SIMULTANEOUS ORIENTATION AND SCALE ESTIMATOR

Yang Cheng* and Adnan Ansar; Jet Propulsion Laboratory, California Institute of Technology. USA.
*ycheng@jpl.nasa.gov

Abstract. *We present a novel, fast and robust descriptor-based feature detection and matching algorithm called Simultaneous Orientation and Scale Estimator (SOSE). The primary innovation in this novel approach is in efficiency of scale estimation. Unlike other descriptor approaches, which rely on image pyramids or convolutions with banks of scale-dependent kernels to estimate a feature scale, SOSE estimates feature scale and orientation simultaneously from the image at its native scale. This novel approach has significant benefits for hardware implementation both because of low computational cost and a dramatic reduction in random data access introduced by the pyramid scheme.*

Introduction. Since David Lowe published his ground-breaking paper on the Scale-Invariant Feature Transform (SIFT)[1], descriptor-based feature matching has become a standard in computer vision and beyond. SIFT leverages earlier work in scale-space theory [5,6] to define scale-stable key points in an image as extrema in a representation of the image formed by convolution with a bank of difference of Gaussian kernels separated by a fixed scale factor. Extrema in this Difference of Gaussian (DoG) space approximate extrema of the scale-normalized Laplacian of Gaussian, which was previously shown [6,7] to produce scale-invariant keypoints. Since Lowe's work, many descriptor-based feature recognition algorithms have been produced. These include computational simplifications to SIFT, such as SURF [4], novel types of descriptors (ORB [8]) and modifications to the scale-space formalism (KAZE [9]), as well as many others.

A common drawback of descriptor-based approaches for efficient hardware implementation is that they use image pyramids or banks of image convolutions to model a scale-space representation. Random data access in the process of exhaustive search in scale-space is not amenable to parallelization or FPGA implementation. However, if the scale-space representation scheme is simplified, these approaches typically suffer from poorer performance in scale-invariance.

In this paper, we propose a novel approach that can estimate feature scale and orientation in a single image layer, which means the detailed scale representation, which is essential for good performance of most other algorithms, becomes unnecessary.

Approach. In this paper we will mainly focus on the efficiency improvement by eliminating the need for multi-level data layers and most of the random memory access operations in key point scale and orientation estimation. We find that the proposed method not only accomplishes this but also addresses issues caused by imprecise feature localization inherent in approaches that perform detection in scale-space whenever coarse scales are use.

Rosin [3] proposed the notion of orientation for an image patch in terms of its first order moments. He defines the moments of a patch as:

$$m_{qp} = \sum_{x,y} x^p y^q I(x,y) \qquad [1]$$

The Intensity Centroid (IC), analogous to an intensity-derived center of mass, of the patch is given by

$$IC = \left(m_{01}/m_{00}, m_{10}/m_{00} \right)$$

and the natural orientation for the patch is given by the angle between the patch center and the IC

$$\theta = atan2(\, m_{01}, m_{10}) \qquad [2]$$

Although the IC gives the most accurate orientation estimate, it is not always stable spatially. Therefore, identifying a moment where the orientation is relatively stable spatially and radially will lead to a stable and accurate orientation estimate.

We define the orientation stability measure (M_1) at P(x, y) as

$$M_1(r) = \left[\left(\frac{\partial\theta}{\partial x}\right)^2 + \left(\frac{\partial\theta}{\partial y}\right)^2 \right] < t1 \quad (x^2+y^2) < r^2 [3]$$

Since M_1(r) is close to zero, the orientation is locally approximated by a constant. Hence the orientation of a patch is stable spatially within a radius ***r*** of the patch center. For example, if ***P*** is moved slightly, dP(dx, dy), the new orientation is then

$$\theta' = \theta + \frac{\partial\theta}{\partial x}dx + \frac{\partial\theta}{\partial y}dy \approx \theta$$

We define another metric to enforce radial stability in IC:

$$M_2(r) = tan\big(\theta(r) - \theta(r+dr)\big) < t_2. \qquad [4]$$

Then we define the scale *S* at point *P* as the radius ***r*** when M_1 and M_2 are smaller than some chosen thresholds.

Computationally, the process described is very simple and does not involve any image pyramids or scale-dependent banks of convolutional kernels. It is also easy to implement in parallel, a critical need for hardware implementations. Further, because the scale and orientation are determined where they are most spatially and radially stable, the approach reduces feature localization error and improves sensitivity. We have named this approach *Simultaneous Orientation and scale Estimator (SOSE).*

Note that the core of SOSE is neither intrinsically a feature detector nor a feature descriptor algorithm. Instead, it provides efficient scale and orientation estimates to any detector with repeatability over scale, rotation and other relevant image transformations. We currently use Harris corner detection as the feature selection algorithm and a descriptor similar to SURF described below.

Performance evaluation

In order to test the effectiveness of SOSE, we scaled down a test image (1024 by 1024pixels) by scale at 0.9, 0.8, 0.7, 0.6. Then we run the SOSE estimator between the original image and scaled down image. The figure 3 shows the linear regression analysis of feature scale between original image and scaled down image and the slopes of the regressions for scale = 0.9, 0.8, 0.7, and 0.6. are 0.863, 0.772, 0.75 and 0.724 and coefficient of determination (R^2) are 0.8697, 0.8589, 0.7320 and 0.6108 respectively. However, SOSE stops working effectively when the scale is less than 0.6 , likely due to insufficient support for the orientation estimate in the down-sampled image. In general, SOSE is sufficient to cover the scale range for many real-world applications. In extreme case, we could employ a simplified pyramid scheme to extend the scale invariance beyond 0.6.

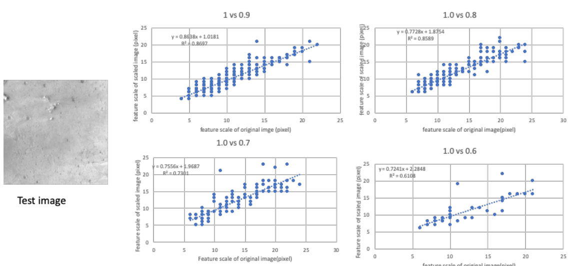

Figure 1: SOSE works when the scale is greater or equal to 0.6.

We evaluated SOSE against four scenarios: synthetic in-plane rotation between 0 to 360 degrees, off-nadir rotation from 0 to 45 degrees, and changes in image image illumination.

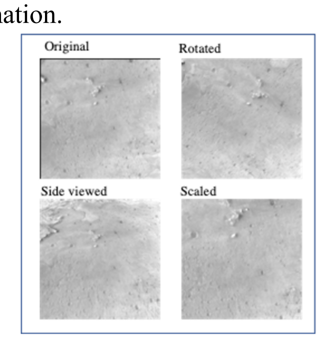

Figure 2: The test data set for evaluating the SOSE and other algorithms.

We also ran SIFT, AKAZE, SURF, and Brisk on the data sets. To make an apples-to-apples comparison, we limited the number of features for each matcher to the best 500. Among these listed algorithms, AKAZE had the best performance. SOSE fluctuates between the performance of SIFT and BRISK due to quantization in the orientation estimate. When quantization introduces small errors, SOSE outperforms SIFT, and when they are large, SIFT does better.

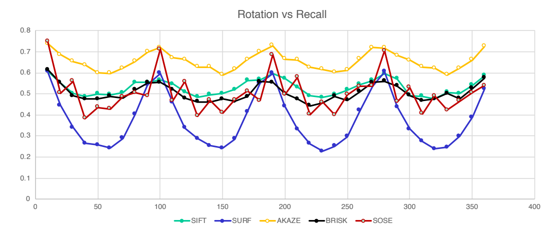

Figure 3: The image rotation vs the recall.

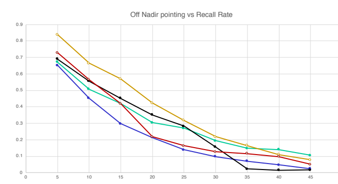

Figure 4: The Off nadir pointing vs recall. SOSE.

We also compared the performance under different lighting conditions. The synthetic images were rended with sun angle between 5 to 90 degrees with 5 degree intervals. We run the matchers between the one with 45 sun elevation with rest of images.

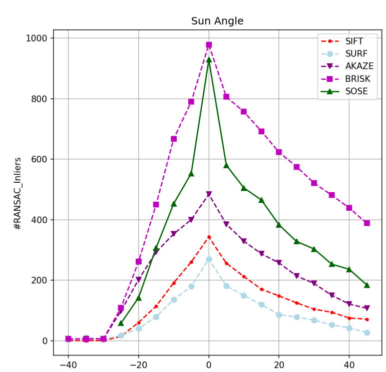

Figure 5: The performance comparison between SOSE and other matchers under different illumination condition.

Overall SOSE's performance is comparable to other SOA matchers. However, the goal of SOSE im implementation in firmware without loss of performance. This analysis shows that it meets that goal.

Even in software implementation, SOSE is among the faster match algorithms of those studies. Its speed is very close to SURF and AKAZE and about 9 times faster than SIFT.

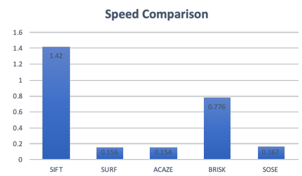

Figure 6: The SOSE is one pf the fastest algorithms.

Conclusions

We presented a novel approach for feature scale and orientation estimation which we call SOSE. Although its performance is slightly worse on average than the classic algorithms such as SIFT, that performance loss is mainly due to the choice of quantization in descriptor space to enhance efficiency. Critically, SOSE avoids the need for a pyramid scheme for scale estimation. As a result, it is both faster and requires less data access than other approaches. Hence, it is particularly suited for implementation in FPGA or other embedded systems.

Acknowledgments

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

[1] David G. Lowe, "Distinctive image features from scale-invariant key points," *International Journal of Computer Vision,* 60, 2 (2004), pp. 91-110.

[2] P. L. Rosin. Measuring corner properties. *Computer Vision and Image Understanding*73(2):291 – 307, 1999.

[3] Ethan Rublee et al, ORB: An efficient alternative to SIFT or SURF, 2011 International Conference on Computer Vision

[4] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool, "Speeded Up Robust Features", ETH Zurich, Katholieke Universiteit Leuven

[5] Koenderink, J.K, 1984. The structure of images. Biological Cybernetics, 50:363-396.

[6] Lindeberg, T. 1994. Scale-space theory: A basic tool for analyzing structures at different scales. Journal of Applied Statistics, 21(2):224-270.

[7] Mikolajcyk, K., and Schmidt, C. 2002. An affine invariant interest point detector. ECCV 2002, Copenhagen, Denmark, pp. 128-142.

[8] Rublee, E., et al., "ORB: An efficient alternative to SIFT or SURF," ICCV 2011, pp. 2564-2571

[9] Alcantarilla, P.F., Bartoli, A, and Davison, A.J., "KAZE Features," ECCV 2012, pp. 214-227.

[10] S. Winder M. Brown, R. Szeliski. Multi-image matching using multi-scale oriented patches. CVPR, 2005.

[11] Cordelia Schmid Krystian Mikolajczyk. Scale and ane invariant interest point detectors. IJCV, 60(1):63{86, 2004.