INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

SIFT: Scale Invariant Feature transform (Review)

Ridhi Jindal[1] ***Dr. Sonia Vatta***[2]

[1,2] ***School of Computer Science and Engineering,***
Bahra University,
Himachal Pradesh, India.
Corresponding Author: ridhijindal136@gmail.com[1]

ABSTRACT

This paper presents a study on SIFT (Scale Invariant Feature transform) which is a method for extracting distinctive invariant features from images that can be used to perform reliable matching between different views of an object or scene. The features are invariant to image scaling, translation, and rotation, and partially invariant to illumination changes and affine or 3D projection. There are various applications of SIFT that includes object recognition, robotic mapping and navigation, image stitching, 3D modeling, gesture recognition, video tracking, individual identification of wildlife and match moving.

KEYWORDS*: SIFT, Keypoints, Scale, Descriptor, DoG.*

I. INTRODUCTION:

Scale-invariant feature transform (or SIFT) is an algorithm in computer vision to detect and describe local features in images. For any object in an image, interesting points on the object can be extracted to provide a "feature description" of the object. The algorithm was published by David Lowe in 1999[1].This paper gives a review and describes how SIFT extract the image features that have many properties that make them suitable for matching differing images of an object or scene. The features are invariant to image scaling and rotation, and partially invariant to change in illumination and 3D camera view point [2]. They are well localized in both the spatial and frequency domains, reducing the probability of disruption by occlusion, clutter, or noise. Large numbers of features can be extracted from typical images with efficient algorithms.

One of the most important characteristic of these features is that the relative positions between them in the original scene shouldn't change from one image to another. For example, if only the four corners of a square box were used as features, they would work regardless of the square box's position; but if points in the frame were also used, the recognition would fail if the square box is opened or closed. Similarly, features located in articulated or flexible objects would typically not work if any change in their internal geometry happens between two images in the set being processed. However, in practice SIFT detects and uses a much larger number of features from the images, which reduces the contribution of the errors caused by these local variations in the average error of all feature matching errors.

II. STAGES OF COMPUTATION IN SIFT

Fig. 1 shows the key stages of SIFT followed by their description

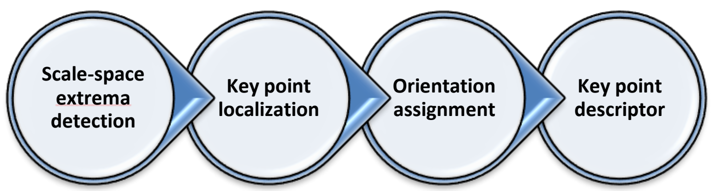

Fig 1: Stages of SIFT

1. **Scale-space extrema detection:** The first stage of computation searches over all scales and image locations. It is implemented efficiently by using a difference-of Gaussian function to identify potential interest points that are invariant to scale and orientation [3] as shown in Fig 2.

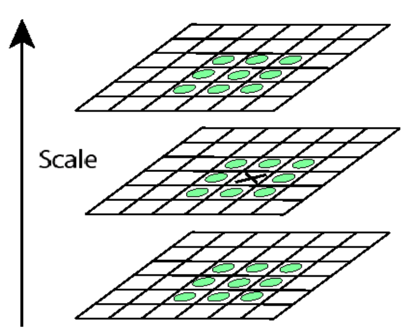

Fig 2: Scale-space extrema detection

2. Key point localization: At each candidate location, a detailed model is fit to determine location and scale. Key points are selected based on measures of their stability [4].

3. Orientation assignment: One or more orientations are assigned to each key point location based on local image gradient directions. All future operations are performed on image data that has been transformed relative to the assigned orientation, scale, and location for each feature, thereby providing invariance to these transformations.

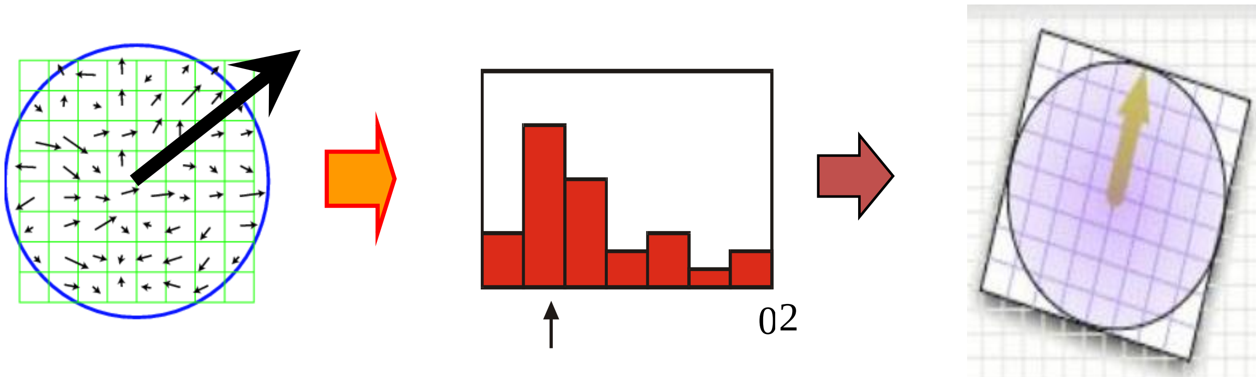

Fig 3: Orientation Assignment

4. Keypoint descriptor: The local image gradients are measured at the selected scale in the region around each key point. These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination.

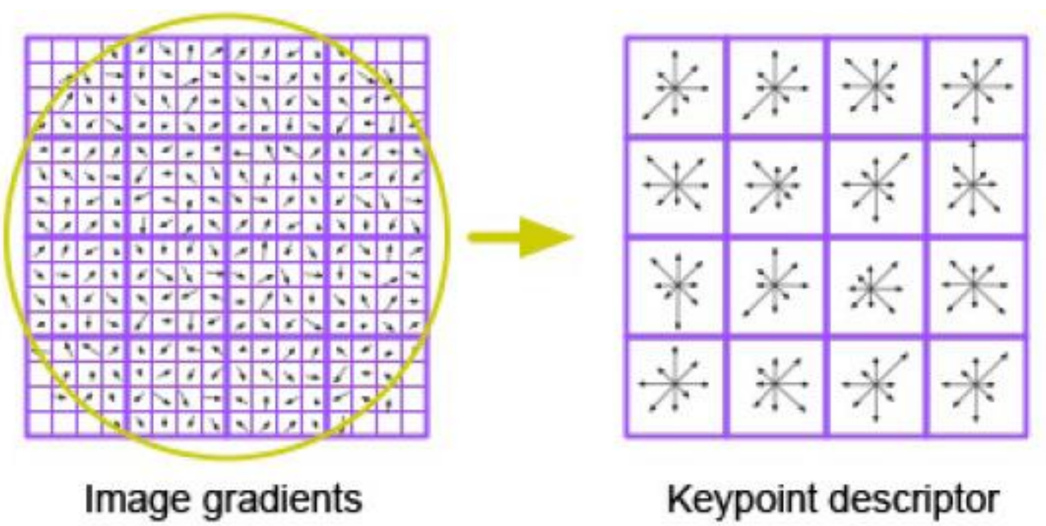

Fig 4: Keypoint Descriptor

III. PROPERTIES OF SIFT

1. Highly distinctive
    a. A single feature can be correctly matched with high probability against a large database of features from many images.
2. Scale and rotation invariant.
3. Partially invariant to 3D camera viewpoint [5]
    a. Can tolerate up to about 60 degree out of plane rotation [6]
4. Partially invariant to changes in illumination
5. Can be computed fast and efficiently.

IV. APPLICATIONS

1. Object recognition using SIFT features [5]

Given SIFT's ability to find distinctive keypoints that are invariant to location, scale and rotation, and robust to affine transformations (changes in scale, rotation, shear, and position) and changes in illumination, they are usable for object recognition.

2. Robot localization and mapping

In this application, a [7] trinocular stereo system is used to determine 3D estimates for keypoints locations. This provides a robust and accurate solution to the problem of robot localization in unknown environments.

3. Panorama stitching

SIFT feature matching can be used in image stitching for fully automated panorama reconstruction from non-panoramic images. The input images can contain multiple panoramas and noise images (some of which may not even be part of the composite image), and panoramic sequences are recognized and rendered as output.[8]

4. 3D scene modeling, recognition and tracking

Extensions of the SIFT descriptor to 2+1-dimensional spatio-temporal data in context of human action recognition in video sequences have been studied. The computation of local position-dependent histograms in the 2D SIFT algorithm are extended from two to three dimensions to describe SIFT features in a spatio-temporal domain [9]. For application to human action recognition in a video sequence, sampling of the training videos is carried out either at spatio-temporal interest points or at randomly determined locations, times and scales. The spatio-temporal regions around these interest points are then described using the 3D SIFT descriptor. These descriptors are then clustered to form a spatio-temporal Bag of words model. 3D SIFT descriptors extracted from the test videos are then matched against these words for human action classification. The authors report much better results with their 3D SIFT descriptor approach than with other approaches like simple 2D SIFT descriptors and Gradient Magnitude.

5. Analyzing the Human Brain in 3D Magnetic Resonance Images

FBM was validated in the analysis of AD using a set of ~200 volumetric MRIs of the human brain, automatically identifying established indicators of AD in the brain and classifying mild AD in new images with a rate of 80% [10].

V. REFRENCES

[1]. Lowe, David G. (1999). "Object recognition from local scale-invariant features". Proceedings of the International Conference on Computer Vision 2. pp. 1150–1157.

[2]. "Method and apparatus for identifying scale invariant features in an image and use of same for locating an object in an image", David Lowe's patent for the SIFT algorithm, March 23, 2004.

[3]. Lowe, D. G., “Distinctive Image Features from Scale-Invariant Keypoints”, International Journal of Computer Vision, 60, 2, pp. 91-110, 2004.

[4]. Serre, T., Kouh, M., Cadieu, C., Knoblich, U., Kreiman, G., Poggio, T., “A Theory of Object Recognition: Computations and Circuits in the Feedforward Path of the Ventral Stream in Primate Visual Cortex”, Computer Science and Artificial Intelligence Laboratory Technical Report, December 19, 2005 MIT-CSAIL-TR-2005-082.

[5]. Beis, J., and Lowe, D.G “Shape indexing using approximate nearest-neighbour search in high-dimensional spaces”, Conference on Computer Vision and Pattern Recognition, Puerto Rico, 1997, pp. 1000–1006.

[6]. Lowe, D.G., Local feature view clustering for 3D object recognition. IEEE Conference on Computer Vision and Pattern Recognition,Kauai, Hawaii, 2001, pp. 682-688.

[7]. Lowe, David G.; Little, J. (2001). "Vision-based mobile robot localization and mapping using scale-invariant features". Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) 2. p. 2051

[8]. Lowe, David G. (2003). "Recognising Panoramas". Proceedings of the ninth IEEE International Conference on Computer Vision 2. pp. 1218–1225.

[9]. Flitton, G.T., Breckon, T.P., Megherbi, N. (2013). "A Comparison of 3D Interest Point Descriptors with Application to Airport Baggage Object Detection in Complex CT Imagery". Pattern Recognition (Elsevier).

[10]. Niebles, J. C. Wang, H. and Li, Fei-Fei (2006). "Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words". Proceedings of the British Machine Vision Conference (BMVC). Edinburgh. Retrieved 2008-08-20.

.