Boston University
CAS CS 640: AI

Lecture on
Introduction to Computer Vision

by Margrit Betke
October 10, 2024

Learning Objectives for this Lecture

Computer Science

- Understand formats of images used as inputs to AI models: greyscale, color, medical scans
- Understand differences and similarities between pre-2012 “traditional computer vision” and post-2012 neural-network-based computer vision & see examples
- Understand why convolution is powerful
- Understand the connection between convolution and correlation
- Understand template matching with image pyramids
- Understand CNNs as a learning hierarchy of features

What is a digital image?

- Digital images are fields of colored dots
- Each dot is called a

  pixel = picture cell
- In the medical domain, a pixel can be a

  voxel = volumetric cell

Image Processing

Computer Science

- The engineering field of image processing proceeds computer vision by decades.
- Goal: Restore noisy images or send encoded image through a channel

Image Processing

Computer Science

- The engineering field of image processing proceeds computer vision by decades.
- Goal: Restore noisy images or send encoded image through a channel
- Standard test image

Lena Soderberg ‘72

Slide credit: Diane Theriault

Image Processing

Computer Science

- The engineering field of image processing proceeds computer vision by decades.
- Goal: Restore noisy images or send encoded image through a channel

- Standard test image with detail,
  shading, texture,
  sharp & blurry regions:
  Lena Soderberg '72
  controversy!

Slide credit: Diane Theriault

Color Models

Computer Science

- Images can be gray scale, color, or

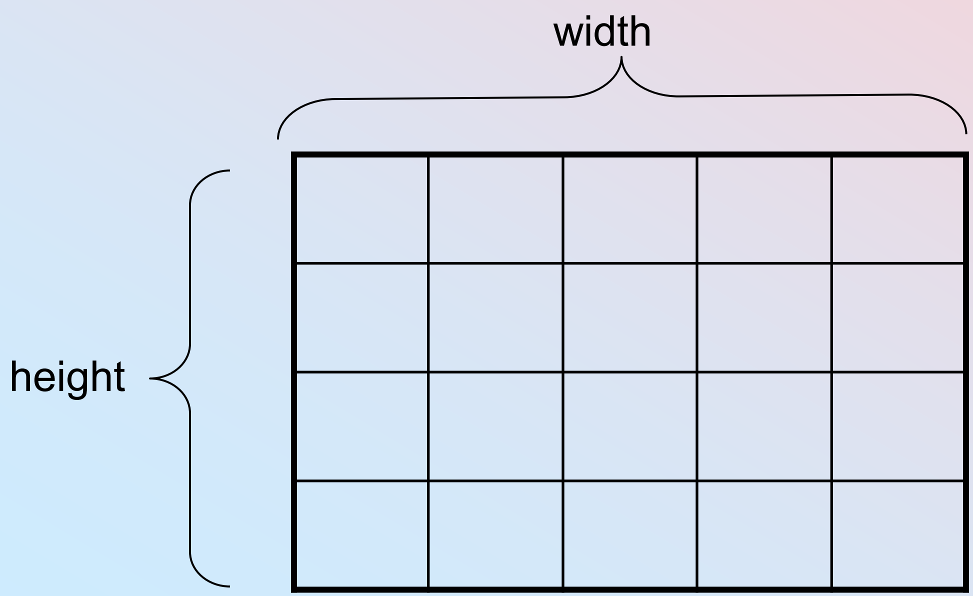
  color with an alpha (transparency) channel (used in graphics for overlaying images)
- Most common color representation is RGB (Red, Green, Blue)
- Other models include CMYK (used for print) and YUV (often used for input from cameras, compression, and transmission)

Slide credit: Diane Theriault

What is an image?

Computer Science

- Images are 2 dimensional arrays of data, with an associated width, height, and color depth.
- Images typically use one byte per color channel per pixel.
- Gray images have 1 channel. RGB images have 3 color channels. RGBA images have 4 color channels.



Slide credit: Diane Theriault

Digital Image File Formats

Image:

Header
Table = Map = Raster of Pixel Values

Size of table, color, compression scheme

Gray-scale images: generally 1 byte per pixel

Color images: 3 numbers (each 1 byte) per pixel

Medical images, e.g., CT, MRI: typically 2 bytes per voxel

Example: PGM Image

Image file

Image ??

```
P2
3   3   255
-----------
0   255   0
220   0  20
0   130   0
```

Example: PGM Image

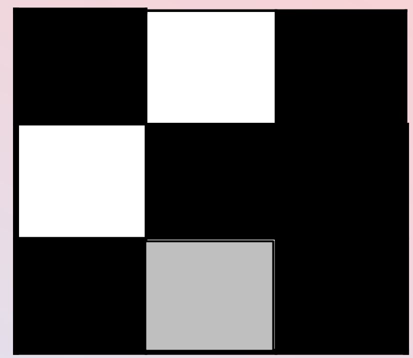
Computer Science

Image file

```
P2
3   3   255

0   255   0
220   0   20
0   130   0
```

Image

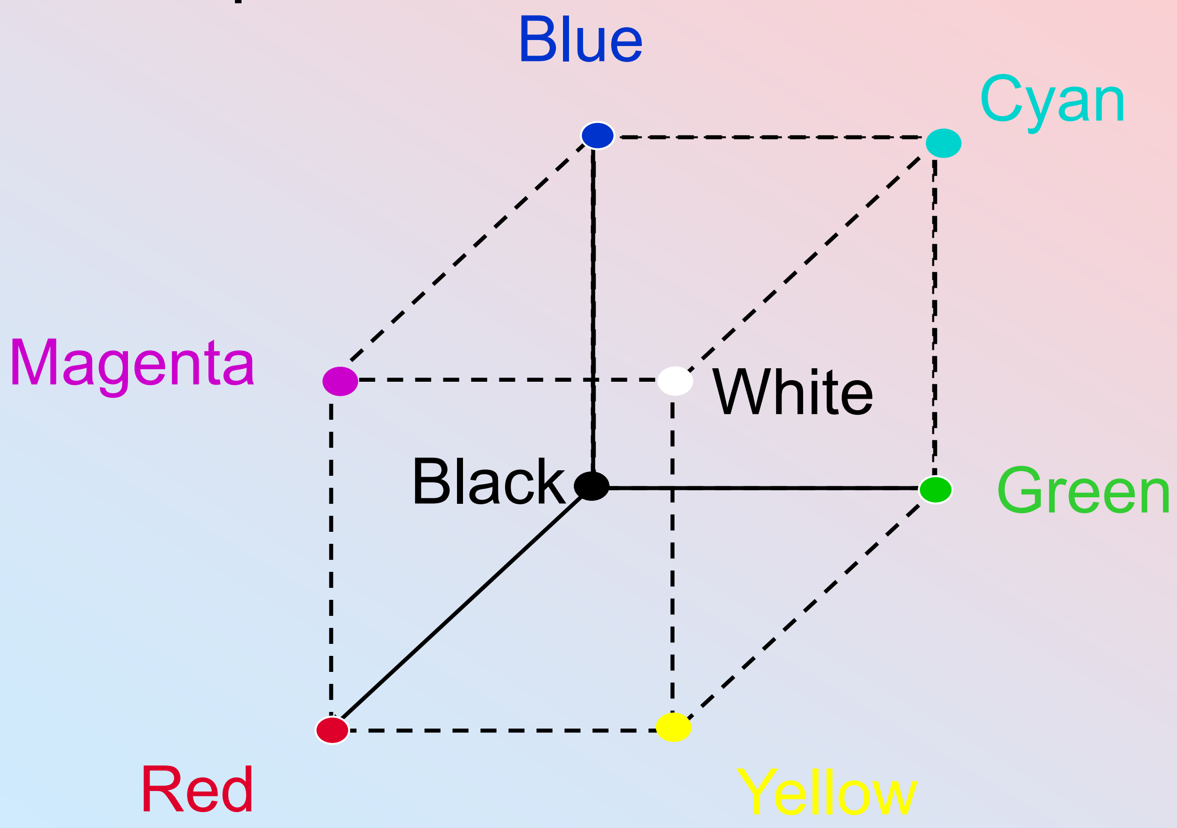


Light: Electromagnetic Waves

Wavelength λ

RGB Color Space

Additive Space



Example: PPM Image

Boston University
Computer Science

Image file

```
P3
3   3   255
```

0	0	0	255	0	0	0	0	0
0	255	0	0	0	0	255	255	0
0	0	0	0	0	255	0	0	0

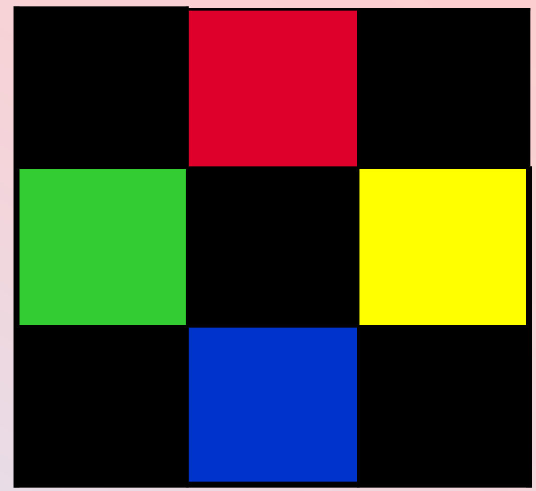
Image ??

Example: PPM Image

Image file

```
P3
3   3   255
-----------------------------------------
0   0    0   255  0    0    0    0    0
0  255   0    0   0    0   255  255   0
0   0    0    0   0   255   0    0    0
```

Image



How do I get at the data?

Boston University
Computer Science

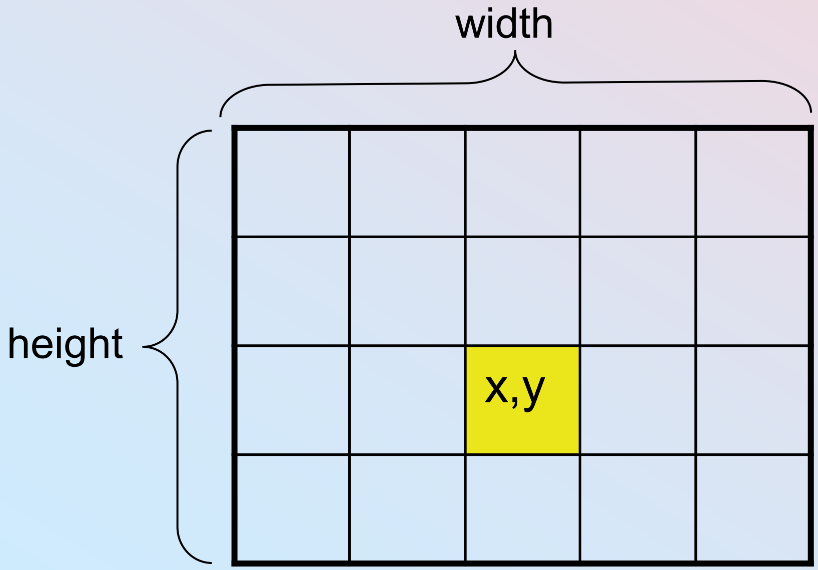
- Some image-handling APIs have nice interfaces, but speed can be a problem.
- You will probably have to handle the bytes of data directly at some point

Slide credit: Diane Theriault

How do I get at the data?

Computer Science

- X = desired row
- Y = desired column
- C = color channel (red, green, blue, …).
- Bpp = Bytes per pixel (color channels)
- Image data is normally stored in row major order
- Note that there may be multiple values associated with each x,y pixel
- Data(x,y,c) = y*(width*Bpp) + x*Bpp + c



Slide credit: Diane Theriault

Example of a "Traditional" Computer Vision Algorithm: Color to Gray Scale Conversion

- Pre-neural-network-revolution Computer Vision: Algorithms

- Example of such a pre-2012 algorithm:

  Converting from color to gray scale,
  a very common operation

Color-to-Grayscale Conversion

Computer Science

Four Conversion Algorithms:

1. “Quick and dirty” conversion: Grab the Green Channel: G
2. Average R, G, B: (R+G+B)/3
3. Max(R, G, B)
4. Weigh them: 0.3*R + 0.6*G + 0.1*B

Slide credit: Diane Theriault

Image File Formats

Computer Science

- PPM / PGM is the simplest file format ever, but not supported by Photoshop or MS Image Viewer. Uncompressed.
- BMP: Microsoft's uncompressed image format
- GIF: Images are compressed using run-length encoding, and reducing the number of colors used. Patent expired 2003.
- JPEG: Images are compressed by removing high frequency information
- DICOM: Digital Imaging and Communications in Medicine

Slide credit: Diane Theriault

Tools of the Trade

Computer Science

- OpenCV is a widely used, open-source computer vision library maintained by Intel
- Provides libraries for image I/O, movie I/O and camera capture
- Industrial strength computer vision and image processing implementations
- "Quick and dirty" GUI toolkit

Slide credit: Diane Theriault

Today's Computer Vision: Mostly (but not all) Neural Networks

Computer Science

- Deep convolutional neural networks
- Transformers
- Diffusion models

\+ traditional computer vision algorithms, representations, geometry, and tricks

Deep learning does not work well for:

Multi-view geometry, i.e., 3D object pose and 3D scene representation, and object tracking

Traditional Computer Vision Algorithm: Template Matching

Computer Science

- *Template matching = search algorithm*
- *Goal: Find object shown in template image in a scene image.*
- *More specific goal: Find the location of the scene subimage that best matches the template image*
- *Mathematical definition of match:*
  - *Scene subimage s, template image m*
  - *Same size images = n pixels*
  - s_i *= ith pixel grey value in subimage of scene*
  - m_i *= ith pixel grey value in template image m*
  - *Sum-squared match equation SSD = Σ_(i=1 to n)* $(s_i - m_i)^2$
  - *Alternative match equation: Normalized Correlation Coefficient*
  - *Both match equations involve multiplying pixel values* $s_i\, m_i$

1D Discrete Convolution

Computer Science

1D Convolution:
Time signal *f* and shifted time signal *g* are multiplied and added:

$$(f * g)[n] \stackrel{\text{def}}{=} \sum_{m=-\infty}^{\infty} f[m]\, g[n-m]$$
$$= \sum_{m=-\infty}^{\infty} f[n-m]\, g[m].$$

2D generalization:
f = input image, g = template image
(or CNN function)

2D Convolution Example

Computer Science

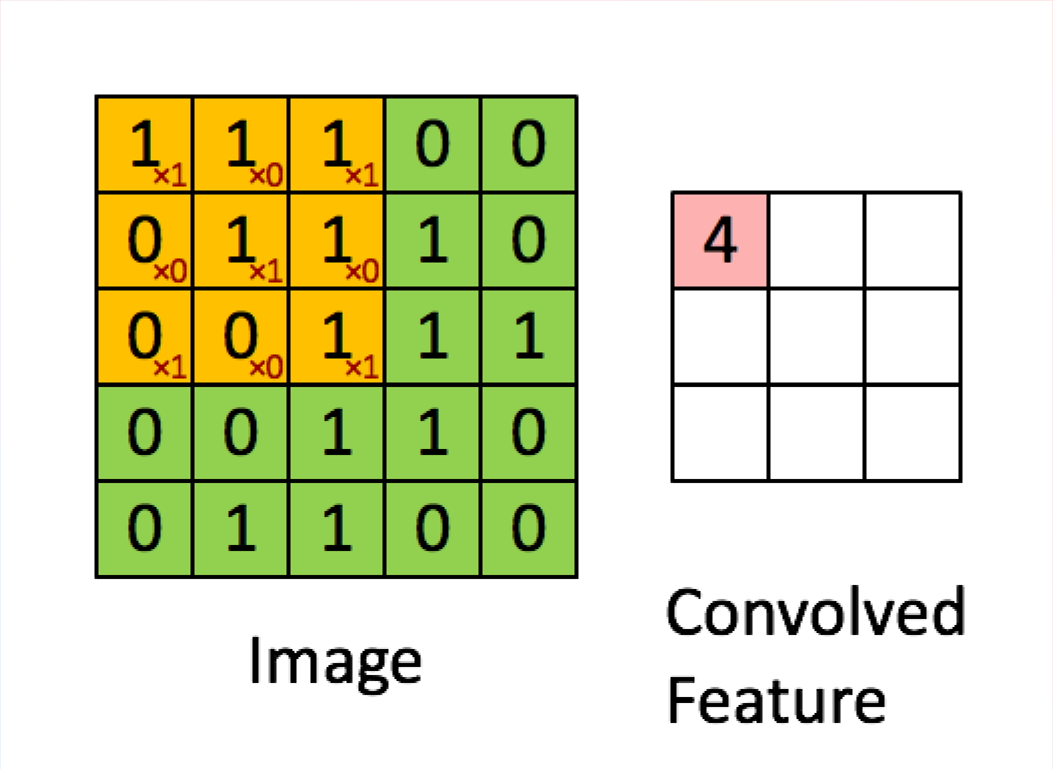
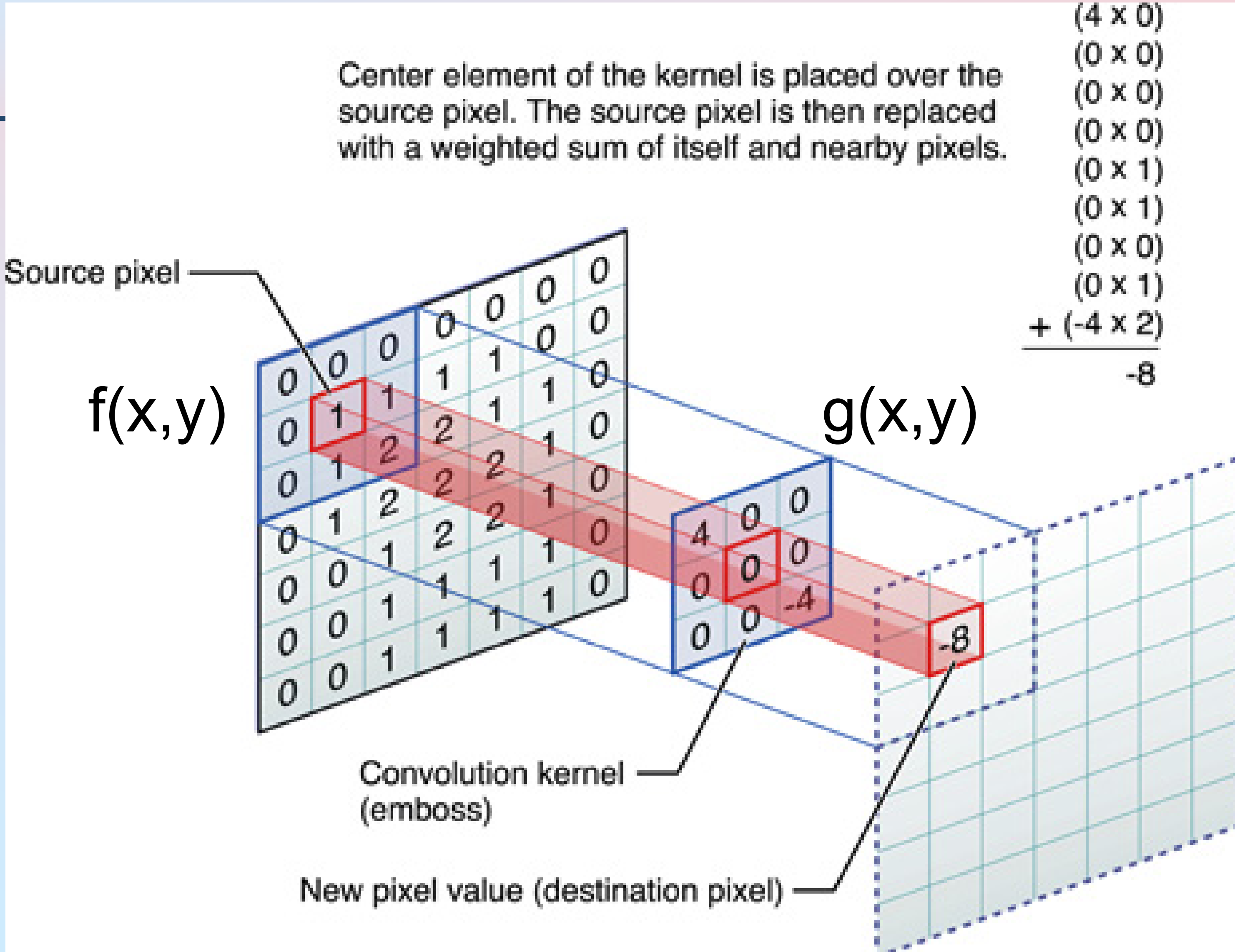


Image Credit: Nvidia



f*g

Image Credit: Madhushree Basavarajaiah

Why is Convolution Powerful?

Signal Processing:

Computer Science

Convolution is used to define a "matched filter" for locating "targets" in time signals

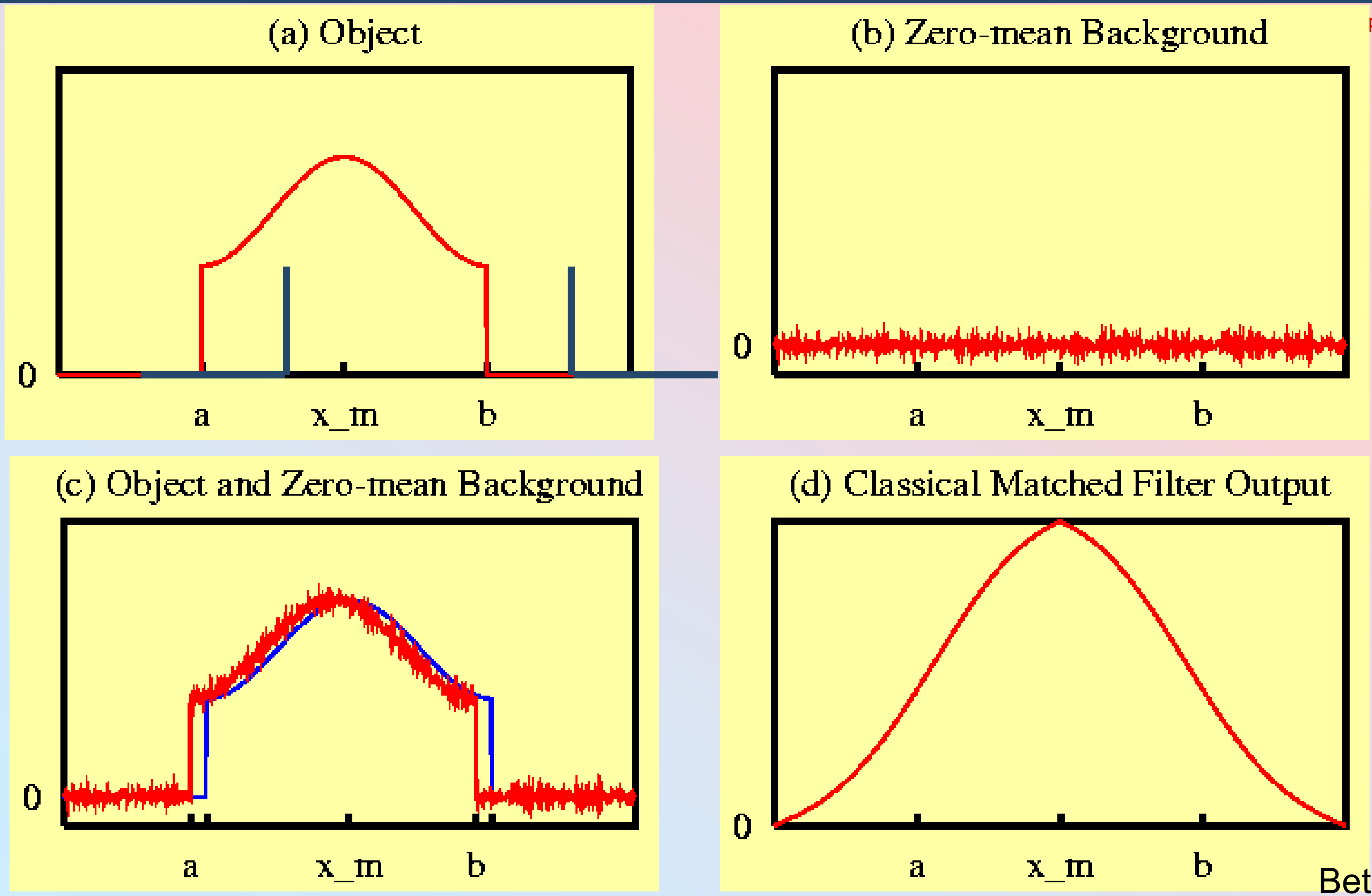
Template matching is optimal algorithm if noise is Gaussian.

Optimality of Template Matching

Computer Science

Betke, Makris, IJCV 2001

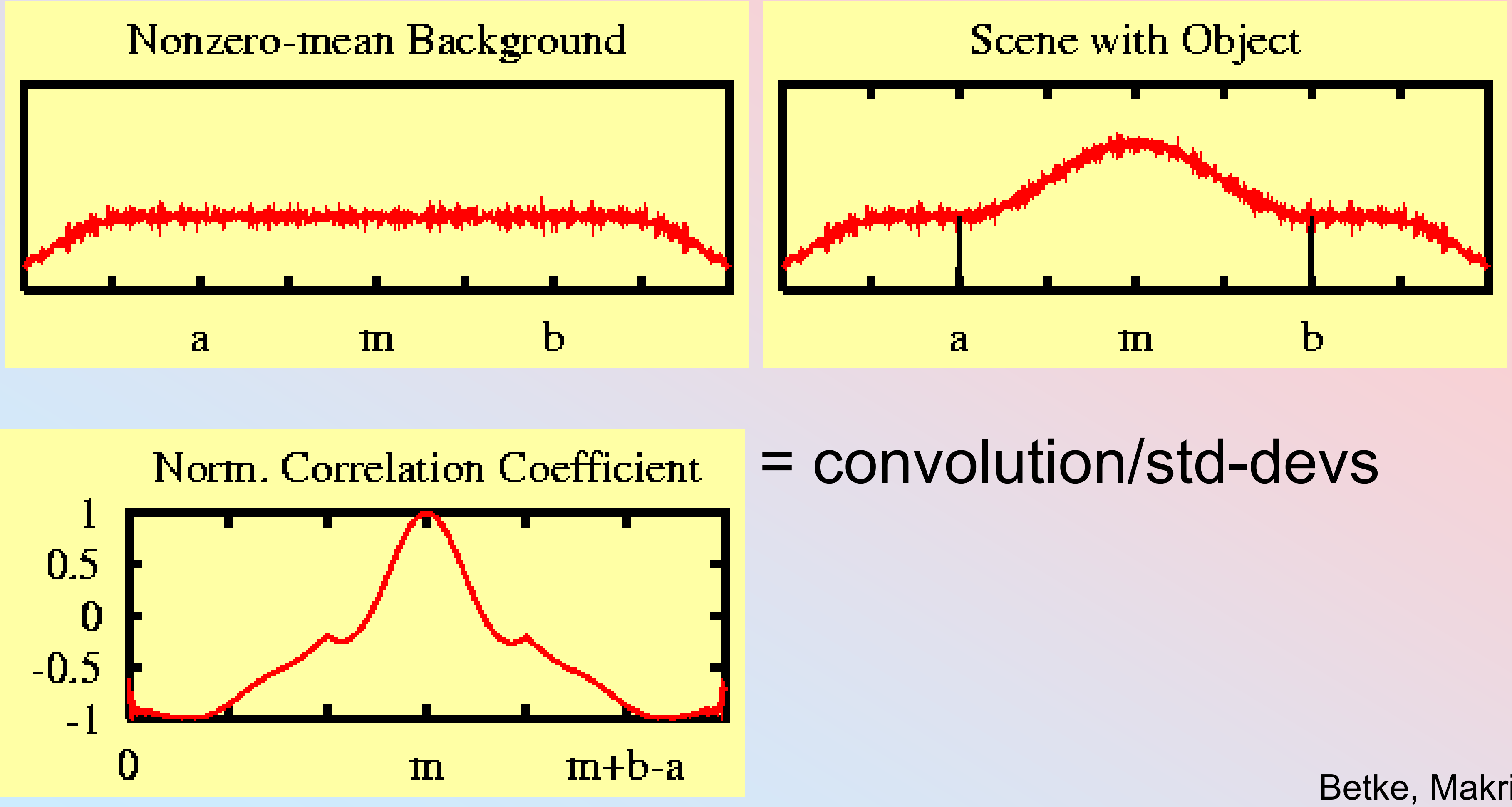
1D Position Estimation: Σ object*background



Betke, Makris, IJCV 2001

Another 1D convolution example:

Computer Science



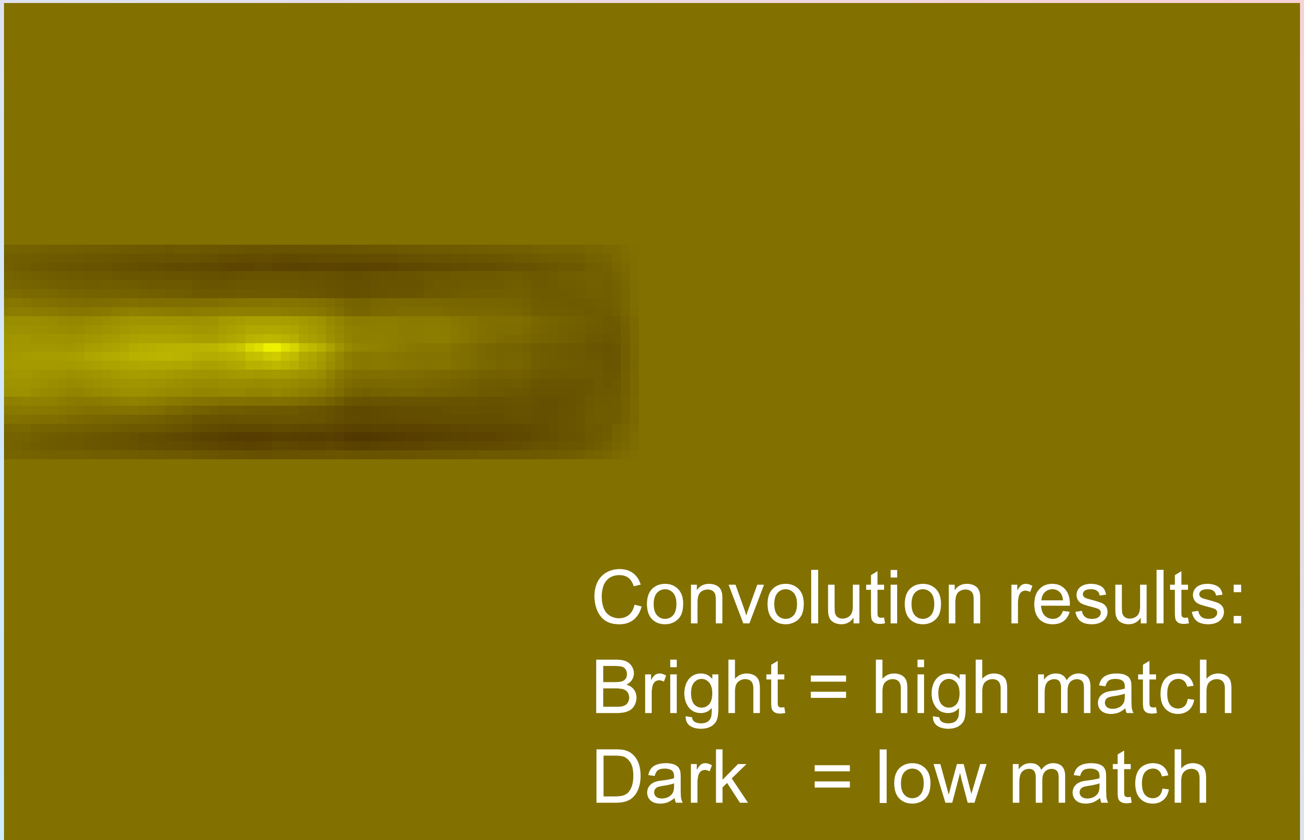
= convolution/std-devs

Betke, Makris, IJCV 2001

2D Position Estimation

Computer Science

Convolution of one-way sign with itself:
Both scene and template image are



Betke, Makris, IJCV 2001

2D Position Estimation

Computer Science

Convolution of one-way sign with itself:
Both scene and template image are

Betke, Makris,
IJCV 2001

2D Position Estimation

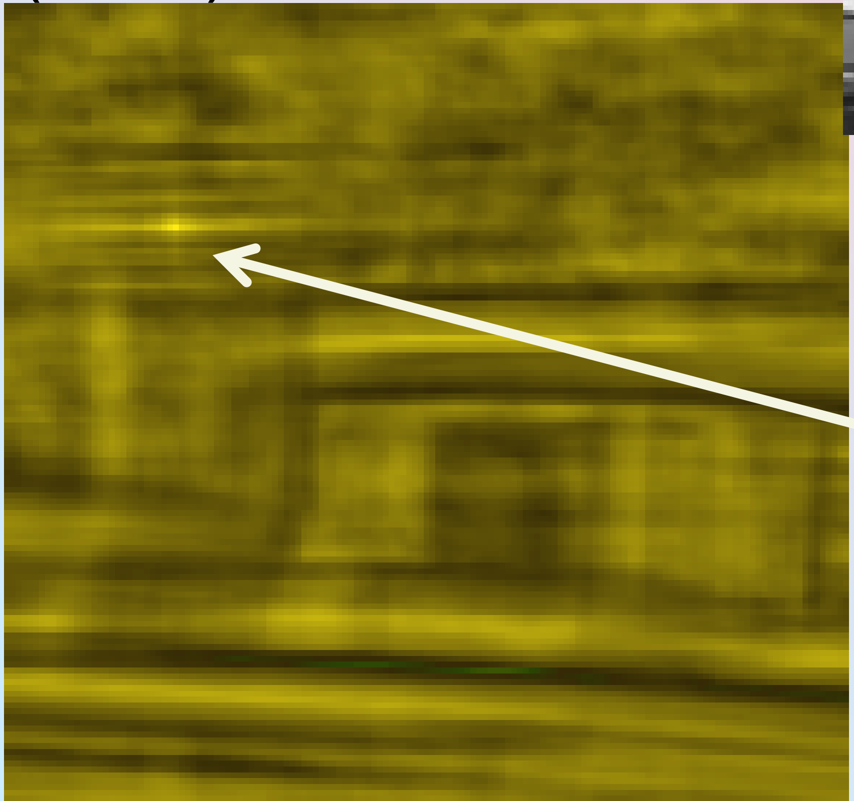
Computer Science

Convolution of one-way sign with itself:
Both scene and template image are

Betke, Makris, IJCV 2001

2 D Position Estimation

Convolution of one-way sign with scene (NCC)

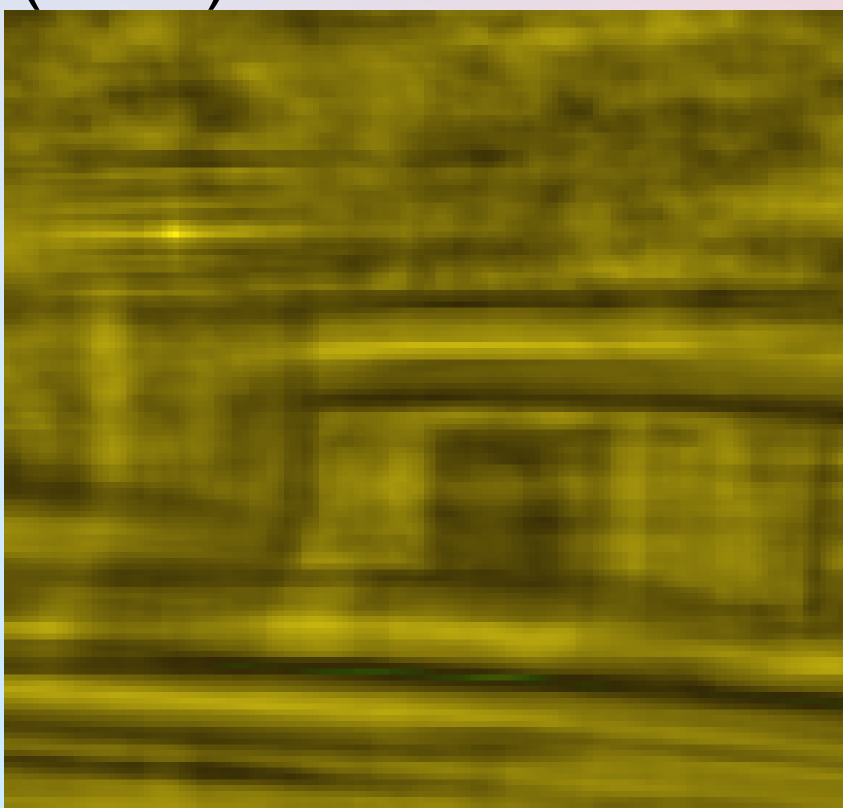


Peak in performance surface (= negative loss fct) at correct location

Betke, Makris, IJCV 2001

2 D Position Estimation

Convolution of one-way sign with scene (NCC)



This performance surface is computed for correct size of one-way sign

Different surfaces for different sizes of object

Sample Performance Surfaces

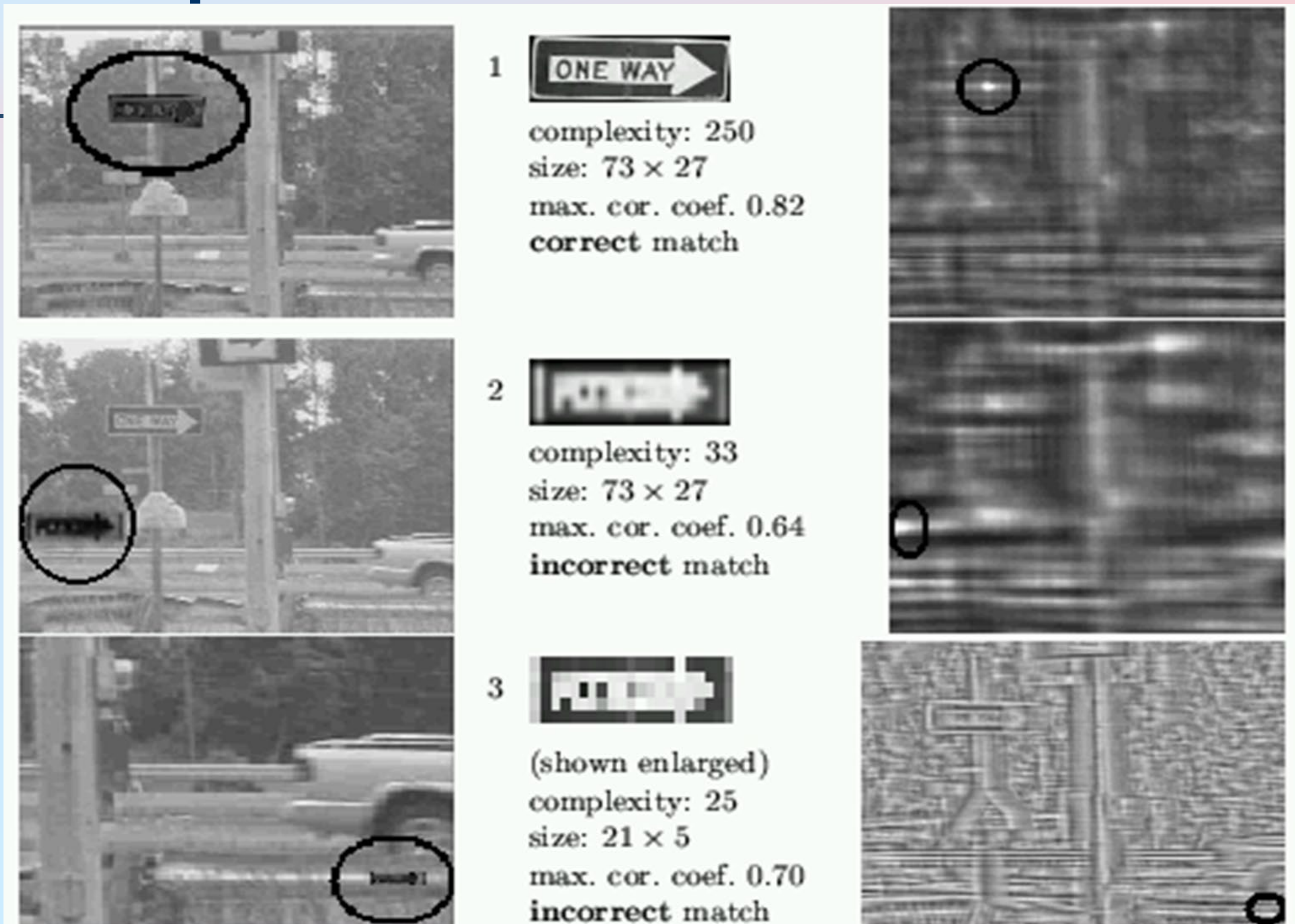
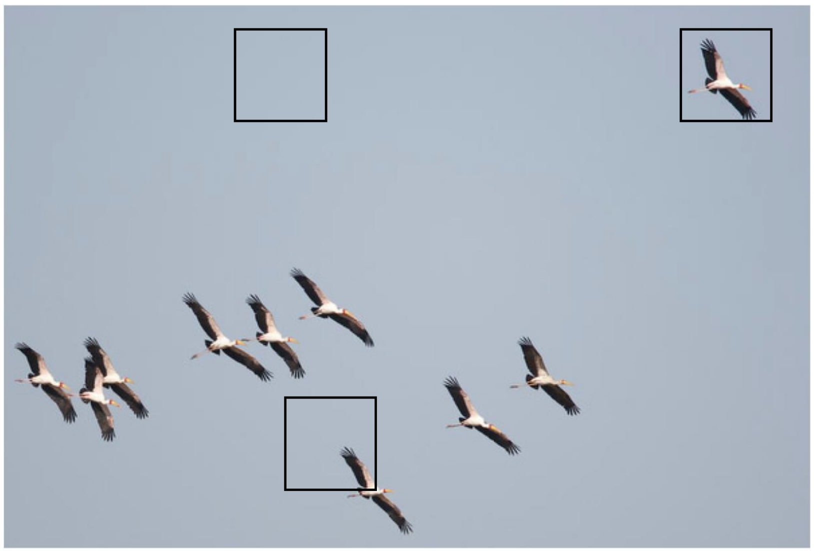


Image Credit: Efros/Freeman

Convolving template with subimage



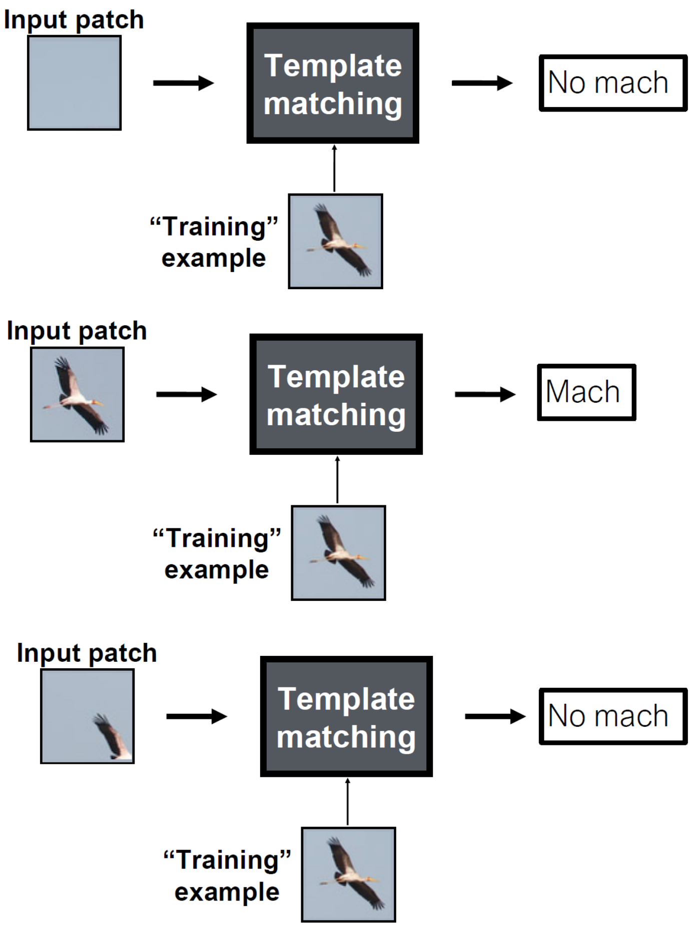
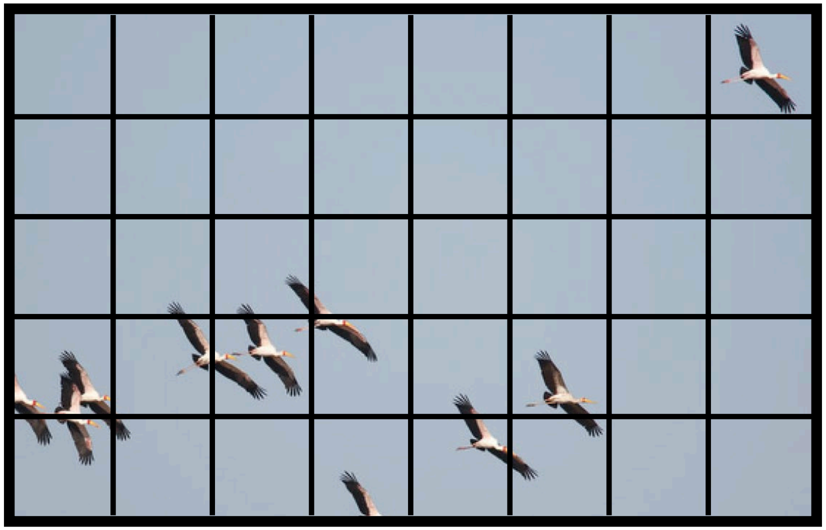


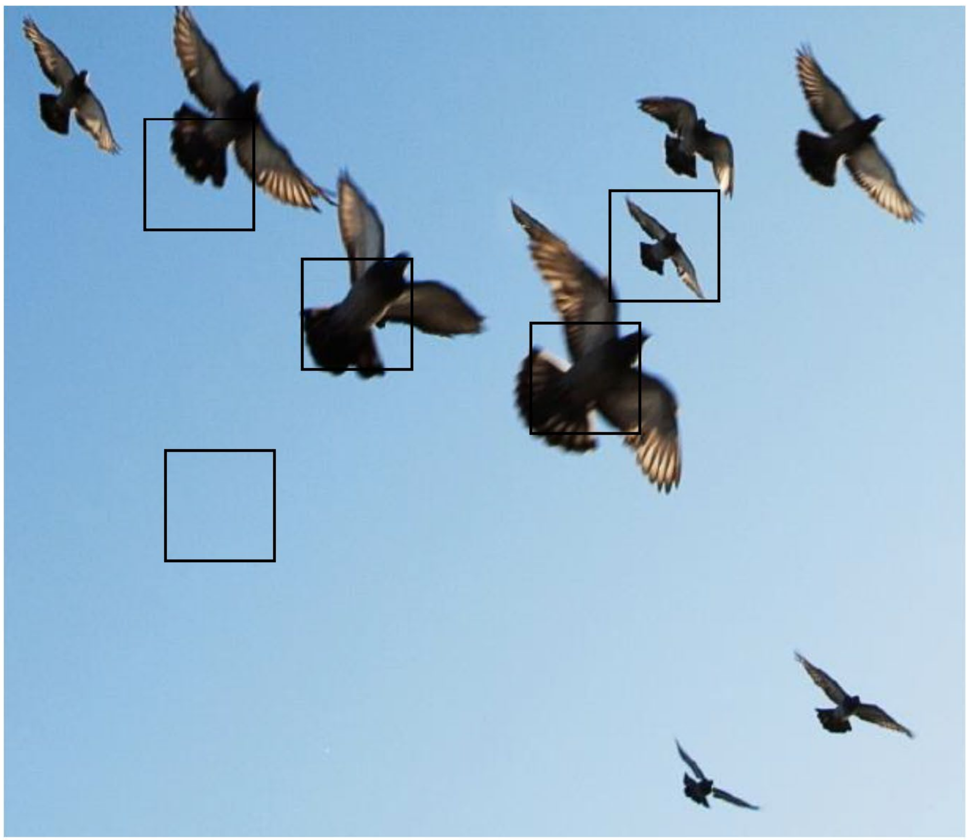
Image Credit: Efros



Sky	Sky	Sky	Sky	Sky	Sky	Sky	Bird
Sky	Sky	Sky	Sky	Sky	Sky	Sky	Sky
Sky	Sky	Sky	Sky	Sky	Sky	Sky	Sky
Bird	Bird	Bird	Sky	Bird	Sky	Sky	Sky
Sky	Sky	Sky	Bird	Sky	Sky	Sky	Sky

Image Credit: Freeman

What if object in image appears in a range of sizes?



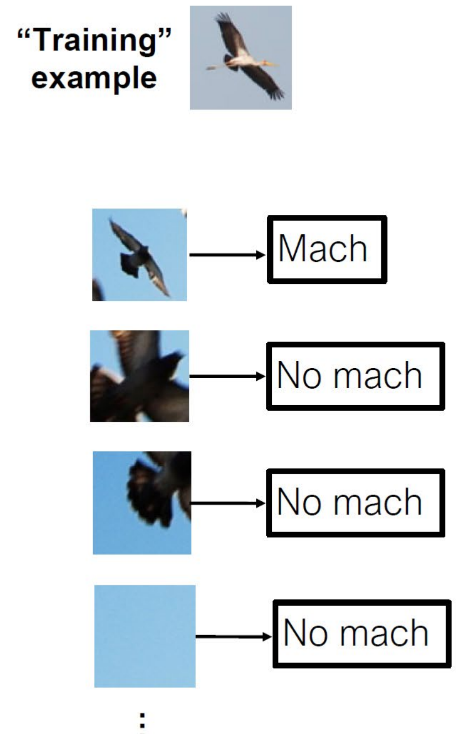
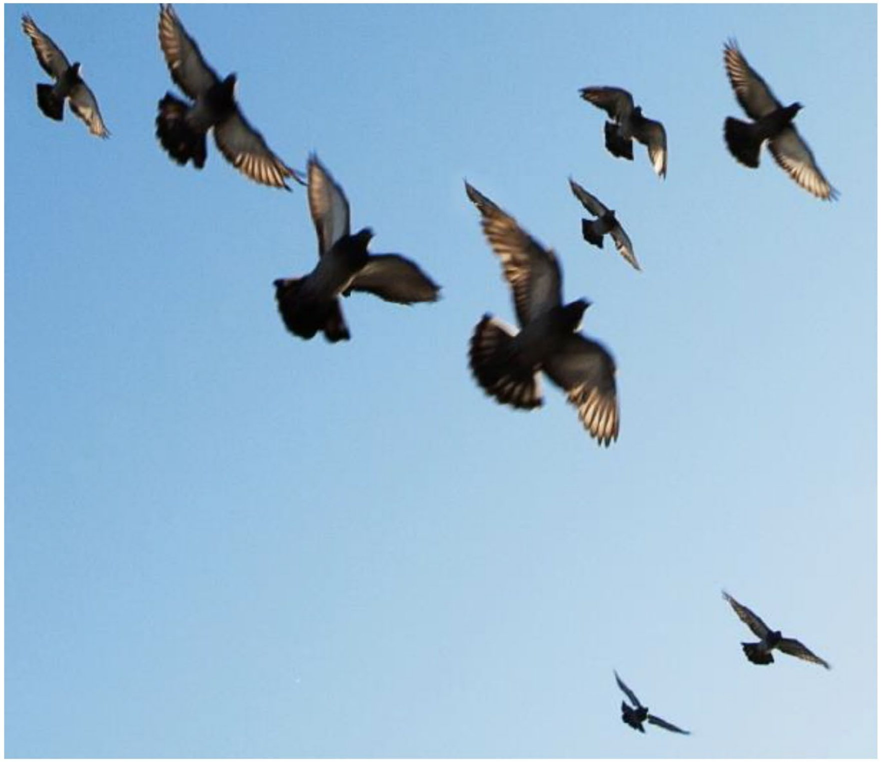


Image Credit: Efros

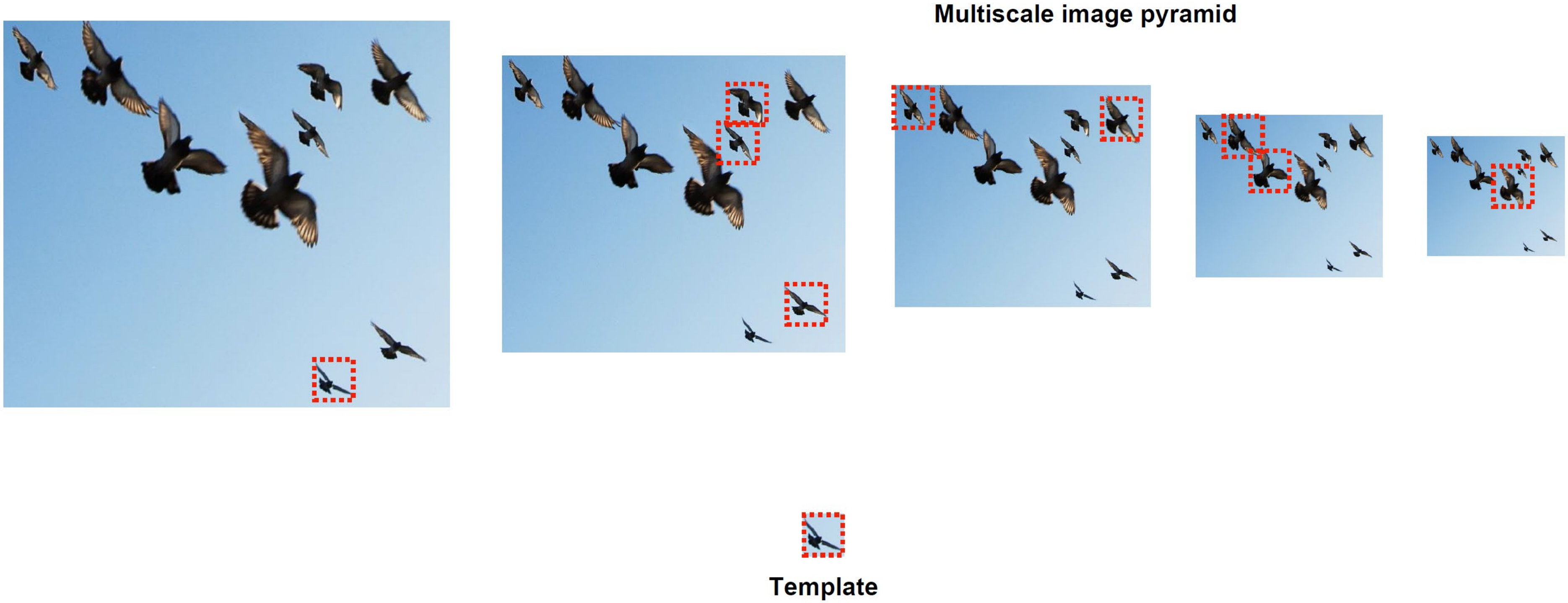
Multi-Scale Pyramids



Template

Image Credit: Efros

Multi-Scale Pyramids



A multiscale image pyramid provides an alternative image representation to achieve translation and scale invariance

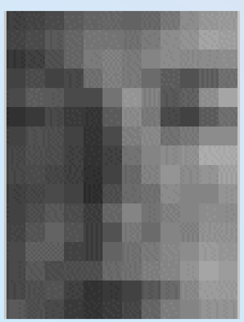
Image Credit: Efros

Multi-Resolution Matching

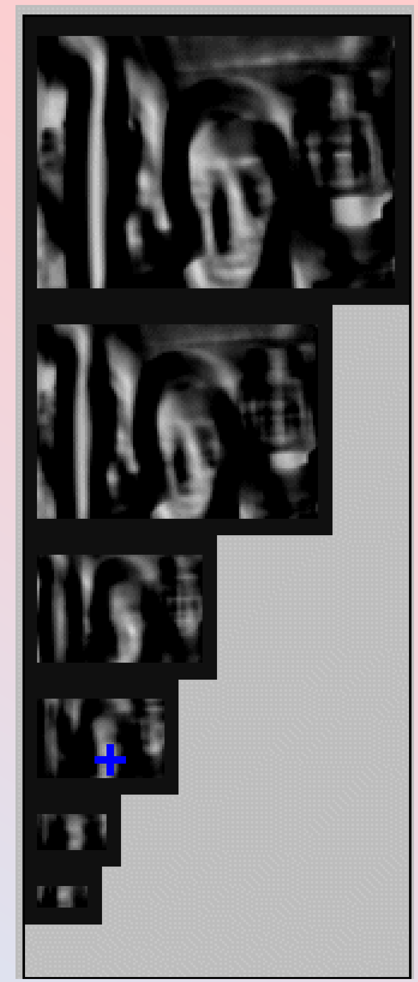
Computer Science

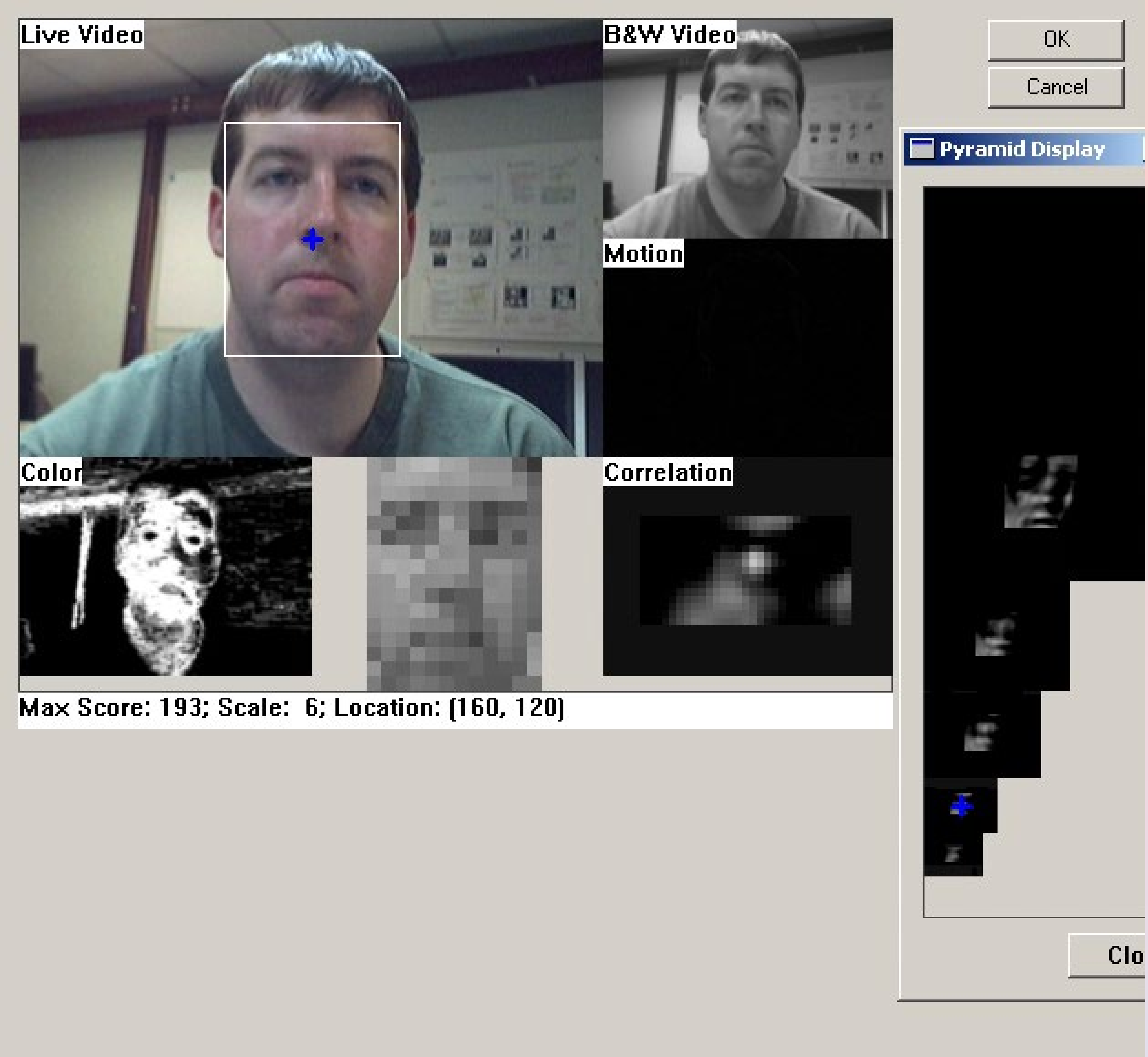
Normalized correlation coefficient over multi-resolution search space:

$$r = \frac{1/n \; \Sigma_i \, (s_i - mean(s)) \, (m_i - mean(m))}{(\sigma_s \, \sigma_m)}$$



← Template matched over all resolutions →




Live Video
B&W Video
OK
Cancel
Pyramid Display
Motion
Color
Correlation
Max Score: 193; Scale: 6; Location: (160, 120)
Clo

Learning Objectives for this Lecture

Computer Science

- Understand formats of images used as inputs to AI models: greyscale, color, medical scans
- Understand differences and similarities between pre-2012 “traditional computer vision” and post-2012 neural-network-based computer vision & see examples
- Understand why convolution is powerful
- Understand the connection between convolution and correlation
- Understand template matching with image pyramids
- Understand CNNs as a learning hierarchy of features