# Distinctive Image Features from Scale-Invariant Keypoints

Summary written by Sahana Srihari
February 4, 2020

**Summary:** The following paper delves into methods of identifying distinctive keypoints in an image that is invariant to scale. The problem addressed is not only invariant to scale but also shows robust performance in the case of rotation, illumination, change in 3D viewpoint and addition of noise. The paper also explores object recognition and matching amidst clutter and occlusion.

**Related work:** The author focuses on the prior work relating to the modification of the corner detection algorithm [4], which was improved by Harris in [2] in terms of repeatability of small invariances. Work of Schmid and Mohr [5] demonstrated the use local image feature matching to a generalized image recognition problem through the use of consistent cluster of matched features that allows a degree of resistance to occlusion and clutter. The author's main focus is related to the adaptation of his prior work [3] with improvements in stability and invariance.

**Approach:** This paper focuses on the detection of scale-space invariant features for various applications.

Step(1): Detection of scale-space extrema - which involves the detection of keypoints using a cascade filtering approach. The first step is to identify locations and scales resistant to viewpoint changes. To assure stable detection of keypoints, the Difference of Gaussian(DoG) function is convolved with the image(approximation to the scale normalised Laplacian of Gaussians). Sample points in the image are locally compared to neighboring pixels across, above and below scale-space which is then selected in comparison to being smaller or larger. Ultimately, this process results in the detection of the local extrema, when chosen, a detailed fit on the neighboring data is performed, resulting in low contrast points to be rejected which is the first step in accurately localizing this keypoint.

Step(2) Keypoint localisation- Brown in [1] has proposed a method in which the interpolated location of the maximum is determined using a 3D quadratic function and the Taylor series expansion of the scale space function $D$(x,y,$\sigma$) shifted so that the origin is on the sample point:

$$D(x)=D+\frac{\partial D^T}{\partial x}x+\frac{1}{2}x^T\frac{\partial^2 D}{\partial x^2}x \tag{1}$$

The location of the extremum is located by taking the derivative and setting it to 0. If the extremum is larger than 0.5 in any dimension, it lies closer to another sample point. In this case, the sample point is changed and the extremum is interpolated about this point. An example is presented for keypoint selection on a natural image of 233 by 189 pixels having 832 keypoints identified by the location, scale, and orientation then threshold-ed on minimum contrast to get 729 keypoints remain. The edge reponses are eliminated using derivative of Hessian matrix following [2]

Step(3): Orientation assignment - Orientation is assigned to keypoints based on local image properties allowing invariance to image rotation. An orientation histogram is constructed and the highest peaks and 80% within it are utilised to create a keypoint.

Step(4): Image Desciptor - Magnitude and orientation of the image gradient around they keypoint is sampled, selecting the amount of blur. Weighted gaussian window covers the regions which yields a summed orientation histogram for the corresponding sub regions. The paper is established a 4x4 descriptor with 8 orientations which is a 128 dimension feature vector.

**Strengths:** The paper provides an original and classical approach for finding distinctive image features which acted as an inspiration for most of the descriptors proposed later. The author proposes an approach to find image features which are invariant to scale and rotation which makes the system robust and cutting edge for its time. The computation of these keypoints is also efficient which leads to several thousands of keypoints to be extracted in real-time. The paper does a good job in explaining without clouding it with excessive mathematical intricacies.

**Weaknesses:** Based on a classical approach with multiple mathematical equations, it is complicated and computationally heavy. Due to the constructed Histogram of Gradients (HOG)- gradient of each pixel needs to be calculated, leads to time intensive computations. It is quite slow compared to other algorithms like SURF which runs faster. Although SIFT boasted its performance with occlusion and blur, it does not really work well as we need it to when there are changes in light and blur. In conclusion, SIFT holds its own to many modern approaches to solve this problem but it is slow, computationally expensive and time intensive, therefore not always the best choice.

**Reflections:** While the authors take a very mathematical approach as compared to existing feature extraction approaches, a simpler approach which which is not as computationally intensive would be more beneficial and also calculated faster. Also, based on the discussion in weaknesses section, if this approach was also imbibed with a better invariability to affine distortion and occlusion, it would justify using this over faster algorithms even though it is expensive.

## References


[1] M. Brown and D. G. Lowe. Invariant features from interest point groups. In *BMVC*, volume 4, 2002.

[2] C. G. Harris, M. Stephens, et al. A combined corner and edge detector. In *Alvey vision conference*, volume 15, pages 10–5244. Citeseer, 1988.

[3] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pages 1150–1157. Ieee, 1999.

[4] H. Moravec. Towards automatic visual obstacle avoidance. In *Proc. 5th Int. Joint Conf. Art. Intell., 1977*, 1977.

[5] C. Schmid, R. Mohr, and C. Bauckhage. Comparing and evaluating interest points. In *Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271)*, pages 230–235. IEEE, 1998.