A PERFORMANCE COMPARISON OF FEATURE DETECTORS FOR PLANETARY ROVER MAPPING AND LOCALIZATION

W. Wan [a, *], M. Peng [a], Y. Xing [b], Y. Wang [a], Z. Liu [a], K. Di [a], B. Teng [b], X. Mao [b], Q. Zhao [a], X. Xin [a], M. Jia [a]

[a] State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China - (wanwh, pengman, wangyx716, liuzq, dikc, zhaoqiang, xinxin, jiamn)@radi.ac.cn
[b] Beijing Institute of Control Engineering, Beijing, China – 13717990916@163.com, tengweb@hotmail.com, cicely_mao@163.com

Commission III, ICWG III/II

KEY WORDS: Feature detector, feature matching, SIFT, planetary rover, stereo images

ABSTRACT:
Feature detection and matching are key techniques in computer vision and robotics, and have been successfully implemented in many fields. So far there is no performance comparison of feature detectors and matching methods for planetary mapping and rover localization using rover stereo images. In this research, we present a comprehensive evaluation and comparison of six feature detectors, including Moravec, Förstner, Harris, FAST, SIFT and SURF, aiming for optimal implementation of feature-based matching in planetary surface environment. To facilitate quantitative analysis, a series of evaluation criteria, including distribution evenness of matched points, coverage of detected points, and feature matching accuracy, are developed in the research. In order to perform exhaustive evaluation, stereo images, simulated under different baseline, pitch angle, and interval of adjacent rover locations, are taken as experimental data source. The comparison results show that SIFT offers the best overall performance, especially it is less sensitive to changes of image taken at adjacent locations.

1. INTRODUCTION

Feature detection and matching are key techniques in computer vision and robotics, and have been successfully implemented in many fields such as object recognition, 3D reconstruction, image retrieval, and camera localization, etc. Tens of features feature detectors and matching methods have been developed and used in different applications. It is valuable and important to evaluate and compare of the detectors and matching methods under different environmental conditions, so that to provide reference for design and development of application systems in the future.

Many investigations and comparisons of detectors and descriptors have been presented. Schmid et al.(2000) categorized the evaluating methods based on ground-truth verification, visual inspection, localization accuracy, theoretical analysis and specific tasks, and introduced two evaluation criteria including repeatability rate and information content for comparing feature detectors. Mikolajczyk and Schmid (2005) compared affine regions detectors while Mikolajczyk et al. (2003) reported an evaluation of local descriptors. Rodehorst and Koschan (2006) compared the SUSAN-2D operator, the Plessey detector and Förstner operator, and the performance of the detectors were compared under three criteria including detection rate, repeatability rate and localization accuracy. In general, previous comparisons were either application oriented or limited in experimentation or in the number of detectors and descriptors compared. Apollonio et al. (2014) verified the efficiency of different feature-based methods in different situations (scale variation, camera rotation, and affine transformation). Jazayeri and Fraser (2010) assessed the performance of the interest operators within an eight-image network on the basis of accuracy of interest-point localization, detection rate and speed. Gil et al. (2010) evaluated the repeatability of the detectors, as well as the invariance and

distinctiveness of the descriptors, under different perceptual conditions using sequences of images representing planar objects as well as 3D scenes. Mukherjee et al. (2015) made a comprehensive review of a large number of popular feature detectors developed in the last three decades, and conducted exhaustive experiments on several datasets for each combination of detectors and descriptors to provide a ranking that can also be weighted to suit specific applications.

It's worthy to note that most evaluation and comparison results are based on standardized reference images; different transformations of a single image, such as viewpoint change, illumination variation, and scale change are simulated to measure the stability of feature detectors.

As far as we know from published literatures, there is no performance comparison of feature detectors and matching methods for planetary mapping and rover localization using rover stereo images.

In this research, we present a comprehensive evaluation and comparison of six feature operators, including Moravec, Förstner, Harris, FAST, SIFT and SURF, aiming for optimal implementation of feature-based matching in planetary surface environment. To facilitate quantitative analysis, a series of evaluation criteria, including distribution evenness of matched points, coverage of detected points, and feature matching accuracy, are developed in the research.

2. INTEREST POINT DETECTORS

In this section the feature detectors including Moravec, Harris, Förstner, SIFT and SURF are introduced, and they have been implemented and applied in our experiments.

---

*   Corresponding author.

2.1 Moravec

Moravec detector (Moravec, 1980) is one of the earliest corner detection algorithms; for each pixel it compares a patch centered on that pixel with 8 local patches which are simply shifted by a small amount (typically 1 pixel in each of the eight possible directions) from the current patch. It computes the sum of squared differences (SSD) in four directions and takes the smallest value as the measure of corner strength. Therefore it detects point where there are large intensity variations in every direction. It's well noted that the operator is not isotropic. Hence the response on the operator is not isotropic.

2.2 Förstner detector

Förstner detector (Förstner and Guelch, 1987) uses the auto-correlation function to classify the pixels into categories (interest points, edges or region); the detection and localization stages are separated into the selection of windows (in which features are known to reside) and feature location within selected windows. Further local statistics allow estimating the thresholds for the classification automatically. To compute the location of a corner with subpixel accuracy, the Förstner algorithm seeks for the point closest to all the tangent lines of the corner in a given window based on a least-square solution.

2.3 Harris detector

Harris detector(Harris and Stephens, 1988) computes a matrix related to the autocorrelation function of the image which is similar to Moravec. The Harris corner detector differs from the Moravec detector in how to determine the cornerness value. Rather than looking at the sum of squared differences it makes use of partial derivatives, a Gaussian weighting function, and the Eigenvalues of a matrix representation of the equation. It is significantly more expensive computationally as compared to the Moravec corner detector.

2.4 FAST detector

The FAST (Features from Accelerated Segment Test) feature detector (Edward et al, 2010) is based on accelerated segment test. Firstly a feature is detected at pixel p by examining a circle of 16 pixels surrounding *p*. The pixel on the circle is considered 'bright' if its intensity is above the intensity of p by at least threshold t, and 'dark' if the intensity value is below its intensity of p by at least threshold t. The algorithm is further accelerated by using ID3 (Iterative Dichotomiser 3) algorithm to classify a candidate pixel as corner or non-corner. As small sets of positive corners are produced after segmentation, to further refine the results, the corner response function V which measures cornerness of detected corners is used, and non-maximal suppression is applied to remove corners that have an adjacent corner with higher V.
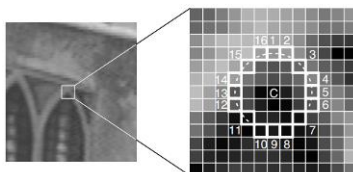

Figure 1. Fast detector

$$V = \max\left(\sum_{x \in S_{bright}} \left| I_{p \rightarrow x} - I_p \right| - t, \sum_{x \in S_{dark}} \left| I_p - I_{p \rightarrow x} \right| - t \right) \quad (1)$$

2.5 SIFT detector

The previous corner detectors examine an image at only a single scale. Lowe (1999) proposed a Scale Invariant Feature Transform (SIFT) detector/descriptor scheme. SIFT keypoints are invariant with respect to scale, translation and orientation in image scale space. The scale-space of an image is first produced from the convolution of the input image with a Difference of Gaussian (DOG) detector. Maxima and minima of this scale-space function are determined by comparing each pixel in the pyramid to its neighbors. Keypoints are taken as maxima and minima of the difference of Gaussian function that occur at multiple scales. Then the interpolation of nearby data is done using the quadratic Taylor expansion of the DOG scale-space function. The keypoints with low contrast are discarded based on the second-order Taylor expansion, such that poorly determined locations and high edge responses are eliminated. As a key step in achieving invariance to rotation, the main orientation for each feature is assigned based on local image gradient directions. A descriptor vector of 128 dimensions is then computed for each keypoint such that the descriptor is highly distinctive and partially invariant to the remaining variations.

2.6 SURF detector

SURF (Speeded Up Robust Features)(Bay et al., 2006) is a high-performance scale and rotation-invariant interest point detector inspired by SIFT. The standard version of SURF is claimed by its authors to be several times faster than SIFT and more robust against different image transformations. SURF relies on integral images for image convolutions. It is based on sums of 2D Haar wavelet responses and employs box filters using integral images, as opposed to the more computationally demanding Gaussian filters employed by SIFT. A fast Hessian matrix-based measure for the detector and a distribution-based descriptor is used. It describes a distribution of Haar wavelet responses within the interest point neighborhood, and only 64 dimensions are used reducing the time for feature computation and matching. The indexing step is based on the sign of the Laplacian, which increases the matching speed and the robustness of the descriptor.

The Hessian matrix at scale σ is defined as follows:

$$H(x, y, \sigma) = \begin{bmatrix} I_{xx}(x, y, \sigma) & I_{xy}(x, y, \sigma) \\ I_{xy}(x, y, \sigma) & I_{yy}(x, y, \sigma) \end{bmatrix} \quad (2)$$

where I_{xx} represents second order Gaussian smoothed image derivatives. Bay et al. (2006) used simple box filters to approximate convolution with the Gaussian second order derivatives. So box filters can be computed in constant time using the integral image. The approximate determinant of Hessian matrix then is

$$Det[H(x, y, \sigma)] \approx D_{xx} D_{yy} - (0.912 D_{xy})^2 \quad (3)$$

Finally, 3×3×3 –neighborhood non-maximum suppression and sub-pixel refinement are then applied.

3. EVALUATION CRITERIA

A common evaluation technique is to measure the repeatability rate which is the number of repeated points between two images considering the total number of extracted points. Keypoints

which are not detected in both images can corrupt the repeatability measure. As a consequence, only points which lie in the common scene parts are used to compute the repeatability. The repeatability rate is defined by

$$r = \frac{CorrectN}{\min(n_1, n_2)} \tag{4}$$

where *CorrectN* represents the number of matched points, n_1, n_2 are the number of points detected in stereo images respectively.

Meanwhile, the percentage of matched inliers is defined by

$$precision = \frac{inliers}{CorrectN} \tag{5}$$

Furthermore, to facilitate quantitative analysis, a series of evaluation criteria, including distribution evenness of matched points, coverage of detected points, and feature matching accuracy, are developed in the research. The matching accuracy is of importance since it measures 3D point accuracy. The coverage of detected points is related to localization accuracy, thus it is a secondary consideration for planetary application.

The matching accuracy measures whether the corresponding point is accurately located at a specific location (ground truth) based on the image attitude and 3D information. As shown in Figure 2, given a pair of stereo images and 3D point P, point p_i and p_j are homologous points after image matching, $\tilde{p}_j$ is calculated value after iteration, then the matching accuracy is the sum of matched point and projected value. Taking the feature points in the right image as base points, calculate the discrepancies between their homologous points (from image matching) and the projected positions of the 3D points in the left image to depict the inconsistencies between the matched points and measured points.

$$\varepsilon = \| \tilde{p}_j^{i+1} - p_j^{i+1} \| \tag{6}$$

The projected coordinates is computed as follows:

Step 1. Based on the initial height value Z_0 and left feature point (x_l, y_l), the approximate value of ground coordinate (X_1, Y_1) can be determined using collinearity equation.

$$X\text{-}X_s = (Z\text{-}Z_s)\frac{a_1 x + a_2 y - a_3 f}{c_1 x + c_2 y - c_3 f}$$
$$Y\text{-}Y_s = (Z\text{-}Z_s)\frac{b_1 x + b_2 y - b_3 f}{c_1 x + c_2 y - c_3 f} \tag{7}$$

Step 2. According to ground coordinate (X_1, Y_1), the corresponding height can be interpolated from the existed DEM.

Step 3. Repeat steps 1 to 2 until the termination condition is met, i.e., the change of the coordinate (X_i+1, Y_i+1, Z_i+1) between two successive iterations is less than a threshold.

Step 4. Calculate the projected point $p(x_r, y_r)$ of ground coordinate (X, Y, Z).
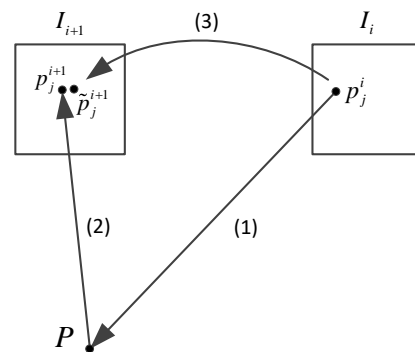

Figure 2. Evaluation method of matching accuracy

Evenness is an important factor to evaluate performance and quality of detectors, it is related to localization accuracy. Given an image of $M*N$, the image is divided into evenly spaced K grids, and the number of matched point in each grid can be calculated in terms of the coordinates of points and grid range, the evenness U is determined as

$$U = \frac{1}{K-1}\sum_{k=1}^{K}(N_k - \frac{1}{K}\sum_{k=1}^{K}N_k)^2 \tag{8}$$

K is the total number of grids, N_k is the number of points located within the k_{th} grid.

Coverage reflects the distribution of matched points within the overlap area between pairs. Firstly the convex hull of a point set is calculated, and the overlapping area of stereo images is computed, the ratio of convex hull area and overlapping area is obtained. In computational geometry, a number of algorithms are known for computing the convex hull for a finite set of points. In this paper the convex hull is determined using PCL (point cloud library). Furthermore, it's noted that the overlapping area is determined by a rectangular area.

The error of localization must be the most useful criterion in motion estimation of rover, so the relationships of different detectors and descriptors with the localization error are also evaluated. The specific steps for rover localization are:

Step 1. Give predefined EOPs (Exterior Orientation Parameters) to the first frame as initial EOPs of the image sequence.

Step 2. Calculate the three dimensional coordinates c based on the matched points and EOPs by space intersection.

Step 3. Repeat Step 2 to obtain the coordinates P_{XYZ}^2 of corresponding point in the second frame

Step 4. The rotation and translation parameters R, T can be determined using the model

$$P_{XYZ}^2 = R \cdot P_{XYZ}^1 + T \tag{9}$$

Step 5. Calculate the EOPs of the second frame based on the initial EOPs and R, T.

Step 6. Repeat the steps above to calculate the EOPs of the sequence images, the localization error is determined by calculating the difference between predefined rover location P_1 and solved location P_2.

$$R = Euler\_Dis(P_1, P_2) / Traverse\_Length \qquad (10)$$

4. EXPERIMENTAL RESLUTS

The performances of feature detection and matching are affected by different geometric factors, including image rotations, scale change, illumination variation, change of view point and camera noise. Due to the complexity of explicit formula derivation, simulated images are used to compare the performance of feature detectors. The simulated images are generated from digital projection with a virtual stereo camera by using the existing Digital Elevation Model (DEM) and Digital Orthophoto Map (DOM) of an area of Martian surface. The DEM and DOM were produced from HiRISE (High Resolution Imaging Science Experiment) stereo images are available at Mars Orbital Data Explorer website of Planetary Data System Geosciences Node (http://ode.rsl.wustl.edu/mars/pagehelp/quickstartguide/index.html?hirise_dtm.htm). The size of the maps is 7186m × 16392m. The resolutions of DOM and DEM are 0.25 m and 1 m respectively, and the vertical accuracy of the DEM is tens of centimeters. Given the interior and exterior orientation parameters, pixel size, image size and lighting parameters for the virtual stereo camera, the corresponding simulated images can be generated using back projection techniques base on the rigorous sensor model, i.e., the collinearity equations. In order for the simulated images to have enough details, the resolution of DEM and DOM is considered as 0.002m in the simulation computation.
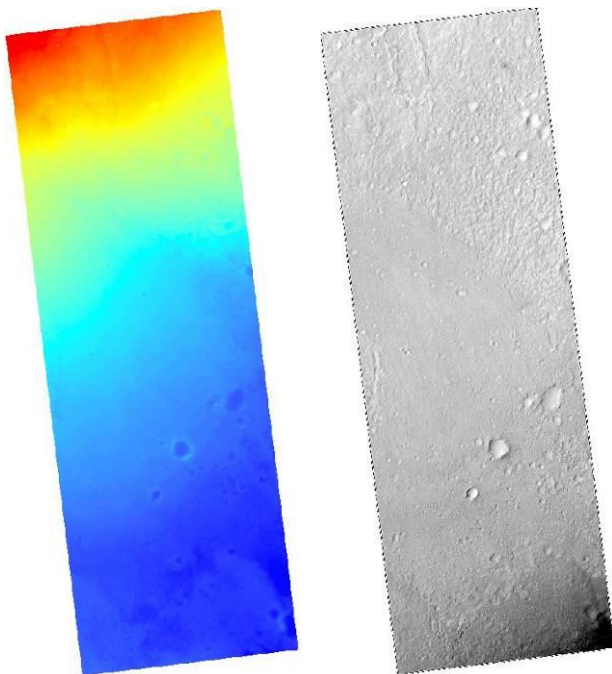

Figure 3. DEM (left) and DOM (right) (ID: DTEEC_010573_1755_010639_1755_U01.IMG, PSP_010573_1755_RED_A_01_ORTHO.JP2) generated from of HiRISE stereo images

In order to establish different parameters of the geometric configuration affects matching results, a "normal case" stereo camera is used, the image size of stereo camera is 1024×1024 pixels and the focal length is 1189 pixels, the coordinates of the principal point in the fiducial axis system is set as (511.5, 511.5). The baseline and pitch angle are set to 0.3 m and −20° respectively. Assuming that the stereo camera is installed 1.5 m above the ground on a rover, the principal optical axis from a stereo camera having the intended pitch angle will intersect the flat ground at a distance of 4.9 m, which is an ordinary setting used in field experiments. The distortion parameters are assumed to be equal to zero, while the exterior orientation parameters (φ, ω, κ, Xs, Ys, Zs) of different photographs are defined differently in terms of the location. A ground reference coordinate system is defined such that the Y axis points in the traverse direction, the Z axis points downward (perpendicular to the plane $Z = 0$), and the X axis forms the right-hand coordinate system.

Furthermore，due to the given DOM is of some fixed sun altitude, the sun altitude and sun orientation parameters are introduced to reflect illumination variation of different image in the simulation experiments. To simplify the implementation, the OpenGL phong model is applied to generate image texture and gray value. Phong model uses three parameters: ambient, diffuse and specular. Ambient reflects light that comes from all directions equally and is scattered in all directions equally by the polygons in the scene, diffuse means the light comes from a particular point source and hits surfaces with an intensity that depends on whether they face towards the light or away from it, and specular lighting is what produces the shiny highlights. The sun altitude and sun orientation parameters can be transferred into the corresponding phong model parameters. The sun altitude are set to 30°, 45°, 60° and 90° respectively, while the sun azimuth are set to 90° and 270°.

Based on the given average height Z_0 and pixel coordinate, the ground coordinates (X,Y,Z) can be obtained taking formulas and the procedure in section 4 as reference, then the gray value of pixel can be interpolated on the basis of known (X,Y) and modified DOM, finally the simulate images can be completed after point-by-point computation. In addition, if the ground coordinates don't converge during iteration, the grey value is set as zero. Figure 4 shows four simulated images of different parameters.
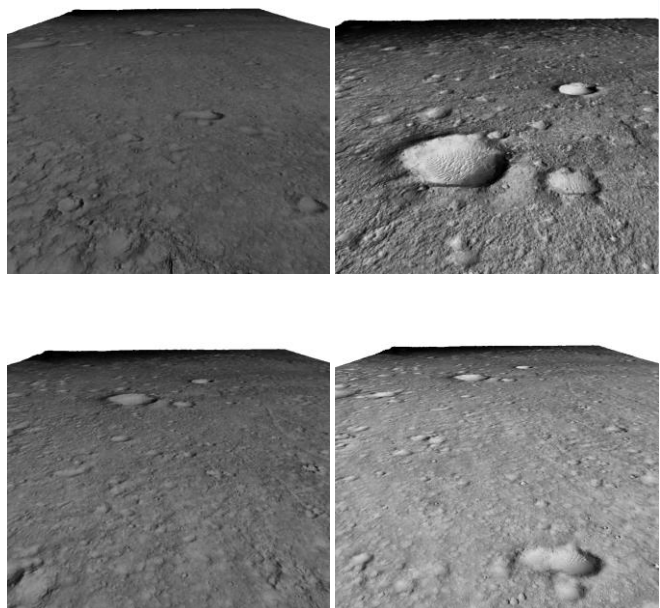

Figure 4. Samples of simulated images

The experimental results on different descriptors are shown in Table 1 under the condition of preserving 1000 feature points. The results show that the performance of SIFT detector is superior to others. This may be explained by the characteristics

of stereo visual odometry, because there are small scale and rotation changes deformations among two consecutive frames. It is worth to note that there are no noises added to the simulated images, meanwhile other elements affecting the EOPs are not considered, so the localization error is up to 0.02%. Further, to compa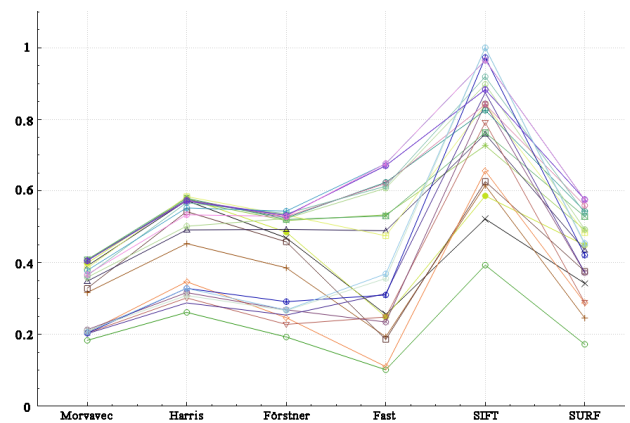re the performances of different descriptors under different parameter setting, Figure 5 shows the results of repeatability rate, precision, accuracy, evenness and coverage after normalization if the max value of the criterion is higher than 1. In these figures, different colors represent different experiment settings.

Table 1. List of detectors which provide best result in each criterion
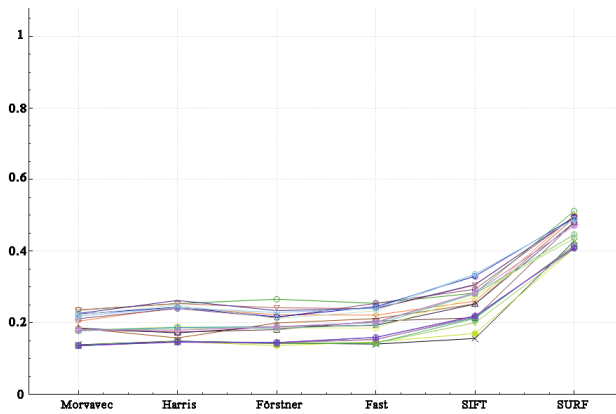
Evaluation criteria	Detector	Best value	Sun altitude	Sun azimuth	Distance of waypoints
Repeatability rate	Fast	0.625	30°	270°	0.2 m
Precision	SIFT	0.999	60°	270°	0.2 m
Accuracy	Moravec	0.135	30°	270°	0.2 m
Evenness	SIFT	0.312	90°	270°	1.0 m
Coverage	Moravec	0.660	45°	90°	0.5 m
Localization	SIFT	0.0002	30°	270°	0.2 m
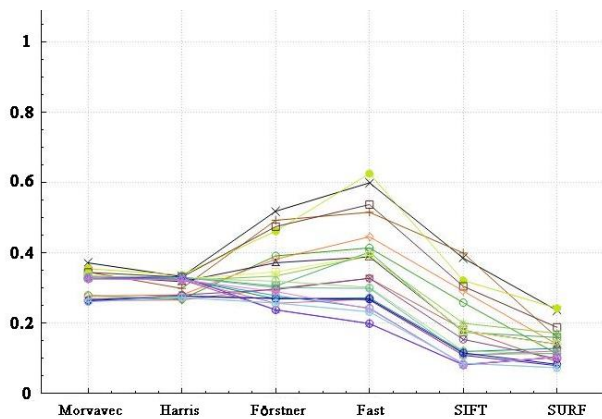
(a) Repeatability rate results

(b) Precision results

(c) Accuracy results

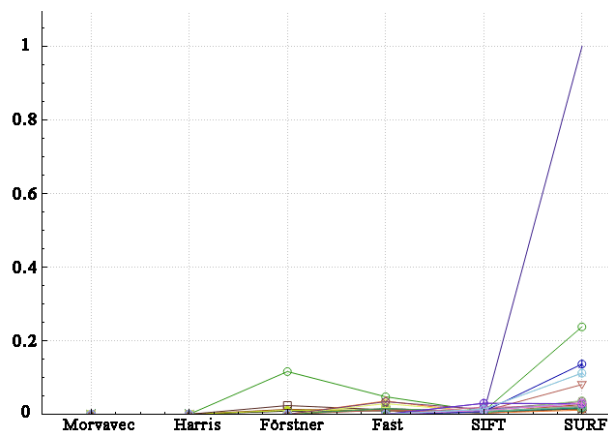(d) Evenness results
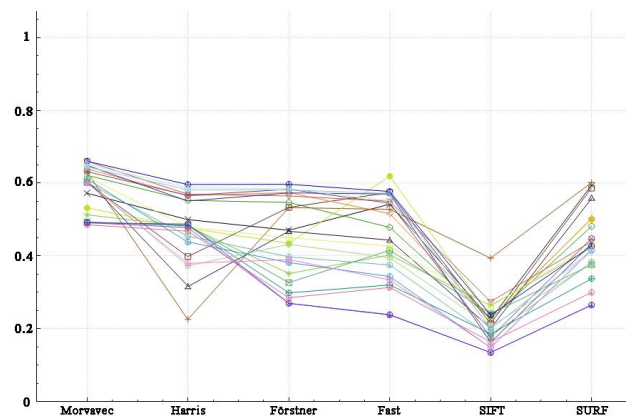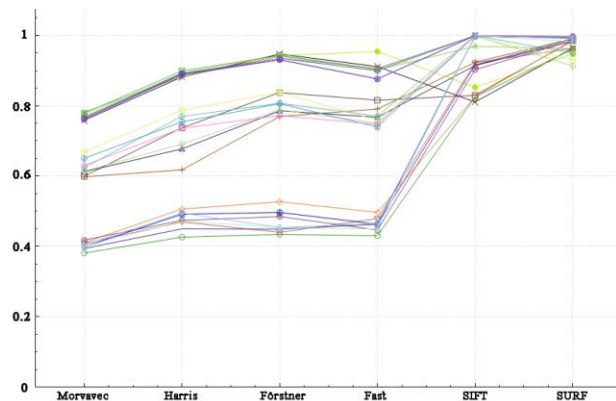
(e) Coverage results

(f) Localization accuracy results

Figure 5. Experimental results

5. CONCLUSIONS

An evaluation and comparison of interest point detectors for rover mapping and localization has been presented. Typical measures used in photogrammetric applications, such as repeatability rate and precision are adopted; moreover, three criterions including distribution evenness of matched points, coverage of detected points, and feature matching accuracy are proposed.

From our experimental results, the SIFT shows an overall good performance among those detectors. Compared with other evaluation papers, we performed a quantitative analysis of the point detectors for rover images based on the simulated images and exterior orientation parameters. Generally, the detectors with scale invariance are more stable than simple corner detectors, so they have less failure rate on motion estimation and high matching accuracy. In the experiment, each operator has its fixed set of parameters for the entire image. An adaptive parameter selection strategy could give better results in point selection and distribution.

The actual performance of detectors will depend on on the real world conditions. The comparative results in this research are indicative of the relative performance, while the exact measures of the performance may be different in real world applications. Nevertheless, the comparison results can be useful as reference in design and development of rover localization and mapping systems in the future, particularly they are valuable in design of rover navigation and mapping system in China's first Mars rover mission, which is planned to be launched in 2020.

ACKNOWLEDGMENT

REFERENCES

Apollonio, F. I., Ballabeni, A., Gaiani, M., et al, 2014. Evaluation of feature-based methods for automated network orientation. In: *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(5), pp. 47.

Bay, H., Tuytelaars, T. and Van, G. L., 2006. Surf: Speeded up robust features. European Conference on Computer Vision, pp. 404-417.

Edward, R., Reid P. and Tom D., 2010. Faster and better: a machine learning approach to corner detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32, pp. 105-119.

Förstner, W. and Gülch, E., 1987. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In: *ISPRS intercommission conference on fast processing of photogrammetric data*, pp. 281-305.

Gil, A., Mozos, O. M., Ballesta, M., et al., 2010. A comparative evaluation of interest point detectors and local descriptors for visual SLAM. *Machine Vision and Applications*, 21(6), pp. 905-920.

Harris, C. and Stephens, M., 1988. A combined corner and edge detector. In: *Proceedings of the 4th Alvey Vision Conference*, pp. 147–151.

Jazayeri, I. and Fraser, C. S., 2010. Interest operators for feature-based matching in close range photogrammetry. *The Photogrammetric Record*, 25(129), pp. 24-41.

Lowe, D. G., 1999. Object recognition from local scale-invariant features. In: *International Conference on Computer Vision*, 2, pp. 1150-1157.

Mikolajczyk, K. and Schmid, C., 2003. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 27(10), pp. 1615-1630.

Mikolajczyk, K., Tuytelaars, T., Schmid, C., et al., 2005. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1), pp. 43-72.

Moravec, H. P., 1980. Obstacle avoidance and navigation in the real world by a seeing robot rover. *Stanford Univ Ca Dept Of Computer Science.*

Mukherjee, D., Wu, Q. M. and Wang, G., 2015. A comparative experimental study of image feature detectors and descriptors. *Machine Vision & Applications*, 26(4).

Rodehorst, V. and Koschan, A., 2006. Comparison and evaluation of feature point detectors. In: *5th International Symposium Turkish-German Joint Geodetic Days.*

Schmid, C., Mohr, R. and Bauckhage, C., 2000. Evaluation of interest point detectors. *International Journal of computer vision*, 37(2), pp. 151-172.