Digital Image Processing ECE 566

Ahmad Ghasemi

Department of Electrical and Computer Engineering

University of Massachusetts Amherst

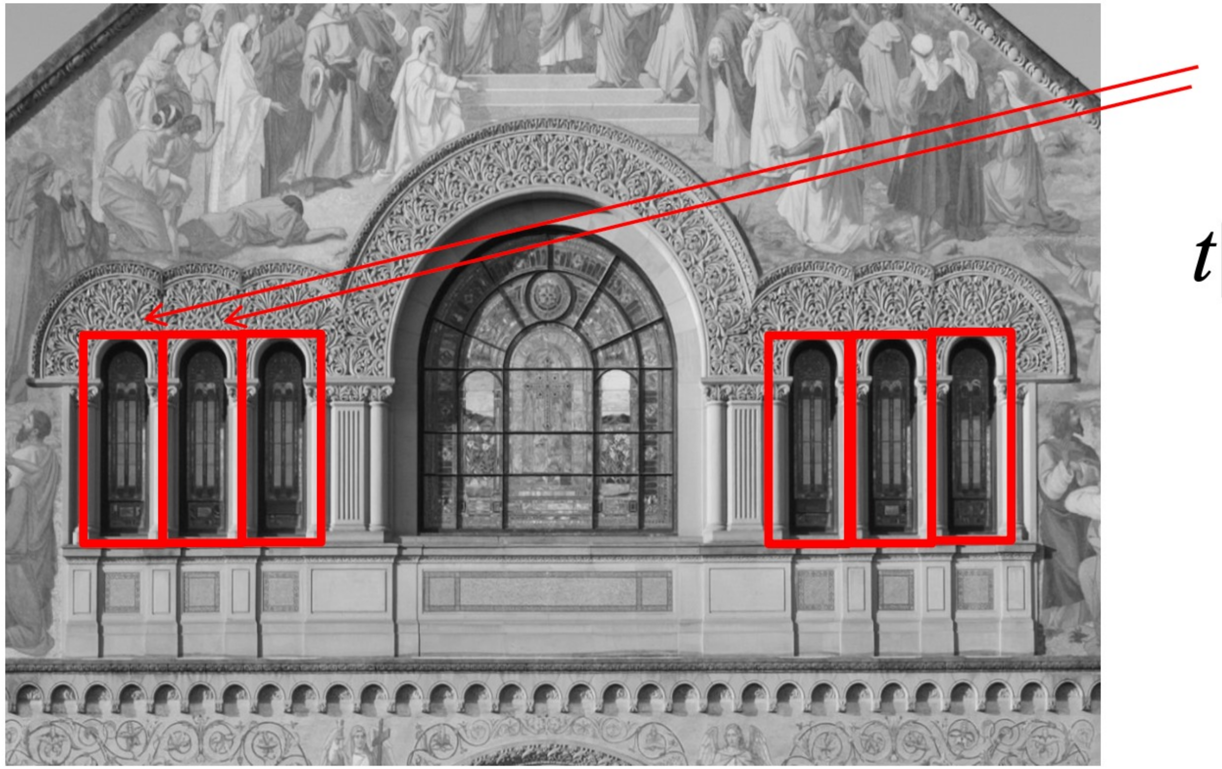
UMass Amherst

Template Matching

Problem: locate an object, described by a template $t[x,y]$, in the image $s[x,y]$

Face recognition and medical image processing

Example



$t[x,y]$

$s[x,y]$

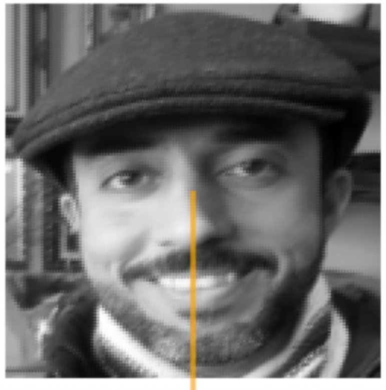
Credit: DIP, Stanford University

SIFT: Scale-space peak Selection

- **Constructing a Scale Space:** To make sure that features are scale-independent
- **Keypoint Localisation:** Identifying the suitable features or keypoints
- **Orientation Assignment:** Ensure the keypoints are rotation invariant
- **Keypoint Descriptor:** Assign a unique fingerprint to each keypoint

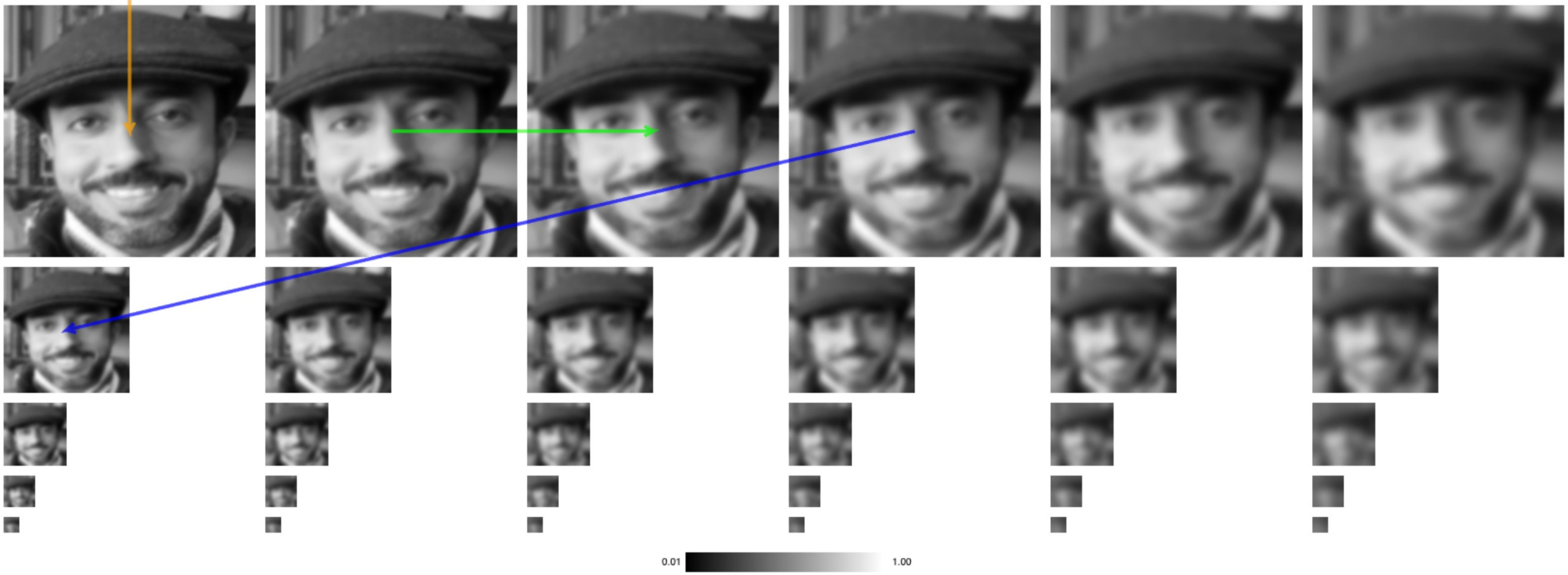
Finally, we can use these keypoints for feature matching!

SIFT: Scale-space



Gaussian Blur

> We use the **Gaussian Blurring technique** to reduce the noise in an image.



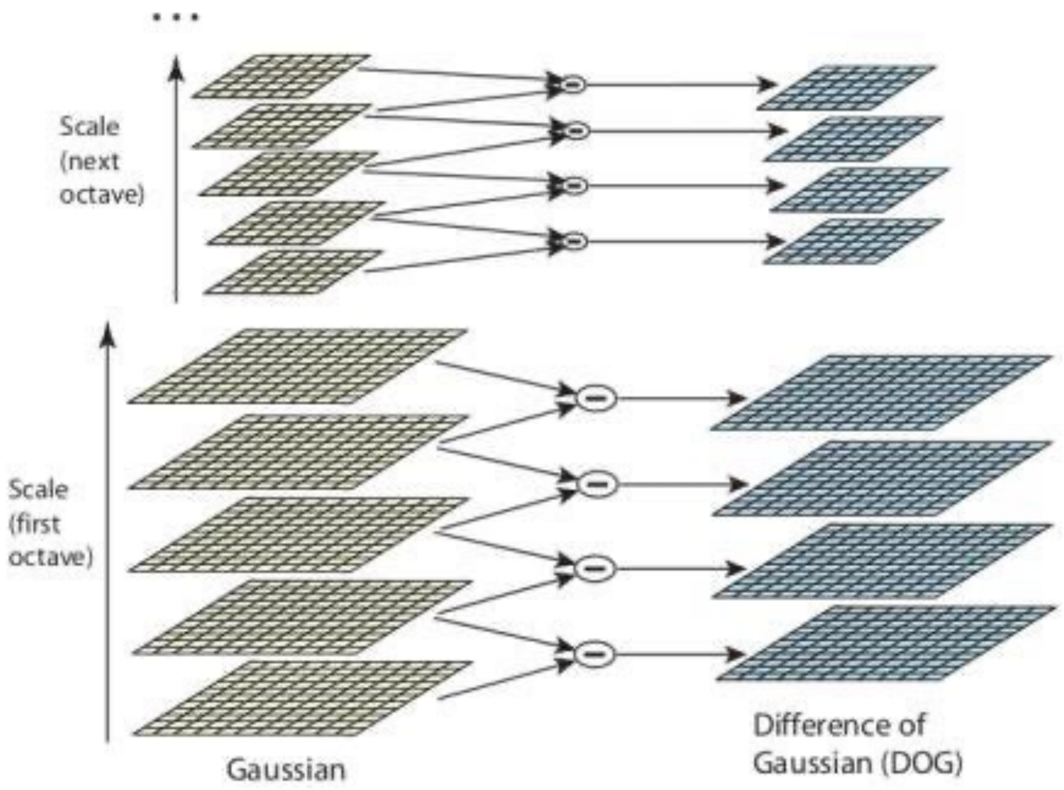
University of Massachusetts Amherst

SIFT: DOG(Difference of Gaussian kernel)

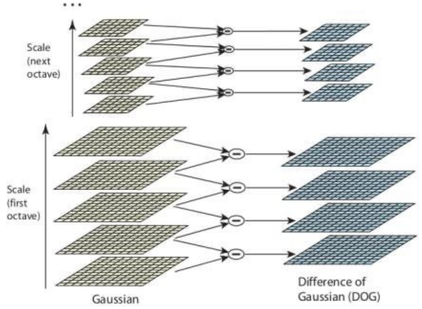
Use the blurred images to generate another set of images, the Difference of Gaussians (DoG).

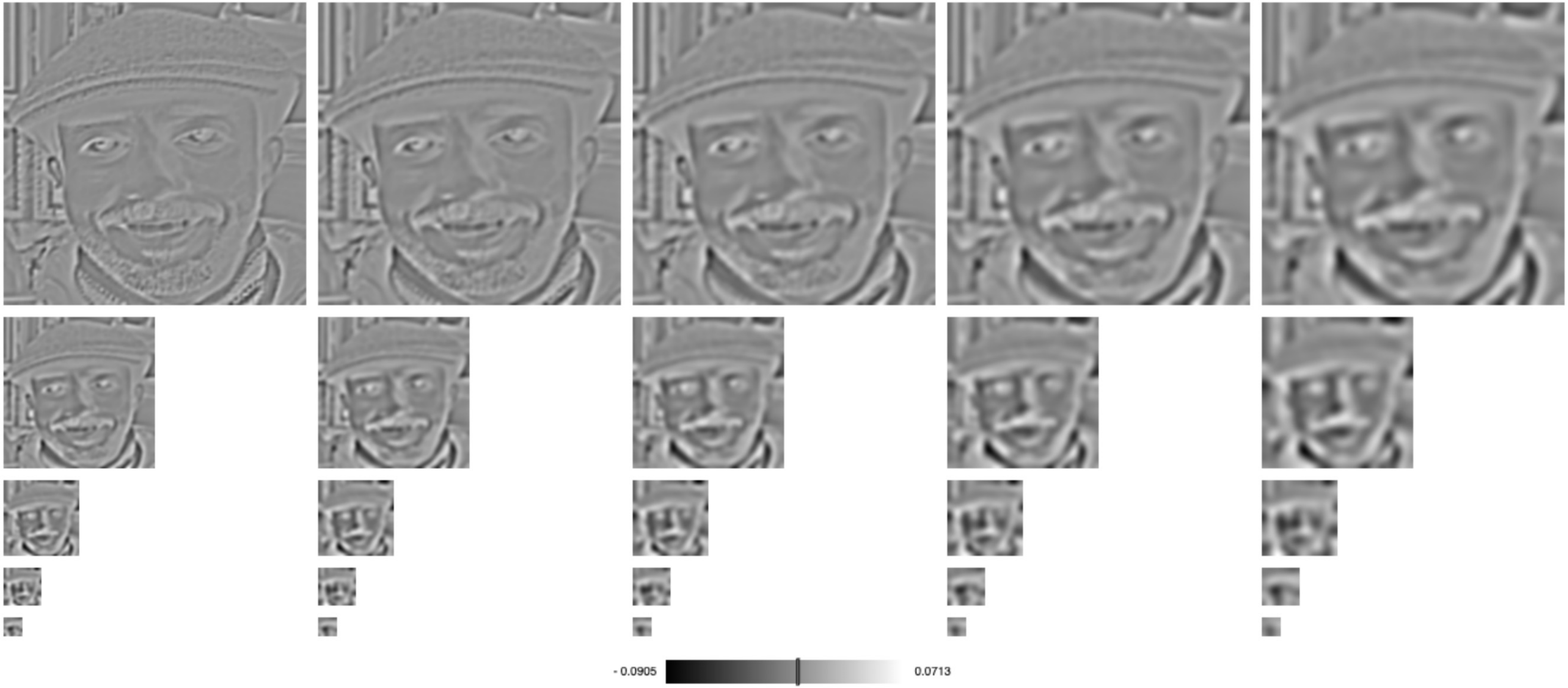
These DoG images are great for finding out interesting keypoints in the image.

The difference of Gaussian is obtained as the difference of Gaussian blurring of an image with two different σ, let it be σ and kσ. This process is done for different octaves of the image in the Gaussian Pyramid.

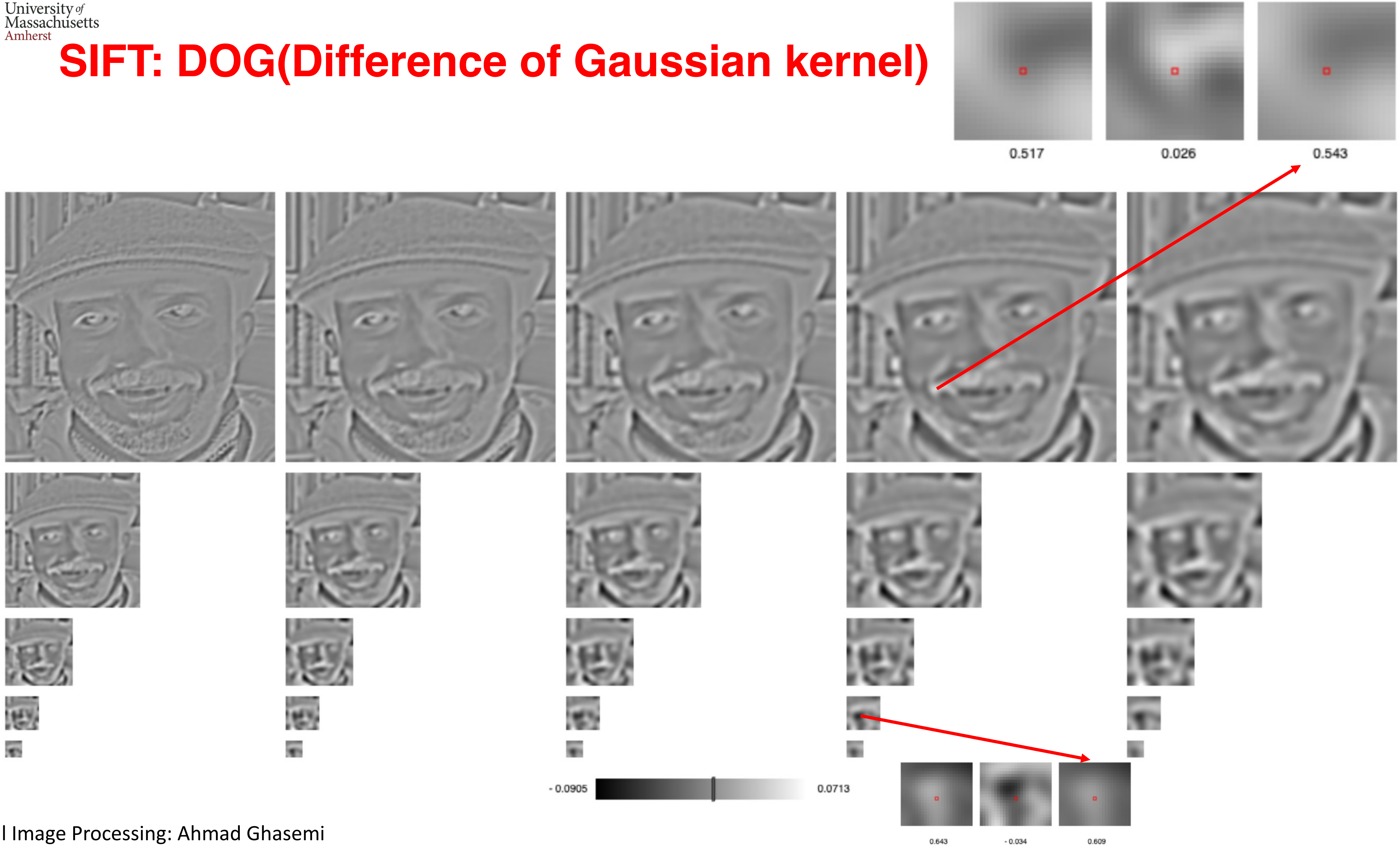


SIFT: DOG(Difference of Gaussian kernel)



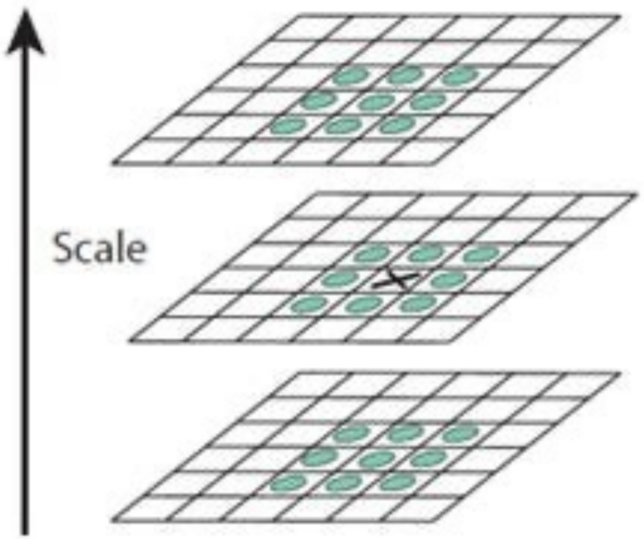


SIFT: DOG(Difference of Gaussian kernel)



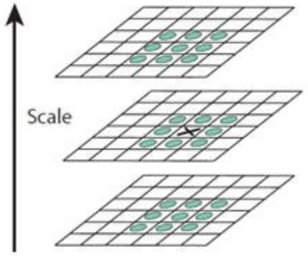
SIFT: Finding Keypoints

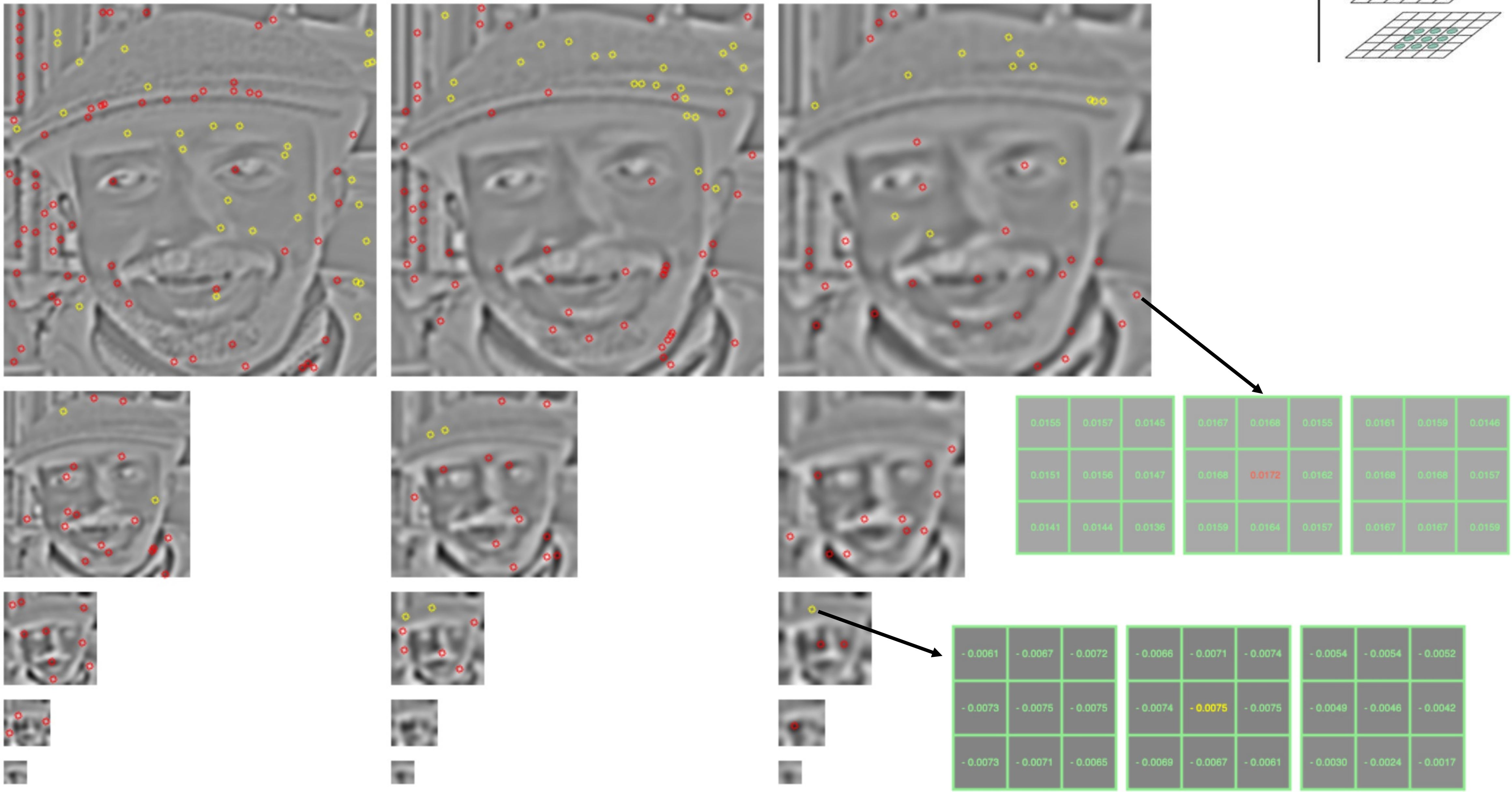
One pixel in an image is compared with its 8 neighbors as well as 9 pixels in the next scale and 9 pixels in previous scales. This way, a total of 26 checks are made.



If it is a local extrema, it is a potential keypoint. It basically means that keypoint is best represented in that scale.

SIFT: Finding Keypoints





SIFT: Interpolation

The interpolation is done using the quadratic Taylor expansion of the Difference-of-Gaussian scale-space function, $D(x, y, \sigma)$ with the candidate keypoint as the origin. This Taylor expansion is given by:

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

where D and its derivatives are evaluated at the candidate keypoint and $\mathbf{x} = (x, y, \sigma)^T$ is the offset from this point.

If the offset $\hat{\mathbf{x}}$ is larger than 0.5 in any dimension, then that's an indication that the extremum lies closer to another candidate keypoint. In this case, the candidate keypoint is changed and the interpolation performed instead about that point. Otherwise the offset is added to its candidate keypoint to get the interpolated estimate for the location of the extremum.

SIFT: Discarding low-contrast keypoints

Discarding low-contrast keypoints

To discard the keypoints with low contrast, the value of the second-order Taylor expansion $D(\mathbf{x})$ is computed at the offset $\hat{\mathbf{x}}$. If this value is less than 0.03, the candidate keypoint is discarded. Otherwise it is kept, with final scale-space location $\mathbf{y} + \hat{\mathbf{x}}$, where $\mathbf{y}$ is the original location of the keypoint.

Eliminating edge responses

The DoG function will have strong responses along edges, even if the candidate keypoint is not robust to small amounts of noise. Therefore, in order to increase stability, we need to eliminate the keypoints that have poorly determined locations but have high edge responses.

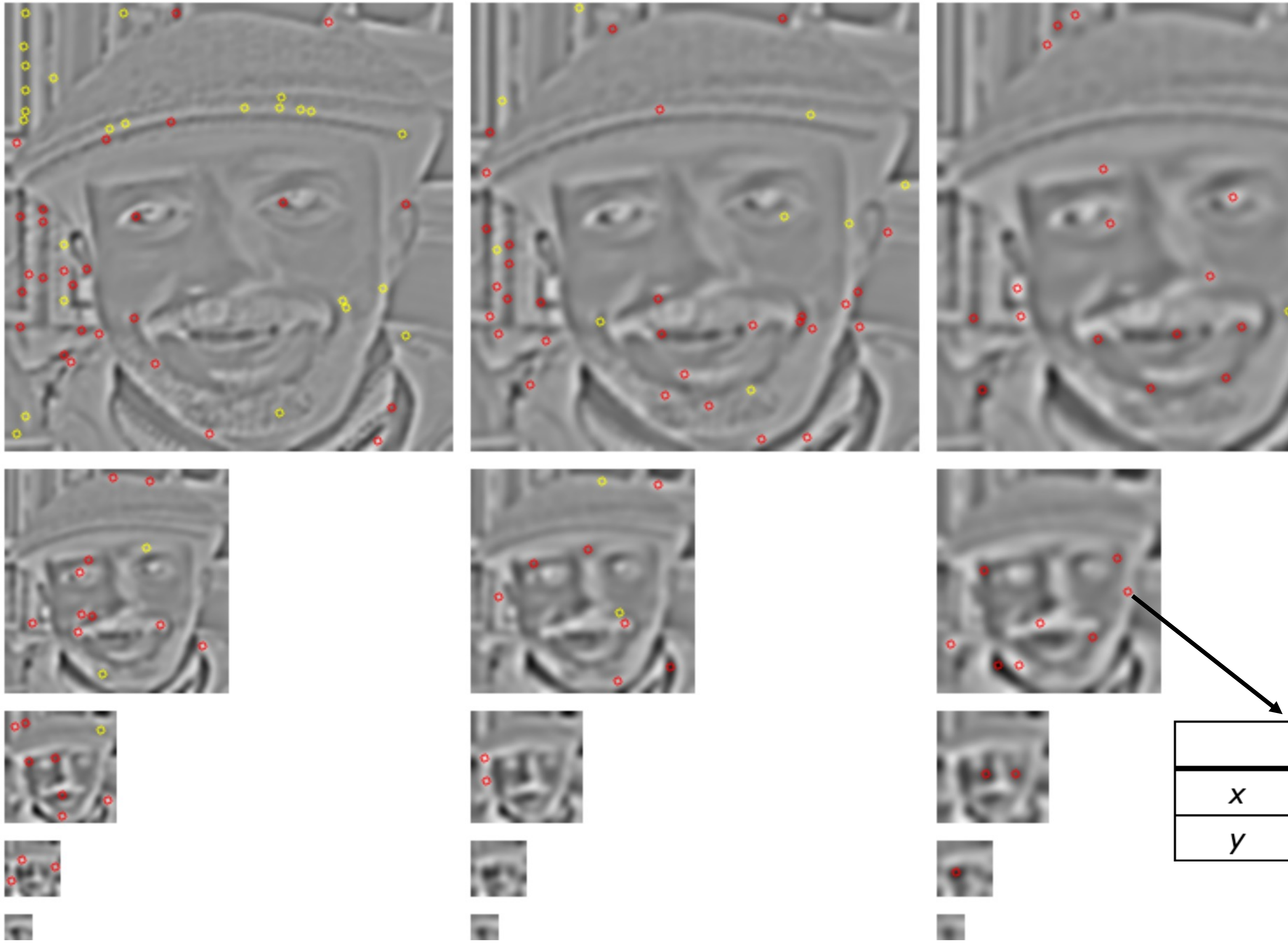
For poorly defined peaks in the DoG function, the principal curvature across the edge would be much larger than the principal curvature along it. Finding these principal curvatures amounts to solving for the eigenvalues of the second-order Hessian matrix, **H**:

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$

SIFT: Discarding low-contrast keypoints

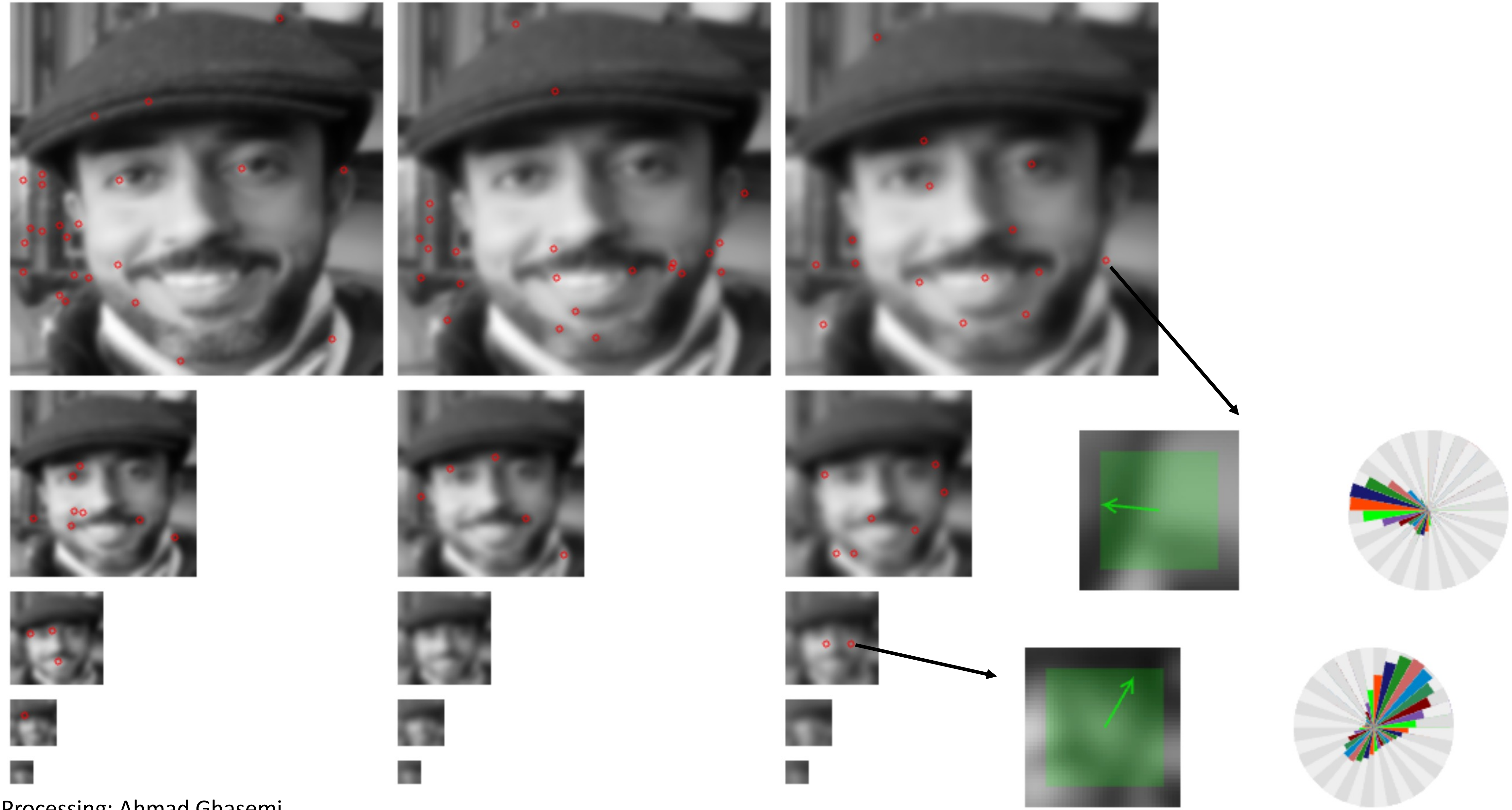
The eigenvalues of **H** are proportional to the principal curvatures of D. It turns out that the ratio of the two eigenvalues, say α is the larger one, and β the smaller one, with ratio $r = \alpha/\beta$, is sufficient for SIFT's purposes. The trace of **H**, i.e., $D_{xx} + D_{yy}$, gives us the sum of the two eigenvalues, while its determinant, i.e., $D_{xx}D_{yy} - D_{xy}^2$, yields the product. The ratio $\mathrm{R} = \mathrm{Tr}(\mathbf{H})^2 / \mathrm{Det}(\mathbf{H})$ can be shown to be equal to $(r+1)^2/r$, which depends only on the ratio of the eigenvalues rather than their individual values. R is minimum when the eigenvalues are equal to each other. Therefore, the higher the absolute difference between the two eigenvalues, which is equivalent to a higher absolute difference between the two principal curvatures of D, the higher the value of R. It follows that, for some threshold eigenvalue ratio r_{th}, if R for a candidate keypoint is larger than $(r_{\mathrm{th}}+1)^2/r_{\mathrm{th}}$, that keypoint is poorly localized and hence rejected. The new approach uses $r_{\mathrm{th}} = 10$.

SIFT: Discarding low-contrast keypoints



	discrete	interpolated
x	109	109.149
y	71	70.406

SIFT: Orientation Assignment



SIFT: Keypoint Descriptor

SIFT: Why does it work?

- Key points are extracted at different scales and blur levels and all subsequent computations are performed within the scale space framework. This will make the descriptors invariant to image scaling and small changes in perspective.

SIFT: Why does it work?

- Key points are extracted at different scales and blur levels and all subsequent computations are performed within the scale space framework. This will make the descriptors invariant to image scaling and small changes in perspective.

- Computation relative to a reference orientation is supposed to make the descriptors robust against rotation.

SIFT: Why does it work?

- Key points are extracted at different scales and blur levels and all subsequent computations are performed within the scale space framework. This will make the descriptors invariant to image scaling and small changes in perspective.

- Computation relative to a reference orientation is supposed to make the descriptors robust against rotation.

- Likewise, the descriptor information is stored relative to the key point position and thus invariant against translations.

SIFT: Why does it work?

- Key points are extracted at different scales and blur levels and all subsequent computations are performed within the scale space framework. This will make the descriptors invariant to image scaling and small changes in perspective.

- Computation relative to a reference orientation is supposed to make the descriptors robust against rotation.

- Likewise, the descriptor information is stored relative to the key point position and thus invariant against translations.

- Many potential key points are discarded if they are deemed unstable or hard to locate precisely. The remaining key points should thus be relatively immune to image noise.

SIFT: Why does it work?

- Key points are extracted at different scales and blur levels and all subsequent computations are performed within the scale space framework. This will make the descriptors invariant to image scaling and small changes in perspective.
- Computation relative to a reference orientation is supposed to make the descriptors robust against rotation.
- Likewise, the descriptor information is stored relative to the key point position and thus invariant against translations.
- Many potential key points are discarded if they are deemed unstable or hard to locate precisely. The remaining key points should thus be relatively immune to image noise.
- The histograms are normalized at the end which means the descriptors will not store the magnitudes of the gradients, only their relations to each other. This should make the descriptors invariant against global, uniform illumination changes.

SIFT: Why does it work?

- Key points are extracted at different scales and blur levels and all subsequent computations are performed within the scale space framework. This will make the descriptors invariant to image scaling and small changes in perspective.

- Computation relative to a reference orientation is supposed to make the descriptors robust against rotation.

- Likewise, the descriptor information is stored relative to the key point position and thus invariant against translations.

- Many potential key points are discarded if they are deemed unstable or hard to locate precisely. The remaining key points should thus be relatively immune to image noise.

- The histograms are normalized at the end which means the descriptors will not store the magnitudes of the gradients, only their relations to each other. This should make the descriptors invariant against global, uniform illumination changes.

- The histogram values are also thresholded to reduce the influence of large gradients. This will make the information partly immune to local, non-uniform changes in illumination.