

DIP Assignment - 2

Ebenezer Isaac Veeraraju (2020178014)

SIFT

David Lowe devised the scale-invariant feature transform (SIFT) in 1999 as a computer vision algorithm for detecting, describing, and matching local features in images. Object recognition, robotic mapping and navigation, picture stitching, 3D modelling, gesture recognition, video tracking, individual wildlife identification, and match moving are some of the applications.

A set of reference photos is used to extract SIFT key points of objects, which are then saved in a database. An object in a new image is recognised by comparing each feature in the new image to this database and discovering candidate matching features based on the Euclidean distance between their feature vectors. Subsets of important points that agree on the item and its location, scale, and orientation in the new image are identified from the whole set of matches to filter out good matches. Using an efficient hash table implementation of the generalised Hough transform, consistent clusters can be determined quickly.

Outliers are removed after each cluster of three or more features that agree on an object and its pose is subjected to more extensive model testing. Finally, given the precision of the fit and the number of likely false matches, the likelihood that a certain set of features indicates the presence of an object is calculated. Object matches that pass all of these checks can be confidently identified as correct.

Interesting spots on any object in an image can be extracted to produce a "feature description" of the thing. When attempting to locate an object in a test image including numerous other objects, this description taken from a training image can be utilised to identify the object. It is critical that the features derived from the training image be visible even under variations in image scale, noise, and illumination in order to conduct reliable recognition. These spots are generally seen in high-contrast areas of the image, such as the edges of objects.

Another key property of these elements is that their relative placements in the original picture should remain constant from one image to the next. If only the four corners of a door were used as features, the recognition would succeed regardless of the door's position; but, if points in the frame were also used, the recognition would fail regardless of whether the door was open or closed. Similarly, features in articulated or flexible objects are unlikely to operate if the interior geometry of the object changes between two photos in the processing set. SIFT, on the other hand, detects and uses a significantly higher number of features from the images in practise, which minimises the impact of local fluctuations in the average error of all feature matching mistakes.

Because the SIFT feature descriptor is invariant to uniform scaling, orientation, illumination changes, and partially invariant to affine distortion, it can reliably recognise objects even among clutter and under partial occlusion. The original SIFT algorithm is summarised in this section, along with a few competing algorithms for object recognition in clutter and partial occlusion.

SURF

Speeded up robust features (SURF) is a proprietary local feature detector and descriptor in computer vision. It can be used for object detection, image registration, categorization, and 3D reconstruction, among other things. The scale-invariant feature transform (SIFT) descriptor influenced it in part. SURF's standard version is several times faster than SIFT, and its inventors believe that it is more resilient against various picture alterations than SIFT.

SURF uses an integer approximation of the determinant of the Hessian blob detector to detect interest points, which may be computed with three integer operations using a precomputed integral image. The sum of the Haar wavelet response surrounding the point of interest is used as its feature descriptor. These can also be calculated using the integral image.

SURF descriptors have been used to find and recognise items, persons, and faces, recreate 3D scenes, track objects, and extract points of interest, among other things.

SURF was first presented at the 2006 European Conference on Computer Vision by Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. In the United States, an application of the algorithm has been patented. Because a "upright" version of SURF (dubbed U-SURF) is not invariant to picture rotation, it is faster to compute and better suited for applications where the camera is kept more or less horizontal.

Using the multi-resolution pyramid approach, the image is translated into coordinates in order to reproduce the original image with a Pyramidal Gaussian or Laplacian Pyramid shape, resulting in an image with the same size but lower bandwidth. This creates a Scale-Space blurring effect on the original image, ensuring that the points of interest remain scale invariant.

The SURF algorithm follows the same concepts and phases as SIFT, however there are some differences in the details of each step. Interest point detection, local neighbourhood description, and matching are the three primary components of the algorithm.