SAIR

Spatial AI & Robotics Lab

CSE 473/573-A

W7: TEXTURE & OBJECT

Chen Wang

Spatial AI & Robotics Lab

Department of Computer Science and Engineering

University at Buffalo The State University of New York

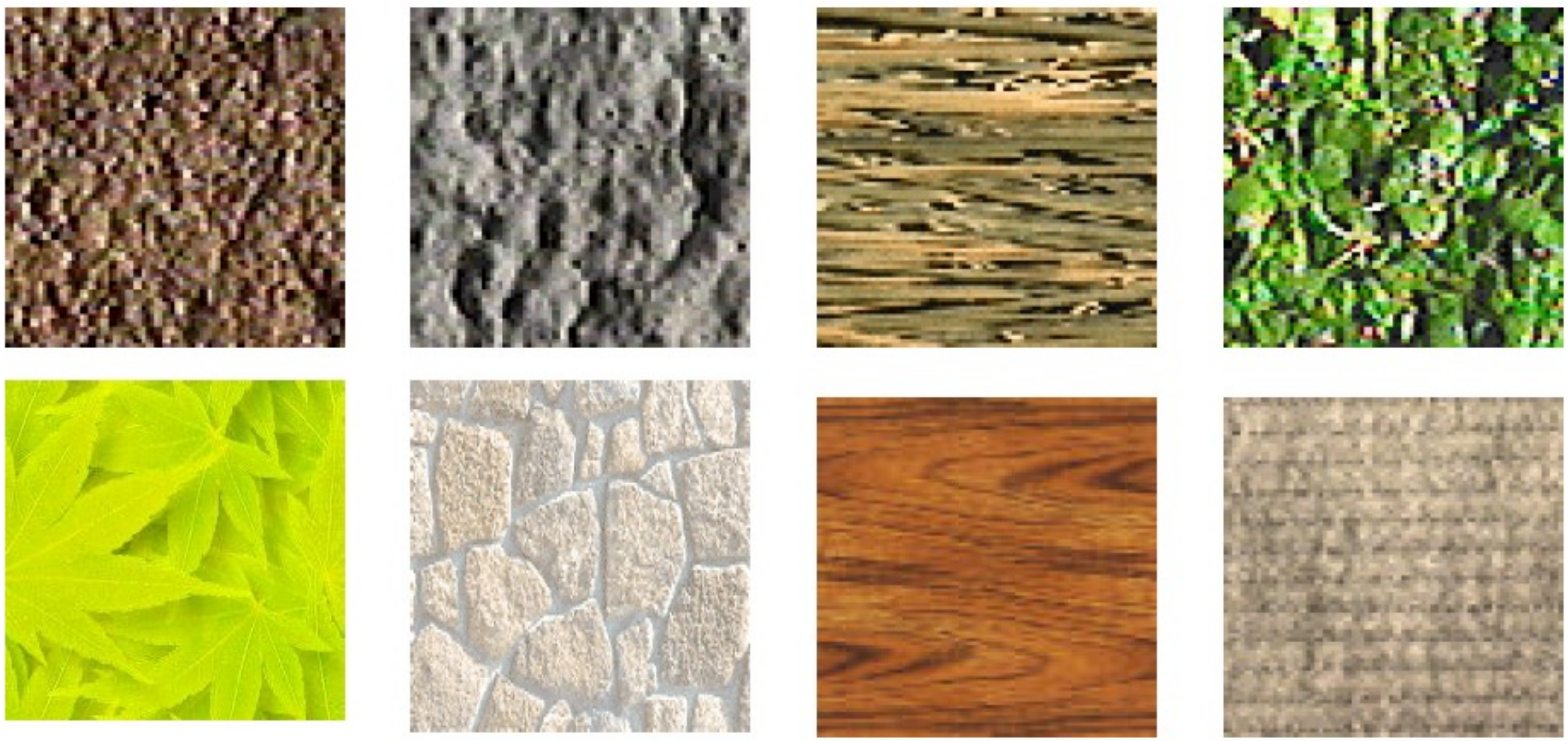
SAIR
Spatial AI & Robotics Lab

COMPUTER VISION

Texture

Texture

- What defines a texture?



What is Texture?

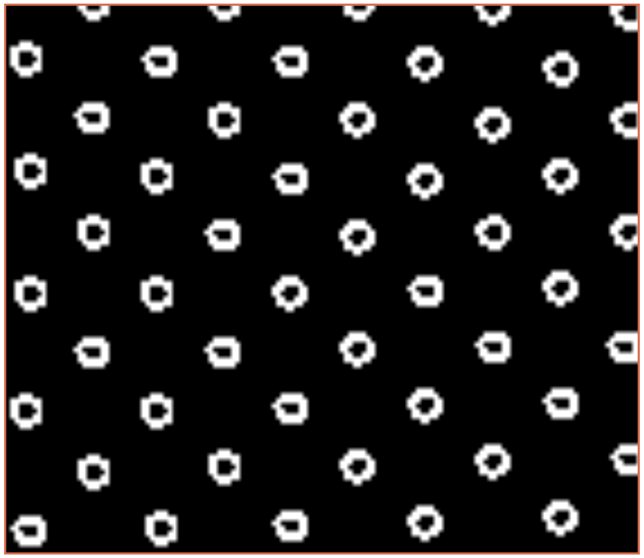
- No Formal Definition
  - There will be **signficant variation** in intensity levels between pixels
  - these variations perform **repetitive** patterns- homogeneous at some scale
  - **local statistics** are constant and slowly varying

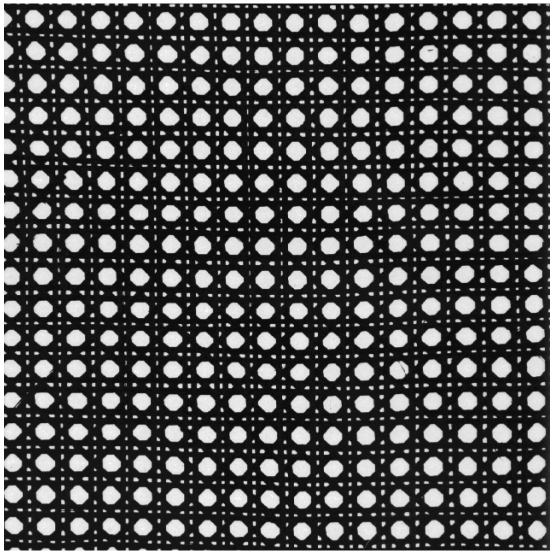
- Human visual systems perceived textures as homogeneous regions even though they don't have the same intensity

SAIR
Spatial AI & Robotics Lab

Includes: regular patterns

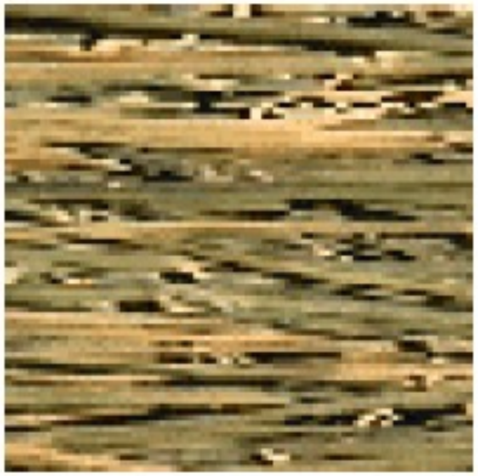
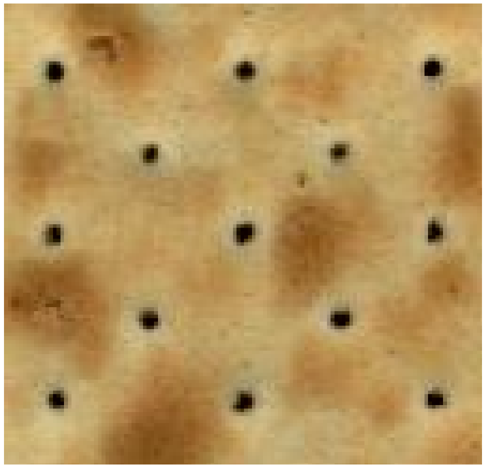
- We interpret this first image as a brick wall





Includes: random patterns



SAIR
Spatial AI & Robotics Lab

Texture

- Is a property of a "group of pixels” or Area
  - a single pixel does not have texture
- It is scale dependent
  - at different scales textures have different properties
- It contains many possibly countless primitive objects
- It involves the spatial distribution of intensities
  - 2D histograms
  - Co-occurrence matrices

Scale

- Scale is important – consider sand
- Close up
  - "small rocks, sharp edges"
  - "rough looking surface"
  - "smoother"

- Far Away
  - "one object
  - ⇒ brown/tan color"

SAIR
Spatial AI & Robotics Lab

How would you describe a Texture?

- Coarseness
- Roughness
- Direction
- Frequency
- Uniformity
- Density

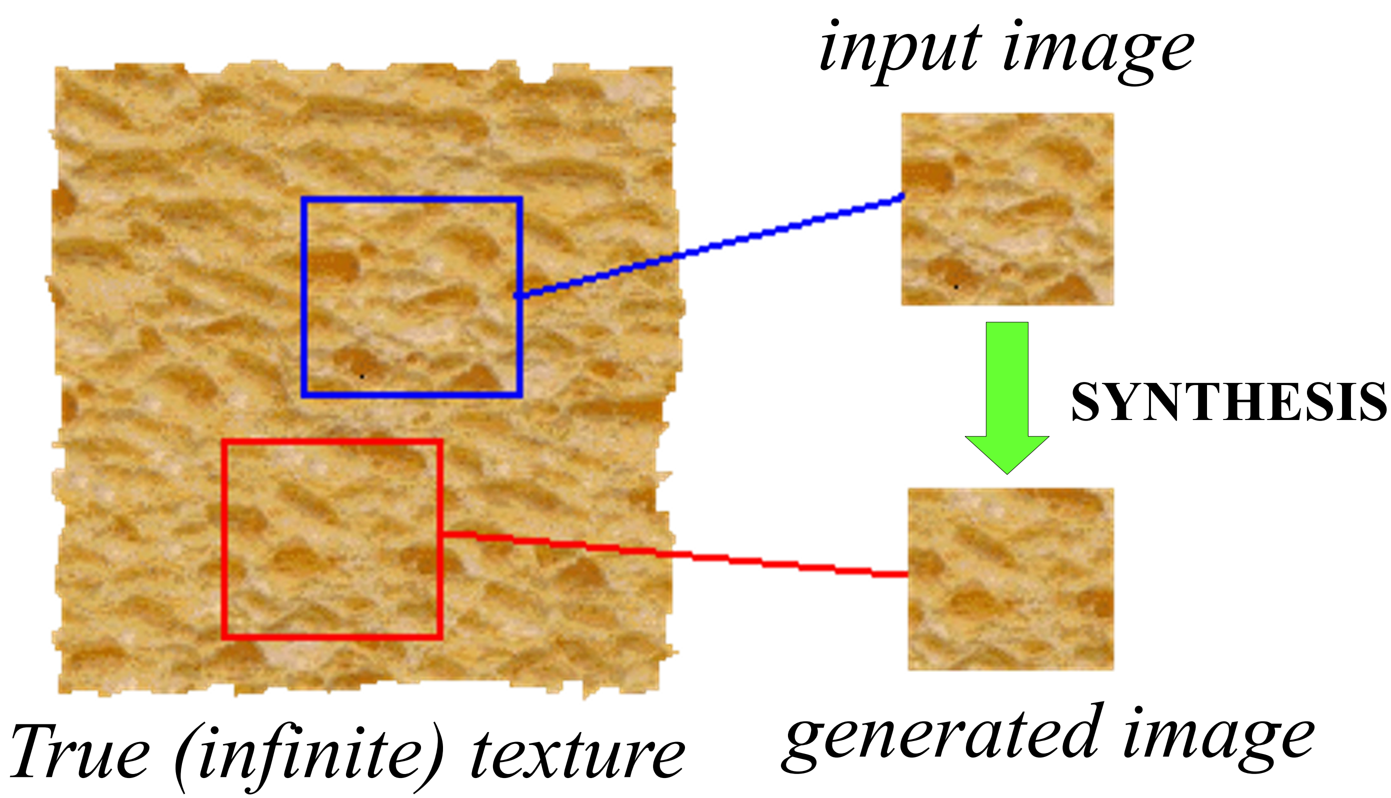
- How do you describe
  - dog fur, cat fur, wood grain, or cloth?

When are two textures similar?

- All images are different instances of the same texture
- We can differentiate between them, but they seem generated by the same process.

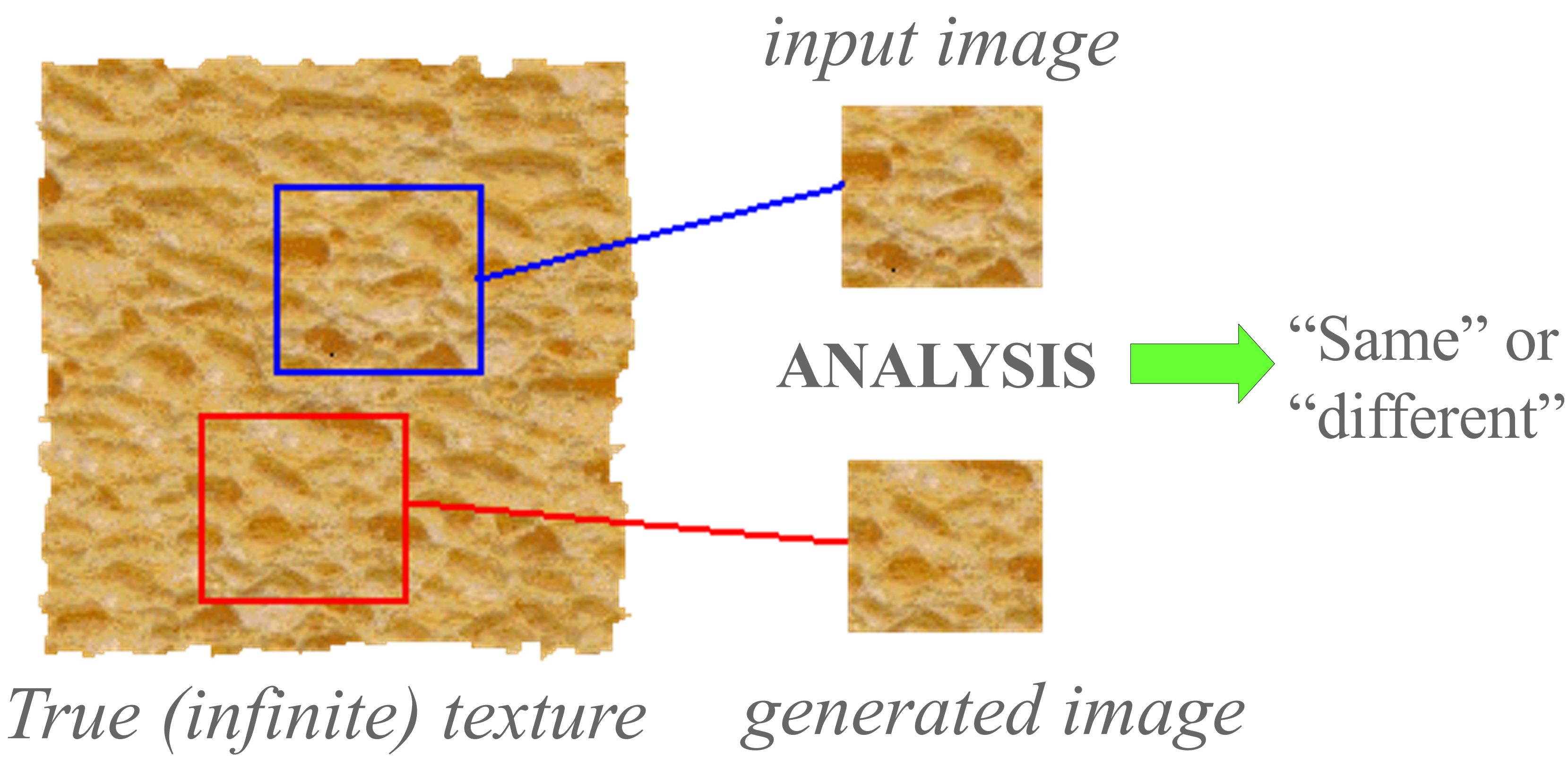
The Goal of Texture Synthesis

- Given a finite sample of some texture, the goal is to synthesize other samples from that same texture
- The sample needs to be "large enough“

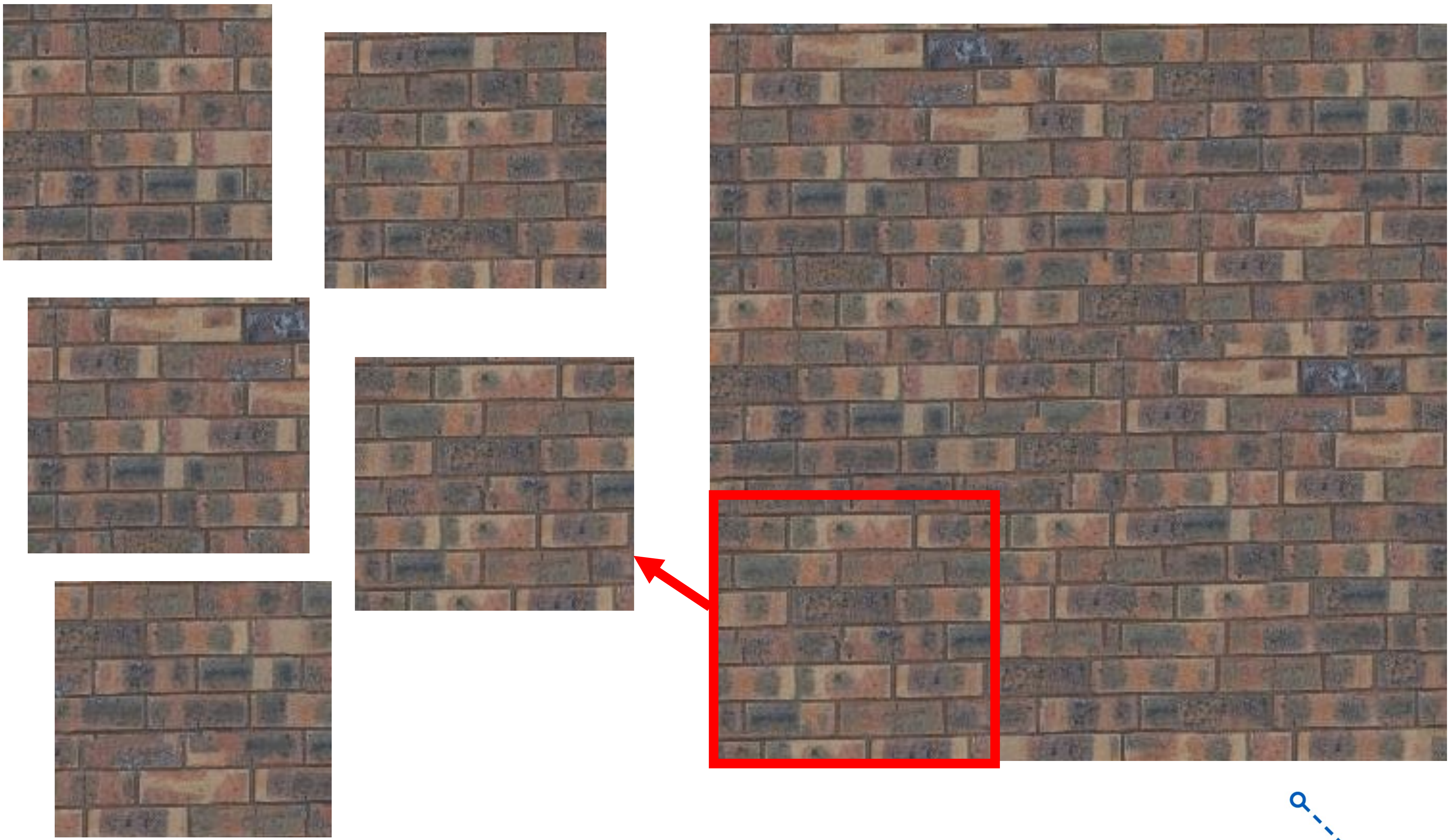


The Goal of Texture Analysis

- Compare textures and decide if they’re made of the same “stuff”.

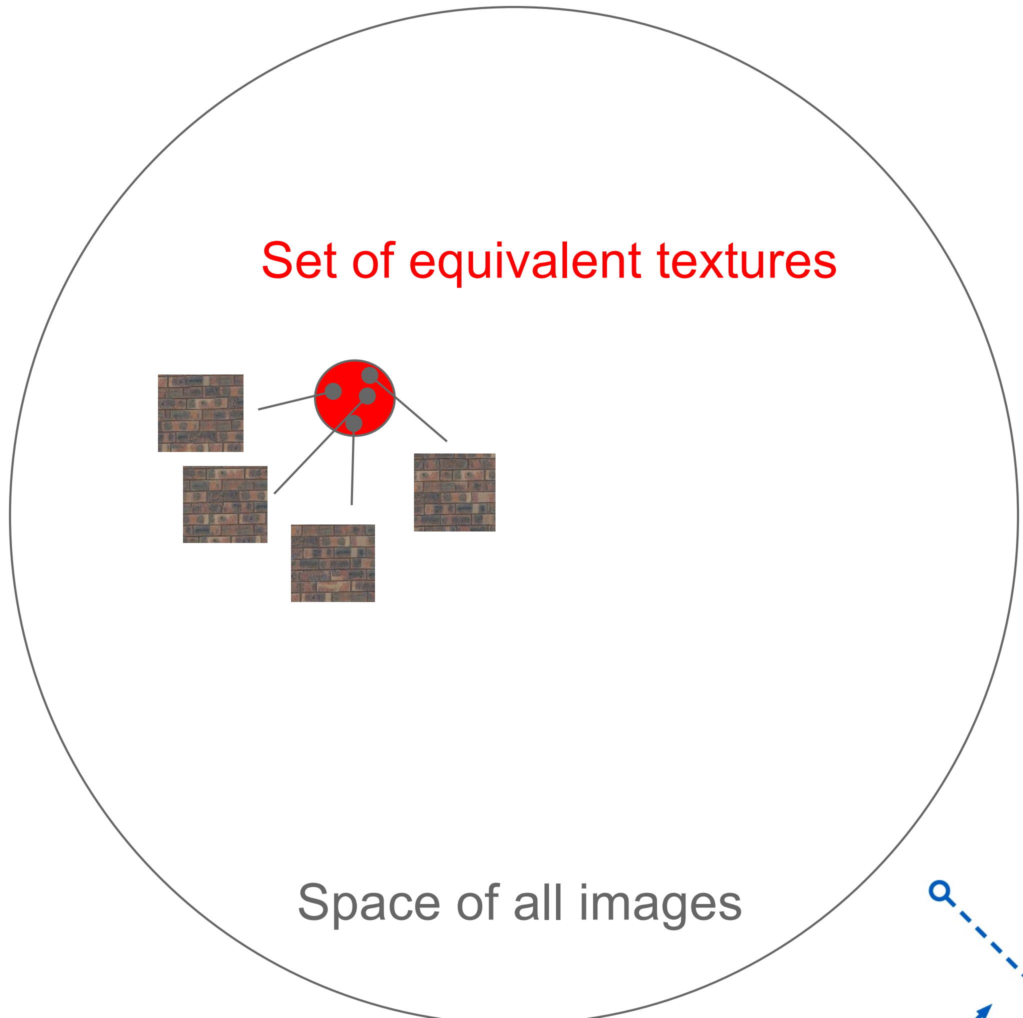


The trivial texture synthesis algorithm



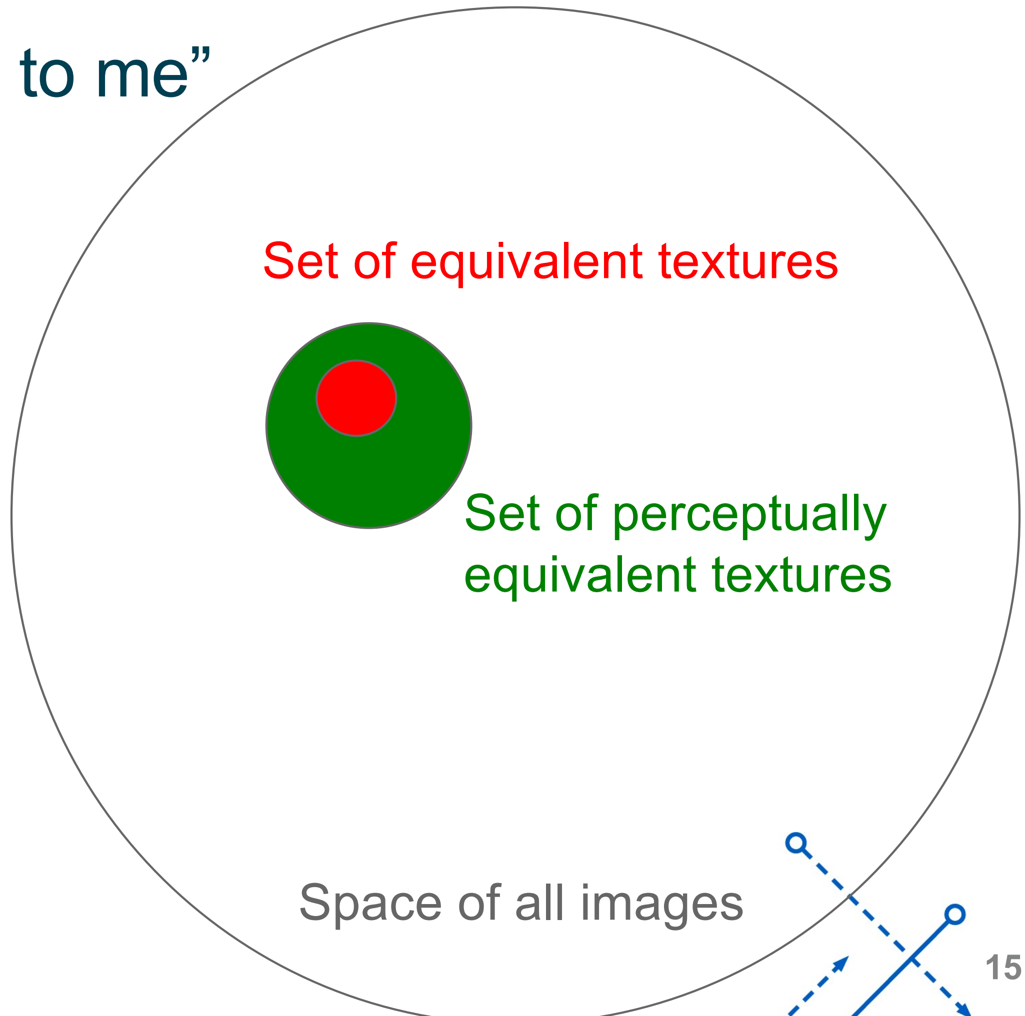
Texture synthesis and representation

- Set of equivalent textures:
  - generated by exactly the same physical process



Texture synthesis and representation

- Set of equivalent textures:
  - Generated by exactly the same physical process
- Set of perceptually equivalent textures:
  - “well, they just look the same to me”



Spatial AI & Robotics Lab

Scale: objects vs. texture

- The same thing can occur as texture or an object
  - depending on the **scale** we are considering.

Why analyze texture?

Importance to perception:

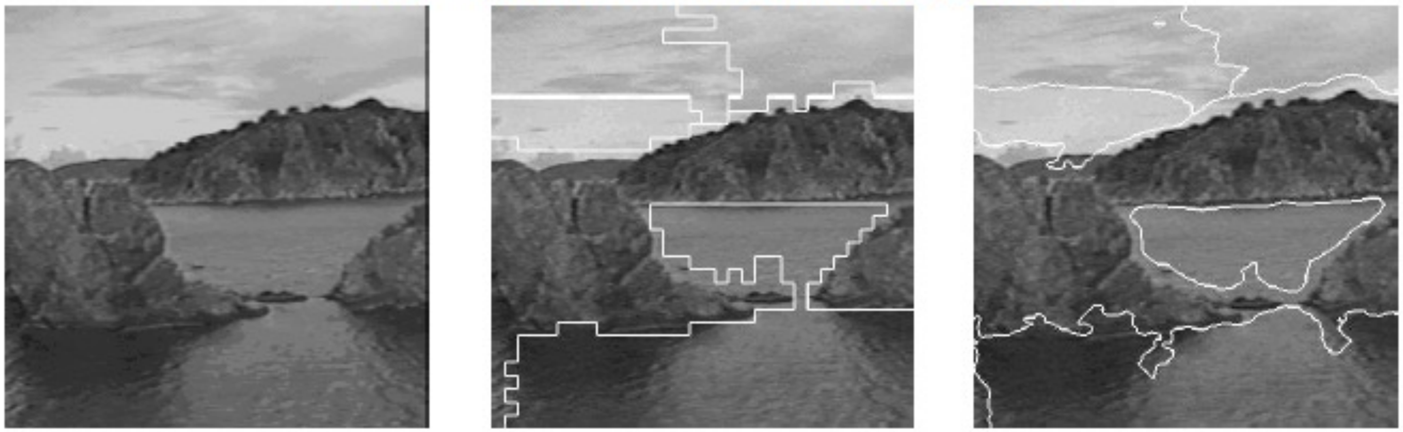
- Often indicative of a material's **properties**
- Can be important **appearance cue**, especially if shape is similar across objects
- Distinguish between shape, boundaries, and texture

Technically:

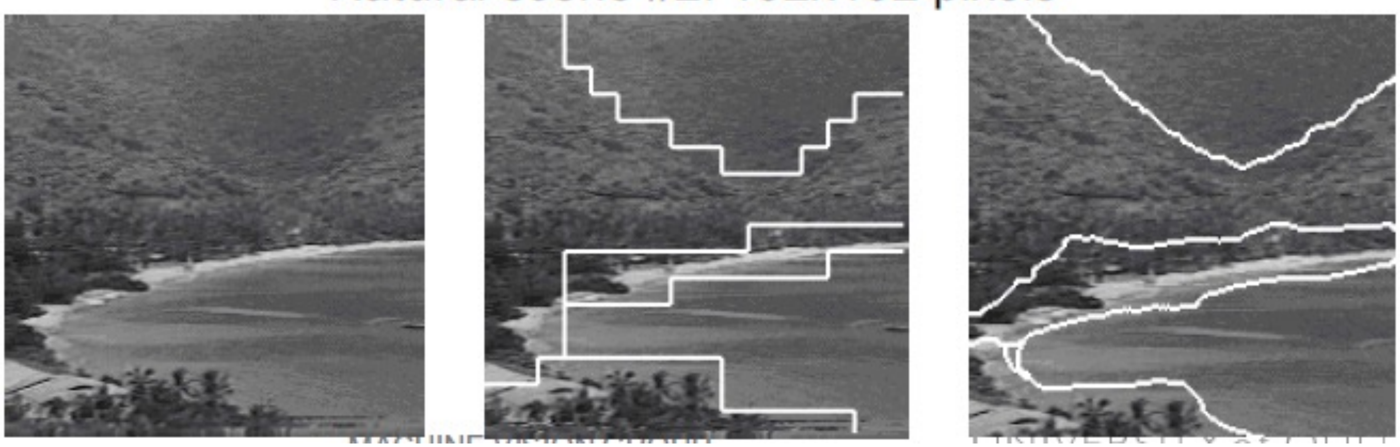
- Representation-wise, we want a feature one step above "building blocks" of filters, edges.

Why analyze texture?

Natural scene #1: 384x384 pixels



Natural scene #2: 192x192 pixels



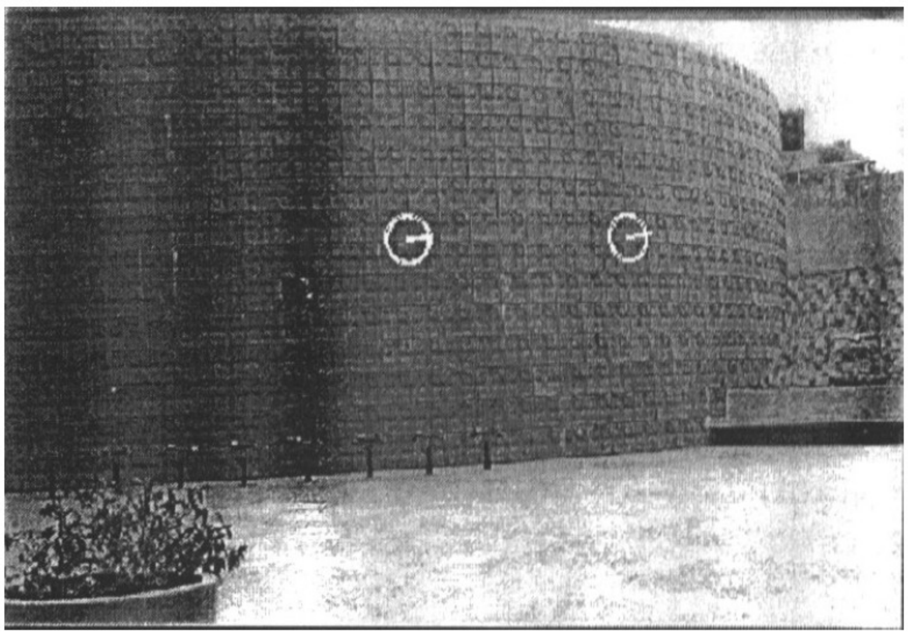
Texture-related tasks

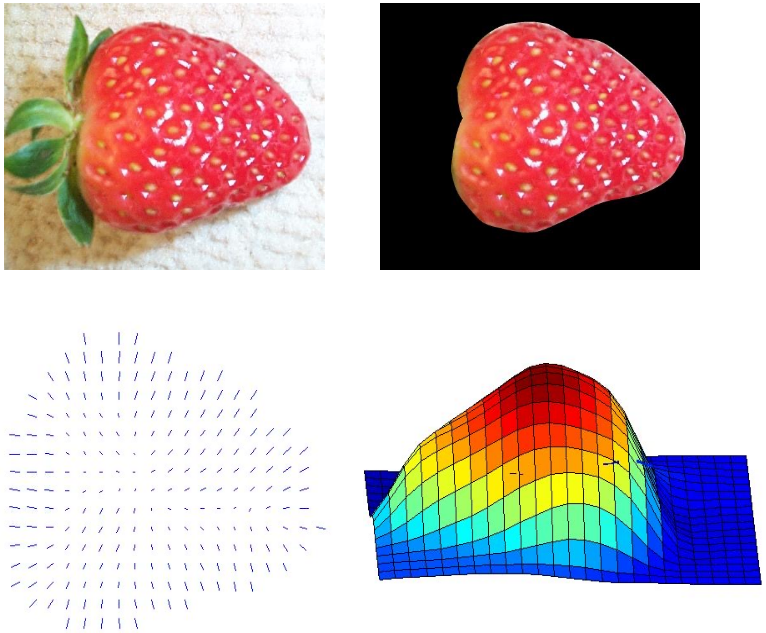
- **Shape from texture**
  - Estimate surface orientation or shape from image texture
- **Segmentation/classification** from texture cues
  - Analyze, represent texture
  - Group image regions with consistent texture
- **Synthesis**
  - Generate new texture patches/images given some examples

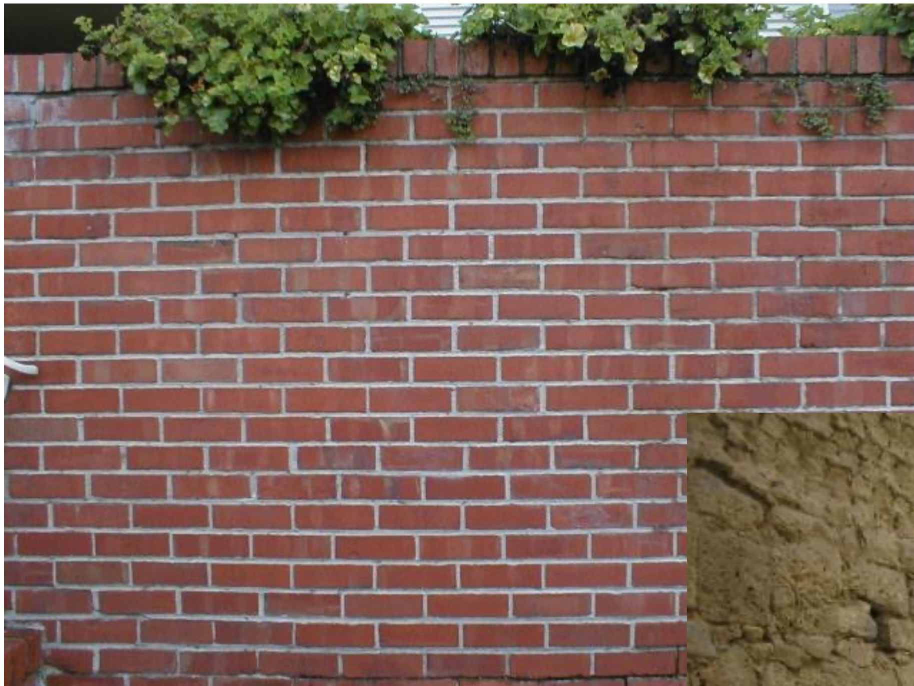
SAIR
Spatial AI & Robotics Lab

Shape from texture

- Use deformation of texture from point to point to estimate surface shape







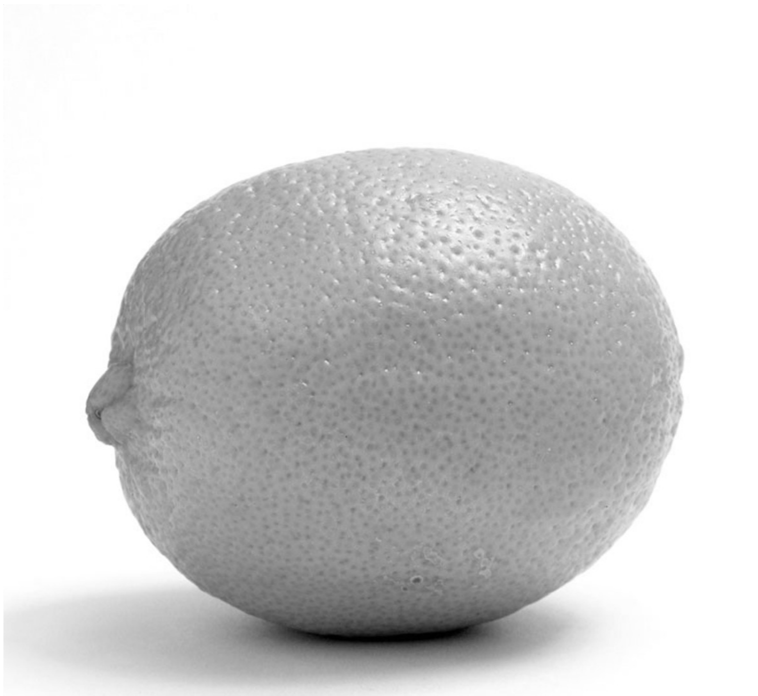


Texture-related tasks

- **Shape from texture**
  - Estimate surface orientation or shape from image texture
- **Segmentation/classification** from texture cues
  - Analyze, represent texture
  - Group image regions with consistent texture
- **Synthesis**
  - Generate new texture patches/images given some examples

SAIR
Spatial AI & Robotics Lab

Texture-related tasks

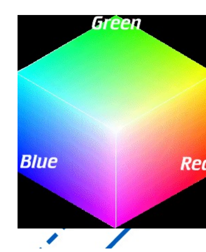


Texture-related tasks

http://animals.nationalgeographic.com/

Color vs. texture

- Recall: These looked very similar in terms of their color distributions (when our features were R-G-B)
- But how would their texture distributions compare?



Psychophysics of texture

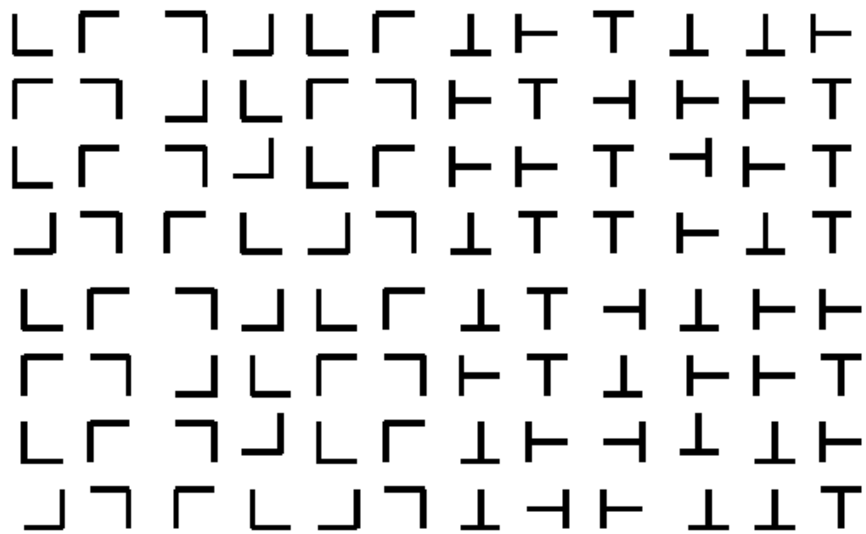
- Some textures distinguishable with *pre-attentive* perception– without scrutiny, eye movements [Julesz 1975]

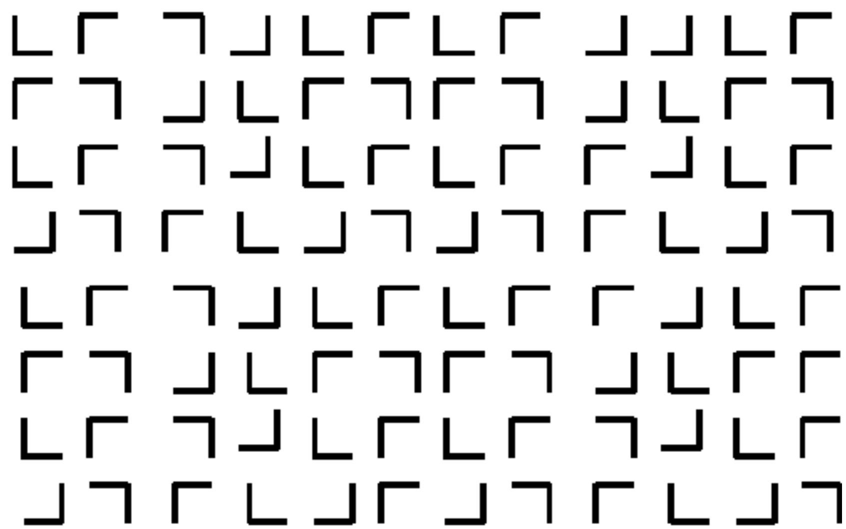
Watch the next screen very carefully.

Are the **left** and **right** sides:

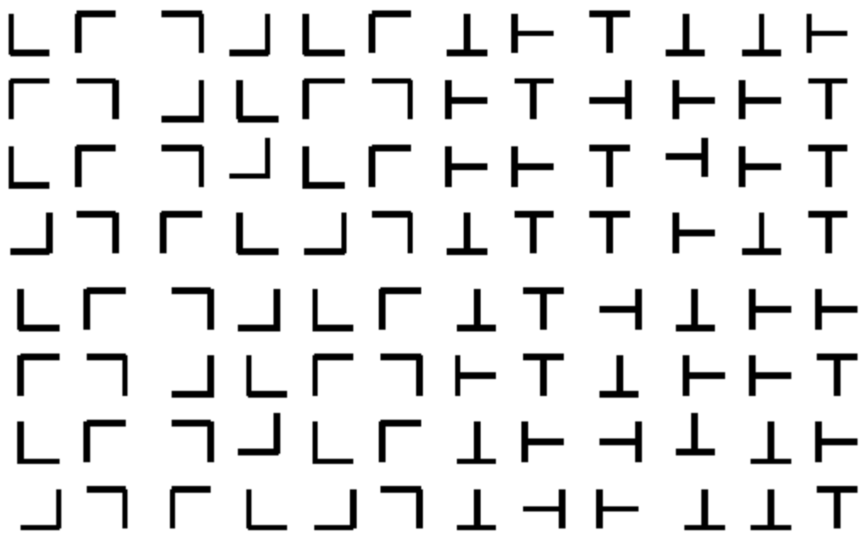
Same or different?

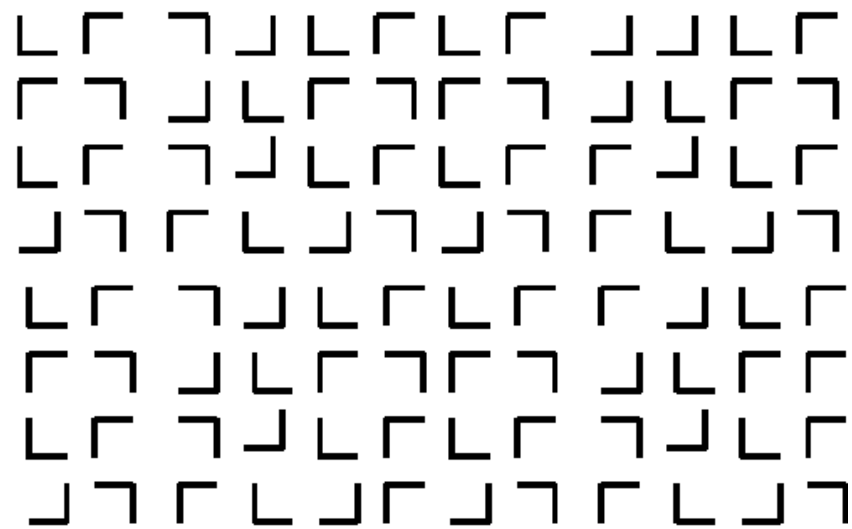
Pre-attentive processing is the subconscious accumulation of information from the environment. All available information is pre-attentively processed.





What you have seen





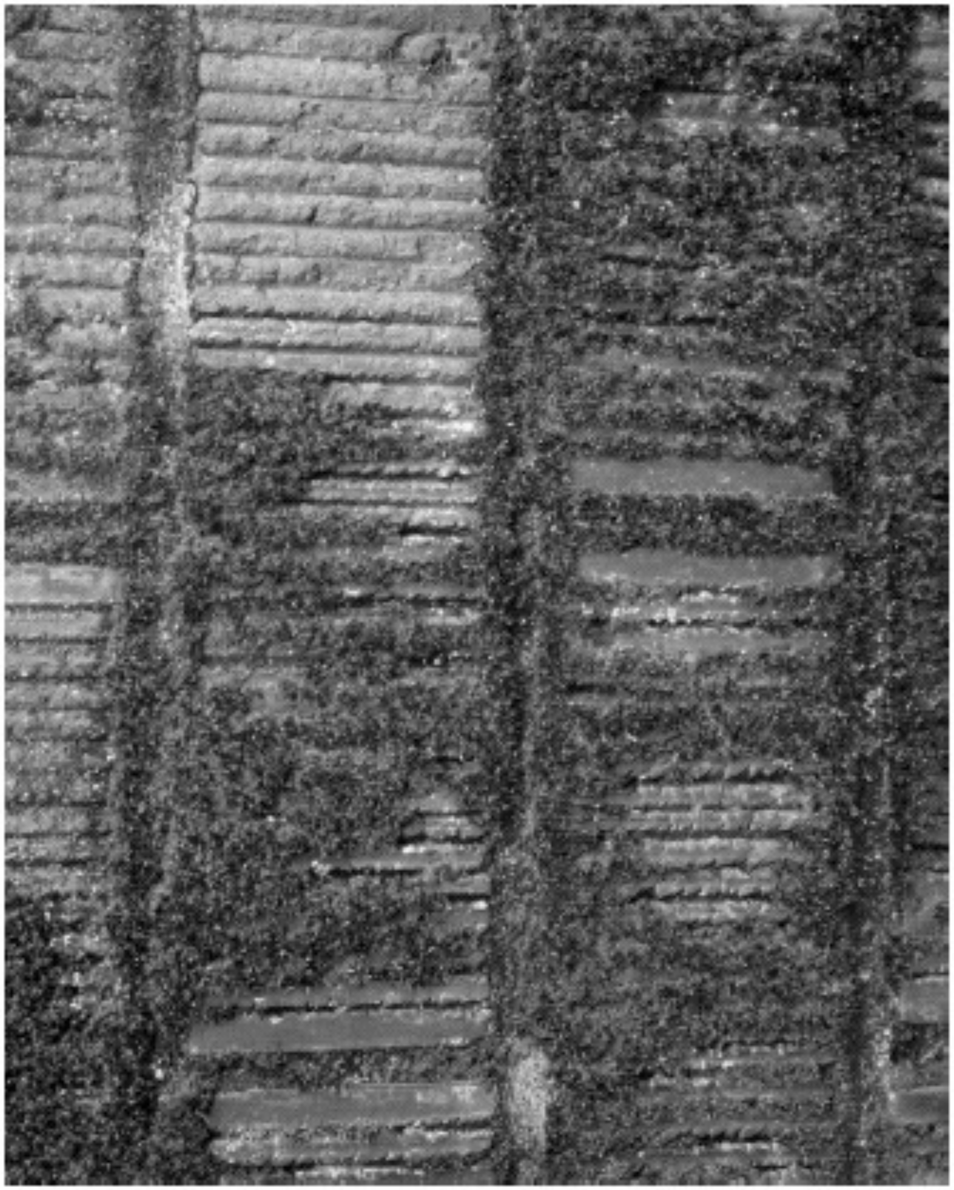
Texture Analysis

- Analyze the texture in terms of statistical relationships between fundamental texture elements
  - Called “textons”.
- It generally required a human to look at the texture in order to decide what those fundamental units were.

Texture representation

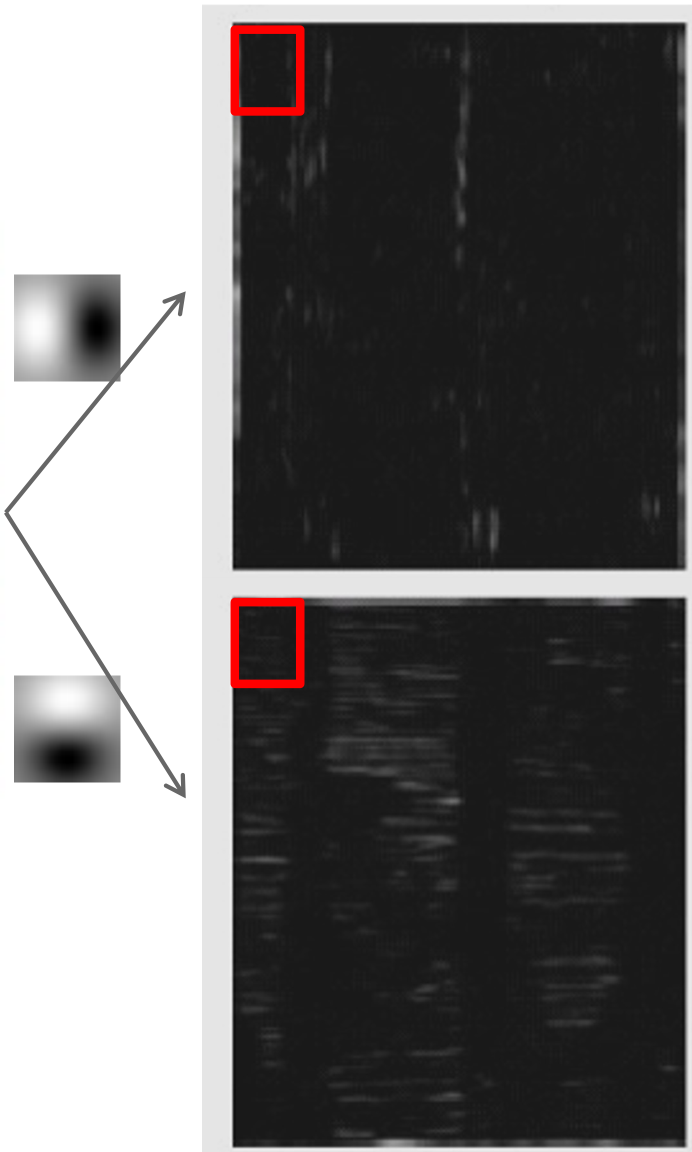
- Textures are made up of repeated local patterns:
  - How to find the patterns
    - Use filters that look like patterns
      - Spots, bars, raw patches…
      - Consider magnitude of response
    - Describe their statistics within each local window
      - Mean, standard deviation, etc.
    - Histogram of "prototypical" feature occurrences

Texture representation



Texture representation

original image



derivative filter responses, squared

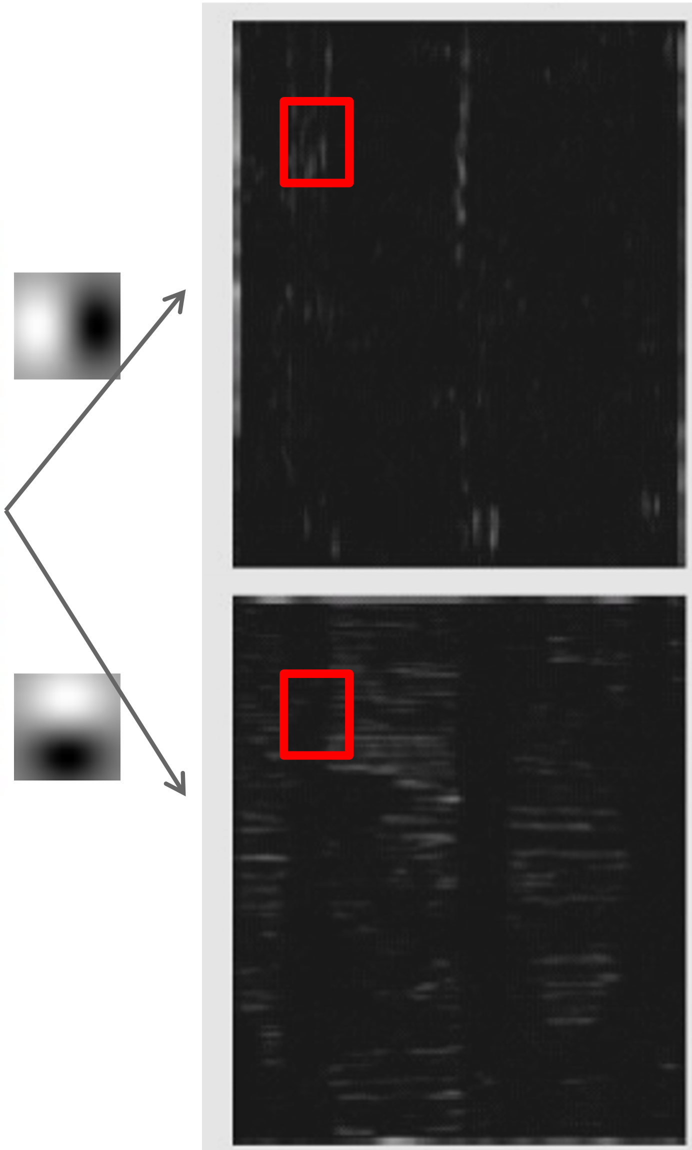
	mean d/dx value	mean d/dy value
Win. #1	4	10

⋮

statistics to summarize patterns in small windows

Texture representation

original image



derivative filter responses, squared

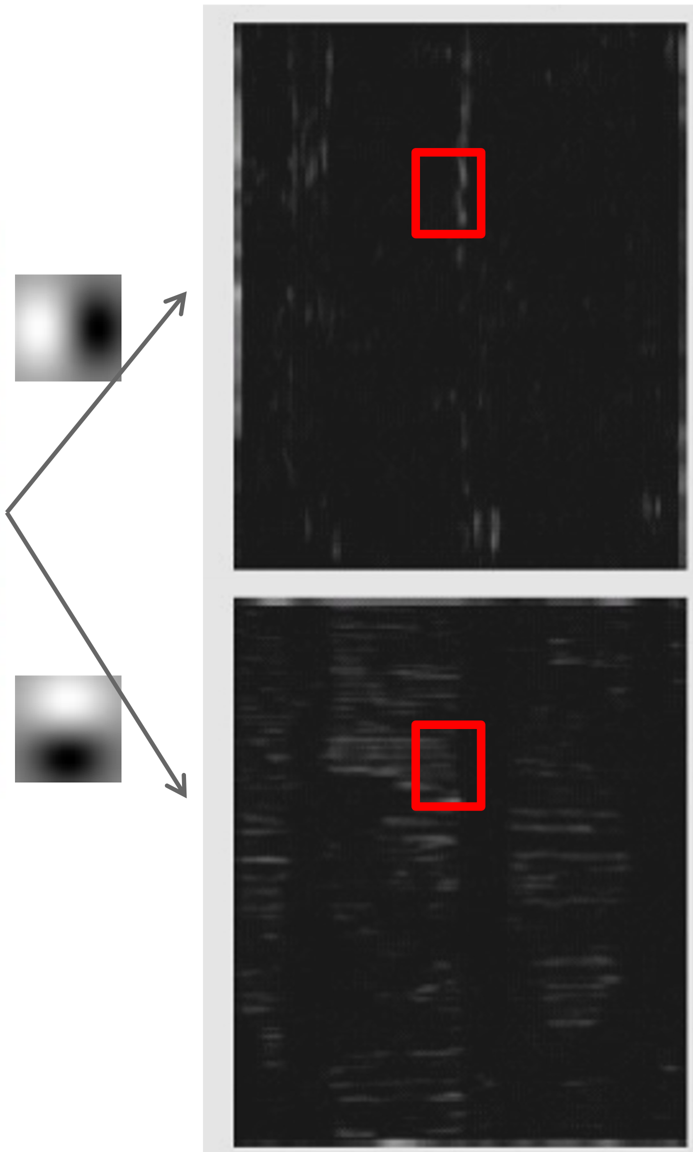
	mean d/dx value	mean d/dy value
Win. #1	4	10
Win.#2	18	7

⋮

statistics to summarize patterns in small windows

Texture representation

original image



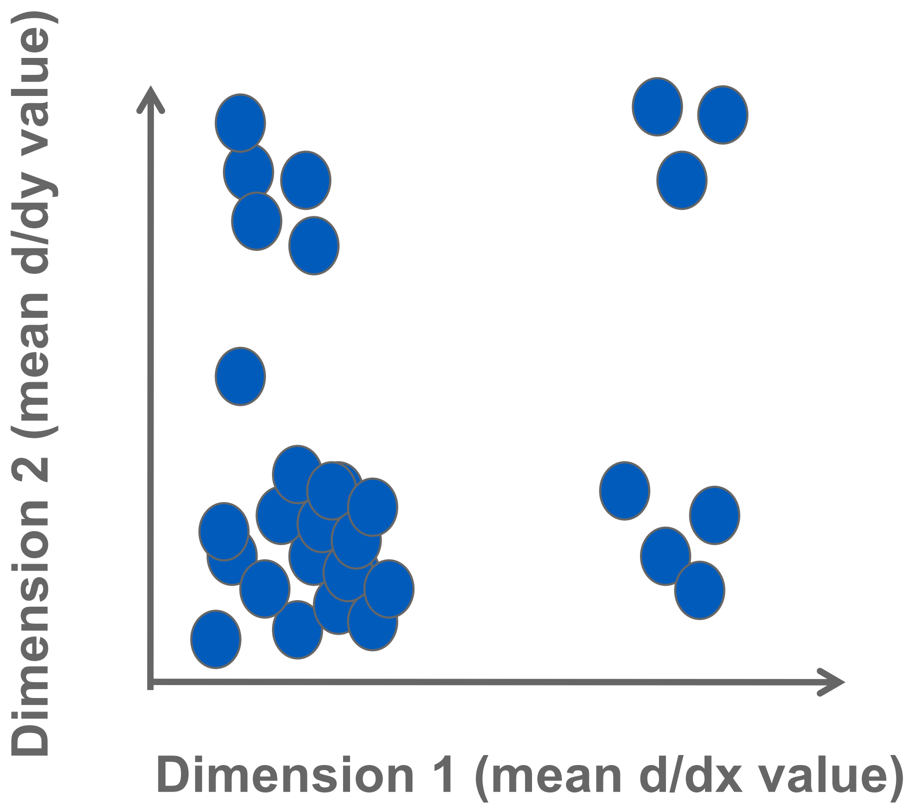
derivative filter responses, squared

	mean d/dx value	mean d/dy value
Win. #1	4	10
Win.#2	18	7
⋮		
Win.#9	20	20

⋮

statistics to summarize patterns in small windows

Texture representation: example



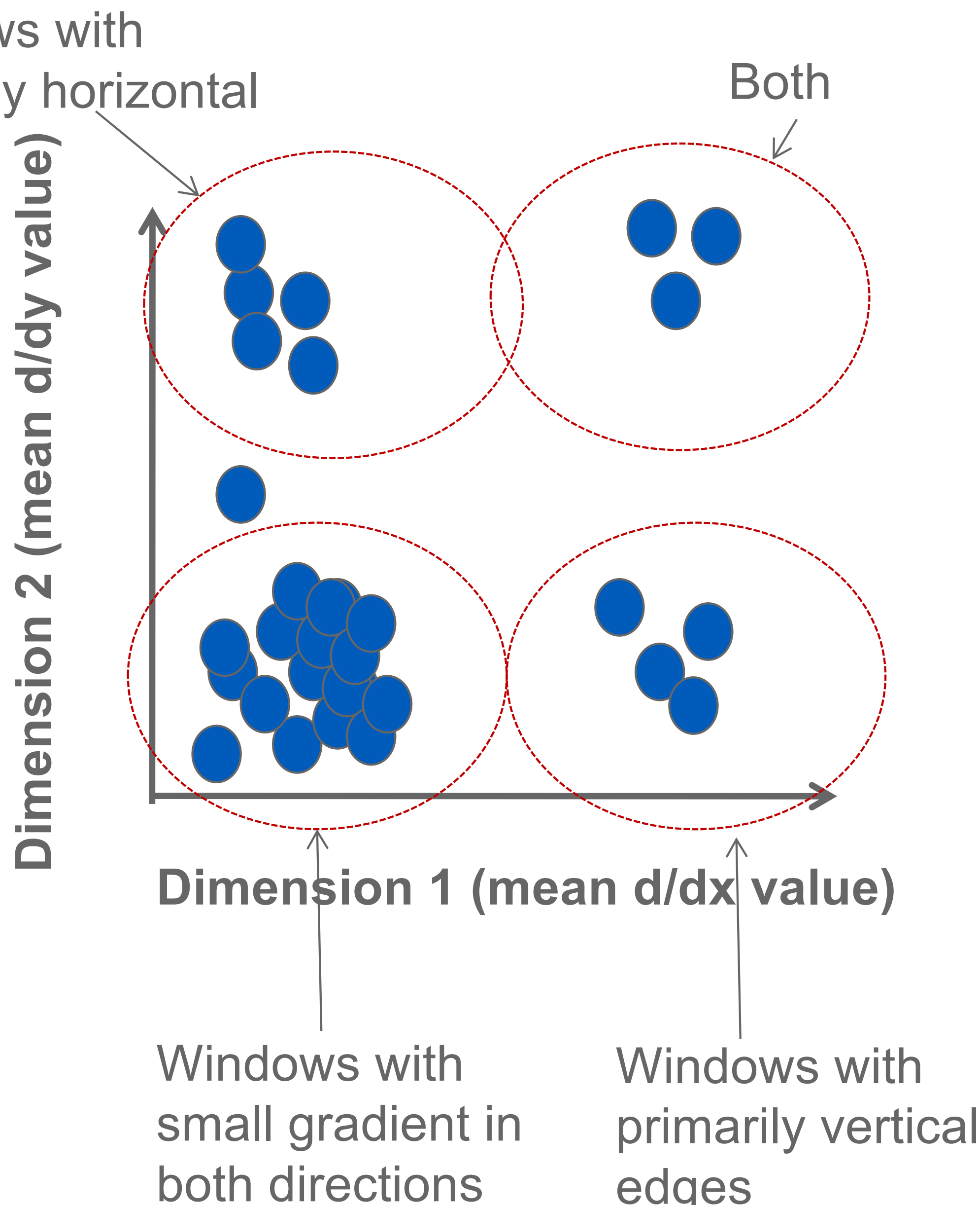
	mean d/dx value	mean d/dy value
Win. #1	4	10
Win.#2	18	7
⋮		
Win.#9	20	20

⋮

statistics to summarize patterns in small windows

SAIR
Spatial AI & Robotics Lab

Texture representation: example



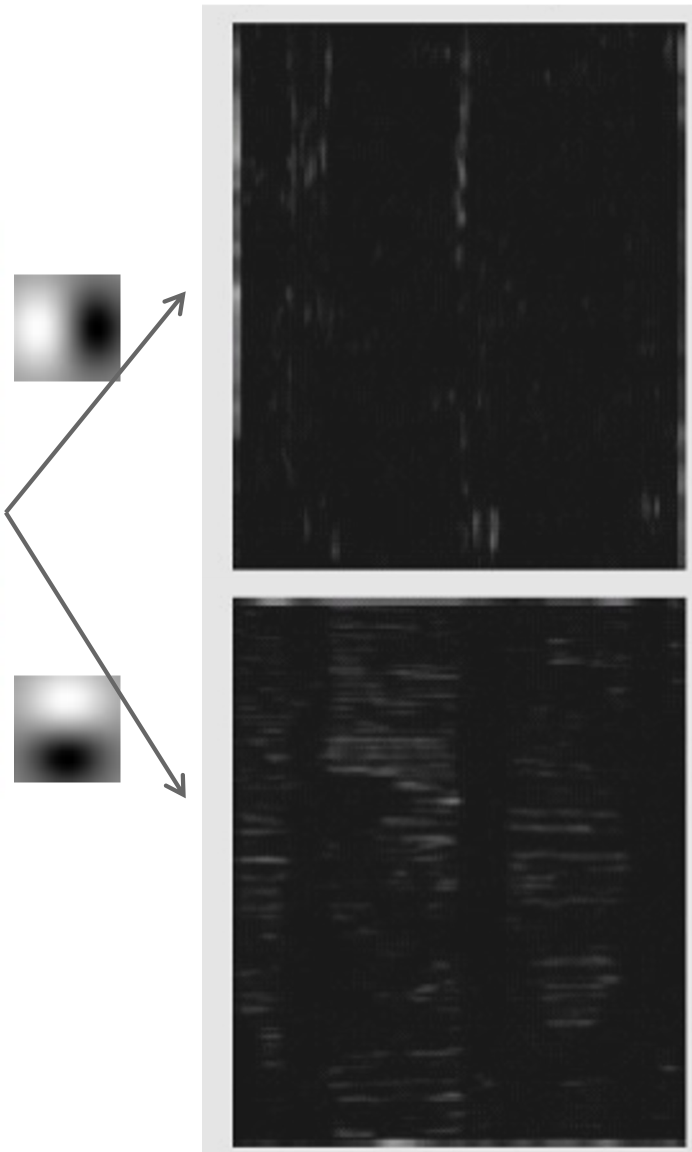
	mean d/dx value	mean d/dy value
Win. #1	4	10
Win.#2	18	7
⋮		
Win.#9	20	20

⋮

statistics to summarize patterns in small windows

Texture representation: example

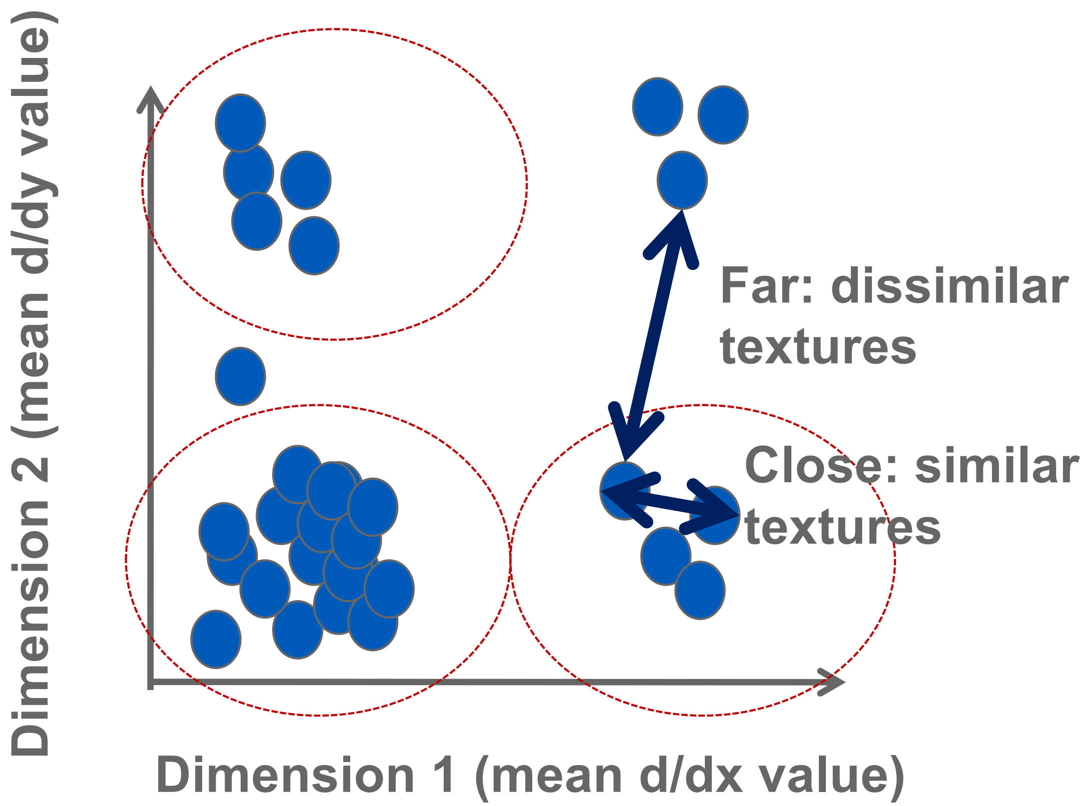
original image



derivative filter responses, squared

visualization of the assignment to texture "types"

Texture representation: example



	mean d/dx value	mean d/dy value
Win. #1	4	10
Win.#2	18	7
⋮		
Win.#9	20	20

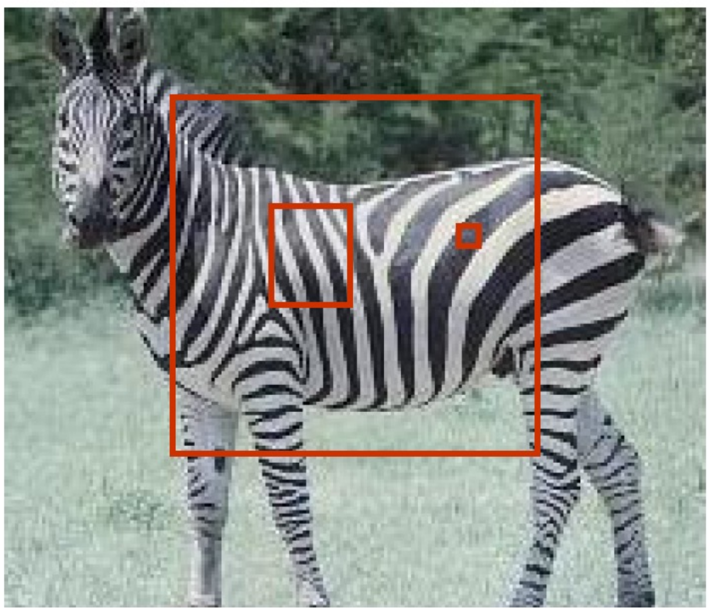
⋮

statistics to summarize patterns in small windows

SAIR
Spatial AI & Robotics Lab

Texture representation: window scale

- We're assuming we know the relevant window size for which we collect these statistics.

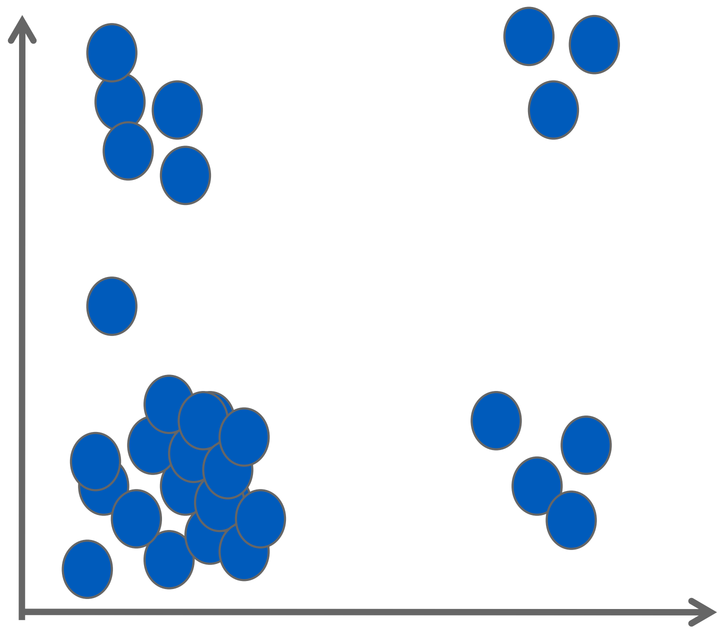


Possible to perform scale selection by looking for window scale where texture description not changing.

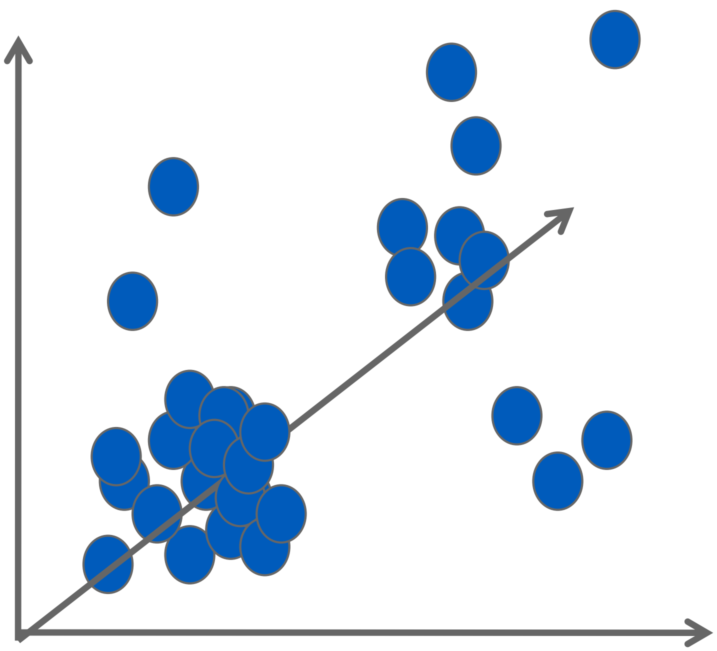
Filter banks

- Our previous example used two filters, and resulted in a 2-dimensional feature vector to describe texture in a window.
  - x and y derivatives revealed something about local structure.
- We can generalize to apply a collection of multiple (*d*) filters: a "filter bank"
  - Recall multi-channel convolution.
- Then our feature vectors will be *d*-dimensional.
  - still can think of nearness, farness in feature space

d-dimensional features



2d

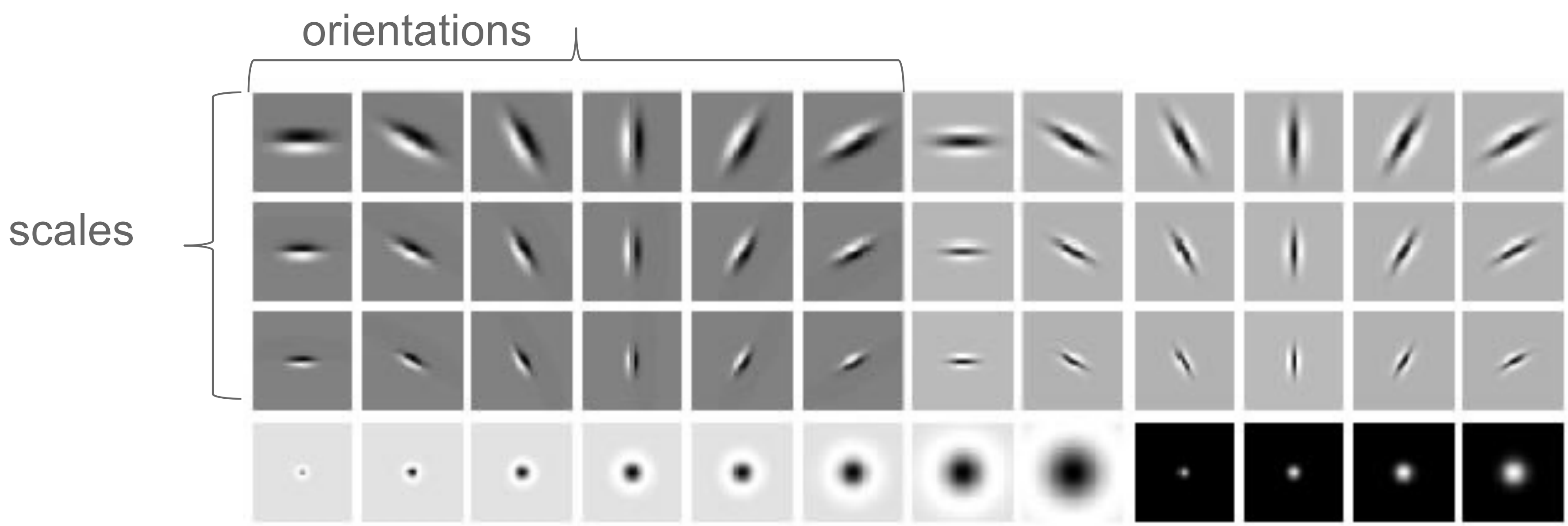


3d

. . .

Filter banks

- What filters to put in the bank?
  - Typically we want a combination of scales and orientations, different types of patterns.



http://www.robots.ox.ac.uk/~vgg/research/texclass/filters.html

Pixel Neighborhood-based Feature

- The most important for texture analysis is to describe the spatial behavior of intensity values in any neighborhood.

- Different methodologies have been proposed.

- **Local binary pattern (LBP)** is one of the most-widely used approach – mainly for face recognition.

- LBP is used for texture analysis too.

Local Binary Pattern (LBP)

For each PIXEL of an image, a BINARY CODE is produced

$$LBP_{p,r}(N_c) = \sum_{p=0}^{P-1} g\left(N_p - N_c\right)2^p$$

where

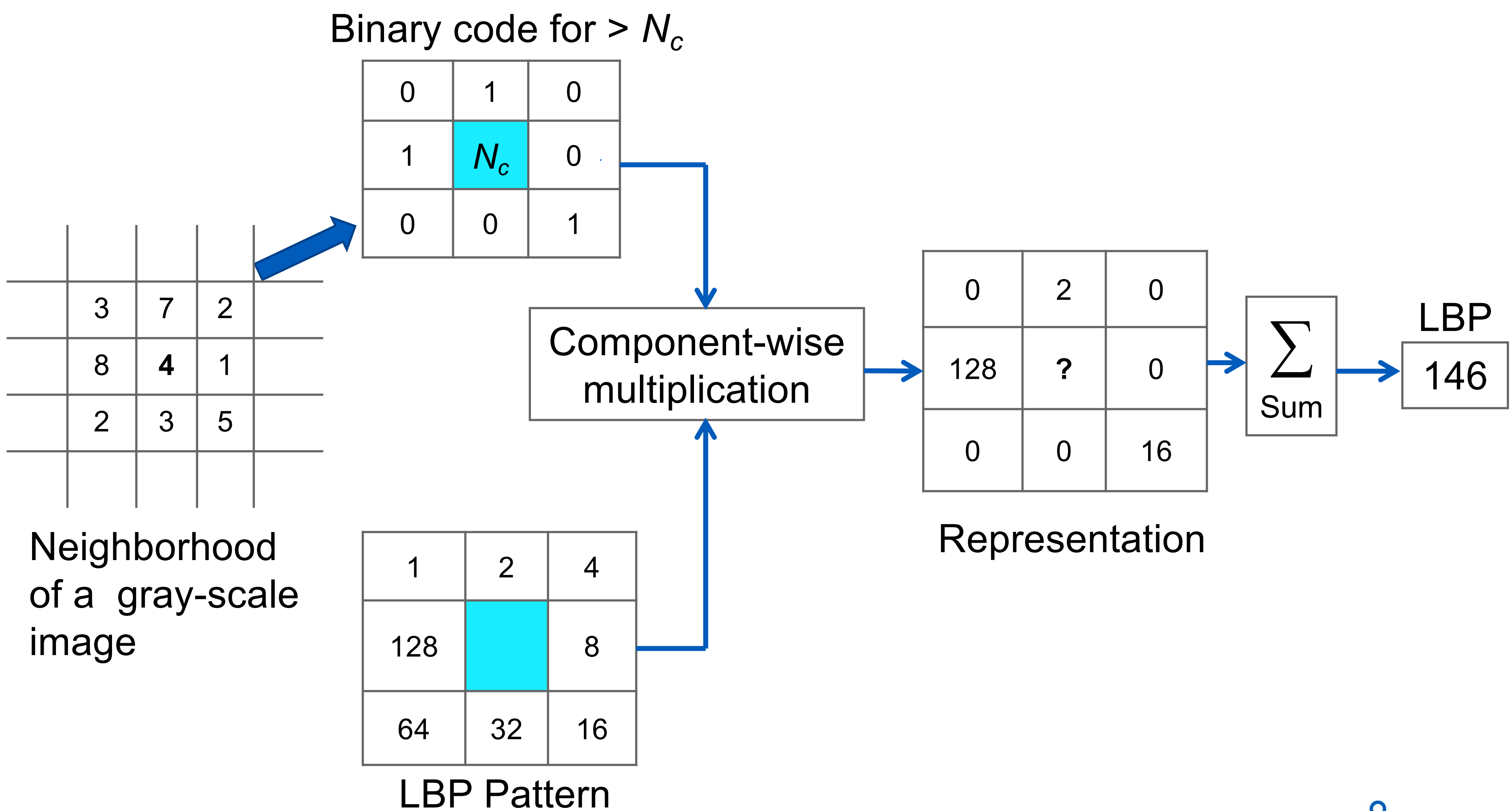
neighborhood pixels $\left(N_p\right)$ in each block → is thresholded by its center pixel (N_c)
p→ sampling points (e.g., $p = 0, 1, \ldots, 7$ for a 3x3 cell, where P = 8)
r → radius (for 3x3 cell, it is 1).

Binary threshold function $g(x)$ is,

$$g(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

SAIR Spatial AI & Robotics Lab

Computation of Local Binary Pattern



Example of how the *LBP operator* works

Computation of Local Binary Pattern

6	5	2
7	6	1
9	8	7

Example

(LSB) 1	0	0
1		0
1	1	1

Threshold

1	2	4
128		8
64	32	16

Weights

Binary Pattern:	1 (MSB)	1	1	1	0	0	0	1 (LSB)
Code/Weight (2^p):	1×2^7	1×2^6	1×2^5	1×2^4	0×2^3	0×2^2	0×2^1	1×2^0
	= 128	= 64	= 32	= 16	= 0	= 0	= 0	= 1

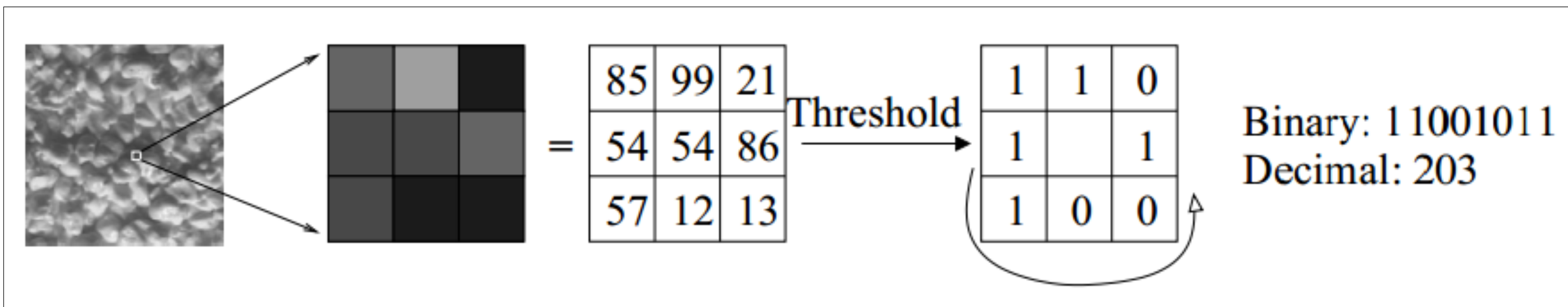
LBP:	1 + 0 + 0 + 0 + 16 + 32 + 64 + 128 = 241

S A I R
Spatial AI & Robotics Lab

MSB stands for most significant bit, while **LSB** is least significant bit.

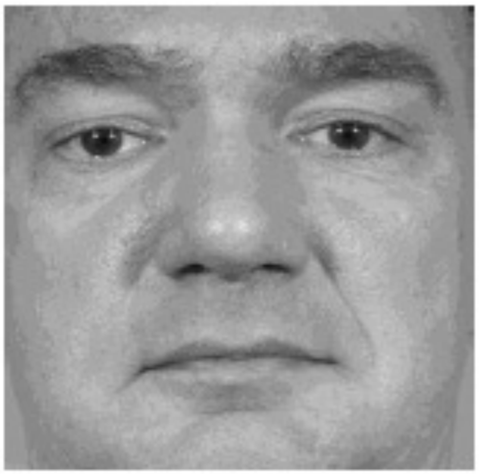
LBP

- One of the best performing texture descriptors
- A label is assigned to every pixel
- Use center pixel to threshold the 3x3 neighborhood
    - Result in binary number
- Histogram of the labels is used as a texture descriptor



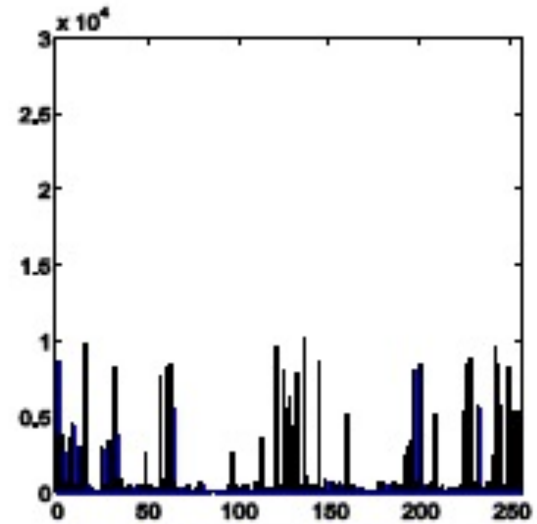
The Basic LBP operator

LBP hitogram



Input image

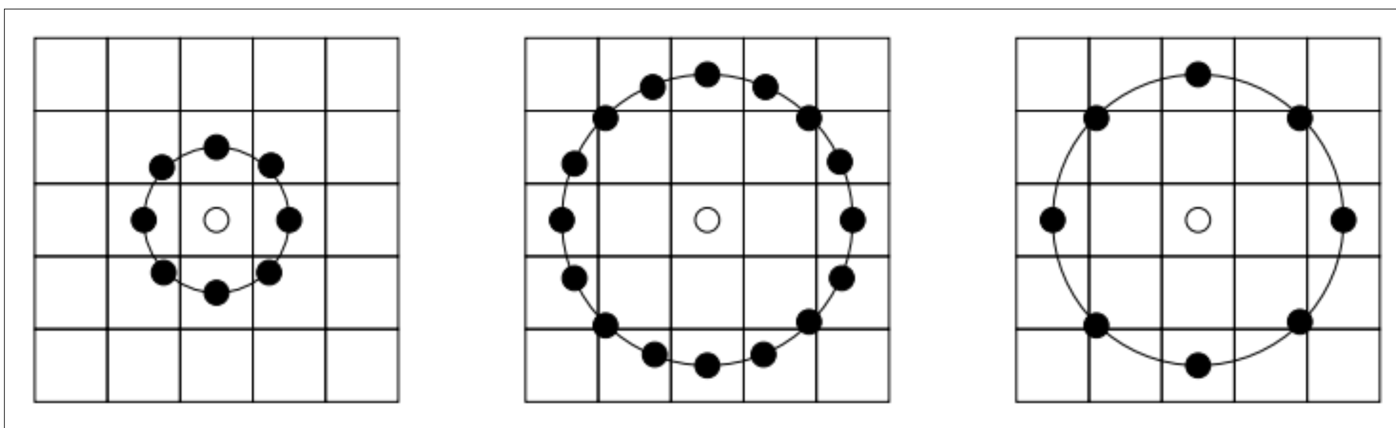
LBP image



LBP histogram

LBP

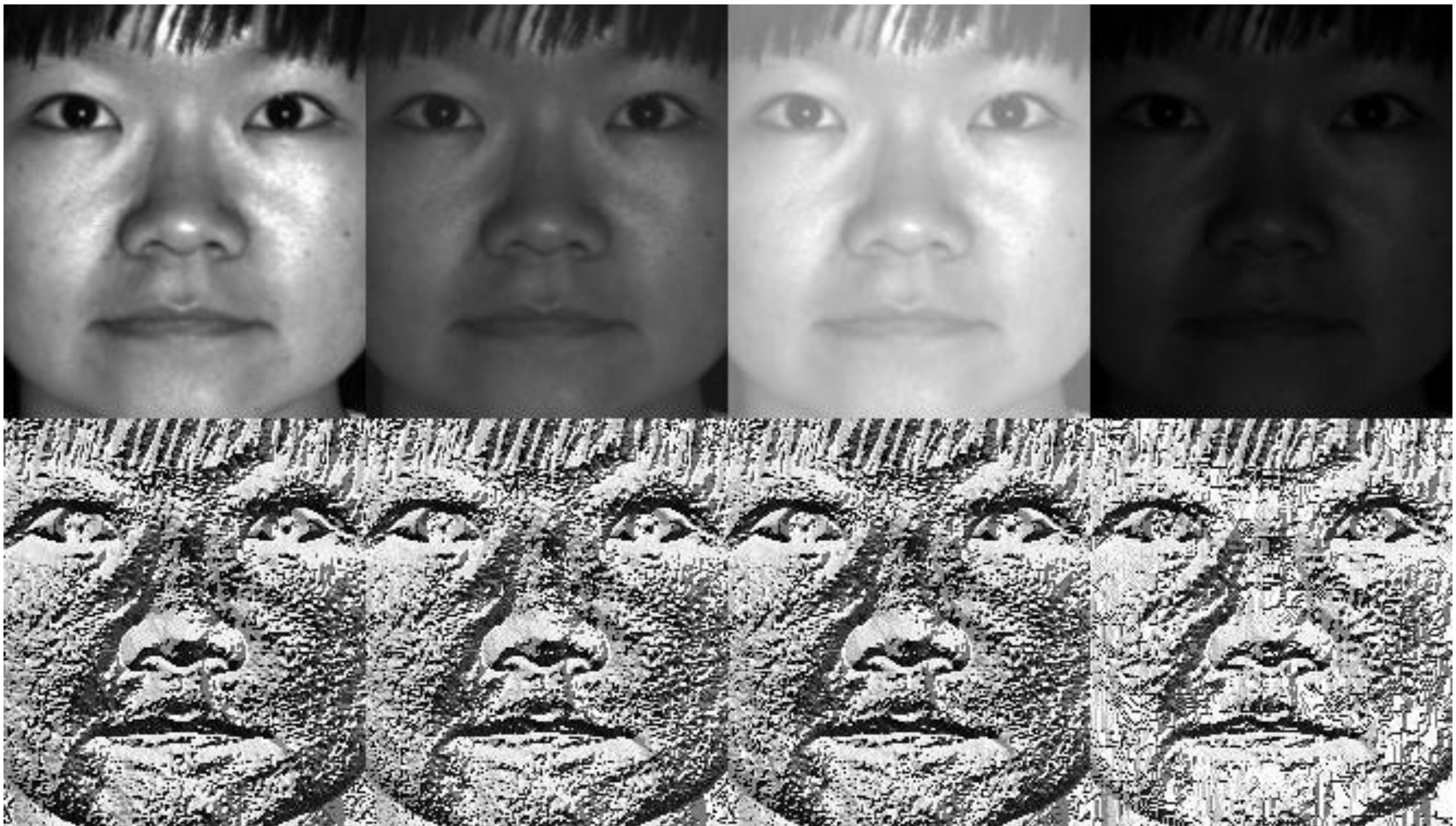
- LBP is extended to use different sizes of neighborhoods.
- Local neighborhoods is defined as a set of **sampling points**.
  - points evenly spaced on a circle centered at the labeled pixel.
- **(P, R) , P = number of sampling points, R = radius**
- **Bilinear interpolation** is used
  - If sampling point is not in the center of the pixel.



The circular (8,1), (16, 2) and (8, 2) neighborhoods.

LBP Example

- Notice how LBP feature are illumination invariant



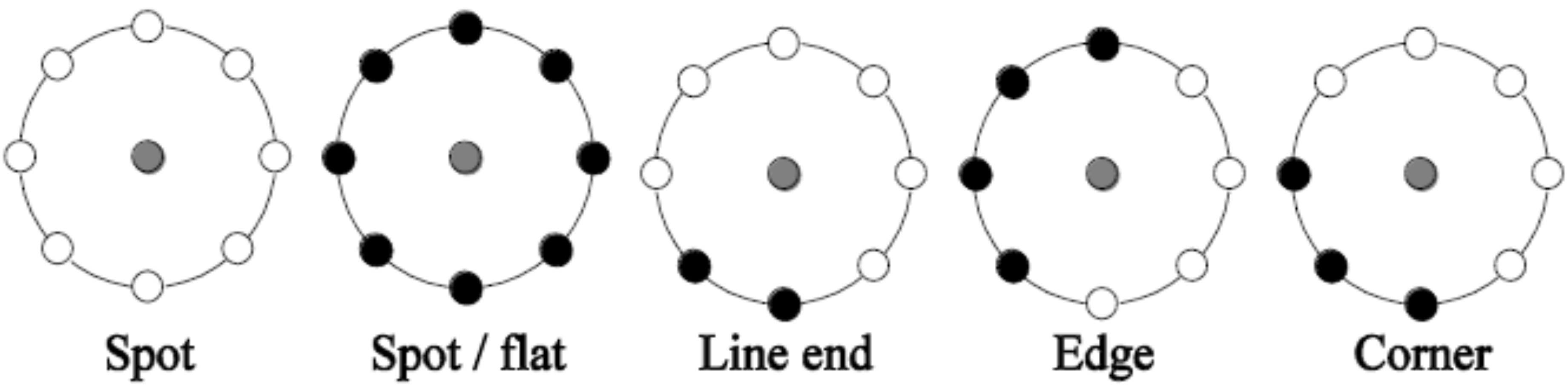
Uniform LBP

- Uniform patterns to further improve LBP.
  - Has at most 2 bitwise transitions in binary pattern.
  - Histogram
    - assigns separate bins for every uniform pattern.
    - assigns a single bin for all non-uniform pattern.

Uniform LBP

Texture primitives (“micro-textons”) detected by the uniform patterns of LBP

1 = black
0 = white



Uniform LBP

- Uniform patterns examples
  - 00000000 (0 transitions)
  - 01110000 (2 transitions)
  - 11001111 (2 transitions)

- Non-uniform patterns examples
  - 11001001 (4 transitions)
  - 01010011 (5 transitions)

- 59 bins histogram
  - 58 uniform bins →
  - 1 non-uniform bin

Length of feature vector

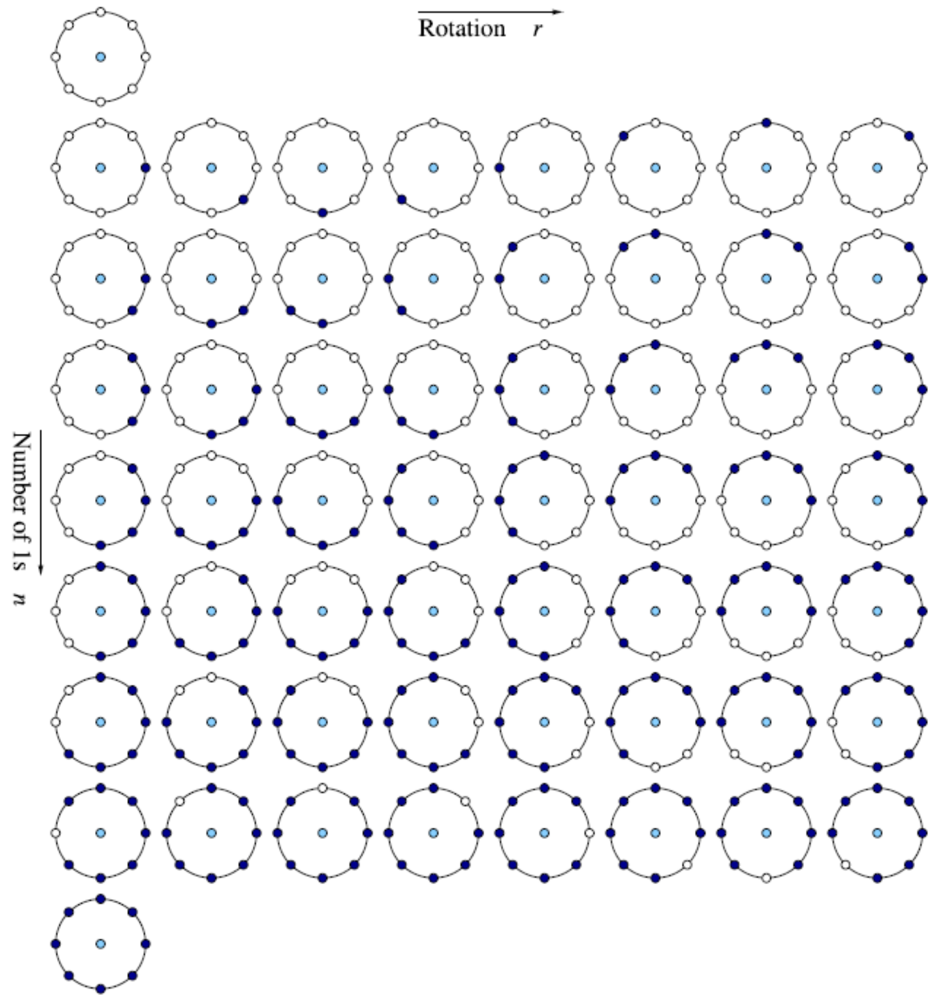
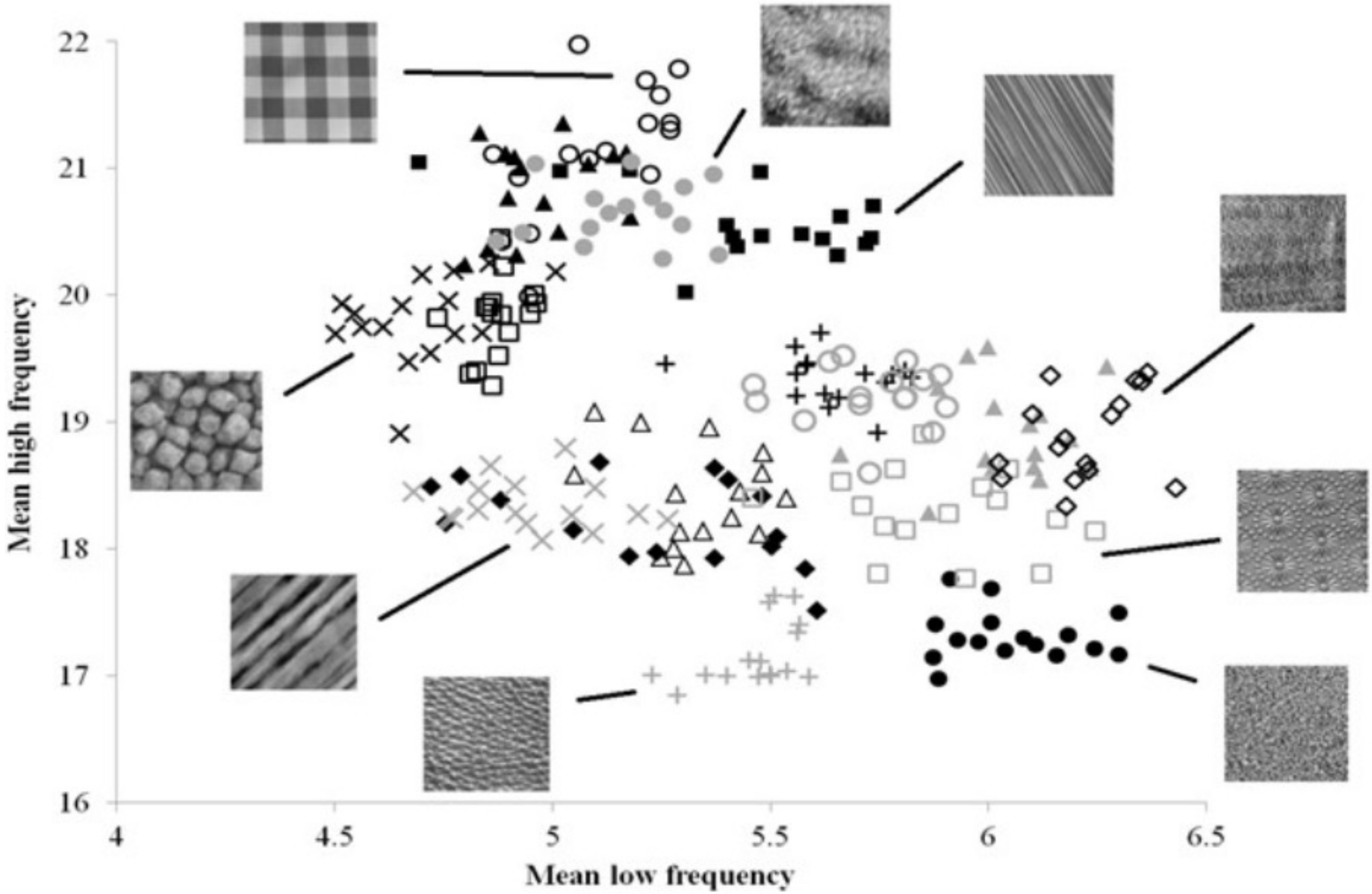


Fig. 2.4 The 58 different uniform patterns in (8, R) neighborhood

Other Measures of Texture (frequency)



Other Measures of Texture

Texture measure	Description	Equation*
Mean	Mean is the average grey level for each sample	$\sum_{i,j=0}^{N-1} i(P_{i,j})$
Variance	Variance is a measure of heterogeneity; it increases when the grey level values differ from their mean. In general, coarse-textured features are associated with higher variances	$\sum_{i,j=0}^{N-1} P_{i,j}(i-\mu_i)^2$
Homogeneity	This parameter measures image homogeneity as it assumes larger values for smaller grey tone differences in pair elements	$\sum_{i,j=0.}^{N-1} \frac{P_{i,j}}{1+(i-j)^2}$
Contrast	Contrast is a measure of spatial frequency, the difference between the highest and the lowest values of a contiguous set of pixels. A high contrast implies high coarse texture	$\sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2$
Dissimilarity	Dissimilarity, akin to contrast, describes the heterogeneity of the grey levels. Higher values of dissimilarity in the GLC matrix indicate coarser textures	$\sum_{i,j=0}^{N-1} P_{i,j}\lvert i-j\rvert$
Entropy	Entropy measures the disorder of an image. When the image is not texturally uniform, many GLC matrix elements have very small values, which imply that entropy is very large	$\sum_{i,j=0}^{N-1} P_{i,j}(-\ln p_{i,j})^2$
Angular second moment	This parameter measures textural uniformity. Thus, high angular second moment values occur when the grey level distribution over the window has either a constant or a periodic form	$\sum_{i,j=0}^{N-1} P_{i,j}{}^2$
Correlation	Correlation is a measure of grey-tone linear dependencies in the image. High correlation values imply a linear relationship between the grey levels of pixel pairs	$\sum_{i,j=0}^{N-1} P_{i,j}\left[\frac{(i-\mu_i)(j}{\sqrt{(\mathrm{SD}_i{}^2)}}\right.$

SAIR
Spatial AI & Robotics Lab

COMPUTER VISION

Object & Scene

Objects vs Texture & Scene

The texture

The object

The scene

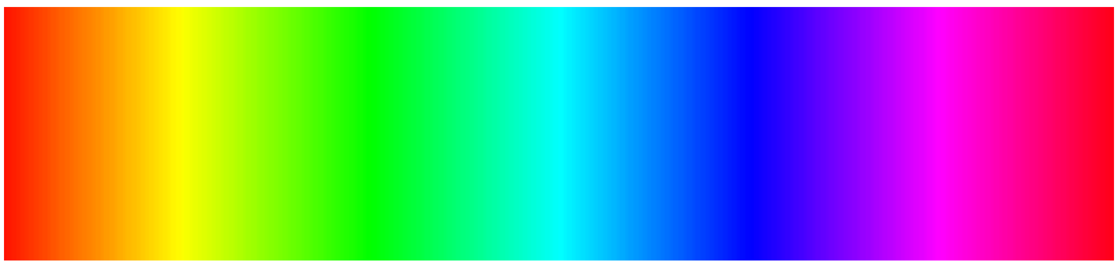
Why be concerned about the difference?

- Features to be extracted?
- Context for Recognition?
- Various categories have features that can be essential for interpretation.

An example of categorical perception

- Continuous perception: graded response

- Categorical perception: “sharp” boundaries



Many perceptual phenomena are a mixture of the two: categorical at an everyday level of magnification, but continuous at a more microscopic level. It can also depend on cultural aspects, expertise, task, attention, …

SAIR
Spatial AI & Robotics Lab

Bar Trivia

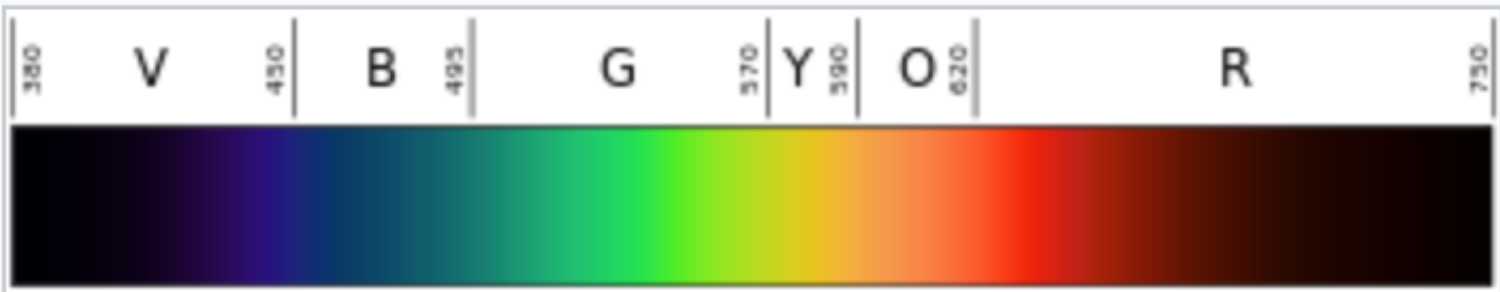
- How many colors can humans see?

  - 3 types of cones
  - Sensitive to red, green, and blue wavelengths of light.
  - Each can detect about 100 different shades
  - A **million (100^3) different color possibilities**

  - Some people have a 4th cone (tetrachromat)
    - A 2010 study suggested that nearly 12 percent of women may have this fourth color perception channel. Men aren't as likely to be tetrachromats. They're actually more likely to be colorblind, or unable to perceive as many colors as women.

Colors are perceivable, but often linguistic

- Note: Name of Colors are Cultural
  - Named for things: “ripe fruit”, “blood”
  - Single word for green/blue, or red/yellow



The notion of "green" in modern European languages corresponds to about 520–570 nm, but many historical and non-European languages make other choices, e.g. using a term for the range of ca. 450–530 nm ("blue/green") and another for ca. 530–590 nm ("green/yellow").

Why do we care about categories?

Perception of function:

- We can perceive the 3D shape, texture, material properties, without knowing about objects.
- But, the concept of category encapsulates also information about what can we do with those objects.

"We therefore include the perception of function as a proper –indeed, crucial- subject for vision science", *from Vision Science, chapter 9, Palmer*.

Why do we care about categories?

- When we recognize an object, we can make
  - predictions about its behavior in the future
  - beyond of what is immediately perceived.

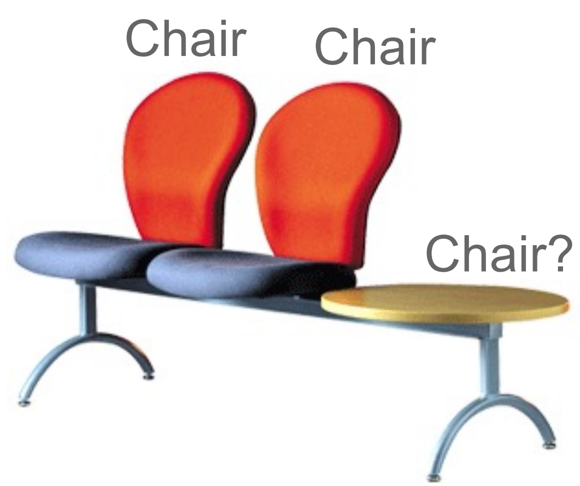
The perception of function

- Direct perception (affordances): [introduced by Gibson]

Flat surface
Horizontal
Knee-high
…
→ Sittable upon

- Mediated perception (Categorization)

Flat surface
Horizontal
Knee-high
…
→ Chair → Sittable upon



- One caveat of this comparison:
  - Deciding that something is a chair might require access to more features than deciding that we can sit on something
  - A different level of categorization.

Direct perception

Some aspects of an object function can be perceived directly

- Functional form:
  - Some forms clearly indicate to a function
  - e.g., “sittable-upon”, container, cutting device, …

Sittable-upon

Sittable-upon

Sittable-upon

It does not seem easy to sit-upon this…

Limitations of Direct Perception

- Objects of similar structure might have very different functions

Figure 9.1.2 Objects with similar structure but different functions. Mailboxes afford letter mailing, whereas trash cans do not, even though they have many similar physical features, such as size, location, and presence of an opening large enough to insert letters and medium-sized packages.

Not all functions seem to be available from direct visual information only.

The functions are the same at some level of description: we can put things inside in both and somebody will come later to empty them. However, we are not expected to put inside the same kinds of things…

Indirect perception of function by categorization

This requires object recognition (Next Lecture)

Which level of categorization is the right one?

- Car is an object composed of:
  - a few doors, four wheels (not always visible), a roof
  - front lights, windshield

If you are thinking in buying a car, you might want to be a bit more specific about your categorization.

SAIR
Spatial AI & Robotics Lab

Entry-level categories

- Typical member of a basic-level category are categorized at the expected level
- Atypical members tend to be classified at a subordinate (finer) level.

A bird

An ostrich

(Jolicoeur, Gluck, Kosslyn 1984)

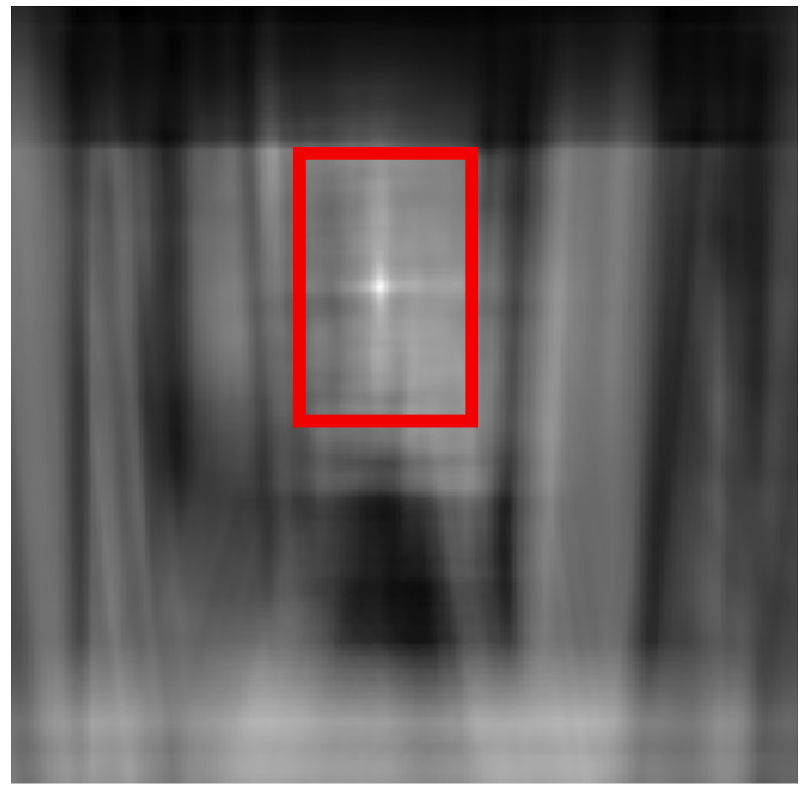
SAIR Spatial AI & Robotics Lab

Object recognition: Is it really so hard?

This is a chair

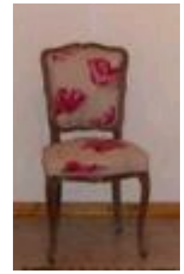
Find the chair in this image

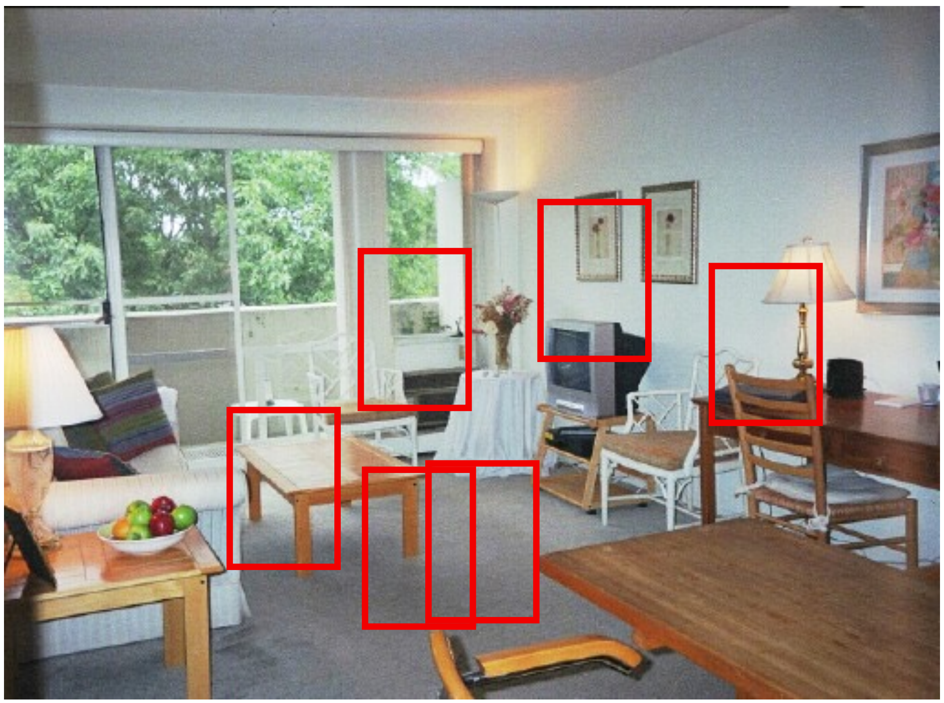
Output of normalized correlation

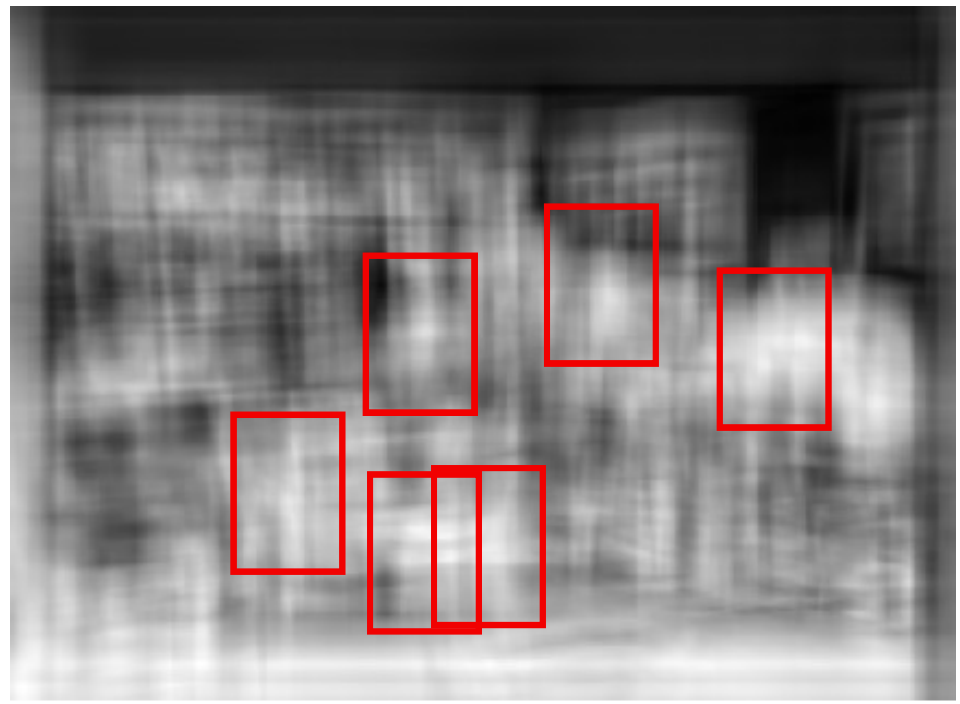


Object recognition: Is it really so hard?

Find the chair in this image







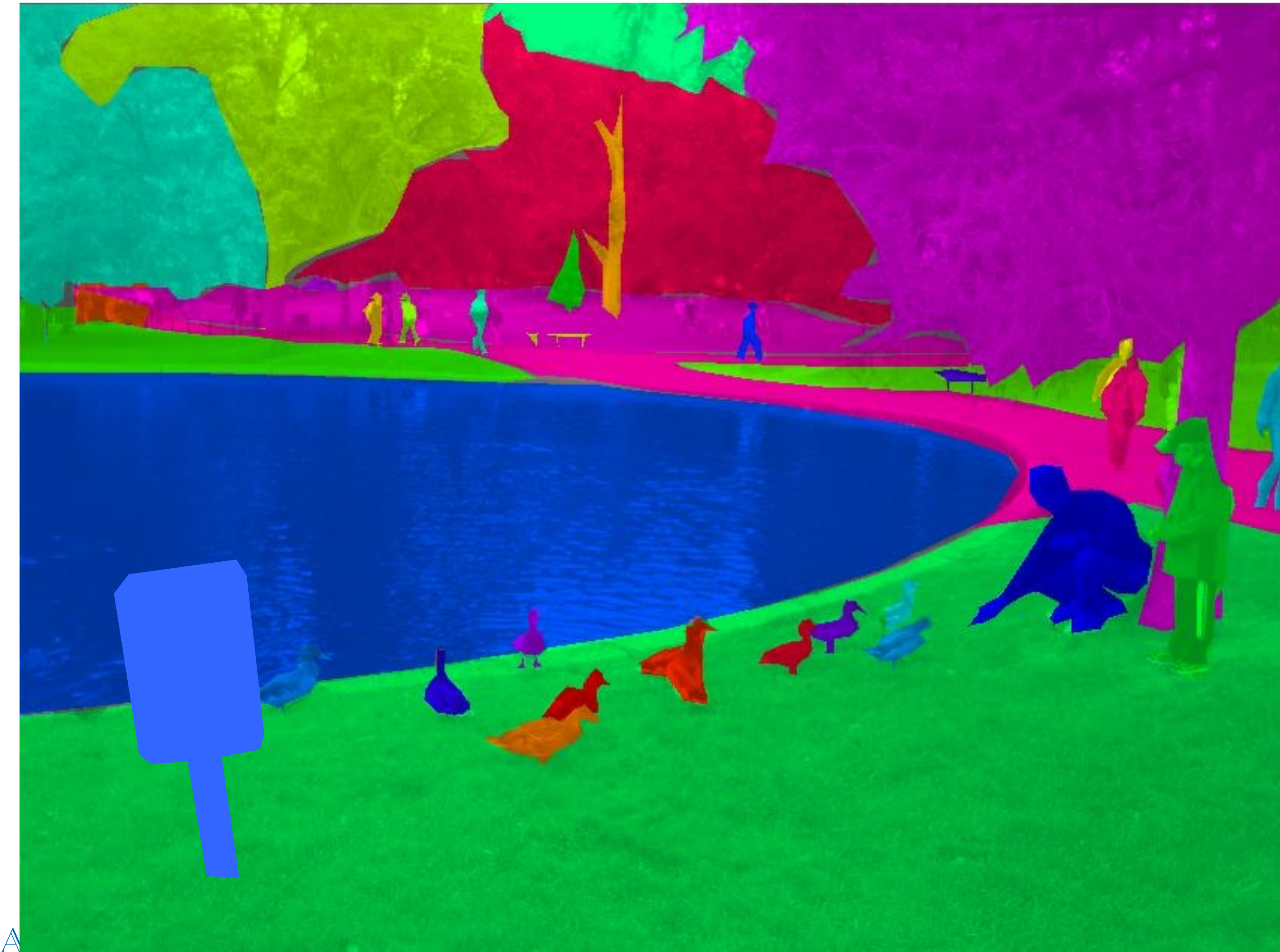
Pretty much garbage
Template matching cannot make it.

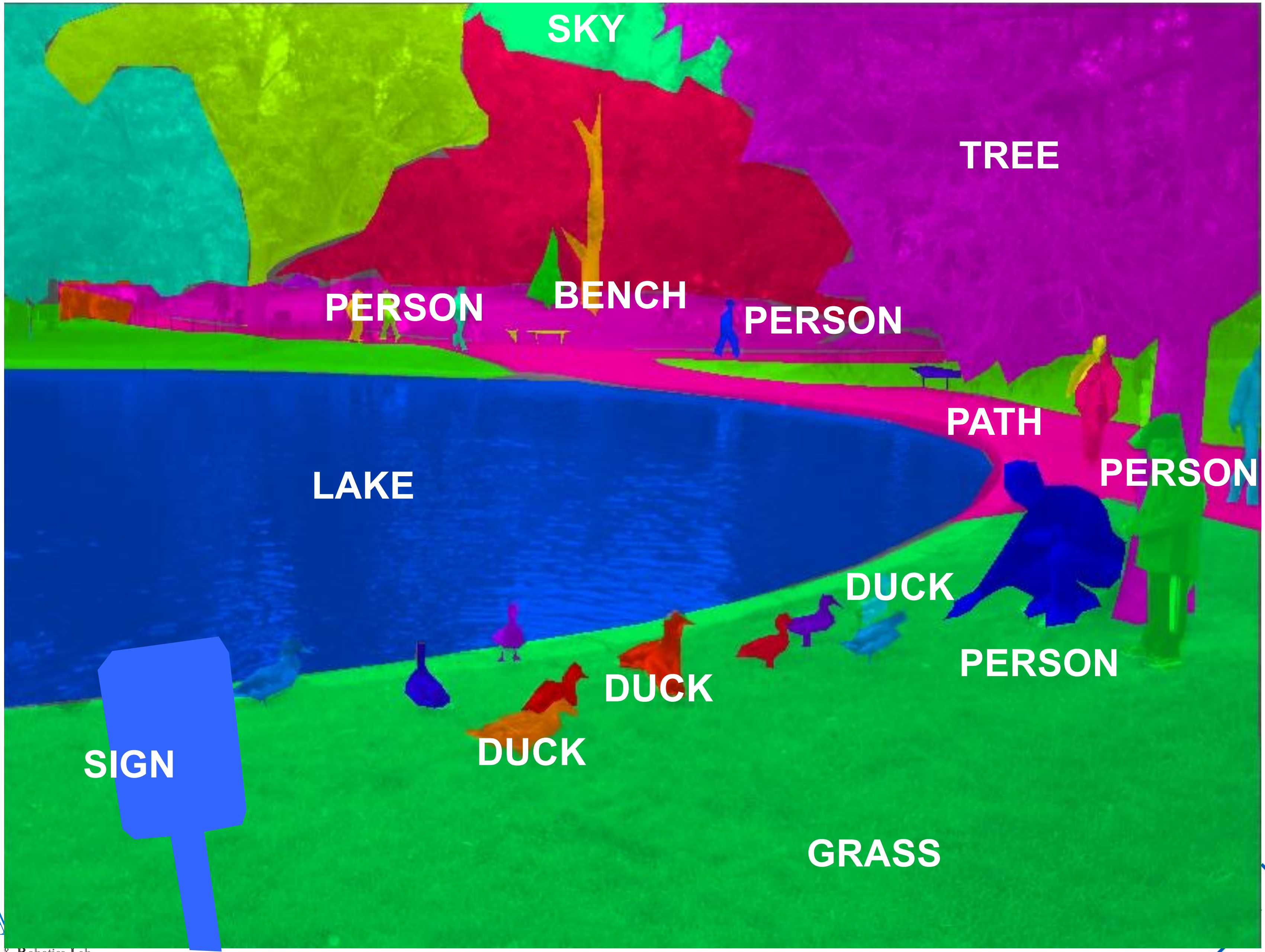
The texture

The object

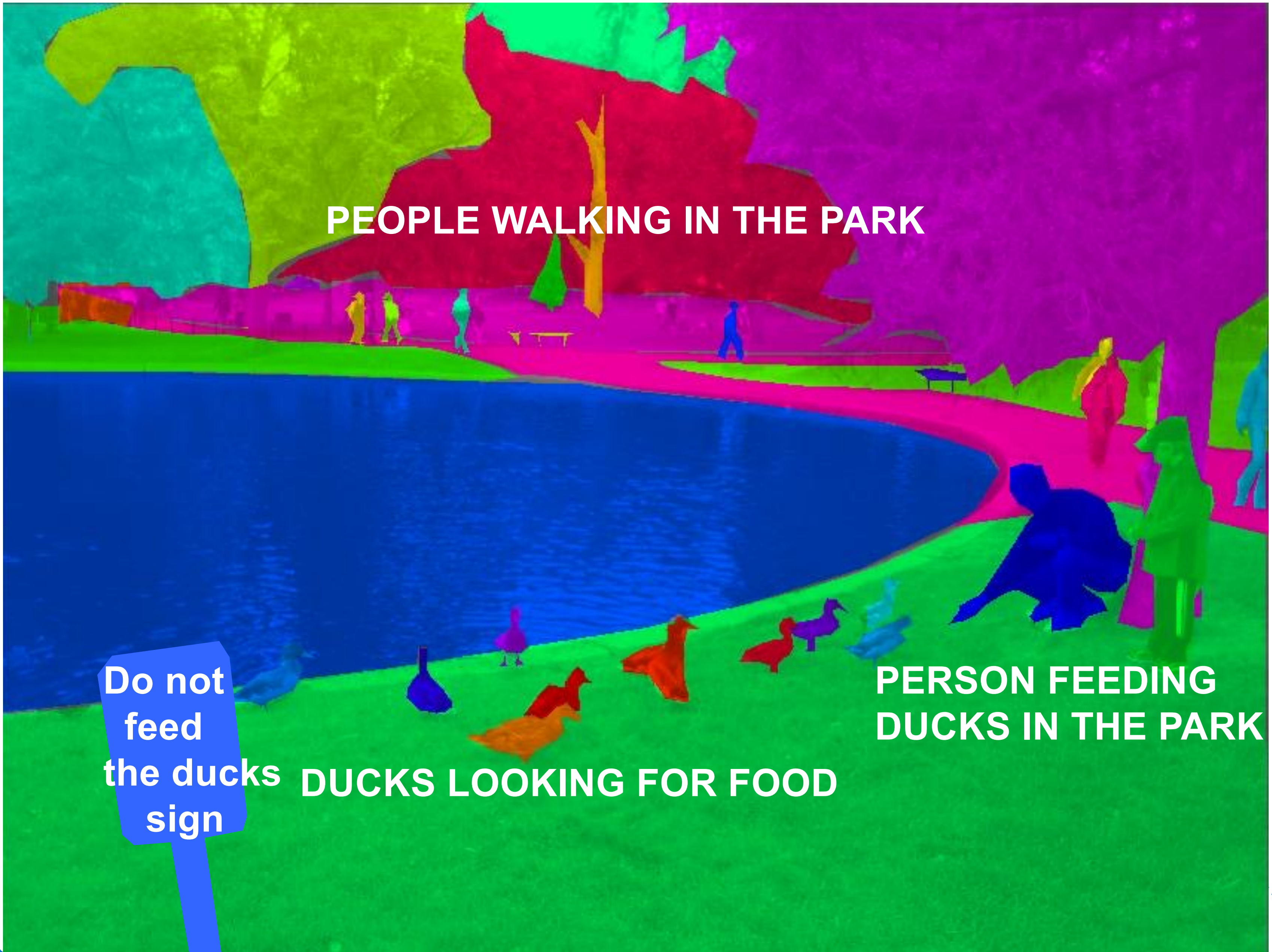
The scene

What do we perceive?




SKY
TREE
PERSON
BENCH
PERSON
PATH
PERSON
LAKE
DUCK
DUCK
PERSON
DUCK
SIGN
GRASS


A VIEW OF A PARK ON A NICE SPRING DAY


PEOPLE WALKING IN THE PARK
Do not feed the ducks sign
DUCKS LOOKING FOR FOOD
PERSON FEEDING DUCKS IN THE PARK


PEOPLE UNDER THE
SHADOW OF THE TREES
PLEASE
DO NOT
FEED
THE
DUCKS
DUCKS ON TOP
OF THE GRASS

Scene views vs. objects

- Scene:
  - a place in which a human can act within
  - or a place to which a human being could navigate.

- A lot more than just a combination of objects
  - just as objects are more than the combinations of their parts.
- Associated with specific functions and behaviors (like objects)
  - E.g., eating in a restaurant, reading in a library,
  - Talking in classroom, playing in a park, etc

Scene views vs. objects

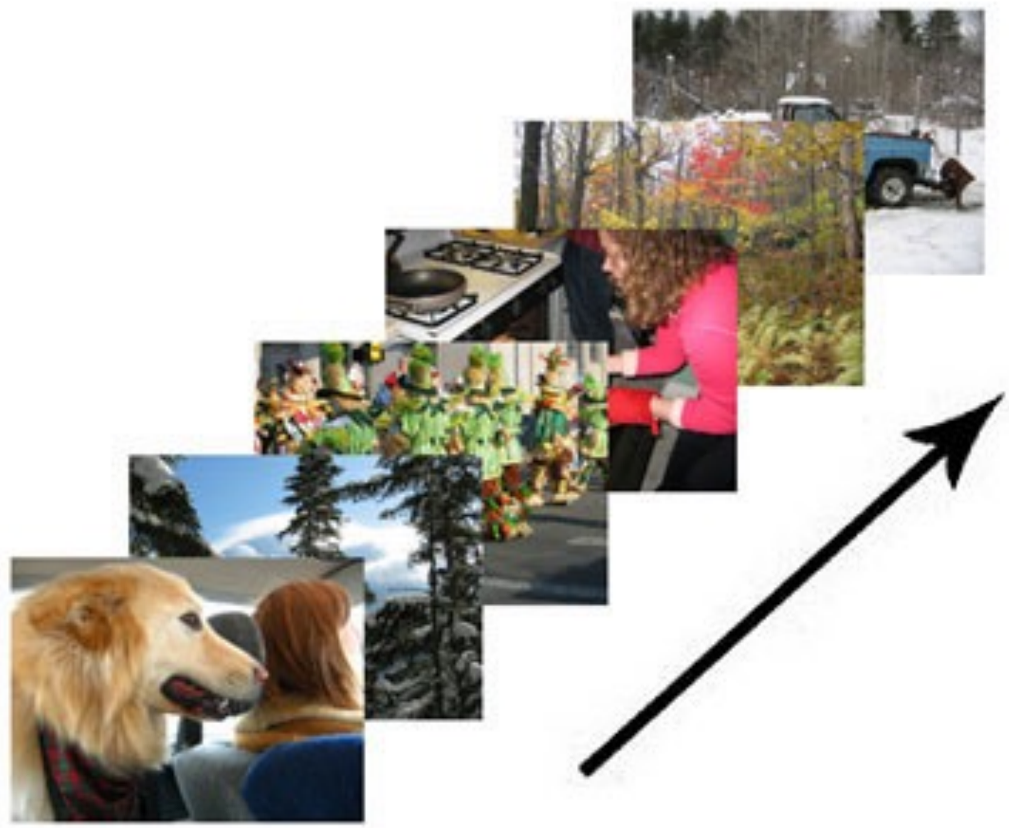
A photograph of a firehydrant

A photograph of a street

Mary Potter (1976)

Mary Potter (1975, 1976) demonstrated

- during a rapid sequential visual presentation (100 msec per image), a novel picture is instantly **understood** and observers seem to comprehend a lot of visual information



Demo : Rapid image understanding

Instructions: 9 photographs will be shown for half a second each. Your task is to **memorize these pictures.**

By Aude Oliva

GET READY

Memory Test

Have you seen this picture ?

Memory Test

Memory Test

Have you seen this picture ?

Memory Test

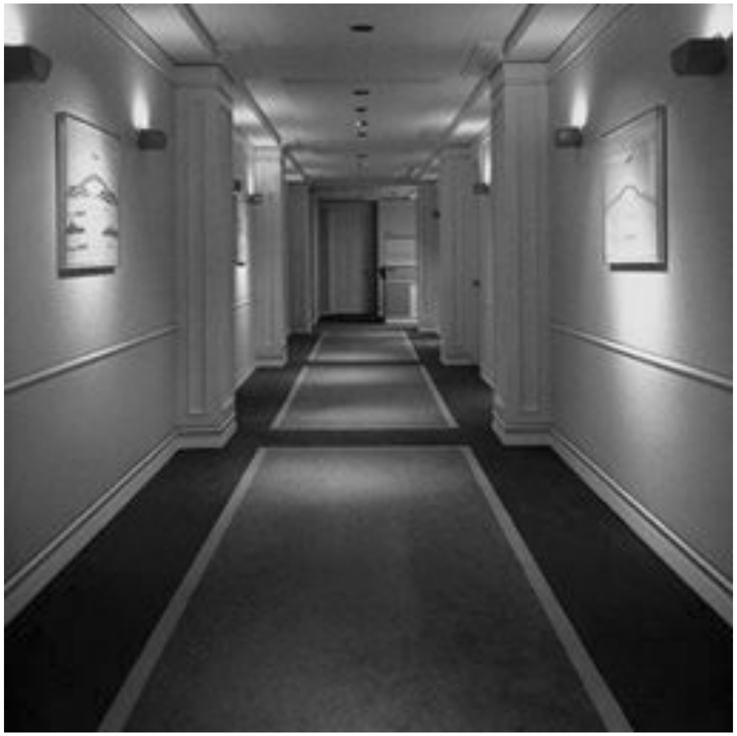
Memory Test

Have you seen this picture ?

S A I R
Spatial AI & Robotics Lab

Memory Test

Memory Test



Have you seen this picture ?

Memory Test

Memory Test

Have you seen this picture ?

Memory Test

Memory Test

Have you seen this picture ?

Memory Test

Memory Test Overview

You have seen these pictures

You were tested with these pictures

The gist of the scene

In a glance, we remember the meaning of an image and its global **layout** but some objects and **details** are forgotten

SAIR
Spatial AI & Robotics Lab

Object Categorization

- How to recognize ANY car

- How to recognize ANY cow

Challenges: robustness

Illumination

Object pose

Clutter

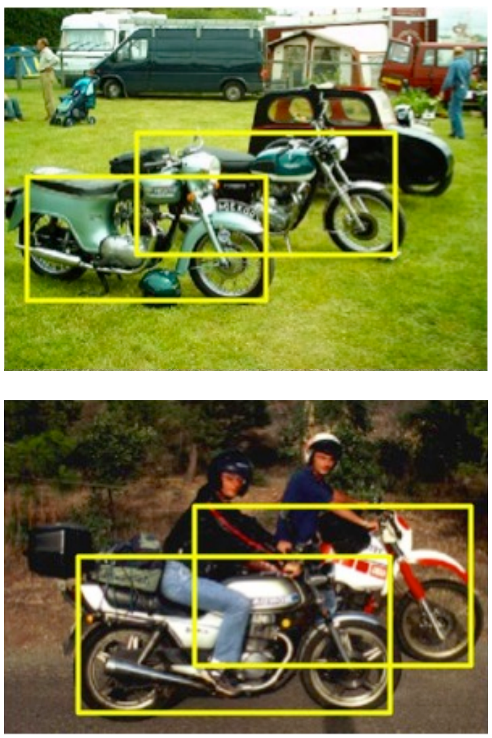
Occlusions

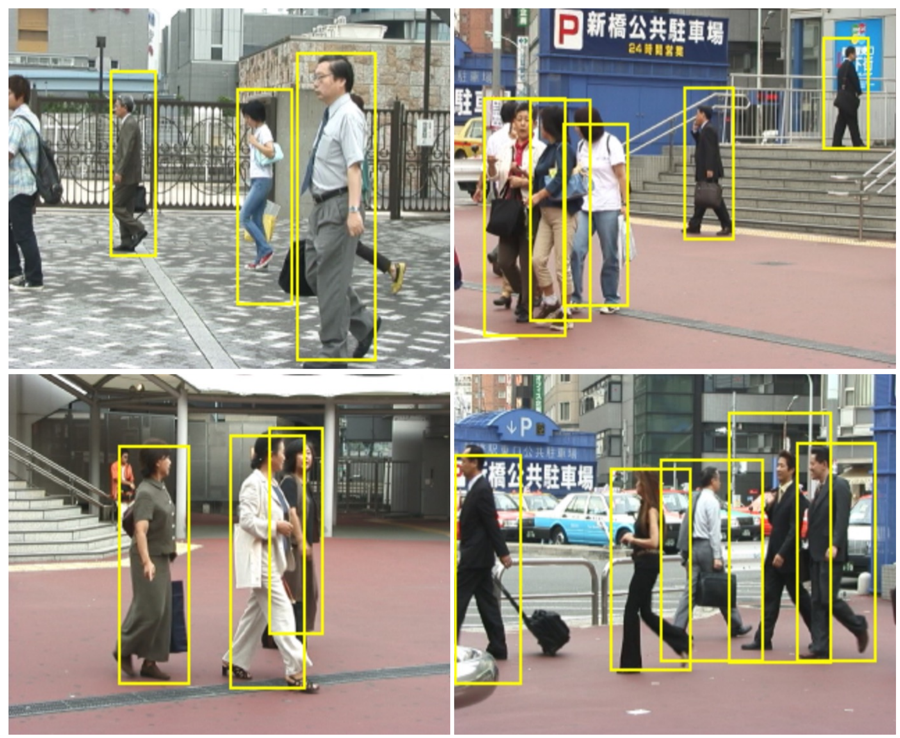
Intra-class appearance

Viewpoint

Challenges: robustness

- Detection in Crowded Scenes
  - Learn object variability
    - Changes in appearance, scale, and articulation
  - Compensate for clutter, overlap, and occlusion



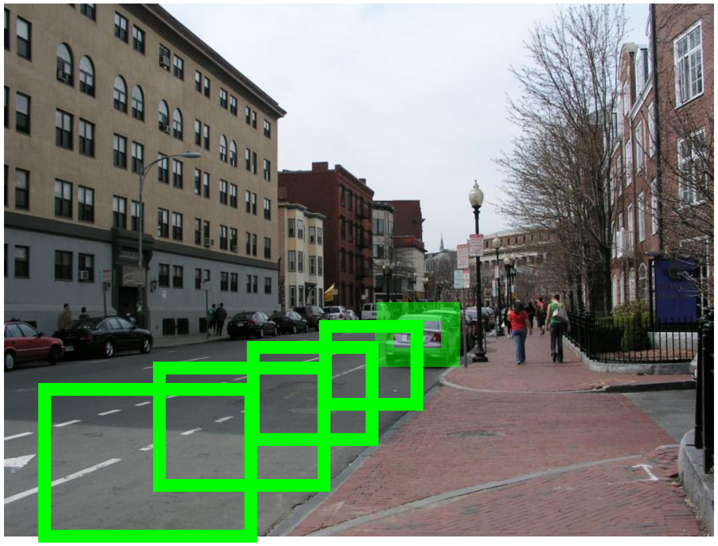


K. Grauman, B. Leibe

Challenges: context and human experience

Challenges: context and human experience

Challenges: context and human experience



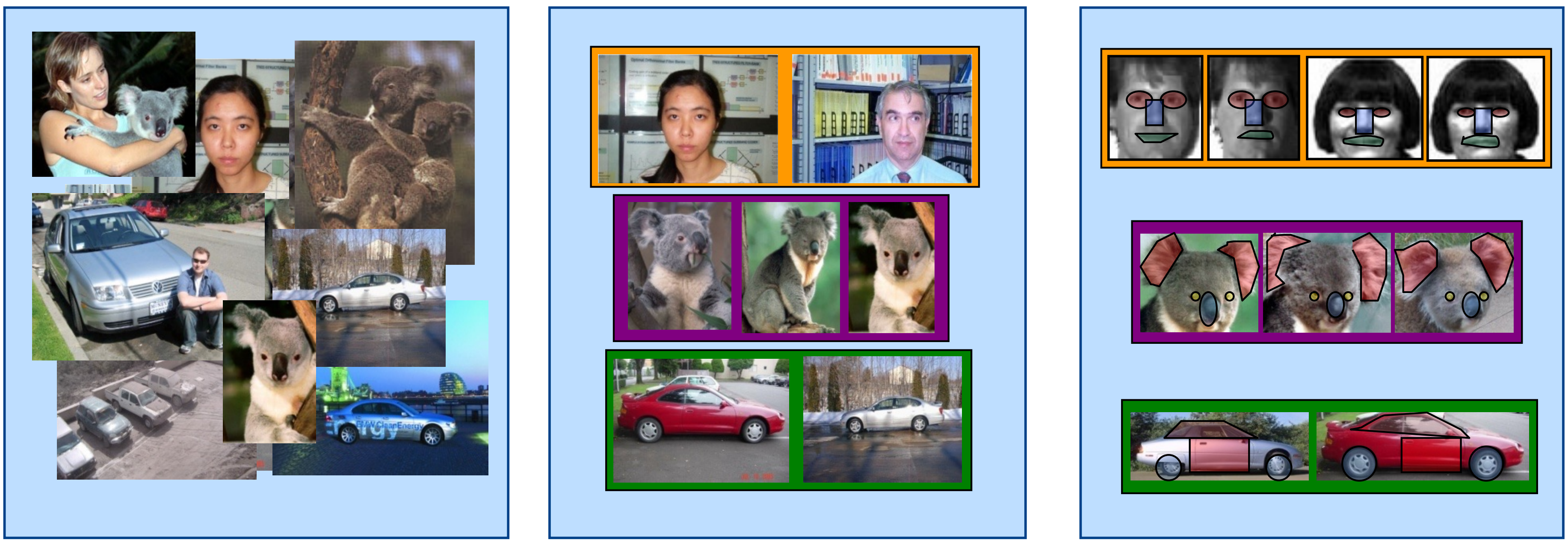
Context cues

Image credit: D. Hoeim

Challenges: learning with minimal supervision

Less

More



Unlabeled,
multiple objects

Classes labeled,
some clutter

Cropped to object,
parts and classes
labeled

Game: Find the pottopod

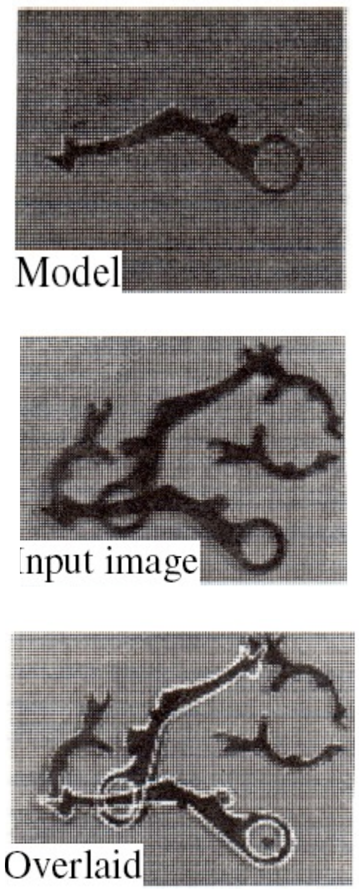
This is a pottopod

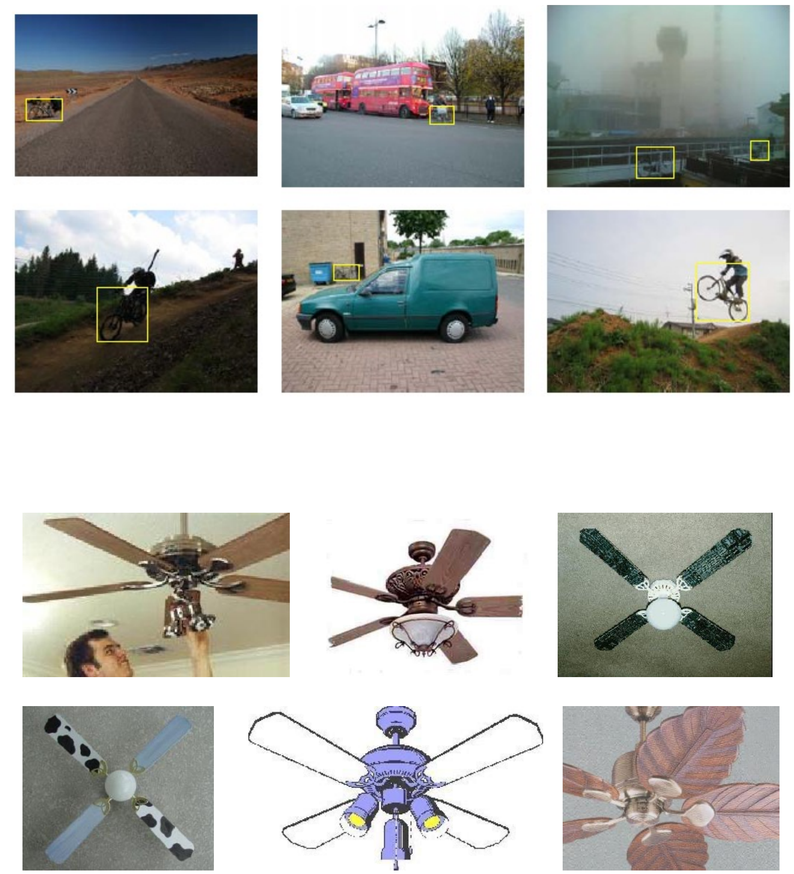
Slide from Pietro Perona, 2004 Object Recognition workshop

Find the pottopod

Slide from Pietro Perona, 2004 Object Recognition workshop

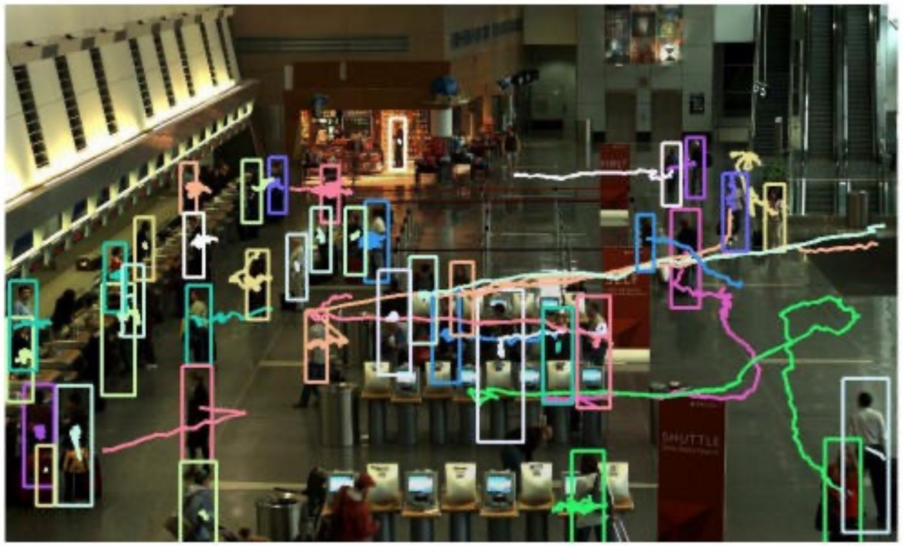
Rough evolution of focus in recognition research

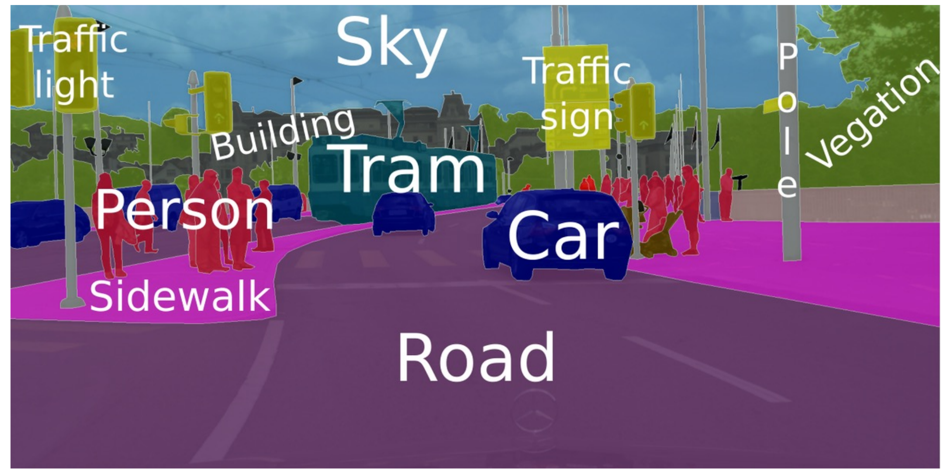


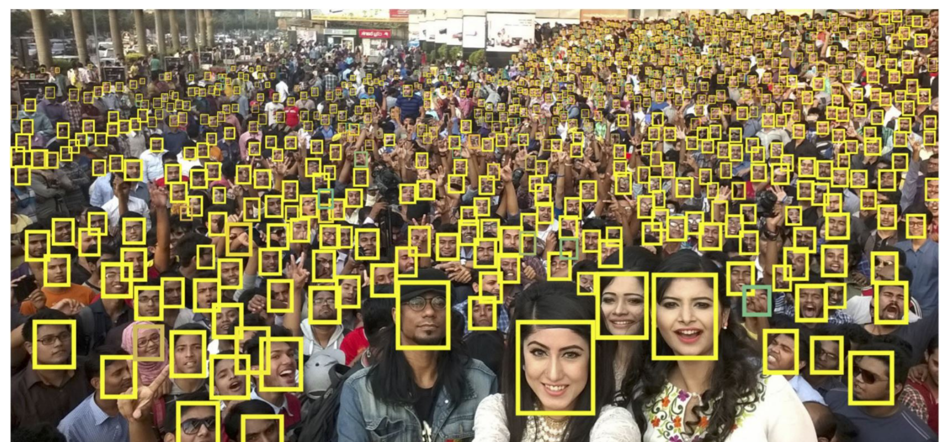


1980s **1990s to early 2000s** **2000-2010…**

Today







In 2014, image recognition by computer outperforms human.

Inputs/outputs/assumptions

What is the **goal**?

- Say yes/no as to whether an object present in image

And/or:

- Categorize all objects
- Forced choice from pool of categories
- Bounding box on object
- Full segmentation
- Build a model of an object category
- Determine pose of an object, e.g., for robot to grasp