



Feature Based Vision Extras

METR 4202: Advanced Control & **Robotics**

Dr Surya Singh -- Lecture # 8

September 16, 2015

metr4202@itee.uq.edu.au

<http://robotics.itee.uq.edu.au/~metr4202/>

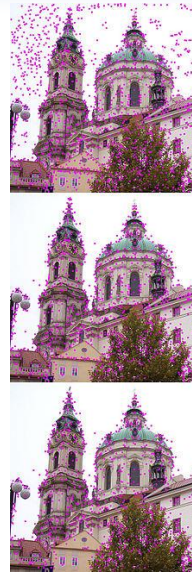
© 2015 School of Information Technology and Electrical Engineering at the University of Queensland



Feature Based Vision Extras

Scale Invariant Feature Transforms

- Goal was to define an algorithm to describe an image with features
- This would enable a number of different applications:
 - Feature Matching
 - Object / Image Matching
 - Orientation / Homography Resolution



Wikipedia: Scale Invariant Feature Transforms (2014)



SIFT: Feature Definition

- SIFT features are defined as the local extrema in a Difference of Gaussian (D) Scale Pyramid.

$$D(x, y, \sigma) = L(x, y, k_i \sigma) - L(x, y, k_j \sigma)$$

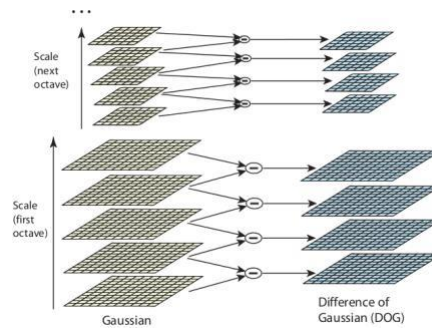
Where

$$L(x, y, k_i \sigma) = G(x, y, k \sigma) * I(x, y)$$



SIFT: Scale Pyramid

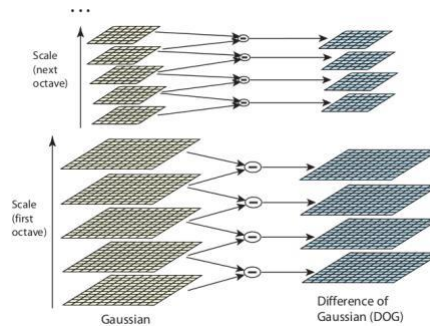
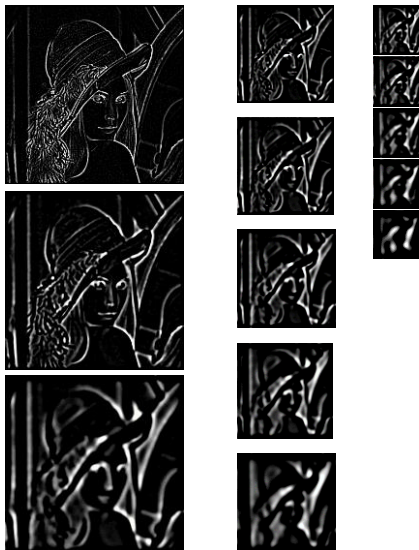
- D images are organised into a pyramid of progressively blurred images.
- Separated into octaves and scale levels per octave.
- Between octaves image is decimated by a factor of 2.



Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91-110.



SIFT: Scale Pyramid

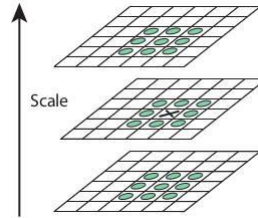


Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91-110.



SIFT: Feature Detection

- Each scale level in the image is evaluated for features.
- A feature is defined as a local maximum or minimum.
- For efficiency the 26 surrounding points are evaluated.



Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91-110.



SIFT: Feature Reduction

- Initial feature detection over detects features descriptive of the image.
- Initially remove features with low contrast.
- Then evaluate features to remove any edge responses.

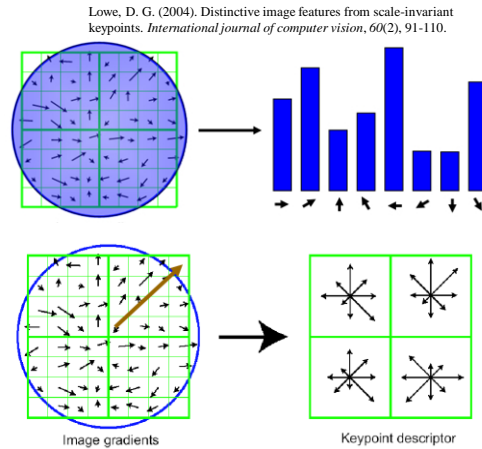


Wikipedia: Scale Invariant Feature Transforms (2014)



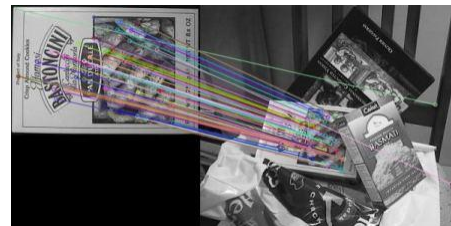
SIFT: Feature Description

- Features are described using the pixel gradients in a 16x16 square centring on the feature point.
- These gradients are then segmented into 4x4 boxes. An 8 bin orientation histogram is created to define the box.



SIFT: Feature Matching

- A match is defined as a pair of features with the closest Euclidian distance to each other.
- Matches above a threshold are culled to improve match.

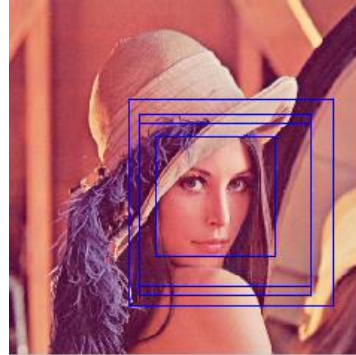


OpenCV: Feature Matching (2014)



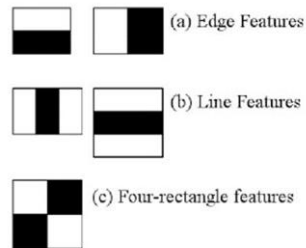
Boosted Cascade Haar-like Weak Classifiers

- Fast object detector designed primarily for use in face detection.
- Uses a cascade of weak classifiers to define object match.



Viola Jones: Feature Definition

- Feature is classified as being the difference between the average intensity of two or more image sections.
- Can be any arithmetic combination of section values.



Viola Jones: Efficient Calculation of Features

- Fast calculation of the feature value is obtained by calculating the integral image.
- This leaves at most 4 sum operations to calculate a feature.

1	2	2	4	1	0	0	0	0	0	0
3	4	1	5	2	0	1	3	5	9	10
2	3	3	2	4	0	4	10	13	22	25
4	1	5	4	6	0	6	15	21	32	39
6	3	2	1	3	0	10	20	31	46	59
					0	16	29	42	58	74

input image integral image



Viola Jones: Boosting

- Iteratively selects best classifier for detection.
- Assigns weights to each classifier to indicate likelihood of classifier indicating positive detection
- If the sum of the weights of positive classifier responses is above a threshold then there is a positive detection.

- Given example images $(x_1, y_1), \dots, (x_n, y_n)$ where $y_i = 0, 1$ for negative and positive examples respectively.
- Initialize weights $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ for $y_i = 0, 1$ respectively, where m and l are the number of negatives and positives respectively.
- For $t = 1, \dots, T$:

1. Normalize the weights,

$$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$$

so that $w_{t,j}$ is a probability distribution.

2. For each feature, j , train a classifier h_j which is restricted to using a single feature. The error is evaluated with respect to w_t , $\epsilon_j = \sum_i w_i |h_j(x_i) - y_i|$.

3. Choose the classifier, h_t , with the lowest error ϵ_t .

4. Update the weights:

$$w_{t+1,i} = w_{t,i} \beta_t^{1-\epsilon_i}$$

where $\epsilon_i = 0$ if example x_i is classified correctly, $\epsilon_i = 1$ otherwise, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.

- The final strong classifier is:

$$h(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise} \end{cases}$$

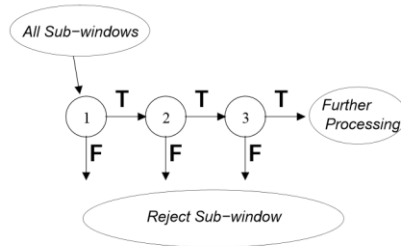
where $\alpha_t = \log \frac{1}{\beta_t}$

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features



Viola Jones: Boosted Cascades

- Effective boosted classifiers require a high number of weak classifiers.
- However, simple low count classifiers offer high rejection rate.
- Solution is to use cascaded classifiers.



Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features

