INITIAL PROJECT PROPOSAL
Distinctive Image Features from Scale-Invariant Keypoints

Project ID - 3 | Team name - Team Agile

Samruddhi Shastri (2019111039)
Pranjali Pathre (2019112002)
Anandhini Rajendran (2019101055)
Ayush Goyal (2019111026)

Github link

https://github.com/pranjali-pathre/SMAI-Project

Problem Statement

Image matching is a fundamental aspect of many problems in computer vision, including object or scene recognition, solving for 3D structure from multiple images, stereo correspondence, and motion tracking.

We present 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 obtained features are invariant to image scaling and rotation, partially invariant to illumination, well localized in spatial and frequency domains, and highly distinct which allows us to match features with high accuracy. The cost of extracting the features is reduced by using a cascade filtering approach.

Goals and Approach

SIFT Feature Extraction is implemented in the following steps:

1. Scale-space extrema detection
2. Keypoint localization
3. Orientation assignment
4. Keypoint descriptor

These sections are briefly described below.

Scale-space extrema detection:

As a first step, we search for potential interest points over all the scales and image locations which in the next step can be used to determine key points. We approach this by using a cascade filtering approach to identify candidate locations. We use scale space to detect locations that are invariant to scale changes to find stable locations. The concept of Scale Space deals with the application of a continuous range of Gaussian Filters to the target image such that the chosen Gaussian have differing values of the sigma parameter.

1. We use the Gaussian function as a scale-space kernel. So the scale space of an image is defined as a result of convolution of a variable-scale Gaussian, *G(x, y, σ)* with an input image.

$$L(x, y, σ) = G(x, y, σ) * I(x, y),$$

2. We use scale-space extrema, i.e., the convolved Gaussian difference to detect the stable points.

$$D(x, y, σ) = (G(x, y, kσ) − G(x, y, σ)) * I(x, y)$$

The maxima and the minima $σ^2 \nabla^2 G$ produce the most stable image features.

$$G(x, y, kσ) − G(x, y, σ) ≈ (k − 1)σ^2 \nabla^2 G$$

In 2D images, we can detect the Interest Points using the local maxima/minima in the Scale Space of Laplacian of Gaussian. A potential SIFT interest point is determined for a given sigma value by picking the potential interest point and considering the pixels in the level above (with higher sigma), the same level, and the level below (with lower sigma than the current sigma level). If the point is maxima/minima of all these 26 neighbouring points, it is a potential SIFT interest point – and it acts as a starting point for interest point detection.

Keypoint localization:

This step compares the pixel to its neighbours for location, scale, and the ratio of principal curvatures.
This step allows us to reject points of low contrast (highly sensitive to noise) or poorly localized along the edge.

Implementation:
1. Initial approach: Locate key points at a location and scale the central sample point.
2. Improved approach: Fitting a 3D quadratic function to the local sample points to determine the interpolated location of the maximum. This method gives us better matching and stability.

Method:

- Scale the space function so that the origin is at the sample point: calculate the D (Taylor's expansion) and its derivatives, offset, and extremum.
- The function value at the extremum is used to reject unstable extrema with low contrast

For stability removing the points with low contrast alone is not sufficient. The difference of gaussian function will be unstable to noise since it has a strong response along the edges. The poorly defined peaks in the Gaussian function are found by the ratio of principal curvatures of the peak. If the ratio is less than a threshold value then the point is discarded. The ratio can be calculated using the eigenvalues of the Hessian matrix computed at the location and scale of the keypoint.

Orientation Assignment

In this step, we assign a consistent orientation to each keypoint. The keypoint descriptor is represented relative to this orientation and therefore achieves invariance to image rotation.

Implementation:
1. The scale of the keypoint is used to select the Gaussian smoothed image with the closest scale ensuring that all computations are performed in a scale-invariant manner.
2. An orientation histogram is formed from the gradient orientations of sample points within a region around the key point. There are 36 bins in total, which span the entire 360-degree range of orientation.
3. The highest peak in the histogram is identified, and then any other local peak within 80% of the highest peak is used to also create a key point with that orientation.
4. Finally, a parabola is fit to the three histogram values closest to each peak to interpolate the peak position for better accuracy.

Keypoint Descriptor

In this step, we compute a descriptor for the local image region that is extremely distinctive but is invariant to other variations, such as a change in illumination or 3D viewpoint.
For each keypoint, a descriptor is created using the key points in the neighbourhood. These descriptors are used for matching keypoints across images. A 16×16 neighbourhood of the keypoint is used for defining the descriptor of that keypoint. This neighbourhood is divided into sub-blocks. Each such sub-block is a non-overlapping, contiguous, 4×4 neighbourhood. Subsequently, for each sub-block, an 8 bin orientation is created similarly as discussed in Orientation Assignment. These 128 bin values (16

sub-blocks * 8 bins per block) are represented as a vector to generate the keypoint descriptor.
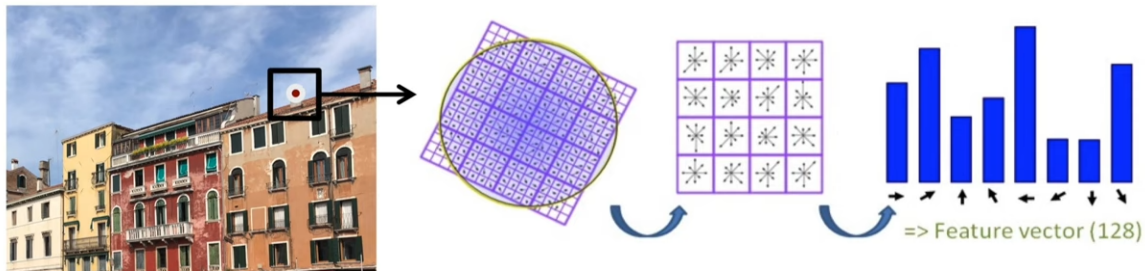
Dataset

We plan to use the Caltech-256 dataset for the given task. It is an object recognition dataset containing 30,607 real-world images, of different sizes, spanning 257 classes (256 object classes and an additional clutter class). Each class is represented by at least 80 images.

Deliverables

1. A working implementation of SIFT algorithm.
2. Features extracted from the images in the dataset.
3. Presentation and description of the project along with instructions to run the demo.
4. A brief comparison with other feature extraction algorithms (Harris Corner Detection and ORB/SURF).
5. A working implementation of the feature matching algorithm. (If time permits)
6. Corresponding features extracted from images after applying feature matching algorithm. (If time permits)

Milestones and Timeline

Timeline		Milestones
26 Oct - 7 Nov	Week 0	Project allocation and project proposal submission, discussions, and initial project layout
7 Nov - 14 Nov	Week 1	Paper and relevant work reading, Implementing Scale Space and Image Pyramids
14 Nov - 21 Nov	Week 2	Finding Scale Space Extrema, Mid evaluation, Implementing Keypoint Orientations, Cleaning Up Keypoints

21 Nov - 28 Nov	Week 3	Generating Descriptors, Implementing applications: Template Matching, Comparison with other algorithms
29 Nov - 30 Nov	Week 3+	Compiling results, Finishing steps and Submission
1 Dec - 4 Dec		Final evaluation

Work distribution

Samruddhi	Implementing Scale Space, Finding Scale Space Extrema, Generating Descriptors, Comparison with other algorithms (ORB/SURF).
Pranjali	Implementing Scale Space, Implementing Keypoint Orientations, Cleaning Up Keypoints, Template Matching.
Anandhini	Implementing Image Pyramid, Finding Scale Space Extrema, Generating Descriptors, Template Matching.
Ayush	Implementing Image Pyramid, Implementing Keypoint Orientations, Cleaning Up Keypoints, Comparison with other algorithms (Harris Corner Detection).

References

Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60, 91–110 (2004). https://doi.org/10.1023/B:VISI.0000029664.99615.94

SIFT - 5 Minutes with Cyrill, Youtube. Uploaded by Cyrill Stachniss, 15 Sept 2020 https://www.youtube.com/watch?v=4AvTMVD9ig0&list=PLgnQpQtFTOGSO8HC48K9sPuNliY1qxzV9&index=20

Cao, L., Liao, D., & Xue, B. D. (2014). Reference Point-Based SIFT Feature Matching. In Applied Mechanics and Materials (Vols. 543–547, pp. 2670–2673). Trans Tech Publications, Ltd. https://doi.org/10.4028/www.scientific.net/amm.543-547.2670