Distinctive Image Features from Scale-Invariant Keypoints

Summary written by Sanika Phatak

February, 02, 2020

Summary: The paper presents a method (SIFT) to extract distinctive features, invariant to image scale, rotation and translation which can be used for reliable matching between different views of an object or a scene. The paper also shows application SIFT for object recognition which robustly identify objects even in presence of clutter and occlusion while achieving near real time performance.

Related work: Image matching development began with the concept of corner detector by Harris [2] which the authors incorporate as a part of SIFT. Invariant local feature matching commenced with research of Schmid and Mohr [6] for feature matching against large database. The author of this paper, Lowe presents his previous work on local feature approach to achieve scale invariance [4] with improvements in this paper.

Approach: The SIFT algorithm has mainly four steps: (i) Scale-space Extrema Detection - The goal is to identify locations and scales that can represent different views of same scene or object. Gaussian pyramids generated by convolving the image with Gaussian kernels of varying scale are used to represent the scale-space. As proposed in [3], Difference-of-Gaussian (DoG) image are produced by subtracting ocataves of adjacent Gaussians in the pyramid to achieve scale invariance. (ii) Keypoint Localization - The maxima and minima (extrema) of DoG is detected by comparing each point to its 8 neighbours in current scale as well as 9 neighbours in the scales above and below in the pyramid. To get more accurate extrema locations, Taylor series expansion of the scale space is done and only the extrema which cross a threshold are considered to remove low contrast keypoints. Hessian matrix, inspired from [2] is used to eliminate strong responses of DoG to edges. (iii) Orientation Assignments - To achieve invariance to image rotation, gradient magnitude of the neighbourhood around every keypoint is calculated to create a histogram of the local gradients. Highest peak of the histogram along with peaks above a threshold are considered to calculate orientation. This creates keypoints with same scale and locations but different orientations which contributes to the stability of matching. (iv) Keypoint Descriptor - To make the keypoints invariant to changes in viewpoints and illumination, descriptor is computed for local region around a keypoint. Inspired by [1], the region around keypoint is normalized and gradient magnitude and orientation at each point is computed which are weighted by a Gaussian window. Orientation histograms are created for 16 subregions of size 4x4 with each subregion having 8 directions forming a vector of size 128. To make the system affine invariant, [5] approach is implemented where additional SIFT features are generated for 4 different viewpoint changes thus increasing the size of feature vector.

Applications and Results: Object Detection - Keypoint Matching is done by identifying the nearest neighbours, the ratio of closest distance to second closest distance is computed and only the cases less than 0.8 are considered to be correct which eliminates 90% false positives. On application for object recognition for cluttered and occluded image gave correct identification of occluded 3D obejcts. The rotation range for 3D objects is limited to 30 degree. The application of all steps of SIFT can be done in less than 0.3 seconds.

Strengths: The paper provide a unique concept to extract scale and rotation invariant features which are also robust against occlusion and clutter and hence is widel used.This paper proposes the basic concept of SIFT which has been a stepping stone for many more robust and efficient feature extraction algorithms. Good stability and invariance are unique advantages which makes SIFT a popular research topic on which constant research is done to make improvements.

Weaknesses: SIFT has a very high dimensionality of 128 vectors and uses the Histogram of gradients method in which gradients of each pixel in a patch is computed which makes it computationally heavy. The algorithm is only partially invariant to illumination and 3D viewpoint changes which leaves a scope of improvement to the algorithm.

Reflections: This is a very widely used algorithm with vast application like object detection, Robot localization, mapping *etc*. While the paper demonstrates a robust approach for invariant feature extraction, it has several parameters used in the algorithm but the theoretical reasons for the choices are not specified and seem to be empirical. As mentioned earlier, this paper was a gateway to many more improved variants of SIFT and still the research on improving the performance prevails.

References

[1] S. Edelman, N. Intrator, and T. Poggio. Complex cell and object recognition.

[2] C. Harris. A combined corner and edge detector. 1988.

[3] T. Lindeberg. Scale-space theory: A basic tool for analysing structures at different scales. 1994.

[4] D. Lowe. Object recognition from local scale invariant features. 1999.

[5] D. Pritchard and W. Heidrich. Cloth motion capture. pages 263–271, 2003.

[6] C. Schmid and R. Mohr. Local greyvalue invariants for image retrieval. 19(5):530–534, 1997.