives of the image point along the x and y directions, so it can be obtained through a second-order difference. SURF using the Fast-Hessian, because the Fast-Hessian can get a rectangular filter[5] to filter during the integral image, so be called Box-Filter, the approximation effect in Figure 4:

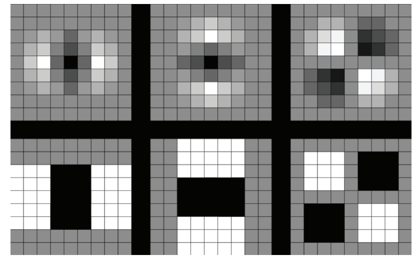

Figure 4. Approximate to LoG using Fast-Hessian

Visible, use the Fast-Hessian approximate to LOG will cause large errors at the edges, in order to limit these errors, use (4) to estimate $\det(H(x,y))$, namely:

$$\det(H(x,y))=\frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2}-0.9\times\left(\frac{\partial^2 f}{\partial x\partial y}\right) \tag{4}$$

Step 2: in each layer of the pyramid, strike each mapping region of the Hessian extreme value response with (4). The size of the region determined by the scale factor, one of the biggest feature of the SURF's detector is sub-sampling to the filter, rather than sub-sampling to the original image, this important improvement making the SURF's speed advantage obviously in the high-resolution image, the size of each layer in the pyramid, according to equation (5):

$$\delta_{current}=\frac{1.2}{9}\times\delta_{last} \tag{5}$$

Inside, $\delta_{current}$ is the scale factor of the current layer, δ_{last} refers to the scale factor of the before layer, the filter scale factor in the first layer is $\delta_{base}=9$, ie, the filter size of the first layer is 9×9. Figure 5 shows the detection of the Hessian response of an actual scene:

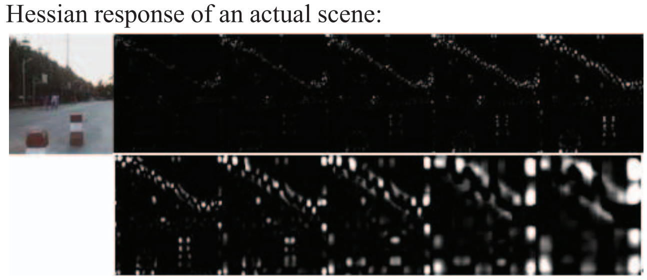

Figure 5. each level's Hessian response map from bottom to top

Thus, the feature points of the bottom are more than the top of the feature points, the more to the bottom , the feature points are often more able to reflect a small difference, the more high-level feature points are often more able to reflect the overall characteristics. Feature points is not only the extreme values on the scale, between the extreme values of the scale layer, which have extremum screening.

Step 3: Screening extreme value from candidate extreme points. This step will carry out the final feature point based on the constraints from the extreme points, the constraints included the boundary constraints, the response threshold constraints, scale extremum constraints. Its mathematical expression such as (6) below:

$$y=b(x)t(x)s(x)=\begin{cases}1 , & b(x)>1\&\&t(x)>1\&\&s(x)>1\\ 0 , & else\end{cases}$$

$$b(x)=r(x)c(x)=\begin{cases}1 , & 0<r(x)<\dfrac{r_c}{\delta_c} \quad \&\& \quad 0<c(x)<\dfrac{c_c}{\delta_c}\\ \\ 0 , & else\end{cases} \tag{6}$$

$$t(x)=\begin{cases}1 , & t(x)>threshold\\ 0 , & else\end{cases}$$

$$s(x)=\begin{cases}1 , & \max\{H(x_r)\}=H(x)\\ 0 , & else\end{cases}$$

Among them, the processing of strike is not include sub-sampling for the original image, so ensure the extreme points on the corresponding scale is not cross-border, we need for the boundary constraints; in order to ensure that the feature point response is strong enough to get a robust characteristics, we need the response threshold constraints, threshold is 12, for different platforms, the threshold may not the same; in order to ensure that the feature point scale extreme values, we need the multi-scale non-extremal constraint .

B. *Feature point description*

For binocular vision platform, if put the left and the right cameras parallel to the place in front of the cameras and correction, the left and the right views in the x and y direction are not a relative rotation, then SURF descriptors don't need to have a rotational invariance, but due to a scale difference about the left and the right cameras are not in the field of view range of objectives, the SURF description need to have scale invariance, so the main control of the characterization stage set the direction to 0^o, and because of the vector of 64-dimensional description, which is still belong to Upright-SURF-64. This paper's SURF descriptor is as follows:

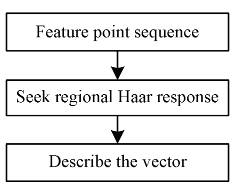

Figure 6. this paper's SURF descriptor

This stage Haar response[6] traverses the feature points around the local area with the Haar wavelet, each local area contains 4×4 sub-areas, each sub-area contains 5×5 sample points. SURF using directions of the Haar wavelet in horizontal and vertical, the Haar wavelet did not use the diagonal direction. Haar response in accordance with each sub-region divided into four categories, and be preserved according to the sum of the kinds, and at last formate of the region feature vector, ie: $\left(\sum H_x \quad \sum H_y \quad \sum|H_x| \quad \sum|H_y|\right)^T$.

III. STEREO MATCHING ALGORITHM[7]

The characteristics of SURF is based on multi-scale analysis of the characteristics, in the binocular vision, this feature has specific scale reproducibility, can search for specific feature points on specific scale layer, this constraint reduces the search range greatly.

This paper proposed a binocular stereo matching algorithm code, it's implementation process can be mathematical expressed in the form of formula (7):

$$f(x)=b(s(e(x)r(x)))=\begin{cases}1 , & b(s(e(x)\&\&r(x)==1)==1)==1\\ 0 , & else\end{cases} \tag{7}$$

For (7), 1 represents true, 0 represents false, b (t) on behalf of the process of symmetry constraints, its domain is the range of s (m); s (m) on behalf of the process of scale constraint, its domain is the value of e (x) and r (x) are all true, e (x) on behalf of the process of epipolar constraint, r (x) on behalf of the process of disparity range constraint, the procedure image of binocular stereo matching process shown in Figure (7):

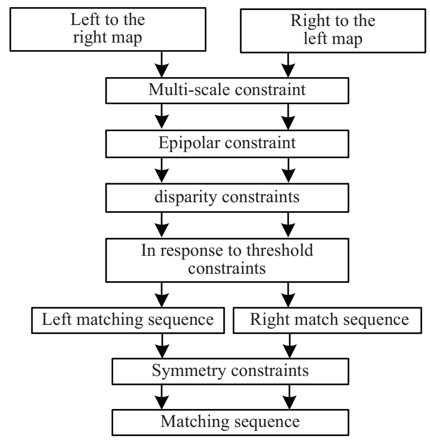

Figure 7. The procedure image of binocular stereo matching

The images include from left to right and from right to left inputed parallelly, the SURF algorithm extract the feature points where the scale level of the qualified basis of the range of the outer loop, and use the epipolar constraint[8] and disparity constraint[9] as a condition to determine specific screening, the points of match threshold more greater, which will be saved into the matching sequence, and contrast to match sequence , save the common part of the two sequences match as the final output sequence.

IV. THE RESULTS AND ANALYSIS OF EXPERIMENT

A. *Extract SURF features from 6 different scene:*

Figure 8. extract SURF features from 6 different scenes

The results can be received from the above experiments: in the region of shape with significant change, such as stair edge, corner, the feature can be extracted by SURF; in the region of the texture changes, such as the rough surfaces of buildings, woods, the feature can also be extracted by SURF. Therefore, for the region with the terrain changes and complexity background texture, SURF can extract more characteristics, in the area of the strong illumination changes, the feature still can be extracted by SURF . So SURF is a feature extraction algorithm，that it suitable for use as the eyes or vision platform or monocular high-resolution platform .

B. Matching algorithm is used to test and sparsely resconstuct the real scene, binocular matching of three different scenarios, shown in Figure 9 matching results for three different scenes:

Figure 9. scene1

Figure 10. scene2

Figure 11. scene3

Fig 9 The consequence of binocular matching

It Can be perceived, although there are still small amounts of mismatch, but the vast majority of feature points can match correctly, and the algorithm have higher efficiency in the implementation, Both enter the scene around 640*480 , and the matching time for complex scene 1, scene 2 within 100ms ,the matching time for uncomplicated scenes less than 80ms. View about the completion of the matching calculation of the disparity, then you can proceed the projection reconstruction of the target .

V. CONCLUSION

Based on the theory of binocular stereo vision, the paper uses the SURF feature point extraction algorithm to make specific screening by using the Haar describer of the principal control direction to 0^{o} and make the epipolar constraint and disparity constraint as the condition. It reduces the scope of the search greatly so as to improve the matching speed and accuracy.

(1) The SURF algorithm is used to construct SURF detector pyramid. Using the integral image and the Hessian matrix to approach on LOG rapidly can increase rate of extract the feature points about 10 times than standard SIFT. Then the speed of match of the binocular stereo vision is improved.

(2) The characteristics of SURF are based on multi-scale analysis. In the binocular vision, this feature has specific scale reproducibility, it can search specific feature points on specific scale layer. The constraint reduces the search range greatly. The paper involves five constraints, such as scale constraints, epipolar constraint, symmetry constraints, disparity constraints and response threshold constraints. To some extent, the more constraints, the better. Not only it can reduce the search range, but reduce the risk of mismatching. So the accuracy of the binocular stereo vision matching can be improved.

ACKNOWLEDGMENT

Firstly, the research is carried out in The Innovation Center of Disabled Rehabilitation in Equipment and Technology is supported by Jilin Provincial Science and Technology Department (No. 20102109). Secondly, the research activities have been funded by Changchun Science and Technology Bureau (No.10PT01).The research activities have also been funded by Jilin Provicial Education Department (No. 2011204, No. 2012243).

REFERENCES

[1] WU L.Robot Vision Navigation s Summarize[J].Computer Knowledge and Technology,2008,1(9):1705-1711.

[2] ZHANG W M,LIU B,LI H B. Characteristic point extracts and the match algorithm based on the binocular vision in three dimensional reconstruction[J].OPTICAL TECHNIQUE,2008,34(2):181-185.

[3] Bay H, Tuytelaars T, Cool L V. SURF: Speeded Up Robust Features. ECCV.2006.

[4] Viola, P.Jones, M.Rapid object detection using a boosted cascade of simple Features.CVPR.2001:511-518

[5] Gradshteyn,I.S.Ryzhik,I.M.Hessian Determinants.Tables of Integrals, Series, and Products.5th ed. San Diego.CA: Academic Press. 1979:1112-1113.

[6] The design of face detection system based on computer vision technology,Electronic Design Engineering. 2011.19(16):38-41.

[7] H Hirschmulle: Stereo Processing by Semiglobal Matching and Mutual Information[J].IEEE Transactions on Pattern Analysisand Machine Intelligence(PAMI),2008.

[8] XU Q Y,YAO H,QI X,YE D,CHE R S.Stereo vision reconstruction method based on common perpendicular constraint[J].Chinese Journal of Scientific Instrument.2008.29(11):2430-2434.

[9] LIN J L,CHEN Y L.3D face reconstruction using binocular vision[J]. Application Research of Computers.2012.29(1):397-400.