

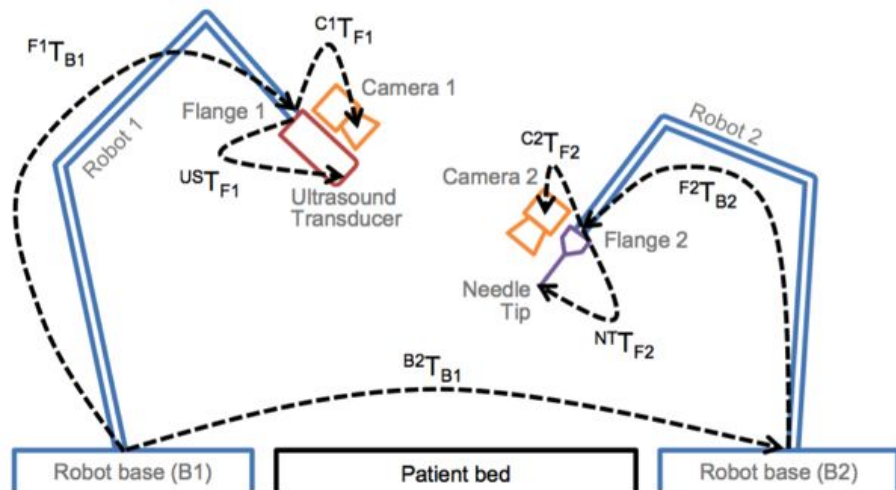
SURF: Speeded-Up Robust Features

Christopher Hunt

Project 17: Robotic Ultrasound Needle Placement and Tracking

Project Introduction

CAMP lab has developed a dual-robotic ultrasound-guided needle placement framework. Our project is the development and exploration of various robot-robot calibration algorithms. One calibration plugin being developed is based on feature matching using an Intel RealSense RGB-D camera.



Paper

Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. "SURF: Speeded Up Robust Features." *Computer Vision – ECCV 2006 Lecture Notes in Computer Science* (2006): 404-17. Web.



What is SURF?

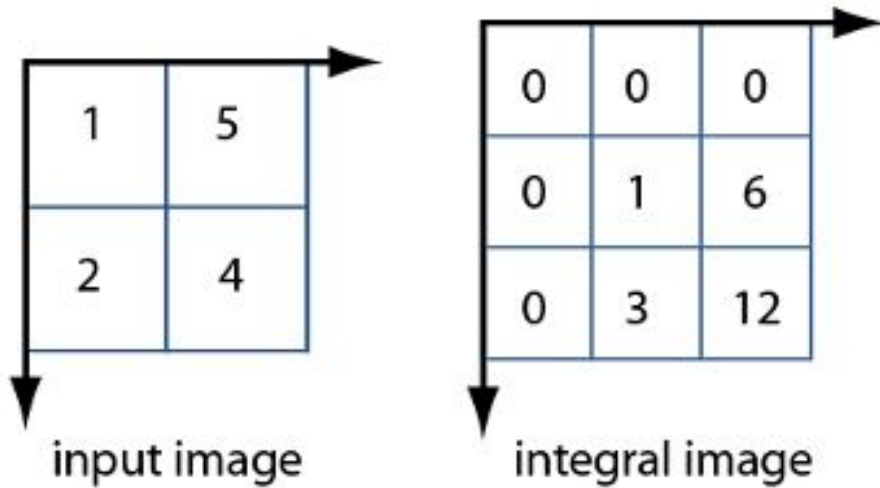
SURF (**S**peeded-**U**p **R**obust **F**eatures) is a feature detection framework introduced by Herbert Bay and his colleagues at ETH Zurich. SURF interest points are in-plane rotation-invariant, robust to noise, and overall, extremely fast to calculate. This procedure can be divided into three steps:

1. Interest Point Detection
2. Interest Point Description
3. Interest Point Matching



Detection: Integral Images

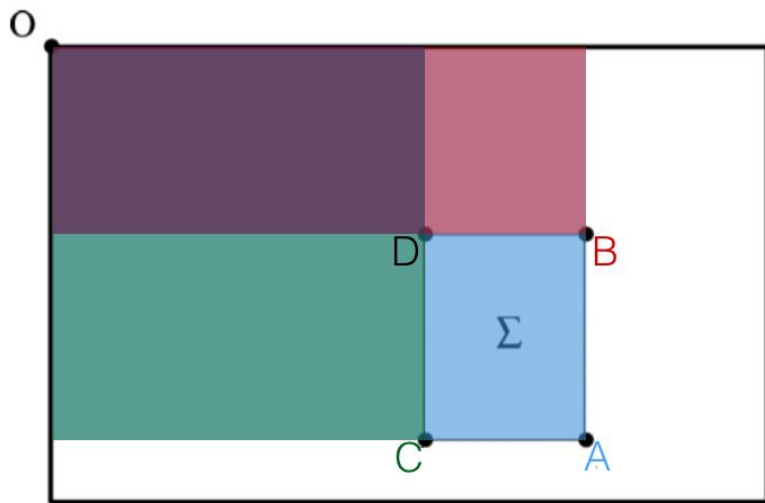
Integral images are an image transform such that any entry of an integral image $I_{\Sigma}(\mathbf{x})$ at a location $\mathbf{x} = (x, y)^T$ represents the sum of all pixels in the input image I within a rectangular region formed by the origin and \mathbf{x} .



$$I_{\Sigma}(\mathbf{x}) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j)$$

Detection: Integral Images

Integral images are incredibly efficient. It is possible to characterize a region of the image using four memory accesses and three operations. This makes it very cheap to detect blobs.



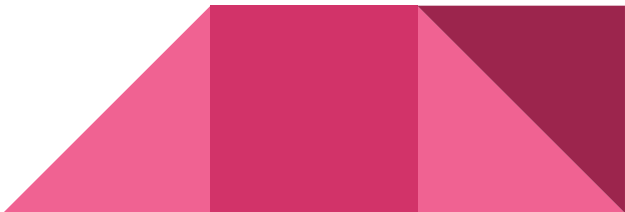
$$\Sigma = A - B - C + D$$

Detection: Hessian-Based Interest Points

The detector detects blob-like structures at locations where the determinant of the Hessian is maximum.

The Hessian is defined as such:
$$H(\mathbf{x}, \sigma) = \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix}$$

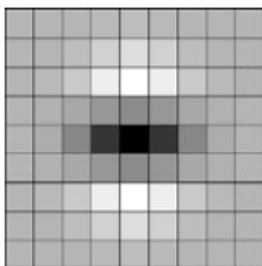
where $L_{xx}(\mathbf{x}, \sigma)$ is the convolution of the Gaussian second order derivative $\frac{\partial^2}{\partial x^2}g(\sigma)$ with the image I in point x .



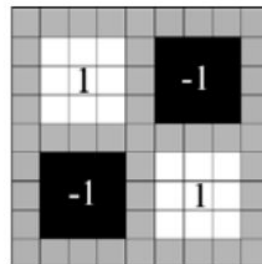
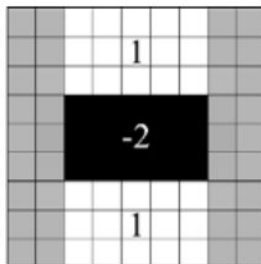
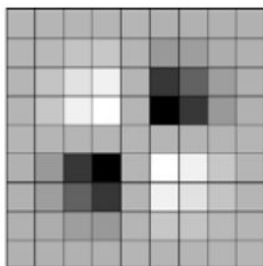
Detection: Hessian Approximation

The actual computation of the Hessian matrix is expensive and slow. Instead, the Hessian can be approximated using box filters!

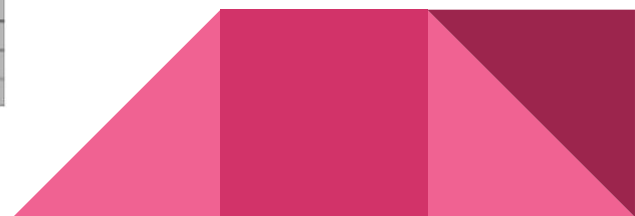
$det(H_{approx}) = D_{xx}D_{yy} - (wD_{xy})^2$ where D_{xx} is the approximation of the Gaussian second order partial derivative in the x-direction and $w = 0.9$.



The Gaussian second order partial derivative in y- and xy-direction.



Box filter approximations of the Gaussian second order partial derivatives.



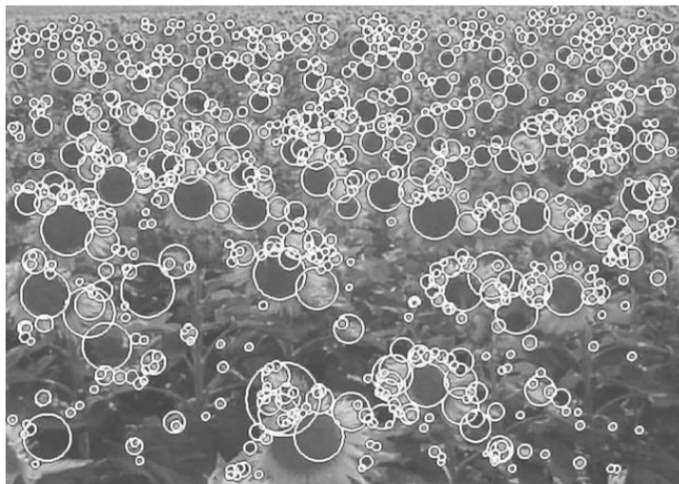
Detection: Scale Space Representation

To match interest points across different scales, a pyramidal scale space is built. Rather than serial downsampling, each successive level of the pyramid is built by upscaling the image in parallel. Each scale is defined as the the response of the image convolved with a box filter of a certain dimension (9x9, 15x15, 27x27 etc.). The scale space is further divided into octaves (sets of filter responses).



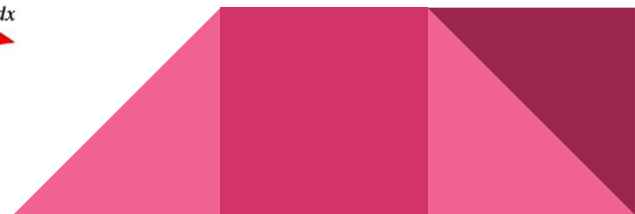
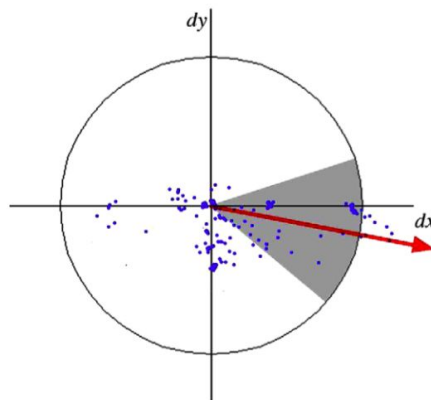
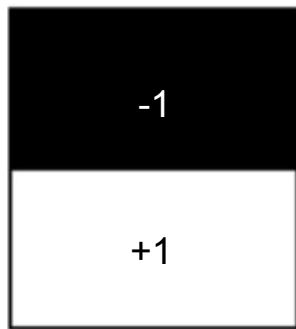
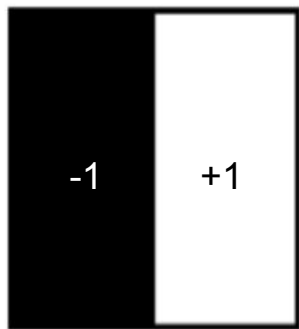
Detection: Interest Point Localization

To localize interest points in the image and over scales, a non-maximum suppression (non-maximum pixels are set to 0) in a $3 \times 3 \times 3$ neighborhood is applied. The maxima of the determinant of the Hessian matrix are then interpolated in scale and image space.



Descriptor: Orientation Assignment

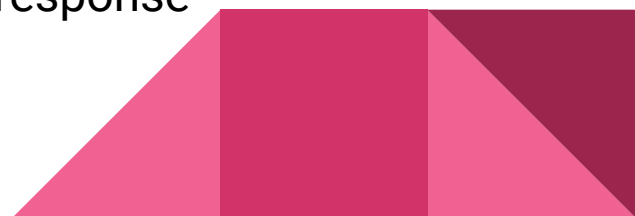
The Haar wavelet responses in x- and y-direction within a circular neighborhood with radius $6s$ is calculated. Responses are weighted with a Gaussian ($\sigma = 2s$) centered at the interest point and then the directional strengths are plotted. These plots are then divided into sliding orientation windows and local orientation vectors are computed as the sum of the x and y responses within each window. The dominant orientation is the largest of all such vectors across all windows.



Descriptor: Feature Vector

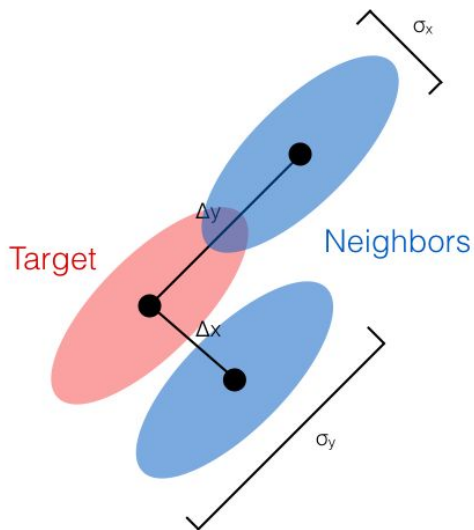
To extract features, an axis-orientated square window of size $20s$ and centered around the interest point is defined. This window is subdivided into a 4×4 grid. The “horizontal” and “vertical” Haar wavelet response is calculated over each subdivision and four metrics are extracted from each subdivision using 5×5 equally spaced points. These metrics are then summed to produce the local feature vector. These local feature vectors are concatenated to form a 64-element feature vector describing the interest point and surrounding neighborhood.

$$\mathbf{v} = \begin{bmatrix} \Sigma d_x \\ \Sigma d_y \\ \Sigma |d_x| \\ \Sigma |d_y| \end{bmatrix} \quad \begin{array}{l} \text{where } d_x \text{ is the “horizontal” Haar wavelet response} \\ d_y \text{ is the “vertical” Haar wavelet response} \end{array}$$



Matching: Nearest Neighbors

Features are matched across frames as the nearest neighbor within a distinct feature threshold. Either Euclidean or Mahalanobis distance may be used to determine “nearest”. In this implementation, uniform precision was assumed and, therefore, Euclidean distance was sufficient.

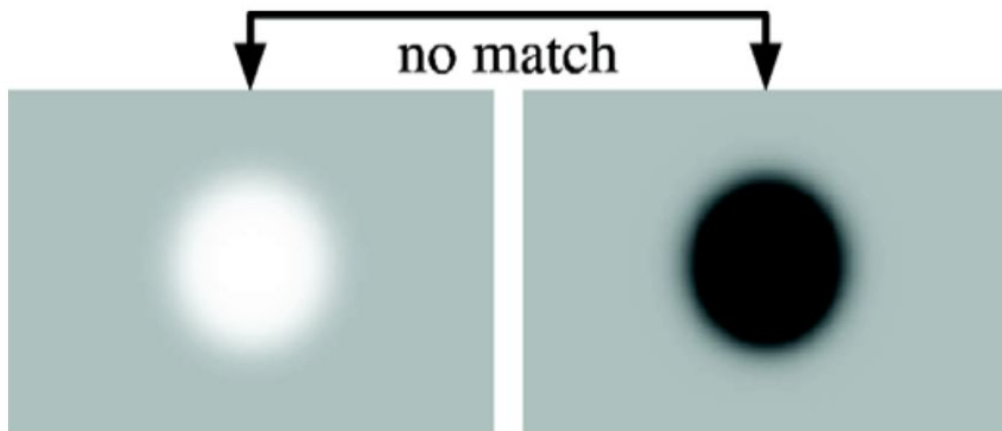


$$D_E = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

$$D_M = \sqrt{\frac{(x_2 - x_1)^2}{\sigma_x^2} + \frac{(y_2 - y_1)^2}{\sigma_y^2}}$$

Matching: Laplacian Indexing

For fast indexing during the matching phase, the sign of the Laplacian ($\text{Tr}(H)$) for the underlying interest point is included in the discrimination cascade. The sign of the Laplacian distinguishes bright blobs on dark backgrounds from the opposite situation and serves as a meaningful metric to divide the set of all interest points.

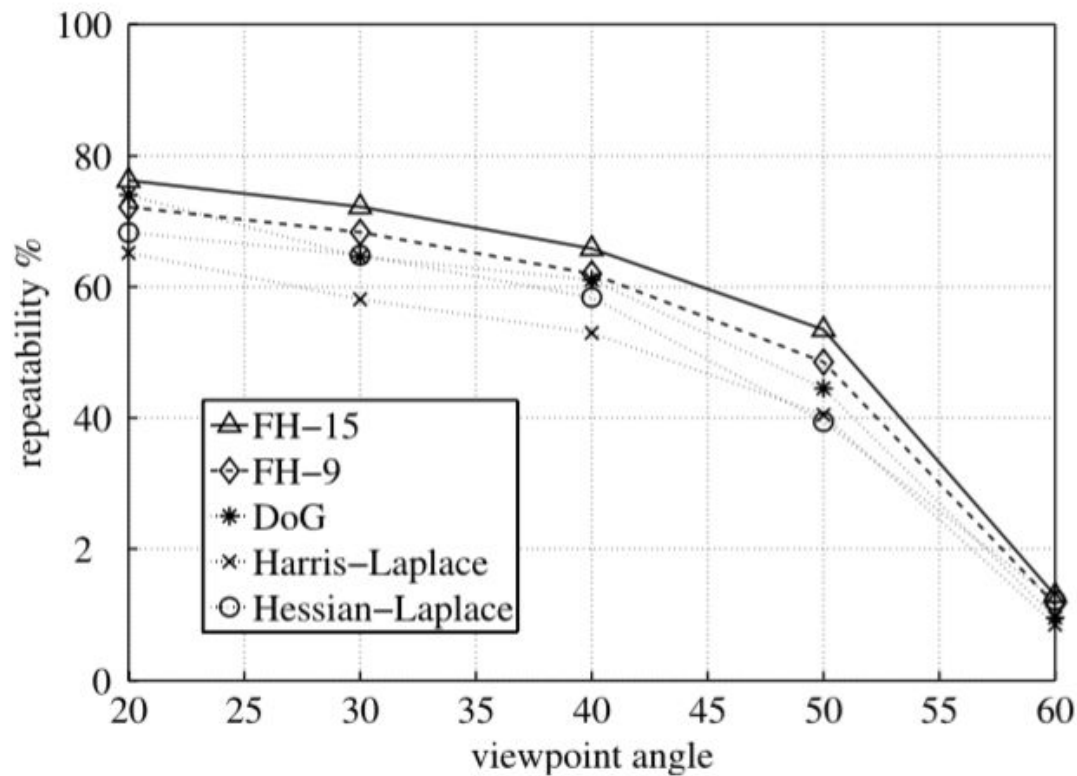


Results: Detector

In order to determine the repeatability of the SURF detector, interest points were generated on two different sequences of images (Graffiti and Wall) where each image is of the same object at a different angle. Repeatability is then defined as the percentage of interest points that remain in the new viewpoint versus the ground truth (image at 0°). Because each sequences contain out-of-plane-rotations, the resulting affine deformations have to be accounted for by the overall robustness of the features.



Results: Detector

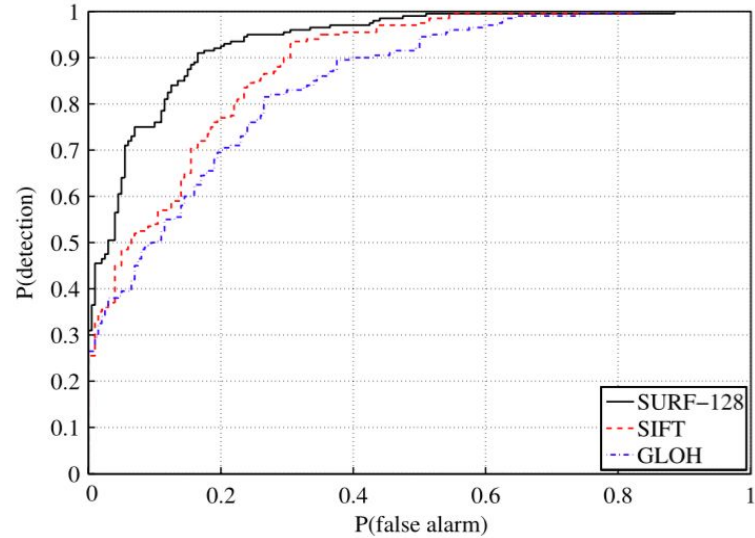
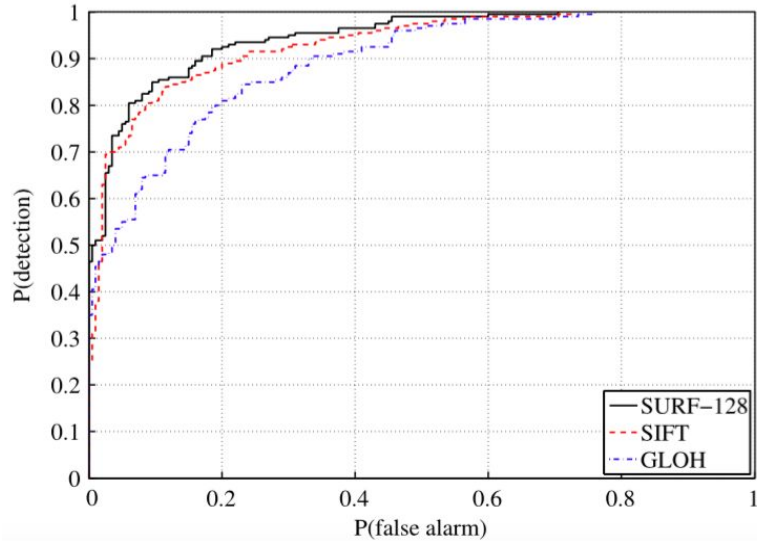


Results: Descriptor

In order to characterize the descriptor, the discriminative power of the feature vector was tested with a publicly available implementation of two bag-of-words classifiers. Using a set 400 images (Caltech background and airplanes), 200 were used for training and 200 were used for testing. The more characteristic the feature vector of a descriptor, the higher the rate of detections and the lower the rate of false positives will be. The SURF-128 descriptor was compared against two viable alternatives, SIFT and GLOH.



Results: Descriptor



Features generated from random edge pixels (left) and features generated from SURF interest points (right).

Opinion

Pros

- Self-contained, for the most part
- Included possible applications for the framework (3D scene reconstruction, object recognition)
- Framework delivers on promises: fast and robust feature generator
- Very detailed on implementation and how implementation decisions affected outcome

Cons

- Didn't give speed metrics for alternative algorithms
- More robust validation of detection and description metrics

