Original Research Paper Engineering

# Performance Evaluation of Various Feature detection Algorithms in VSLAM

**S.SriVidhya** Assistant Professor, Dept of ISE, BNM Institute of Technology, Bangalore-560071,India.

**Prakash S** Professor, Dept of CSE, Dayanandsagar University, Bangalore, India.

**ABSTRACT**

Visual SLAM is a part of robotics community which is used by an autonomous vehicle in an unknown environment to orient themselves and navigate in their environment. The fundamental problem in robotics community is how to detect, identify and recognize the feature points in an image. These features are invariant to variety of effects like rotation, scale changes, view point changes, noise or illumination change. A typical SLAM procedure may be divided into four different parts: Feature extraction, feature matching, map generation and loop closing. In this paper we compare and evaluate how well different available implementations of various feature detectors such as FAST, MSER, SURF, Harris, and MinEigen perform in terms of rotation, scale change, noise, etc., on features such as corner and blobs. Further, their performance has been evaluated based on the number of matches and elapsed time. This review provides a brief introduction for providing a new research in robotics field for detecting features in an image.

**KEYWORDS** *detectors, features, FAST, MSER, SURF, Harris, MinEigen*

**1. Introduction:**

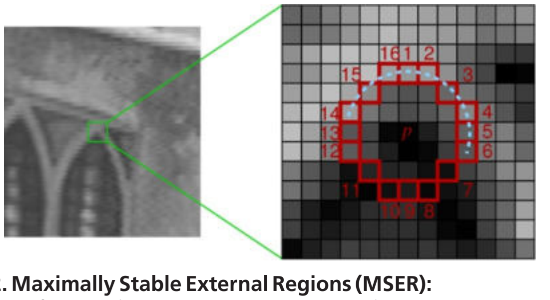
Local features detectors play an important role in many applications like mapping, text recognition, image registration (J. Bauer et al., 2004), object recognition (A.Berg et al., 2005), object categorization (Dorko and Schmid,2003), texture classification (S. Lazebnik et al.,2005), robot localization (S. Se et al.,2001), and video shot retrieval (J. Sivic et al.,2006). There are many researches that build new fast and robust detector (SIFT (D. Lowe, 2004), SURF(H. Bay et al.,2008),Fast (Guo,2011), BRISK (Leutenegger, 2011) and descriptors SIFT(D. Lowe, 2004),SURF (H. Bay et al.,2008), BRISK(Leutenegger,2011), Harris(C. Harris and M. Stephens,1988), FREAK (A. Alahi et al.,2012), MinEigen, MSER,HOG). Local features can be utilized into two different methods. First method includes three steps: feature detection, feature description, and feature matching (S.SriVidhya et al., 2015 ). Second method is bag-of features (E. Nowak et al.,2006) and hyper features (Agarwal et al.,2006) that includes feature detection, feature description, feature clustering, and frequency histogram construction for image representation. A local feature extraction is composed of feature detector and a feature descriptor. In this paper we have discussed about the performance of various features detectors such as FAST, MSER, SURF, Harris, and MinEigen.

**2. Feature Detectors**

**2.1. Features From Accelerated Segment Test (FAST):**

It was proposed originally by Rosten and Drummond (E. Rosten, and T. Drummond, 2006) for identifying interest points in an image. An interest point in an image is a pixel which has a well-defined position and can be robustly detected. Interest points have high local information content and they should be ideally repeatable between different images (Edward Rosten et al.,2010).It is proven that FAST detector performs well on images acquired by mobile devices in the context of visual navigation (Michał Nowicki and Piot,2014). Interest point detection has applications in image matching, object recognition, tracking etc. Segment test detector uses a circle of 16 pixels (a Bresenham circle of radius 3) to classify whether a candidate point p is actually a corner. Each pixel in the circle is labelled from integer number 1 to 16 clockwise. If a set of N contiguous pixels in the circle are all brighter than the intensity of candidate pixel p (denoted by $I_p$) plus a threshold value t or all darker than the intensity of candidate pixel p minus threshold value t, then p is classified as corner.

The high-speed test for rejecting non-corner points is operated by examining 4 example pixels, namely pixel 1, 9, 5 and 13. Because there should be at least 12 contiguous pixels that are whether all brighter or darker than the candidate corner, so there should be at least 3 pixels out of these 4 example pixels that are all brighter or darker than the candidate corner. Firstly pixels 1 and 9 are examined, if both $I_1$ and $I_9$ are within [$I_p$ - t, $I_p$ + t], then candidate p is not a corner. Otherwise pixels 5 and 13 are further examined to check whether three of them are brighter than $I_p$ + t or darker than $I_p$ - t. If there exists 3 of them that are either brighter or darker, the rest pixels are then examined for final conclusion. Average 3.8 pixels are needed to check for candidate corner pixel (E. Rosten, and T. Drummond, 2006). Compared with 8.5 pixels for each candidate corner, 3.8 is really a great reduction which could highly improve the performance.




**2.2. Maximally Stable External Regions (MSER):**

MSERs features (O.Chum and J. Matas, 2005) are regions that are either darker, or brighter than their surroundings, and that are stable across a range of thresholds of the intensity function. MSERs have also been defined on other scalar functions (S. Obdrzalek, 2007), and have been extended to colour (P.-E. Forssen, 2007), this is used to detect blobs in an image. This technique was proposed by Matas et al (2002) to find correspondences between image elements from two images with different viewpoints. This method of extracting a comprehensive number of corresponding image elements contribute to the wide-baseline matching, and it has led to better stereo matching and object recognition algorithms. MSER can efficiently extract crosswalk regions under various illumination conditions, which can avoid the selection of thresholds according to the current environment situation and greatly improve the system flexibility and robustness (Yuqiang Zhai et al., 2015).

**2.3. Speeded Up Robust Features (SURF):**

It was proposed by Herbert Bay et al. at European Conference on Computer Vision (Ryuji et al., 2009). It is a local feature detector and descriptor that can be used for tasks such as object recognition or registration or classification or 3D reconstruction. It is partly inspired by the scale-invariant feature transform (SIFT) descriptor. The standard version of SURF is several times faster than SIFT and claimed by its authors to be more robust against different image transformations than SIFT. To detect interest points, SURF uses an integer approximation of the determinant of Hessian blob detector, which can be computed with 3 integer operations using a precomputed integral image. Its feature descriptor is based on the sum of the wavelet response around the point of interest. These can also be computed with the aid of the integral image. SURF descriptors can be used to locate and recognize objects, people or faces, to make 3D scenes, to track objects and to extract points of interest.

**2.4. Harris Detector:**

Harris and Stephens (C. Harris and M. Stephens, 1988), improved upon Moravec's corner detector by considering the differential of the corner score with respect to direction directly, instead of using shifted patches. This corner score is referred to as autocorrelation. Without loss of generality, they have assumed a gray scale 2-dimensional image. Let this image be given by I. Consider taking an image patch over the area ), $(u,v)$ and shifting it by, $(u,v)$ The weighted *sum of squared differences* (SSD) between these two patches, denoted $S$, is given by:

$$S(x,y)=\sum_u \sum_v w(u,v)(I(u+x,v+y)-I(u,v))^2$$

$I(u+x,v+y)$ can be approximated by a Taylor expansion. Let $I_x$ and $I_y$ be the partial derivatives of I, such that

$$I(u+x,v+y)\approx I(u,v)+I_x\ (u,v)x+I_y(u,v)y$$

This produces the approximation

$$S(x,y)\approx\sum_u \sum_v w(u,v)(I_x(u,v)x+I_y(u,v)y)^2$$

This can be written in matrix form:

$$S(x,y)\approx(x,y)A\begin{pmatrix}x\\y\end{pmatrix}$$ where A is the structure tensor,

$$A=\sum_u\sum_v w(u,v)\begin{bmatrix} I_x^2 & I_xI_y \\ I_xI_y & I_y^2\end{bmatrix}=\begin{bmatrix}\langle I_x^2\rangle & \langle I_xI_y\rangle \\ \langle I_xI_y\rangle & \langle I_y^2\rangle\end{bmatrix}$$

This matrix is a Harris matrix, and angle brackets denote averaging. A corner is characterized by a large variation of $S$ in all directions of the vector),$(x,y)$ By analyzing the eigenvalues of $A$, this characterization can be expressed in the following way: $A$ should have two "large" eigenvalues for a corner. Based on the magnitudes of the eigenvalues, the following inferences can be made based on this argument:

1. If $\lambda_1\approx 0$ and $\lambda_2\approx 0$ then this pixel (x,y) has no features of interest.

2. If $\lambda_1\approx 0$ and $\lambda_2\approx 0$ has some large positive value, then an edge is found

3. If $\lambda_1$ and $\lambda_2$ have large positive values, then a corner is found.

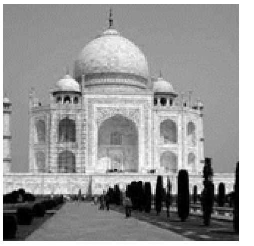
**2.5. MinEigen:**

Detect corners using minimum eigenvalue algorithm and return corner Points object. The object contains information about the feature points detected in a 2-D gray scale input image, I. The detectMinEigenFeatures function in Matlab uses the minimum eigenvalue algorithm developed by Shi and Tomasi to find feature points (Shi, J., and C. Tomasi, 1994).

**3. Experiments**

This paper aims to evaluate the various feature detection algorithms. The implementation was done on Intel® core(TM) i3 processor with 3GB RAM and speed of 2.53GHz. The code was written in Matlab R2013a on Windows 7 professional 64 bits. It consists of various tests by introducing effects like rotation, scale change and noise. The sample picture considered for all the tests is shown below in Fig 3.1of size 35.2KB.

**Fig 3.1 Sample Picture**



First, experiment will evaluate the detectors by number of captured key-points against elapsed time with rotational changes of 25, 45, 75, and 100 in an image. The following table's shows number of detected key points and time needed to capture them.

**Table 3.1: Detected feature points and elapsed time with rotational changes**

| Detectors | Detected Feature points | | | | | Elapsed time | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Original image | Rotate 25 | Rotate 45 | Rotate 75 | Rotate 100 | Original image | Rotate 25 | Rotate455 | Rotate 75 | Rotate100 |
| FAST | 223 | 192 | 199 | 190 | 185 | 0.074 | 0.090 | 0.538 | 0.506 | 0.119 |
| MSER | 130 | 102 | 109 | 102 | 107 | 0.419 | 0.483 | 0.400 | 0.397 | 0.397 |
| SURF | 150 | 156 | 167 | 173 | 170 | 0.674 | 0.511 | 0.439 | 0.379 | 0.373 |
| Harris | 163 | 194 | 184 | 188 | 173 | 0.144 | 0.133 | 0.144 | 0.145 | 0.133 |
| MinEigen | 563 | 734 | 483 | 758 | 658 | 0.193 | 0.158 | 0.138 | 0.160 | 0.158 |

Second, experiment will evaluate the detectors by number of captured key-points against elapsed time with scale changes of 1.2, 1.5, 1.7, and 1.9 in an image. The following table's shows number of detected key points and time needed to capture them.

**Table 3.2: Detected feature points and elapsed time with scale changes**

| Detectors | Detected Feature points | | | | | Elapsed time | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Original image | Scale 1.2 | Scale 1.5 | Scale 1.7 | Scale 1.9 | Original image | Scale 1.2 | Scale 1.5 | Scale 1.7 | Scale 1.9 |
| FAST | 223 | 219 | 246 | 252 | 256 | 0.086 | 0.161 | 0.136 | 0.113 | 0.137 |
| MSER | 130 | 156 | 202 | 234 | 273 | 0.497 | 0.474 | 0.732 | 0.742 | 1.139 |
| SURF | 150 | 214 | 260 | 361 | 408 | 0.440 | 0.503 | 0.583 | 0.845 | 0.892 |
| Harris | 163 | 216 | 323 | 351 | 412 | 0.181 | 0.161 | 0.160 | 0.240 | 0.248 |
| MinEigen | 563 | 792 | 1193 | 1419 | 1805 | 0.192 | 0.176 | 0.312 | 0.212 | 0.239 |

Third, experiment will evaluate the detectors by number of captured key-points against elapsed time with introducing various types of noises like Gaussian, Poisson, Salt & Pepper in an image. The following table shows number of detected key points and time needed to capture them.

**Table 3.3: Detected feature points and elapsed time with various noises**

| | Original image | Gaussian | Poisson | Salt & Pepper | Original image | Gaussian | Poisson | Salt & Pepper |
|---|---|---|---|---|---|---|---|---|
| FAST | 223 | 776 | 294 | 1780 | 0.094 | 0.093 | 0.142 | 0.136 |
| MSER | 130 | 152 | 142 | 124 | 0.425 | 0.445 | 0.425 | 0.377 |
| SURF | 150 | 194 | 169 | 232 | 0.453 | 0.434 | 0.490 | 0.531 |
| Harris | 163 | 968 | 226 | 761 | 0.153 | 0.217 | 0.157 | 0.181 |
| MinEigen | 563 | 1432 | 1299 | 1014 | 0.157 | 0.199 | 0.206 | 0.183 |

**4. Conclusion:**

The main purpose of this paper is to find the best detector in terms of rotation, scale change and various noises. We have deduced from the performance analysis that Min Eigen is the ideal choice amongst FAST, SURF, MSER, Harris feature detector algorithms. Min Eigen detects more features and provides better results, even under the rotational, scaling changes and introduction of noise. As shown in performance evaluation tables above, we see that FAST is taking less time but the number of features detected is less. Min Eigen takes marginally high time when compared to FAST but provides better results(time difference – scale:0.015,rotation: 0.068,Noise:0.106).Thus ignoring this marginal increase we suggest that Min Eigen is the optimal feature detection algorithm which can be used efficiently used in the SLAM procedure. We are planning to combine these detectors and descriptors to match the image precisely in order to attain better feature extraction results.

**References:**

1. J. Bauer, H. Bischof, A. Klaus, K. Karner(2004), "Robust and fully automated image registration using invariant features," ISPRS, DVD Proc, pp. 12–23.
2. A.Berg, T. Berg, J. Malik(2005)," Shape matching and object recognition using low distortion correspondence," in: Proceedings of the IEEE International Conference Computer Vision Pattern Recognition, vol. 1, pp. 26–33.
3. G. Dorko, C. Schmid(2003), "Selection of scale-invariant parts for object class recognition," in: Proceedings of the IEEE International Conference on Computer Vision, vol. 1, pp. 634– 639.
4. S. Lazebnik, C. Schmid, J. Ponce(2005), "A sparse texture representation using local affine regions", IEEE Trans. Pattern Anal. Mach. Intell. , vol. 27 (8), pp. 1265–1278.
5. S. Se, D. Lowe, J. Little(2001), "Vision-based mobile robot localization and mapping using scale-invariant features," in: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 2051–2058.
6. J. Sivic, F. Schaffalitzky, A. Zisserman(2006), "Object level grouping for video shots," Int. J. Comput. Vis., vol. 67 (2),pp. 189–210.
7. D. Lowe(2004), "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, vol.60(2), pp. 91–110.
8. H. Bay, A. Ess, T. Tuytelaars, L. Van Gool(2008), "SURF: Speeded Up Robust Features, "Computer Vision and Image Understanding, vol. 110(3), pp. 346–359.
9. Guo, Lisha, L.Junshan, Z.YingHong, and Z.Tang(2011), "A novel Features from Accelerated Segment Test algorithm based on LBP on image matching." In: Proceedings of "IEEE International Conference on Communication Software and Networks (ICCSN)", pp. 355-358.
10. Leutenegger, Stefan, M.Chli, and R.Siegwart(2011), "BRISK: Binary robust invariant scalable keypoints.", In: Proceedings of "IEEE International Conference on Computer Vision (ICCV)", pp. 2548-2555.
11. C. Harris and M. Stephens(1988), "A combined corner and edge detector," in Proceedings of the 4th Alvey Vision Conference, pp. 147–151.
12. A. Alahi, R. Ortiz, P. Vandergheynst(2012)," FREAK: Fast Retina Keypoint," In Proc. IEEE Conference on Computer Vision and Pattern Recognition.
13. S.SriVidhya, Dr.C.B. Akki, Dr.Prakash S(2015),"A Survey on Data Association Methods in VSLAM", International Journal of Engineering Trends and Technology (IJETT), V30(2),pp.83-88.
14. E. Nowak, F. Jurie, B. Triggs(2006), "Sampling strategies for bag-offeatures image classification," in: Proceedings of the European Conference on Computer Vision, Lecture Notes in Computer Science, vol. 3954, pp. 490–503.
15. Agarwal, Ankur, and B. Triggs(2006), "Hyperfeatures–multilevel local coding for visual recognition," In: Proceedings of "European Conference on Computer Vision–ECCV", pp. 30-43.
16. E. Rosten, and T. Drummond(2006), "Machine learning for high speed corner detection," 9th Euproean Conference on Computer Vision, vol. 1, pp. 430–443.
17. Edward Rosten, Reid Porter and Tom Drummond(2010), "FASTER and better: A machine learning approach to corner detection" in IEEE Trans. Pattern Analysis and Machine Intelligence, vol 32, pp. 105-119.
18. Michał Nowicki, Piot(2014)," Performance Comparison of Point Feature Detectors and Descriptors for Visual Navigation on Android Platform", In: Proceedings of IEEE Conference on Wireless Communications and Mobile Computing Conference (IWCMC), pp. 116 - 121.
19. O.Chum and J. Matas(2005),"Matching with PROSAC - progressive sample consensus" In: Proceedings of " IEEE Conference on Computer Vision and Pattern Recognition (CVPR)", pages 220–226.
20. S. Obdrzalek(2007), "Object Recognition Using Local Affine Frames", PhD thesis, Czech Technical University.
21. P.-E. Forssen(2007), "Maximally stable colour regions for recognition and matching," In: Proceedings of "IEEE Conference on Computer Vision and Pattern Recognition", Minneapolis, USA.
22. J. Matas, O. Chum, M. Urban, and T. Pajdla(2002), "Robust wide baseline stereo from maximally stable extremal regions," Proc. of British Machine Vision Conference, pages 384-396.
23. Yuqiang Zhai,Guolong Cui,Qin Gu,Lingjiang Kong(2015),"Crosswalk Detection Based on MSER and ERANSAC",In: Proceedings of IEEE 18th International Conference on Intelligent Transportation Systems (ITSC),pp. 2770 – 2775.
24. Ryuji Funayama, Hiromichi Yanagihara, Luc Van Gool, Tinne Tuytelaars, Herbert Bay(2009), "Robust Interest Point Detector And Descriptor", In: Proceedings of "European Conference on Computer Vision".
25. C. Harris and M. Stephens (1988), "A combined corner and edge detector", (PDF) ,In: Proceedings of Alvey Vision Conference, pp. 147–151.
26. Shi, J., and C. Tomasi(1994), "Good Features to Track," In: Proceedings of the "IEEE Conference on Computer Vision and Pattern Recognition", pp. 593–600.