Computer Vision Applications

COMP 388-002/488-002 Computer Science Topics

Daniel Moreira

Fall 2022

LOYOLA UNIVERSITY CHICAGO

Image Description

COMP 388-002/488-002 Computer Science Topics
Computer Vision Applications

Daniel Moreira
Fall 2022

LOYOLA UNIVERSITY CHICAGO

Today you will…

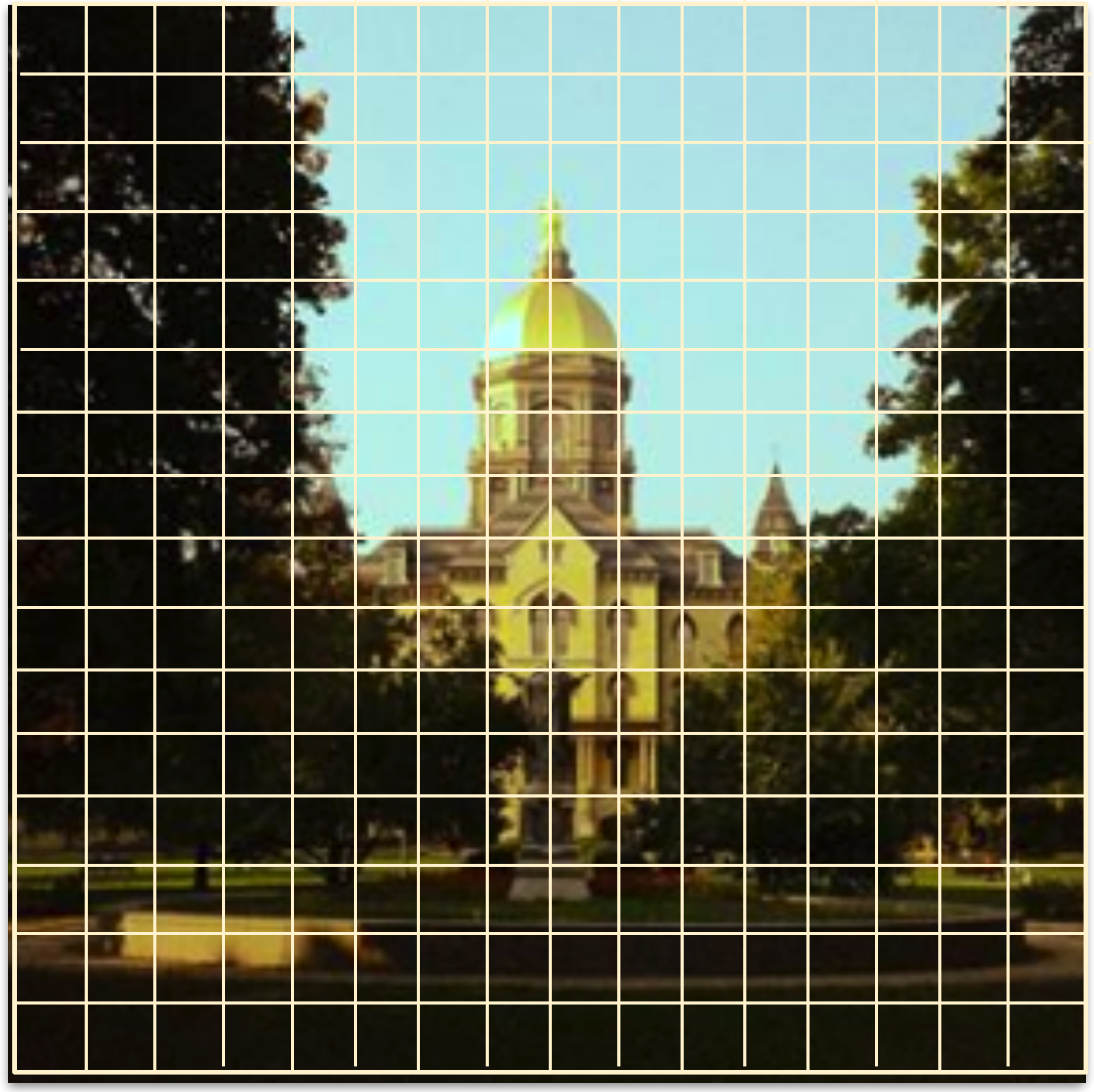
Get to know global and local image description.

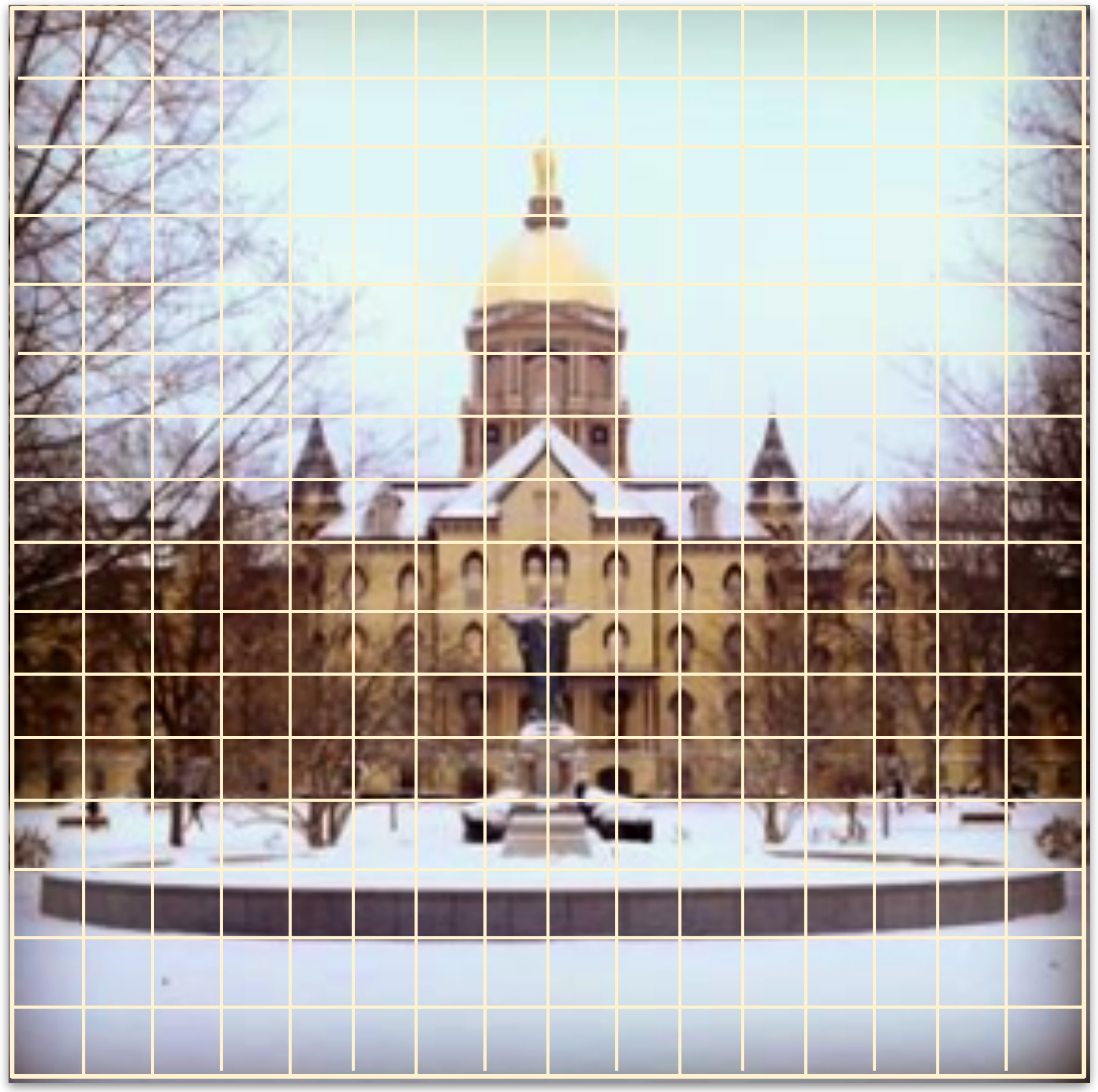
Why do We Need Image Description?

Can a computer system decide if these images depict the same building?

LOYOLA UNIVERSITY CHICAGO

Why do We Need Image Description?





Can a computer system decide if these images depict the same building?

Yes, but not directly based on the pixel values.

capture position

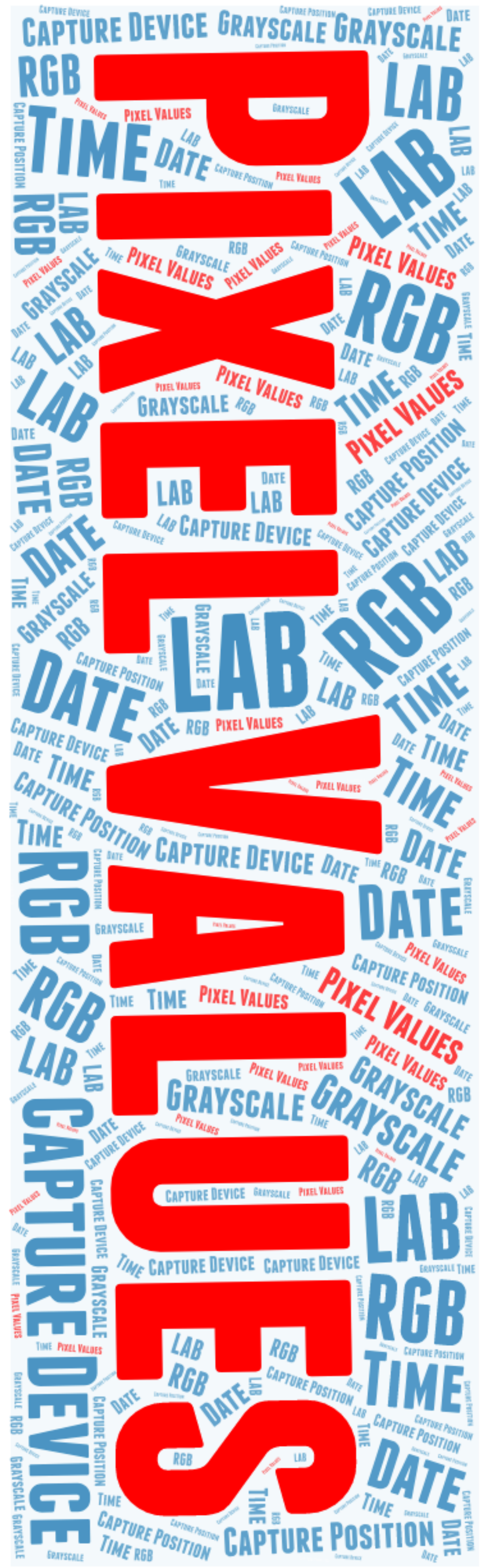
date and time

device

Pixel values depend on complex settings.

Semantic Gap

RECAP



Level 0

?

Task

LOYOLA UNIVERSITY CHICAGO

Semantic Gap

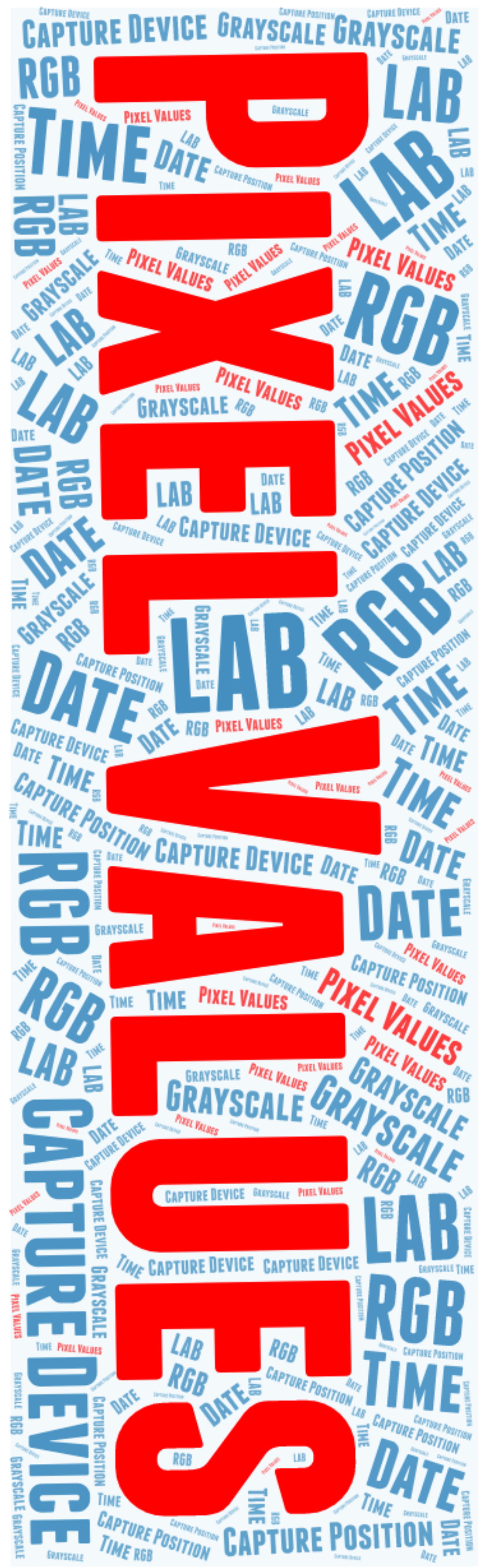
RECAP

Level 0 Level 1 Level 2 Level 3 Task

LOYOLA
UNIVERSITY CHICAGO

Semantic Gap

RECAP



Level 0

Level 1

Level 2

Level 3

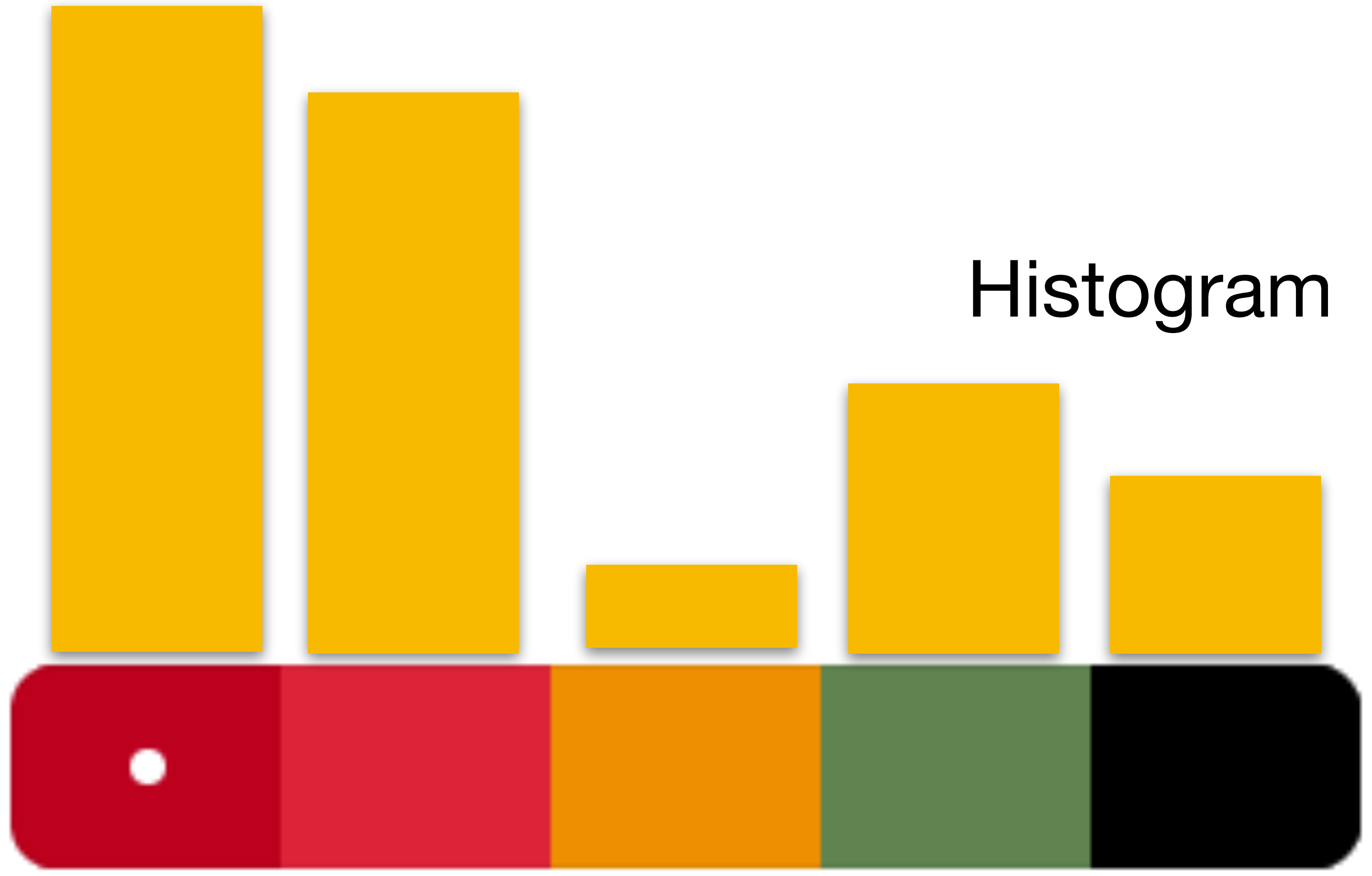
Task

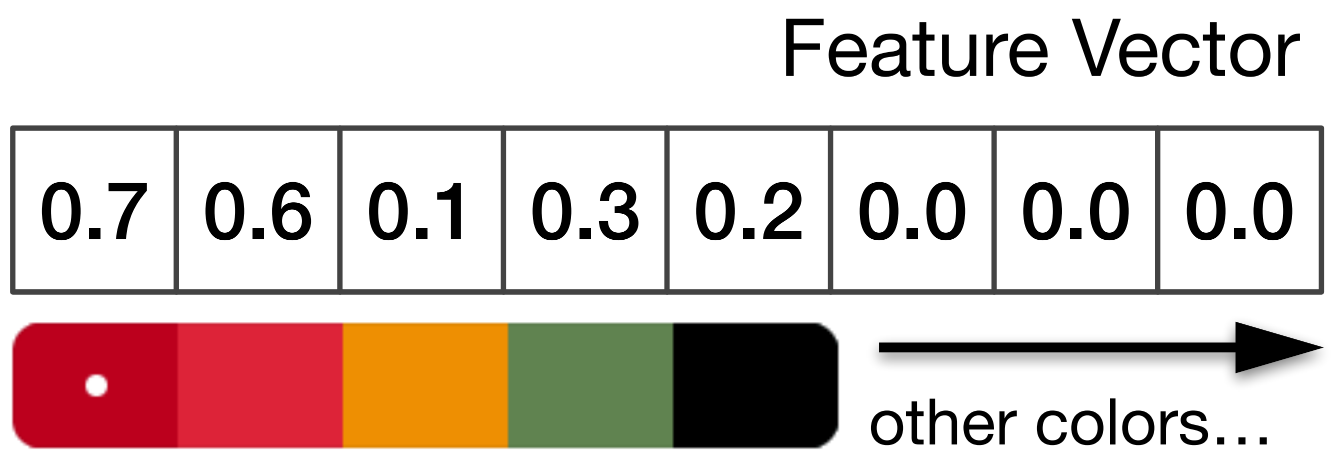
LOYOLA UNIVERSITY CHICAGO

Global Features

The entire image is represented by a single tensor.

Example: **Color Histogram**





Global Features

The entire image is represented by a single tensor.

Example: **Color Histogram**

query

database similar images

https://www.pantechsolutions.net/matlab-code-for-image-retrieval

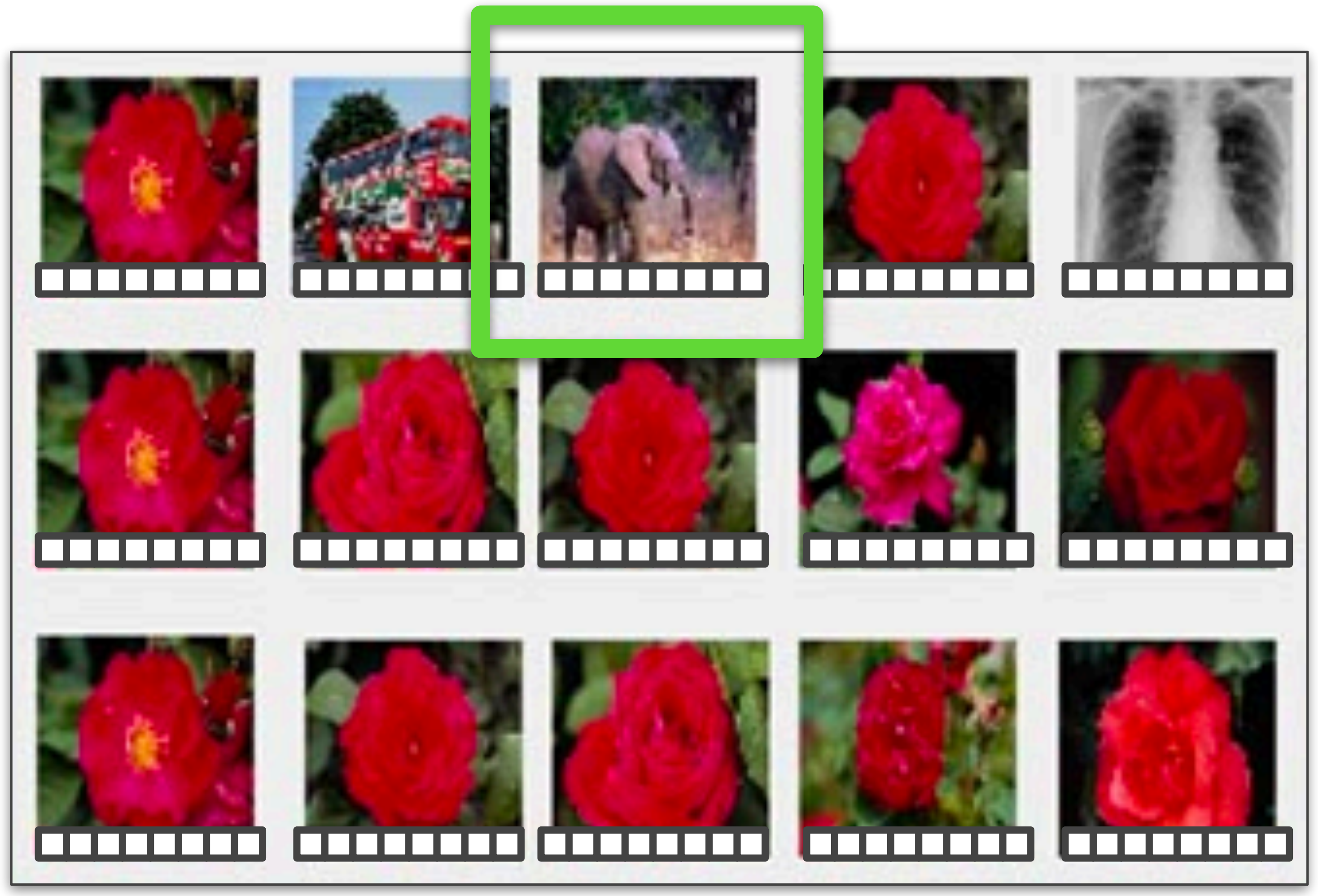
LOYOLA UNIVERSITY CHICAGO

Global Features

The entire image is represented by a single tensor.

Example: **Color Histogram**

query



Cons

No distinction between foreground and background.

Global Features

The entire image is represented by a single tensor.

Example: **Color Histogram**

query

Cons

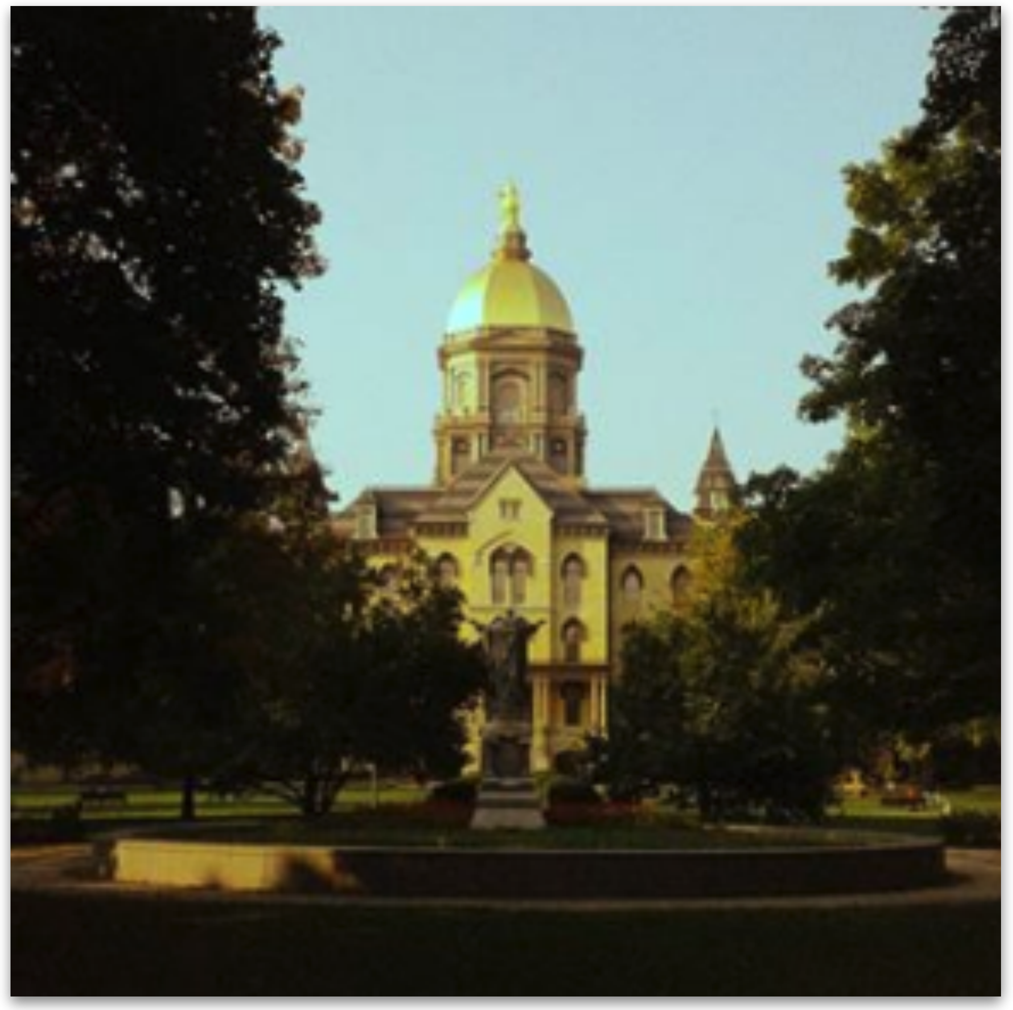
No semantics.

LOYOLA UNIVERSITY CHICAGO

Global Features

The entire image is represented by a single tensor.

Example: **Color Histogram**



A lot of dark green.

A lot of white.

Cons

No semantics.

Not robust to occlusions.

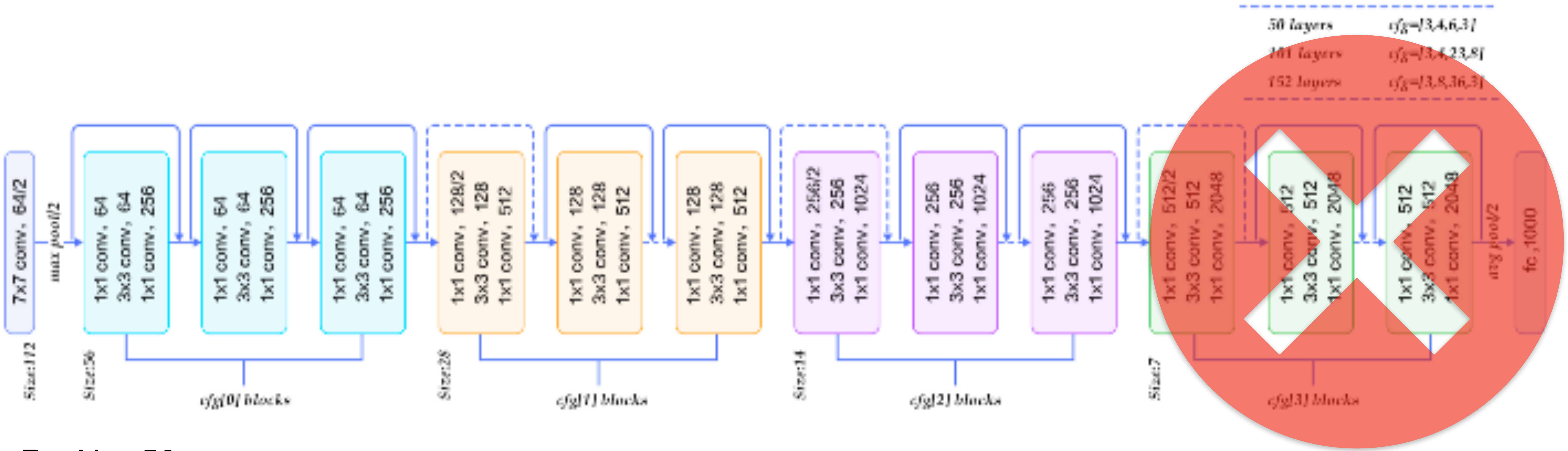
No match.

LOYOLA UNIVERSITY CHICAGO

Global Features

The entire image is represented by a single tensor.

Example: **CNN-based**



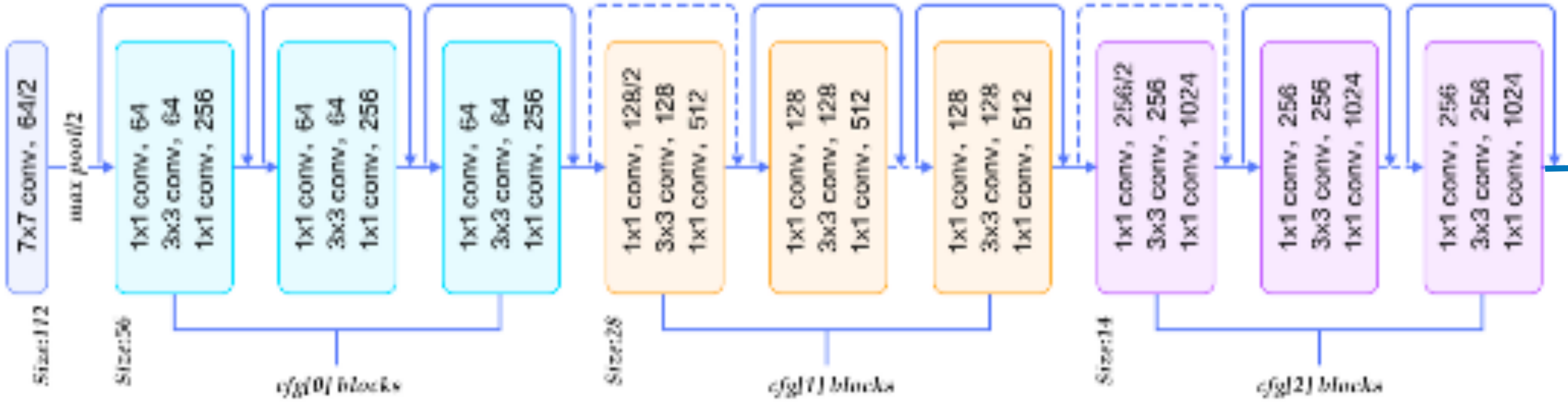
ResNet-50

LOYOLA UNIVERSITY CHICAGO

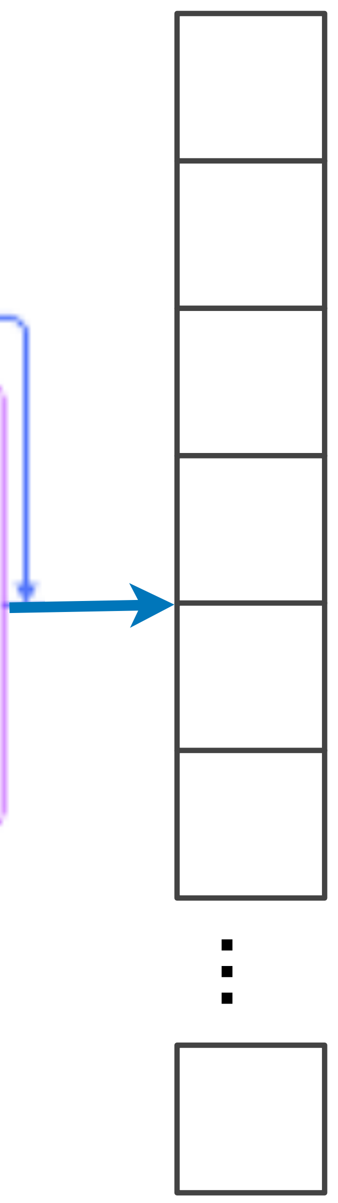
Global Features

The entire image is represented by a single tensor.

Example: **CNN-based**



ResNet-50



1024D

Pros

Inherited semantic awareness.

Global Features

The entire image is represented by a single tensor.

Example: **CNN-based**

query

database semantically similar images

Pros

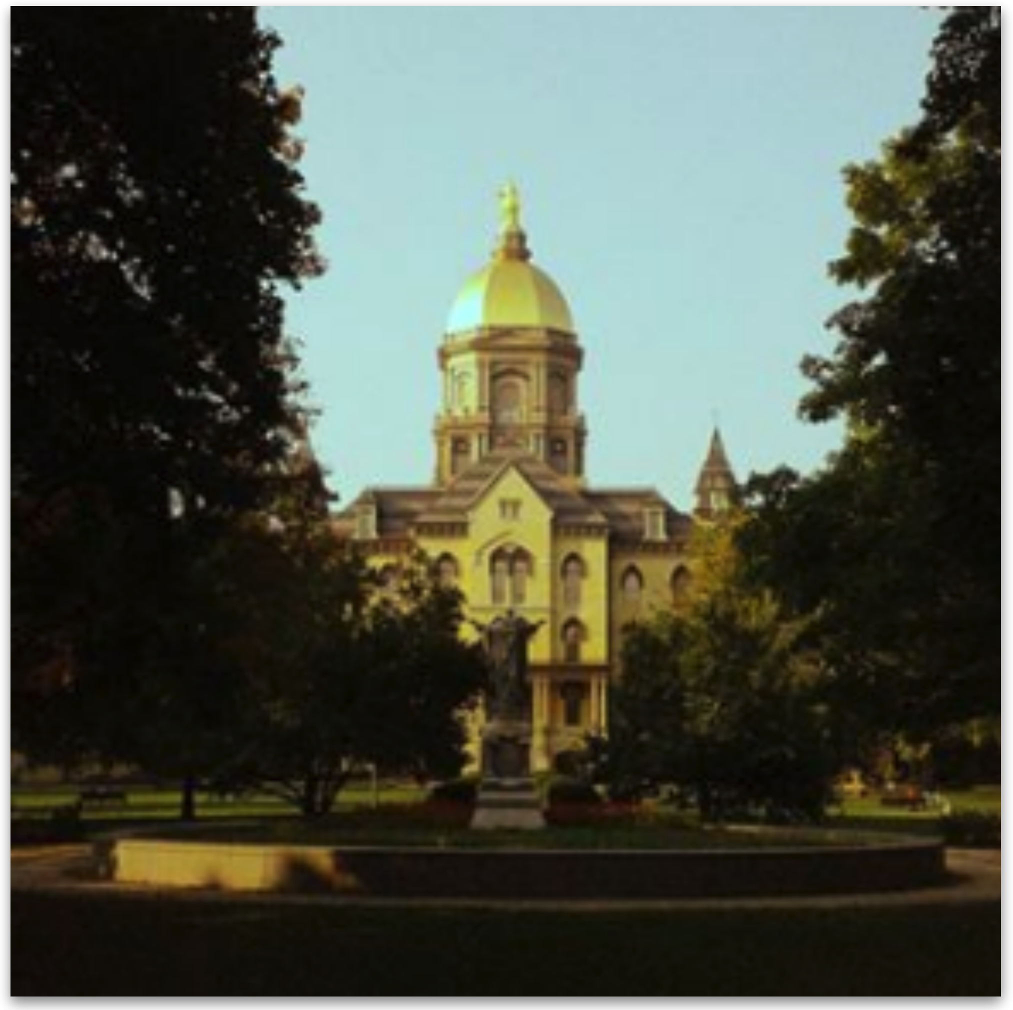
Inherited semantic awareness.

LOYOLA UNIVERSITY CHICAGO

Global Features

The entire image is represented by a single tensor.

Example: **CNN-based**



A lot of occlusion.

Cons

Can it help to decide if these images depict the same building?

Global Features

The entire image is represented by a single tensor.

Example: **CNN-based**

A lot of occlusion.

Unnecessary regions.

Cons

Can it help to decide if these images depict the same building?

LOYOLA UNIVERSITY CHICAGO

Local Features

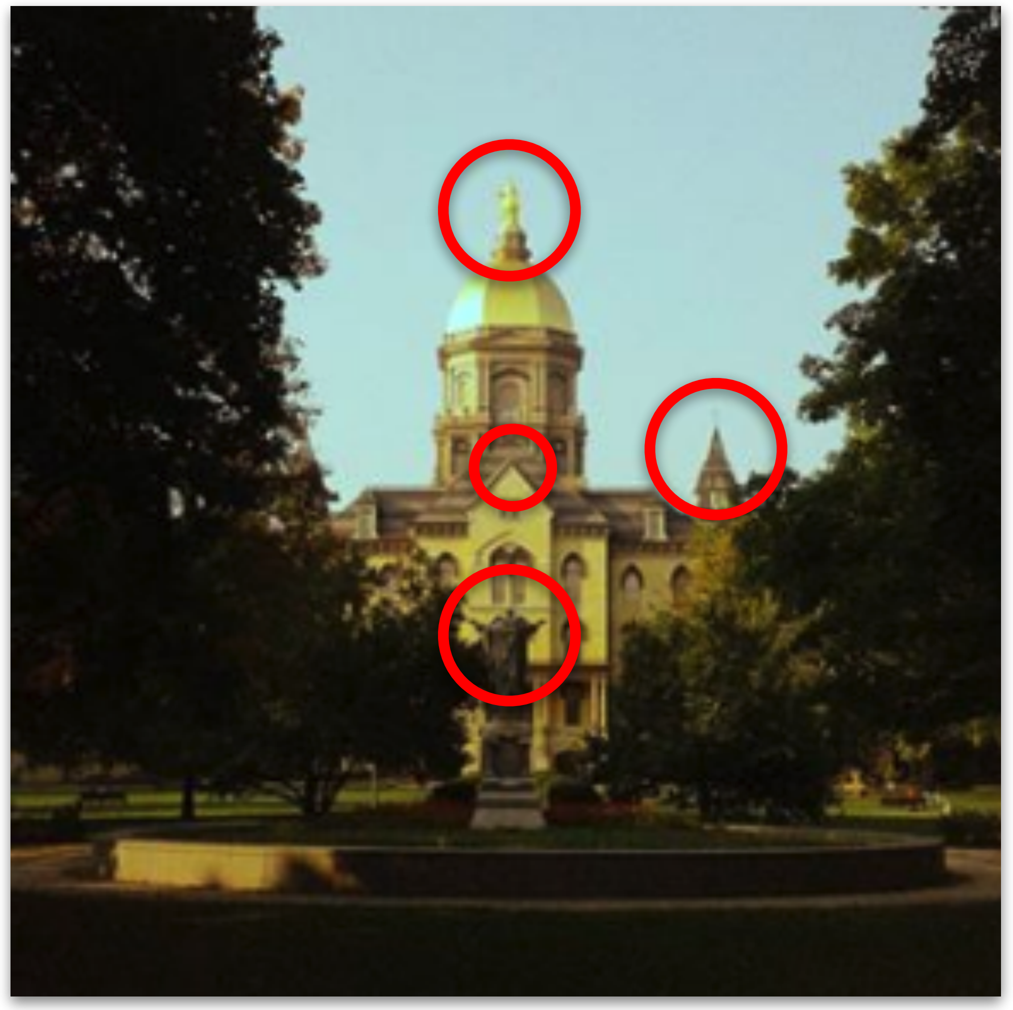
What are Local Features?

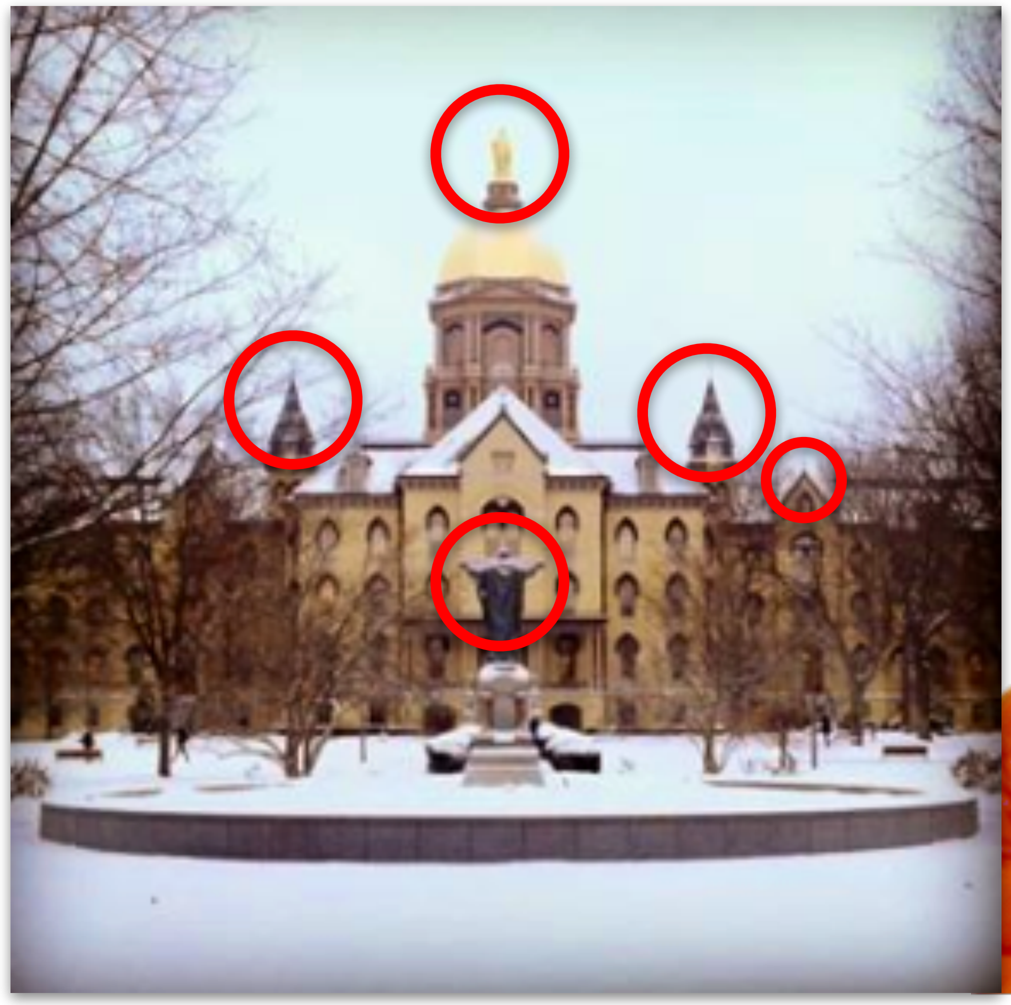
Image patterns that differ from their immediate neighborhood.

LOYOLA UNIVERSITY CHICAGO

Local Features

What are Local Features?
Image patterns that differ from their immediate neighborhood.

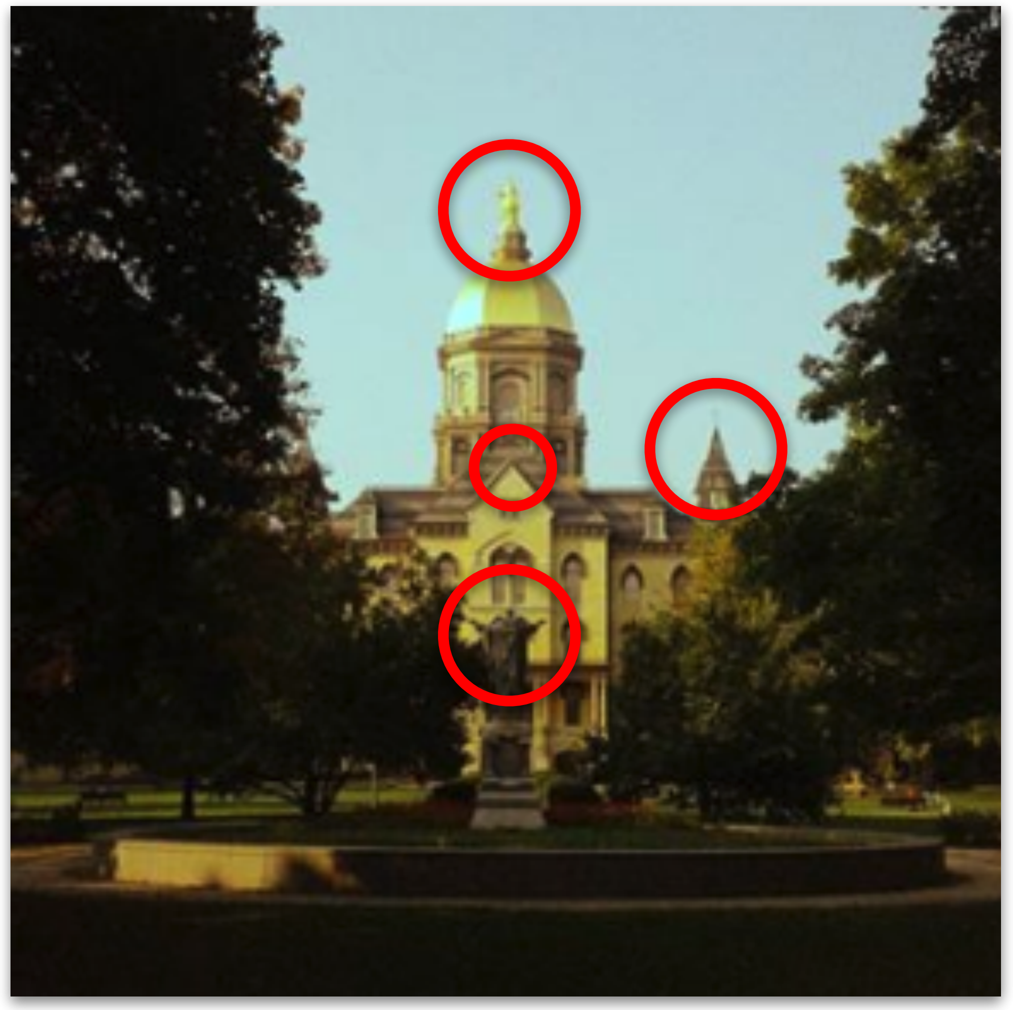


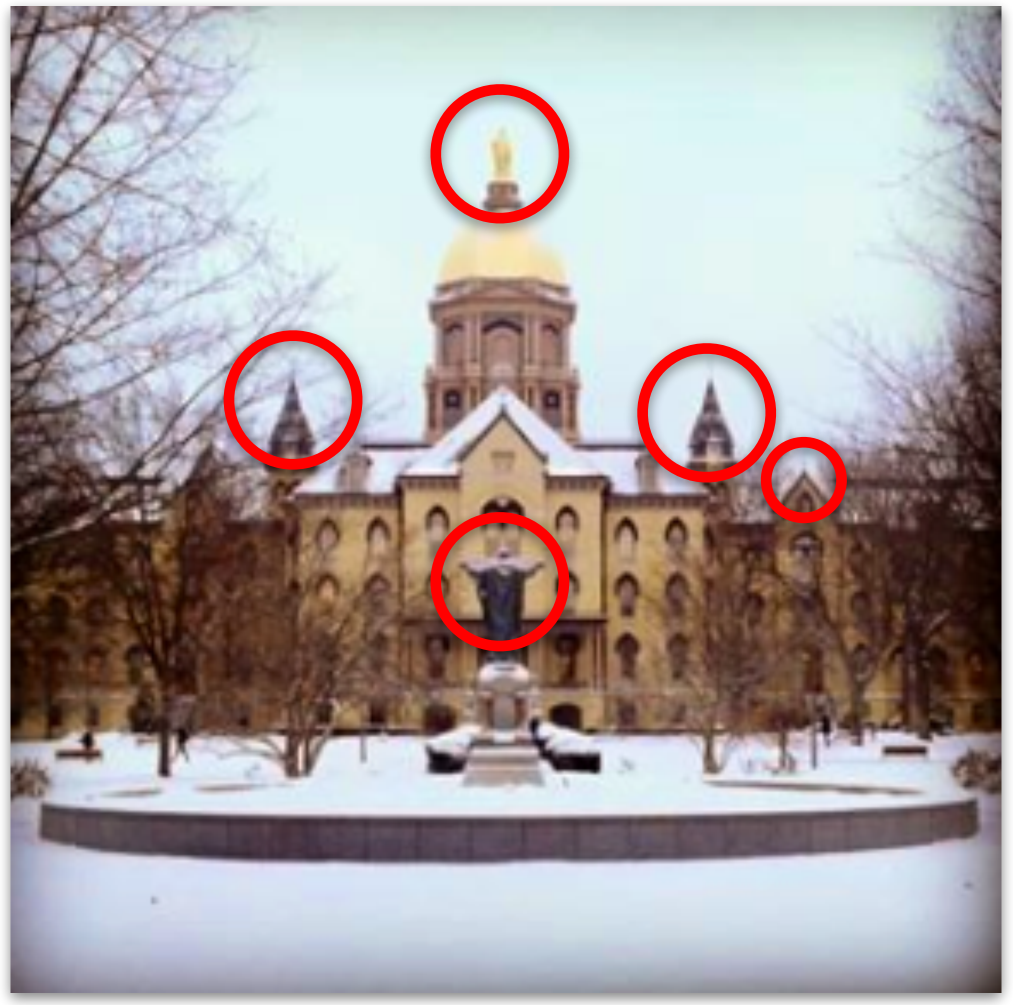


Local Features

What are Local Features?
Image patterns that differ from their immediate neighborhood.

Possible targets: points, edges, corners, junctions, blobs, etc.





Local Features

Why should one use Local Features?

Relevance

(i) Edges are usually enough for humans' object recognition.

(ii) Removing the corners hinders humans' abilities.




craiyon.com

LOYOLA UNIVERSITY CHICAGO

Local Features

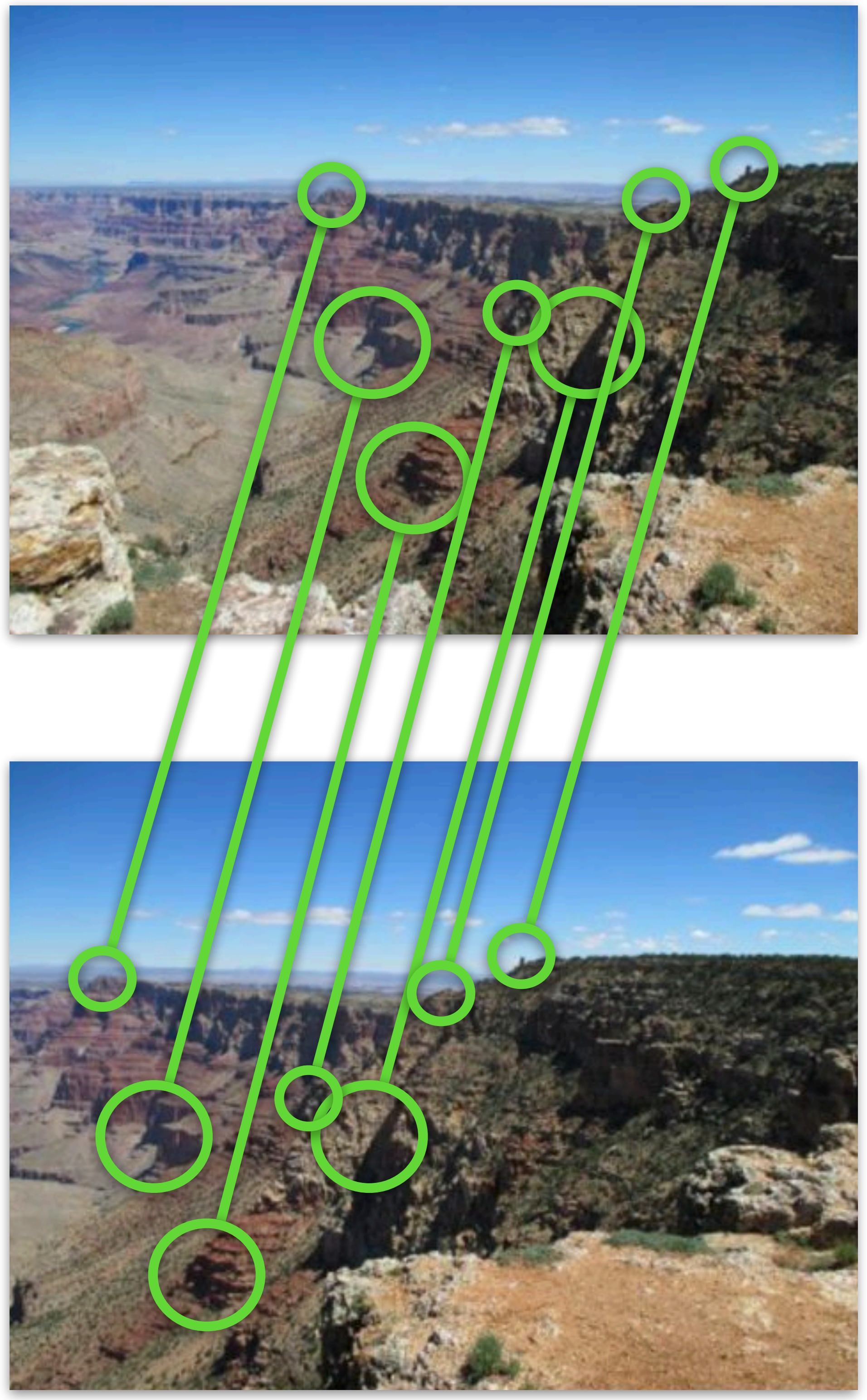
Why should one use Local Features?
Establish anchor points for image registration.

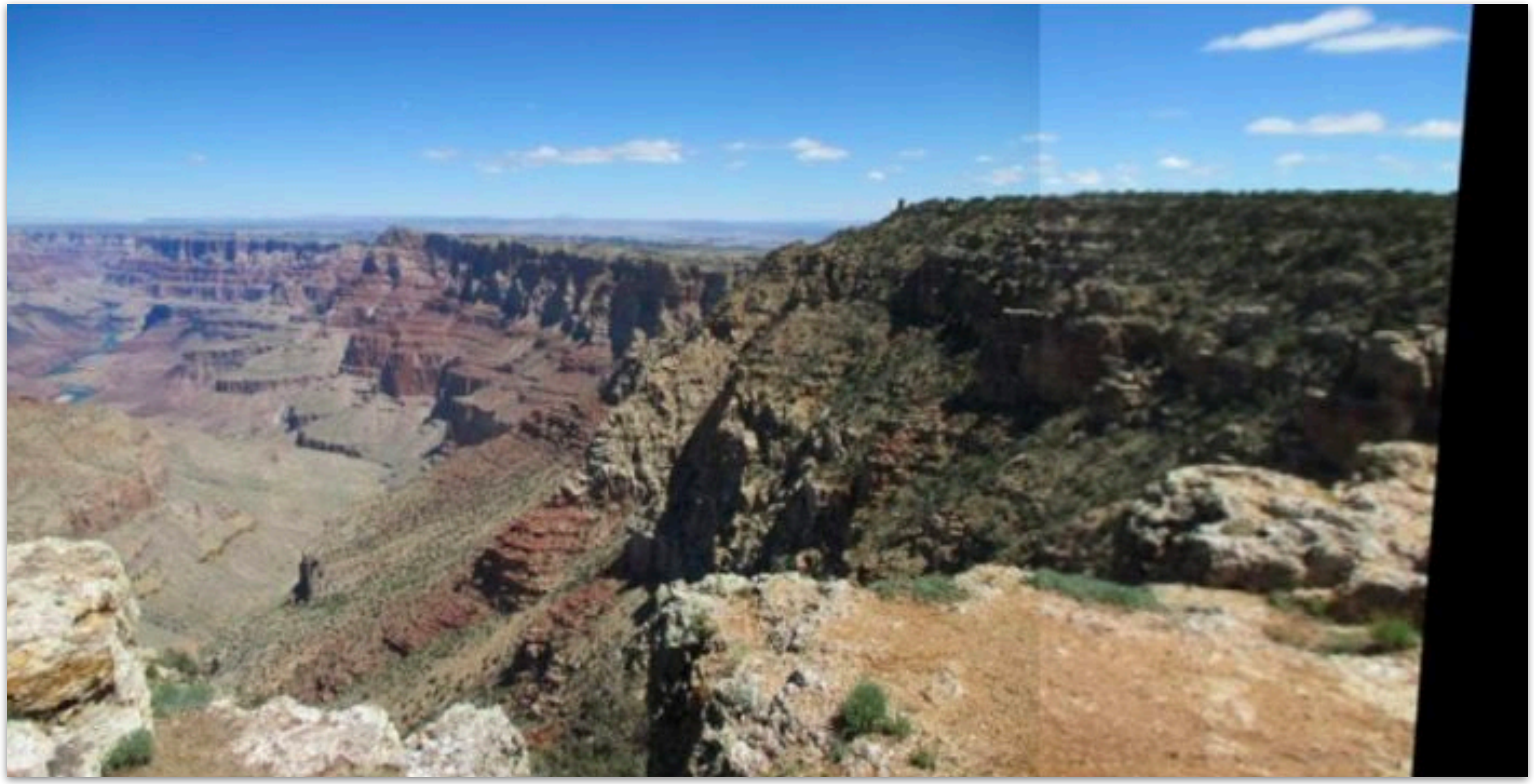
Local Features

Why should one use Local Features?

Establish anchor points for image registration.

pyimagesearch.com





LOYOLA UNIVERSITY CHICAGO

Local Features

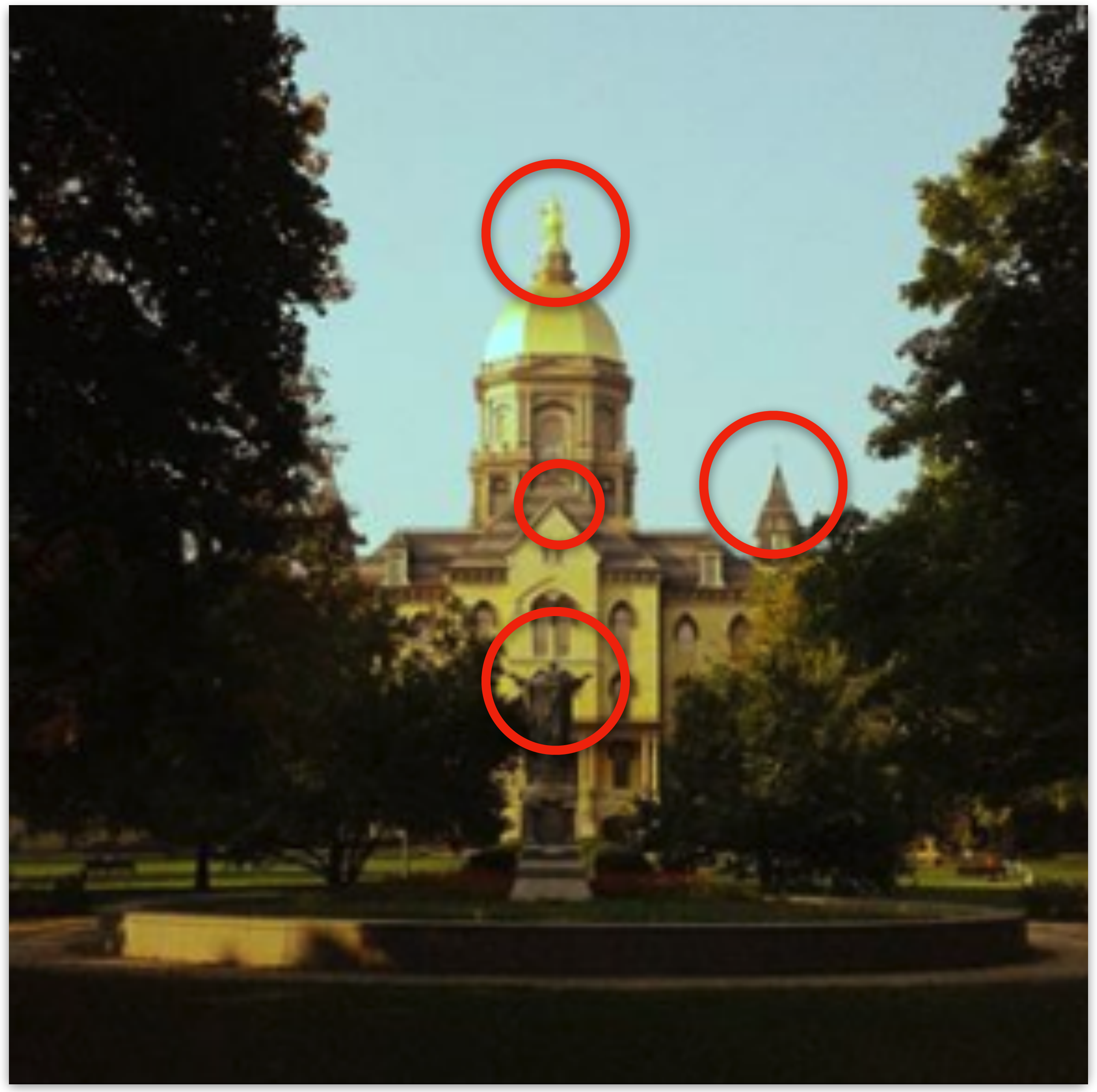
Why should one use Local Features?
Obtain robust and compact image representation for many CV tasks.

Can a computer system decide if these images depict the same building?

LOYOLA UNIVERSITY CHICAGO

Local Features

Why should one use Local Features?
Obtain robust and compact image representation for many CV tasks.

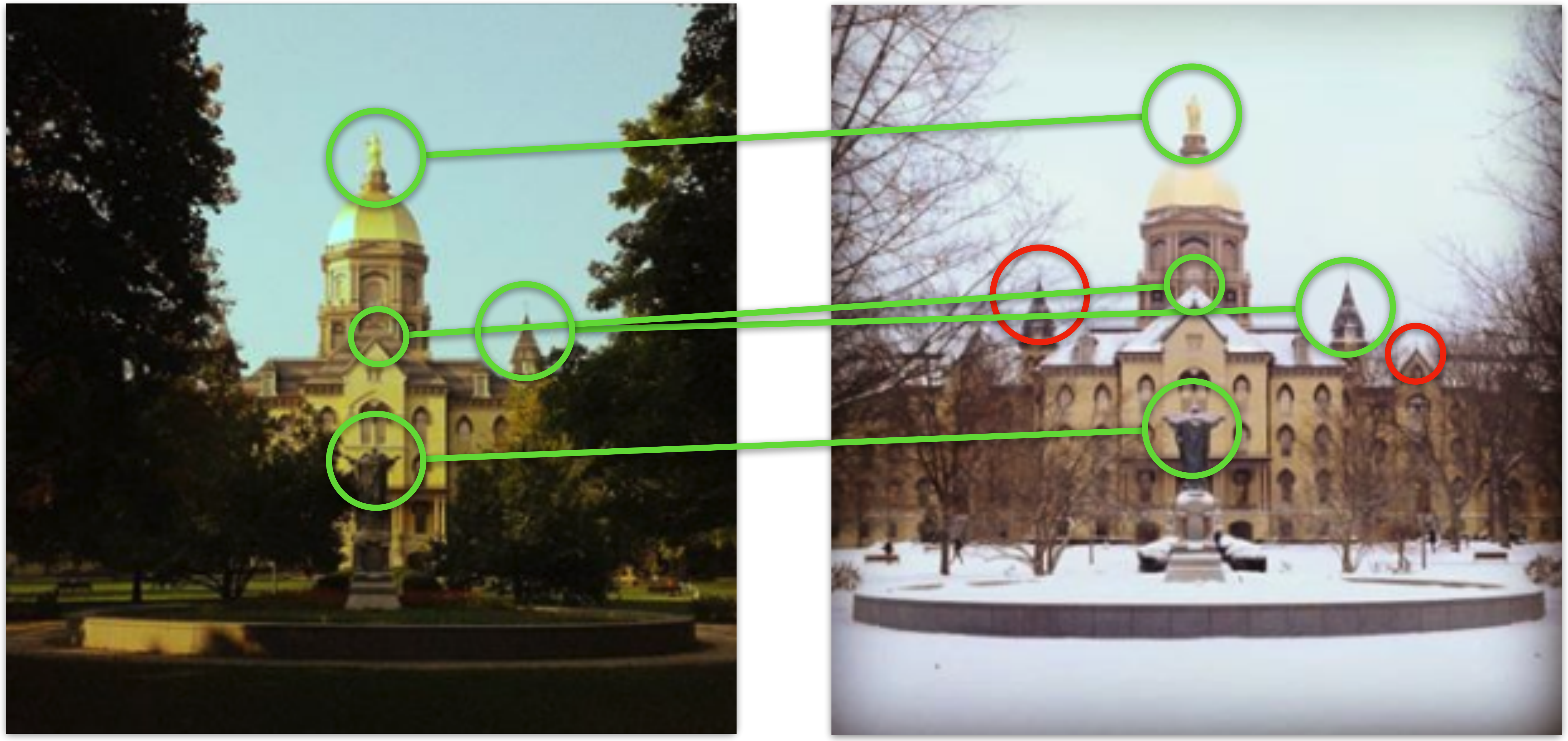


Can a computer system decide if these images depict the same building?

LOYOLA UNIVERSITY CHICAGO

Local Features

Why should one use Local Features?
Obtain robust and compact image representation for many CV tasks.



Can a computer system decide if these images depict the same building?

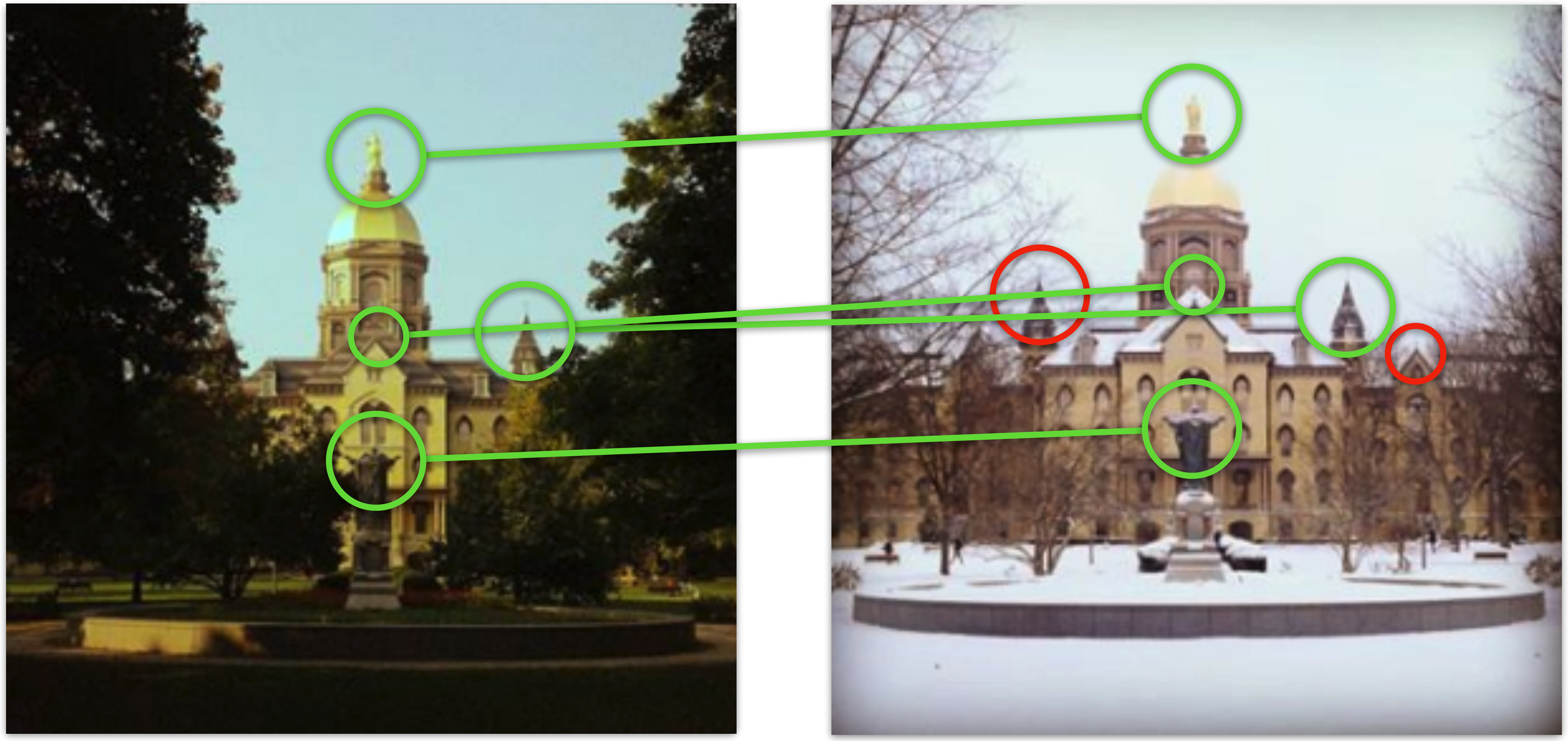
```
Yes_
```

In spite of occlusions.

LOYOLA UNIVERSITY CHICAGO

Local Features

Why should one use Local Features?
Obtain robust and compact image representation for many CV tasks.



Can a computer system decide if these images depict the same building?

```
Yes_
```

Index the local features instead of the entire images.

LOYOLA UNIVERSITY CHICAGO

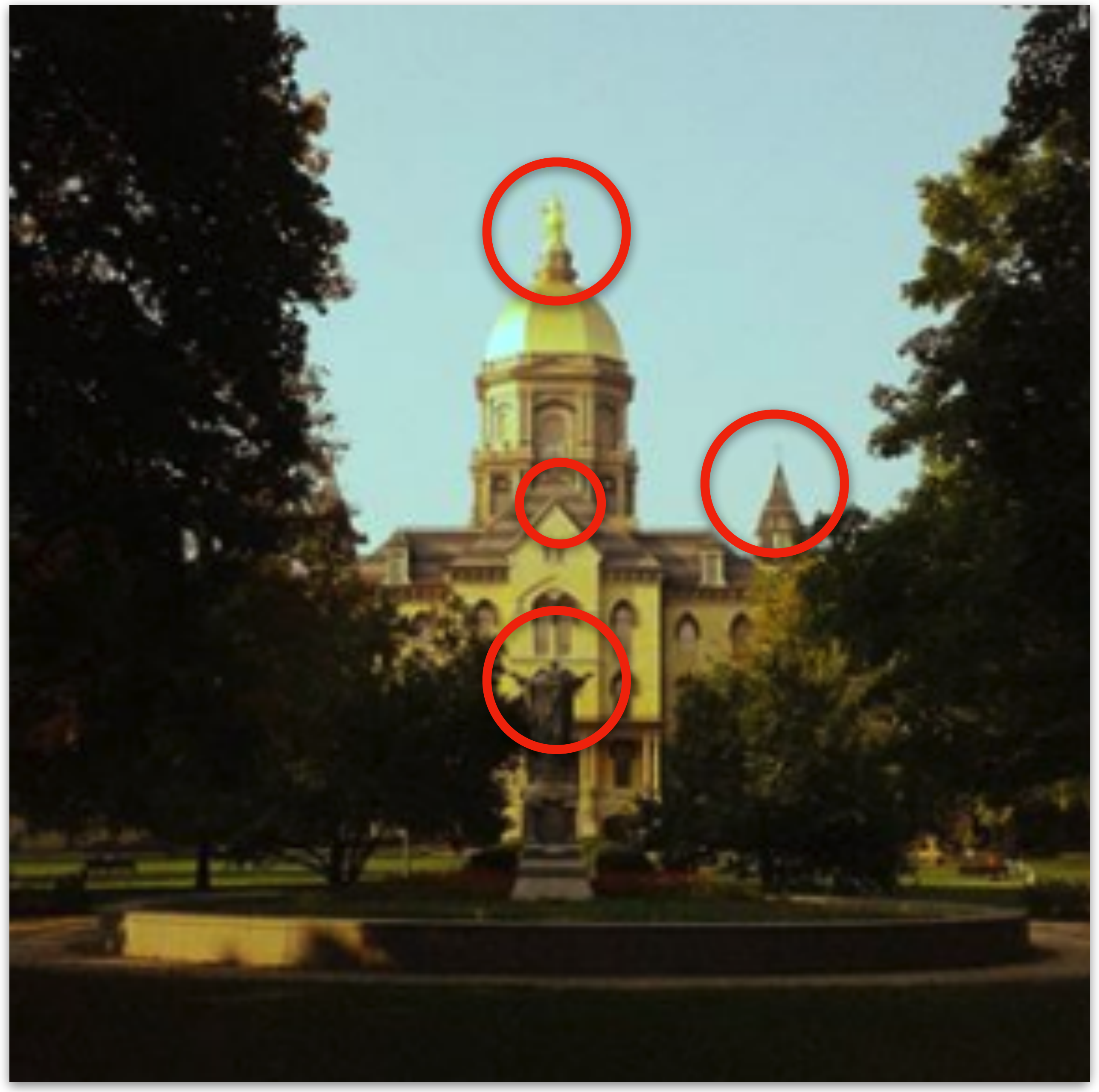
Local Features

Local Feature Steps

Feature Detection
Interest points, keypoints, or regions of interest are identified within the image.

Desired elements: location *(x, y)*, scale, orientation, strength.

Desired properties: repeatability and distinctiveness.



LOYOLA UNIVERSITY CHICAGO

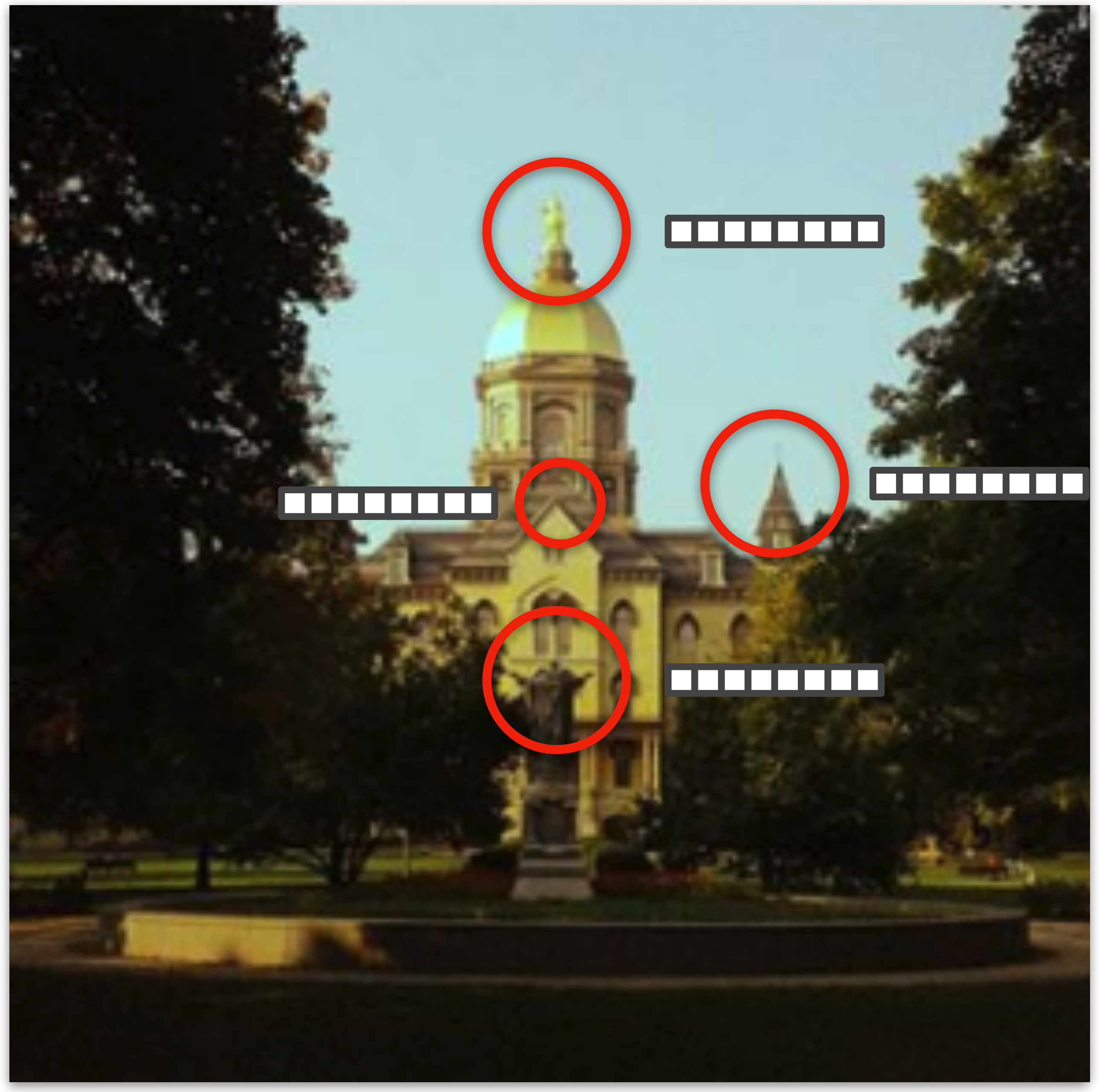
Local Features

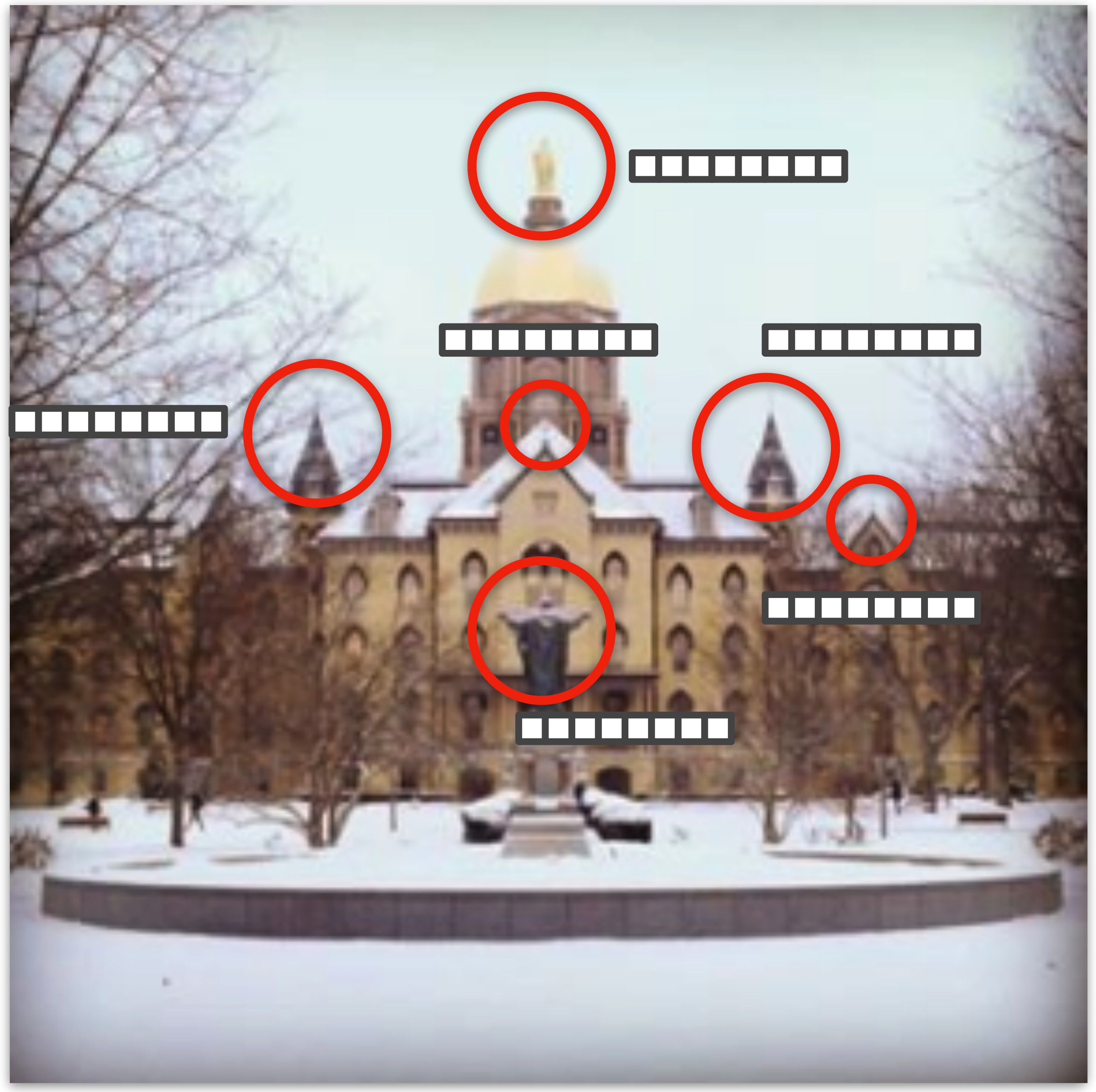
Local Feature Steps

Feature Description
Tensors are computed over the regions of the detected local features.

Desired element: *N*-Dimensional feature vector.

Desired properties: efficiency and robustness to different capture conditions.





LOYOLA UNIVERSITY CHICAGO

Harris Corner Detector

Focus on Corners

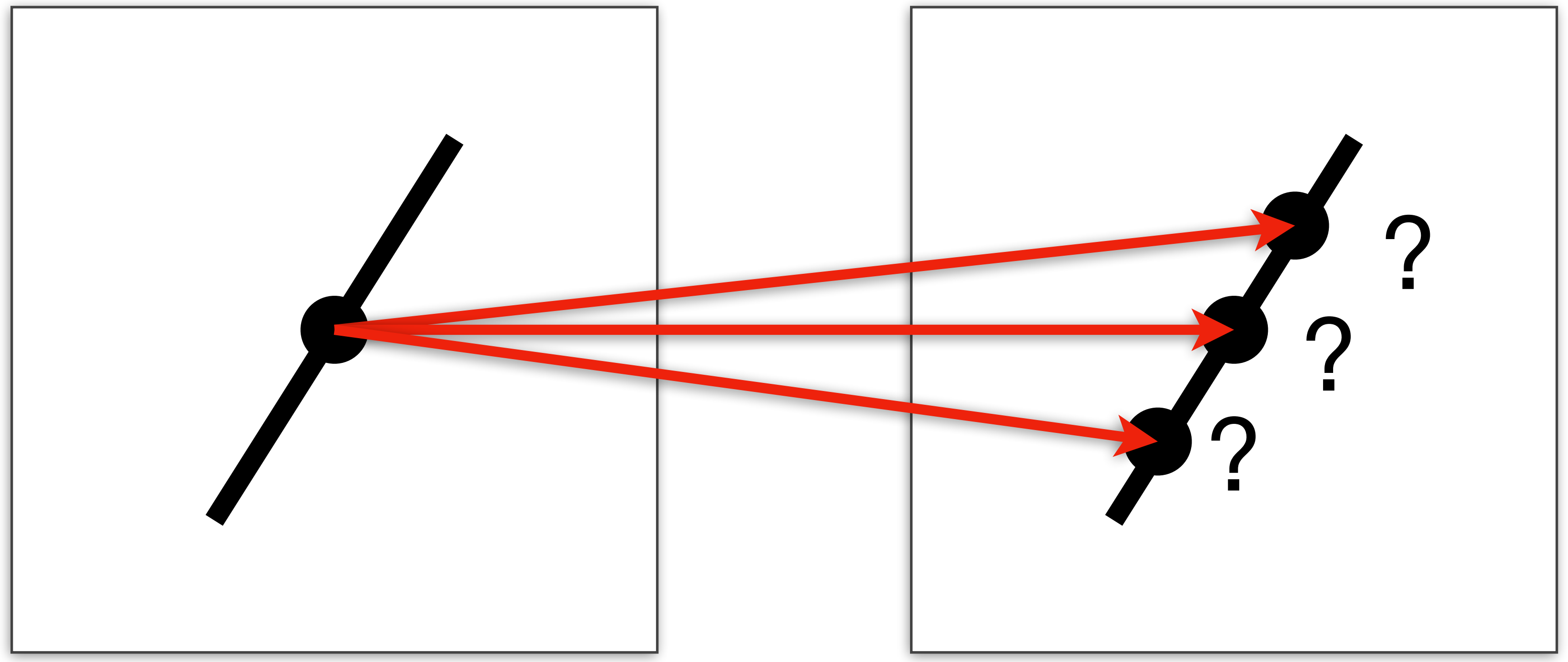
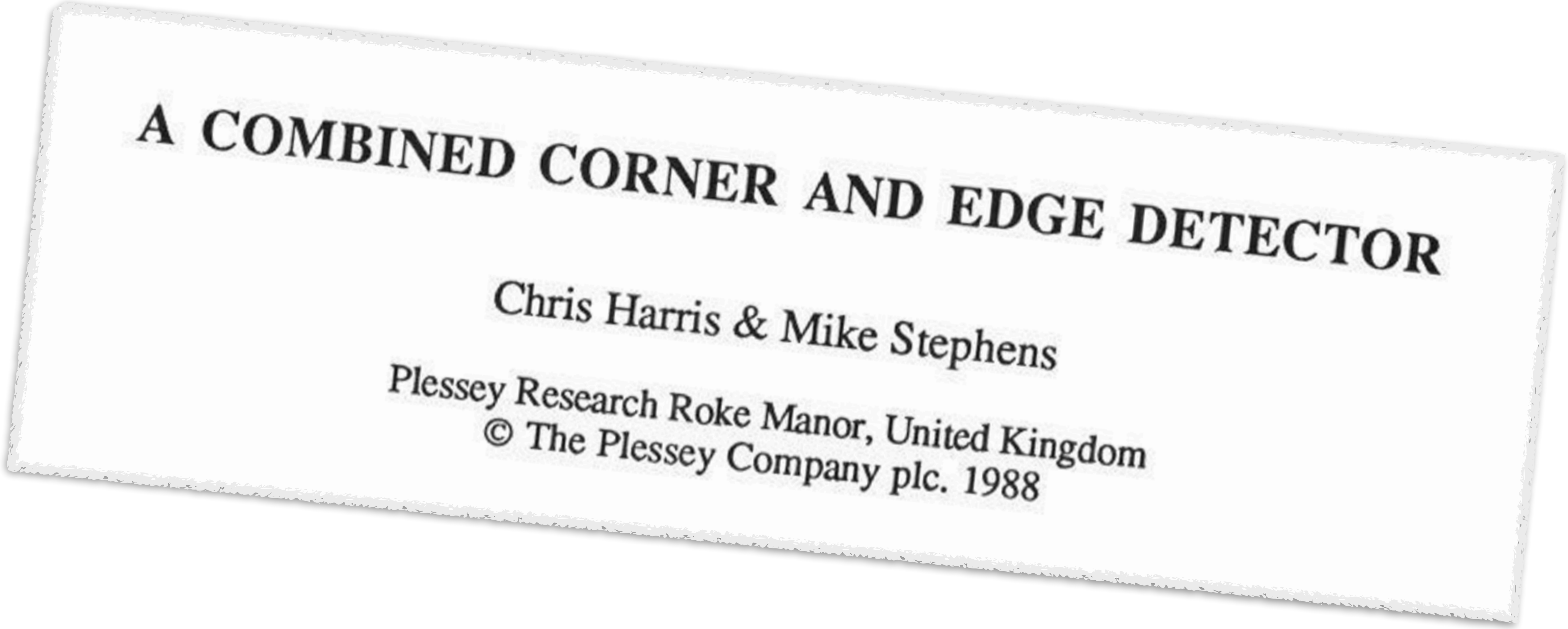


Image 1 Image 2



A COMBINED CORNER AND EDGE DETECTOR

Chris Harris & Mike Stephens

Plessey Research Roke Manor, United Kingdom
© The Plessey Company plc. 1988

Harris Corner Detector

Focus on Corners

Corners are easier to find and match.

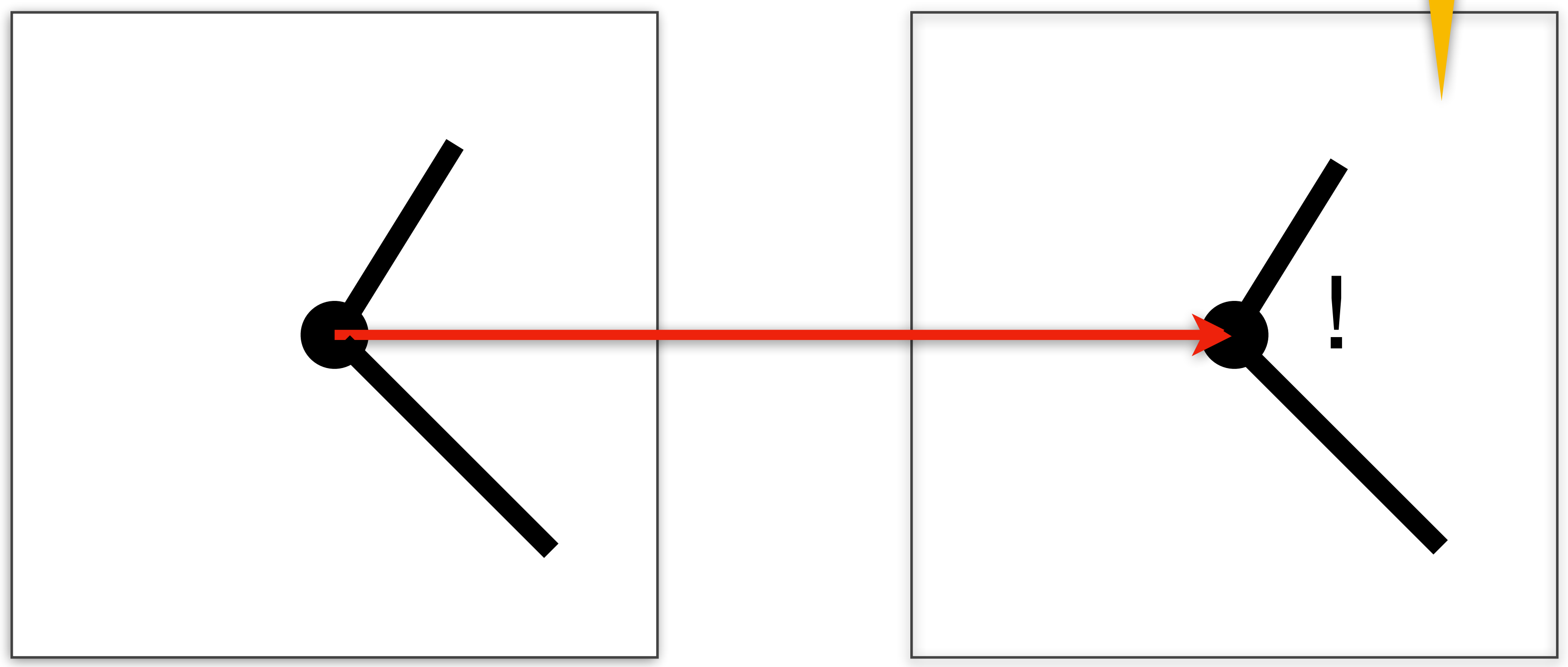
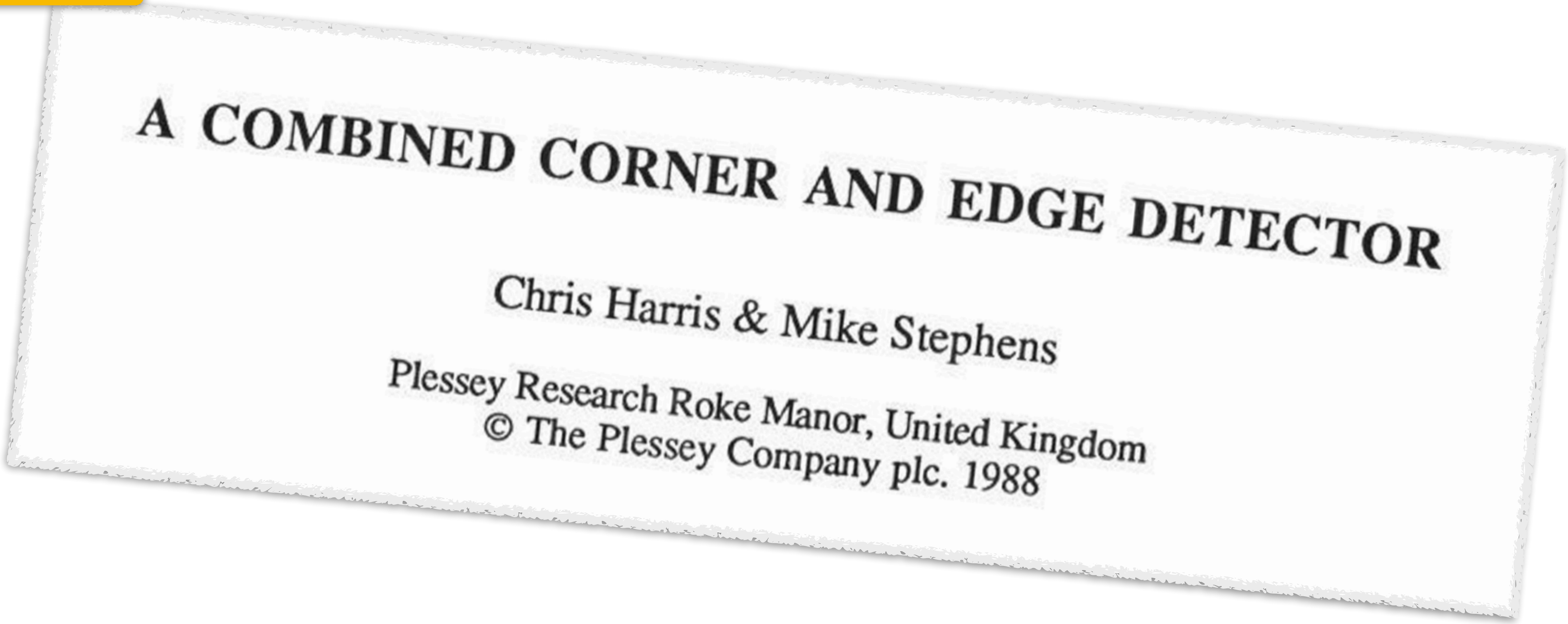


Image 1 Image 2



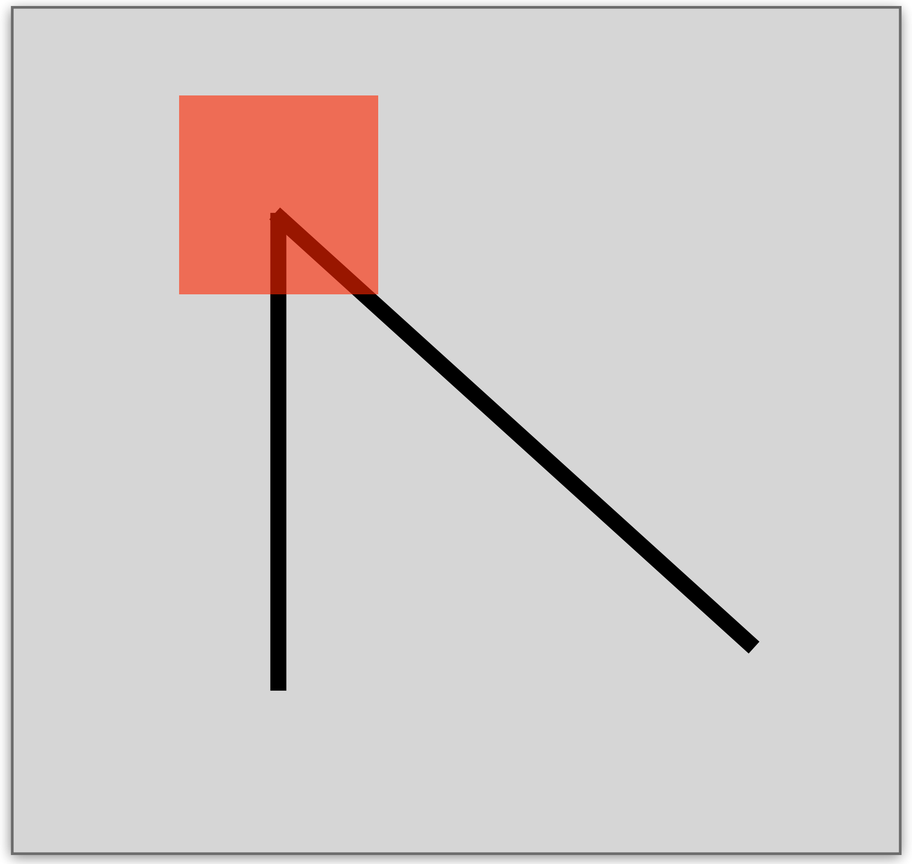
A COMBINED CORNER AND EDGE DETECTOR

Chris Harris & Mike Stephens

Plessey Research Roke Manor, United Kingdom
© The Plessey Company plc. 1988

Harris Corner Detector

Focus on Corners

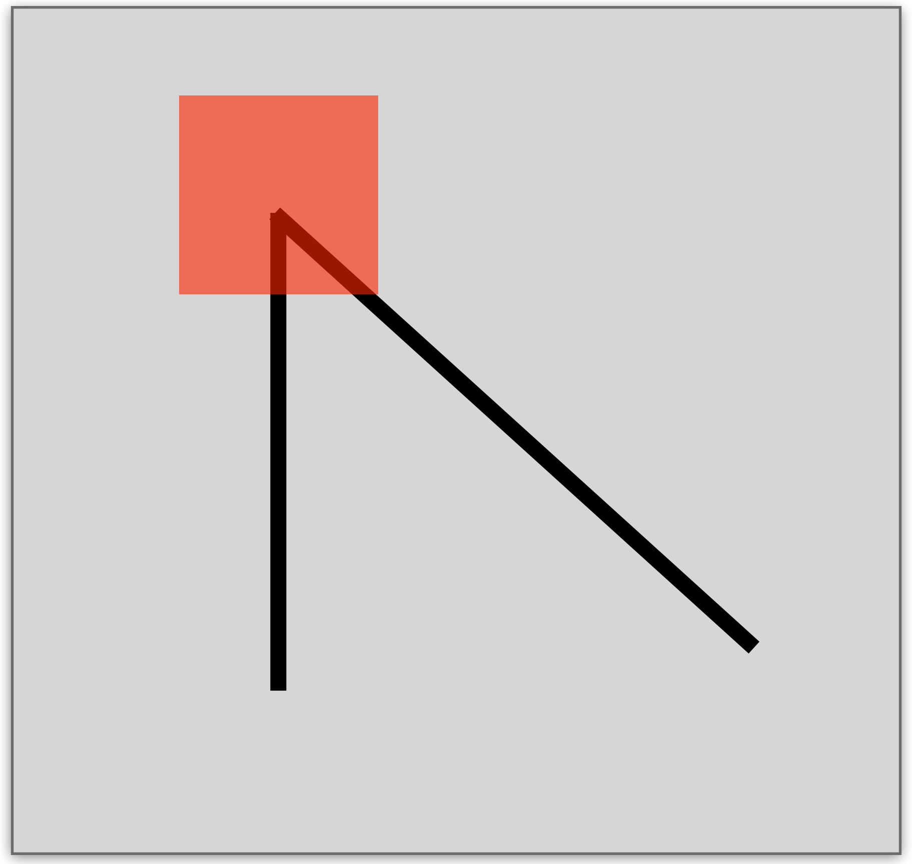


Consider a small $(n \times n)$-pixel **window** $w(x, y)$ around each pixel $I(x, y)$ of a target image I.

Which of the windows depict corners?

Harris Corner Detector

Focus on Corners



Consider a small $(n \times n)$-pixel window $w(x, y)$ around each pixel $I(x, y)$ of a target image I.

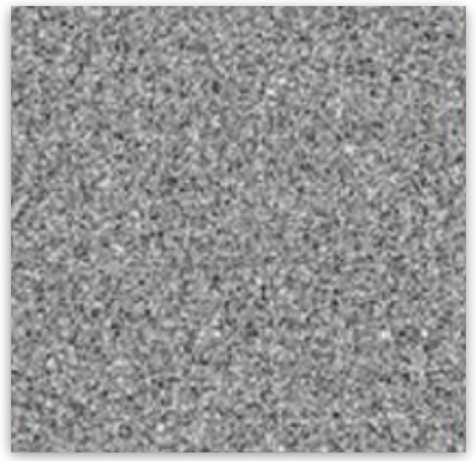
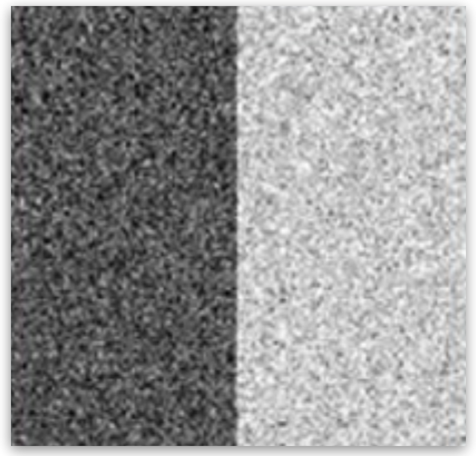
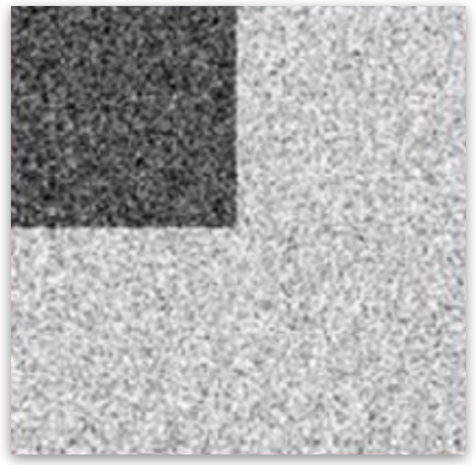
For each window, consider the **image gradients**:

- In x direction: $I_x = \dfrac{\delta I}{\delta_x}$ - In y direction: $I_y = \dfrac{\delta I}{\delta_y}$

LOYOLA UNIVERSITY CHICAGO

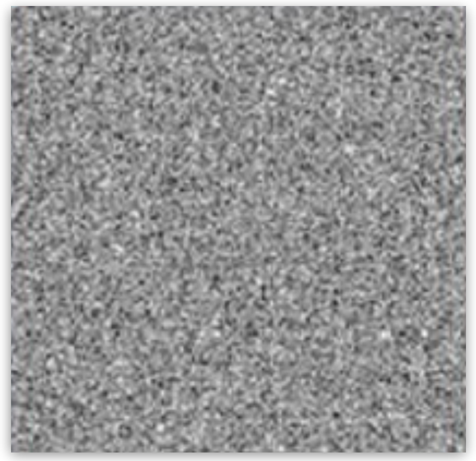
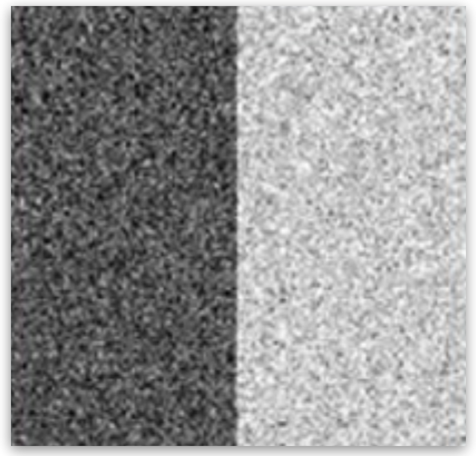
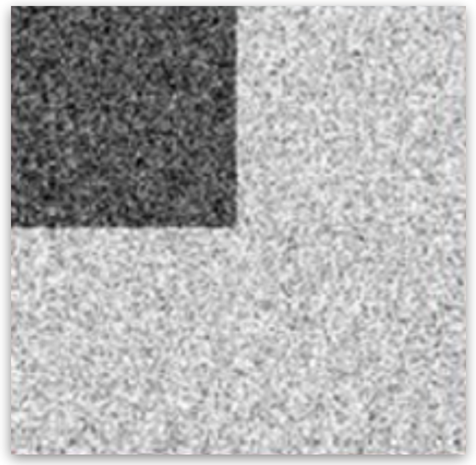
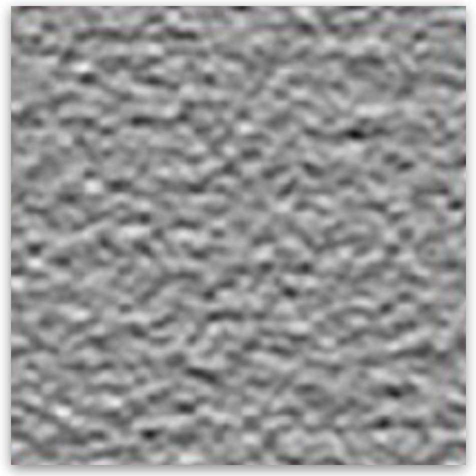
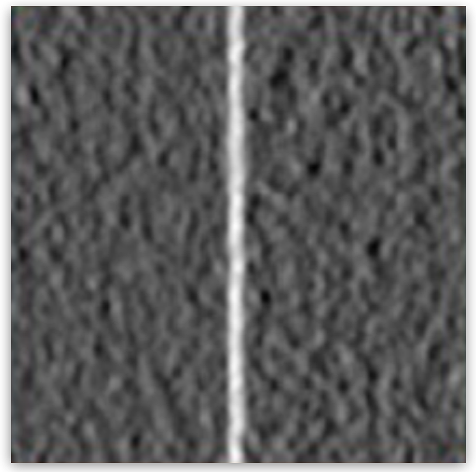
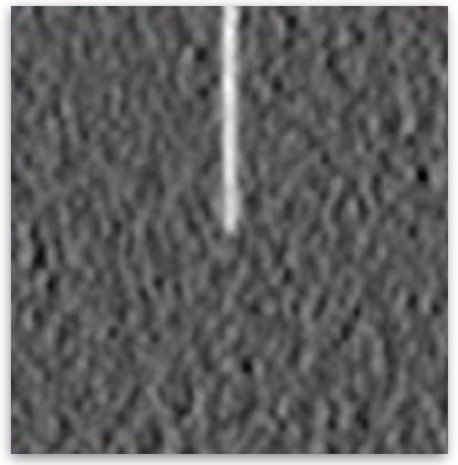
Harris Corner Detector

Image Gradients

	Flat	*Edge (vertical)*	*Corner*
Image			

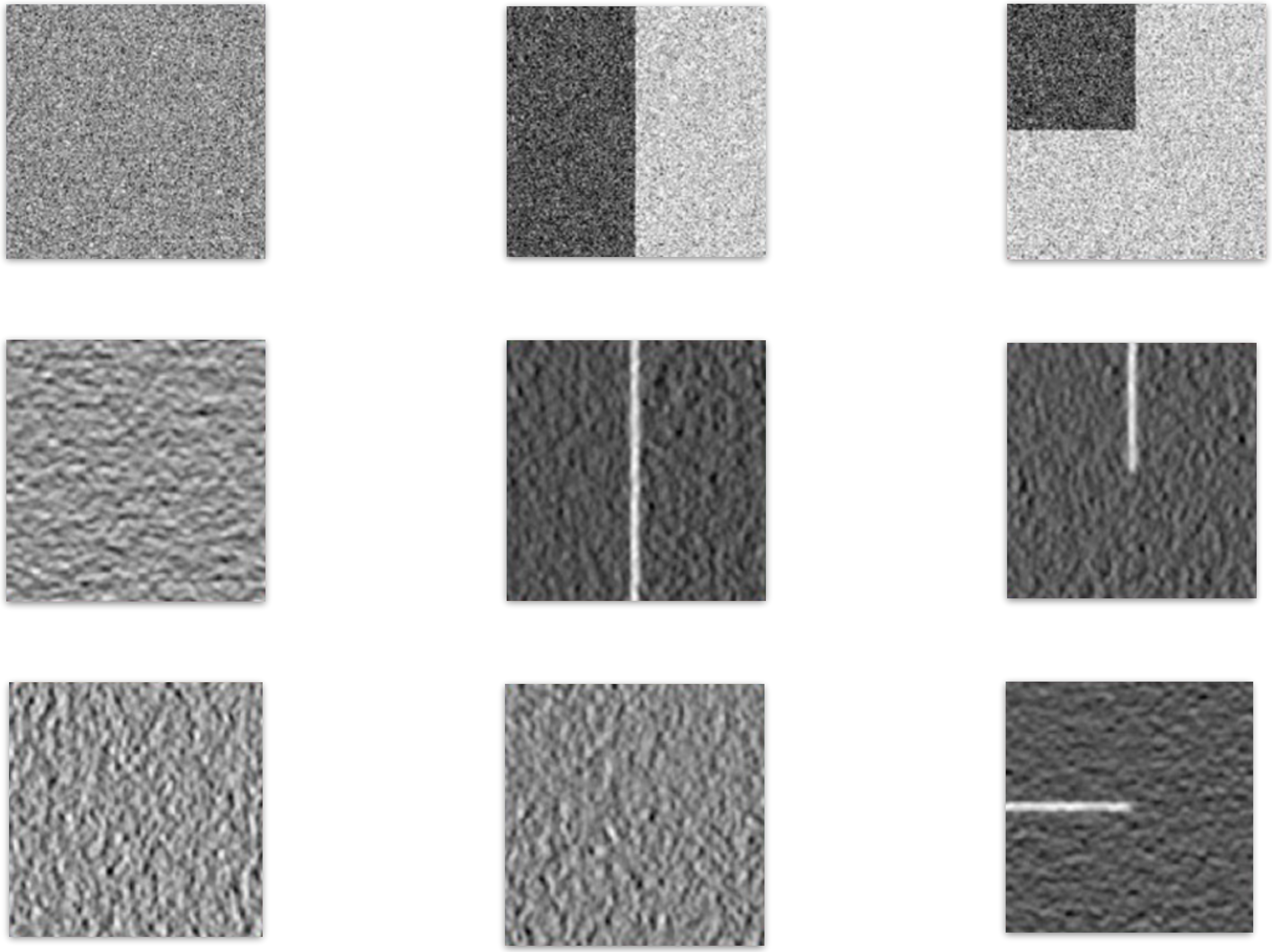
Harris Corner Detector

Image Gradients

	Flat	*Edge (vertical)*	*Corner*	*Possible Implementation*
Image				
I_x				Image convolution with Sobel filter $\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$

Harris Corner Detector

Image Gradients



	Flat	*Edge (vertical)*	*Corner*	*Possible Implementation*
Image				
I_x				Image convolution with Sobel filter $\begin{bmatrix}1 & 0 & -1\\ 2 & 0 & -2\\ 1 & 0 & -1\end{bmatrix}$
I_y				Image convolution with Sobel filter $\begin{bmatrix}1 & 2 & 1\\ 0 & 0 & 0\\ -1 & -2 & -1\end{bmatrix}$

Both I_x and I_y should be large.

Harris Corner Detector

Image Gradients

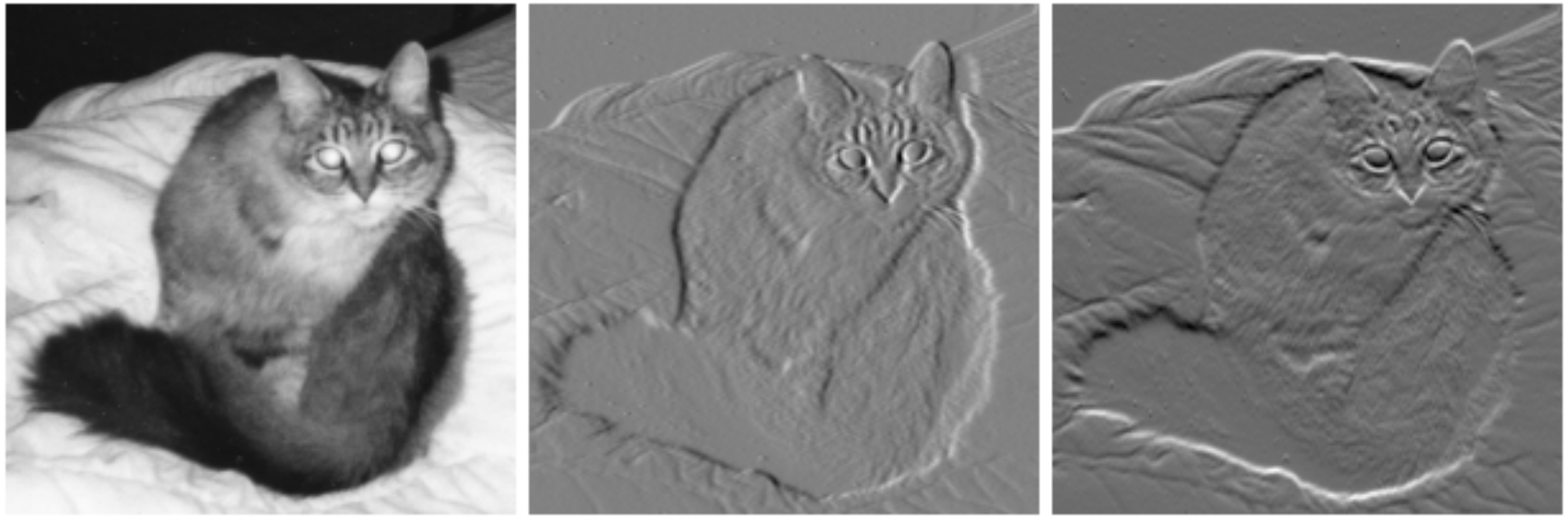
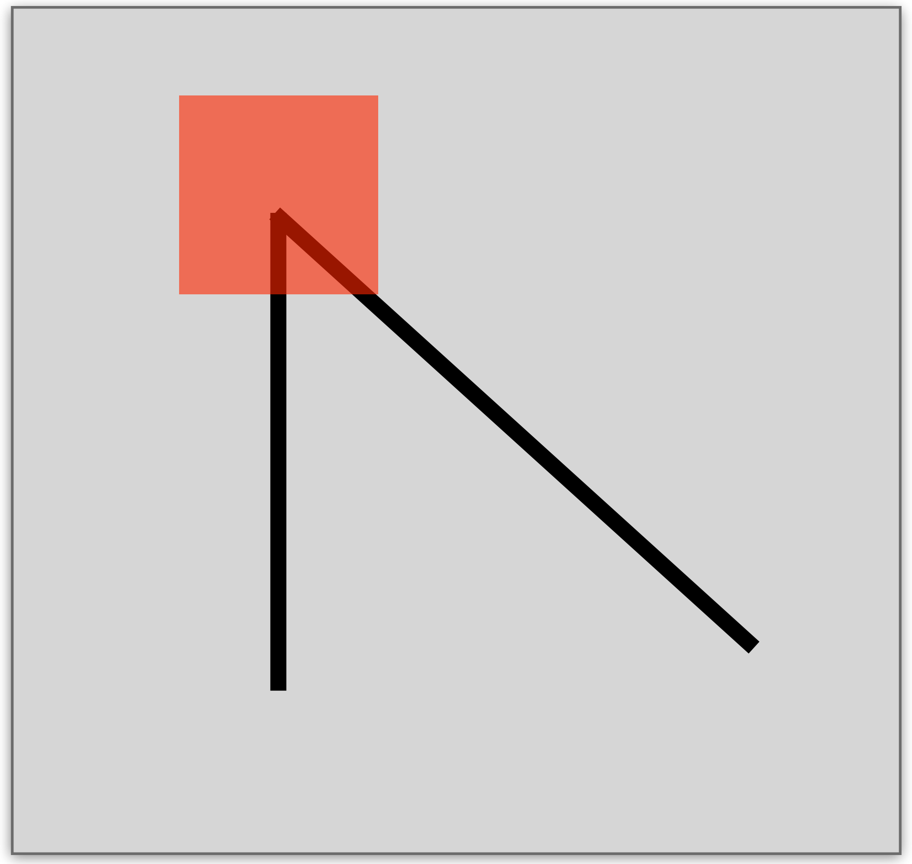


Image $\quad\quad$ I_x $\quad\quad$ I_y

Harris Corner Detector

Focus on Corners



Consider a small $(n \times n)$-pixel window $w(x, y)$ around each pixel $I(x, y)$ of a target image I.

For each window, compute the **structure tensor**:

$$M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x I_x & I_x I_y \\ I_x I_y & I_y I_y \end{bmatrix} = \begin{bmatrix} \sum_{x,y} w(x,y) I_x^2 & \sum_{x,y} w(x,y) I_x I_y \\ \sum_{x,y} w(x,y) I_x I_y & \sum_{x,y} w(x,y) I_y^2 \end{bmatrix}$$

LOYOLA UNIVERSITY CHICAGO

Harris Corner Detector

Structure Tensor
(a.k.a. second-moment matrix)

$$M = \begin{bmatrix} \sum_{x,y} w(x,y) I_x^2 & \sum_{x,y} w(x,y) I_x I_y \\ \sum_{x,y} w(x,y) I_x I_y & \sum_{x,y} w(x,y) I_y^2 \end{bmatrix}$$

According to linear algebra principles, the eigenvalues λ_1 and λ_2 of M express the spread of image gradient values in two different directions.

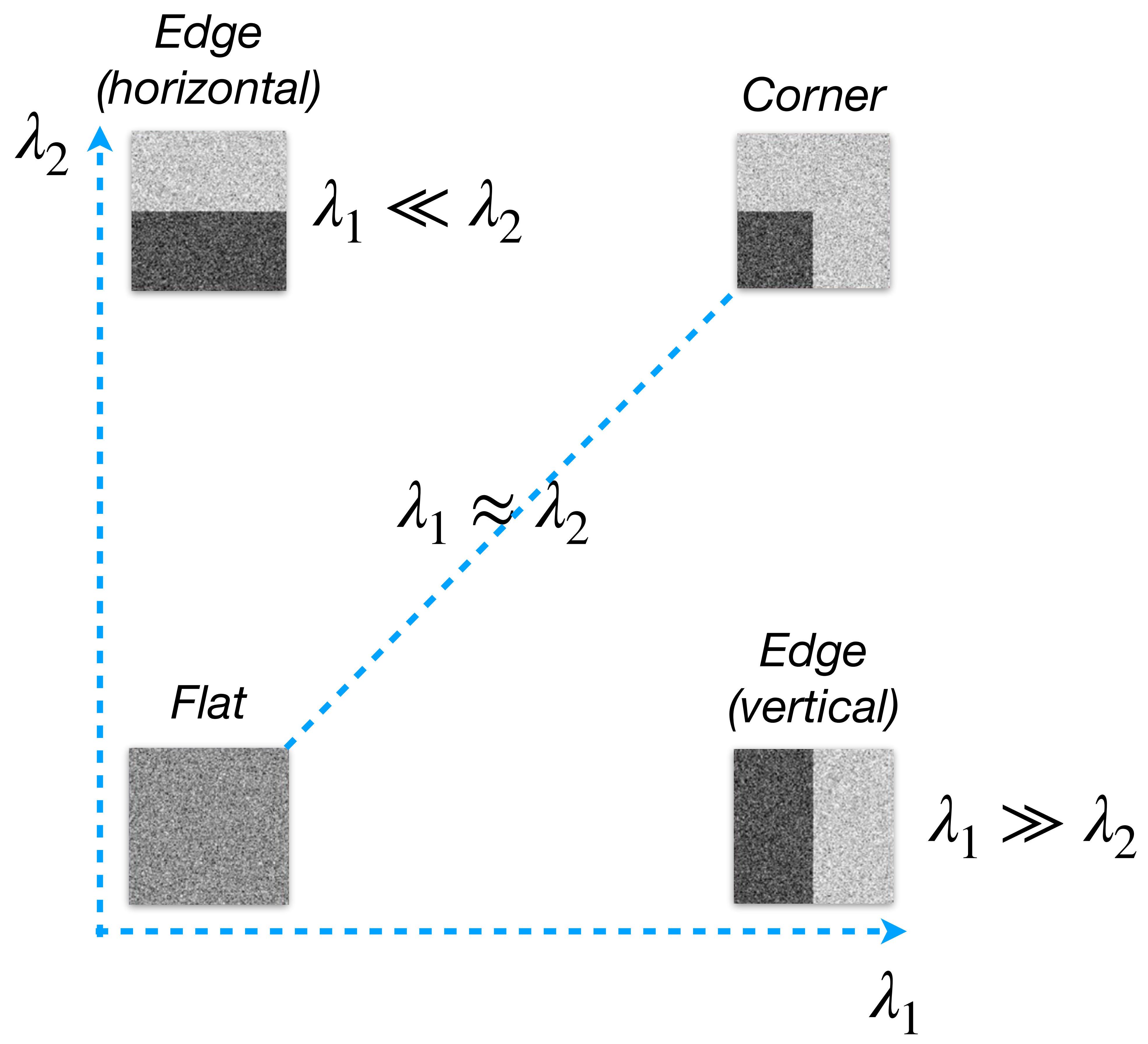
We want both values large, since two different directions define what a corner is.

Harris Corner Detector

Structure Tensor
(a.k.a. second-moment matrix)

$$M = \begin{bmatrix} \sum_{x,y} w(x,y) I_x^2 & \sum_{x,y} w(x,y) I_x I_y \\ \sum_{x,y} w(x,y) I_x I_y & \sum_{x,y} w(x,y) I_y^2 \end{bmatrix}$$

Eigenvalues: λ_1 and λ_2



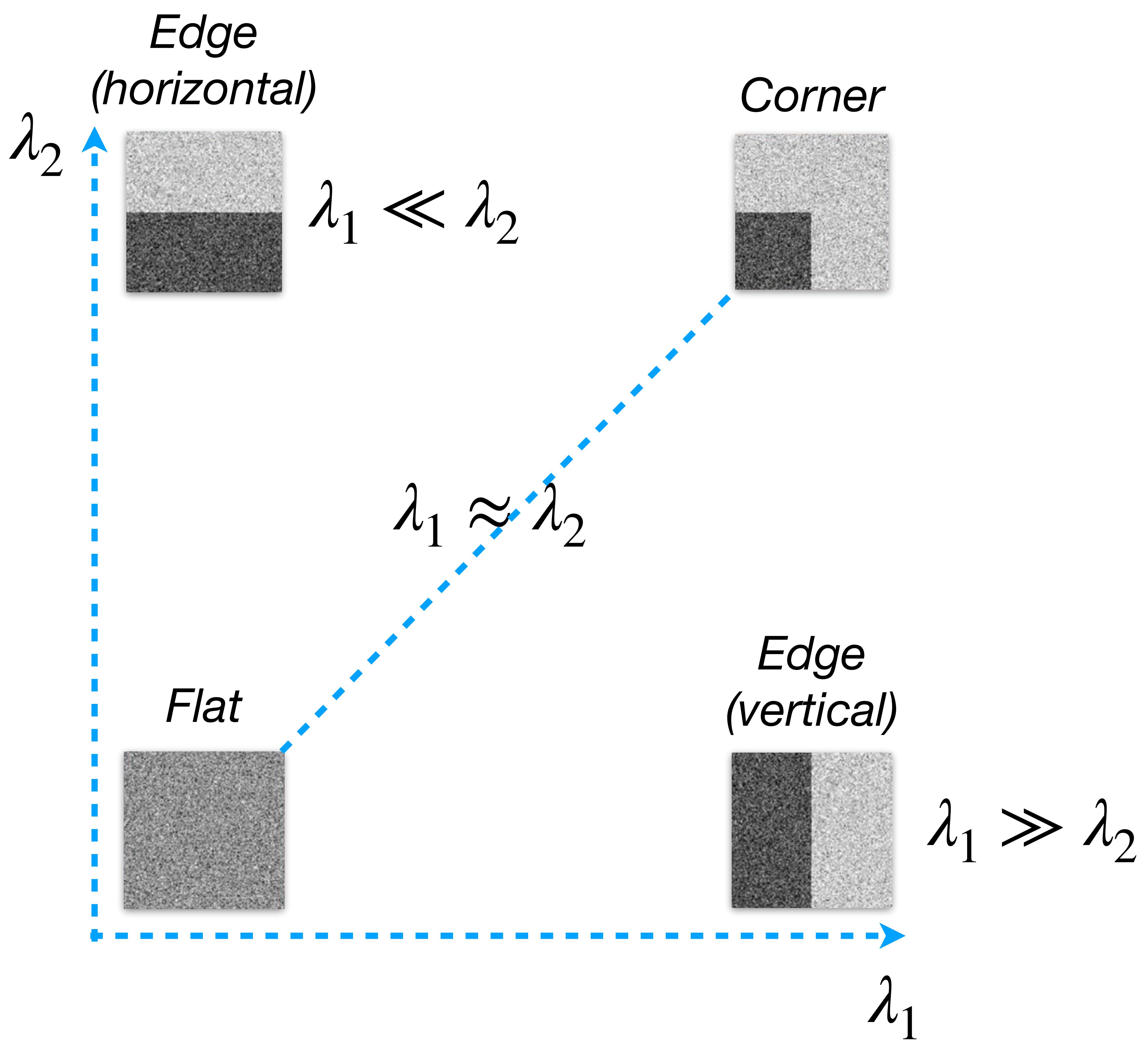
Harris Corner Detector

Structure Tensor
(a.k.a. second-moment matrix)

$$M = \begin{bmatrix} \sum_{x,y} w(x,y) I_x^2 & \sum_{x,y} w(x,y) I_x I_y \\ \sum_{x,y} w(x,y) I_x I_y & \sum_{x,y} w(x,y) I_y^2 \end{bmatrix}$$

Eigenvalues: λ_1 and λ_2

Harris and Stephens' idea
Leverage the following properties:
$det(M) = \lambda_1 \lambda_2$ and $trace(M) = \lambda_1 + \lambda_2$



LOYOLA UNIVERSITY CHICAGO

Harris Corner Detector

Structure Tensor
(a.k.a. second-moment matrix)

$$M = \begin{bmatrix} \sum_{x,y} w(x,y) I_x^2 & \sum_{x,y} w(x,y) I_x I_y \\ \sum_{x,y} w(x,y) I_x I_y & \sum_{x,y} w(x,y) I_y^2 \end{bmatrix}$$

Eigenvalues: λ_1 and λ_2

"Cornerness" Score R

No need to compute λ_1 and λ_2 explicitly but leverage:

$det(M) = \lambda_1 \lambda_2$ and $trace(M) = \lambda_1 + \lambda_2$

$$R = det(M) - k(trace(M))^2$$

LOYOLA UNIVERSITY CHICAGO

Harris Corner Detector

Structure Tensor
(a.k.a. second-moment matrix)

$$M = \begin{bmatrix} \sum_{x,y} w(x,y) I_x^2 & \sum_{x,y} w(x,y) I_x I_y \\ \sum_{x,y} w(x,y) I_x I_y & \sum_{x,y} w(x,y) I_y^2 \end{bmatrix}$$

"Cornerness" Score R

No need to compute λ_1 and λ_2 explicitly but leverage:

$det(M) = \lambda_1 \lambda_2$ and $trace(M) = \lambda_1 + \lambda_2$

$$R = det(M) - k(trace(M))^2$$

$R \gg 0$**: we have a corner!** (λ_1 and λ_2 are both large)

$R < 0$: we have an edge, so ignore. ($\lambda_1 \gg \lambda_2$ or vice versa)
$|R| \approx 0$: we have a flat region, so ignore.

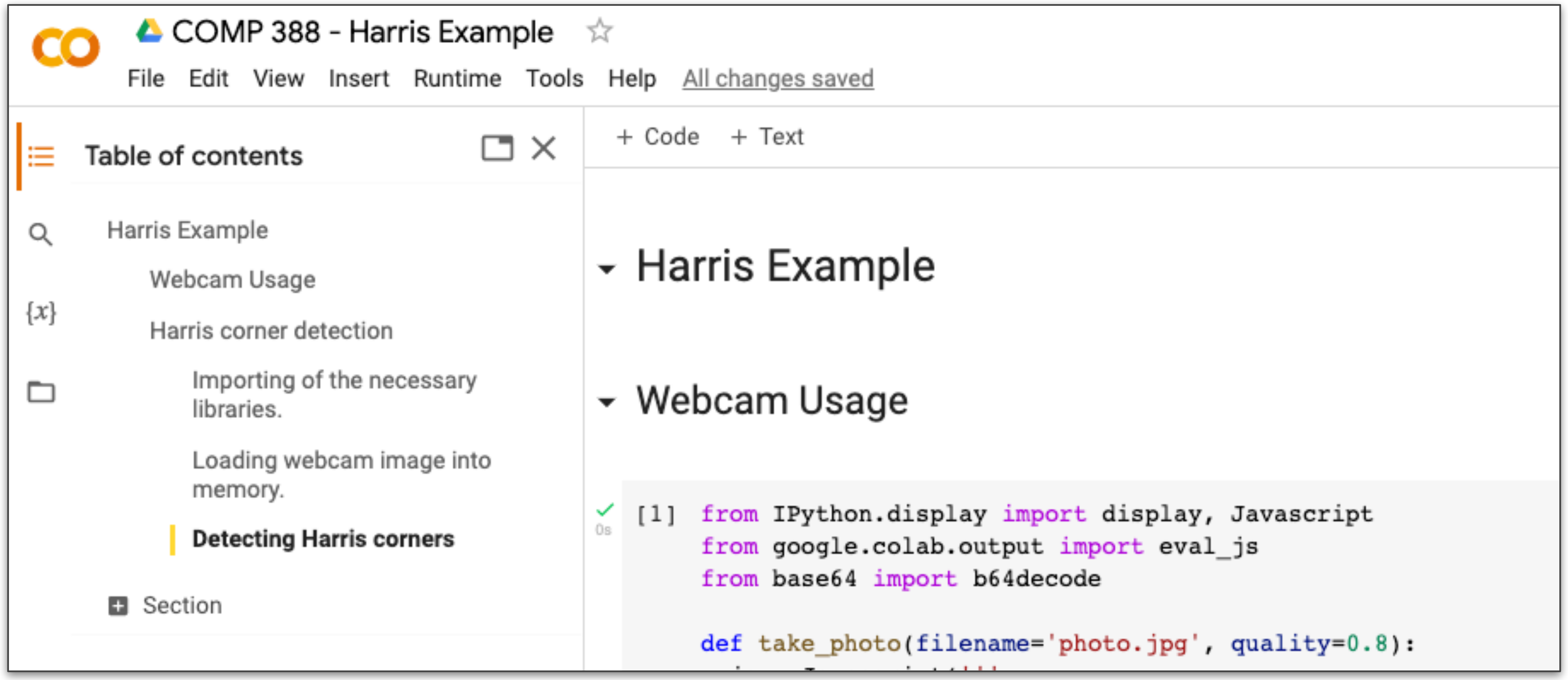
Handcrafted value:
$k = 0.04$

LOYOLA UNIVERSITY CHICAGO

Harris Corner Detector

Example

https://bit.ly/3qRBt2U



LOYOLA UNIVERSITY CHICAGO

Scale Invariant Feature Transform (SIFT)

How to match the same content captured with different resolutions?

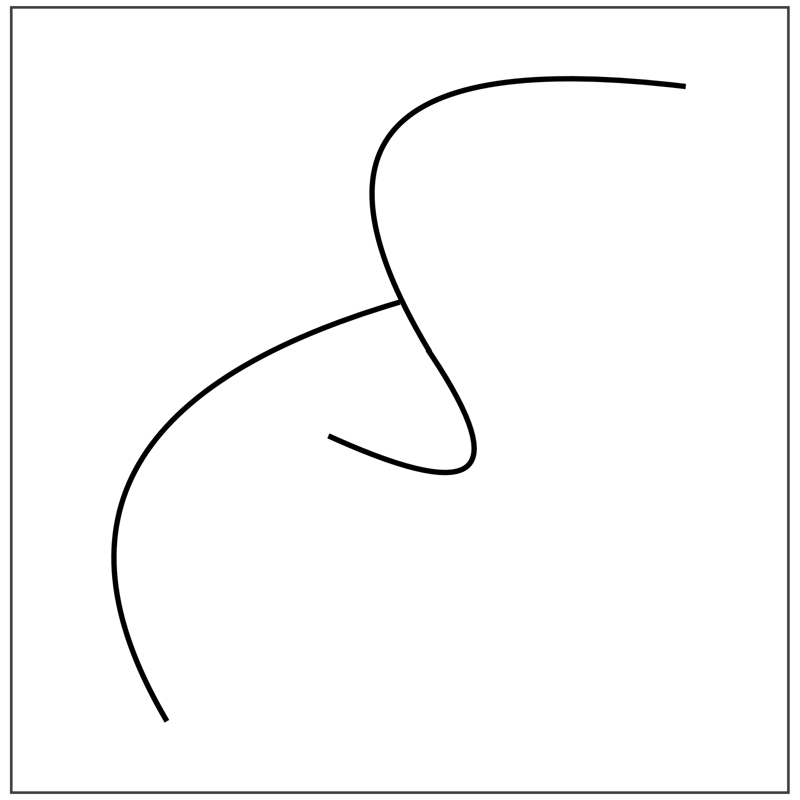


Image 1

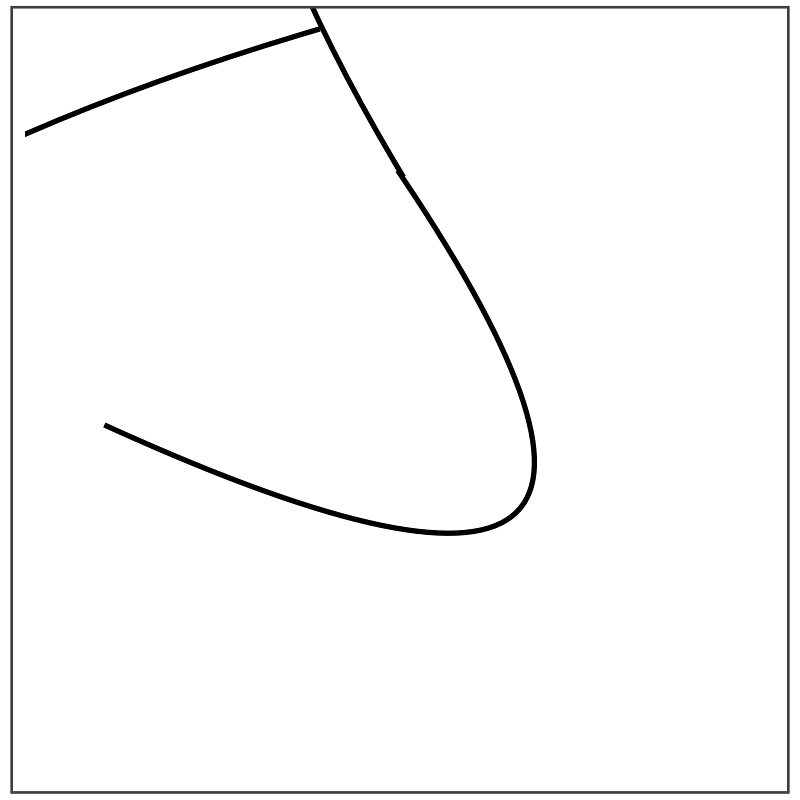


Image 2

International Journal of Computer Vision 60(2), 91–110, 2004
© 2004 Kluwer Academic Publishers. Manufactured in The Netherlands.

Distinctive Image Features from Scale-Invariant Keypoints

DAVID G. LOWE

Computer Science Department, University of British Columbia, Vancouver, B.C., Canada

Lowe@cs.ubc.ca

Received January 10, 2003; Revised January 7, 2004; Accepted January 22, 2004

Scale Invariant Feature Transform (SIFT)

How to match the same content captured with different resolutions?

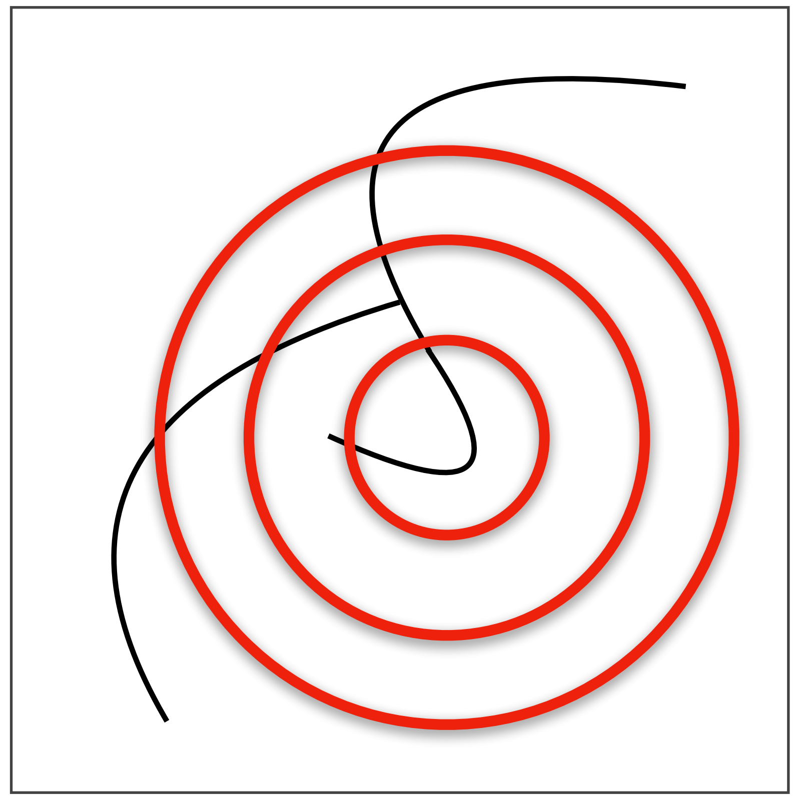


Image 1

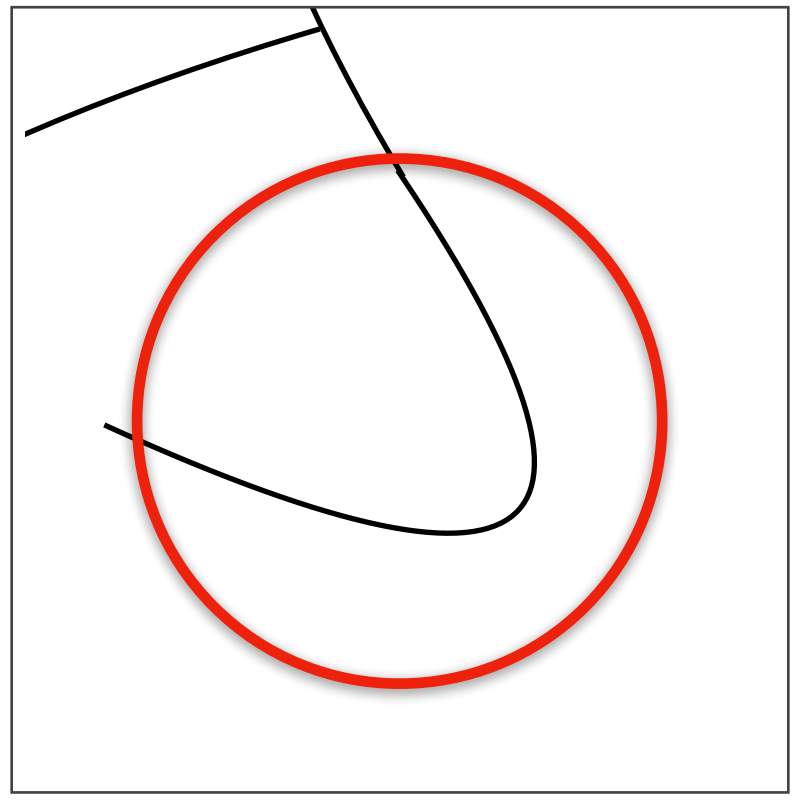


Image 2

International Journal of Computer Vision 60(2), 91–110, 2004

Distinctive Image Features from Scale-Invariant Keypoints

DAVID G. LOWE
Computer Science Department, University of British Columbia, Vancouver, B.C., Canada
Lowe@cs.ubc.ca

Received January 10, 2003; Revised January 7, 2004; Accepted January 22, 2004

Scale Invariant Feature Transform (SIFT)

How to match the same content captured with different resolutions?

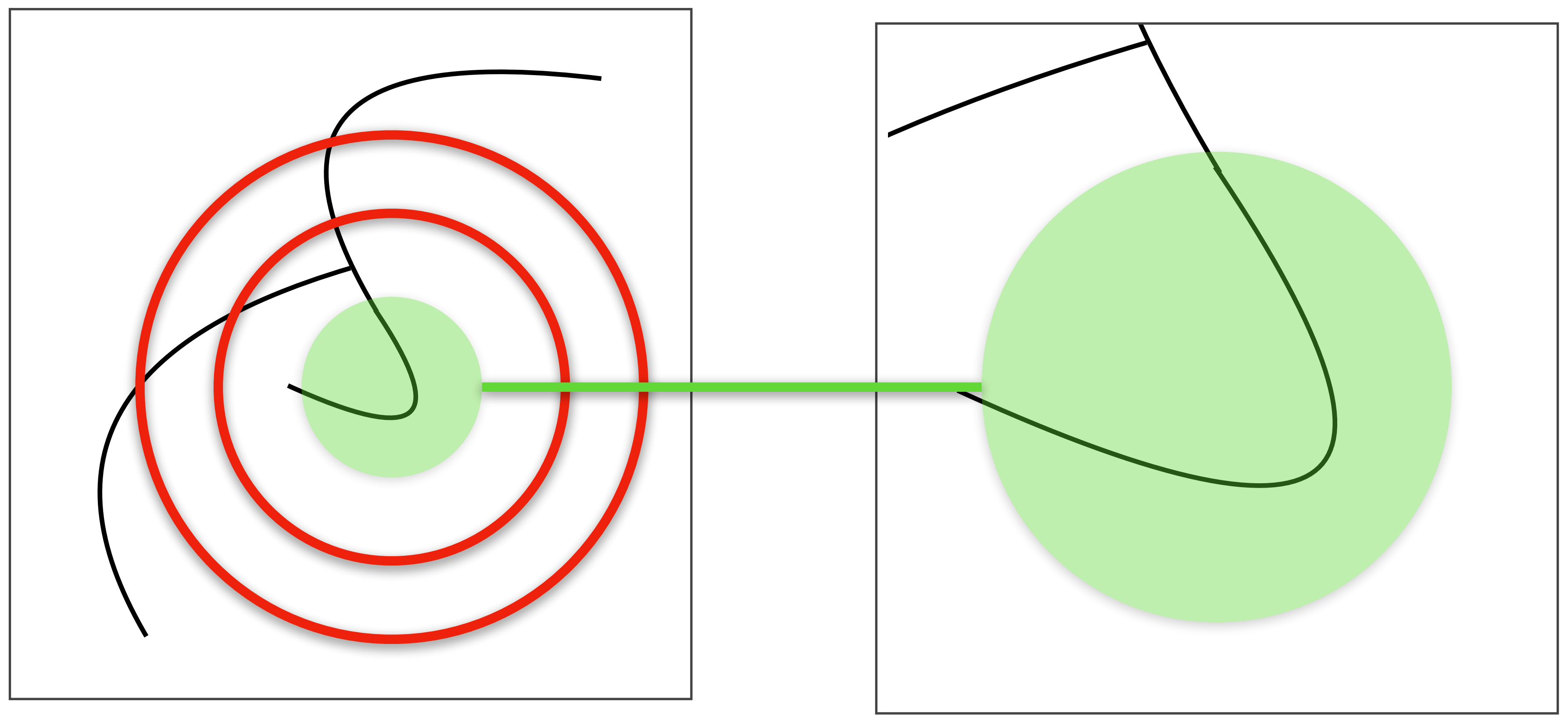
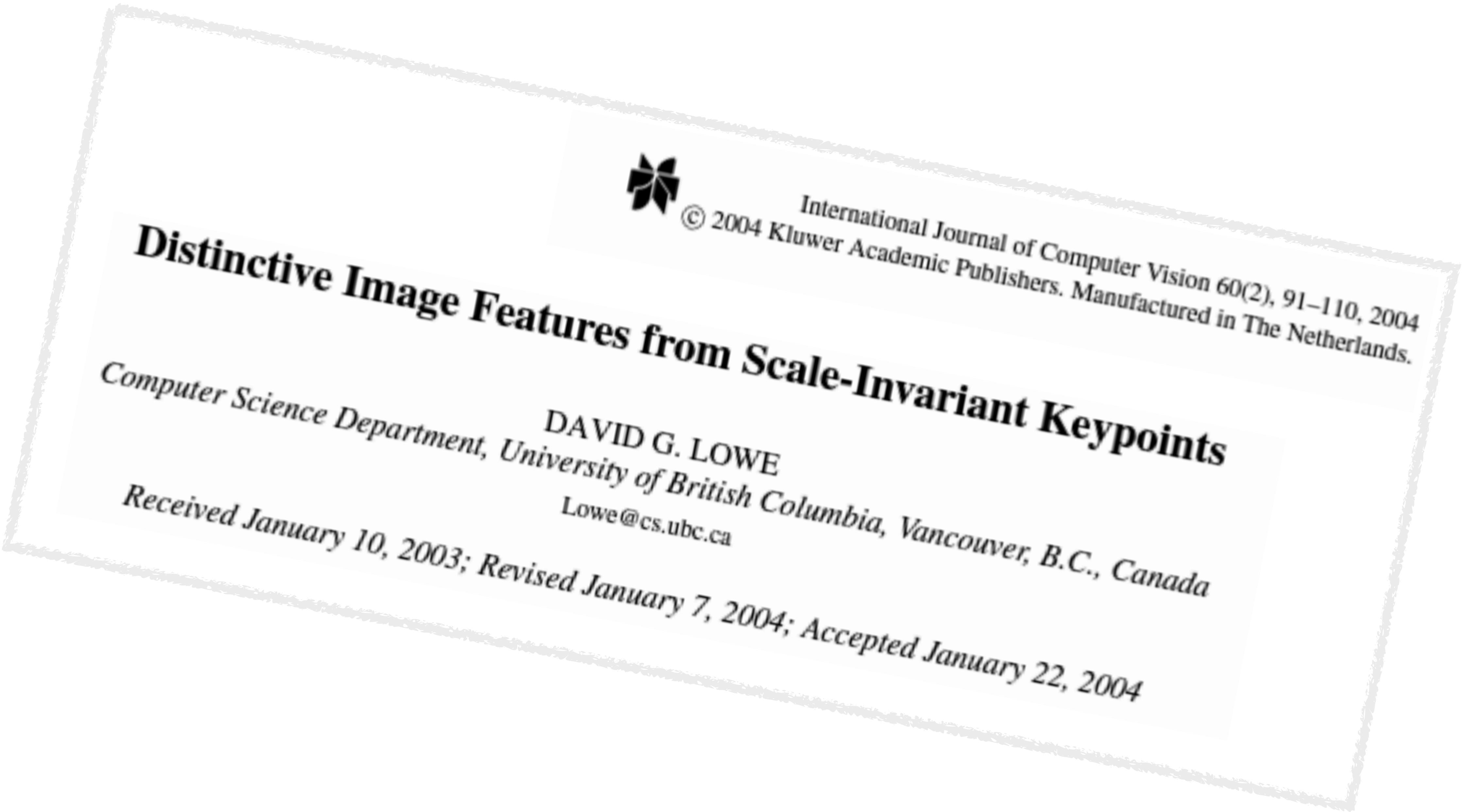


Image 1 Image 2



International Journal of Computer Vision 60(2), 91–110, 2004
© 2004 Kluwer Academic Publishers. Manufactured in The Netherlands.

Distinctive Image Features from Scale-Invariant Keypoints

DAVID G. LOWE
Computer Science Department, University of British Columbia, Vancouver, B.C., Canada
Lowe@cs.ubc.ca

Received January 10, 2003; Revised January 7, 2004; Accepted January 22, 2004

Scale Invariant Feature Transform (SIFT)

How to match the same content captured with tilt?

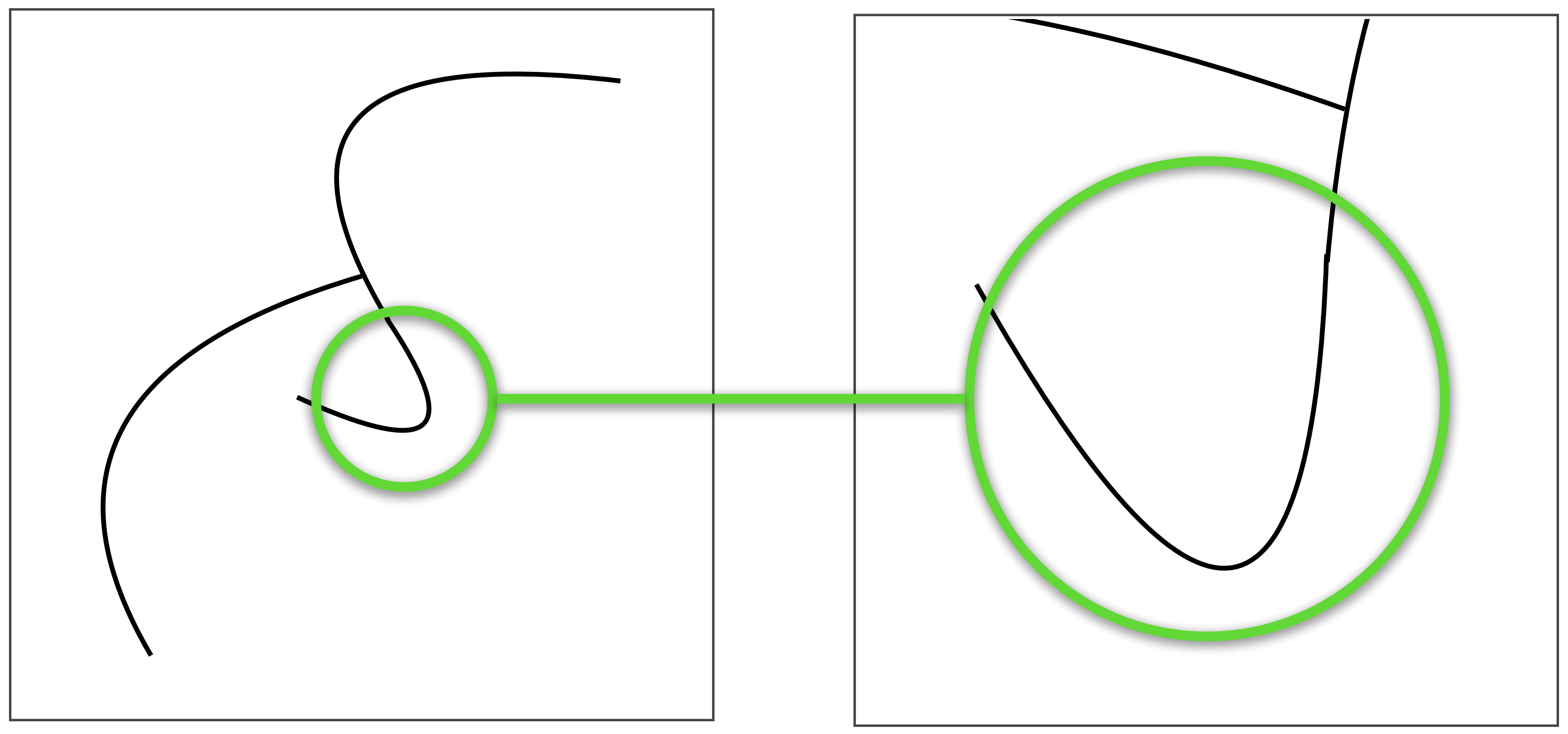
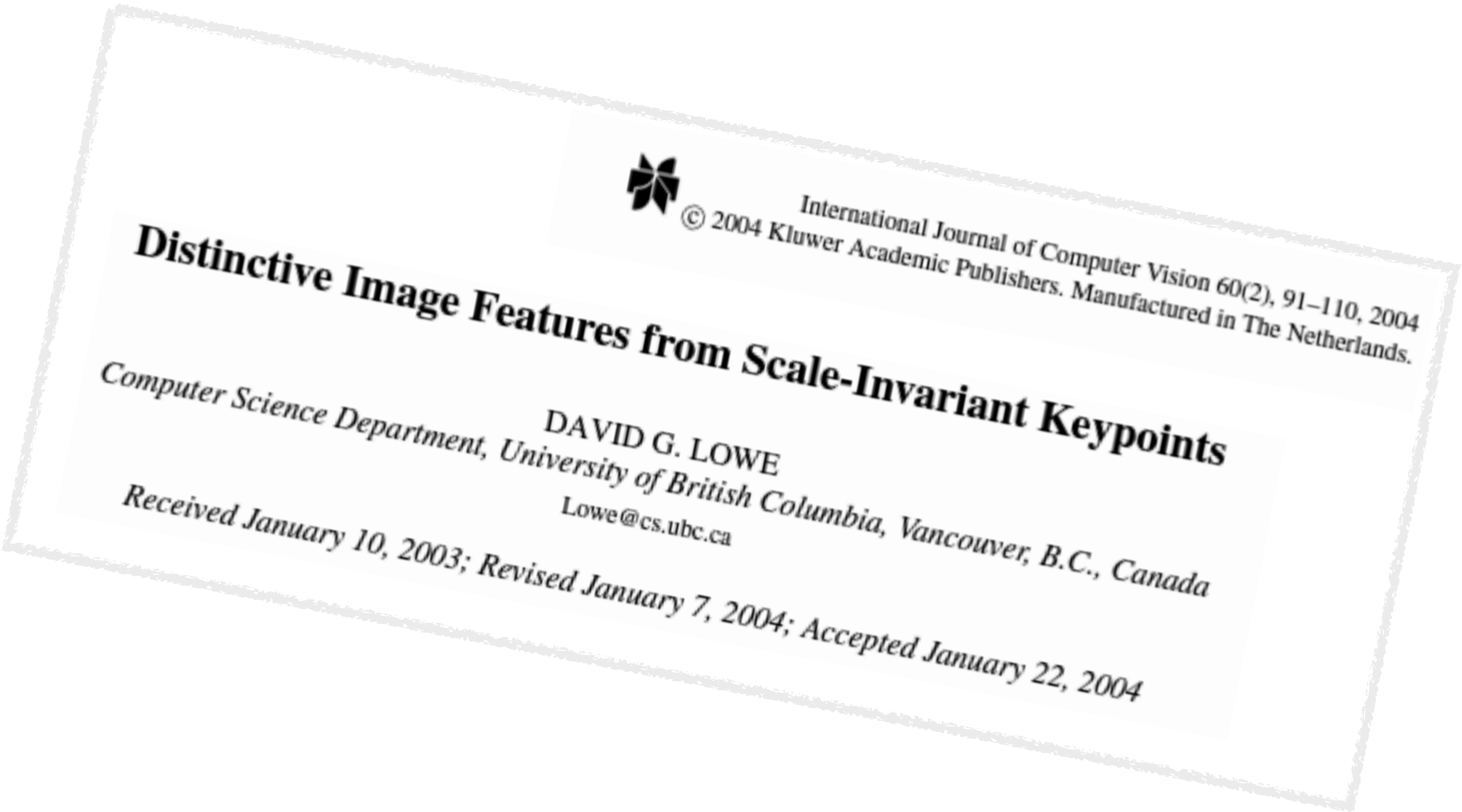


Image 1 Image 2



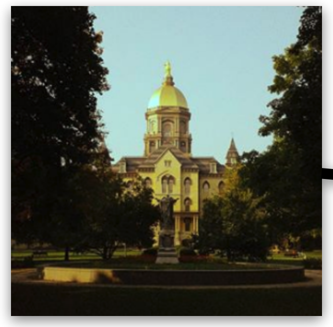
International Journal of Computer Vision 60(2), 91–110, 2004

Distinctive Image Features from Scale-Invariant Keypoints

DAVID G. LOWE
Computer Science Department, University of British Columbia, Vancouver, B.C., Canada
Lowe@cs.ubc.ca

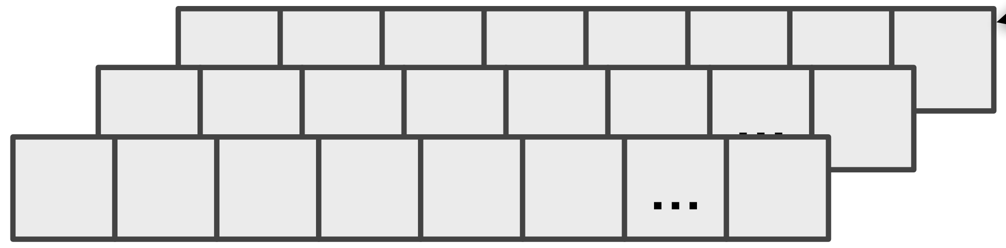
Received January 10, 2003; Revised January 7, 2004; Accepted January 22, 2004

Scale Invariant Feature Transform (SIFT)

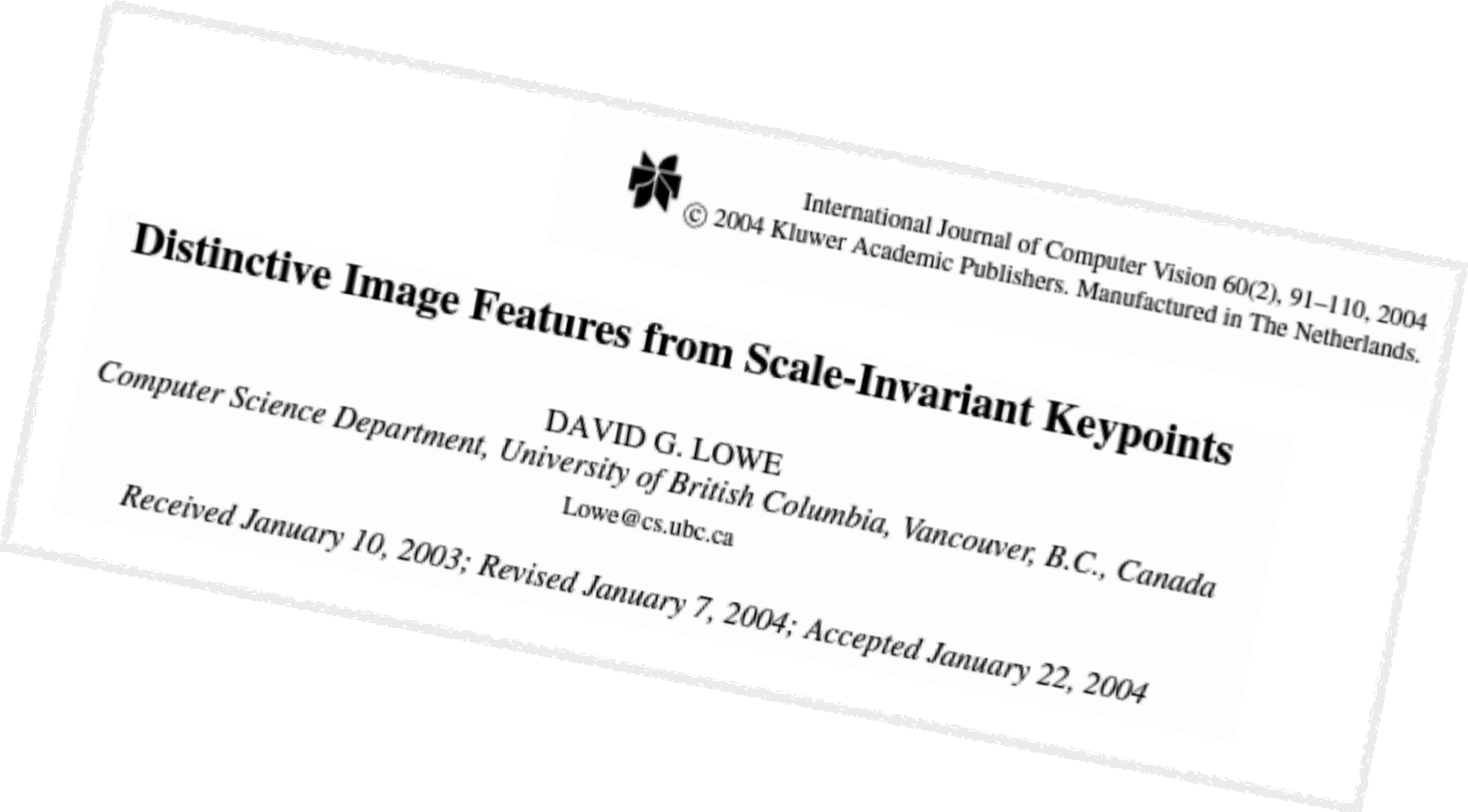


Keypoint Detection

Keypoint Description



128D feature vectors



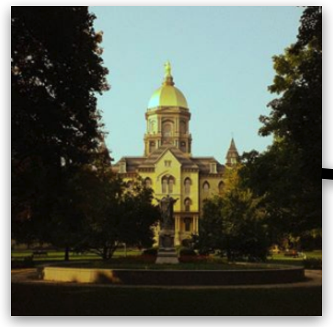
International Journal of Computer Vision 60(2), 91–110, 2004

Distinctive Image Features from Scale-Invariant Keypoints

DAVID G. LOWE
Computer Science Department, University of British Columbia, Vancouver, B.C., Canada
Lowe@cs.ubc.ca

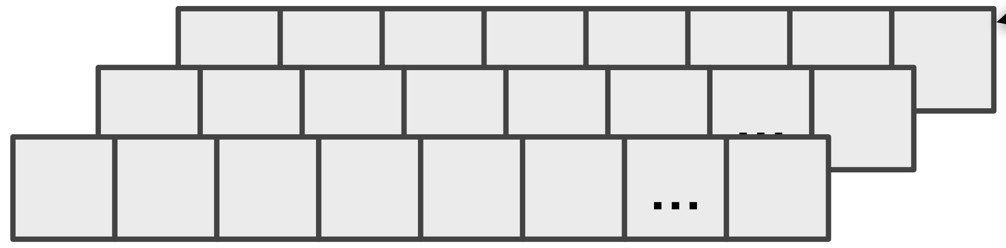
Received January 10, 2003; Revised January 7, 2004; Accepted January 22, 2004

Scale Invariant Feature Transform (SIFT)

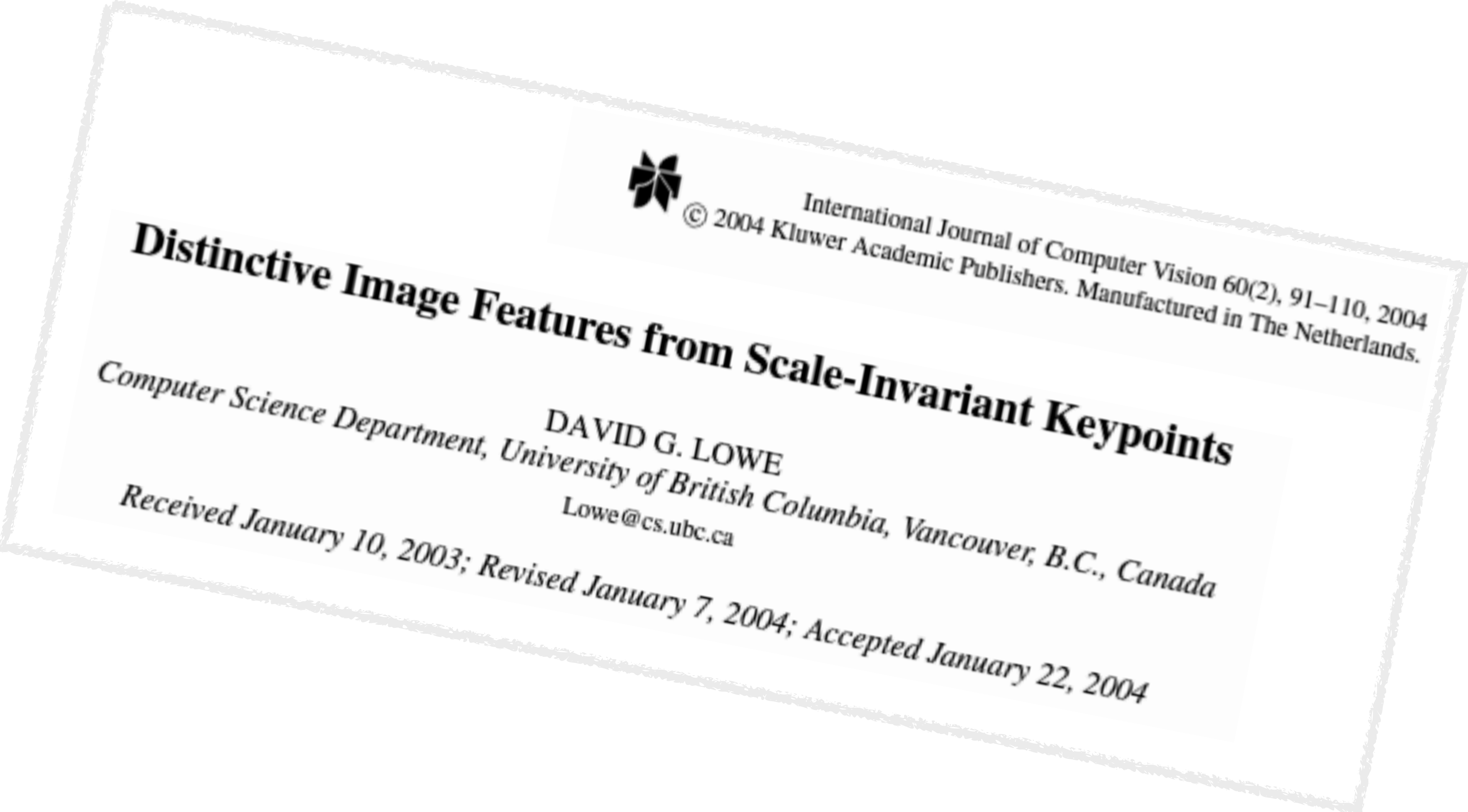


Keypoint Detection

Keypoint Description



128D feature vectors


International Journal of Computer Vision 60(2), 91–110, 2004
© 2004 Kluwer Academic Publishers. Manufactured in The Netherlands.

Distinctive Image Features from Scale-Invariant Keypoints

DAVID G. LOWE
Computer Science Department, University of British Columbia, Vancouver, B.C., Canada
Lowe@cs.ubc.ca

Received January 10, 2003; Revised January 7, 2004; Accepted January 22, 2004

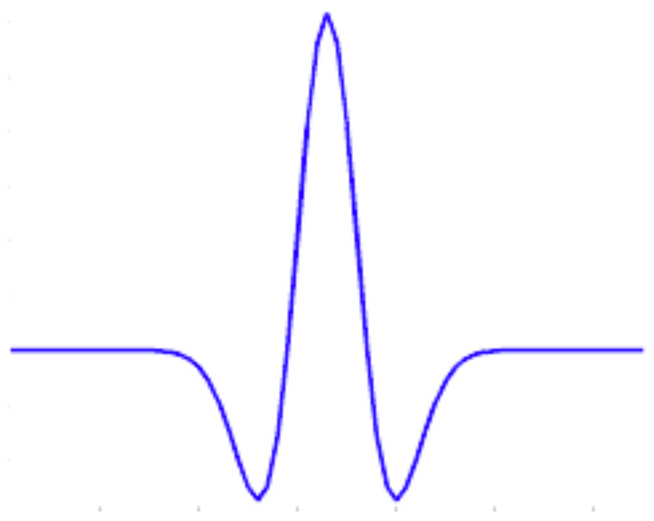
LOYOLA UNIVERSITY CHICAGO

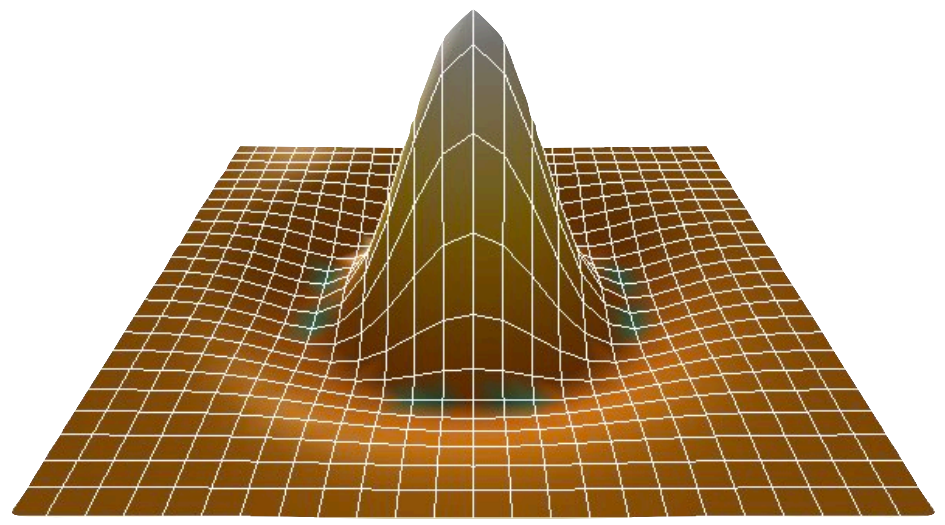
Scale Invariant Feature Transform (SIFT)

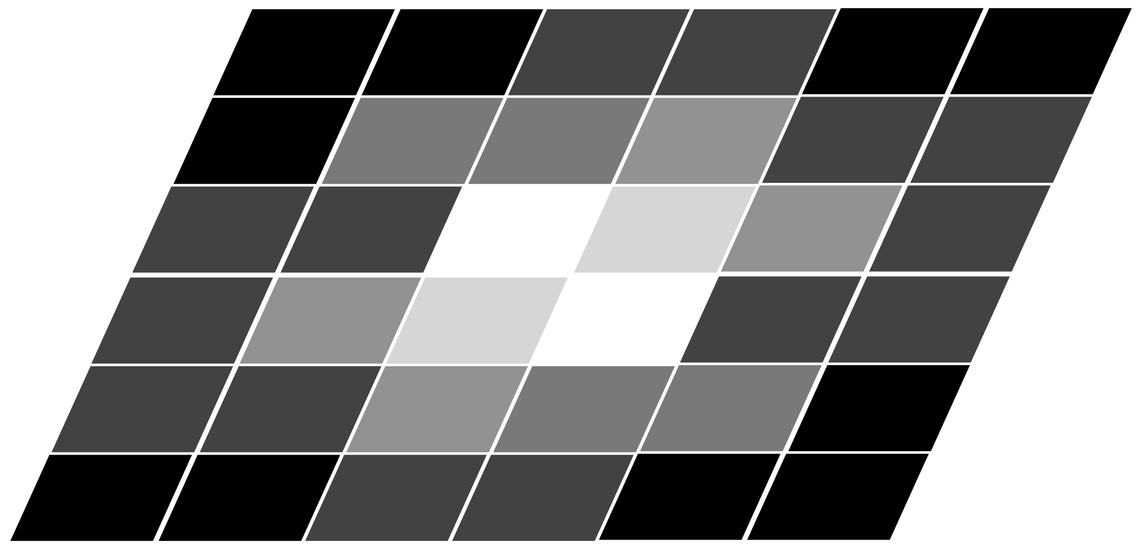
Keypoint Detection
How to detect interest points (a.k.a. keypoints) in an image?

Previous Literature
To focus on blobs:
1. Apply Gaussian to remove noise (blur).
2. Apply Laplacian to detect good regions.

Good regions will have high values after convolution.
This is known as Laplacian of Gaussian (LoG).





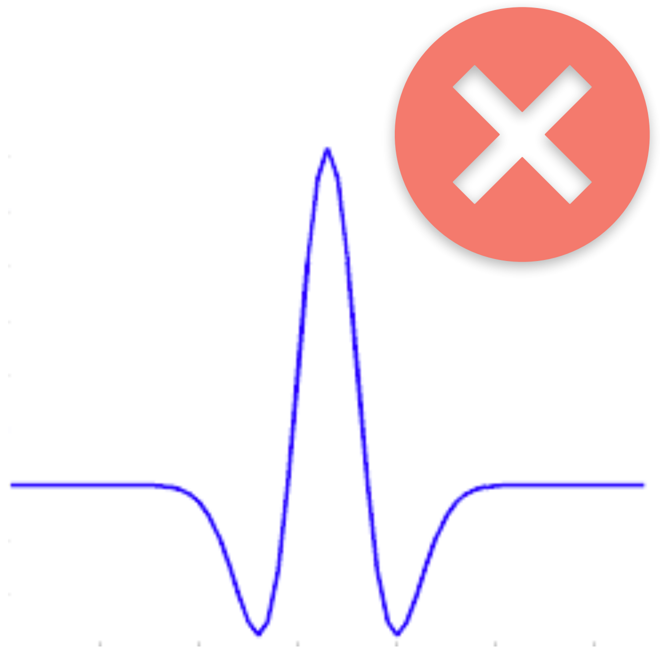


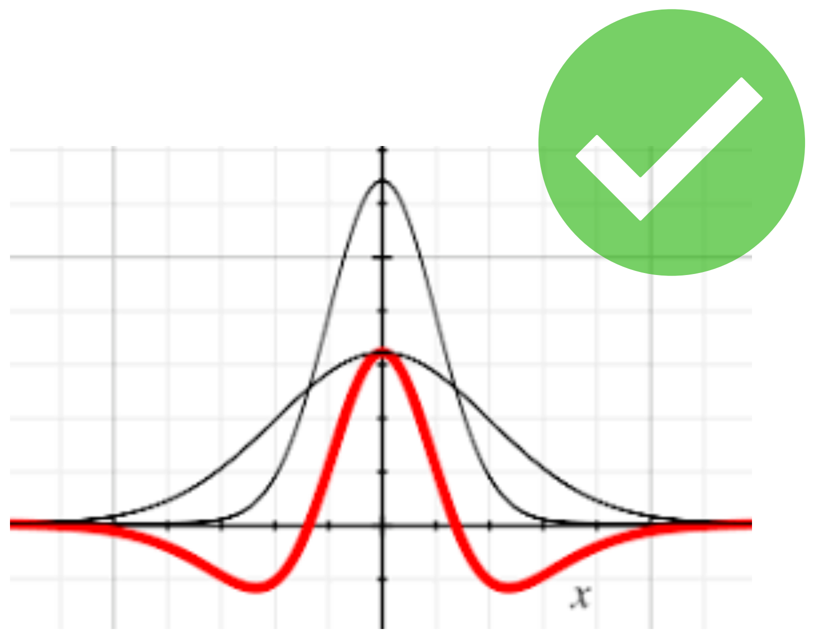
LOYOLA UNIVERSITY CHICAGO

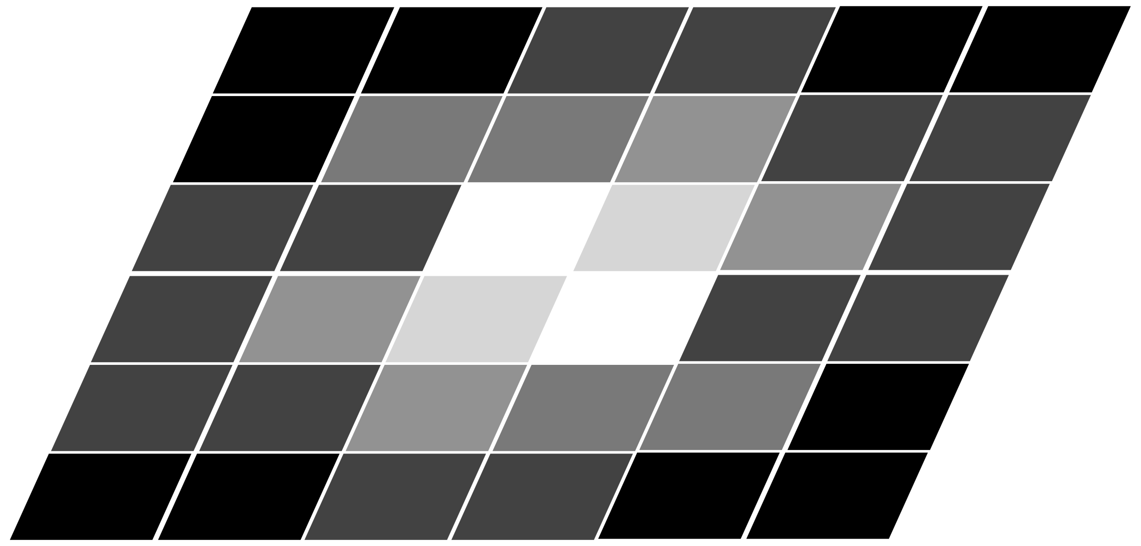
Scale Invariant Feature Transform (SIFT)

Keypoint Detection
How to detect interest points (a.k.a. keypoints) in an image?

Lowe's ideas
Approximate LoG by a Difference of Gaussians (DoG).







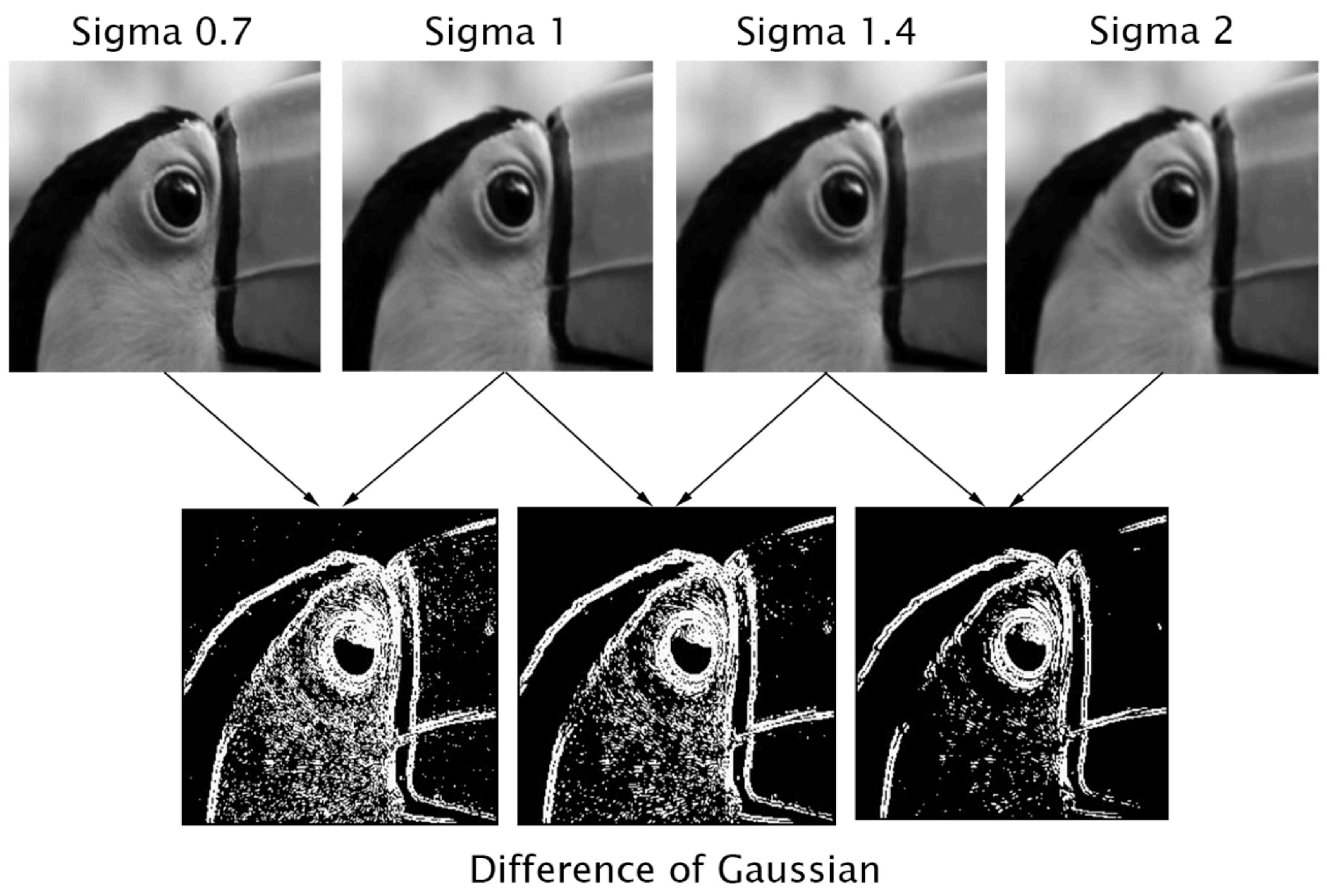
Scale Invariant Feature Transform (SIFT)

https://medium.com/@vad710/cv-for-busy-devs-improving-features-df20c3aa5887

Keypoint Detection
How to detect interest points (a.k.a. keypoints) in an image?

Lowe's ideas
Approximate LoG by a Difference of Gaussians (DoG).

The two subtracted Gaussians come from distinct variances (scales).



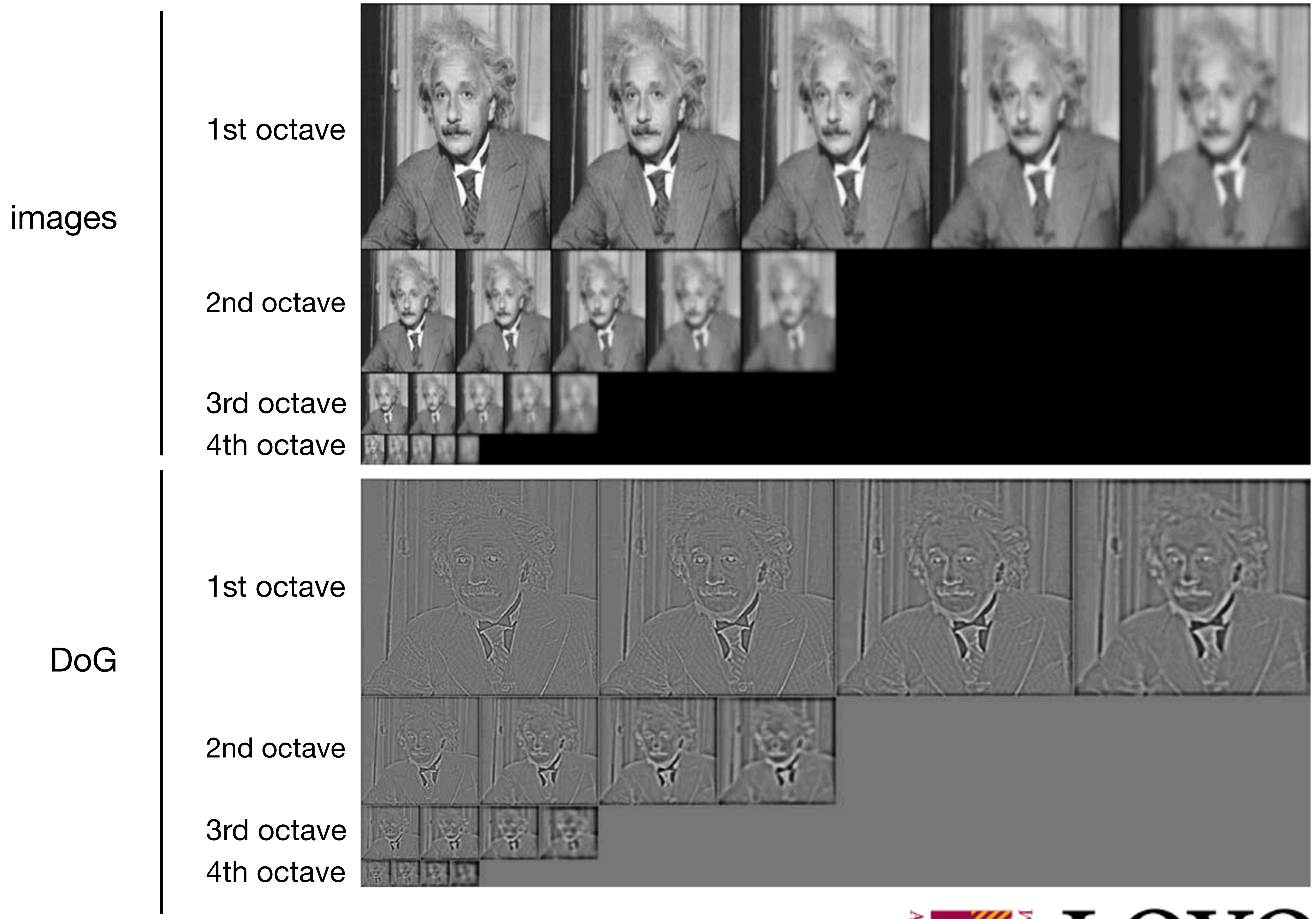
Scale Invariant Feature Transform (SIFT)

https://faculty.cc.gatech.edu/~afb/classes/CS4495-Fall2013/slides/CS4495-11-Features2.pdf

Keypoint Detection
Scale Invariance

Process the image of interest
at different scales.

This is called resolution pyramid.

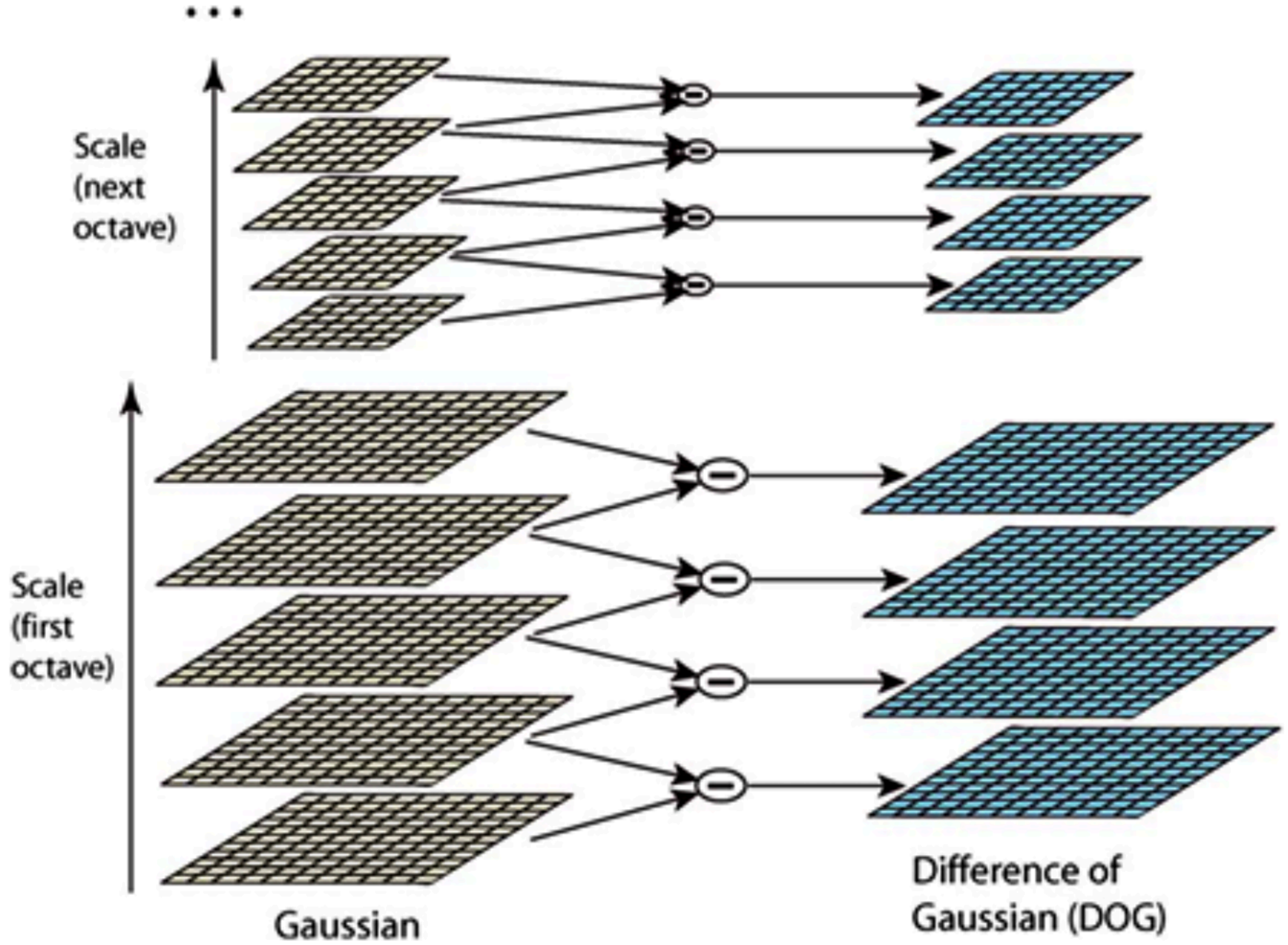


Scale Invariant Feature Transform (SIFT)

Keypoint Detection
Scale Invariance

Process the image of interest
at different scales.

Lowe's

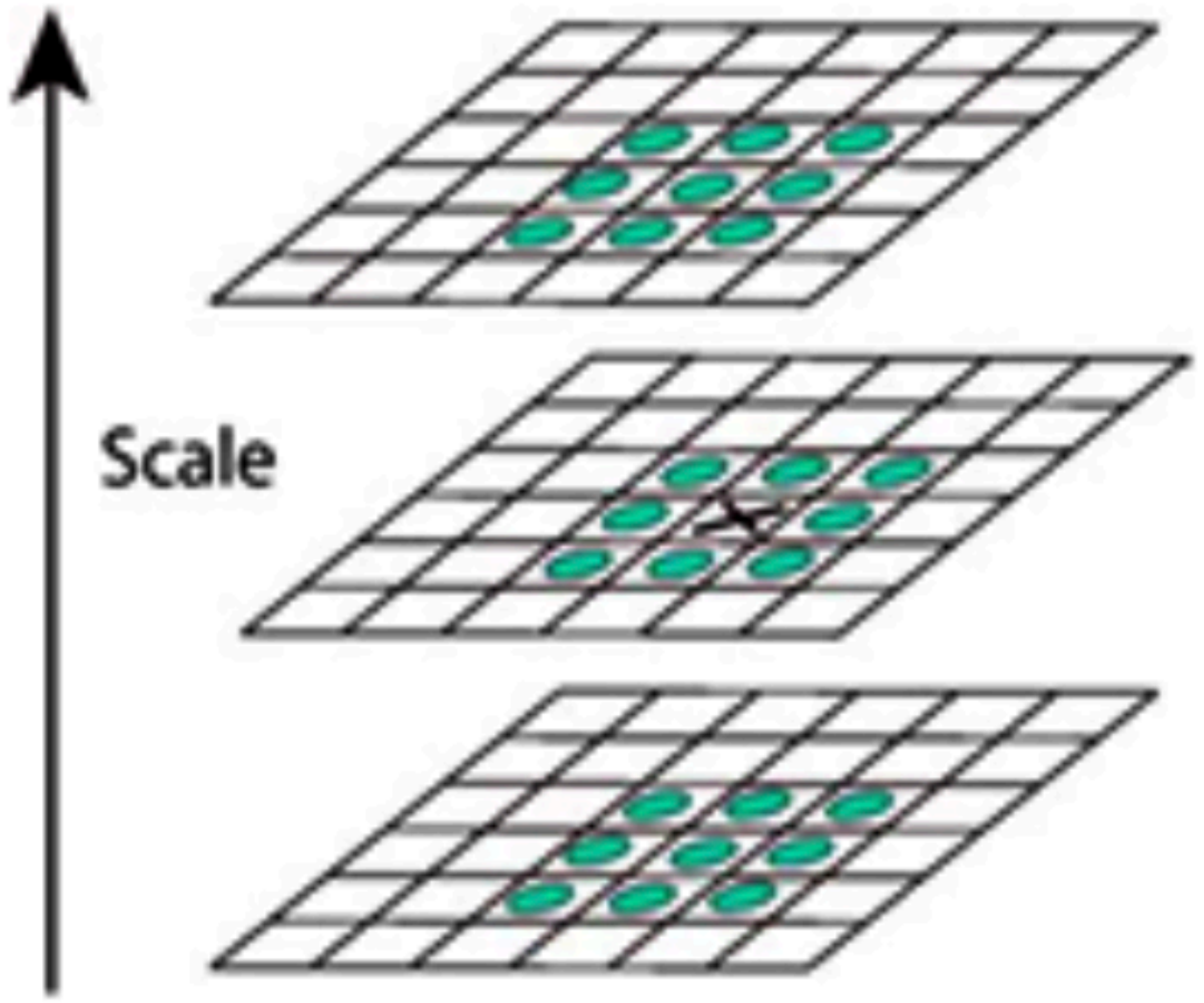


Scale Invariant Feature Transform (SIFT)

Lowe's

Keypoint Detection
Non-maximal Suppression

Good SIFT keypoints present the highest DoG value among their immediate (3 x 3 x 3) scale-space neighborhood.



LOYOLA UNIVERSITY CHICAGO

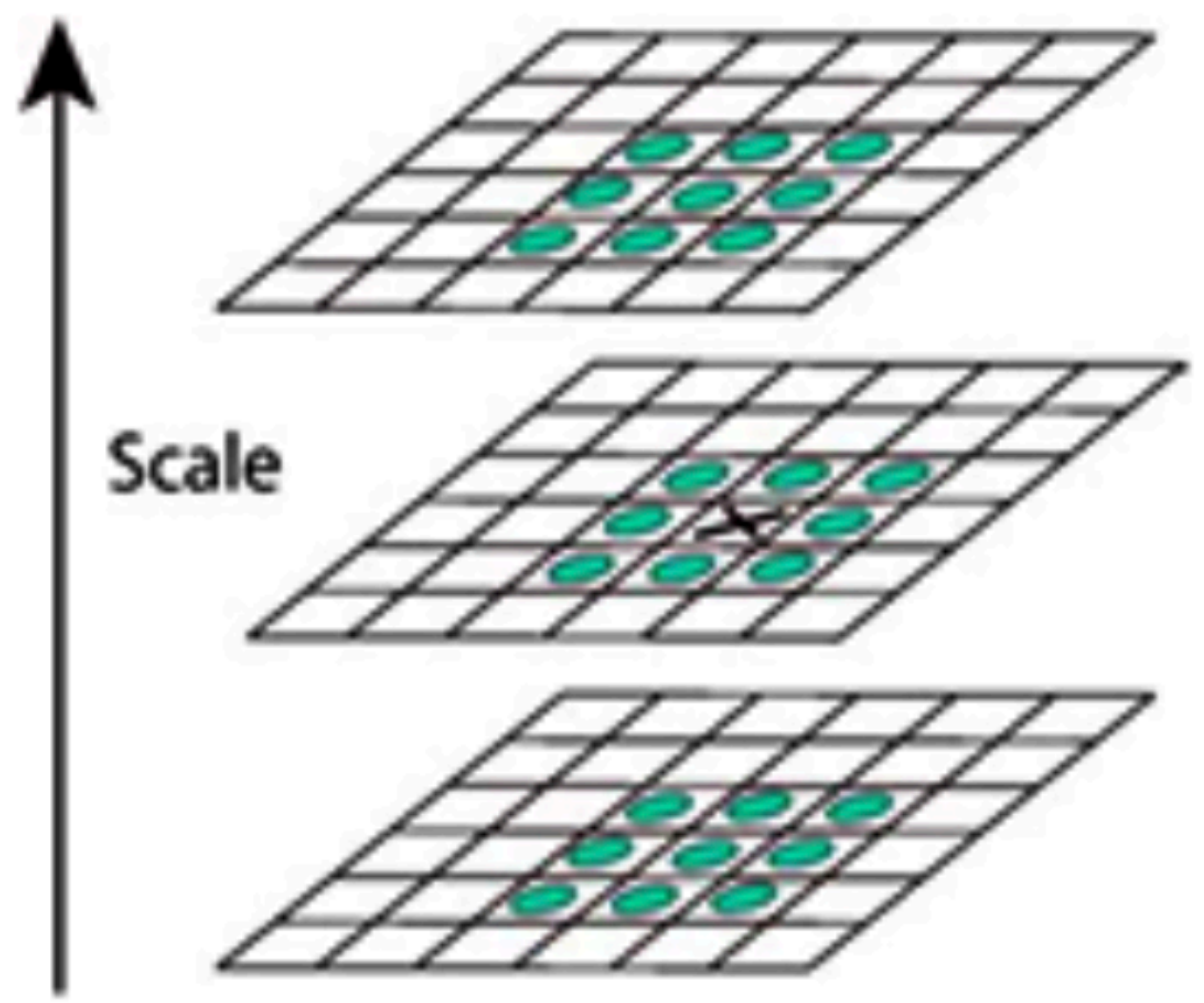
Scale Invariant Feature Transform (SIFT)

Lowe's

Keypoint Detection
Non-maximal Suppression

Good SIFT keypoints present the highest DoG value among their immediate (3 x 3 x 3) scale-space neighborhood.

Local scale-space extrema have $I(x, y)$ position and inherit the scale from the level/octave it belongs to within the resolution pyramid.



LOYOLA UNIVERSITY CHICAGO

Scale Invariant Feature Transform (SIFT)

Keypoint Detection
Orientation Assignment

To become robust to tilt (rotation), compute the gradient angle for all the pixels within the keypoint neighborhood considering its scale.

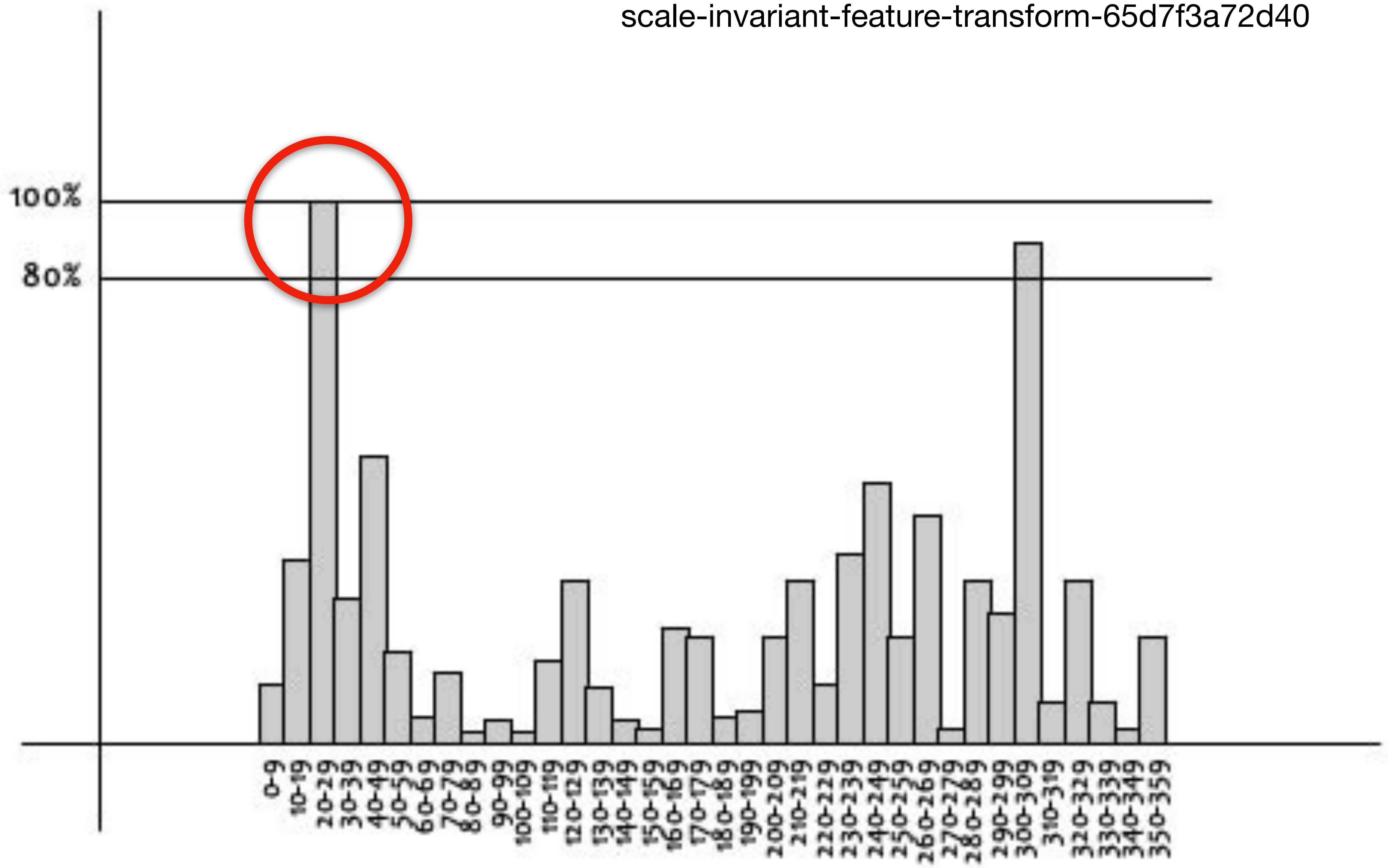
Scale Invariant Feature Transform (SIFT)

https://medium.com/data-breach/introduction-to-sift-scale-invariant-feature-transform-65d7f3a72d40

Keypoint Detection
Orientation Assignment

To become robust to tilt (rotation), compute the gradient angle for all the pixels within the keypoint neighborhood considering its scale.

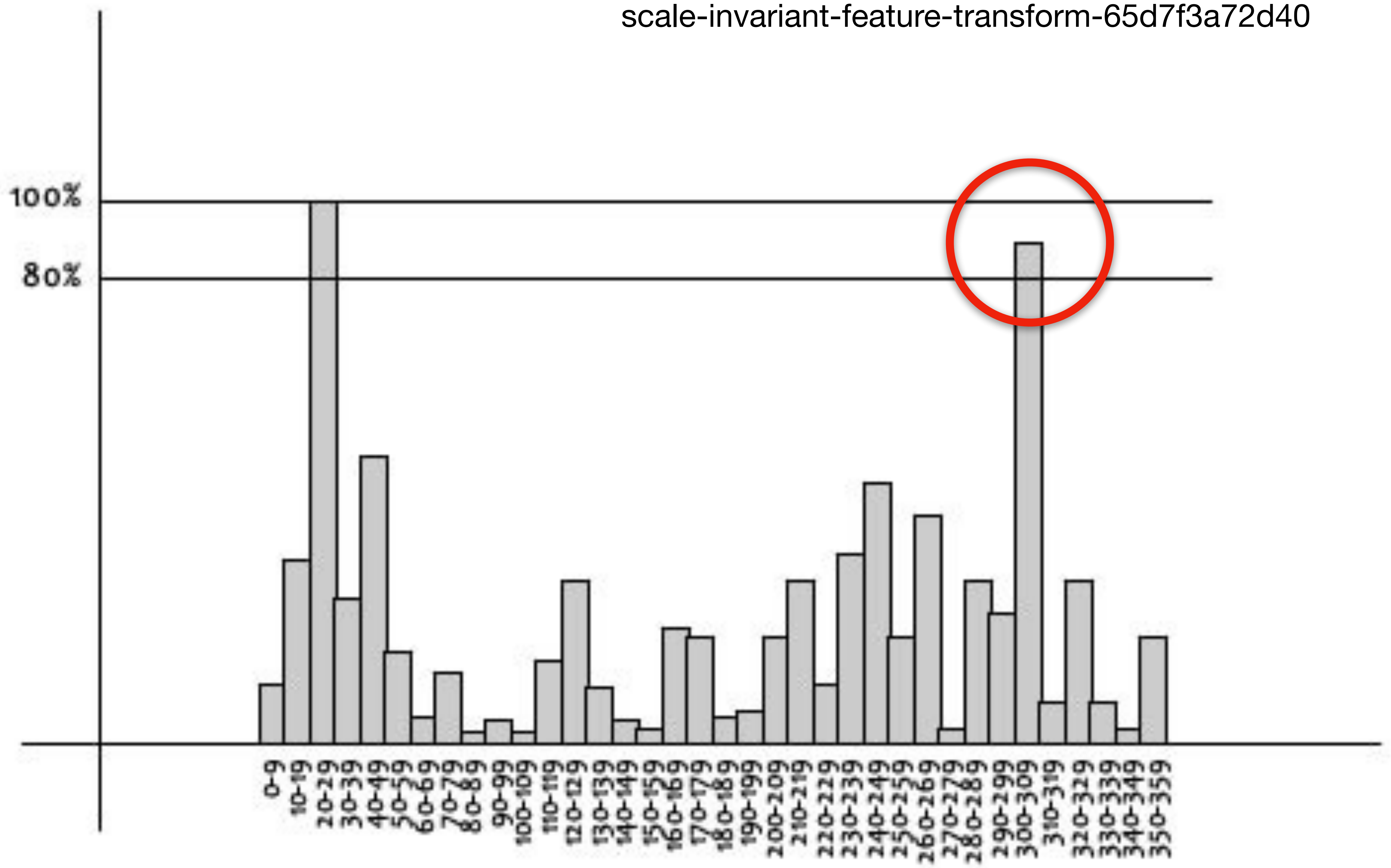
Create an angle histogram with 36 bins. Take the dominant angle as the keypoint orientation.



LOYOLA UNIVERSITY CHICAGO

Scale Invariant Feature Transform (SIFT)

https://medium.com/data-breach/introduction-to-sift-scale-invariant-feature-transform-65d7f3a72d40



Keypoint Detection
Orientation Assignment

To become robust to tilt (rotation), compute the gradient angle for all the pixels within the keypoint neighborhood considering its scale.

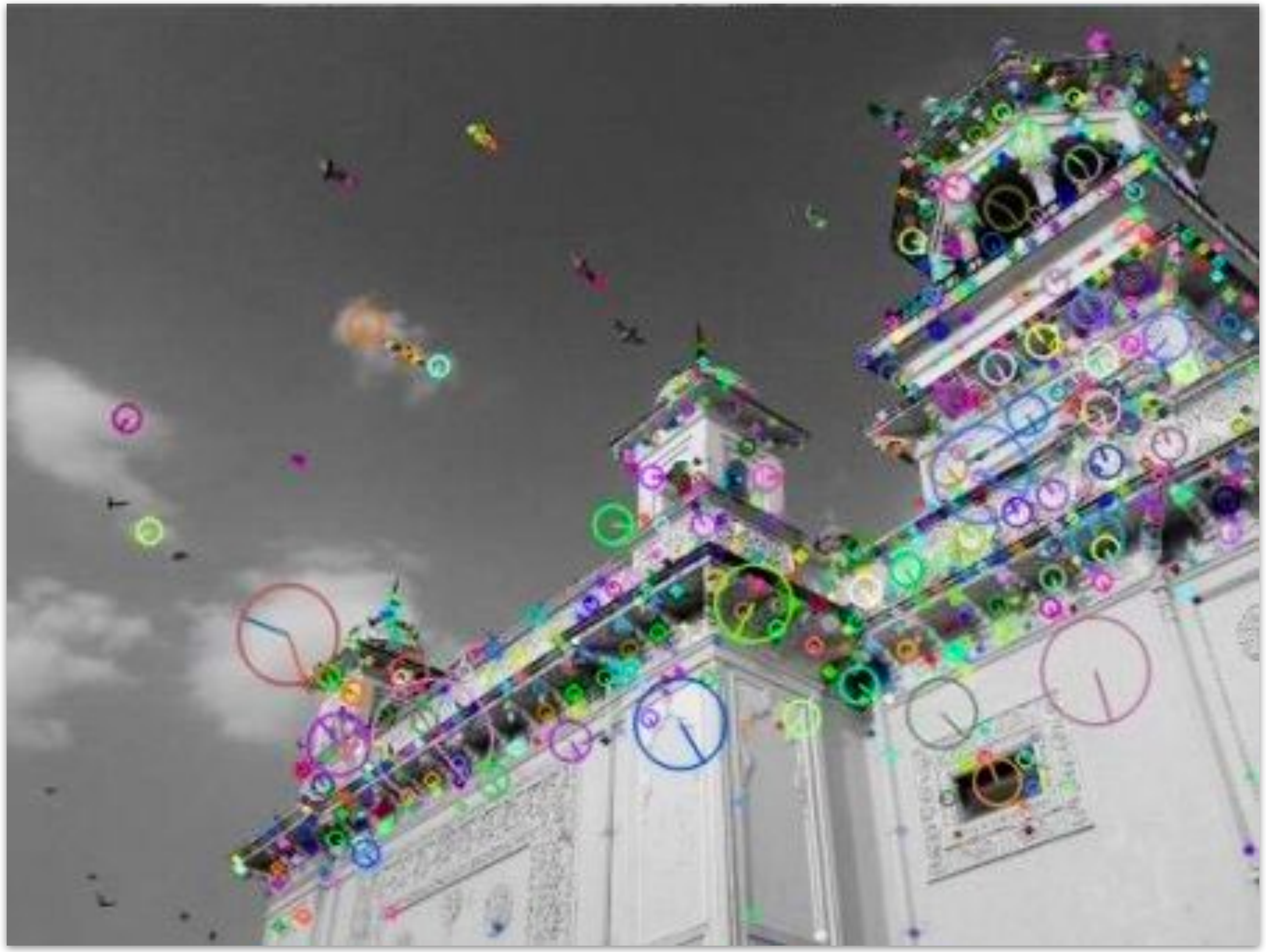
If other angles have at least 80% of the frequency of the maximum angle, create other keypoints with the same location and scale but different orientation.

LOYOLA UNIVERSITY CHICAGO

Scale Invariant Feature Transform (SIFT)

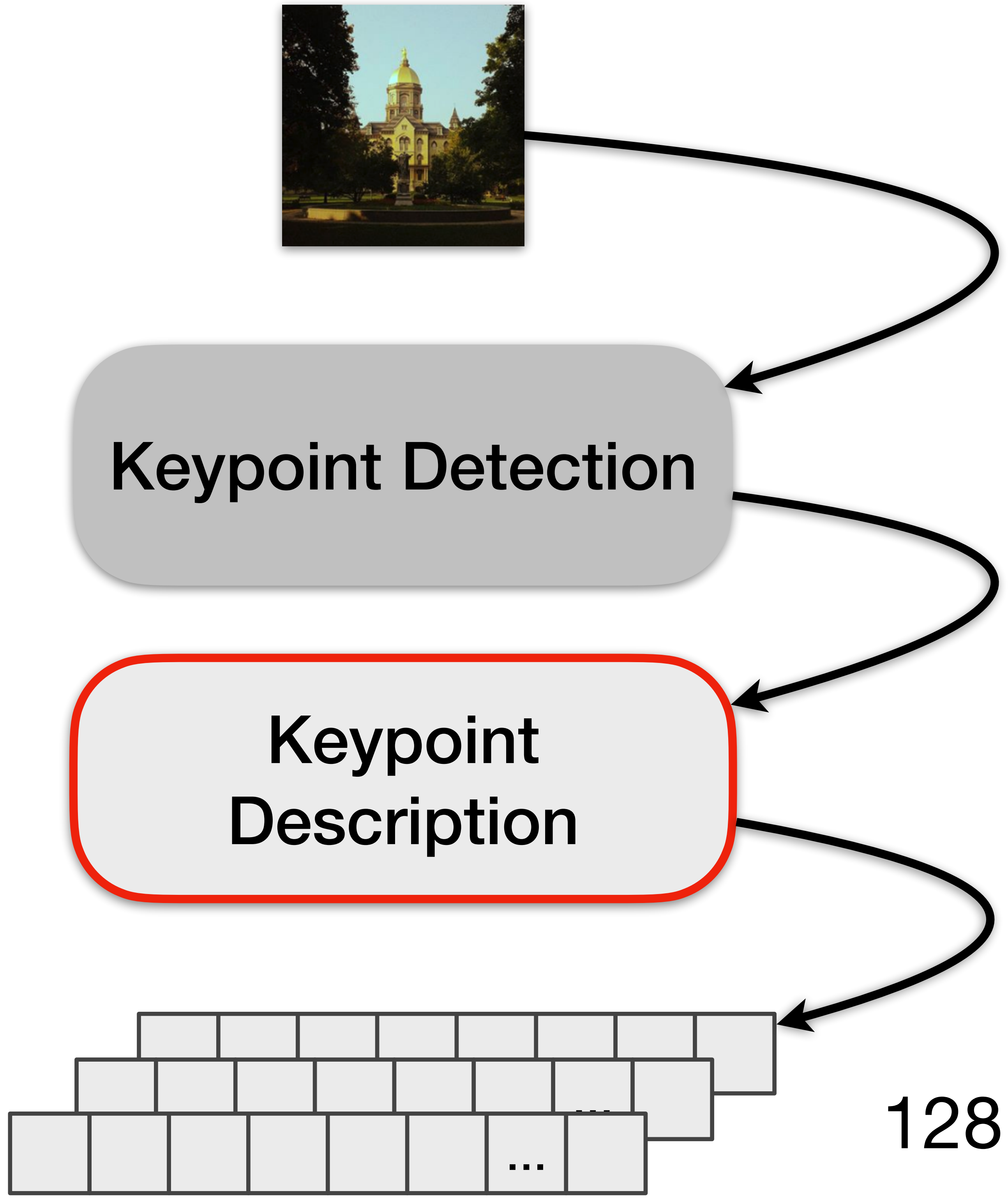
Keypoint Detection

SIFT keypoints have:
1. Location *(x,y)*
2. Scale (from resolution pyramid)
3. Orientation (dominant local gradient angle)
4. Strength (DoG value)

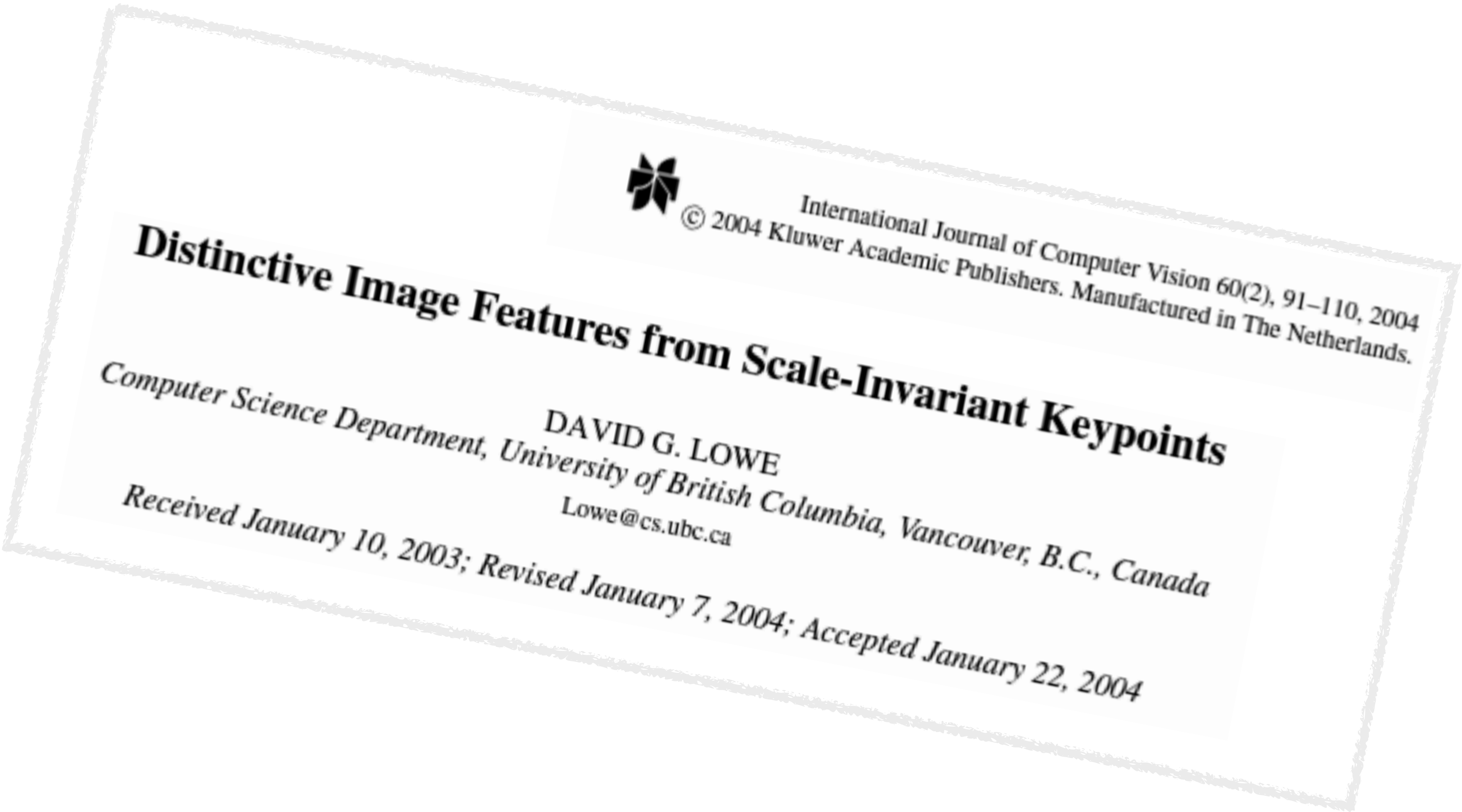


https://docs.opencv.org/4.x/da/df5/tutorial_py_sift_intro.html

Scale Invariant Feature Transform (SIFT)



128D feature vectors


International Journal of Computer Vision 60(2), 91–110, 2004

Distinctive Image Features from Scale-Invariant Keypoints

DAVID G. LOWE
Computer Science Department, University of British Columbia, Vancouver, B.C., Canada
Lowe@cs.ubc.ca

Received January 10, 2003; Revised January 7, 2004; Accepted January 22, 2004

Scale Invariant Feature Transform (SIFT)

Keypoint Description

For each keypoint, rotate the image according to the keypoint orientation.

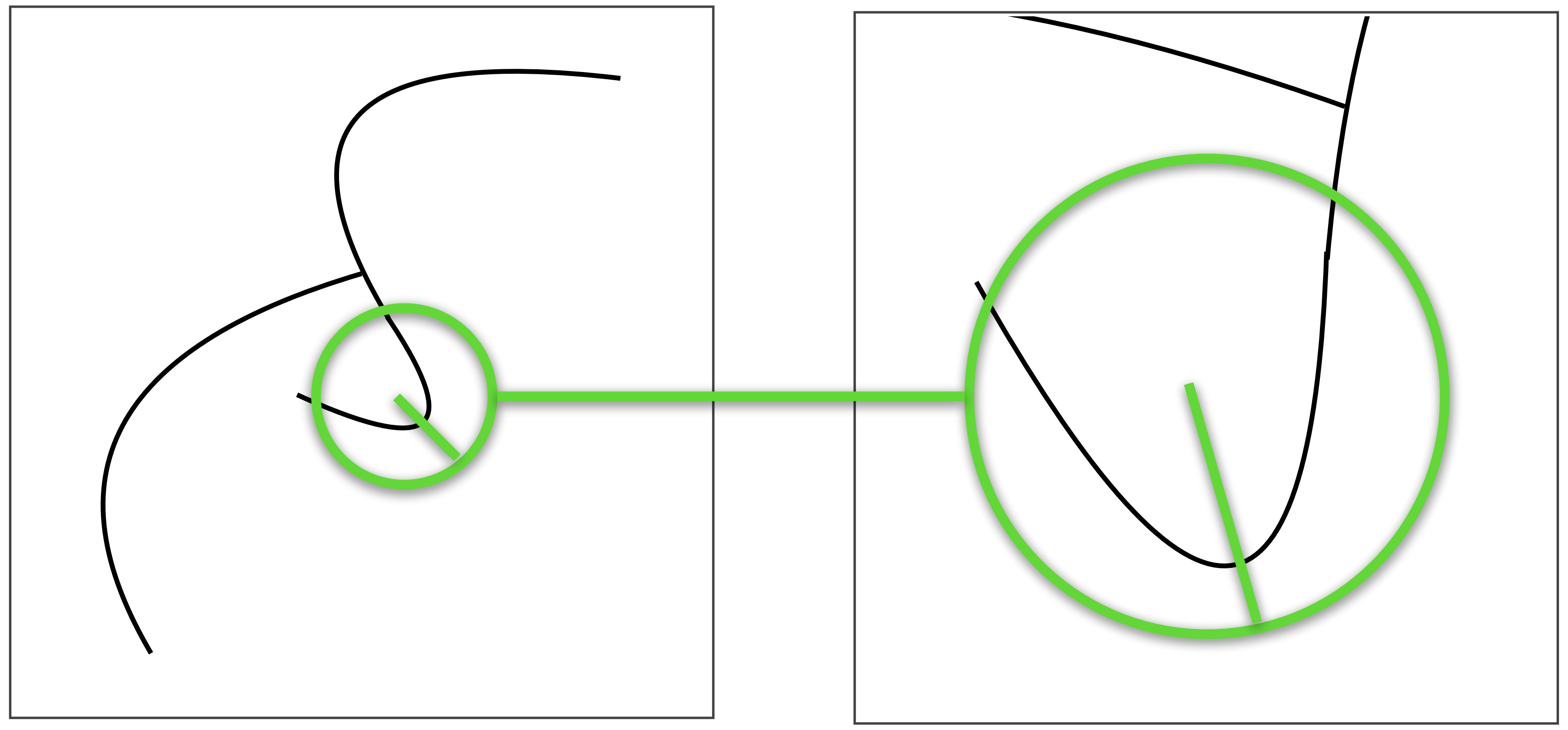
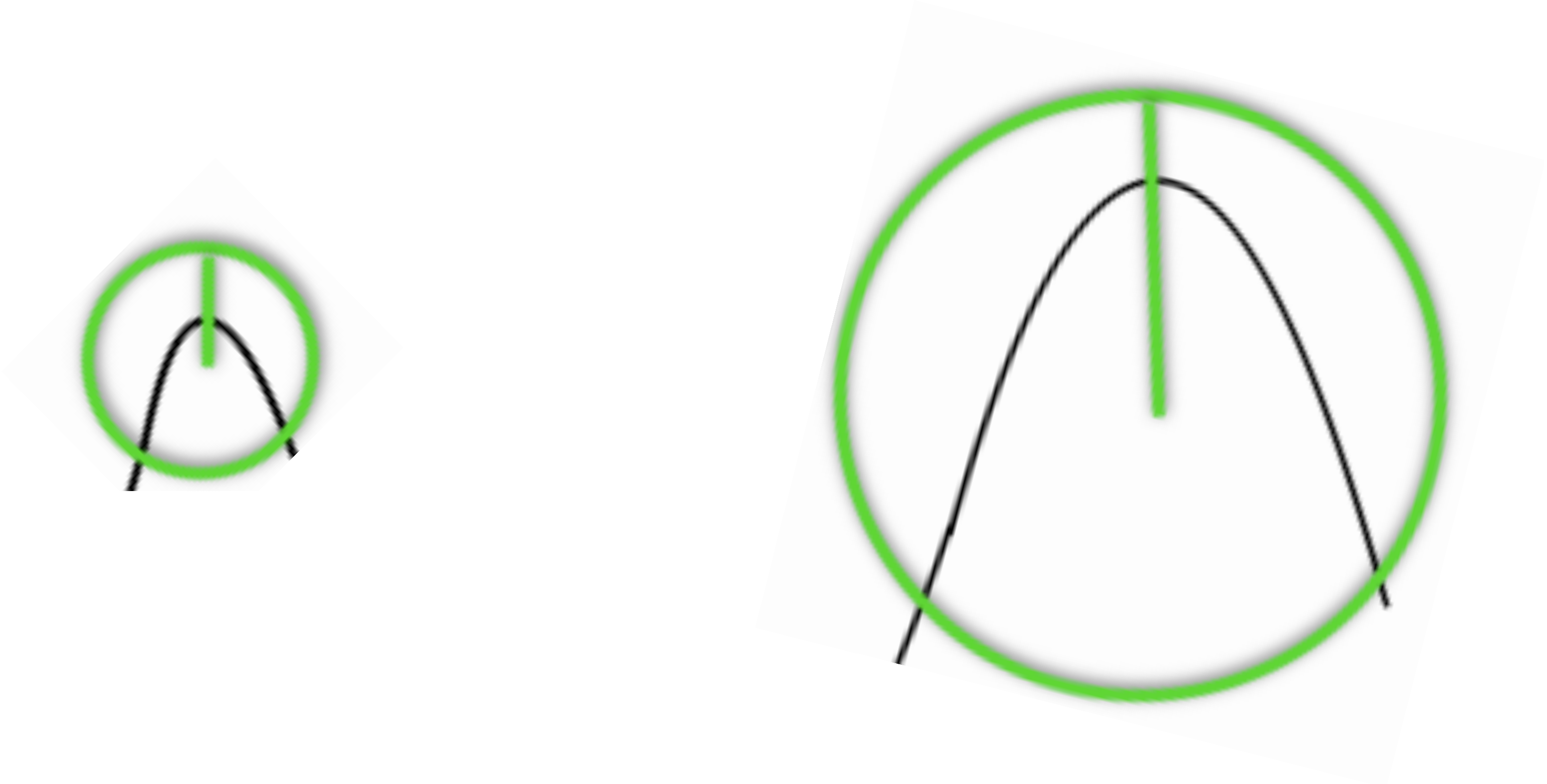


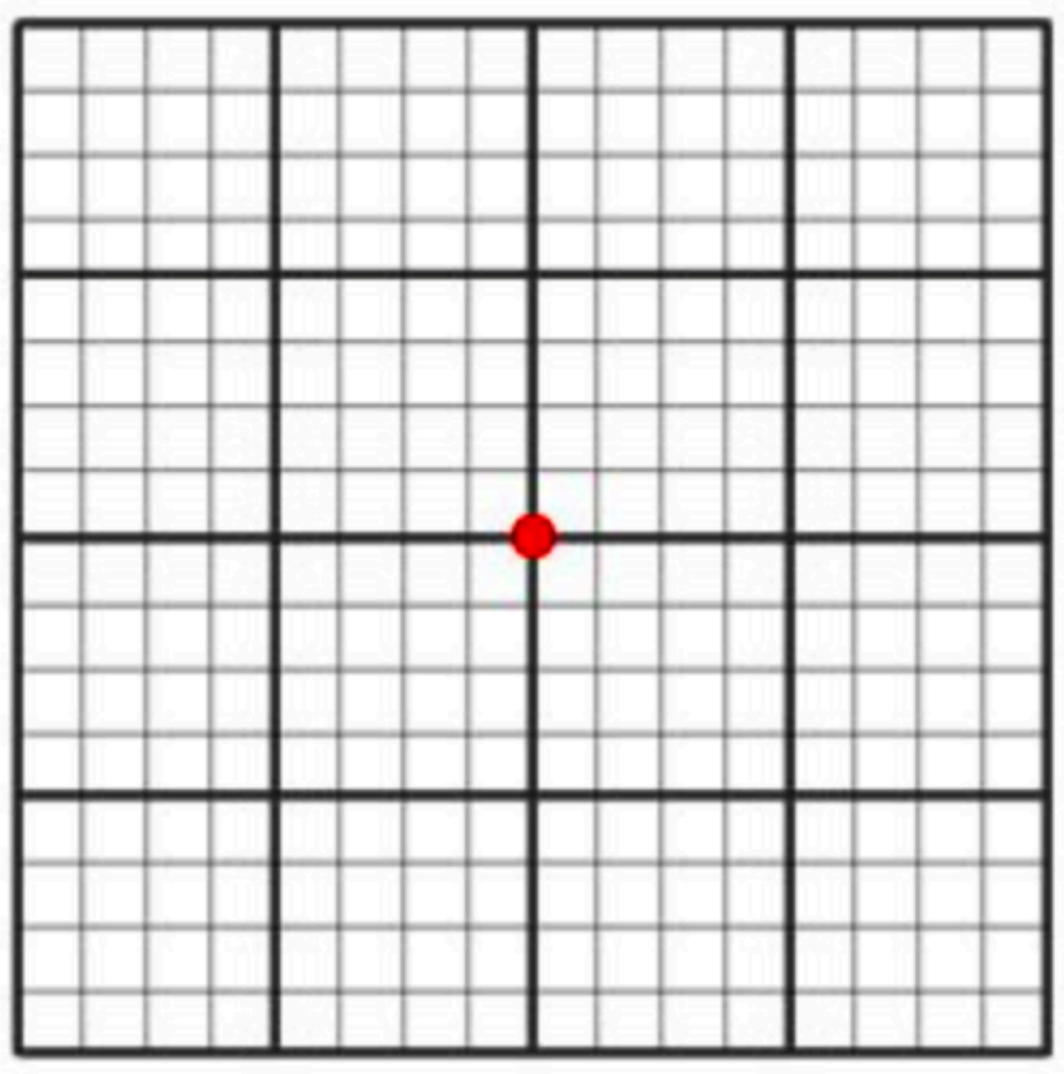
Image 1 Image 2



Scale Invariant Feature Transform (SIFT)

Keypoint Description

For each rotated keypoint, sample a *4 x 4* window on its neighborhood, according to the keypoint scale.

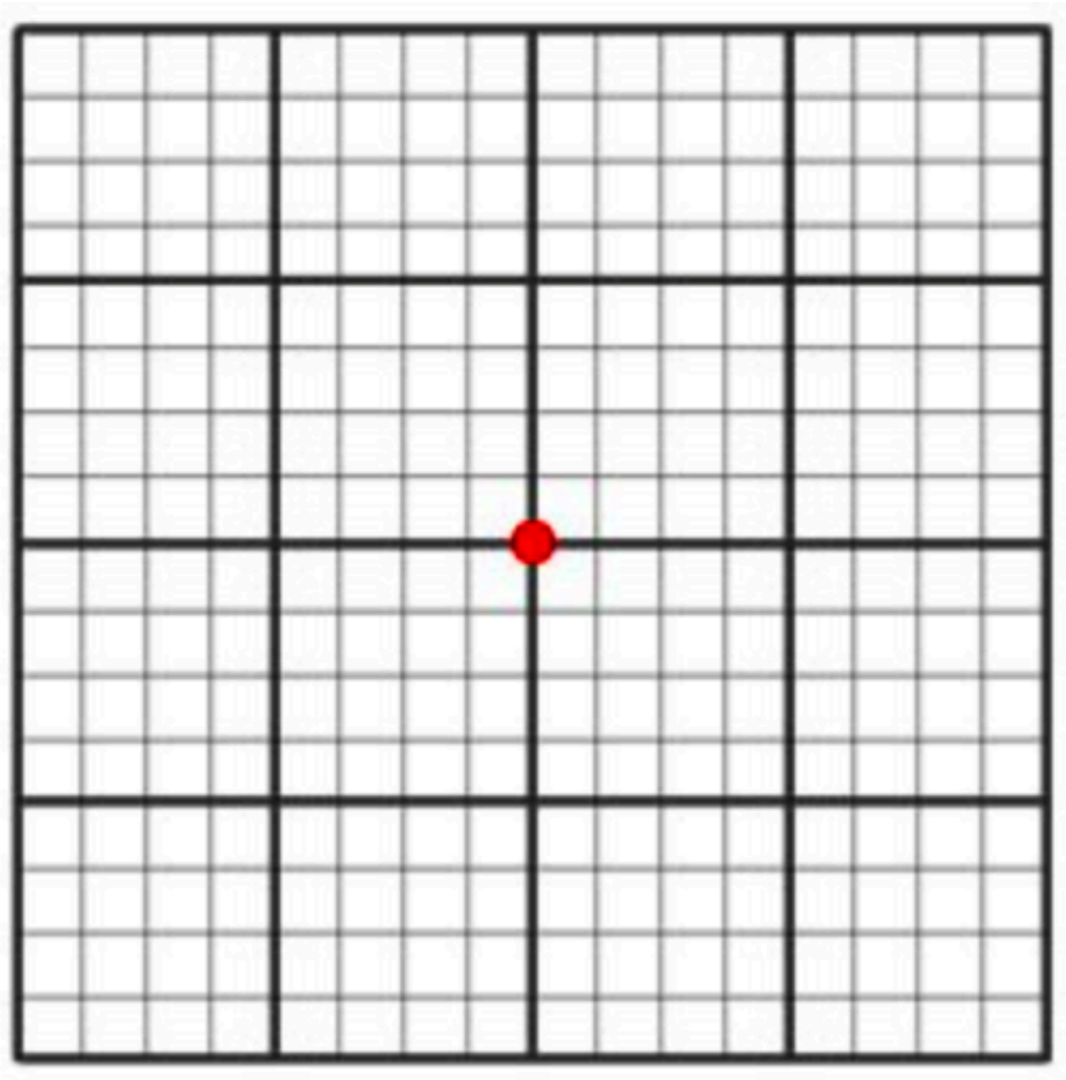


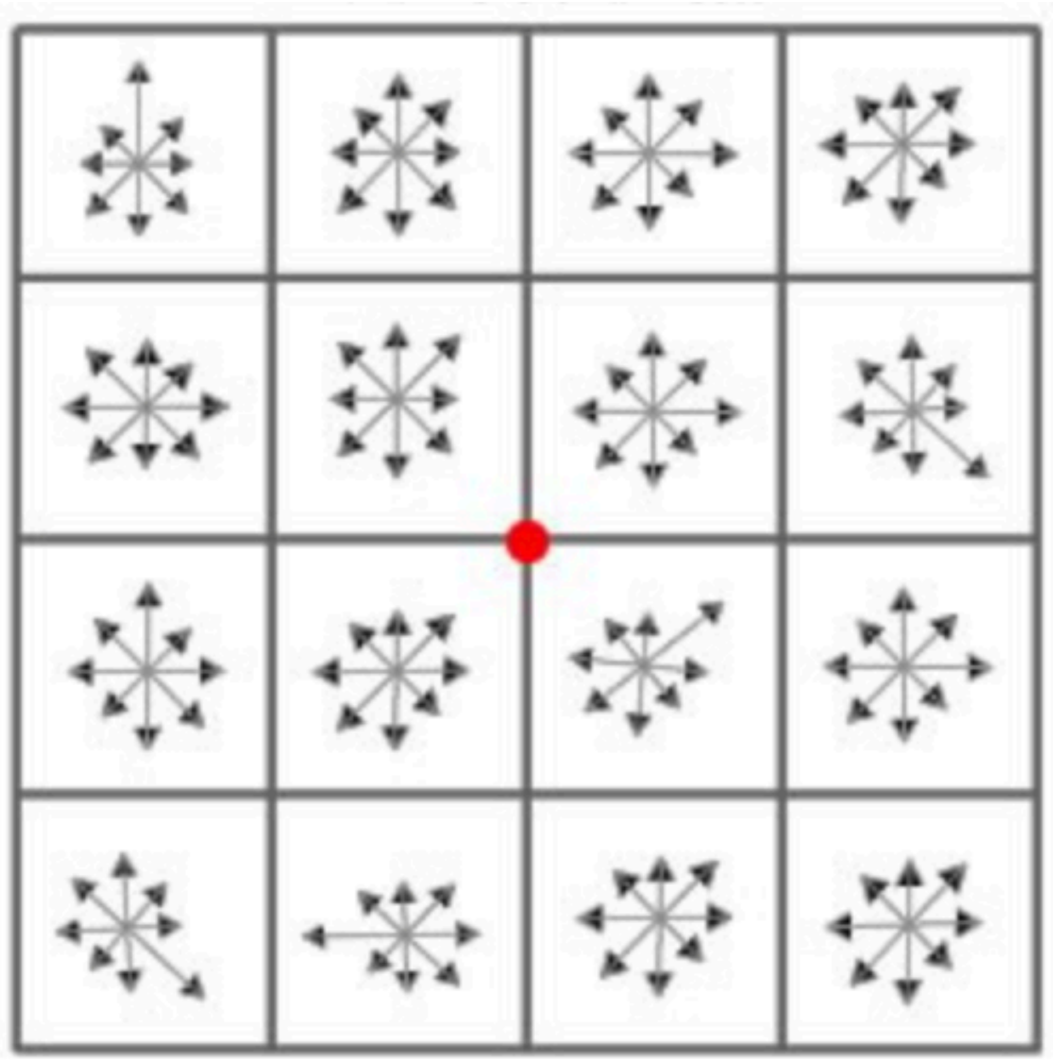
LOYOLA UNIVERSITY CHICAGO

Scale Invariant Feature Transform (SIFT)

Keypoint Description

For each rotated keypoint, sample a *4 x 4* window on its neighborhood, according to the keypoint scale. For each one of the *4 x 4* cells, compute a 8-bin histogram of gradient directions.



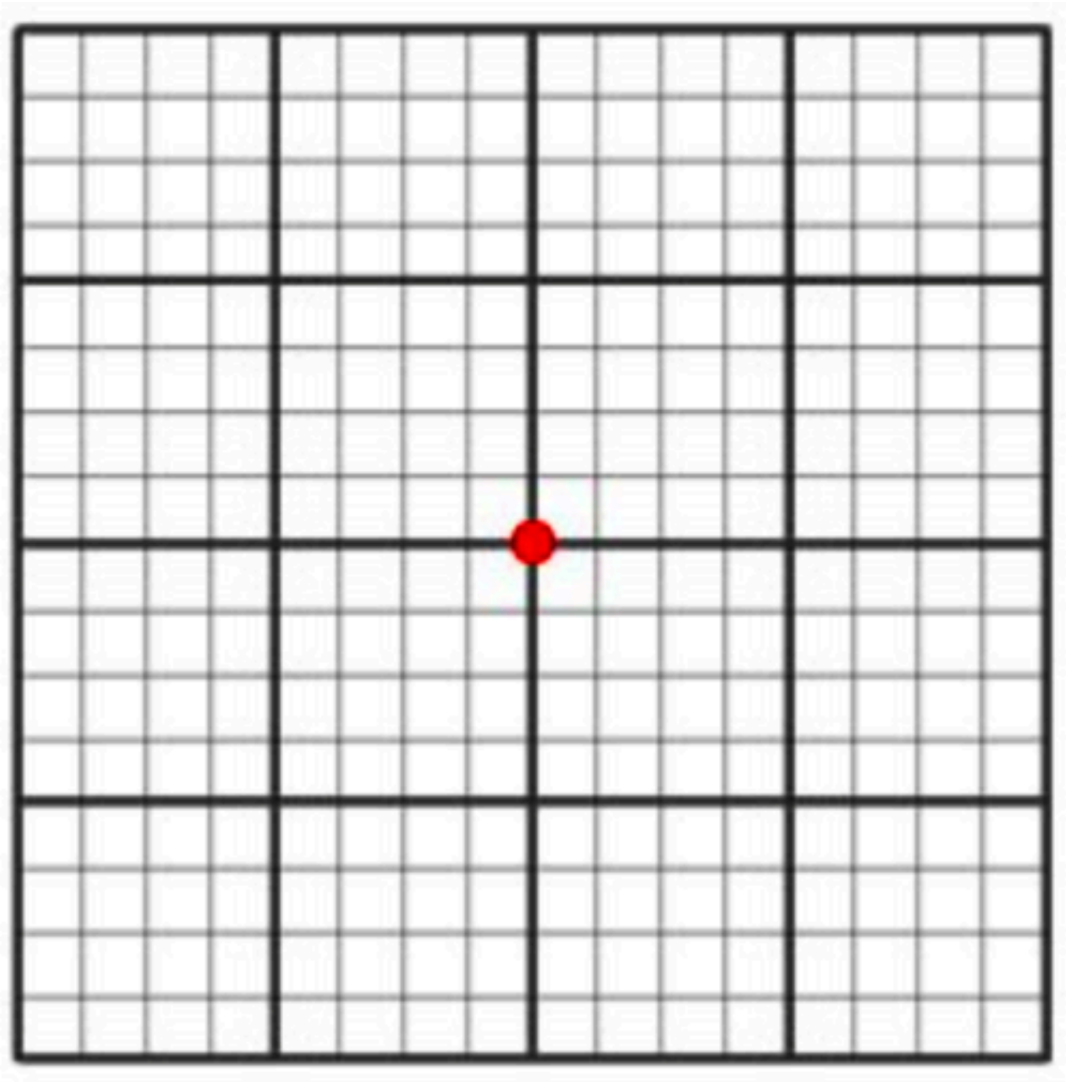


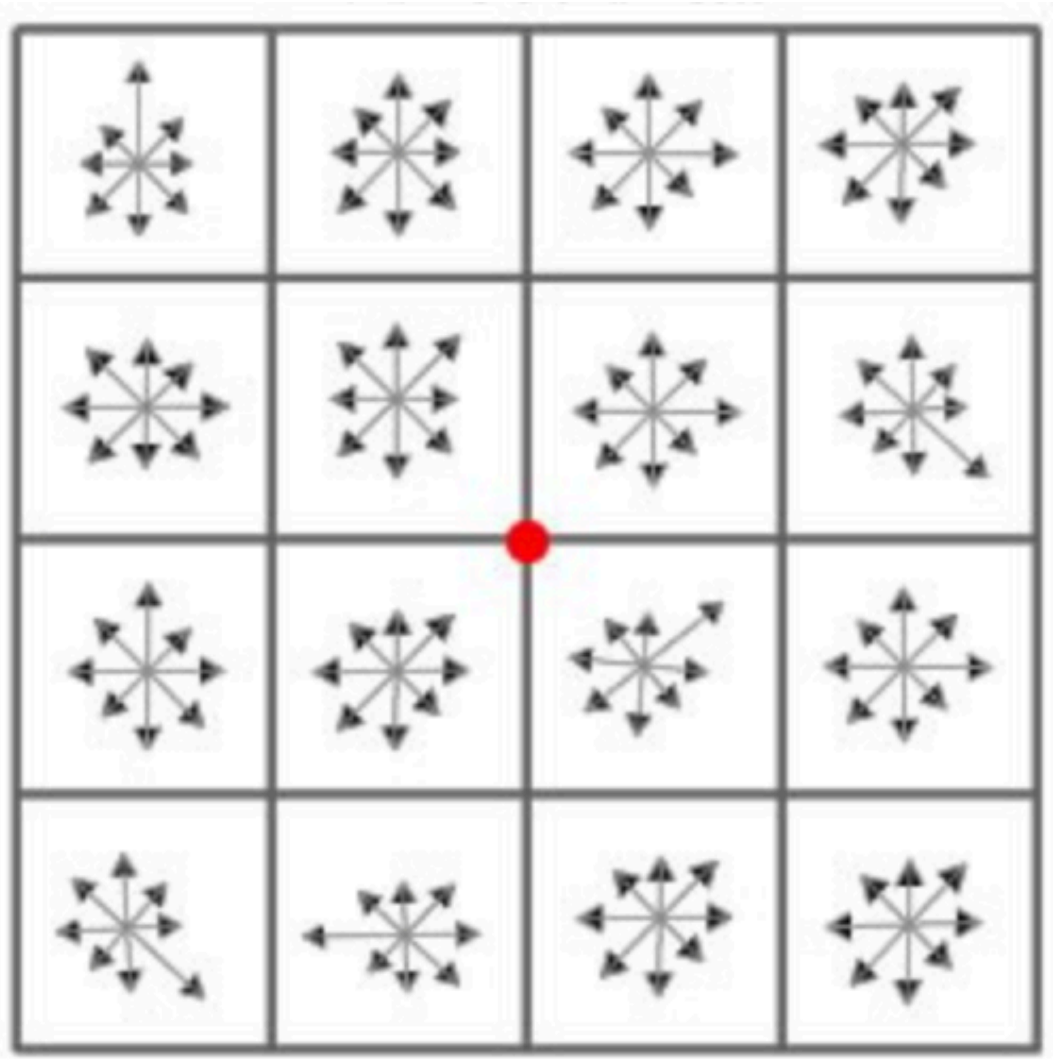
LOYOLA UNIVERSITY CHICAGO

Scale Invariant Feature Transform (SIFT)

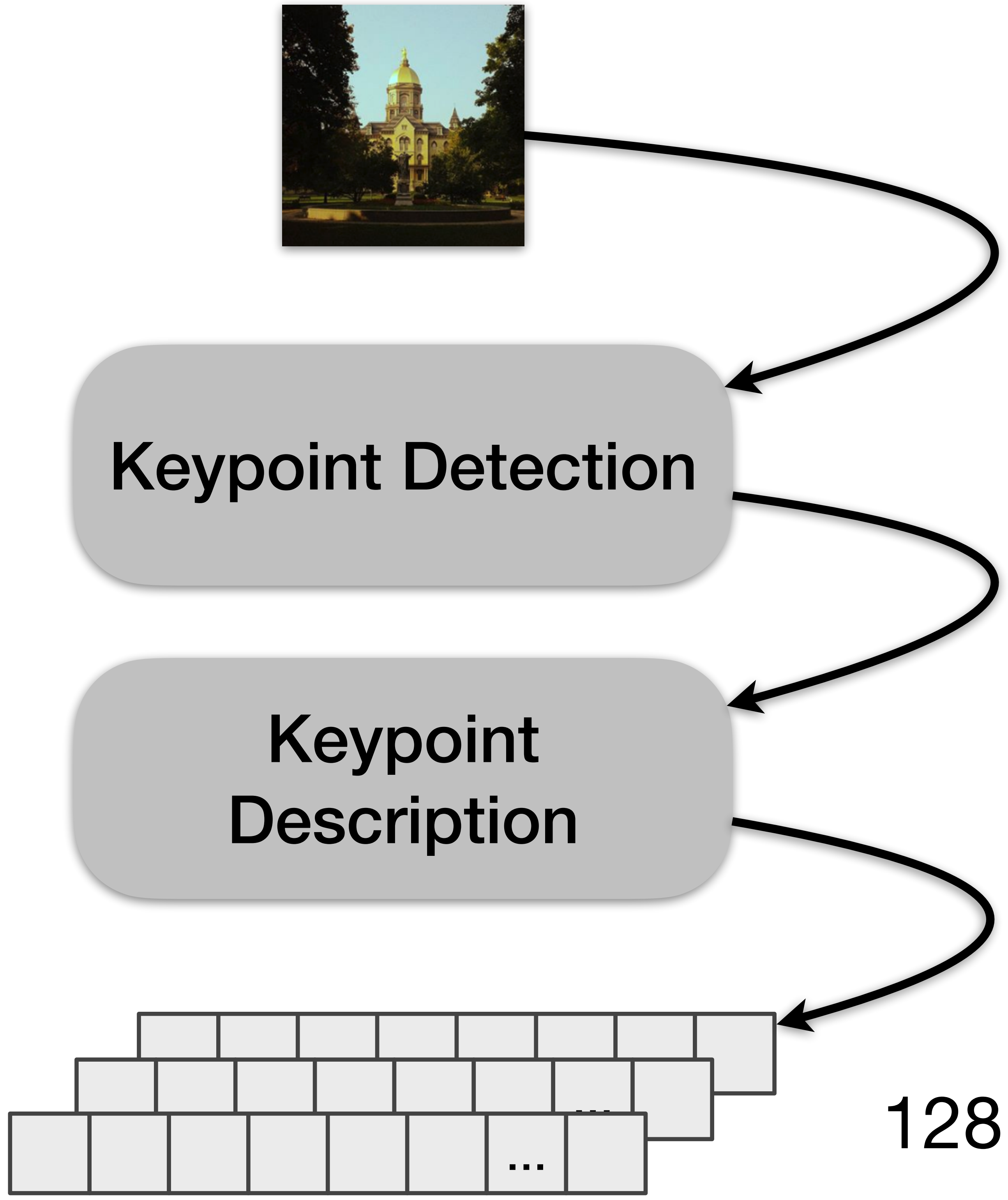
Keypoint Description

For each rotated keypoint, sample a *4 x 4* window on its neighborhood, according to the keypoint scale.
For each one of the *4 x 4* cells, compute a 8-bin histogram of gradient directions.
Fill out a feature vector with the *4 x 4 x 8 =128* histogram values.





Scale Invariant Feature Transform (SIFT)



International Journal of Computer Vision 60(2), 91–110, 2004

Distinctive Image Features from Scale-Invariant Keypoints

DAVID G. LOWE
Computer Science Department, University of British Columbia, Vancouver, B.C., Canada
Lowe@cs.ubc.ca

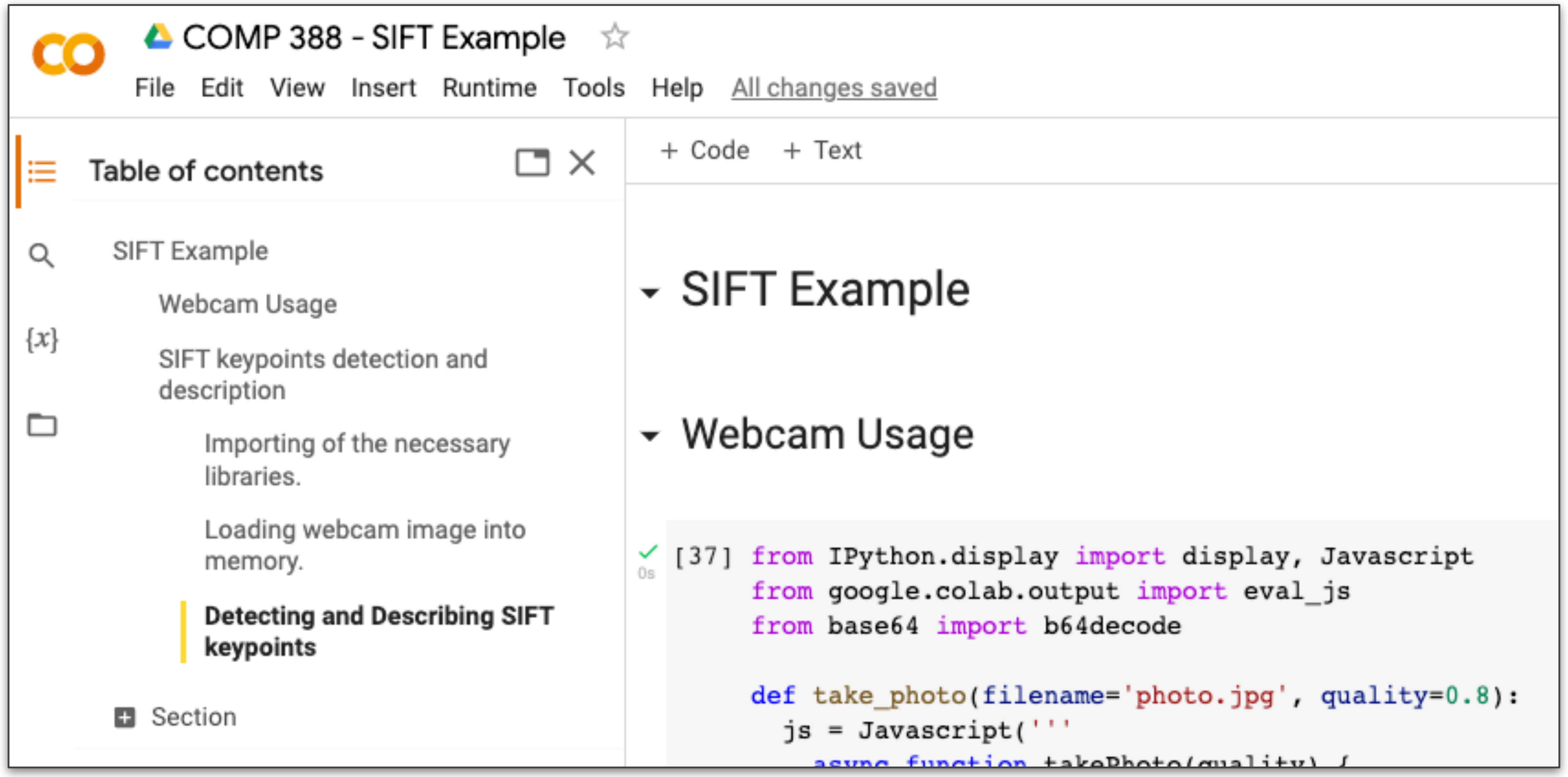
Received January 10, 2003; Revised January 7, 2004; Accepted January 22, 2004

128D feature vectors, compare two vectors with L2-distance.

Scale Invariant Feature Transform (SIFT)

Example

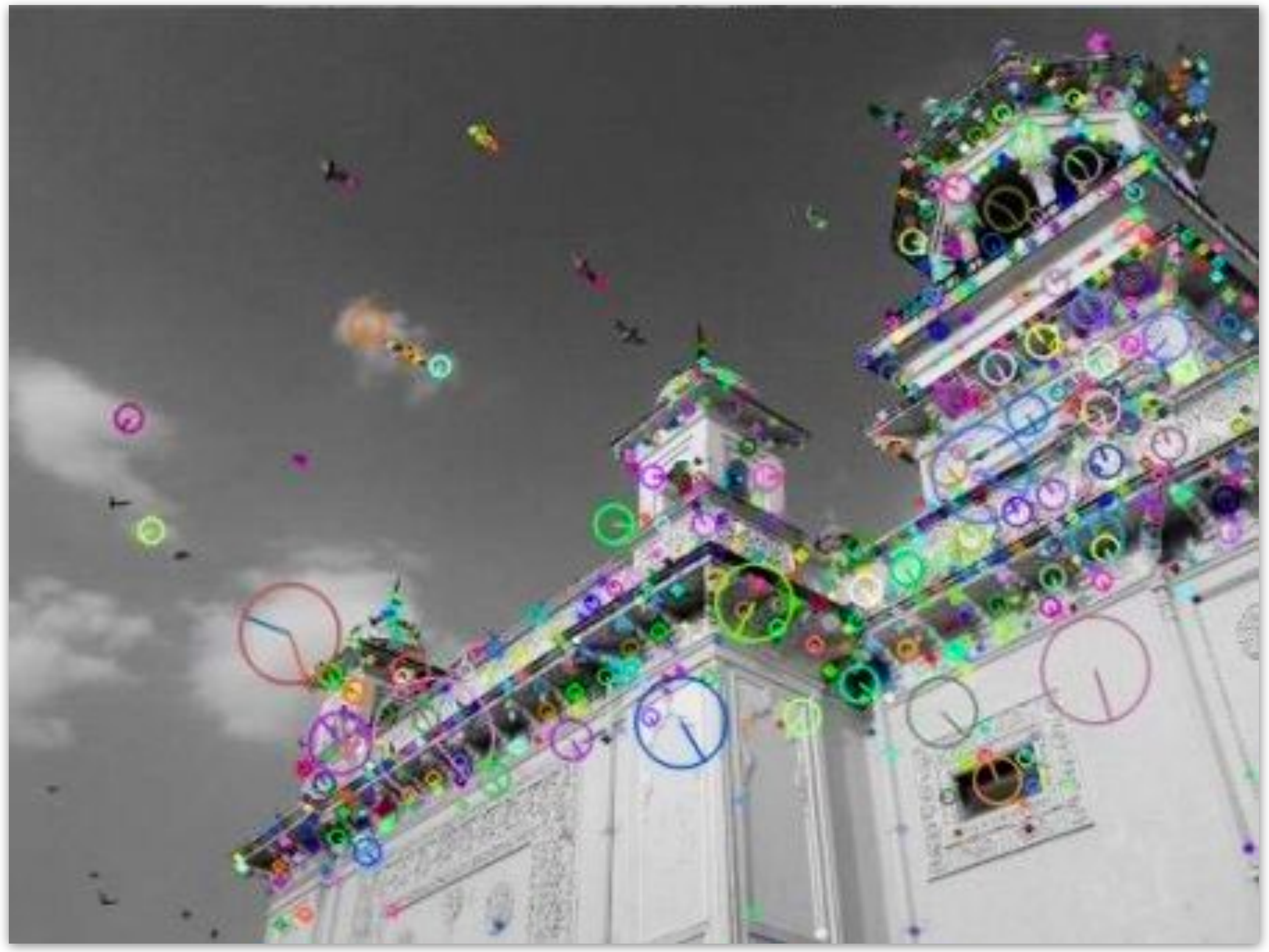
https://bit.ly/3QP3EKw



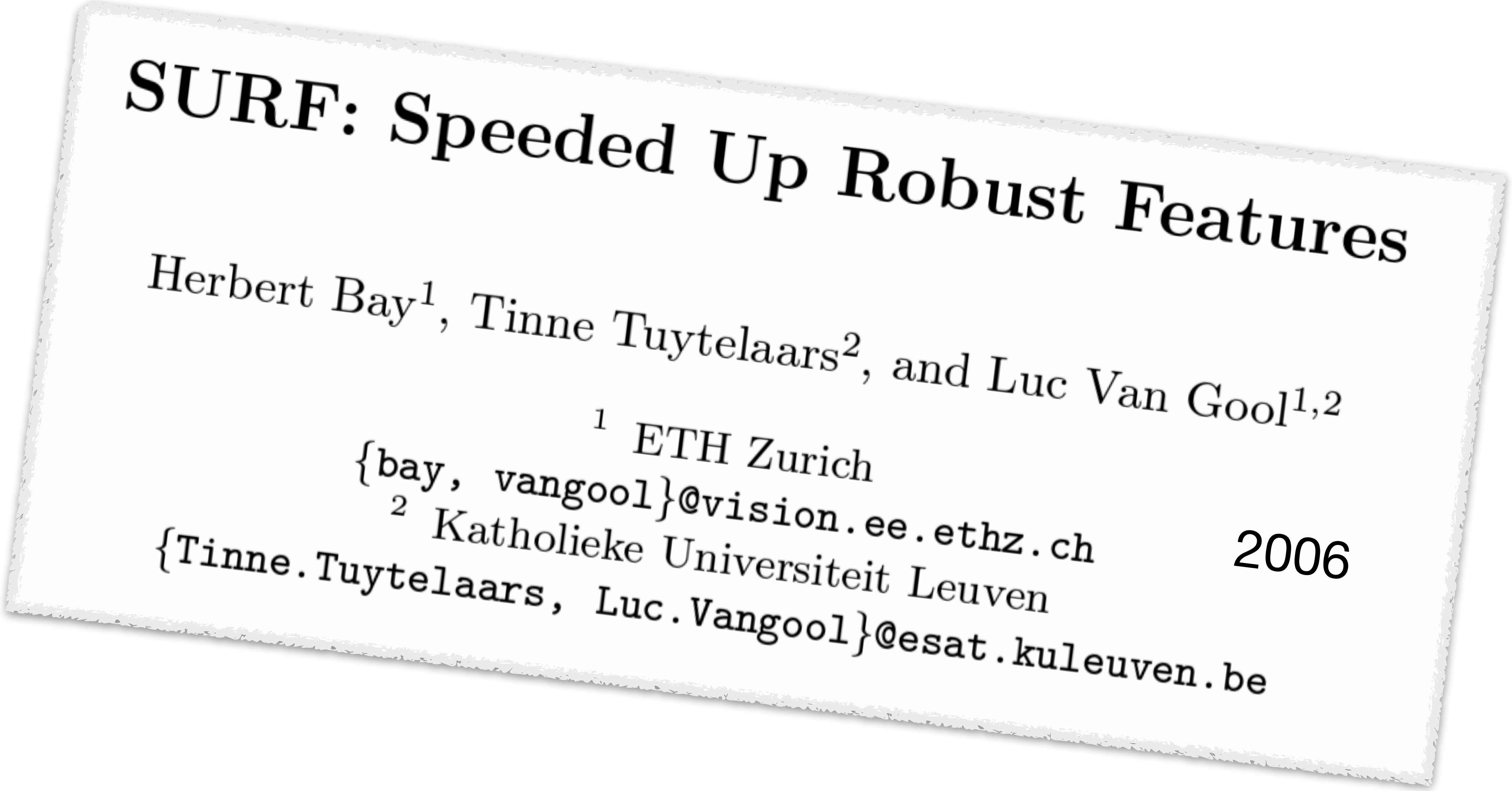
LOYOLA UNIVERSITY CHICAGO

Speeded-Up Robust Features (SURF)

How to develop a faster alternative to SIFT?



https://docs.opencv.org/4.x/da/df5/tutorial_py_sift_intro.html



SURF: Speeded Up Robust Features

Herbert Bay[1], Tinne Tuytelaars[2], and Luc Van Gool[1,2]

[1] ETH Zurich
{bay, vangool}@vision.ee.ethz.ch
[2] Katholieke Universiteit Leuven
{Tinne.Tuytelaars, Luc.Vangool}@esat.kuleuven.be

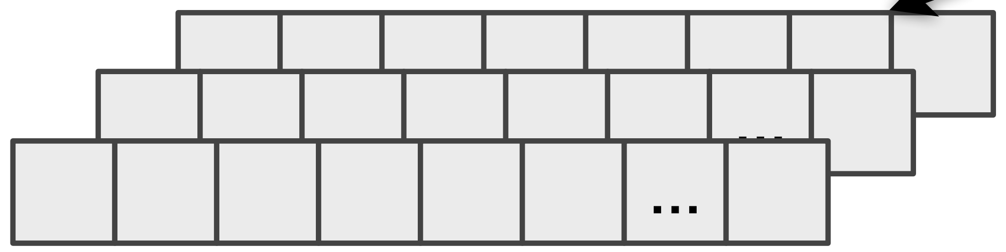
2006

Speeded-Up Robust Features (SURF)

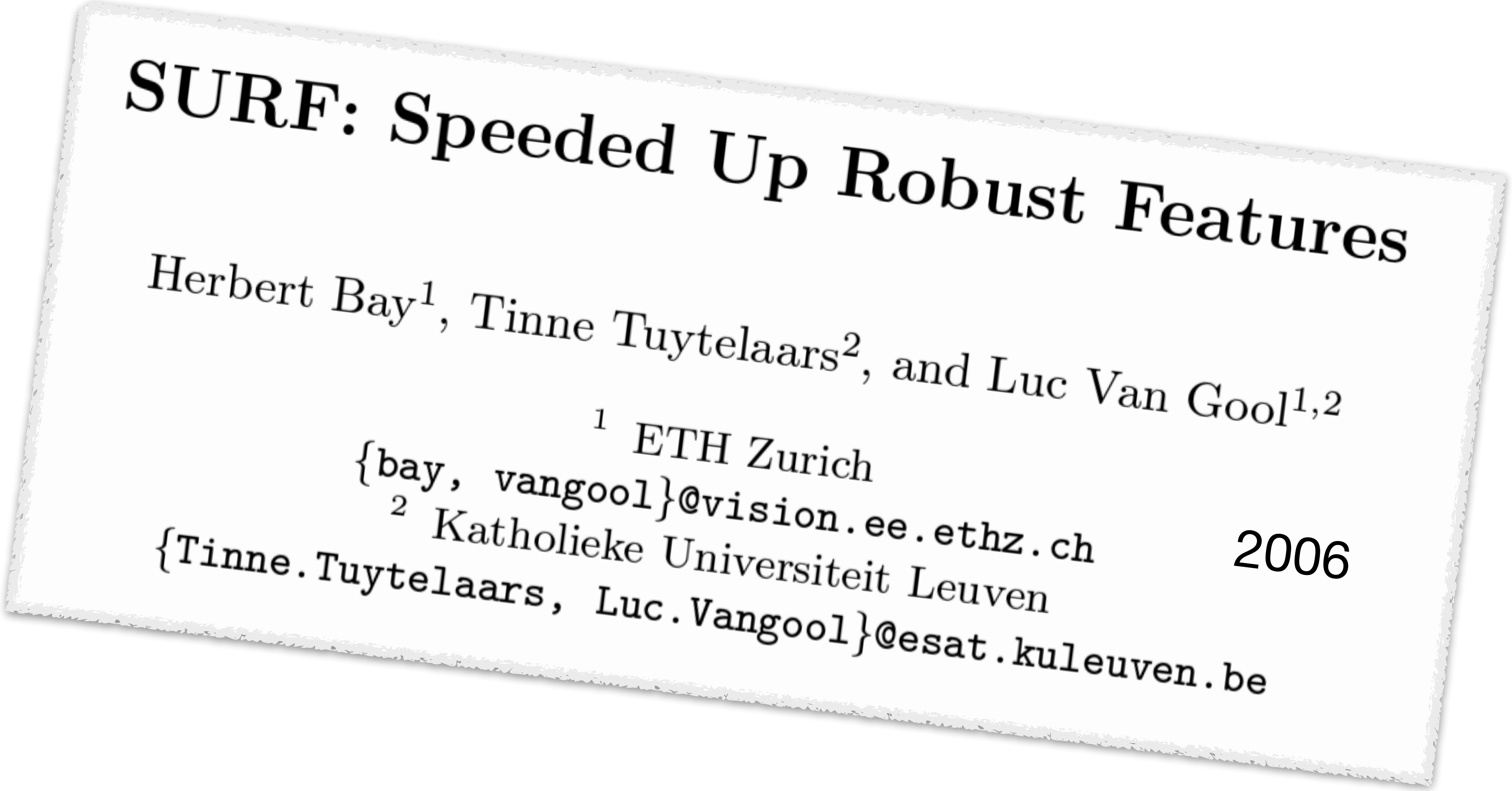
Stages

Faster Keypoint Detection

Faster Keypoint Description



64D feature vectors



SURF: Speeded Up Robust Features

Herbert Bay[1], Tinne Tuytelaars[2], and Luc Van Gool[1,2]

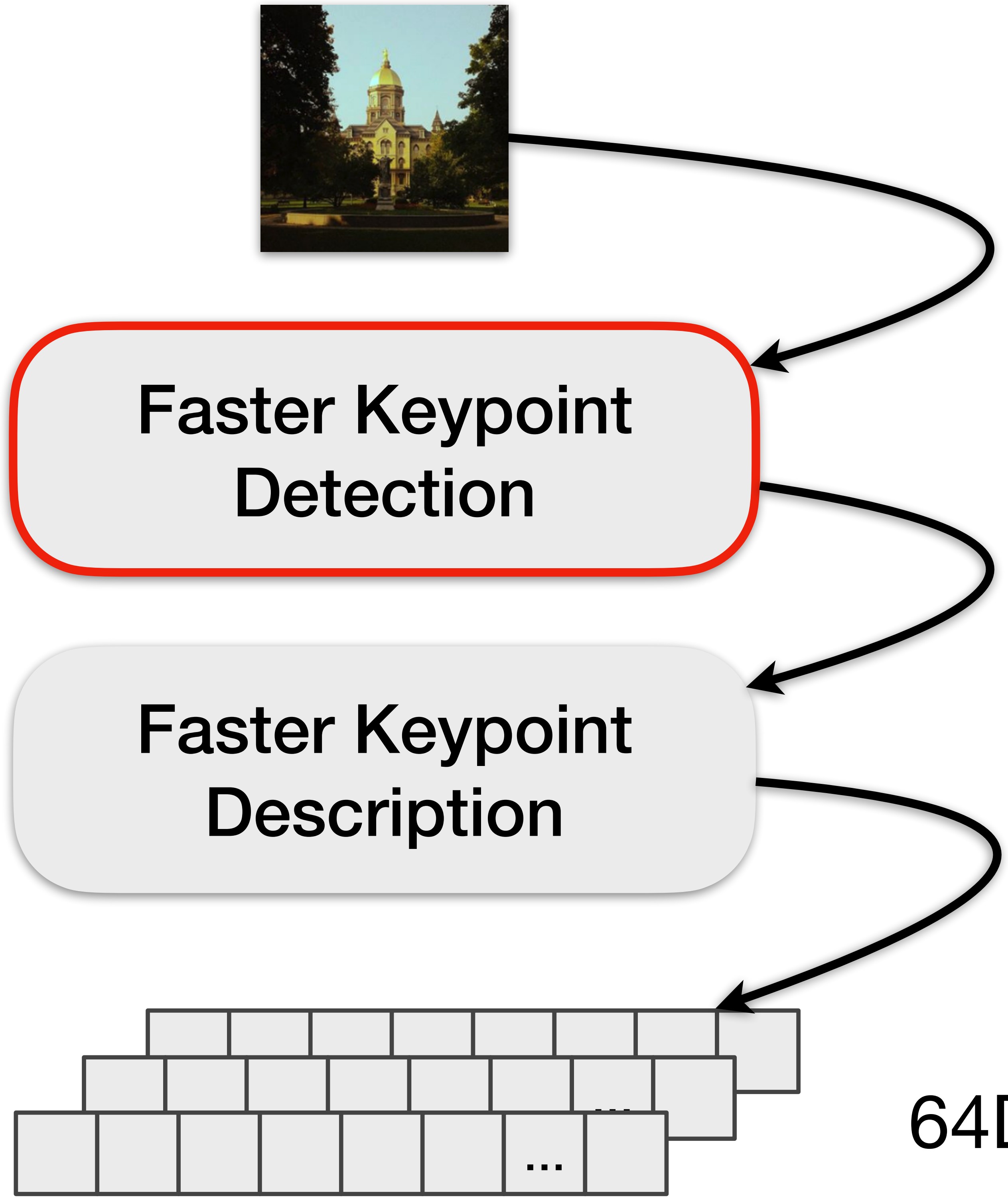
[1] ETH Zurich
{bay, vangool}@vision.ee.ethz.ch
[2] Katholieke Universiteit Leuven
{Tinne.Tuytelaars, Luc.Vangool}@esat.kuleuven.be

2006

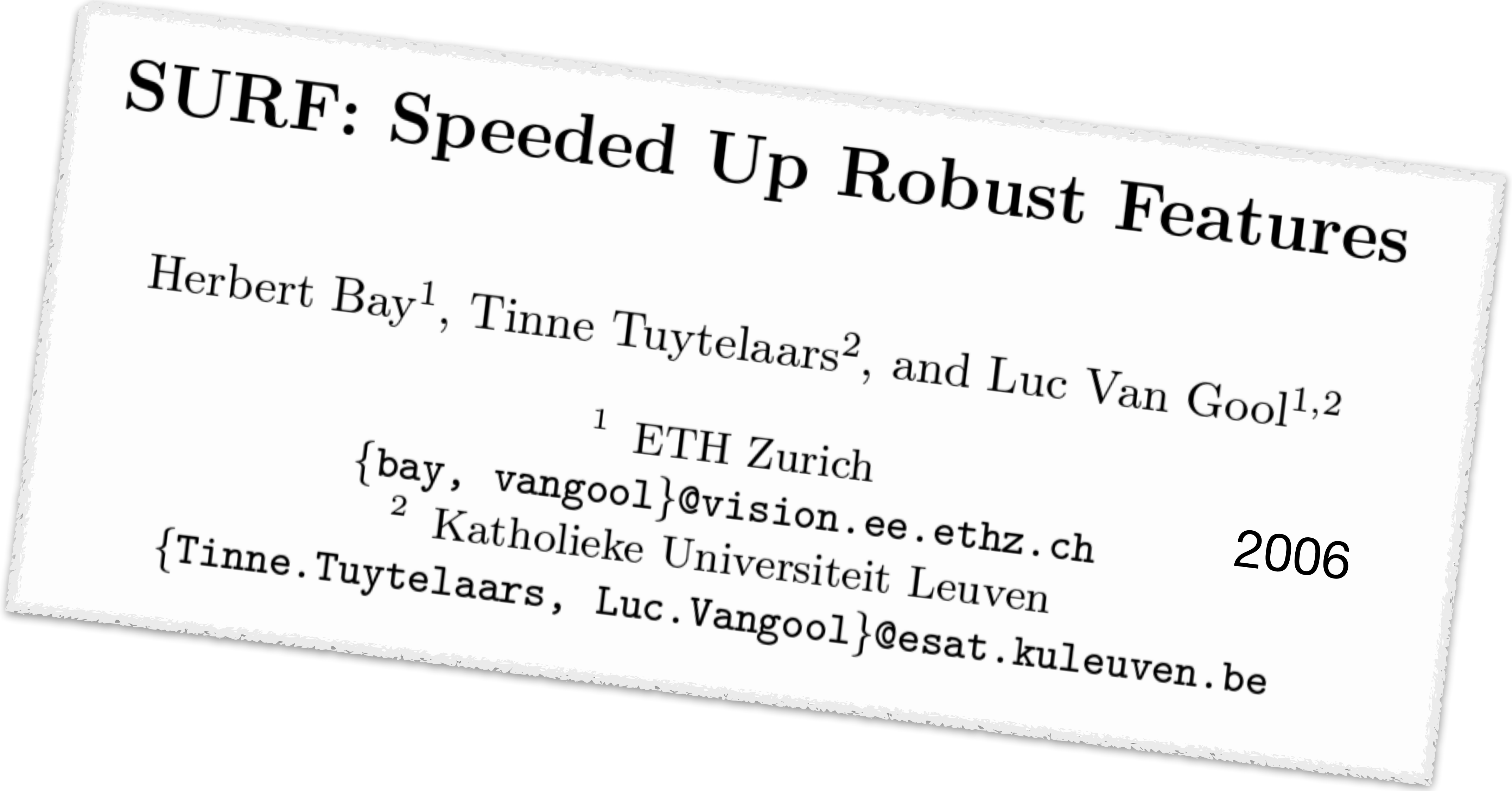
LOYOLA UNIVERSITY CHICAGO

Speeded-Up Robust Features (SURF)

Stages



64D feature vectors


SURF: Speeded Up Robust Features

Herbert Bay[1], Tinne Tuytelaars[2], and Luc Van Gool[1,2]

[1] ETH Zurich
{bay, vangool}@vision.ee.ethz.ch
[2] Katholieke Universiteit Leuven
{Tinne.Tuytelaars, Luc.Vangool}@esat.kuleuven.be

2006

Speeded-Up Robust Features (SURF)

Faster Keypoint Detection
Hessian Matrix

Given an image pixel $I(x, y)$, a scale of interest σ,

and Gaussian second order derivative functions $\dfrac{\delta^2}{\delta x^2}G(\sigma)$, $\dfrac{\delta^2}{\delta y^2}G(\sigma)$, and $\dfrac{\delta^2}{\delta xy}g(\sigma)$,

the Hessian matrix H is given by:

$$H(x, y, \sigma) = \begin{bmatrix} \frac{\delta^2}{\delta x^2}g(\sigma) * I(x, y) & \frac{\delta^2}{\delta xy}g(\sigma) * I(x, y) \\ \frac{\delta^2}{\delta xy}g(\sigma) * I(x, y) & \frac{\delta^2}{\delta y^2}g(\sigma) * I(x, y) \end{bmatrix}$$

Speeded-Up Robust Features (SURF)

Faster Keypoint Detection
Hessian Matrix

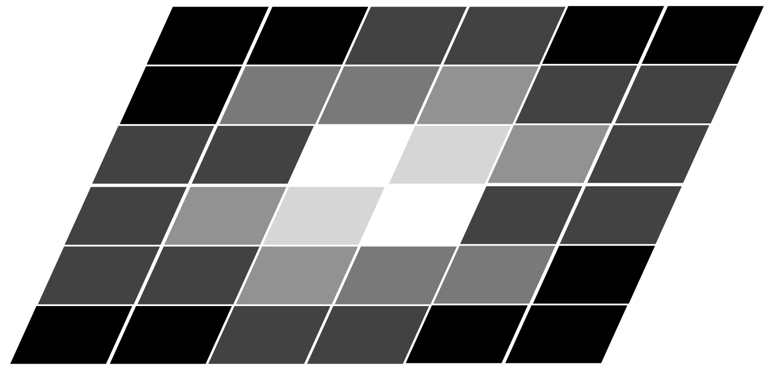
Given an image pixel $I(x, y)$, a scale of interest σ,

and Gaussian second order derivative functions $\frac{\delta^2}{\delta x^2}G(\sigma)$, $\frac{\delta^2}{\delta y^2}G(\sigma)$, and $\frac{\delta^2}{\delta xy}g(\sigma)$,

the Hessian matrix H is given by:

$$H(x, y, \sigma) = \begin{bmatrix} \frac{\delta^2}{\delta x^2}g(\sigma) * I(x, y) & \frac{\delta^2}{\delta xy}g(\sigma) * I(x, y) \\ \frac{\delta^2}{\delta xy}g(\sigma) * I(x, y) & \frac{\delta^2}{\delta y^2}g(\sigma) * I(x, y) \end{bmatrix}$$

Property: blobs with scale σ and centered at $I(x, y)$ will lead to a large $det(H)$.



Take the regions with large $det(H)$ as candidate keypoints.

LOYOLA
UNIVERSITY CHICAGO

Speeded-Up Robust Features (SURF)

Faster Keypoint Detection
Integral Image

Data structure $I_{\sum}$ computed from a given image I that shares the same resolution (i.e., same number of rows and of columns).

Each "pixel" of $I_{\sum}$ has the following value:

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

i.e., it holds the sum of all the pixel values of I that spatially precede the position (x, y).

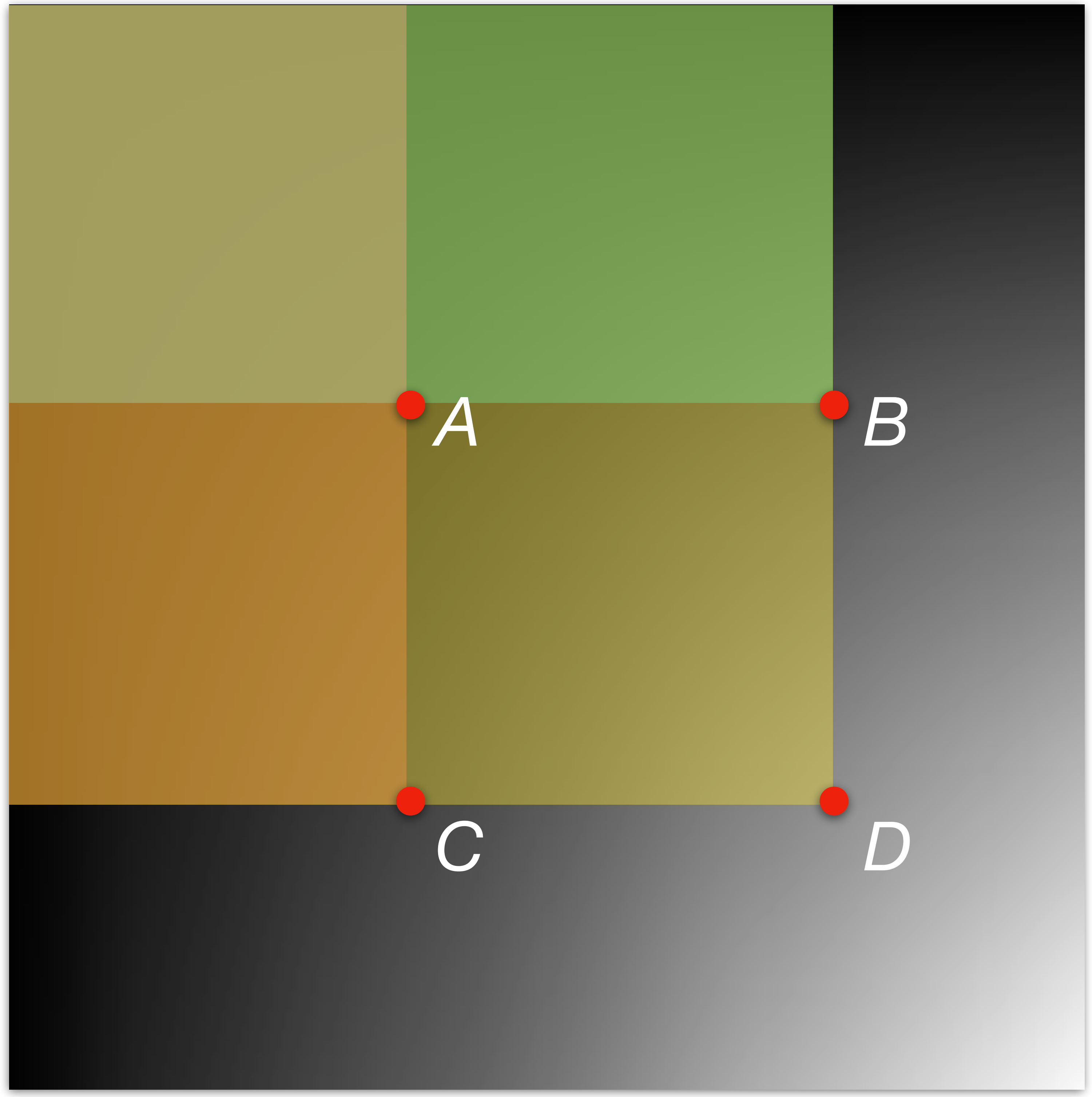
Speeded-Up Robust Features (SURF)

Faster Keypoint Detection
Integral Image

What is the utility?

It is easy to compute the sum of pixel values within any region regardless of the region size.

(0,0)



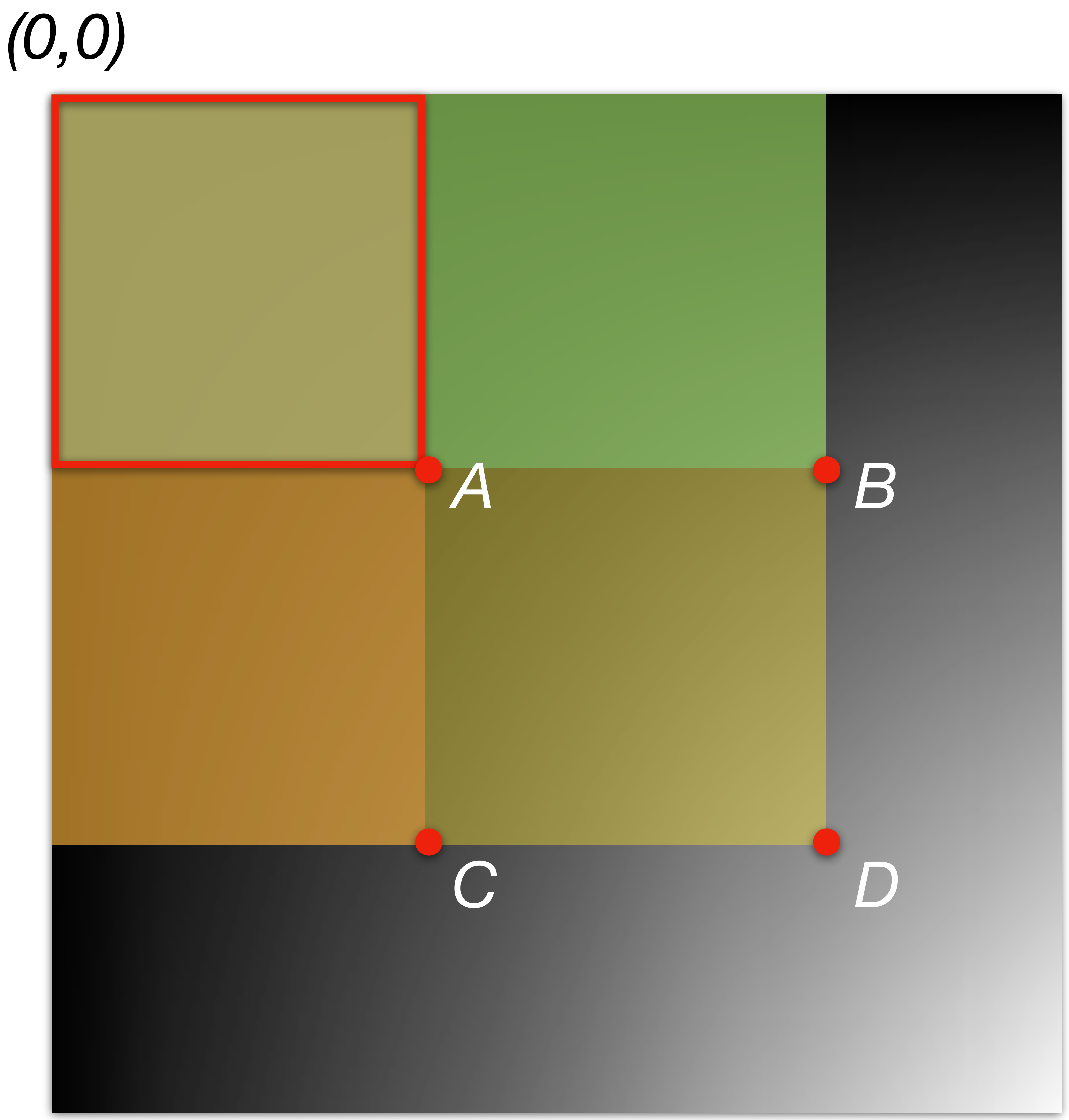
LOYOLA UNIVERSITY CHICAGO

Speeded-Up Robust Features (SURF)

Faster Keypoint Detection
Integral Image

What is the utility?

It is easy to compute the sum of pixel values within any region regardless of the region size.



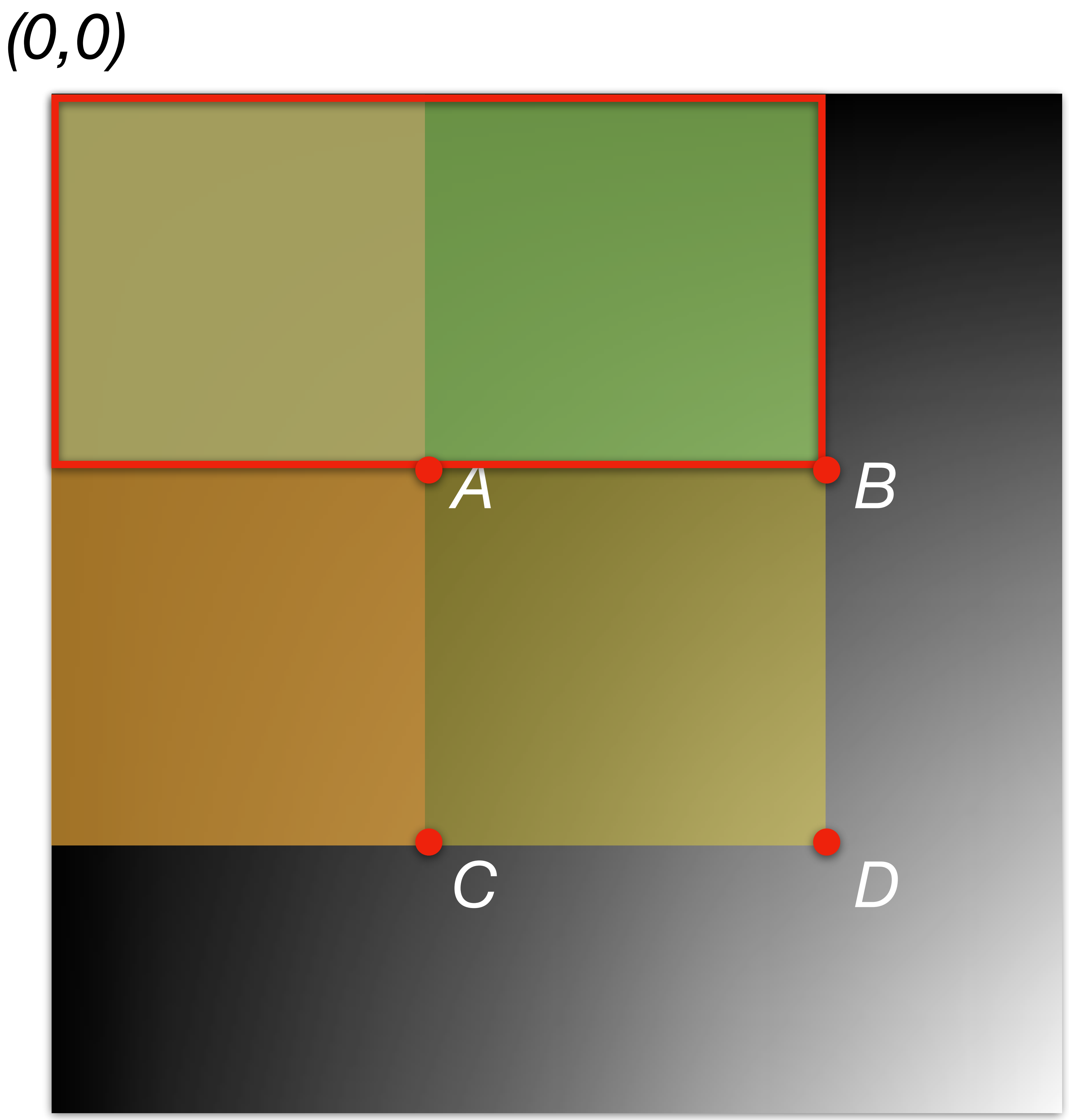
Sum of region between *(0,0)* and *A*?
Answer: $I_{\Sigma}(A)$

Speeded-Up Robust Features (SURF)

Faster Keypoint Detection
Integral Image

What is the utility?

It is easy to compute the sum of pixel values within any region regardless of the region size.



Sum of region between *(0,0)* and *A*?
Answer: $I_{\sum}(A)$

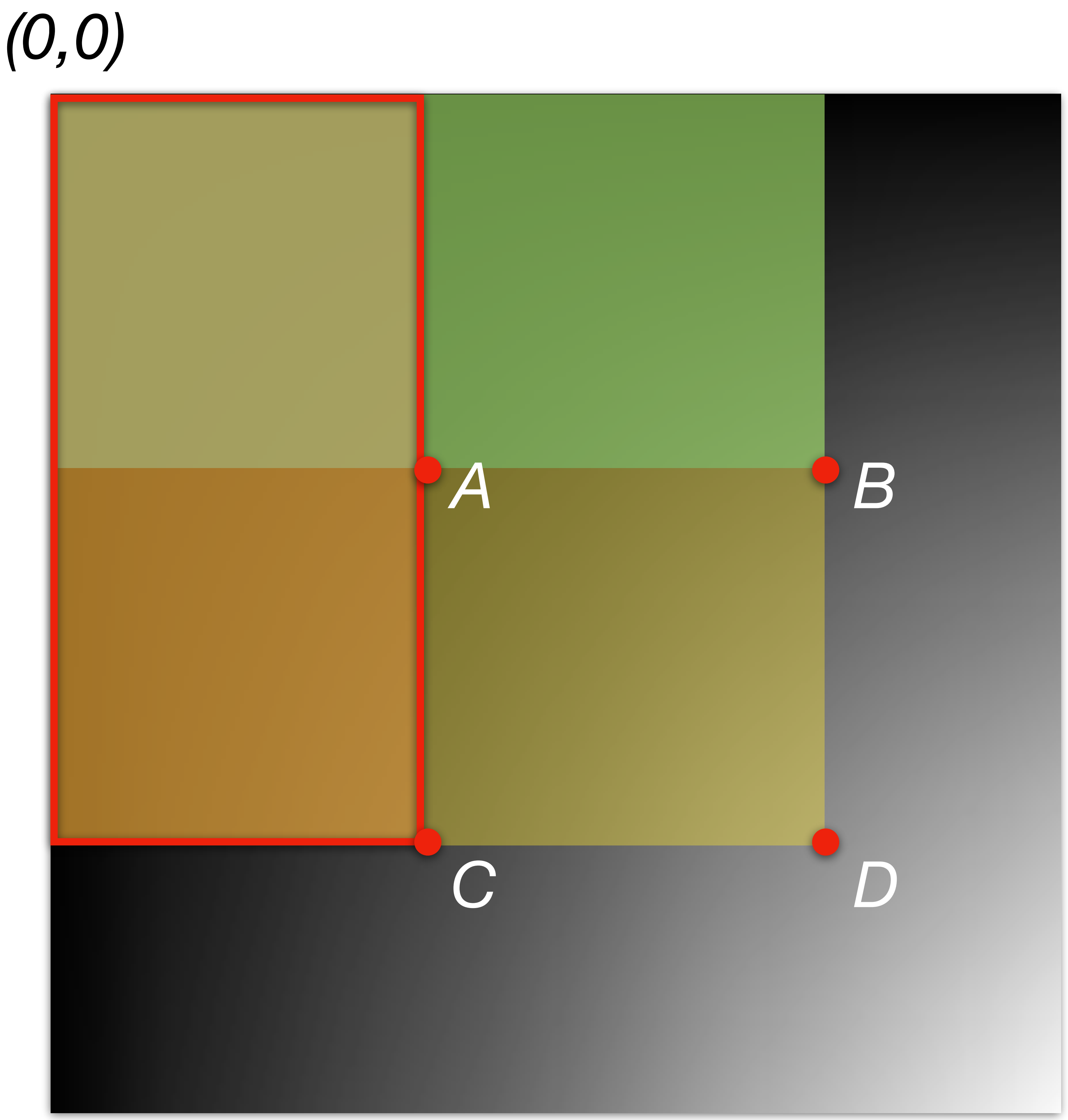
Sum of region between *(0,0)* and *B*?
Answer: $I_{\sum}(B)$

Speeded-Up Robust Features (SURF)

Faster Keypoint Detection
Integral Image

What is the utility?

It is easy to compute the sum of pixel values within any region regardless of the region size.



Sum of region between *(0,0)* and *A*?
Answer: $I_{\Sigma}(A)$

Sum of region between *(0,0)* and *B*?
Answer: $I_{\Sigma}(B)$

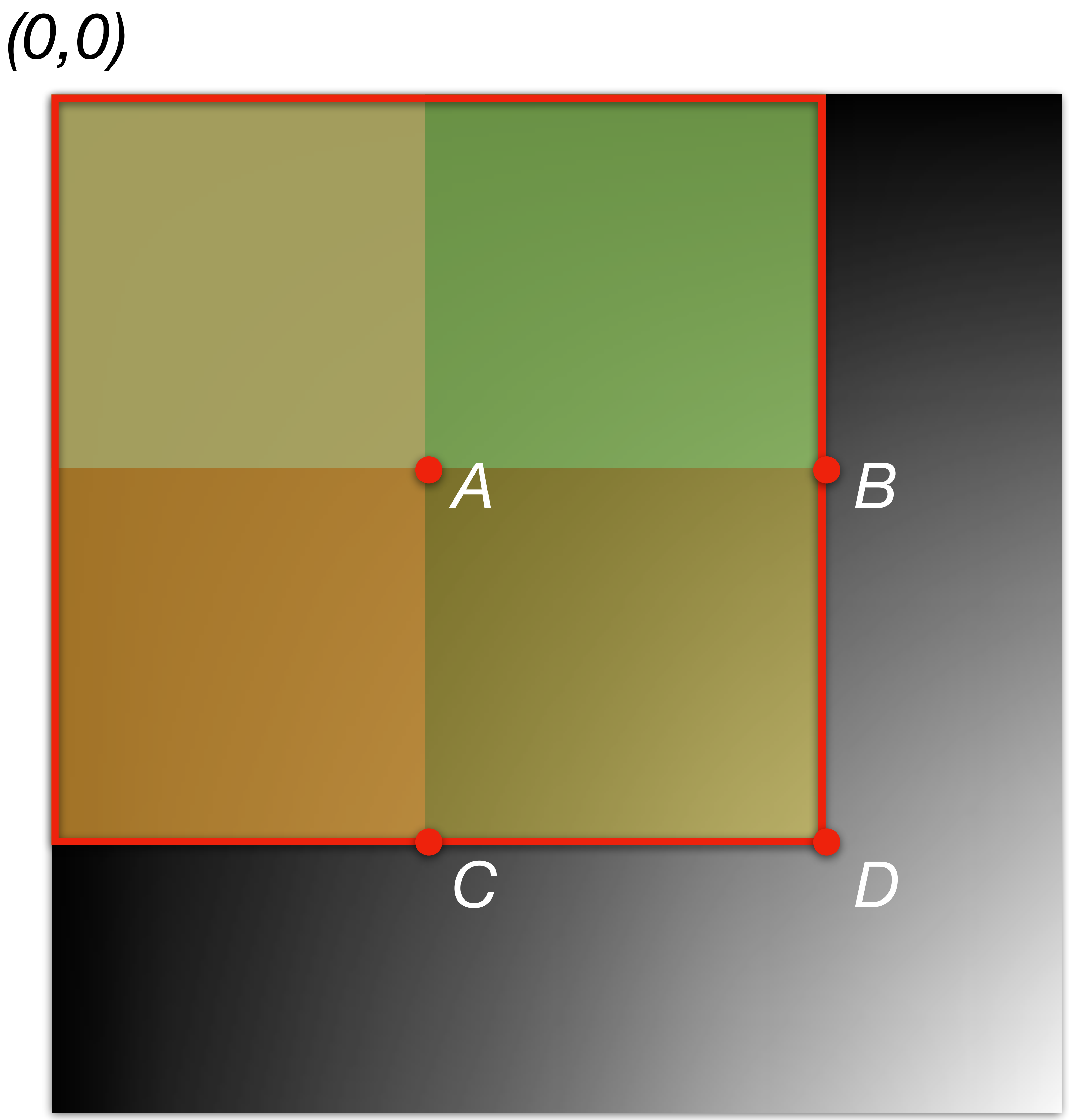
Sum of region between *(0,0)* and *C*?
Answer: $I_{\Sigma}(C)$

Speeded-Up Robust Features (SURF)

Faster Keypoint Detection
Integral Image

What is the utility?

It is easy to compute the sum of pixel values within any region regardless of the region size.



Sum of region between *(0,0)* and *A*?
Answer: $I_{\sum}(A)$

Sum of region between *(0,0)* and *B*?
Answer: $I_{\sum}(B)$

Sum of region between *(0,0)* and *C*?
Answer: $I_{\sum}(C)$

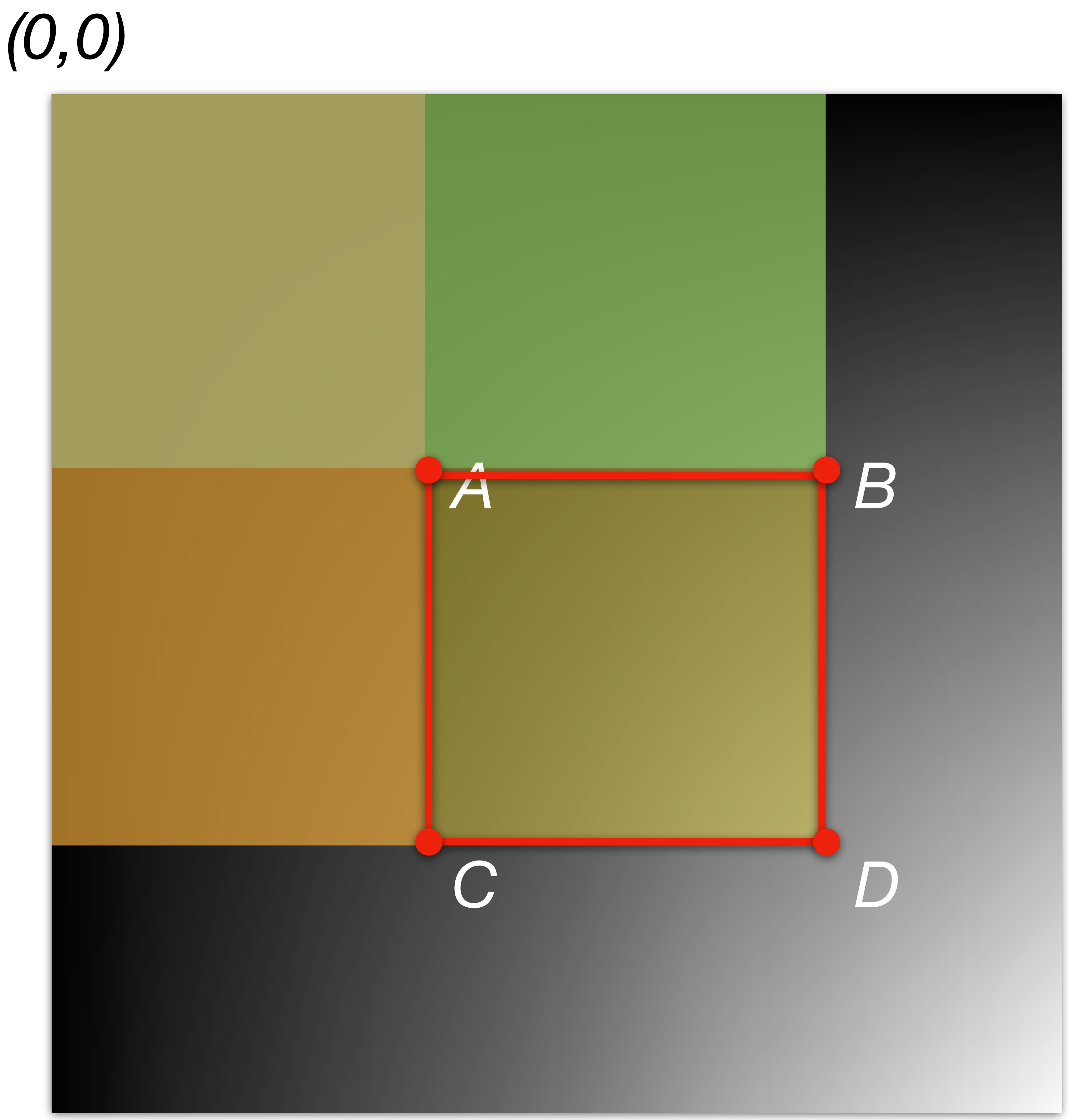
Sum of region between *(0,0)* and *D*?
Answer: $I_{\sum}(D)$

Speeded-Up Robust Features (SURF)

Faster Keypoint Detection
Integral Image

What is the utility?

It is easy to compute the sum of pixel values within any region regardless of the region size.



Sum of region between *A* and *D*?

Answer:

$$I_{\sum}(D) - I_{\sum}(B) - I_{\sum}(C) + I_{\sum}(A)$$

One can get any sum with at most 4 accesses, regardless of the resolution.

LOYOLA UNIVERSITY CHICAGO

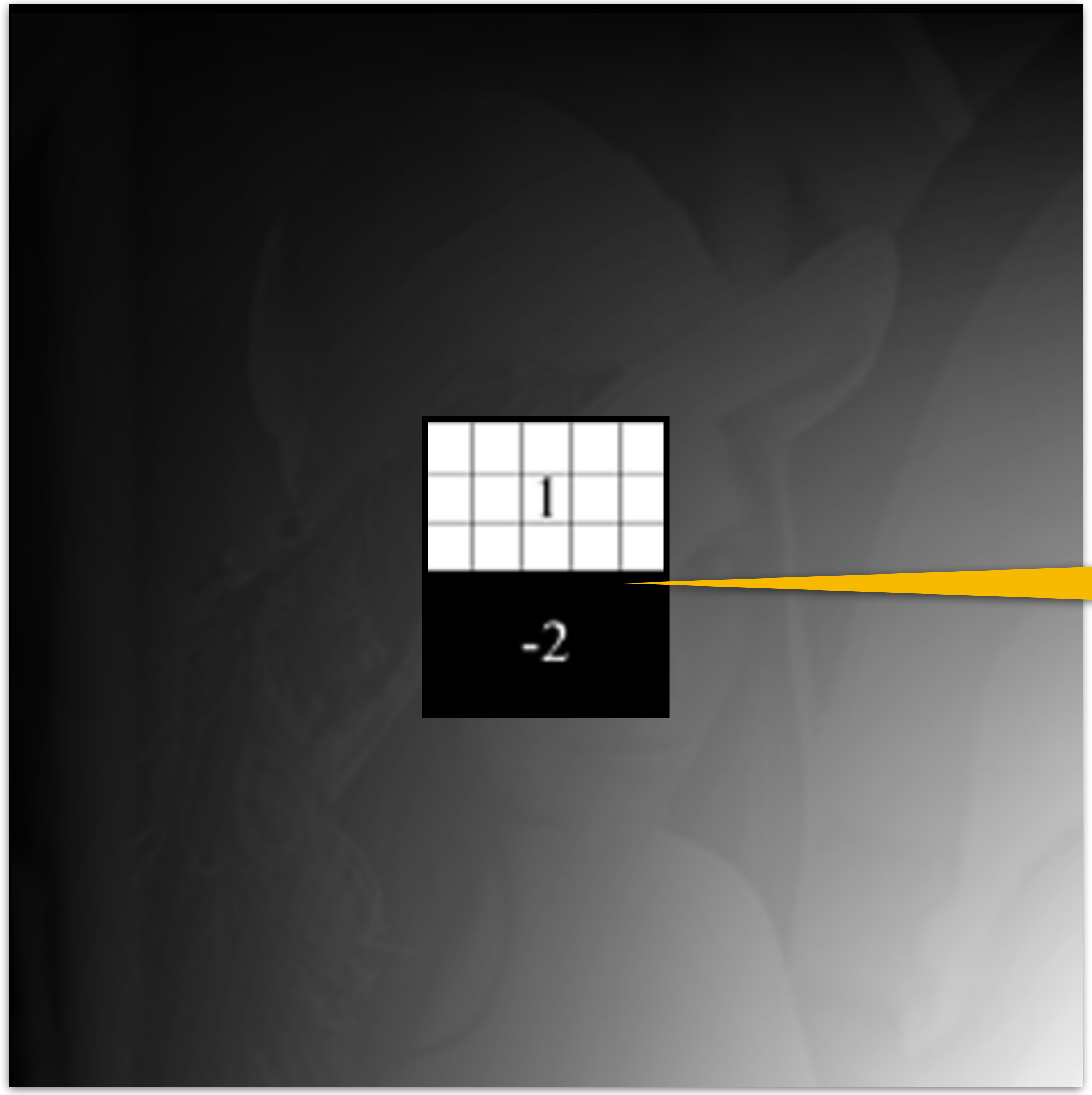
Speeded-Up Robust Features (SURF)

Faster Keypoint Detection
Integral Image

What is the utility?

It is easy to compute the sum of pixel values within any region regardless of the region size.

(0,0)



Box Filters can be easily convoluted with the integral image.

One can quickly test a horizontal edge here (8 accesses)

LOYOLA UNIVERSITY CHICAGO

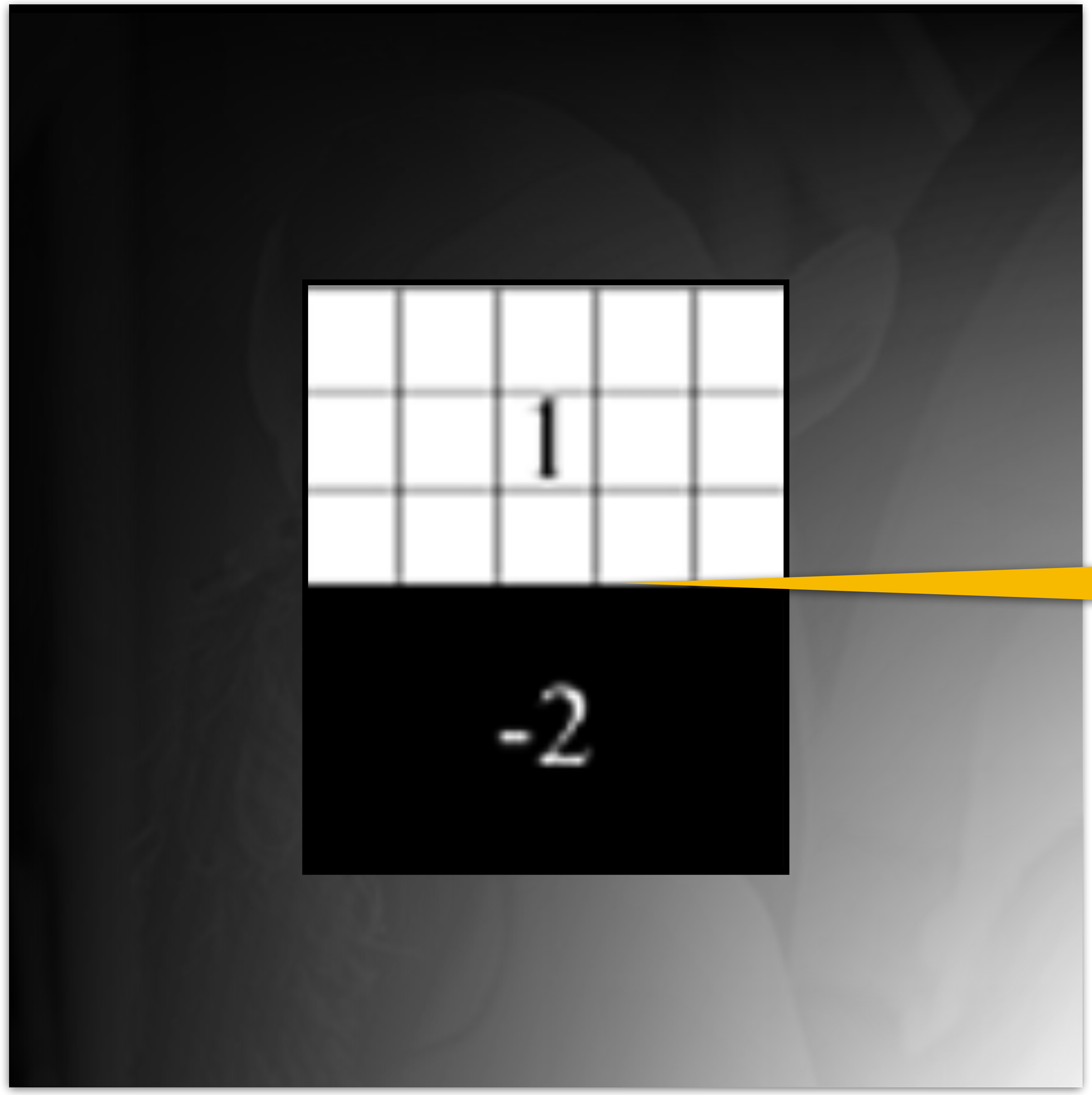
Speeded-Up Robust Features (SURF)

Faster Keypoint Detection
Integral Image

What is the utility?

It is easy to compute the sum of pixel values within any region regardless of the region size.

(0,0)



Box Filters can be easily convoluted with the integral image.

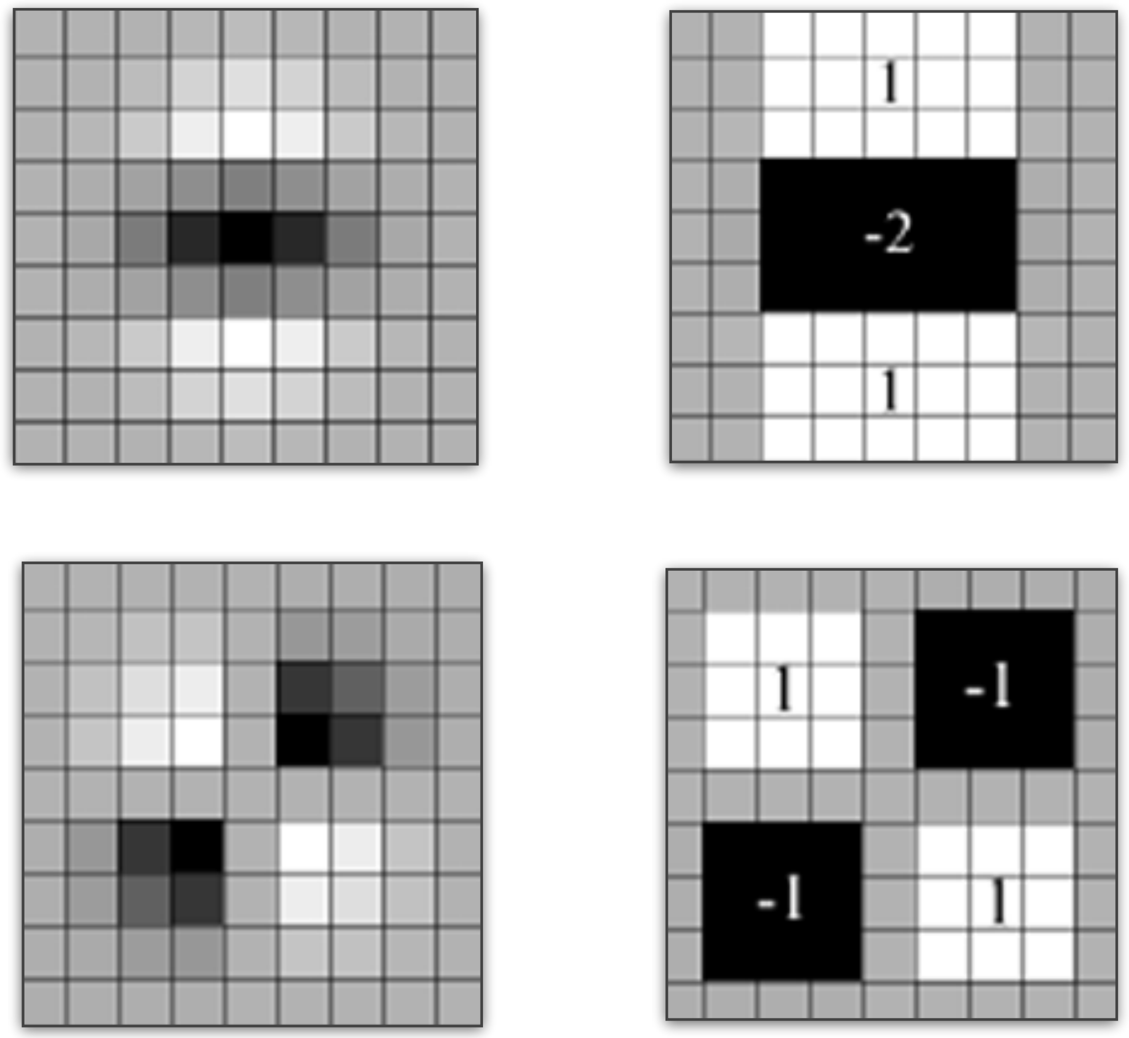
One can quickly test a larger scale horizontal edge here (still 8 accesses)

Speeded-Up Robust Features (SURF)

Faster Keypoint Detection
Box Filters

The Gaussian second order derivative functions $\frac{\delta^2}{\delta x^2}G(\sigma)$, $\frac{\delta^2}{\delta y^2}G(\sigma)$, and $\frac{\delta^2}{\delta xy}g(\sigma)$ can be approximated by box filters.

Bay's



$$\frac{\delta^2}{\delta y^2}G(\sigma)$$

$$\frac{\delta^2}{\delta xy}g(\sigma)$$

Speeded-Up Robust Features (SURF)

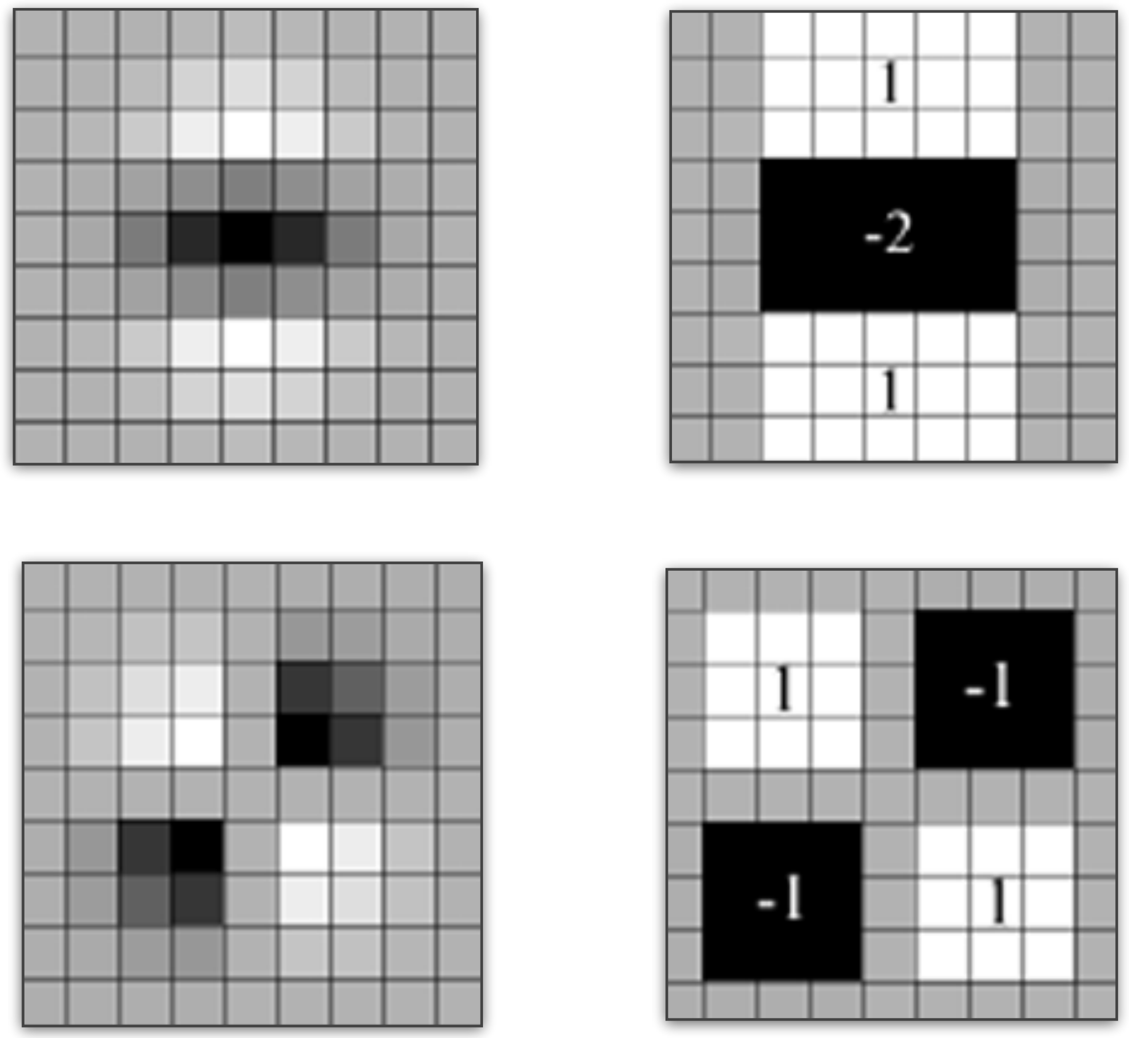
Faster Keypoint Detection
Box Filters

The Gaussian second order derivative functions $\frac{\delta^2}{\delta x^2}G(\sigma)$, $\frac{\delta^2}{\delta y^2}G(\sigma)$, and $\frac{\delta^2}{\delta xy}g(\sigma)$ can be approximated by box filters.

Compute the $det(H)$ quickly by using the box filters and the integral image!

$$H(x, y, \sigma) = \begin{bmatrix} \frac{\delta^2}{\delta x^2}g(\sigma) * I(x, y) & \frac{\delta^2}{\delta xy}g(\sigma) * I(x, y) \\ \frac{\delta^2}{\delta xy}g(\sigma) * I(x, y) & \frac{\delta^2}{\delta y^2}g(\sigma) * I(x, y) \end{bmatrix}$$

Bay's



$$\frac{\delta^2}{\delta y^2}G(\sigma)$$

$$\frac{\delta^2}{\delta xy}g(\sigma)$$

LOYOLA UNIVERSITY CHICAGO

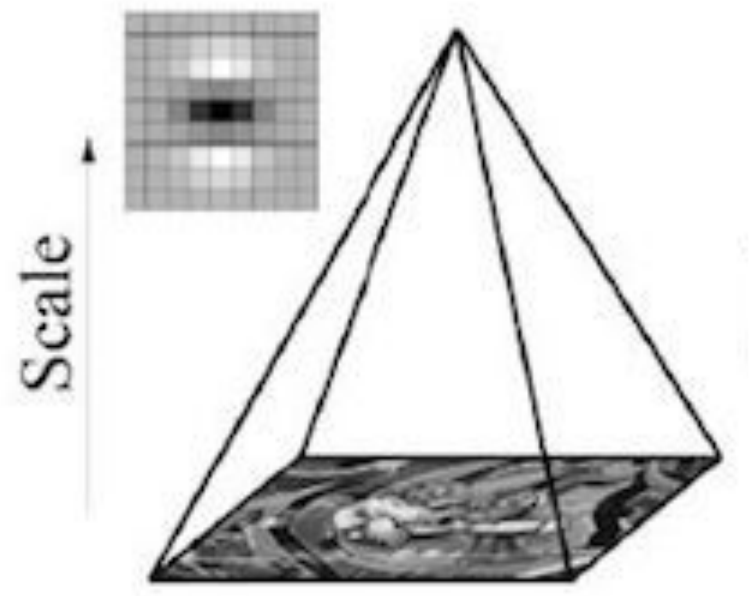
Speeded-Up Robust Features (SURF)

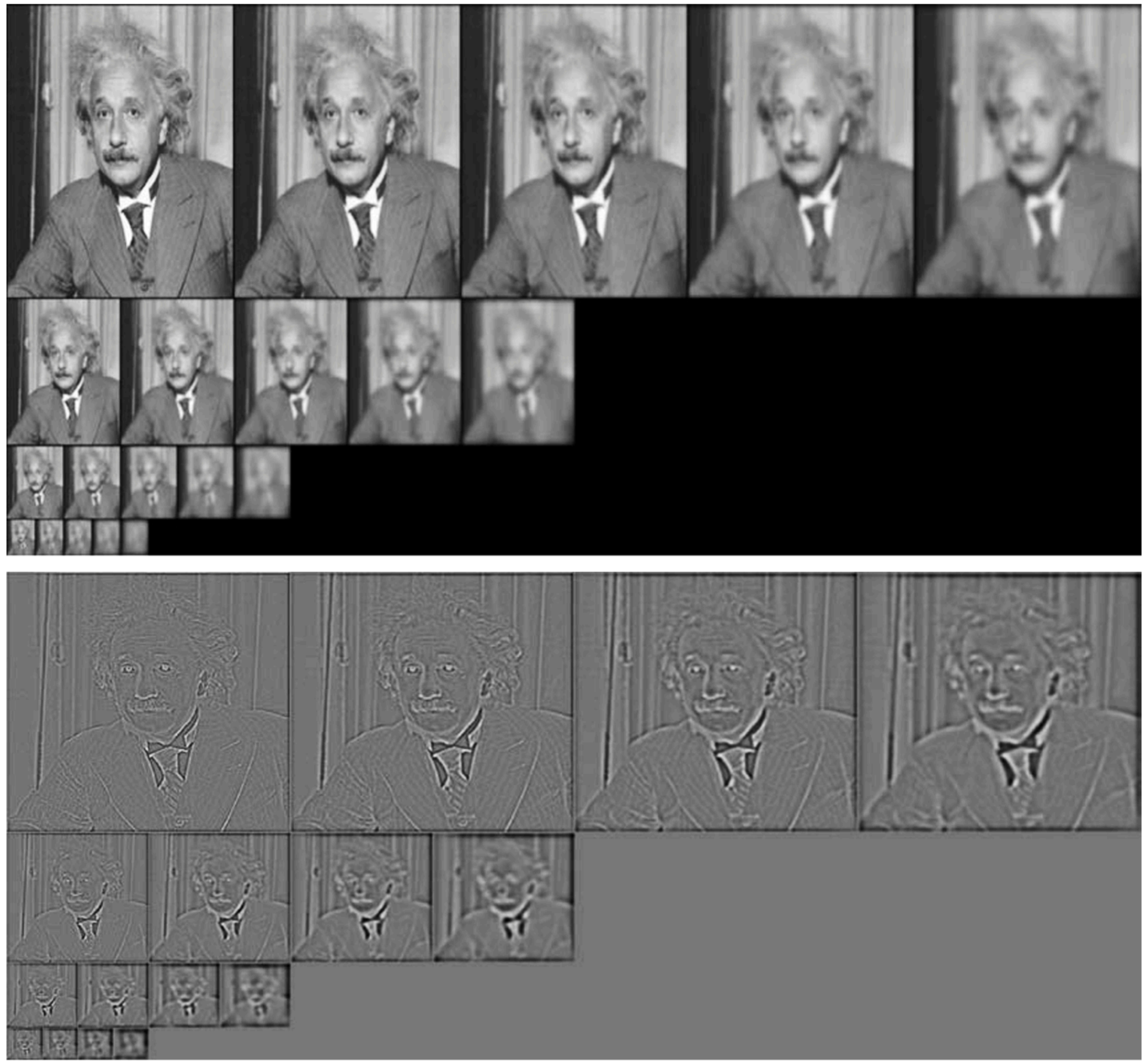
https://faculty.cc.gatech.edu/~afb/classes/CS4495-Fall2013/slides/CS4495-11-Features2.pdf

Faster Keypoint Detection
Scale Invariance

How to obtain scale invariance?

From SIFT: Use resolution pyramid.





https://medium.com/@deepanshut041/introduction-to-surf-speeded-up-robust-features-c7396d6e7c4e

Speeded-Up Robust Features (SURF)

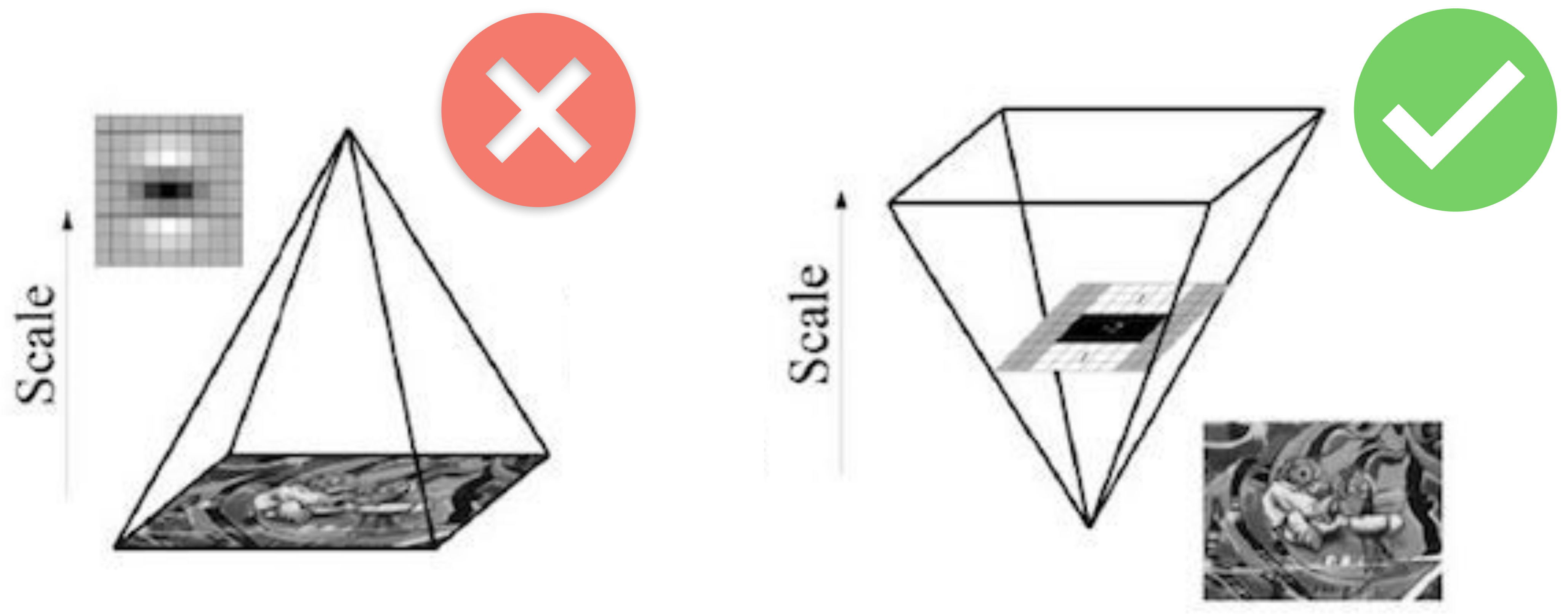
https://faculty.cc.gatech.edu/~afb/classes/CS4495-Fall2013/slides/CS4495-11-Features2.pdf

Faster Keypoint Detection
Scale Invariance

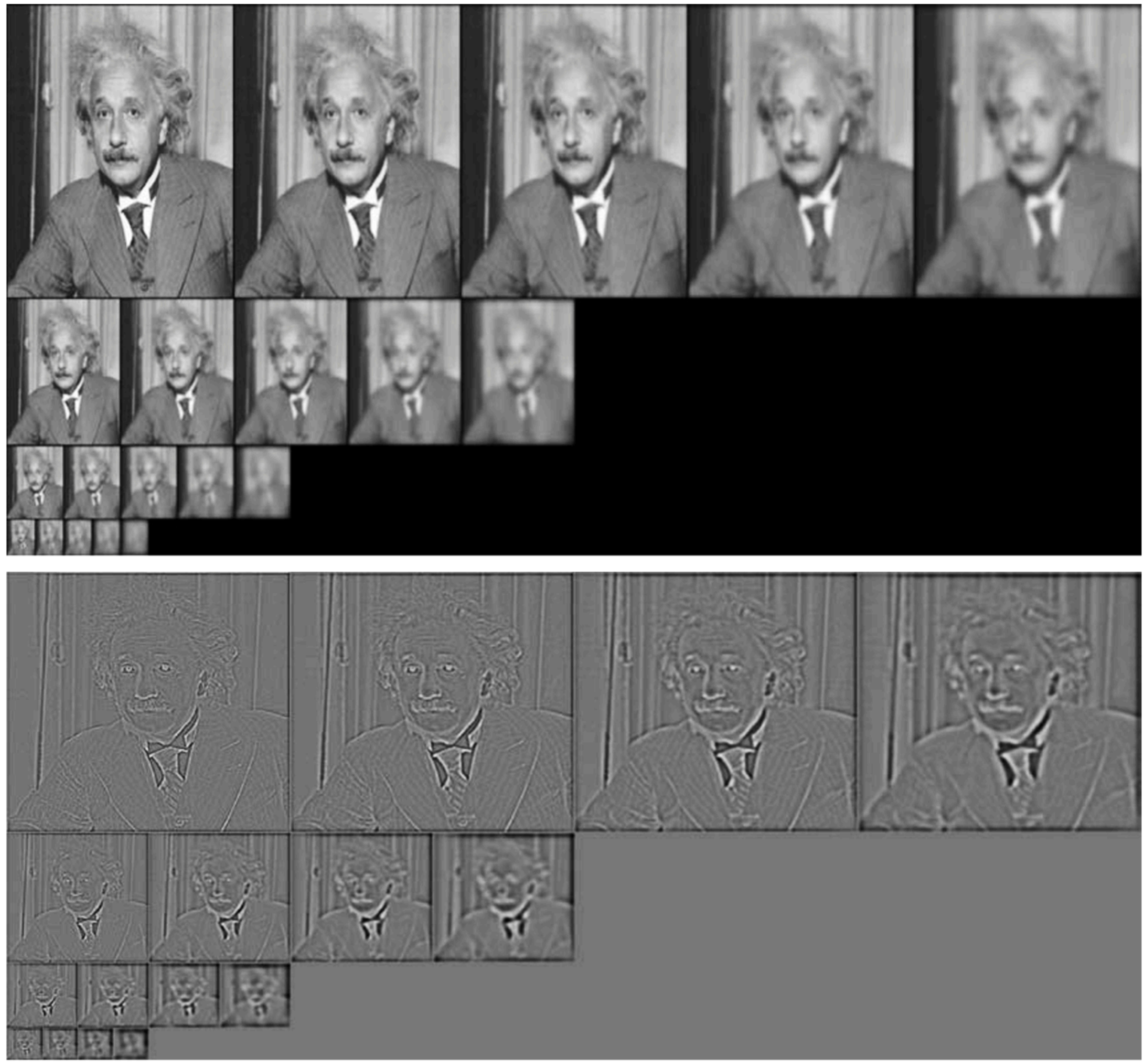
How to obtain scale invariance?

From SIFT: Use resolution pyramid.

But instead of reducing the images, reduce the box filters!



https://medium.com/@deepanshut041/introduction-to-surf-speeded-up-robust-features-c7396d6e7c4e

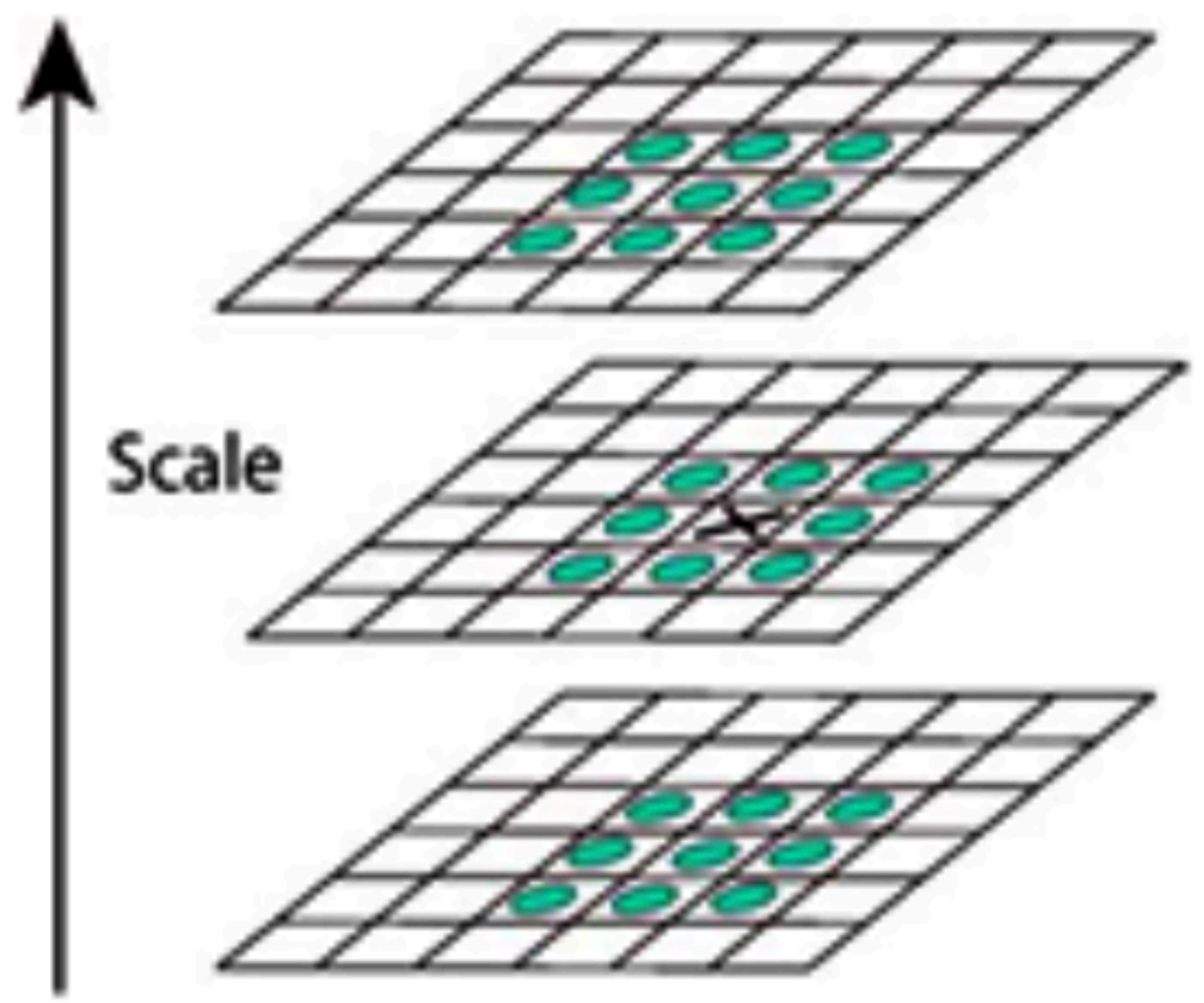


Speeded-Up Robust Features (SURF)

Lowe's

Faster Keypoint Detection
Non-maximal Suppression

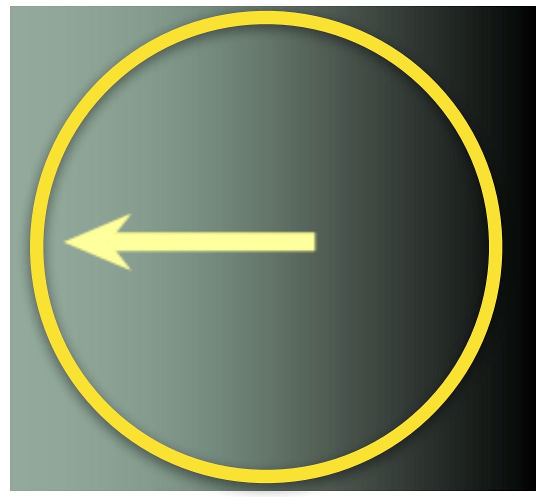
Good SURF keypoints present the highest $det(H)$ values among their immediate (3 x 3 x 3) scale-space neighborhood.



Speeded-Up Robust Features (SURF)

Faster Keypoint Detection
Orientation Assignment

Compute the gradient angle for all the pixels within the keypoint neighborhood considering its scale σ (from the Hessian matrix).



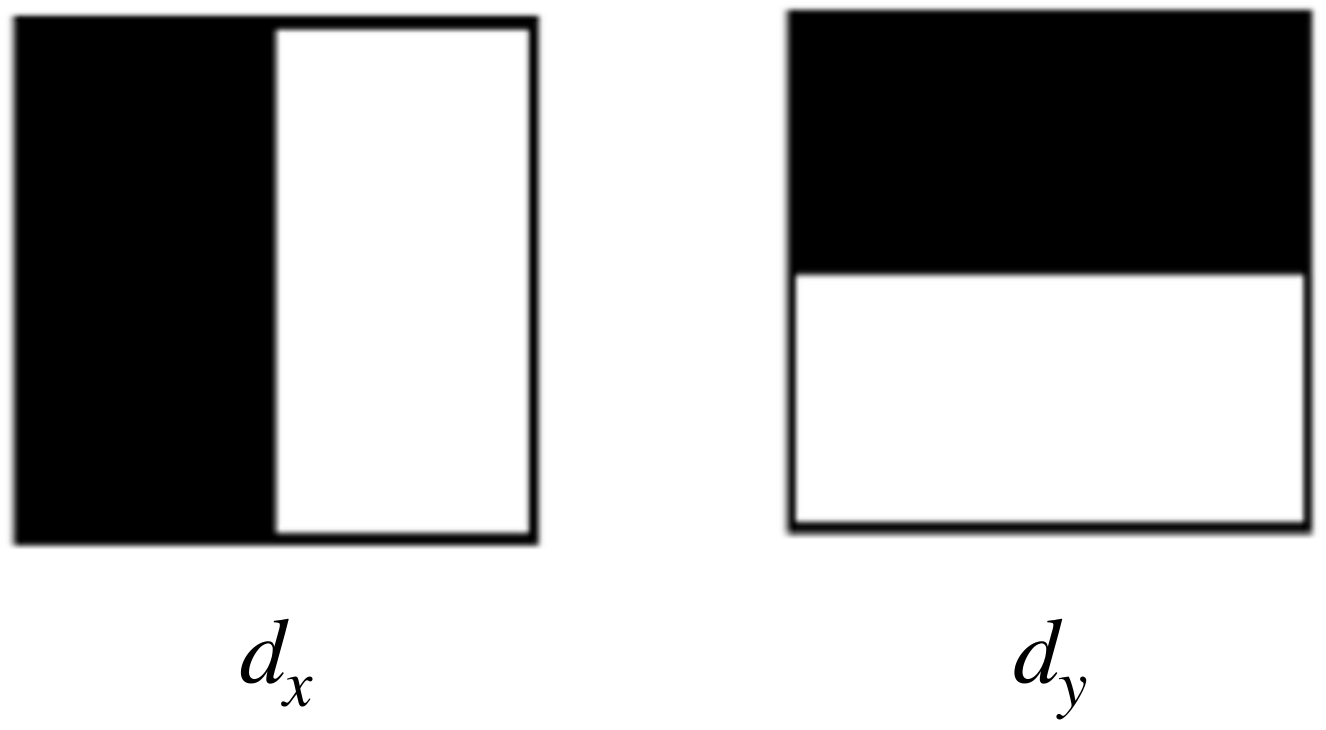
LOYOLA UNIVERSITY CHICAGO

Speeded-Up Robust Features (SURF)

Faster Keypoint Detection
Orientation Assignment

Compute the gradient angle for all the pixels within the keypoint neighborhood considering its scale σ (from the Hessian matrix).

Use box filters d_x and d_y with scale proportional to σ.



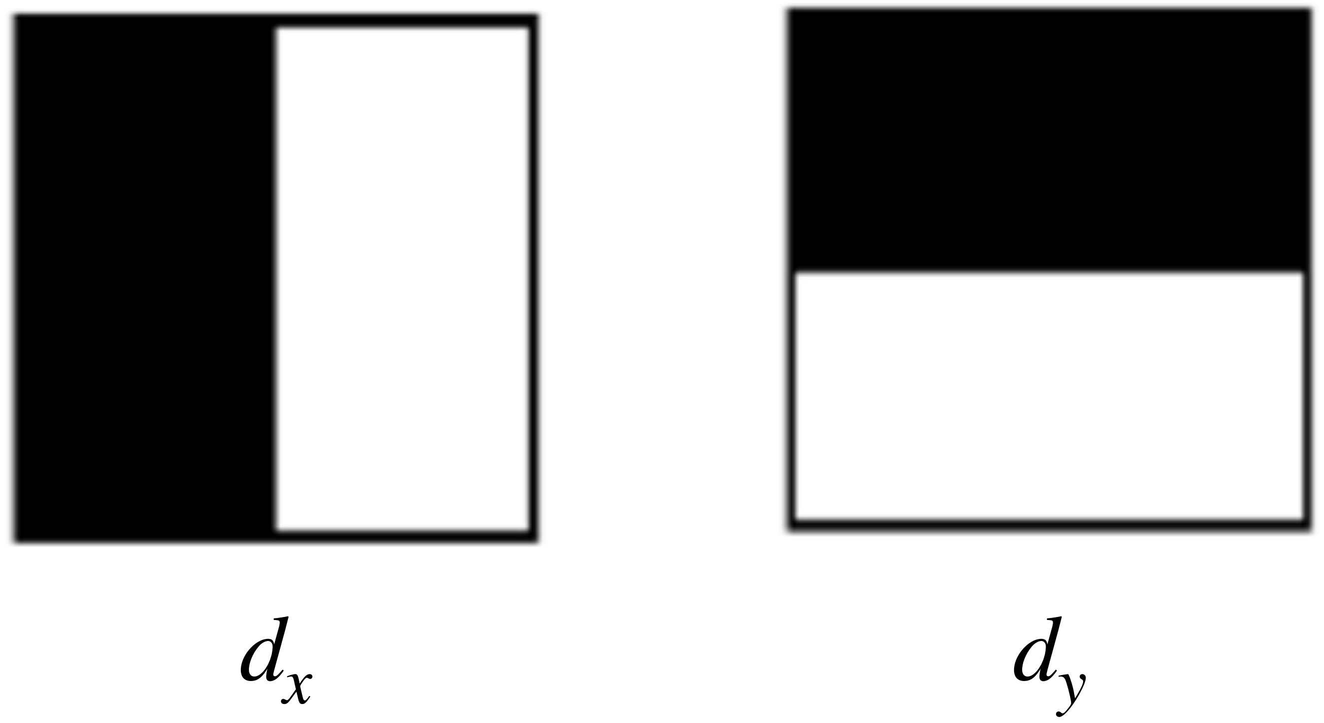
Speeded-Up Robust Features (SURF)

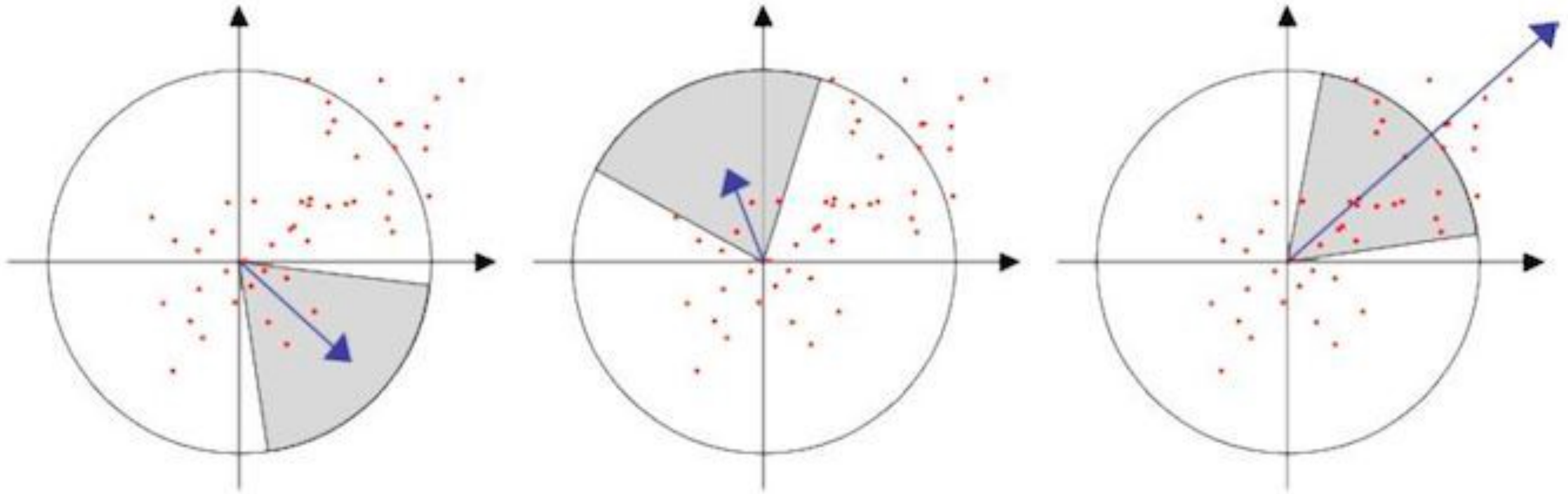
Faster Keypoint Detection
Orientation Assignment

Compute the gradient angle for all the pixels within the keypoint neighborhood considering its scale σ (from the Hessian matrix).

Use box filters d_x and d_y with scale proportional to σ.

Create an angle histogram with 6 bins. Take the dominant angle as the keypoint orientation.





https://medium.com/@deepanshut041/introduction-to-surf-speeded-up-robust-features-c7396d6e7c4e

LOYOLA UNIVERSITY CHICAGO

Speeded-Up Robust Features (SURF)

Faster Keypoint Detection

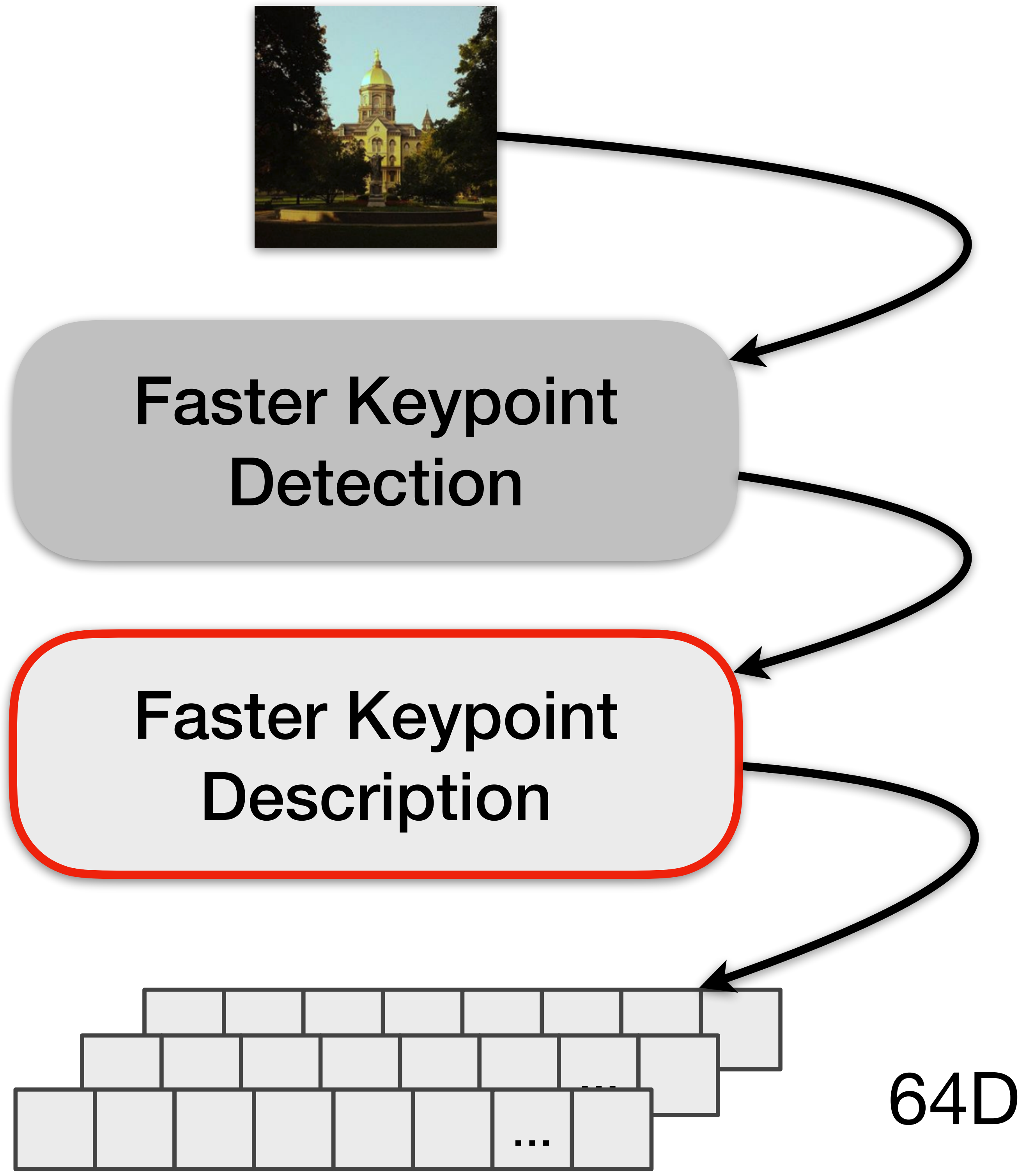
SURF keypoints have:
1. Location *(x,y)*
2. Scale (σ from Hessian matrix)
3. Orientation (dominant local gradient angle)
4. Strength ($det(H)$)

https://docs.opencv.org/3.4/df/dd2/tutorial_py_surf_intro.html

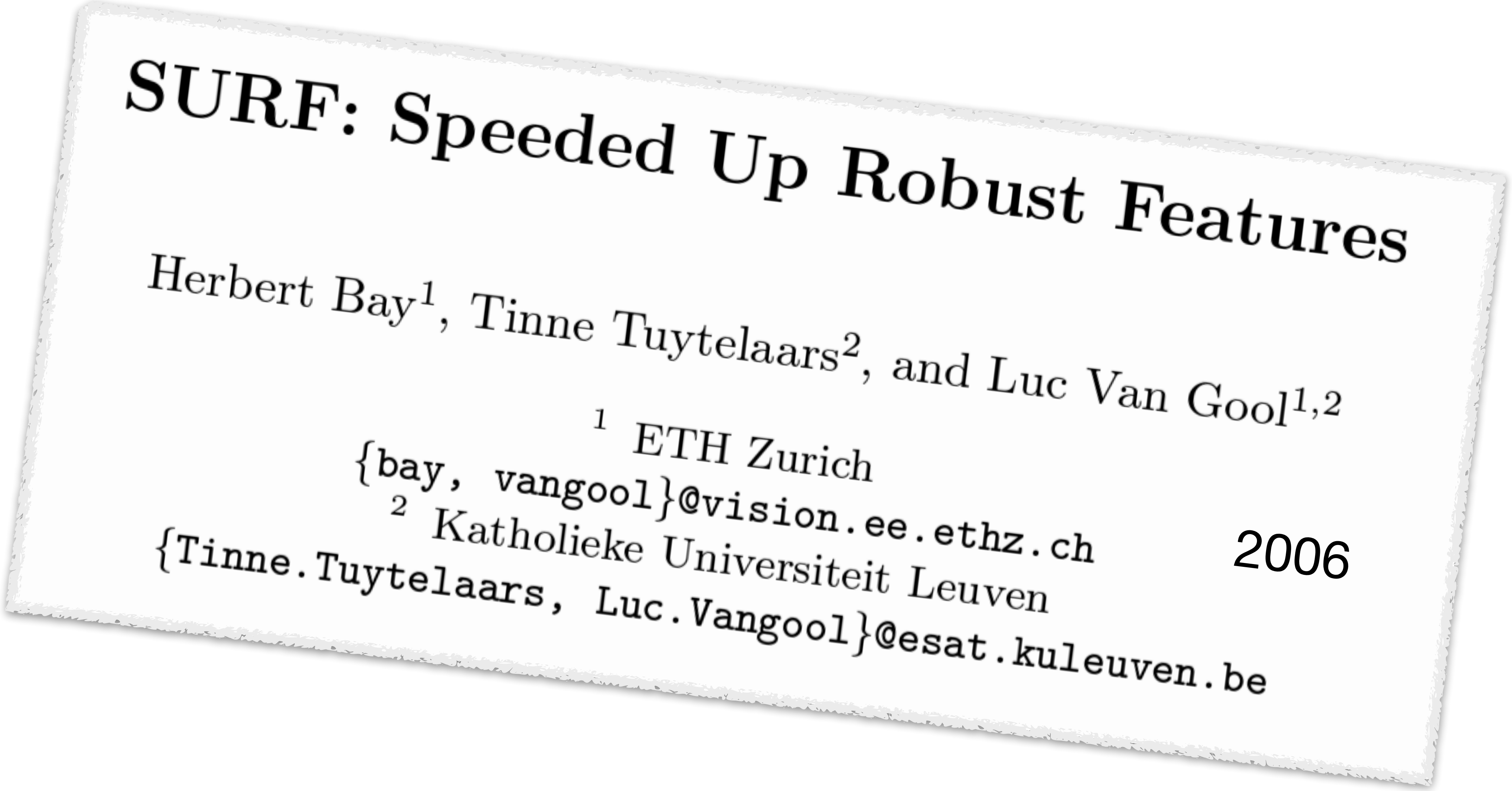
LOYOLA UNIVERSITY CHICAGO

Speeded-Up Robust Features (SURF)

Stages



64D feature vectors


SURF: Speeded Up Robust Features

Herbert Bay[1], Tinne Tuytelaars[2], and Luc Van Gool[1,2]

[1] ETH Zurich
{bay, vangool}@vision.ee.ethz.ch
[2] Katholieke Universiteit Leuven
{Tinne.Tuytelaars, Luc.Vangool}@esat.kuleuven.be

2006

Speeded-Up Robust Features (SURF)

Faster Keypoint Description

For each keypoint, rotate the image according to the keypoint orientation.

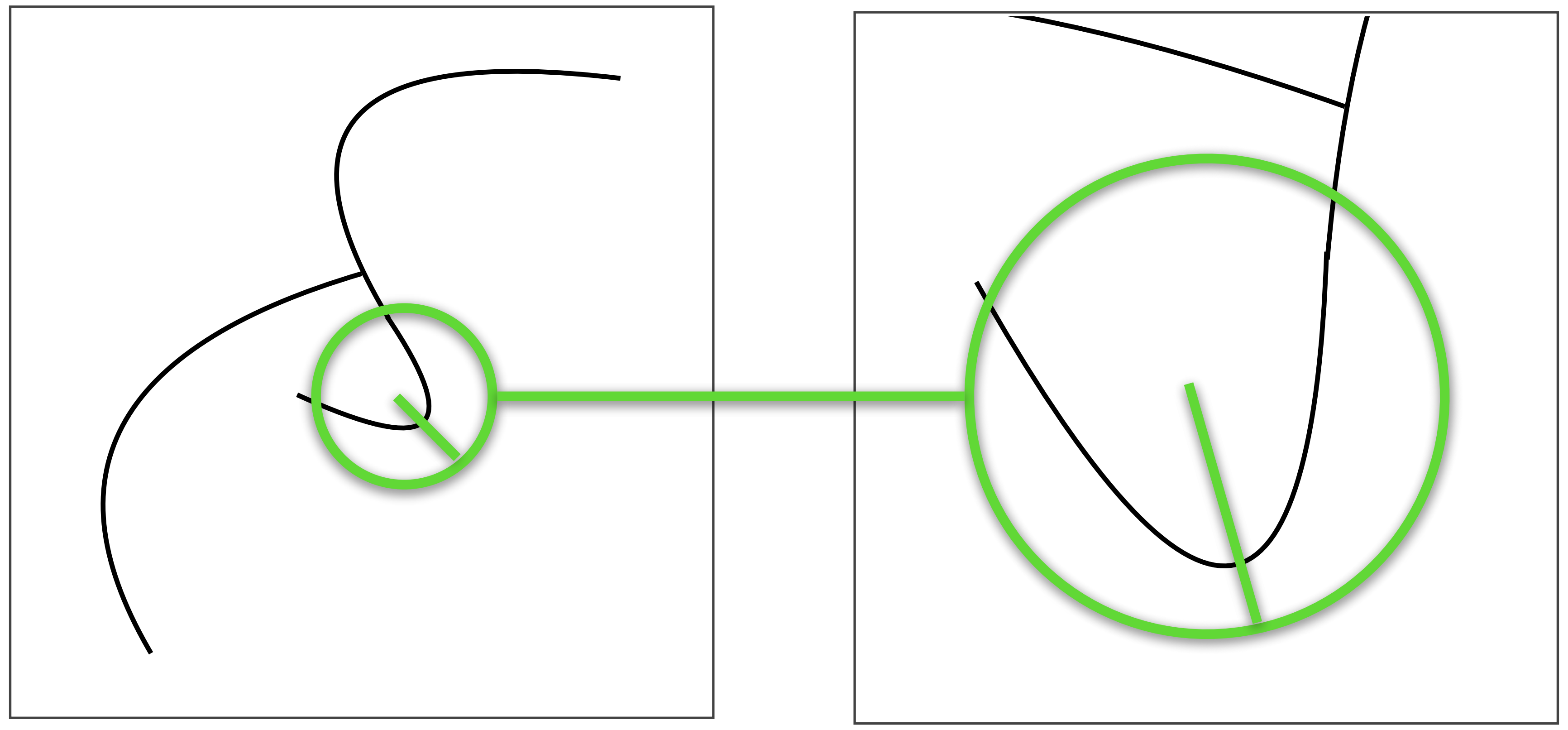
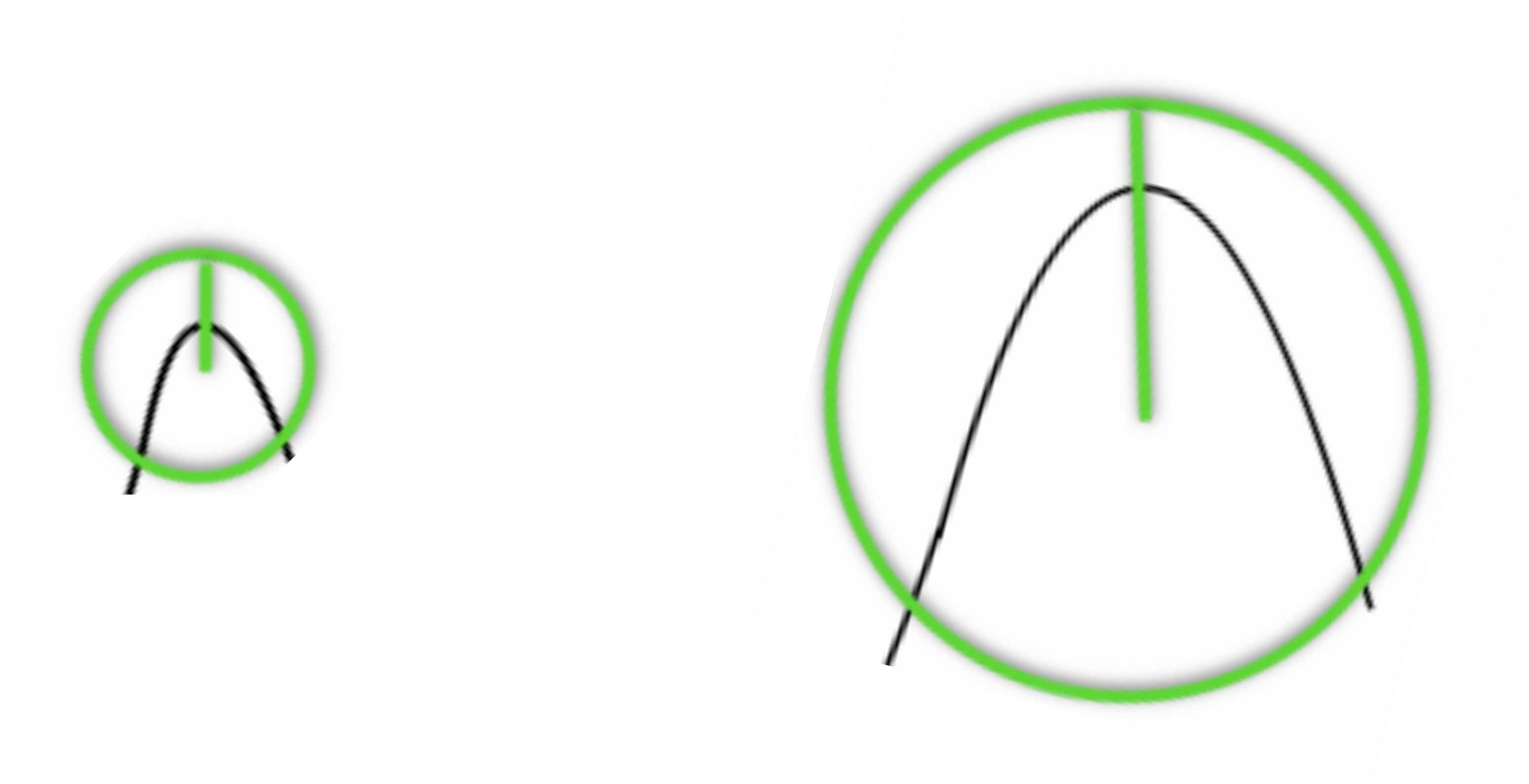


Image 1

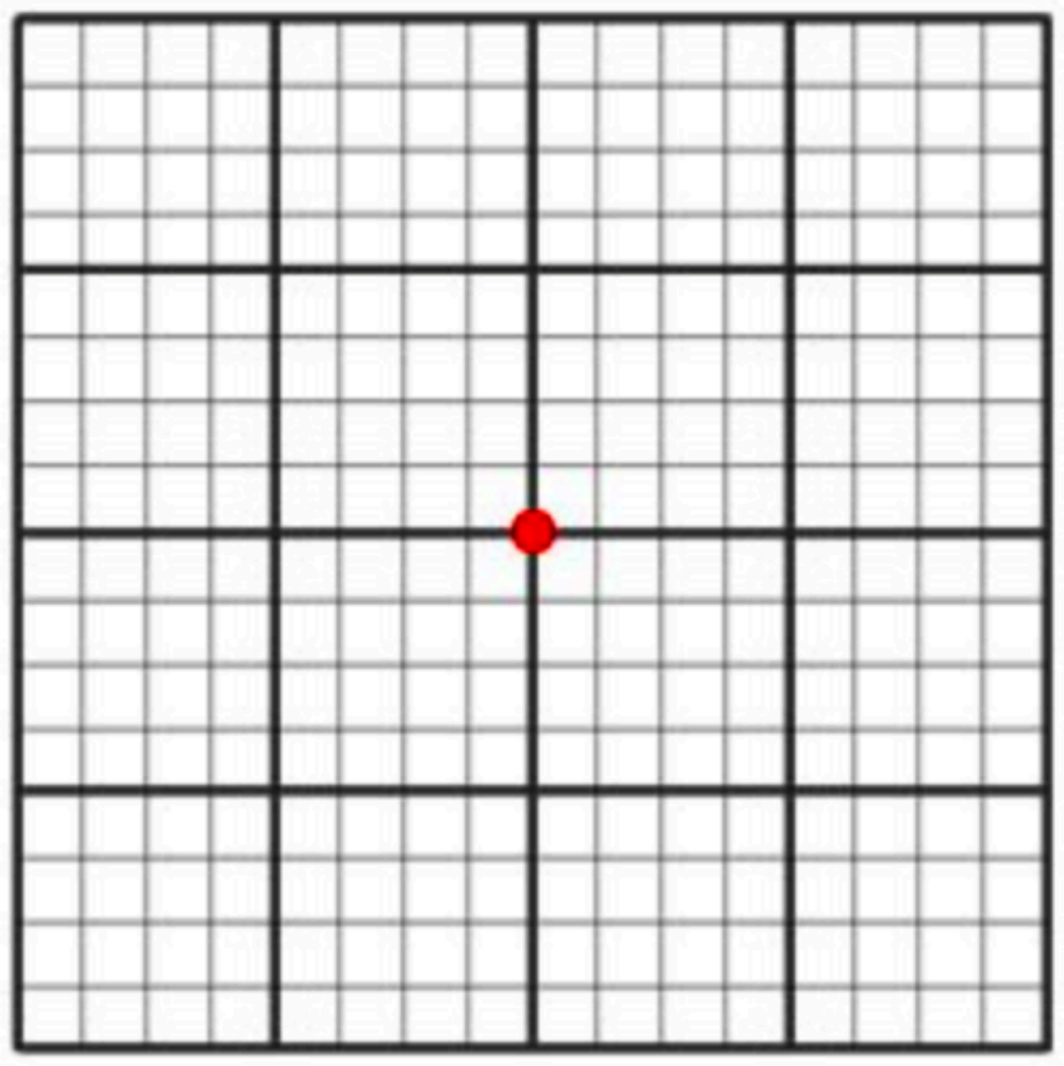
Image 2



Speeded-Up Robust Features (SURF)

Faster Keypoint Description

For each rotated keypoint, sample a *4 x 4* window on its neighborhood, according to the keypoint scale.



Speeded-Up Robust Features (SURF)

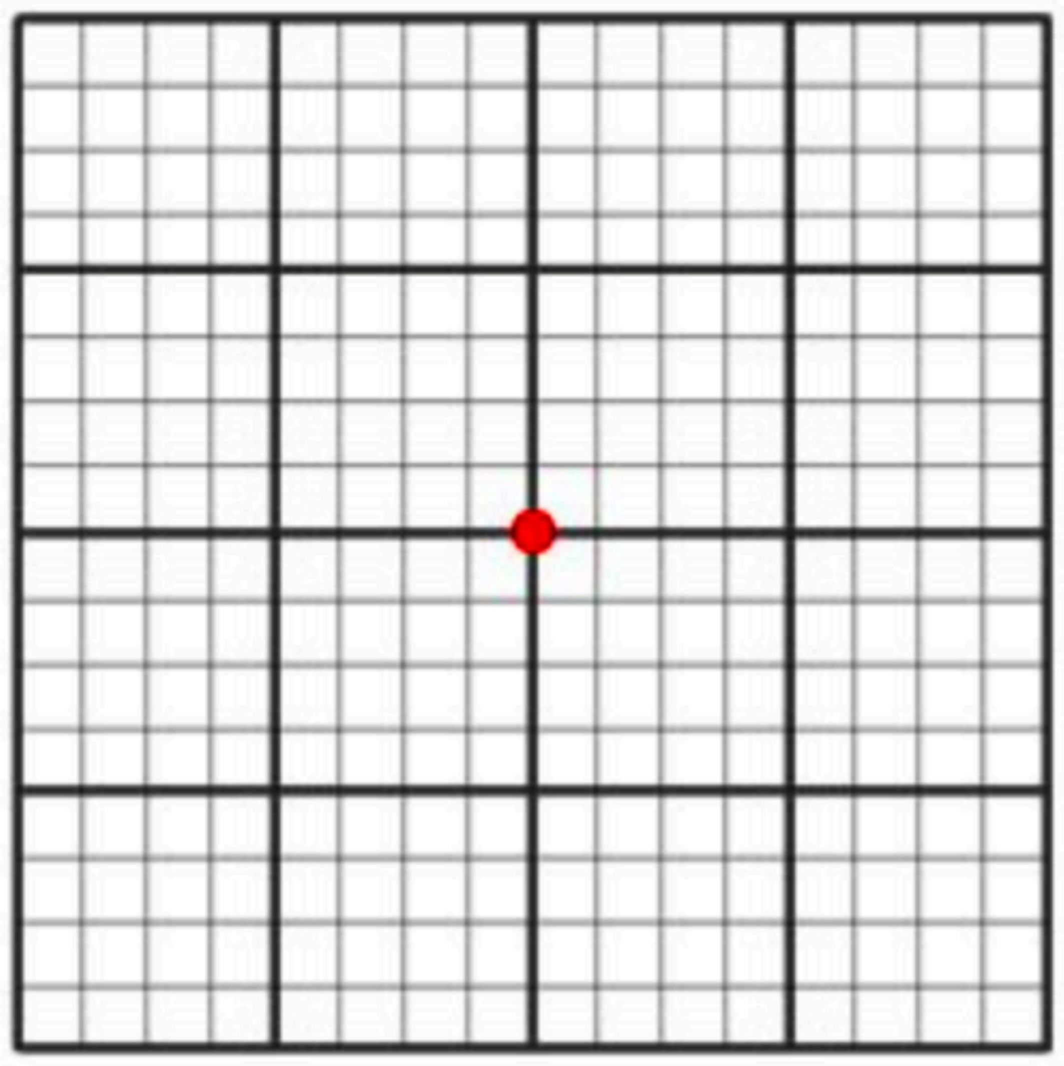
Faster Keypoint Description

For each rotated keypoint, sample a *4 x 4* window on its neighborhood, according to the keypoint scale.

For each one of the *4 x 4* cells, compute 4 sums:
(1) $\sum d_x$, (2) $\sum |d_x|$, (3) $\sum d_y$, and (4) $\sum |d_y|$.

d_x

d_y



Speeded-Up Robust Features (SURF)

Faster Keypoint Description

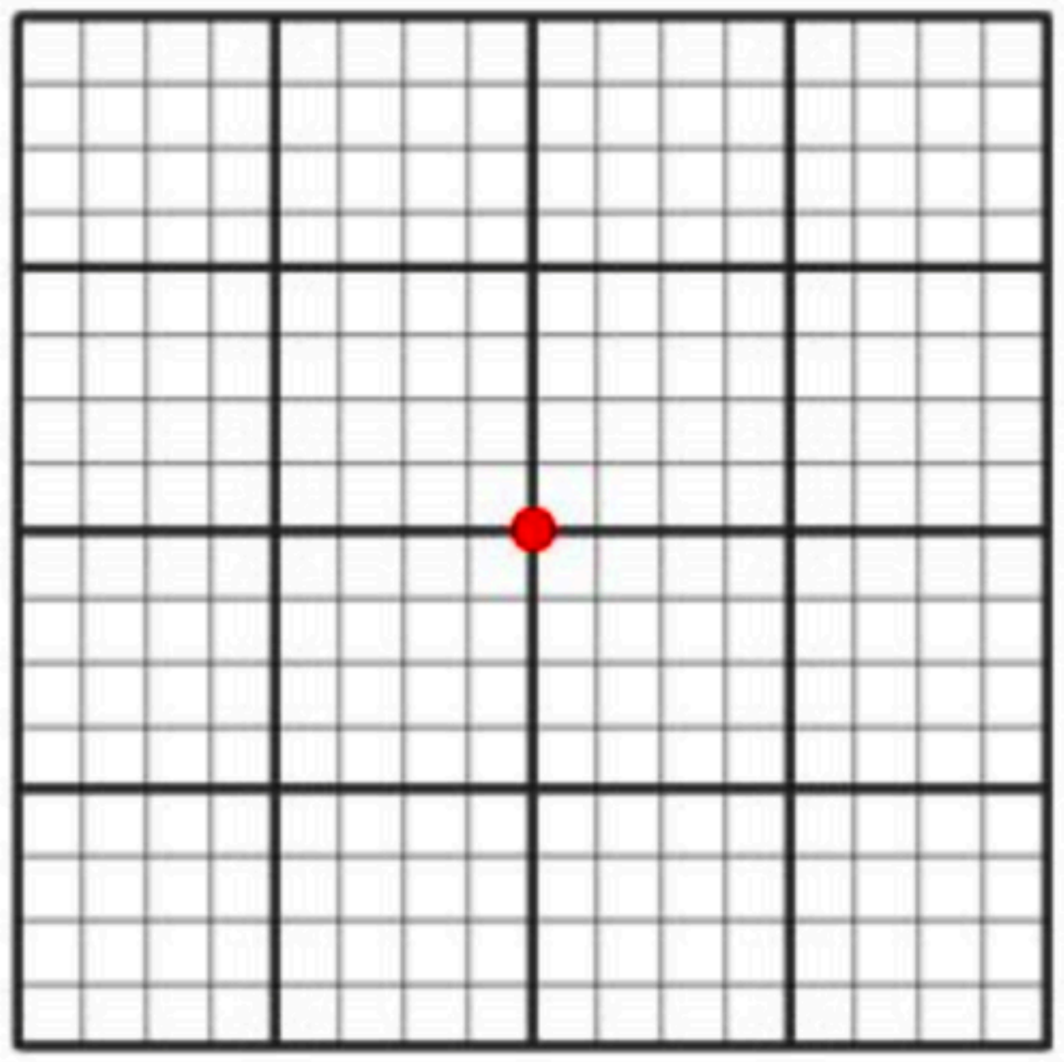
For each rotated keypoint, sample a *4 x 4* window on its neighborhood, according to the keypoint scale.

For each one of the *4 x 4* cells, compute 4 sums:
(1) $\sum d_x$, (2) $\sum |d_x|$, (3) $\sum d_y$, and (4) $\sum |d_y|$.

Fill out a feature vector with the *4 x 4 x 4 = 64* values.

d_x

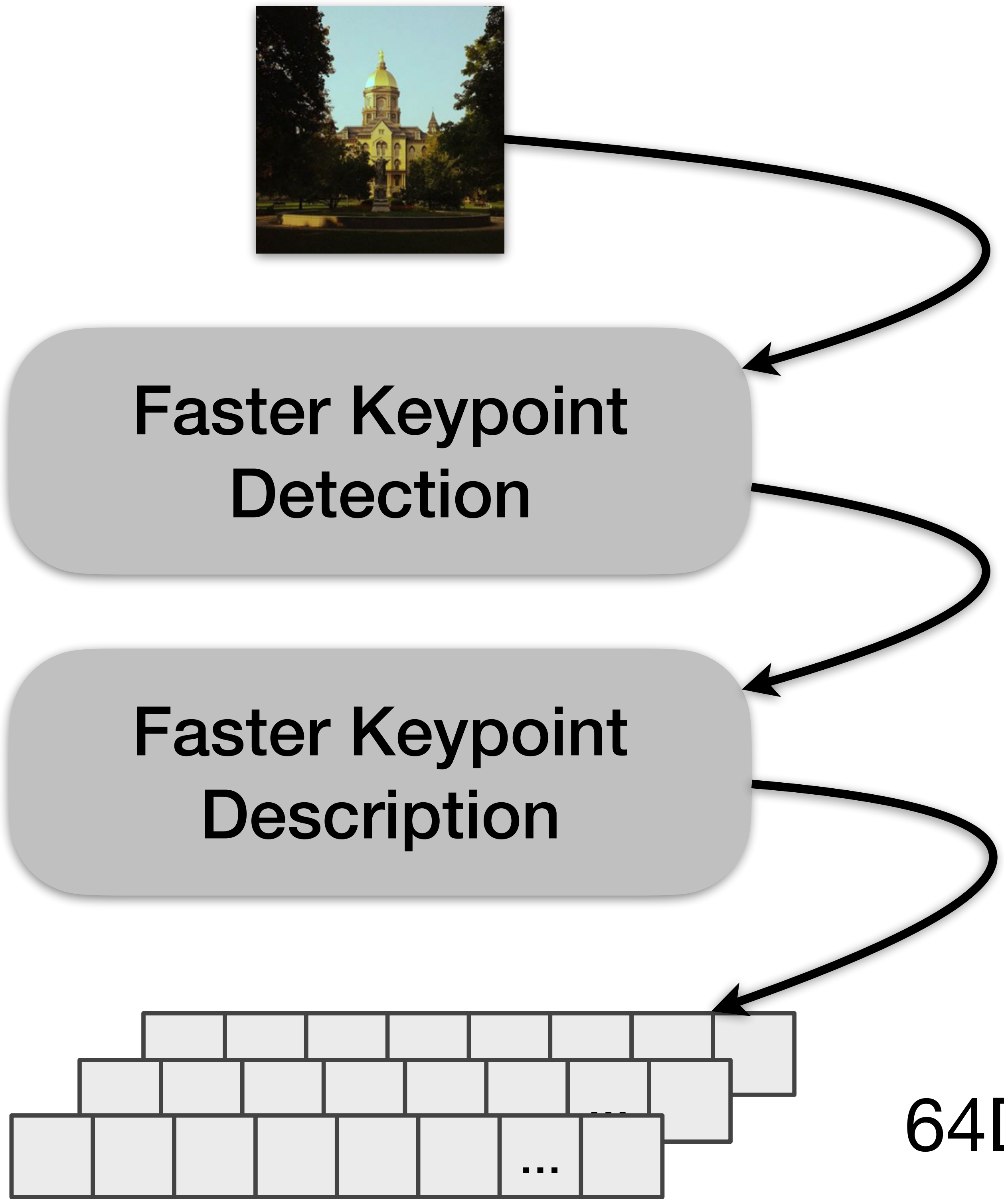
d_y

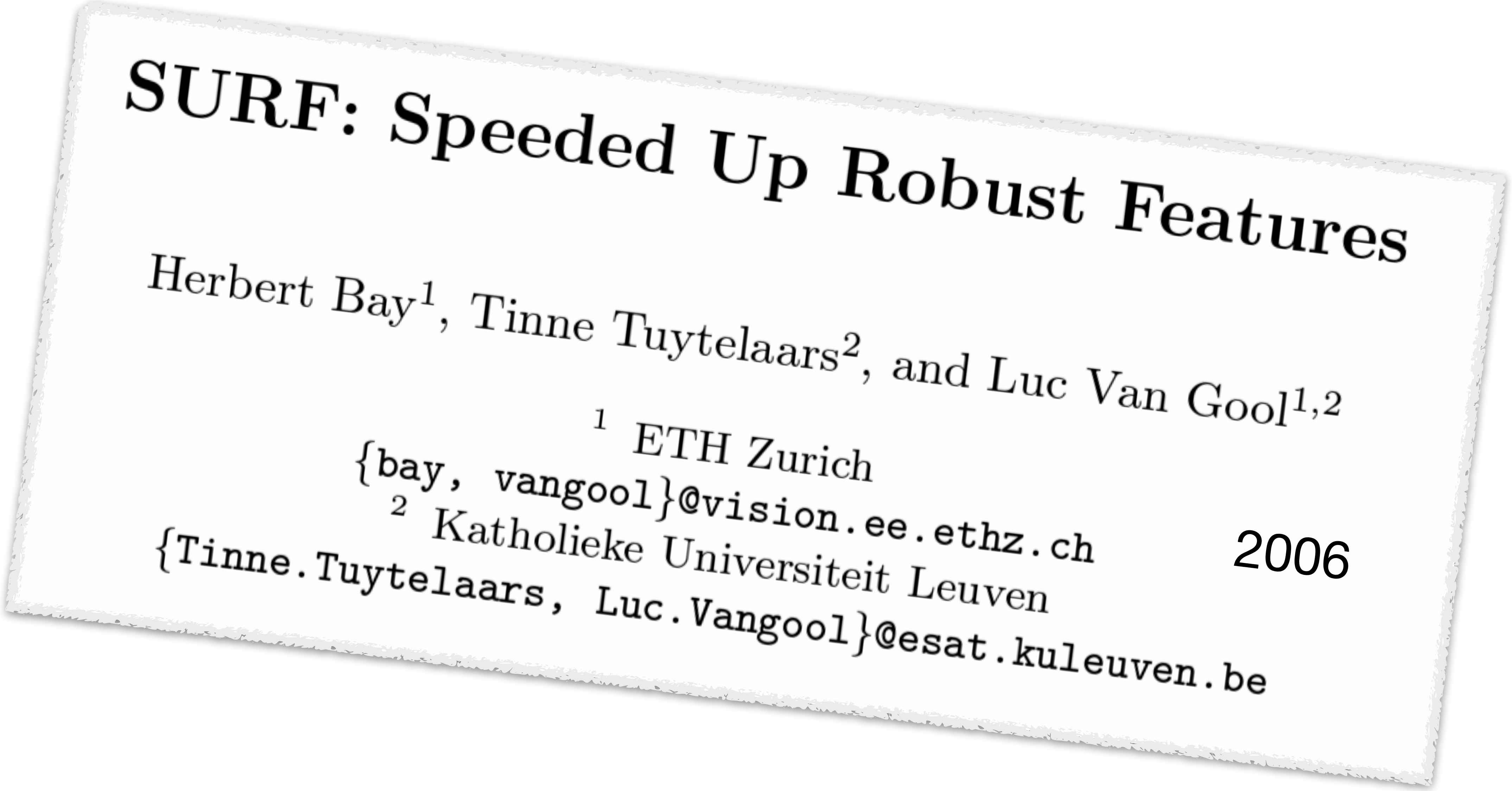


LOYOLA UNIVERSITY CHICAGO

Speeded-Up Robust Features (SURF)

Stages




SURF: Speeded Up Robust Features

Herbert Bay[1], Tinne Tuytelaars[2], and Luc Van Gool[1,2]

[1] ETH Zurich
{bay, vangool}@vision.ee.ethz.ch
[2] Katholieke Universiteit Leuven
{Tinne.Tuytelaars, Luc.Vangool}@esat.kuleuven.be

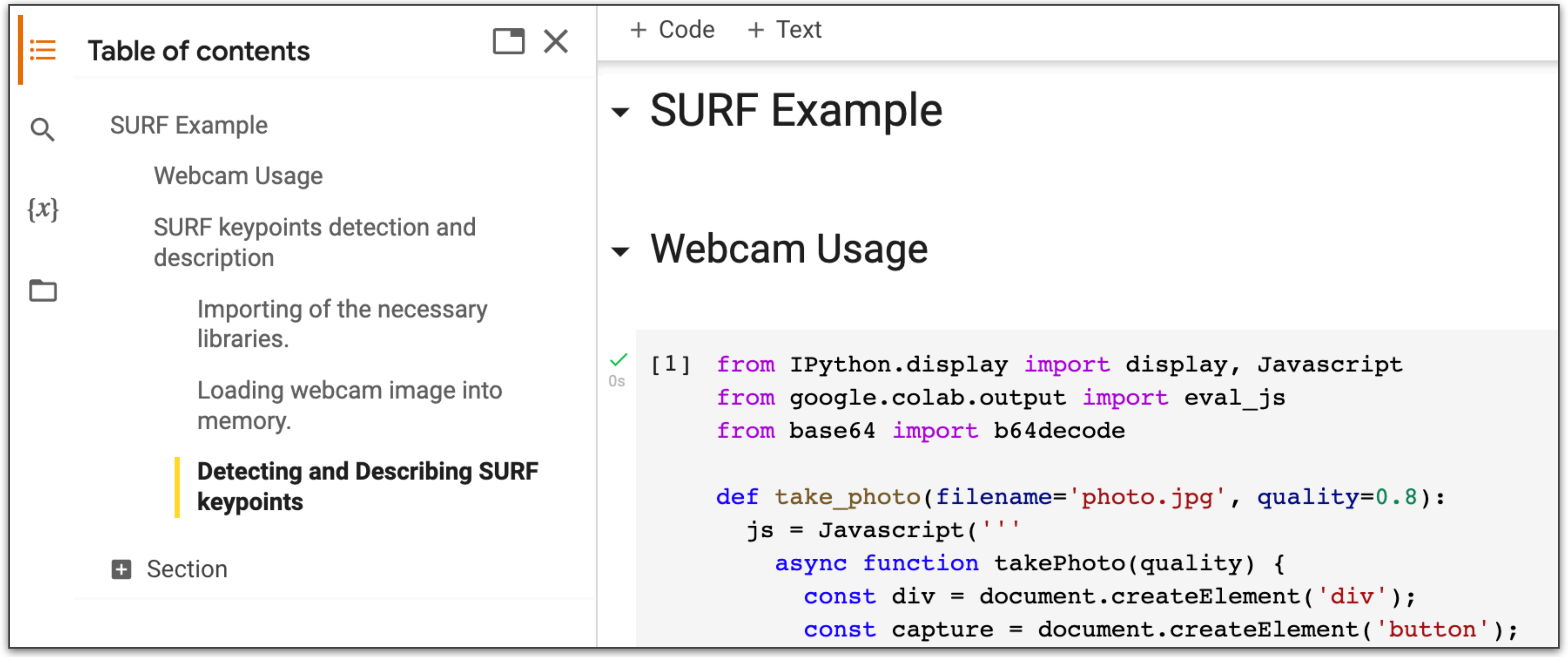
2006

64D feature vectors, compare two vectors with L2- or cosine distance.

Speeded-Up Robust Features (SURF)

Example

https://bit.ly/3qOj3jx



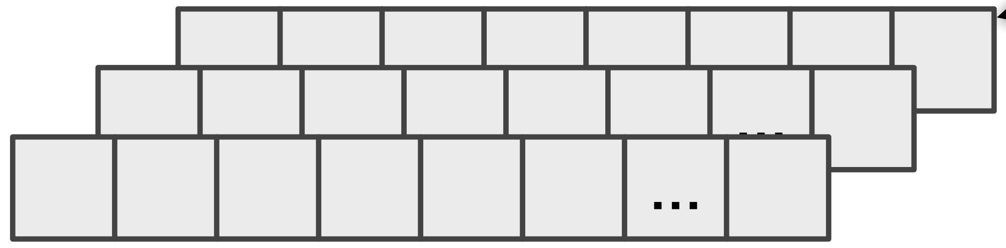
Deep Local Features (DELF)

Stages

Dense Feature Extraction

Keypoint Selection

Dimensionality Reduction



40D feature vectors

Large-Scale Image Retrieval with Attentive Deep Local Features

Hyeonwoo Noh† Andre Araujo* Jack Sim* Tobias Weyand* Bohyung Han†

†POSTECH, Korea
{shgusdngogo,bhhan}@postech.ac.kr

*Google Inc.
{andrearaujo,jacksim,weyand}@google.com

2016

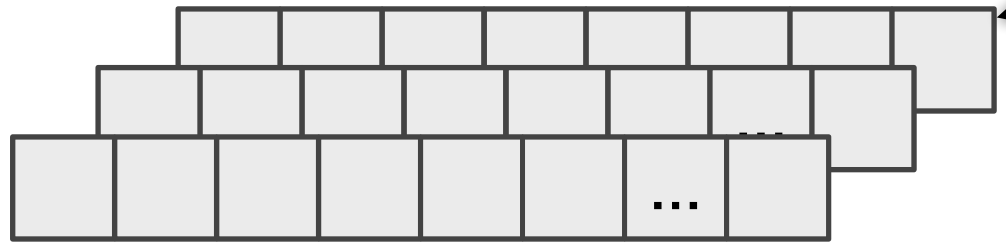
Deep Local Features (DELF)

Stages

Dense Feature Extraction

Keypoint Selection

Dimensionality Reduction



40D feature vectors

Large-Scale Image Retrieval with Attentive Deep Local Features

Hyeonwoo Noh† Andre Araujo* Jack Sim* Tobias Weyand* Bohyung Han†

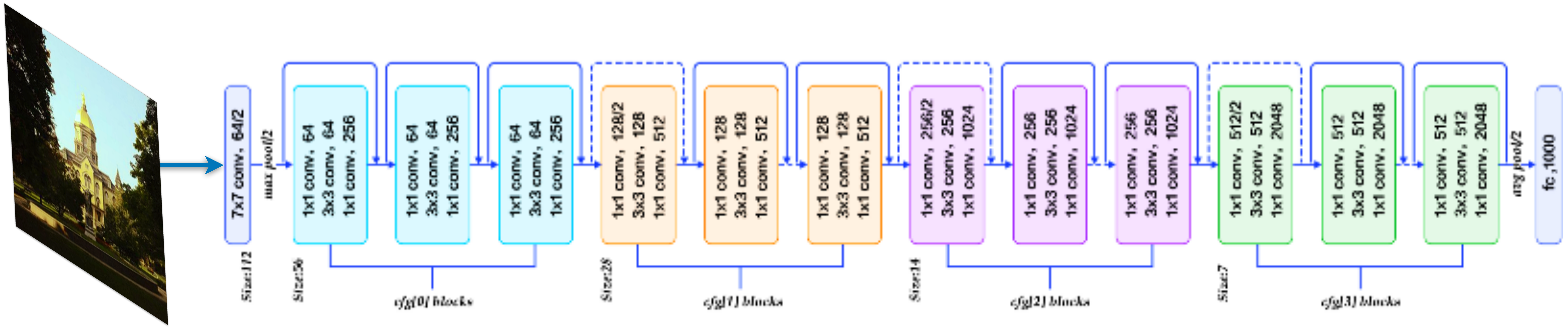
†POSTECH, Korea
{shgusdngogo,bhhan}@postech.ac.kr

*Google Inc.
{andrearaujo,jacksim,weyand}@google.com

2016

Deep Local Features (DELF)

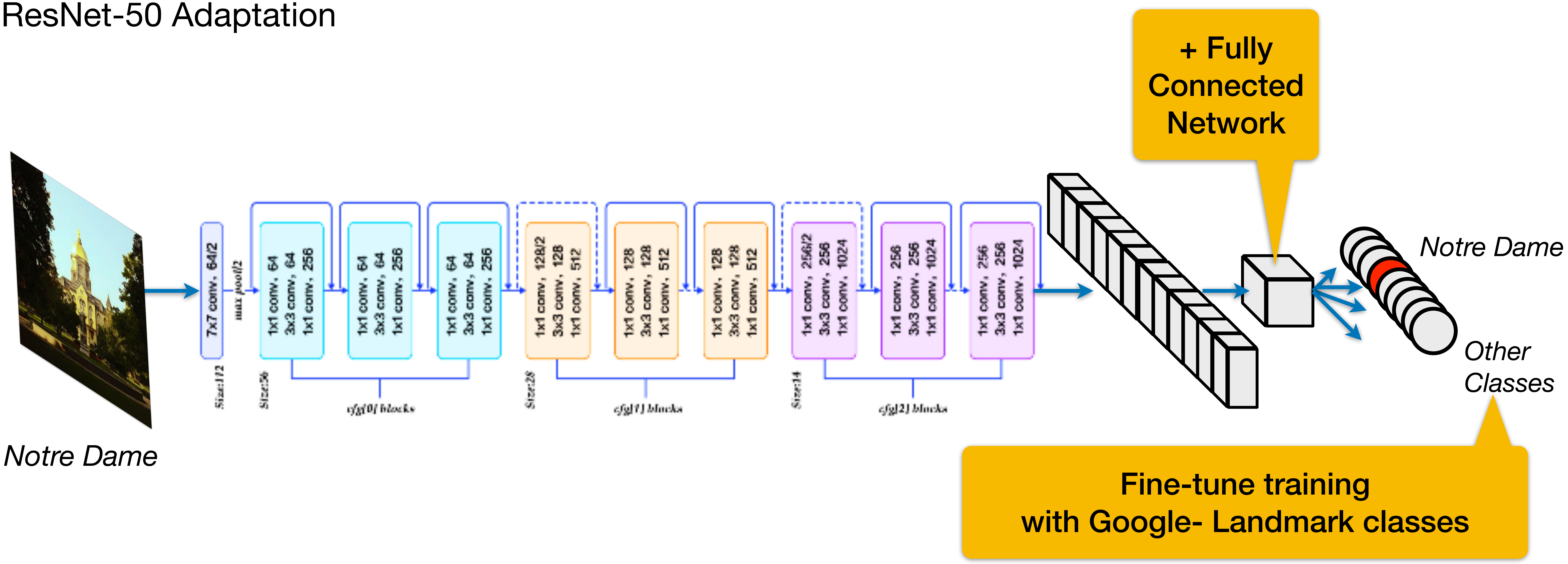
Dense Feature Extraction
Starting Point: ResNet-50



ResNet-50

Deep Local Features (DELF)

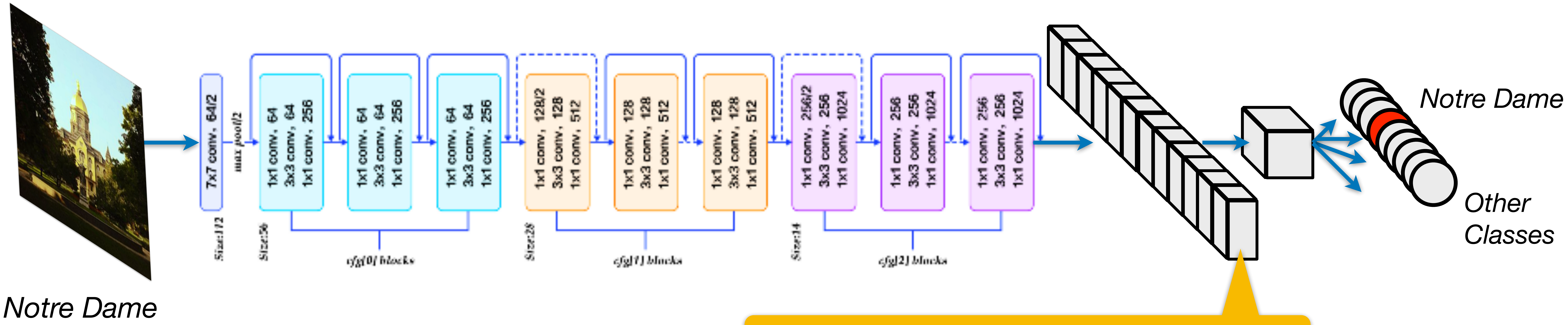
Dense Feature Extraction
ResNet-50 Adaptation



LOYOLA UNIVERSITY CHICAGO

Deep Local Features (DELF)

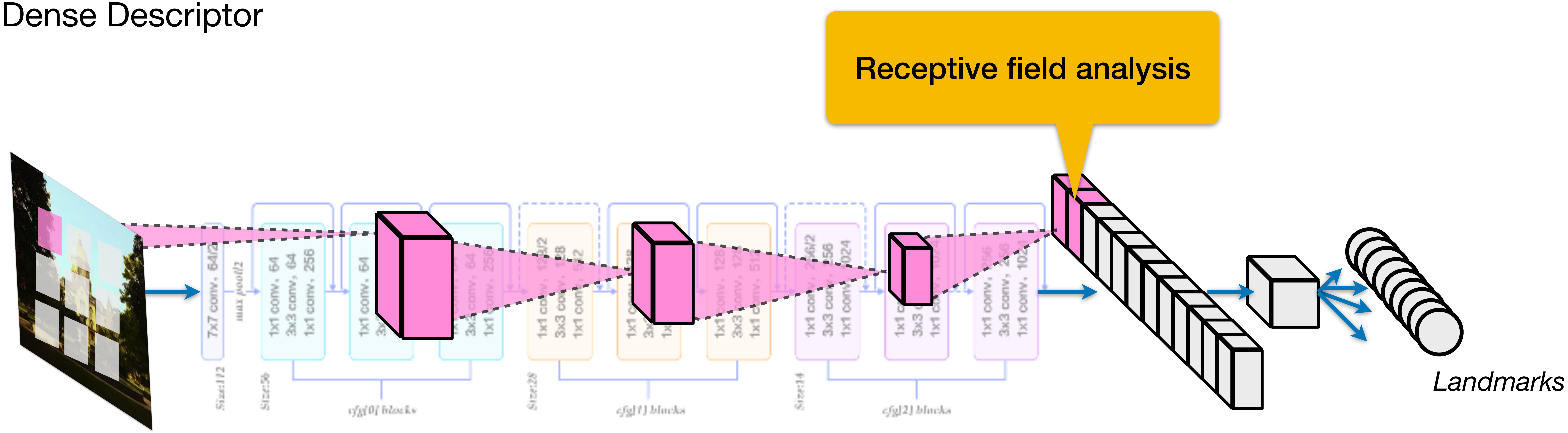
Dense Feature Extraction
How to convert global description to dense description?



Solution: receptive field analysis

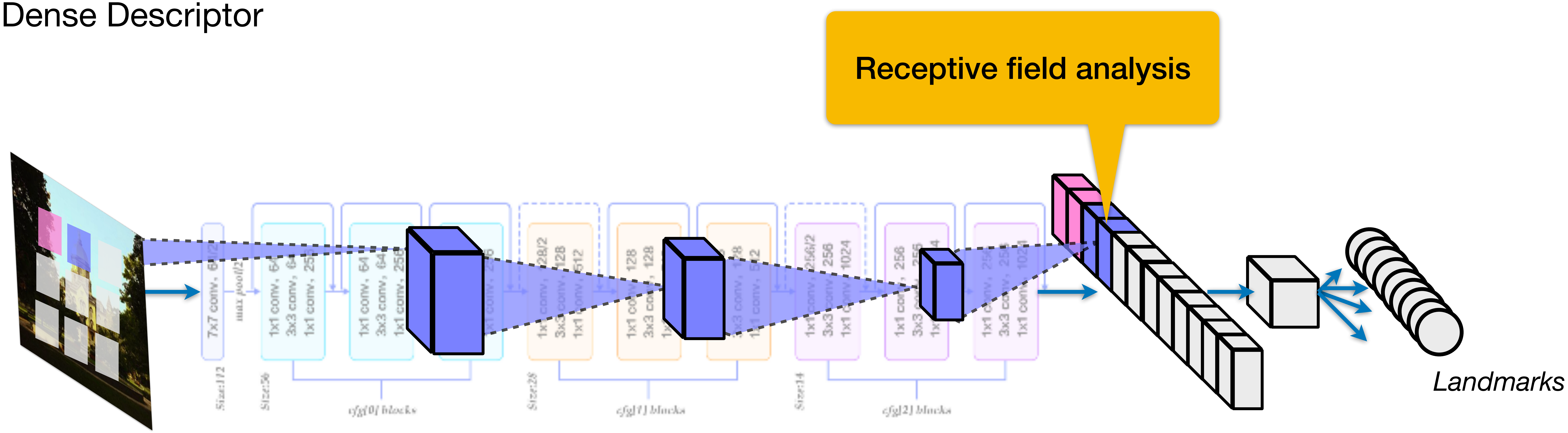
Deep Local Features (DELF)

Dense Feature Extraction
Dense Descriptor



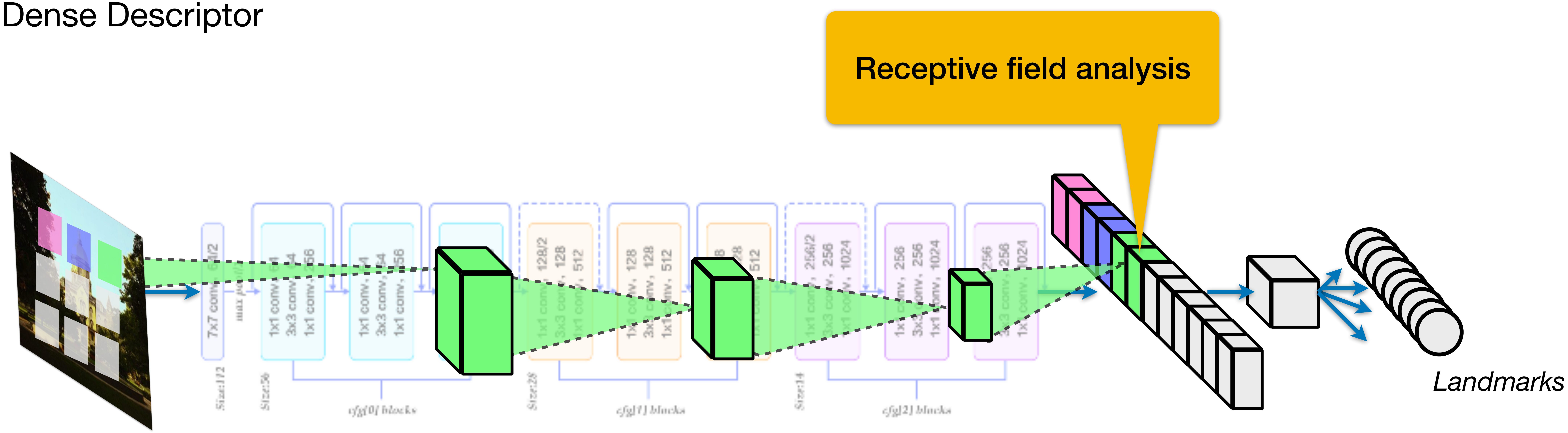
Deep Local Features (DELF)

Dense Feature Extraction
Dense Descriptor



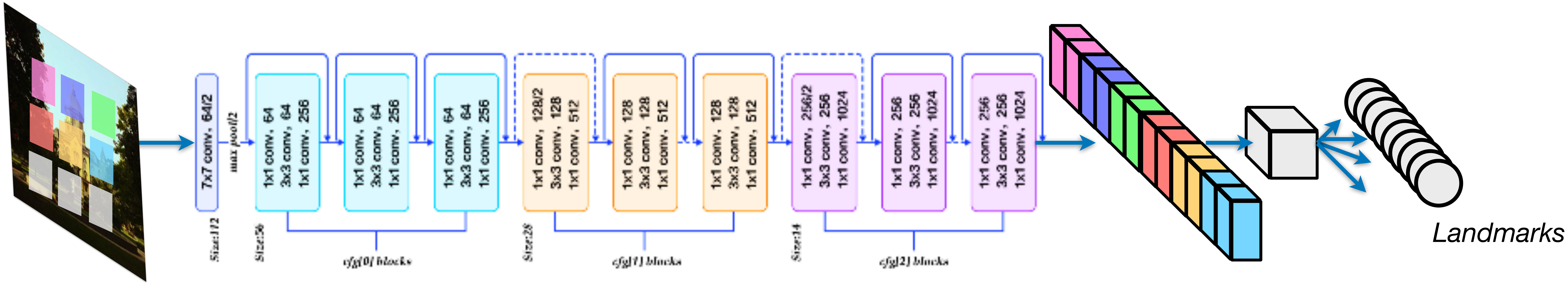
Deep Local Features (DELF)

Dense Feature Extraction
Dense Descriptor



Deep Local Features (DELF)

Dense Feature Extraction
Dense Descriptor

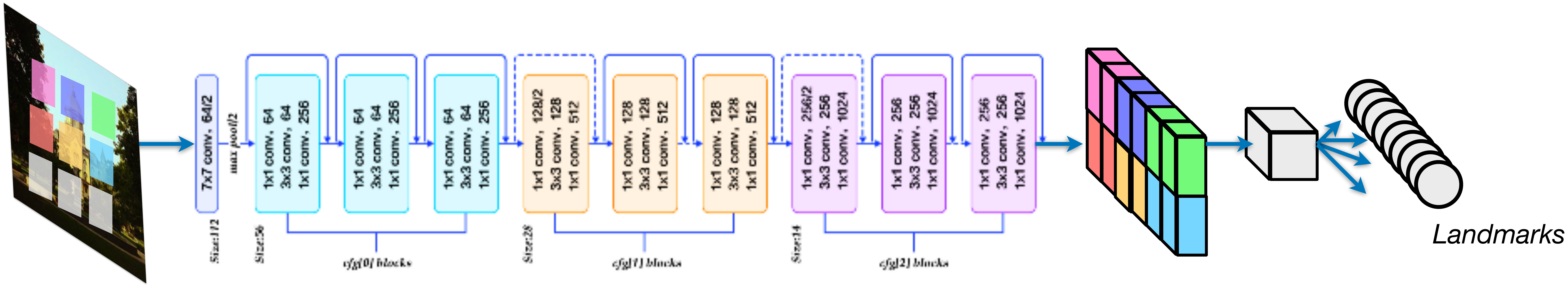


LOYOLA UNIVERSITY CHICAGO

Deep Local Features (DELF)

Dense Feature Extraction
Dense Descriptor

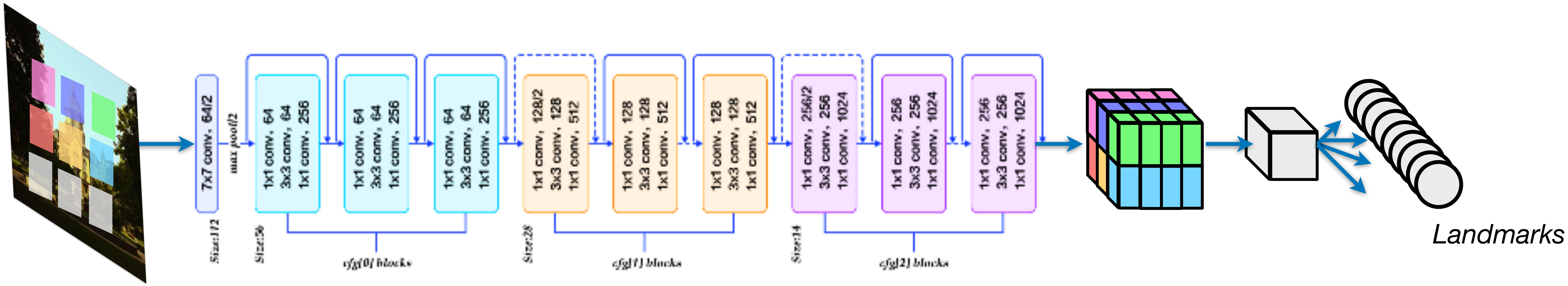
Logical meaning: one feature vector for each densely sampled image region.



Deep Local Features (DELF)

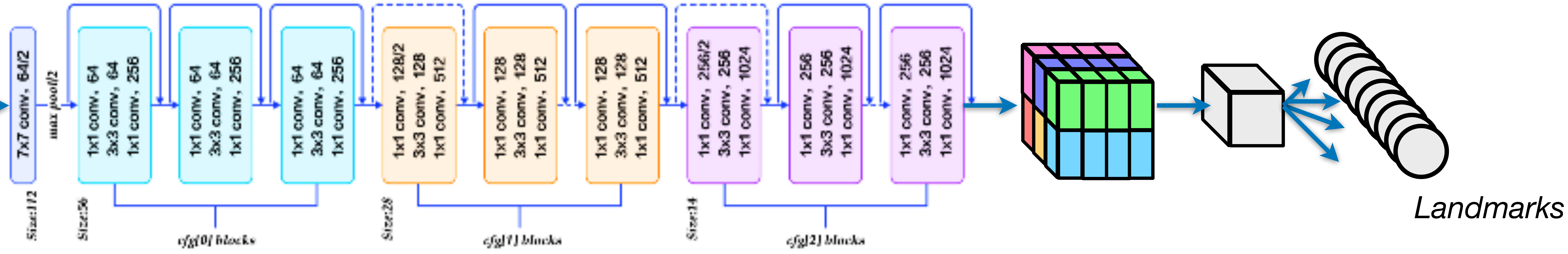
Dense Feature Extraction
Dense Descriptor

Logical meaning: one feature vector for each densely sampled image region.



Deep Local Features (DELF)

Dense Feature Extraction
How to deal with multiple resolutions?



Solution: train with resolution pyramid.

Each level is presented independently.

LOYOLA UNIVERSITY CHICAGO

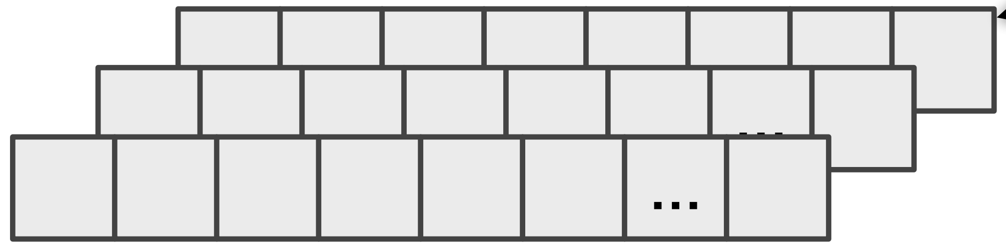
Deep Local Features (DELF)

Stages

Dense Feature Extraction

Keypoint Selection

Dimensionality Reduction



40D feature vectors


Large-Scale Image Retrieval with Attentive Deep Local Features

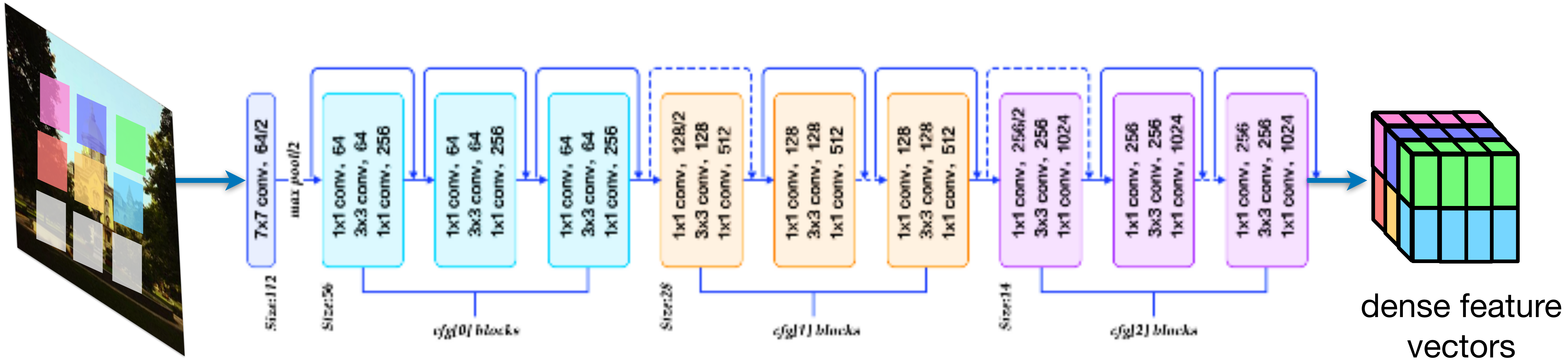
Hyeonwoo Noh† Andre Araujo* Jack Sim* Tobias Weyand* Bohyung Han†
†POSTECH, Korea *Google Inc.
{shgusdngogo,bhhan}@postech.ac.kr {andrearaujo,jacksim,weyand}@google.com

2016

LOYOLA UNIVERSITY CHICAGO

Deep Local Features (DELF)

Attention-based Keypoint Selection
How to convert dense description to keypoint selection?

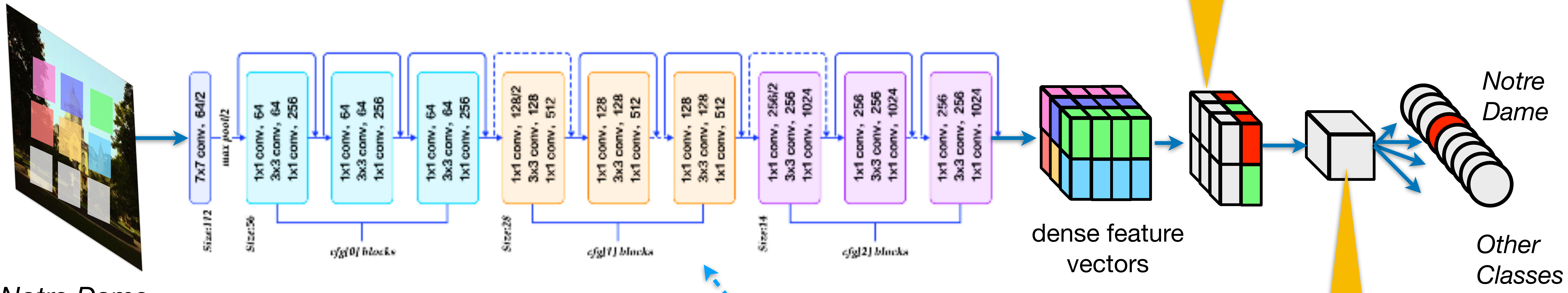


LOYOLA UNIVERSITY CHICAGO

Deep Local Features (DELF)

Attention-based Keypoint Selection
How to convert the dense description to keypoint selection?

Solution: train an attention network.

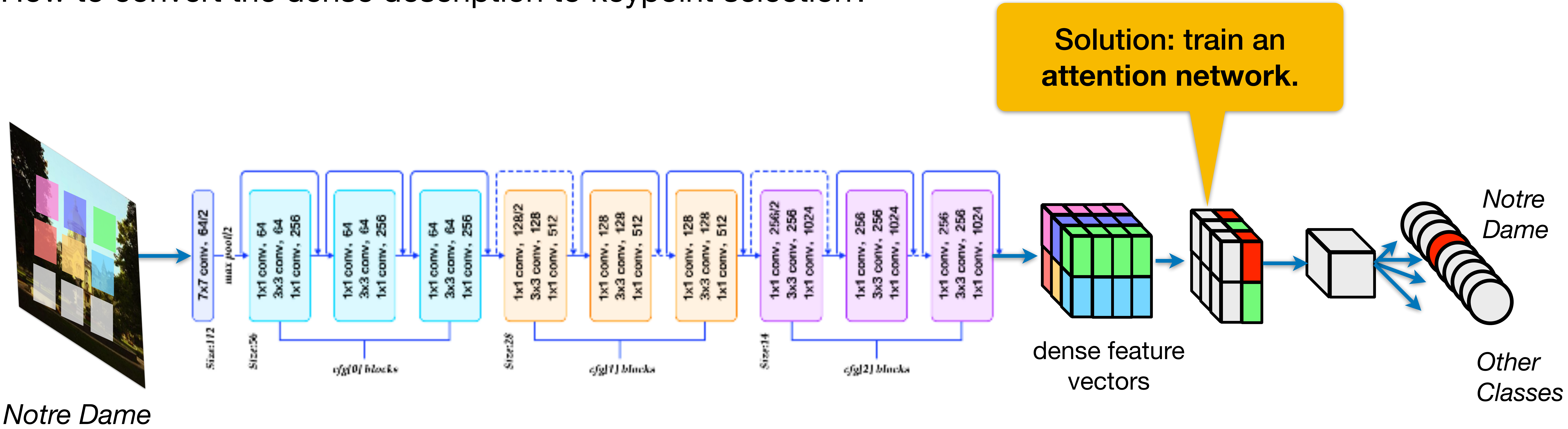


Notre Dame

The Fully Connected Network is trained again but with focus on the attention network. ResNet-50 is left as is.

Deep Local Features (DELF)

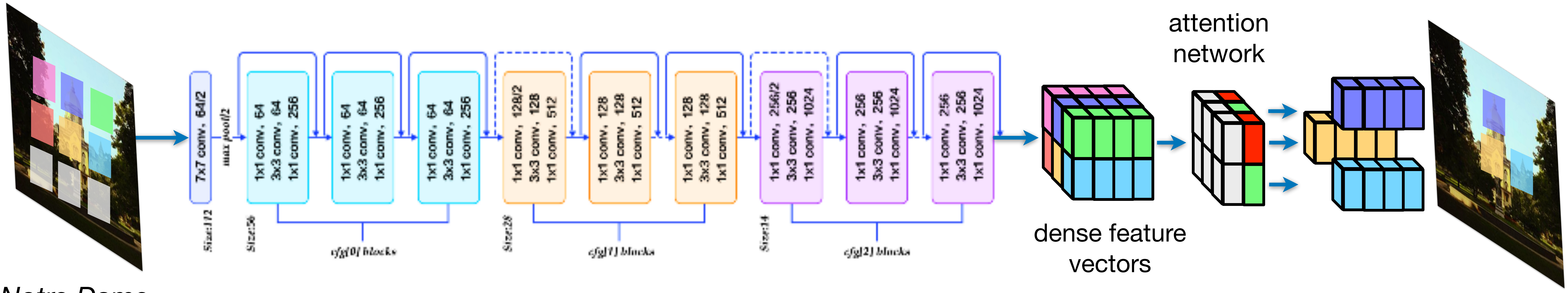
Attention-based Keypoint Selection
How to convert the dense description to keypoint selection?



Attention Network: Weights the dense feature vectors by either suppressing or inciting them.
Rationale: only the locations helpful for classification are kept.

Deep Local Features (DELF)

Attention-based Keypoint Selection
How to convert the dense description to keypoint selection?



Attention Network: Weights the dense feature vectors by either suppressing or inciting them.
Rationale: only the locations helpful for classification are kept.

LOYOLA UNIVERSITY CHICAGO

Deep Local Features (DELF)

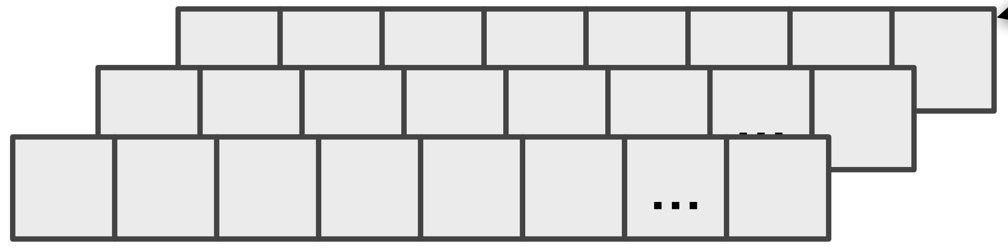
Stages

Dense Feature Extraction

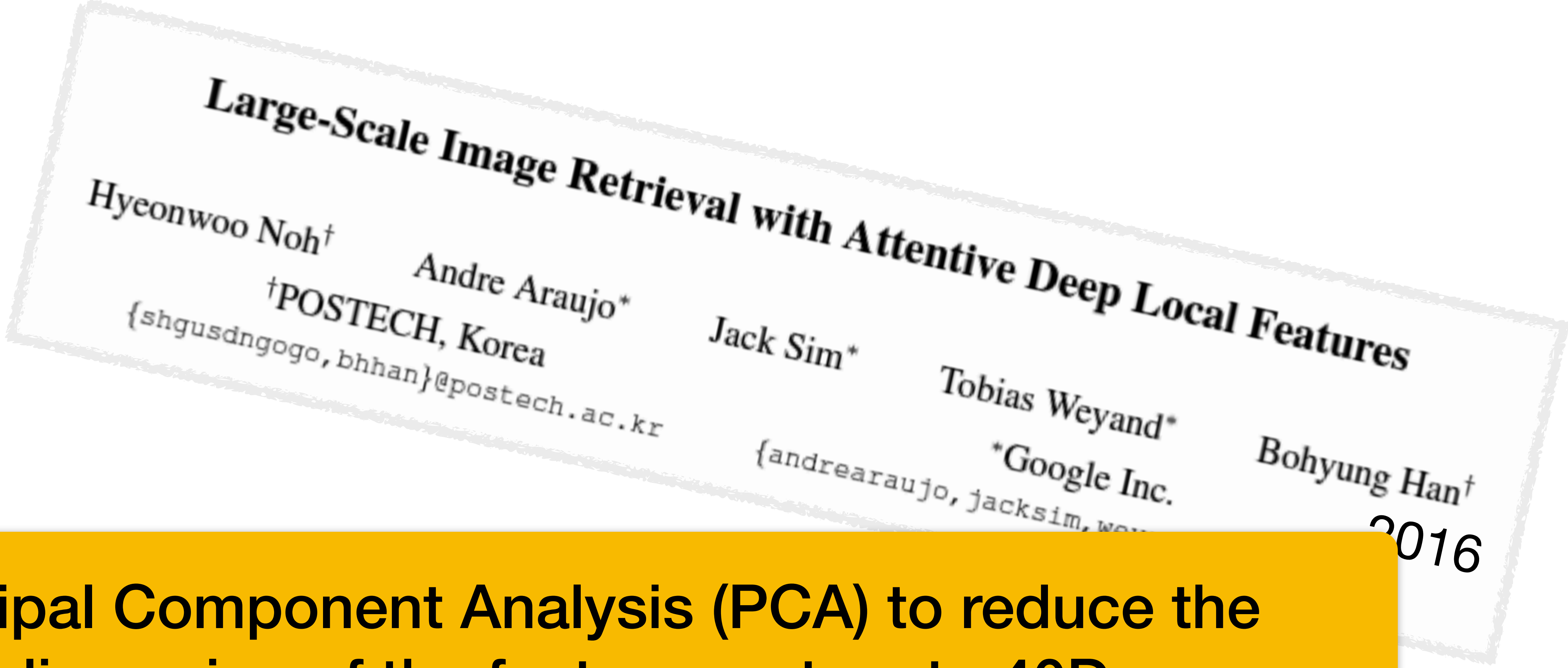
Keypoint Selection

Dimensionality Reduction

Principal Component Analysis (PCA) to reduce the dimension of the feature vectors to 40D



40D feature vectors



Large-Scale Image Retrieval with Attentive Deep Local Features

Hyeonwoo Noh† Andre Araujo* Jack Sim* Tobias Weyand* Bohyung Han†

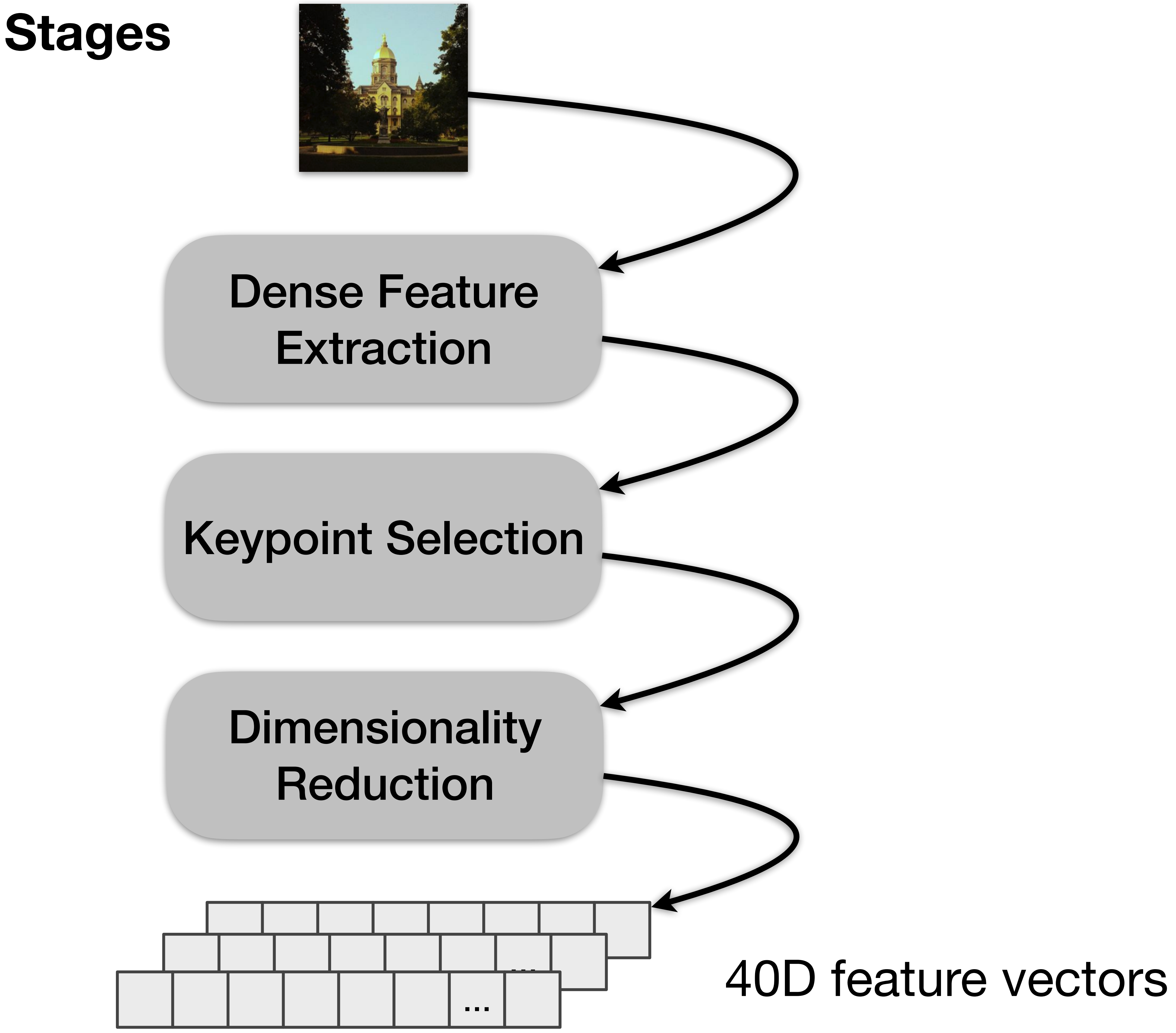
†POSTECH, Korea *Google Inc.

{shgusdngogo,bhhan}@postech.ac.kr {andrearaujo,jacksim,...

2016

LOYOLA UNIVERSITY CHICAGO

Deep Local Features (DELF)



Pros
Learning-based method
(not handcrafted).

Cons
DELF keypoints have *(x, y)* position and response (attention network weight).
They have neither scale nor orientation.

What do you think?
Are they scale and rotation invariant?

LOYOLA UNIVERSITY CHICAGO

What's Up Next?

Usage of Feature Vectors
Nick and Jesus will lead the discussion of Image Retrieval.

Deadline of Assignment #2
Tomorrow is the deadline for submitting assignment #2. Let's share our thoughts with Nick and Jesus.

Release of Assignment #3
Topic: Image classification.