Beer Label Classification for Mobile Applications

Team Members

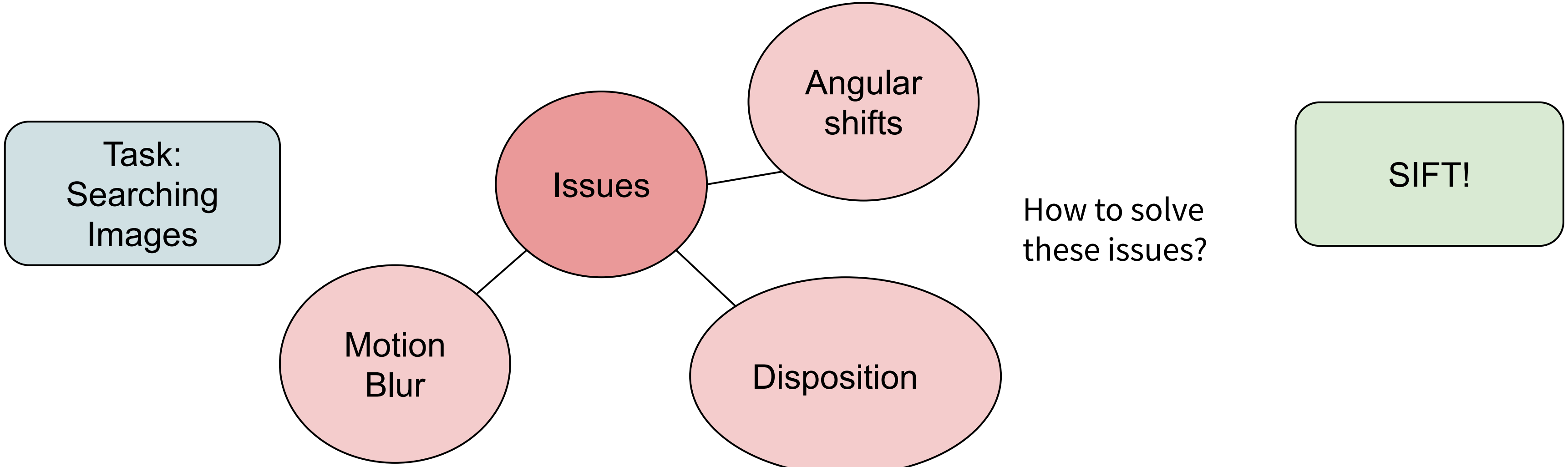
- Madhav Agarwal (2020900022)
- Siddhant Bansal (2019900091)
- Garima Nishad (2019701029)
- Mundru Yallamanda Rao (2019201029)

Goal of the project

- Create an image processing algorithm for the automated identification of beer types using SIFT-based image matching of bottle labels.

Problem definition

- The main problems associated with this task are challenges that are faced in image search in general i.e., searching images irrespective of angle, disposition in the input image.
- In this project we choose to implement SIFT to address the issues w.r.t image search.



DATABASE CREATION AND PRE-PROCESSING

- We **have generated** the database containing bottle and label images.
- No more than 5 labels are from the same brewery.
- For each database image, a corresponding query (test) image of a beer bottle with that label was found.

Query Images

Database

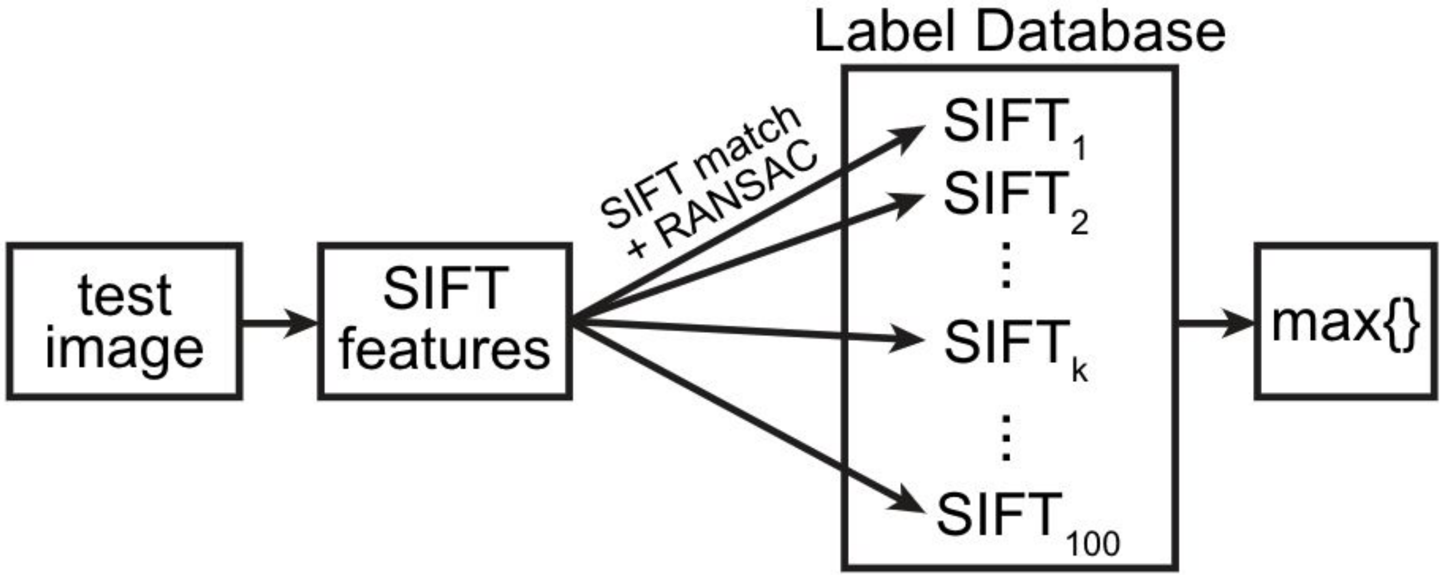
Current data collection status

- We have collected * images in both database and query. Here are some samples:

Query images

Database images

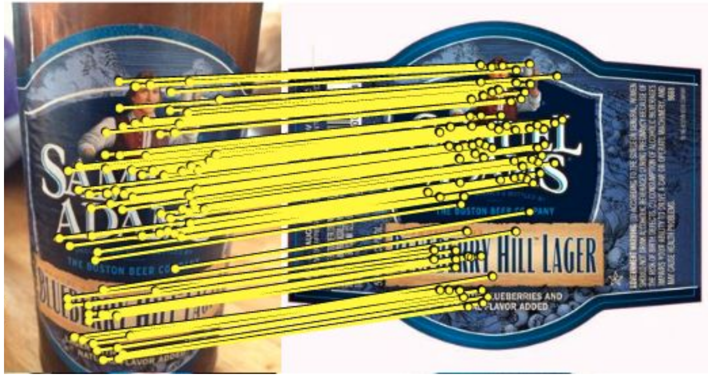
Processing strategy

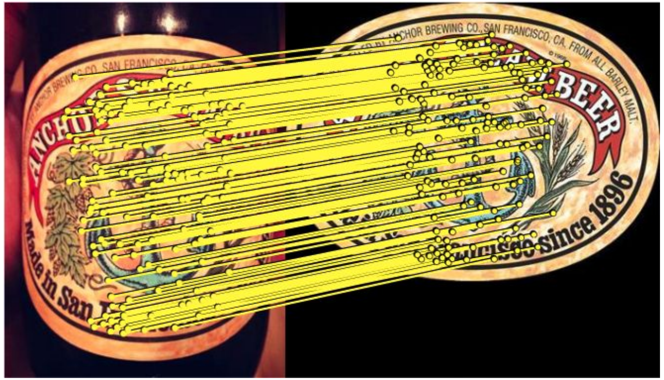


As shown, we'll be extracting SIFT features from the test image. Then, we will match those features with the SIFT features of images from the database. Images having similar features will be queried.

SIFT and Descriptor

- **SIFT keypoints** will be first extracted from all the database images.
- Once SIFT keypoints are identified, a **descriptor** is computed for each of them.
- A descriptor in our case is an **8-bin histogram** will be created for a 4x4 space around the keypoint at its specific scale.
- At a high level, our algorithm will operate by finding the database image that shares the **highest number of SIFT feature** matches with the query image.





SIFT

SIFT - Keypoint Extraction

- Stands for Scale Invariant Feature Transform;
    - Harris Detector is not a scale invariant transform. However, SIFT is.

- Process is similar to how primate visual system works;

- Transforms image data into scale-invariant coordinates.

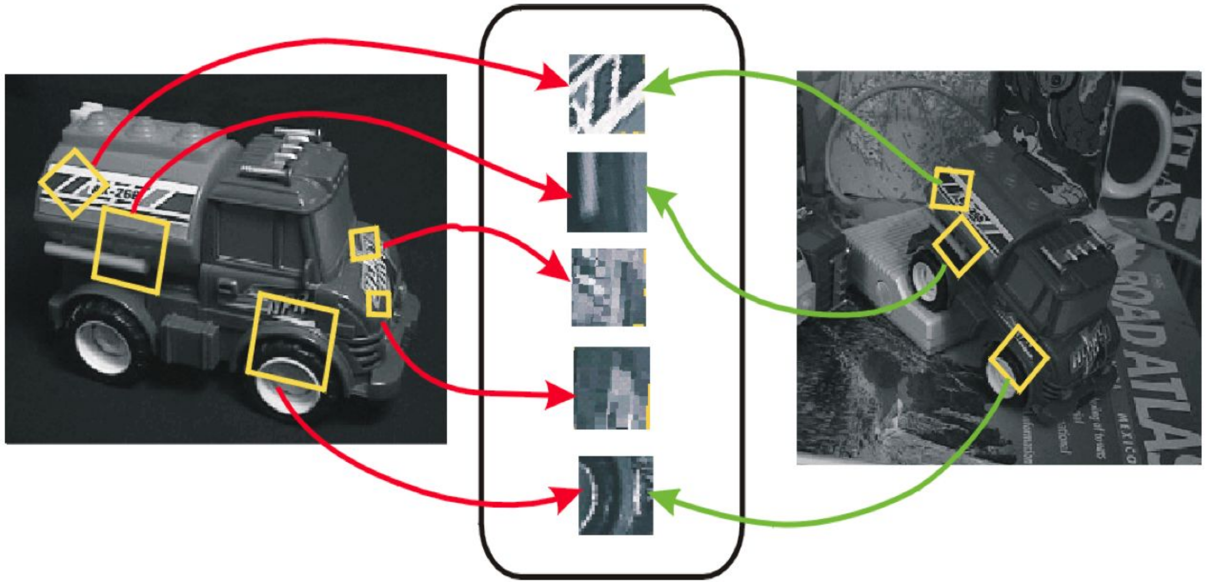
What do we want to achieve?

- Extract distinctive invariant features;
  - Which will help in matching a single image with a large database of features from many images.
- Invariance to image scale and rotation;
- Robustness to
  - Affine distortion,
  - Change in 3D viewpoint,
  - Addition to noise,
  - Changes in illumination.

Advantages of SIFT

- Locality
  - Features are local, so robust to occlusion and clutter
- Distinctiveness
  - Individual features can be matched to a large database of images
- Quantity
  - Many features can be generated for small objects
- Efficiency
  - Provides close to real-time performance

Features used for Object Recognition



Feature Descriptors

https://courses.cs.washington.edu/courses/cse455/09wi/Lects/lect6.pdf

Steps for Extracting the Keypoints

- Scale-space peak selection
  - Construct a set of progressively Gaussian blurred images
  - Take differences to get a “difference of Gaussian” pyramid
  - Find local-extrema in this scale-space.
- Keypoint localization
  - Accurately locating the feature key points
- Orientation Assignment
  - Assigning orientation to the key points
- Keypoint descriptor
  - Describing the keypoint as a high dimensional vector

Scale-space peak selection

- Construct a set of progressively Gaussian blurred images;
- Take differences to get a difference-of-Gaussian pyramid;
- Now the task is to detect local maxima and minima.

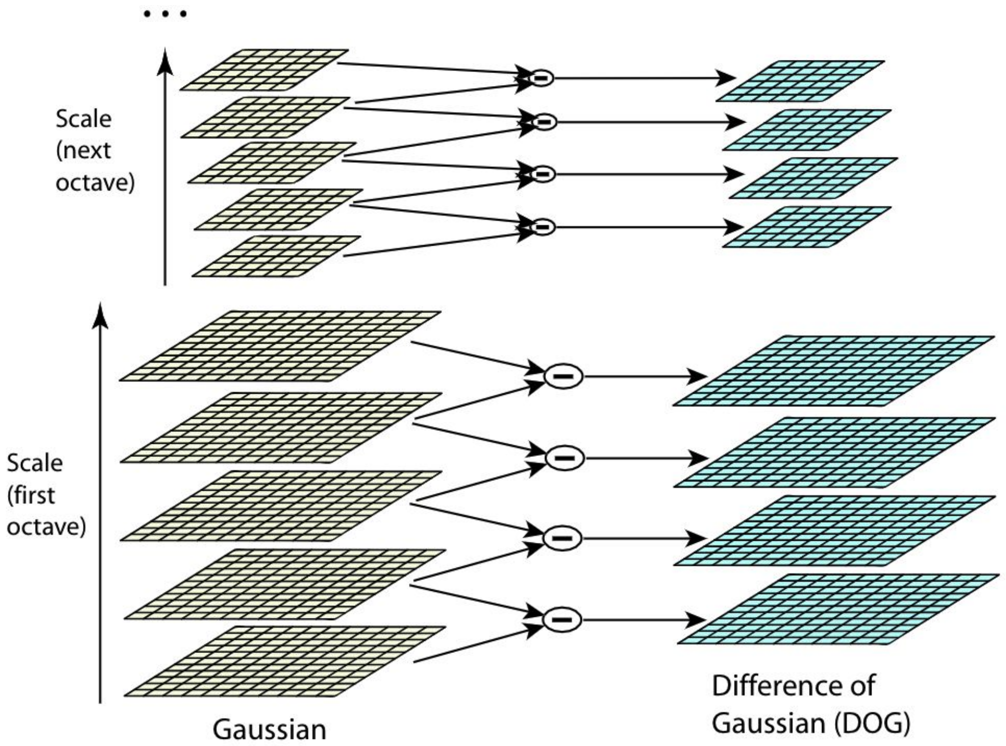


Figure 1: For each octave of scale space, the initial image is repeatedly convolved with Gaussians to produce the set of scale space images shown on the left. Adjacent Gaussian images are subtracted to produce the difference-of-Gaussian images on the right. After each octave, the Gaussian image is down-sampled by a factor of 2, and the process repeated.

https://www.ijcai.org/Proceedings/83-2/Papers/091.pdf

Local extrema detection

- To detect the local maxima and minima, each sample point is compared to its eight neighbors in the current image and nine neighbors in the scale above and below;
- This point will only be selected if it is larger or smaller than all of these neighbors;
- It is important to determine the frequency of sampling in the image and scale domains that is needed to reliably detect the extrema.

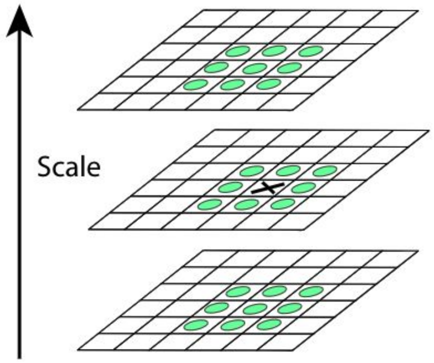
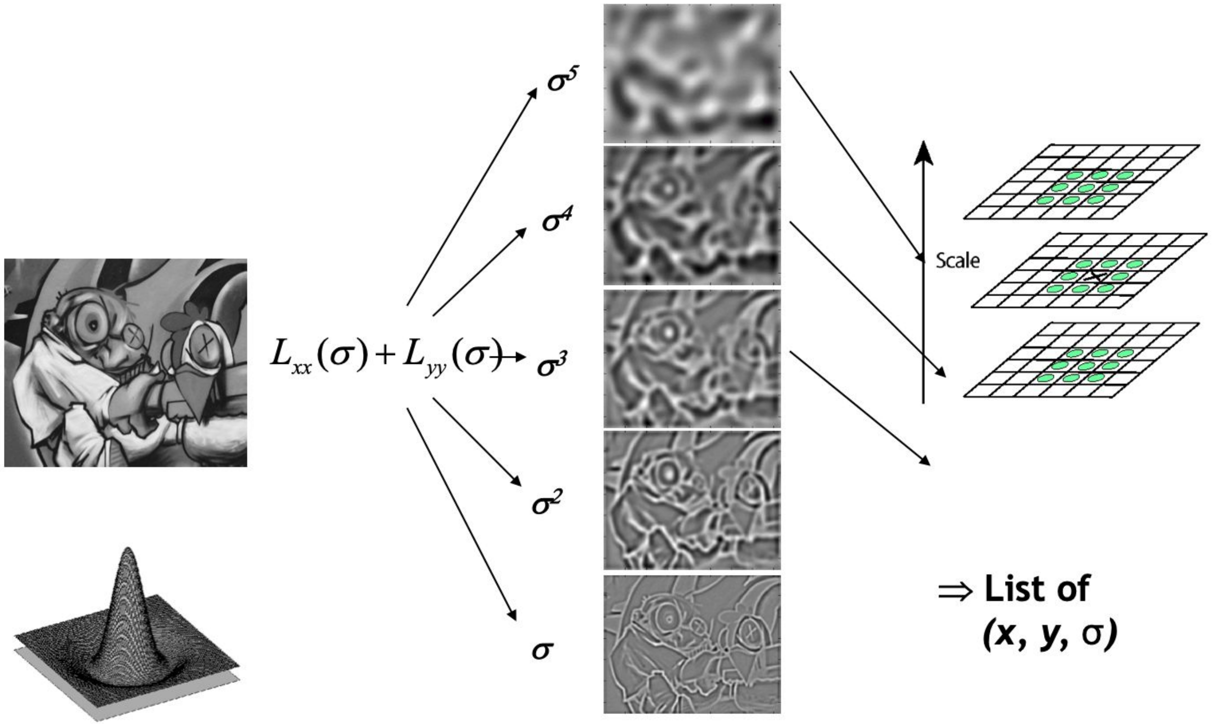


Figure 2: Maxima and minima of the difference-of-Gaussian images are detected by comparing a pixel (marked with X) to its 26 neighbors in 3x3 regions at the current and adjacent scales (marked with circles).

How is scale selection helping?



https://youtu.be/U0wqePj4Mx0

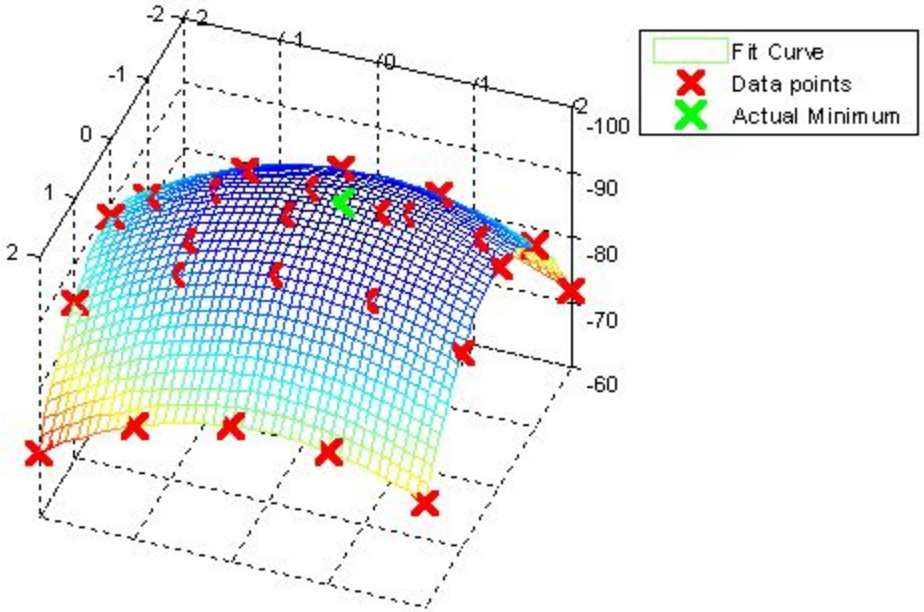
Keypoint Localization

- Fit a quadratic function to the surrounding values (of extremas found in the previous steps) for sub-pixel and sub-scale interpolation.
- Perform Taylor series expansion around that point.

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

- The local point of extremum, is determined by taking the derivative of this function and setting it to 0.

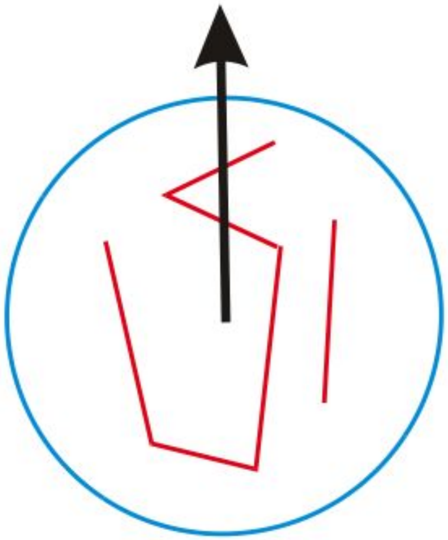
$$\hat{\mathbf{x}} = -\frac{\partial^2 D}{\partial \mathbf{x}^2}^{-1} \frac{\partial D}{\partial \mathbf{x}}.$$

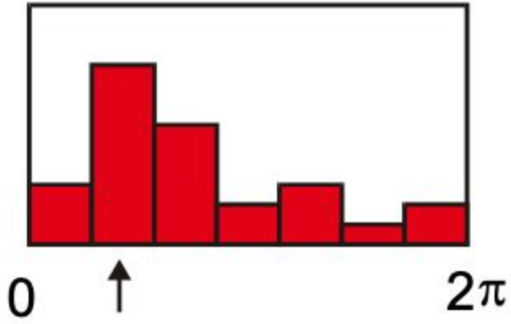


Brown, Matthew, and David G. Lowe. "Invariant features from interest point groups." BMVC. Vol. 4. 2002.

Orientation Assignment

- A consistent orientation is assigned to keypoint for achieving invariance to image rotation;
- Histogram of local gradient directions is computed at the selected scale;
- Now, each keypoint specifies stable 2D coordinates. Which are, x, y, scale, and orientation;
- Select the dominant orientation and normalizing by rotating to the fixed orientation.





https://filebox.ece.vt.edu/~jbhuang/teaching/ece5554-4554/fa16/lectures/Lecture_07_InterestPoints.pdf

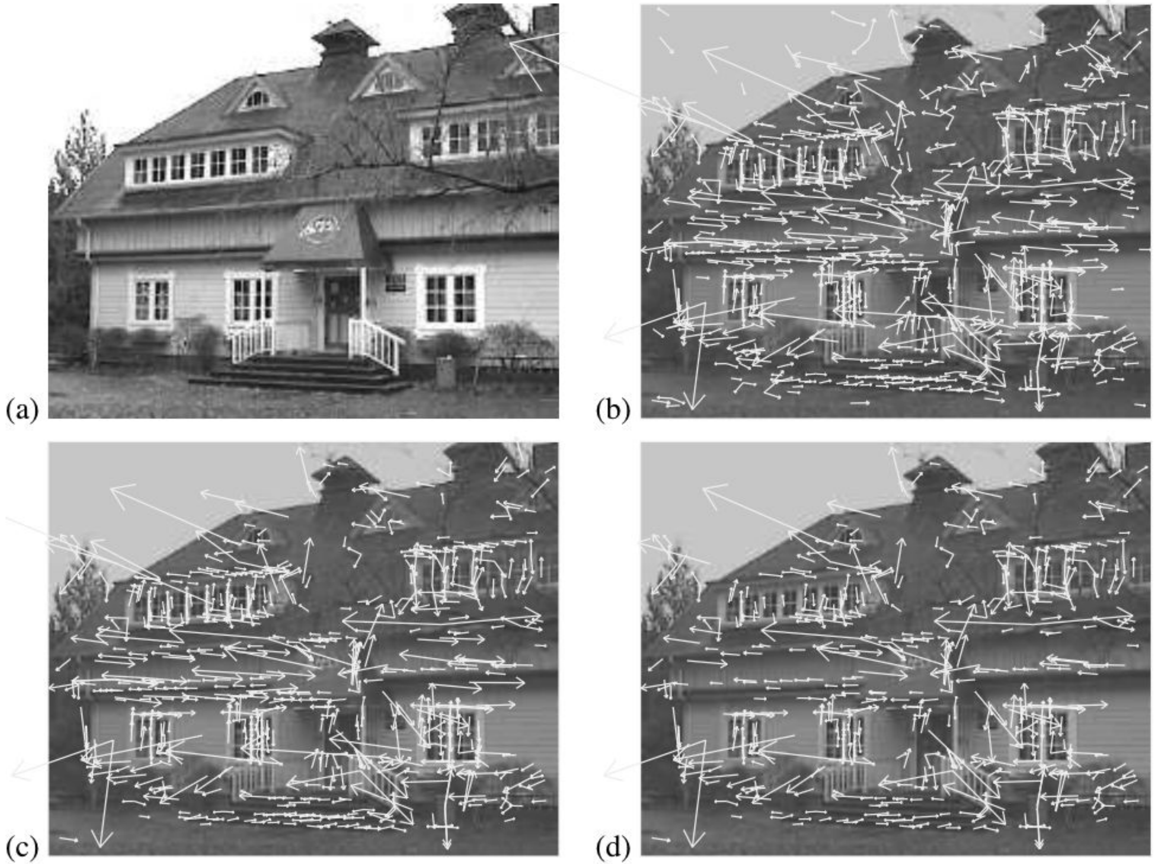
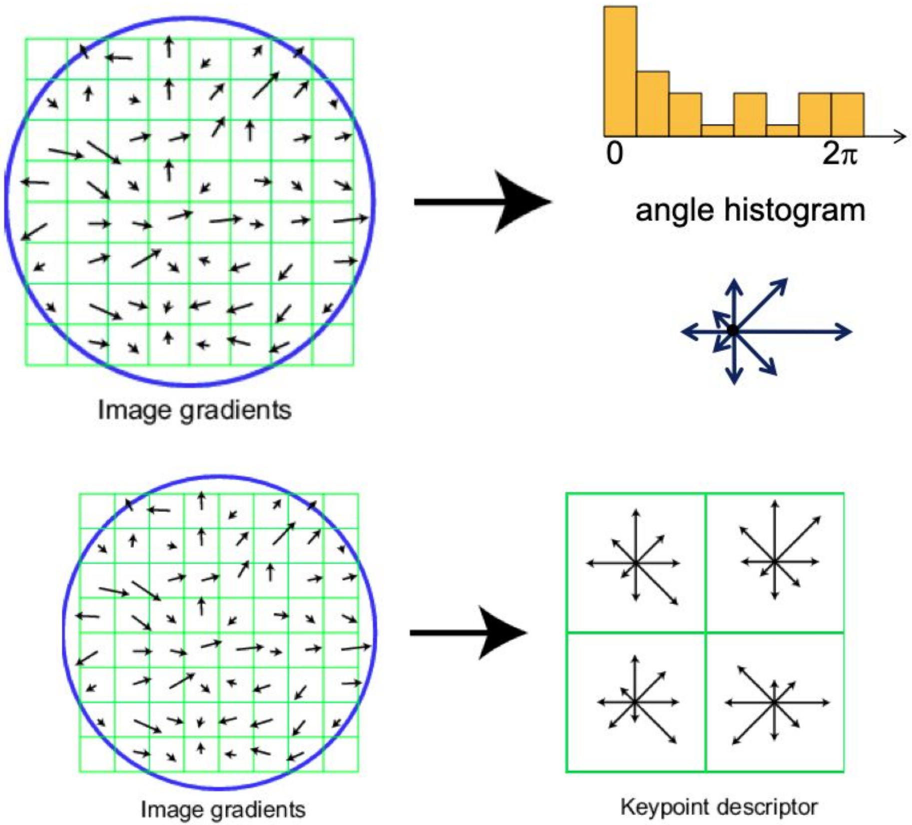


Figure 5: This figure shows the stages of keypoint selection. (a) The 233x189 pixel original image. (b) The initial 832 keypoints locations at maxima and minima of the difference-of-Gaussian function. Keypoints are displayed as vectors indicating scale, orientation, and location. (c) After applying a threshold on minimum contrast, 729 keypoints remain. (d) The final 536 keypoints that remain following an additional threshold on ratio of principal curvatures.

Keypoint descriptor

- Take 16x16 square window around detected feature;
- Compute edge orientation (angle of the gradient - 90°) for each pixel;
- Throw out weak edges;
- Create histogram of surviving edge orientations;
- For SIFT, 8 orientations x 4x4 histograms array results in 128 dimensions.



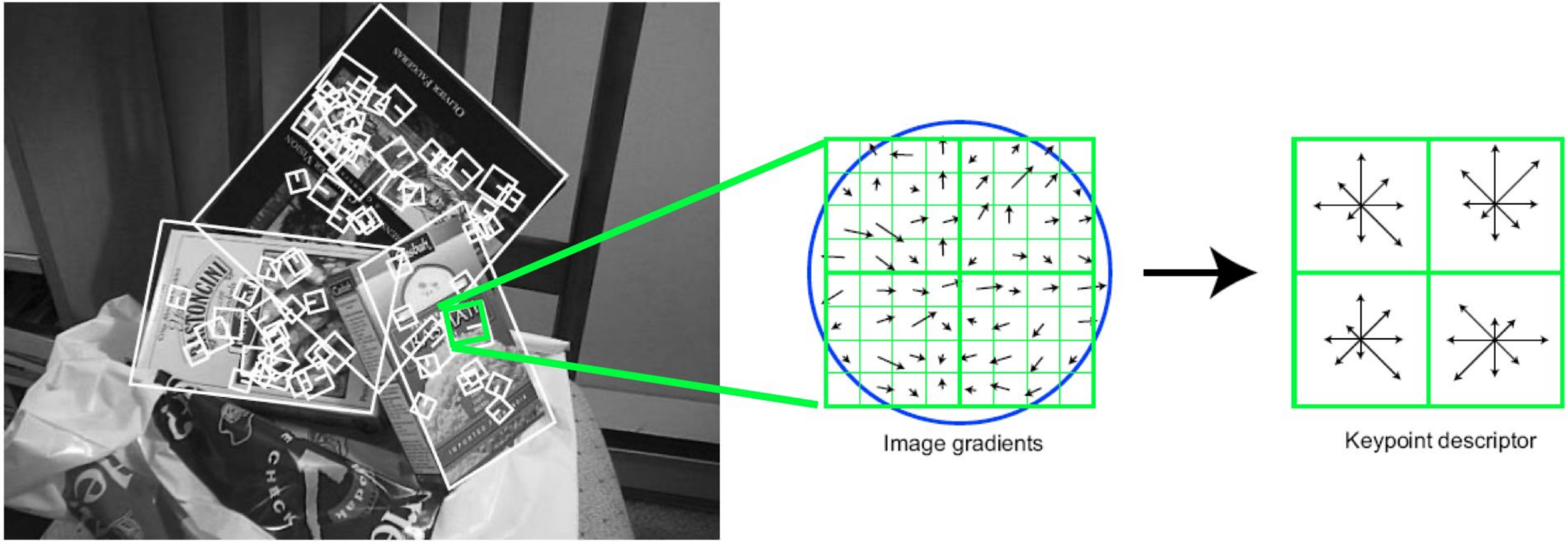


Image gradients
Keypoint descriptor

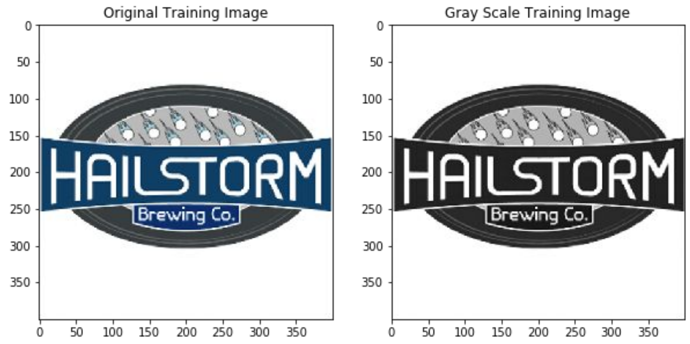
ORB(Oriented FAST and Rotated BRIEF)

- ORB is a very fast algorithm that creates feature vectors from detected keypoints.
- It is invariant to rotations, changes in illumination, and noise.
- The first step in the ORB algorithm is to locate all the keypoints in the training image.
- After the keypoints have been located, ORB creates their corresponding binary feature vectors and groups them together in the ORB descriptor.

ORB(Oriented FAST and Rotated BRIEF)

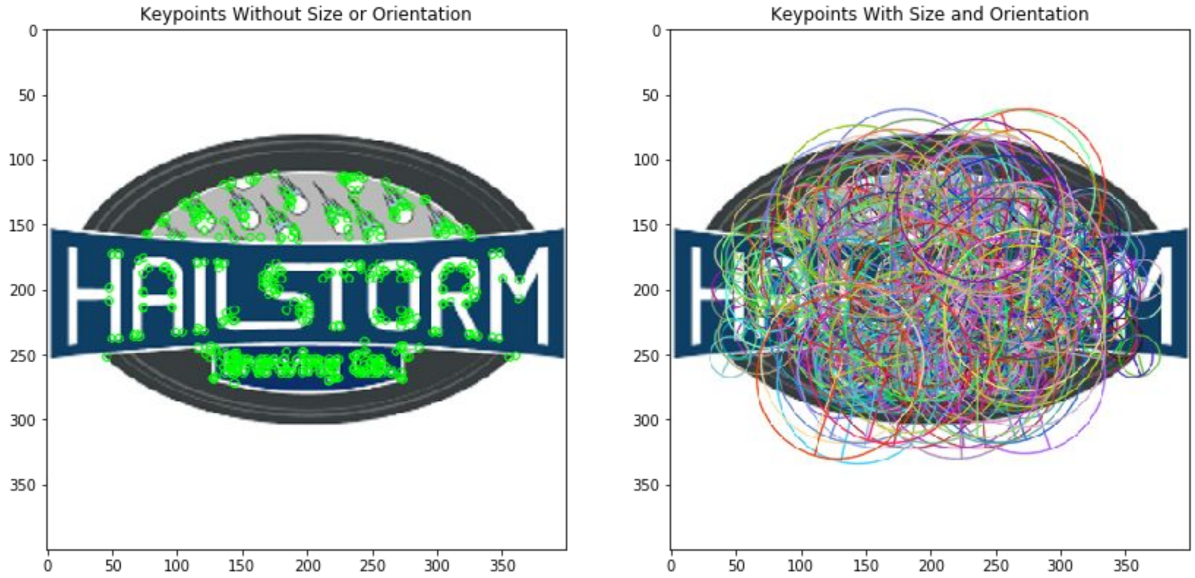
Algorithm:

- Take the query image and convert it to grayscale.
- Now Initialize the ORB detector and detect the keypoints in query image and scene.
- Compute the descriptors belonging to both the images.
- Match the keypoints using Brute Force Matcher.
- Show the matched images.



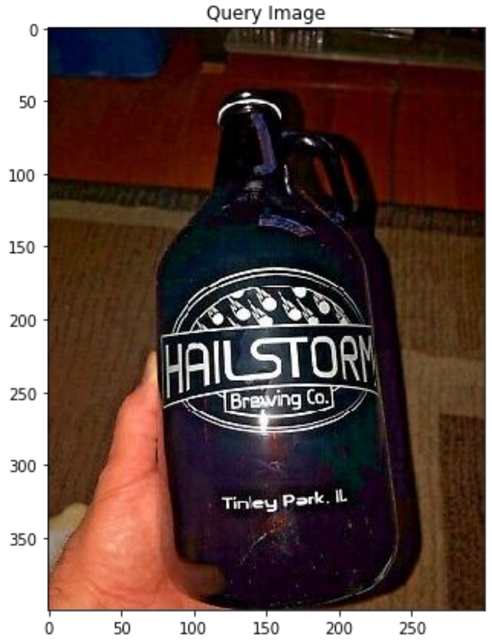
Locating Keypoint

- Rright image, every keypoint has a center, a size, and an angle.
- Once the keypoints for the training image have been found and their corresponding ORB descriptor has been calculated, the same thing can be done for the query image.



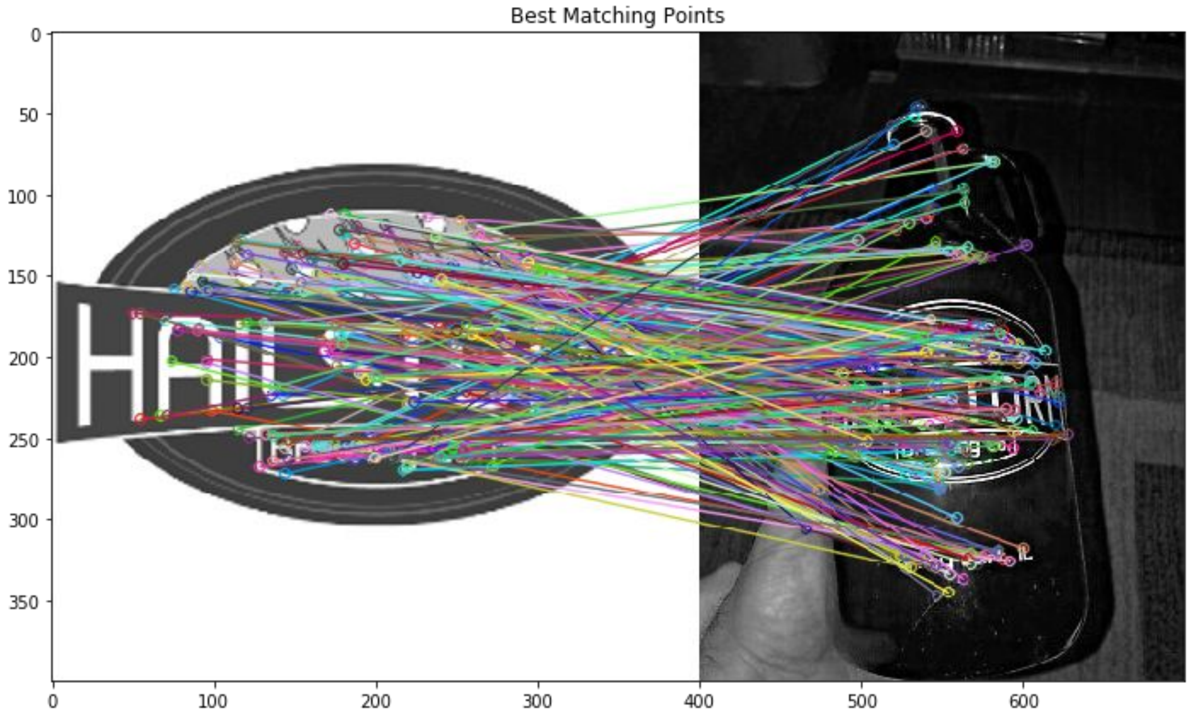
Feature Matching

- The final step is to perform keypoint matching between the two images using their corresponding ORB descriptors.



Feature Matching

- This *matching* is usually performed by a matching function.
- One of the most commonly used matching functions is called *Brute-Force*.

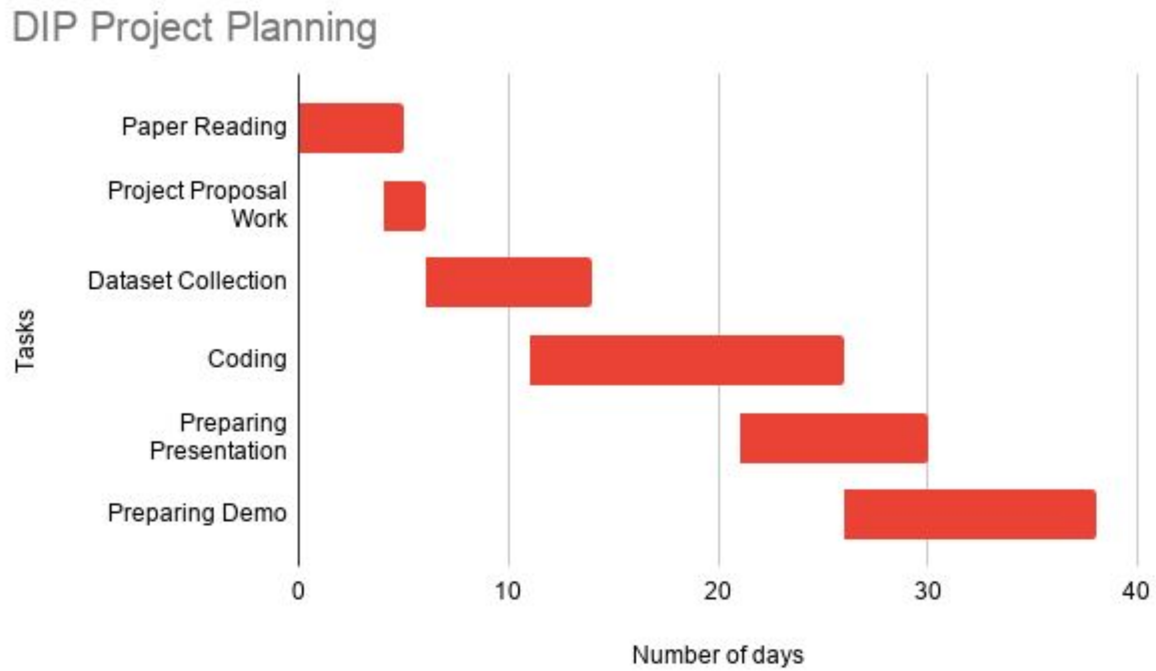


Next Steps

- Finalize the dataset;
- Implement the SIFT algorithm and descriptor;
- Run experiments mentioned in the paper;
- Generate various insights;
- Check other use cases.

Action Plan

- This is how we plan to finish the project.



Cheers!