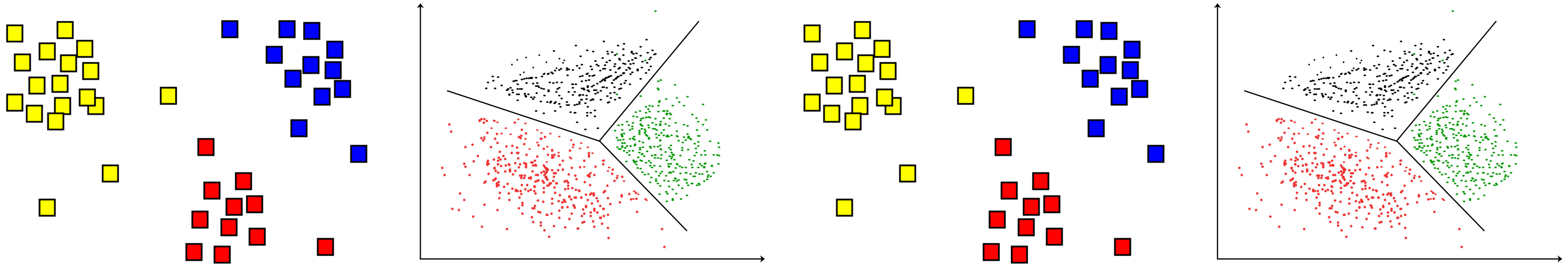
CPSC 425: Computer Vision



Lecture 18: Image Classification

Image Classification

We next discuss **image classification**, where we pass a whole image into a classifier and obtain a class label as output.

What Makes Image Classification **Hard**?




Intra-class variation, viewpoint, illumination, clutter, and occlusion (among others!)

Figure source: Jianxiong Xiao. Original credit: ?

Image Classification

In addition to images containing single **objects**, the same techniques can be applied to classify natural **scenes** (e.g. beach, forest, harbour, library).

Why might classifying scenes be useful?

Image Classification

In addition to images containing single **objects**, the same techniques can be applied to classify natural **scenes** (e.g. beach, forest, harbour, library).

Why might classifying scenes be useful?

Visual perception is influenced by expectation. Our expectations are often conditioned on the **context**.

What is This **Object**?

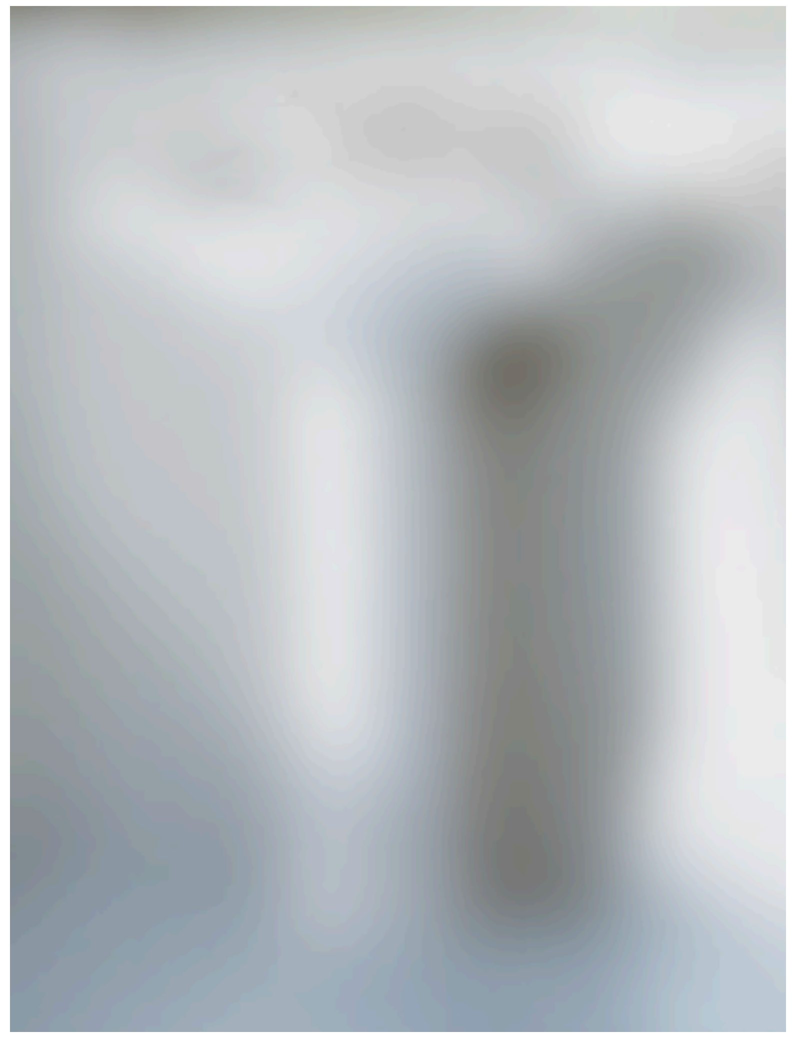


Figure source: Jianxiong Xiao

What is This **Object**?

Figure source: Jianxiong Xiao

What is This **Object**?

Figure source: Jianxiong Xiao

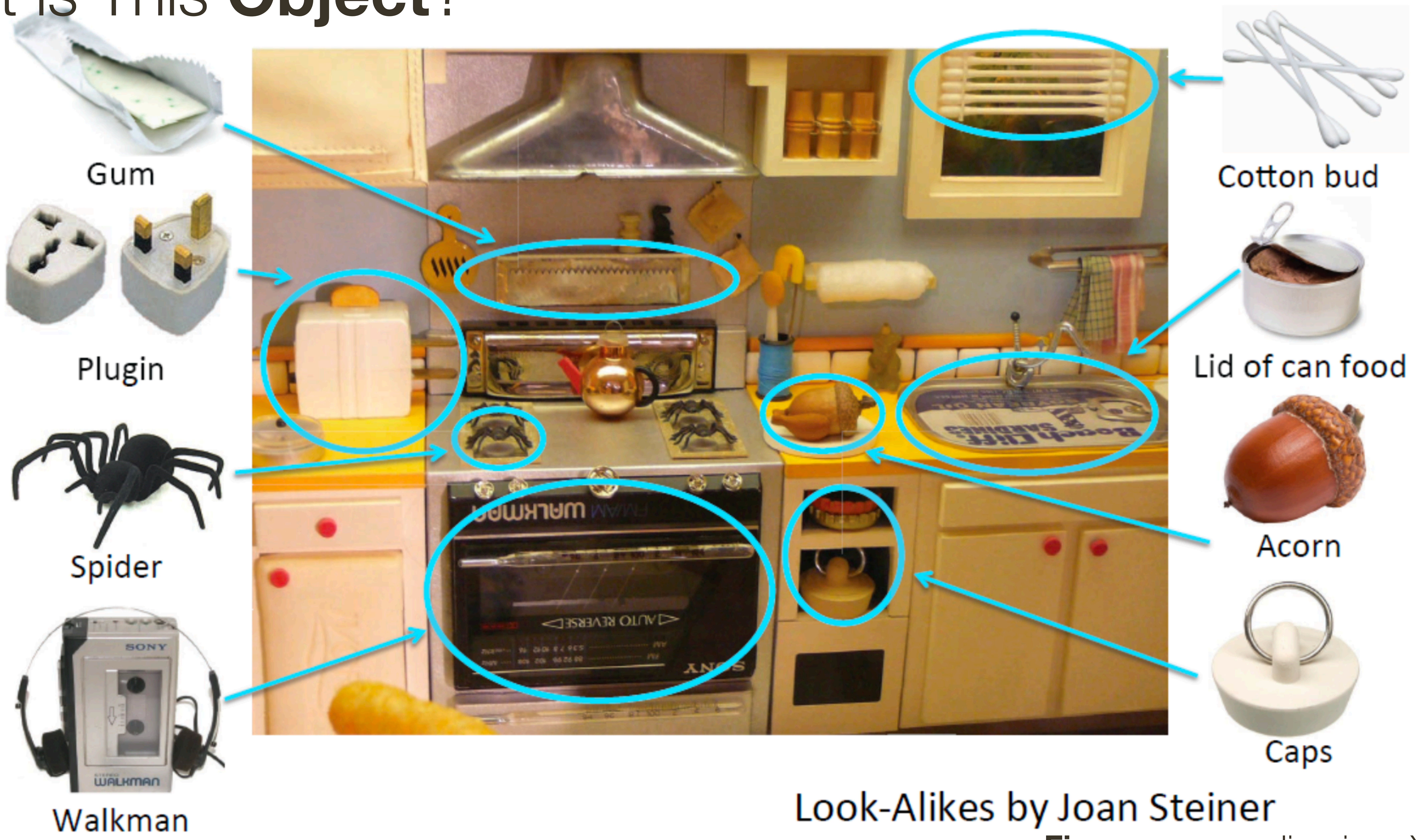
What is This **Object**?

Figure source: Jianxiong Xiao

What is This **Object**?

Figure source: Jianxiong Xiao

What is This **Object**?



Look-Alikes by Joan Steiner

Figure source: Jianxiong Xiao

Visual **Words**

Many algorithms for image classification accumulate evidence on the basis of **visual words**.

To classify a text document (e.g. as an article on sports, entertainment, business, politics) we might find patterns in the occurrences of certain words.

Vector Space Model

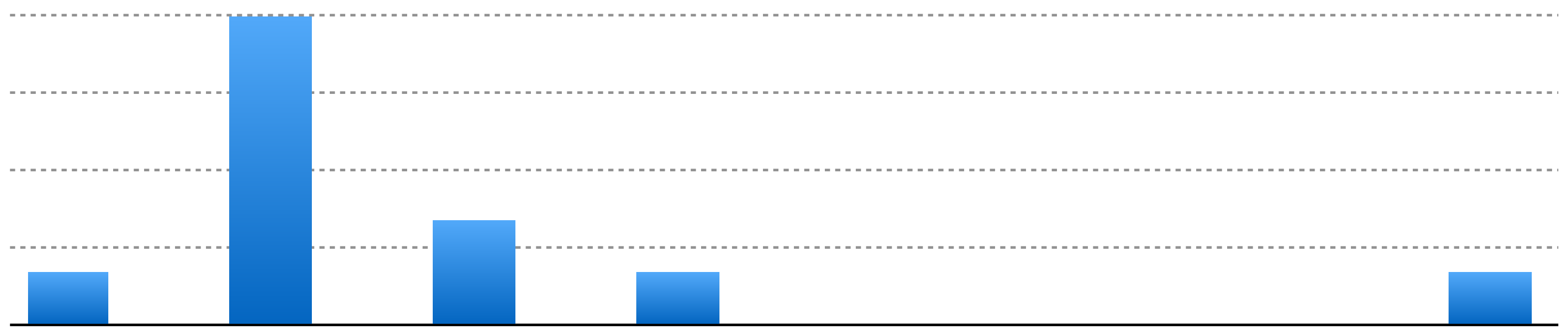
G. Salton. ‘Mathematics and Information Retrieval’ Journal of Documentation,1979


The Newspa

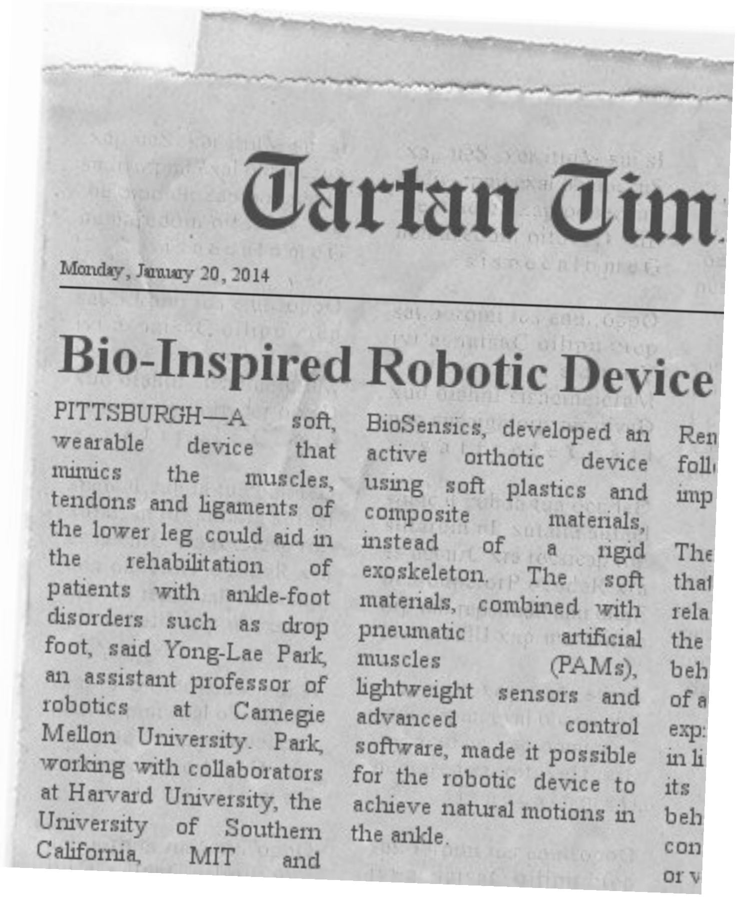
Sunday, December 22, 2013

DARPA Selects Carnegie Me

The Tartan Rescue Team from Carnegie Mellon University's National Robotics Engineering Center ranked third among teams competing in the Defense Advanced Research Projects Agency (DARPA) Robotics Challenge Trials this weekend in Homestead, Fla., and was selected by the agency as one of eight teams eligible for DARPA funding to prepare for next December's finals. The team's four-limbed CMU Highly Intelligent Mobile Platform, or CHIMP, robot scored 18 out of a possible 32 points during the two-day trials. It demonstrated its ability to perform such tasks as removing debris, cutting a hole through a wall and closing a series of valves.



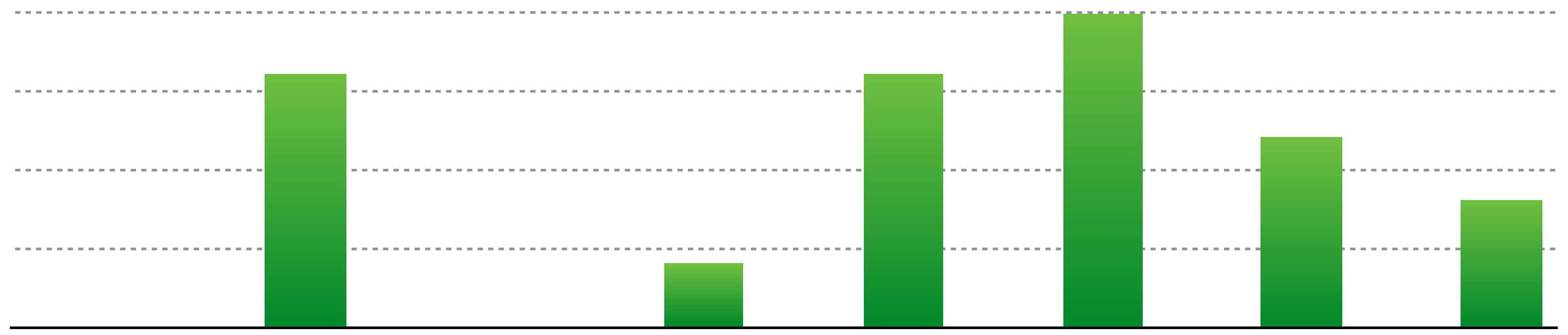
1	6	2	1	0	0	0	1
Tartan	robot	CHIMP	CMU	bio	soft	ankle	sensor


Tartan Tim

Monday, January 20, 2014

Bio-Inspired Robotic Device

PITTSBURGH—A soft, wearable device that mimics the muscles, tendons and ligaments of the lower leg could aid in the rehabilitation of patients with ankle-foot disorders such as drop foot, said Yong-Lae Park, an assistant professor of robotics at Carnegie Mellon University. Park, working with collaborators at Harvard University, the University of Southern California, MIT and BioSensics, developed an active orthotic device using soft plastics and composite materials, instead of a rigid exoskeleton. The soft materials, combined with pneumatic artificial muscles (PAMs), lightweight sensors and advanced control software, made it possible for the robotic device to achieve natural motions in the ankle.



0	4	0	1	4	5	3	2
Tartan	robot	CHIMP	CMU	bio	soft	ankle	sensor

http://www.fodey.com/generators/newspaper/snippet.asp

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Vector Space Model

A document (datapoint) is a vector of counts over each word (feature)

$$\boldsymbol{v}_d = [n(w_{1,d}) \quad n(w_{2,d}) \quad \cdots \quad n(w_{T,d})]$$

$n(\cdot)$ counts the number of occurrences

just a histogram over words

What is the similarity between two documents?

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Vector Space Model

A document (datapoint) is a vector of counts over each word (feature)

$$\boldsymbol{v}_d = [n(w_{1,d}) \quad n(w_{2,d}) \quad \cdots \quad n(w_{T,d})]$$

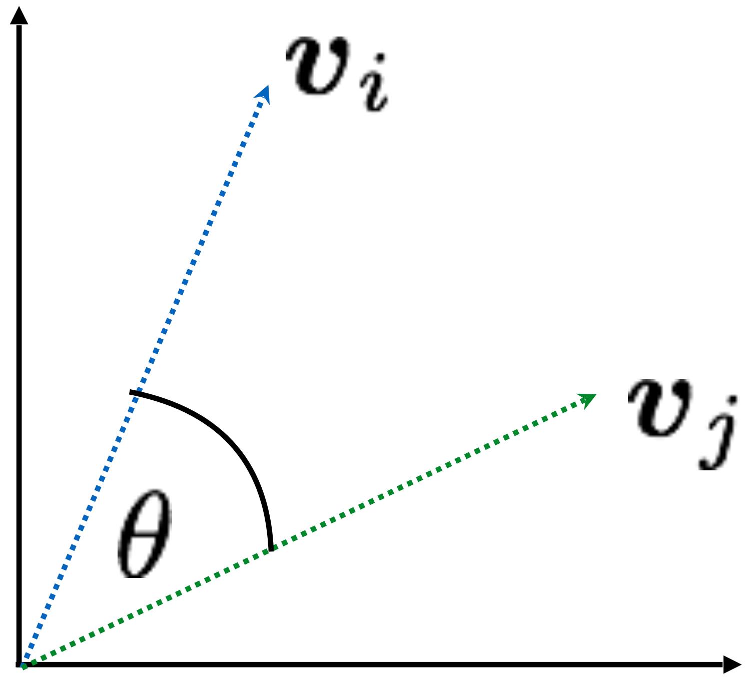
$n(\cdot)$ counts the number of occurrences

just a histogram over words

What is the similarity between two documents?

Use any distance you want but the cosine distance is fast and well designed for high-dimensional vector spaces:

$$\begin{aligned} d(\boldsymbol{v}_i, \boldsymbol{v}_j) &= \cos\theta \\ &= \frac{\boldsymbol{v}_i \cdot \boldsymbol{v}_j}{\|\boldsymbol{v}_i\| \|\boldsymbol{v}_j\|} \end{aligned}$$



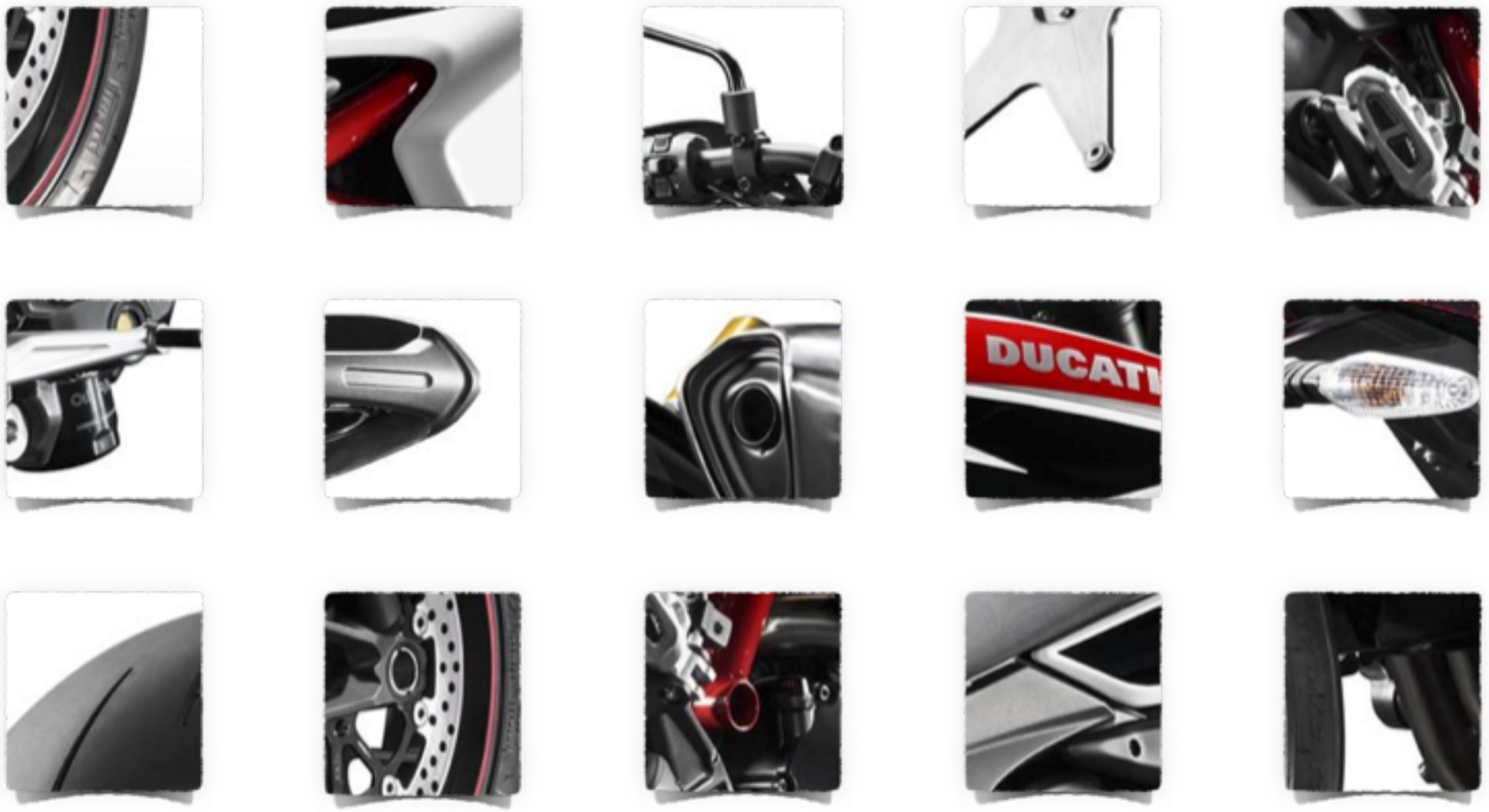
Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Visual **Words**

In images, the equivalent of a word is a local image patch. The local image patch is described using a descriptor such as SIFT.

We construct a **vocabulary** or **codebook** of local descriptors, containing representative local descriptors.

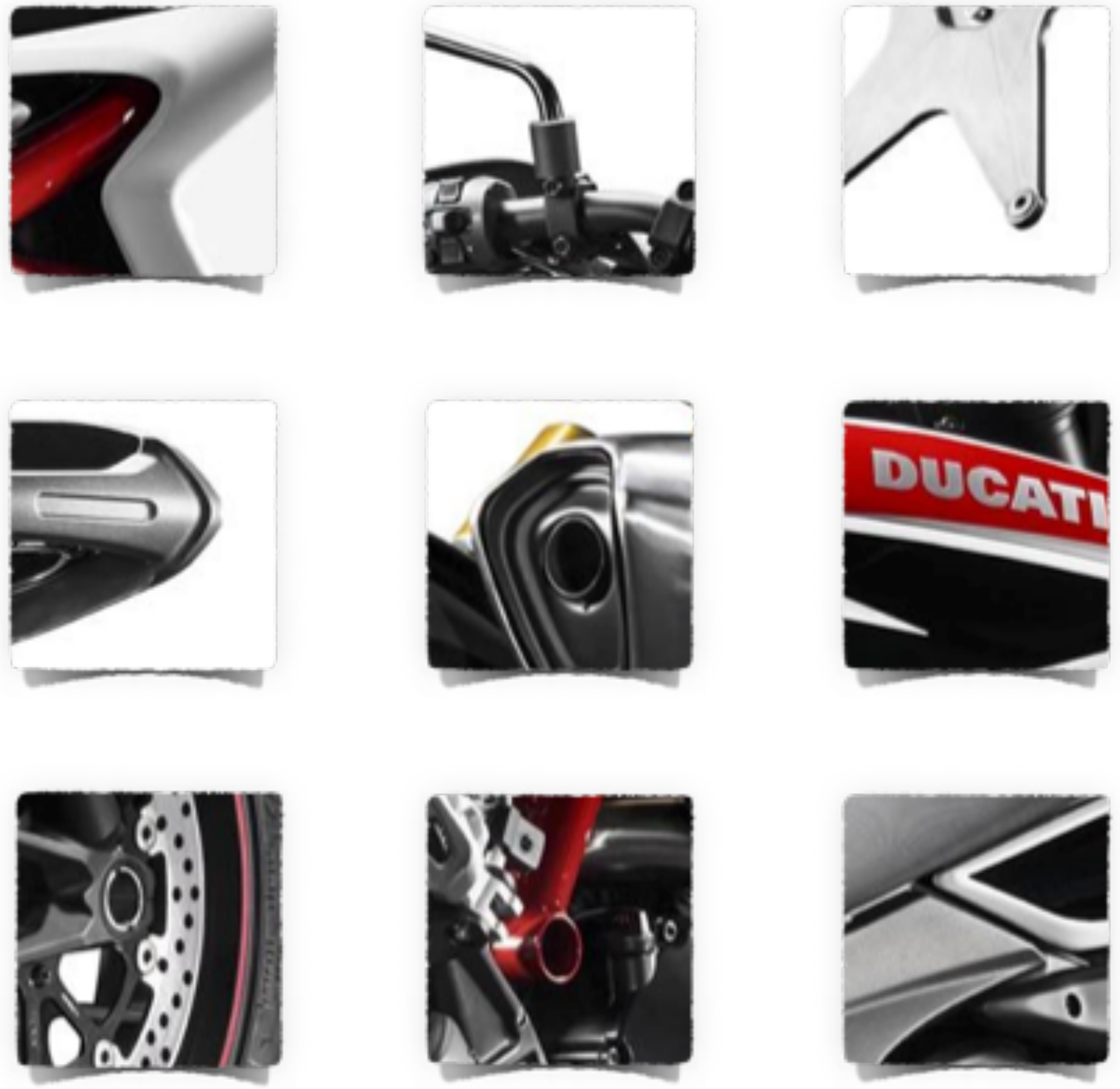
What **Objects** do These Parts Belong To?



Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Some local feature are very informative

An object as



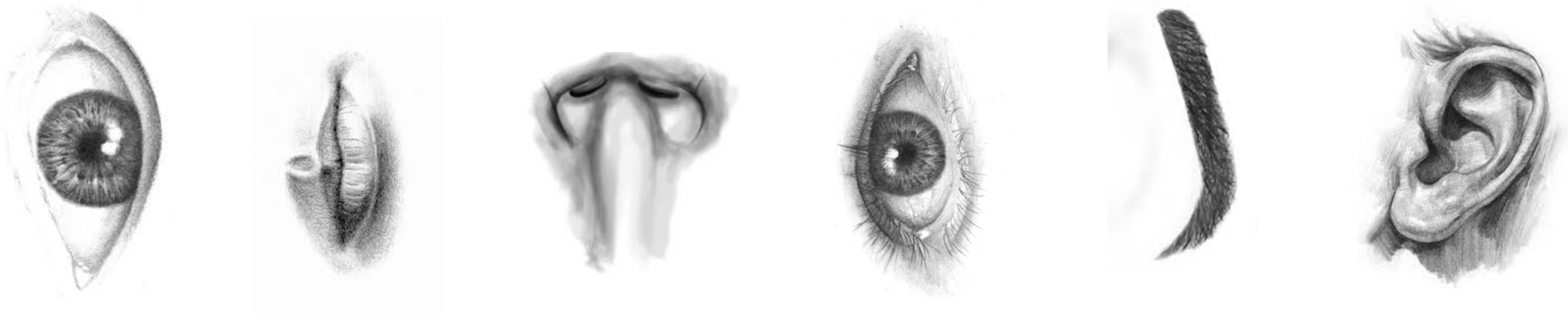
a collection of local features

(bag-of-features)

- deals well with occlusion
- scale invariant
- rotation invariant

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

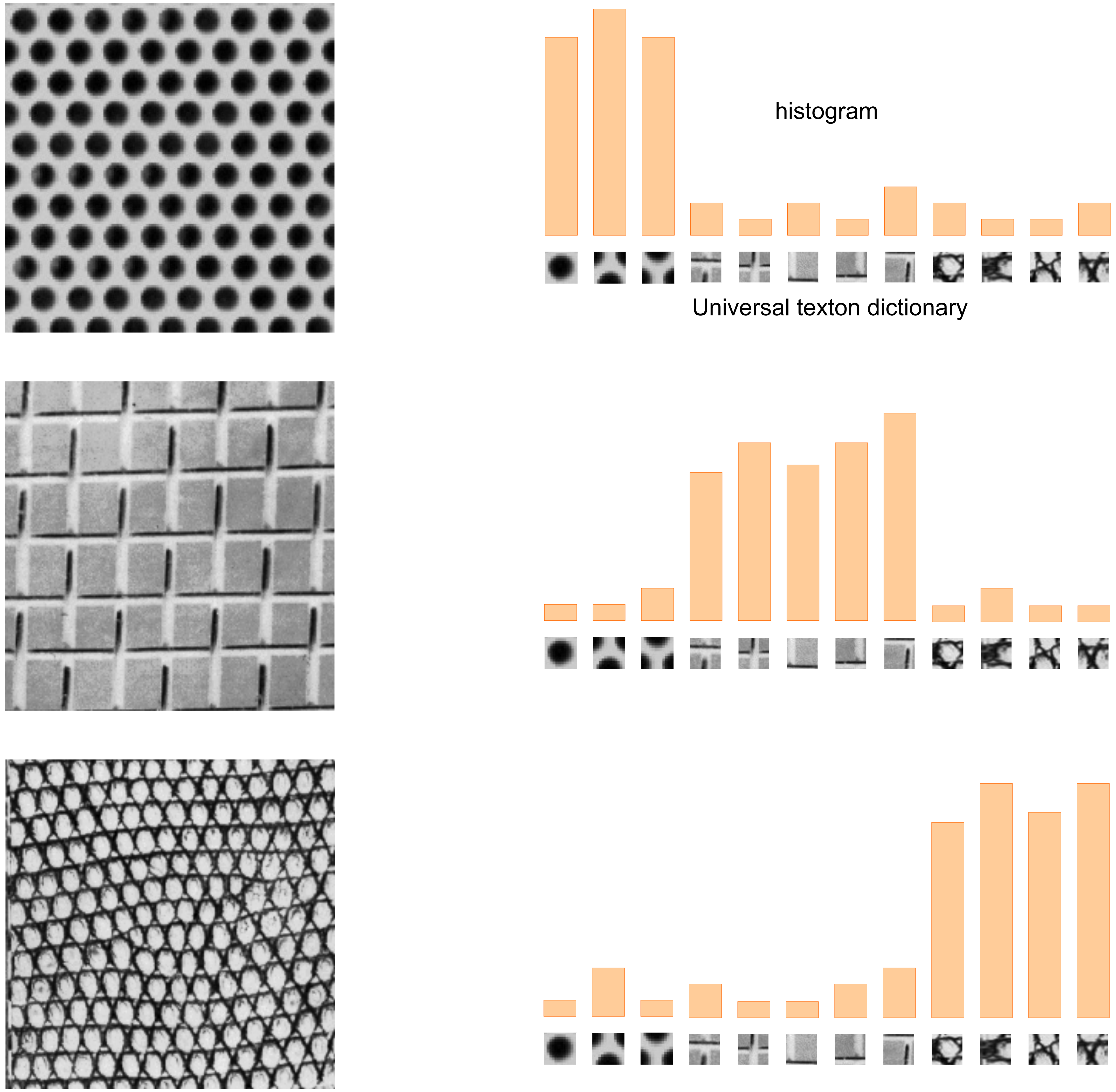
(**not so**) Crazy Assumption



spatial information of local features
can be ignored for object recognition (i.e., verification)

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Recall: Texture Representation



Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Visual **Words**

In images, the equivalent of a word is a local image patch. The local image patch is described using a descriptor such as SIFT.

We construct a **vocabulary** or **codebook** of local descriptors, containing representative local descriptors.

Question: How might we construct such a codebook? Given a large sample of SIFT descriptors, say 1 million, how can we choose a small number of 'representative' SIFT codewords, say 1000?

Standard **Bag-of-Words** Pipeline (for image classification)

Dictionary Learning:
Learn Visual Words using clustering

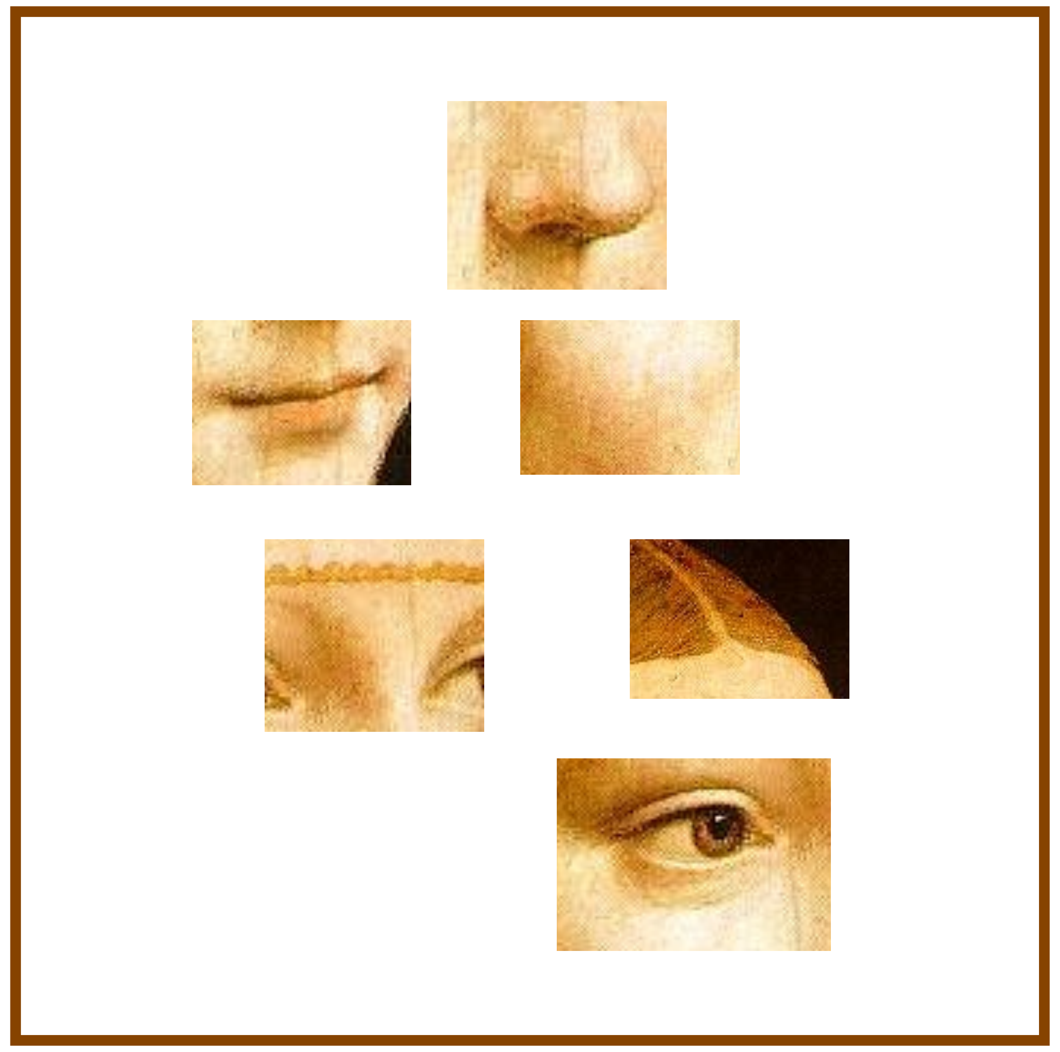
Encode:
build Bags-of-Words (BOW) vectors
for each image

Classify:
Train and test data using BOWs

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

1. **Dictionary Learning**: Learn Visual Words using Clustering

1. **extract features** (e.g., SIFT) from images



Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

1. **Dictionary Learning**: Learn Visual Words using Clustering

2. **Learn visual dictionary** (e.g., K-means clustering)

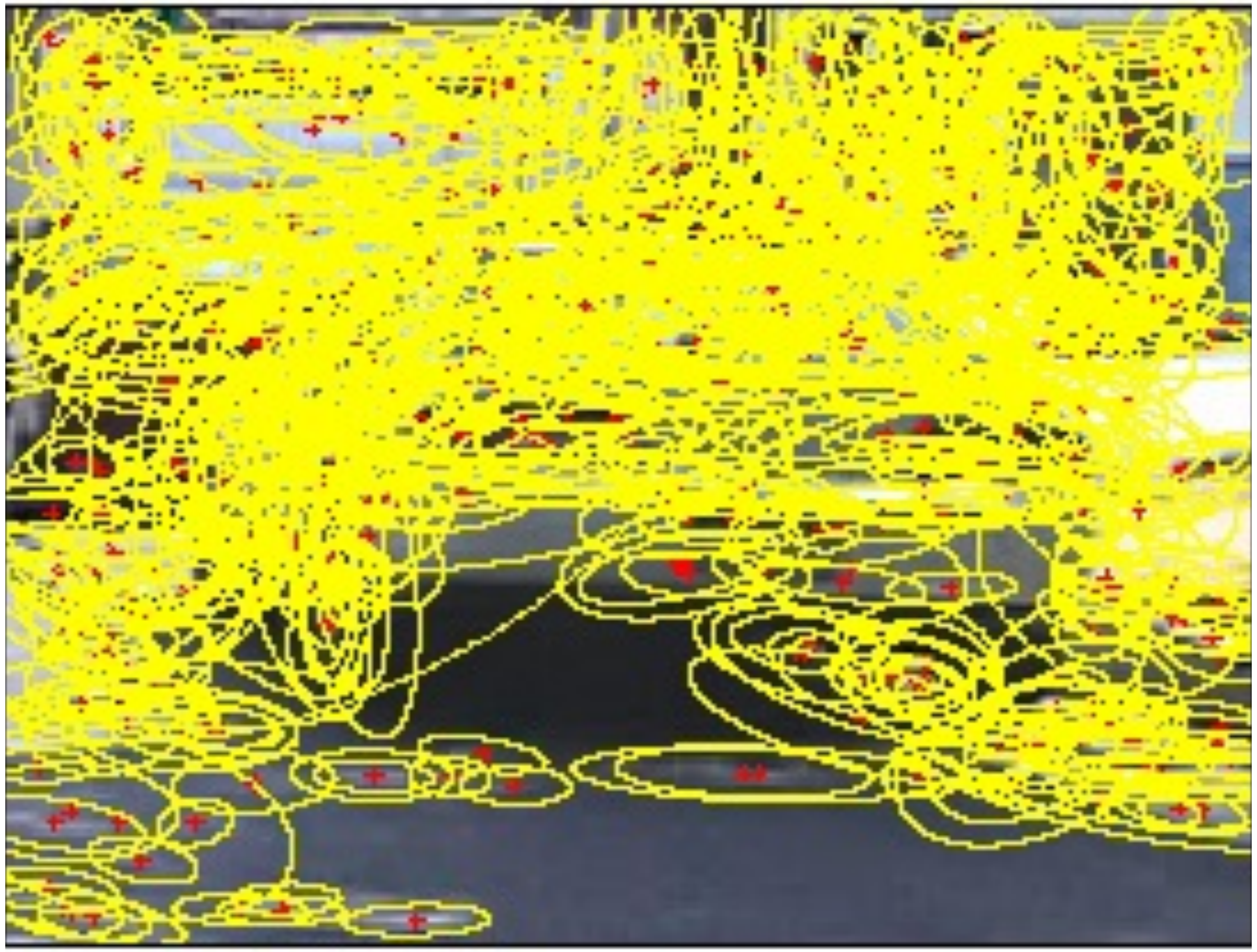
Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

What **Features** Should We Extract?

- Regular grid
  Vogel & Schiele, 2003
  Fei-Fei & Perona, 2005

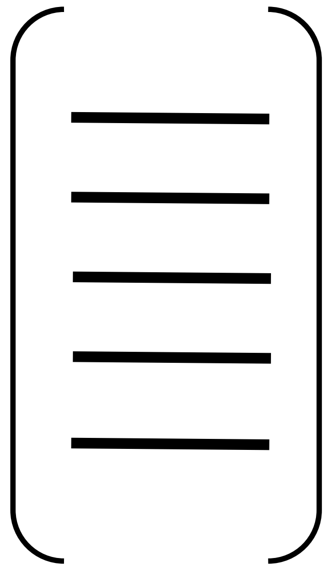
- Interest point detector
  Csurka et al. 2004
  Fei-Fei & Perona, 2005
  Sivic et al. 2005

- Other methods
  Random sampling (Vidal-Naquet & Ullman, 2002)
  Segmentation-based patches (Barnard et al. 2003)



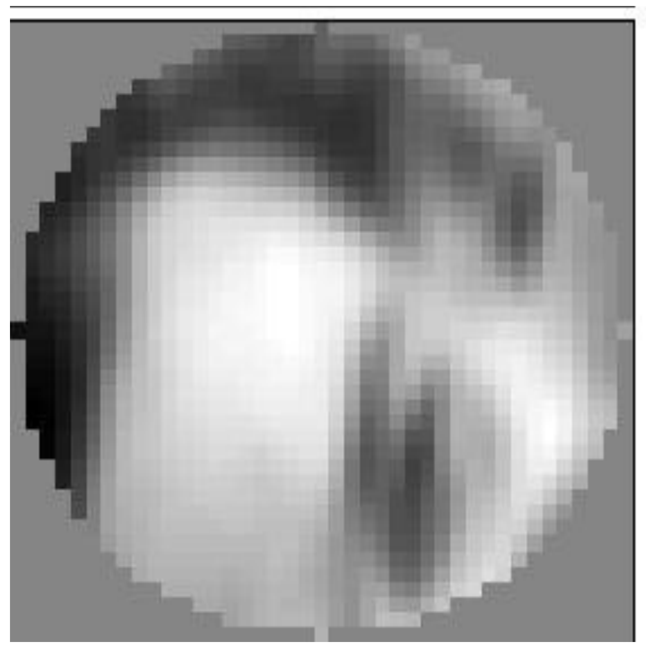
Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Extracting **SIFT** Patches

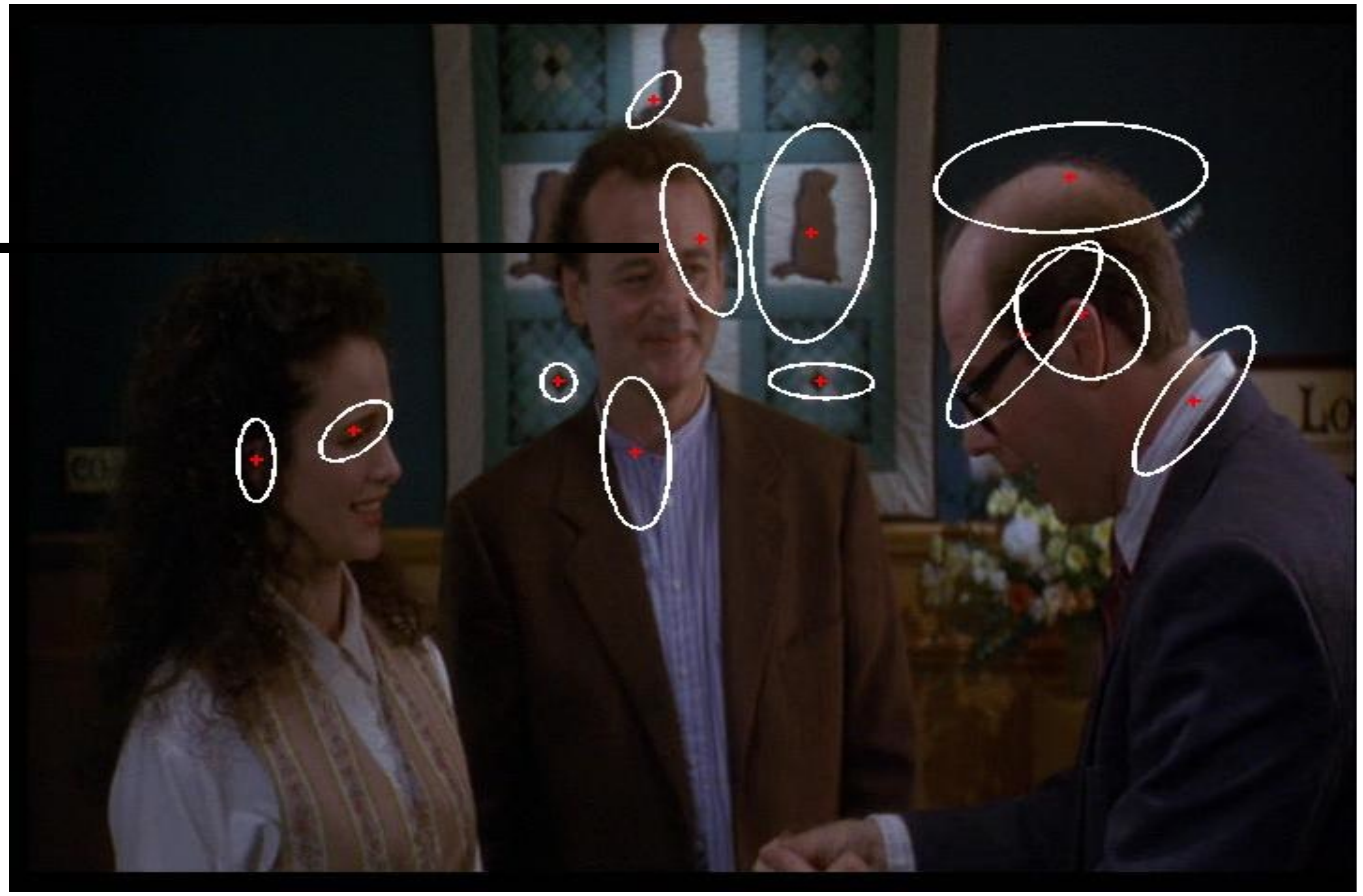


Compute SIFT descriptor

[Lowe'99]



Normalize patch



Detect patches

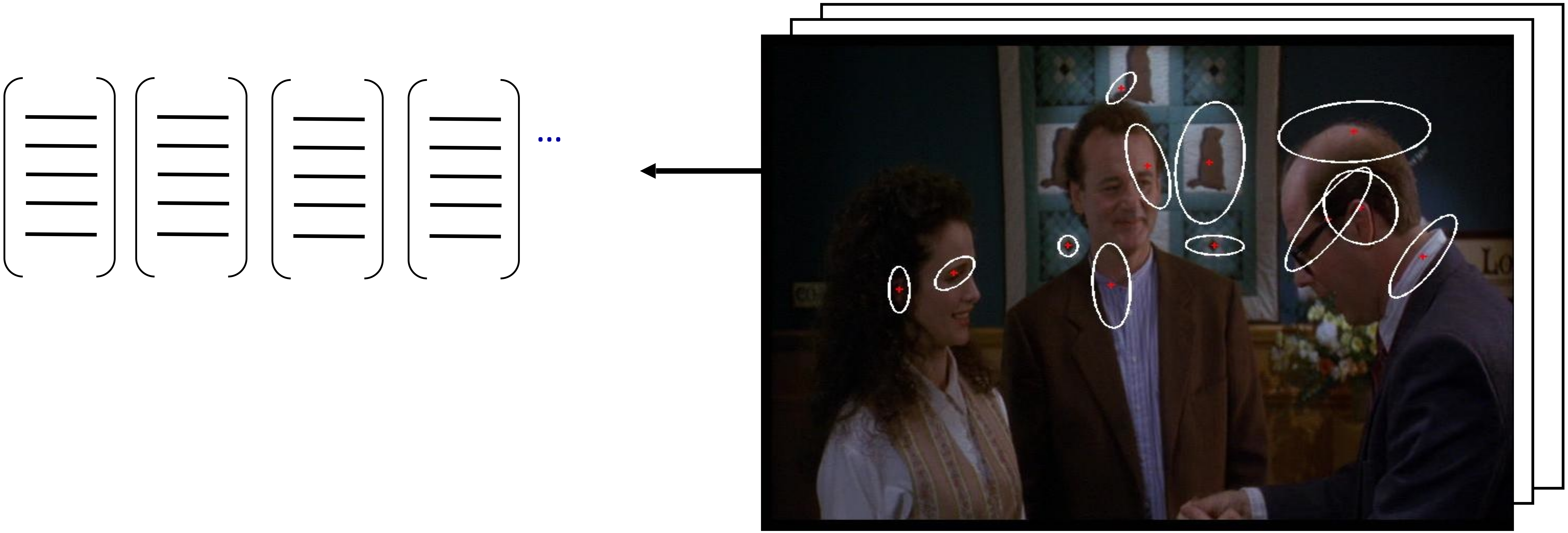
[Mikojaczyk and Schmid '02]

[Mata, Chum, Urban & Pajdla, '02]

[Sivic & Zisserman, '03]

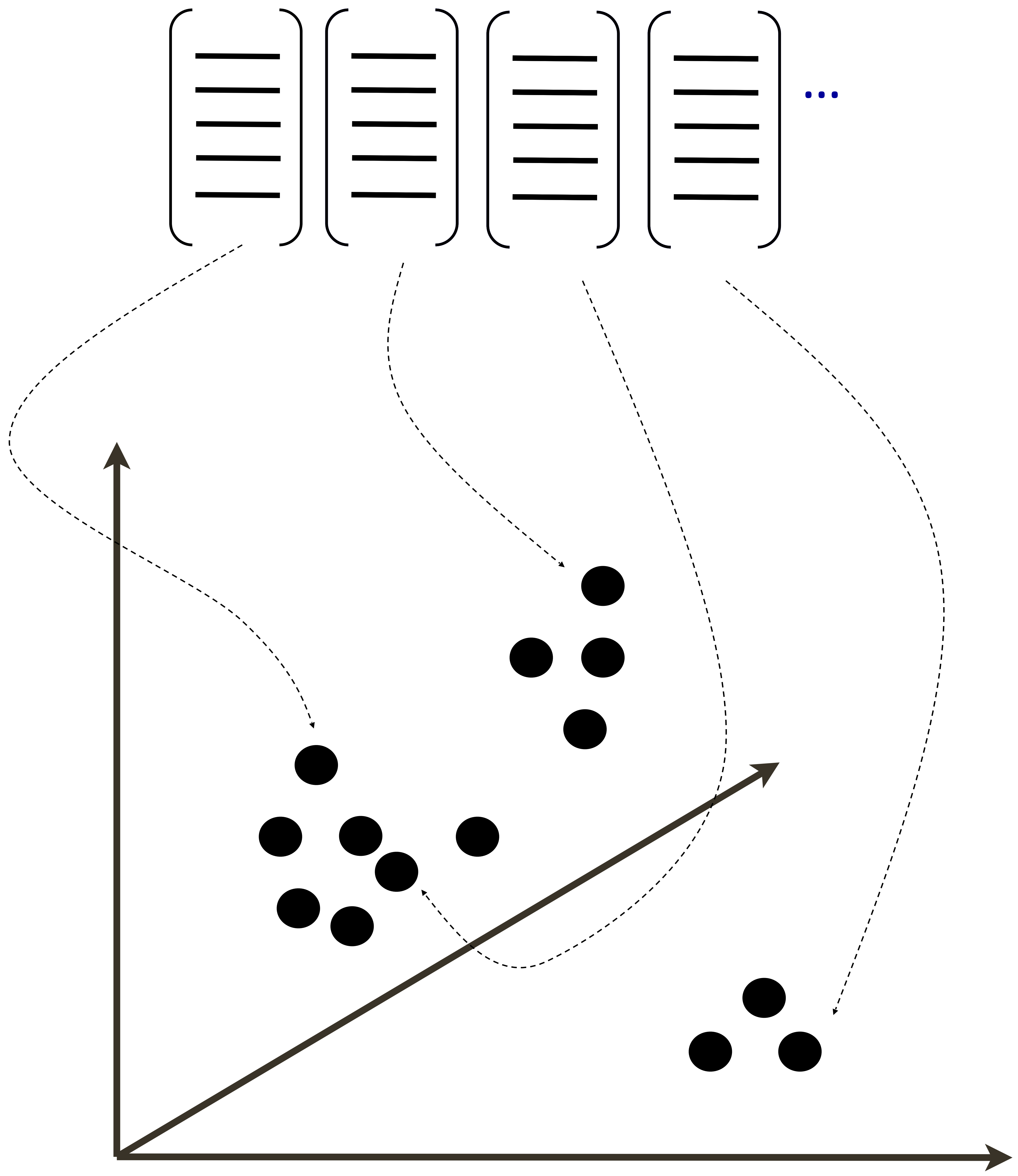
Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Extracting **SIFT** Patches



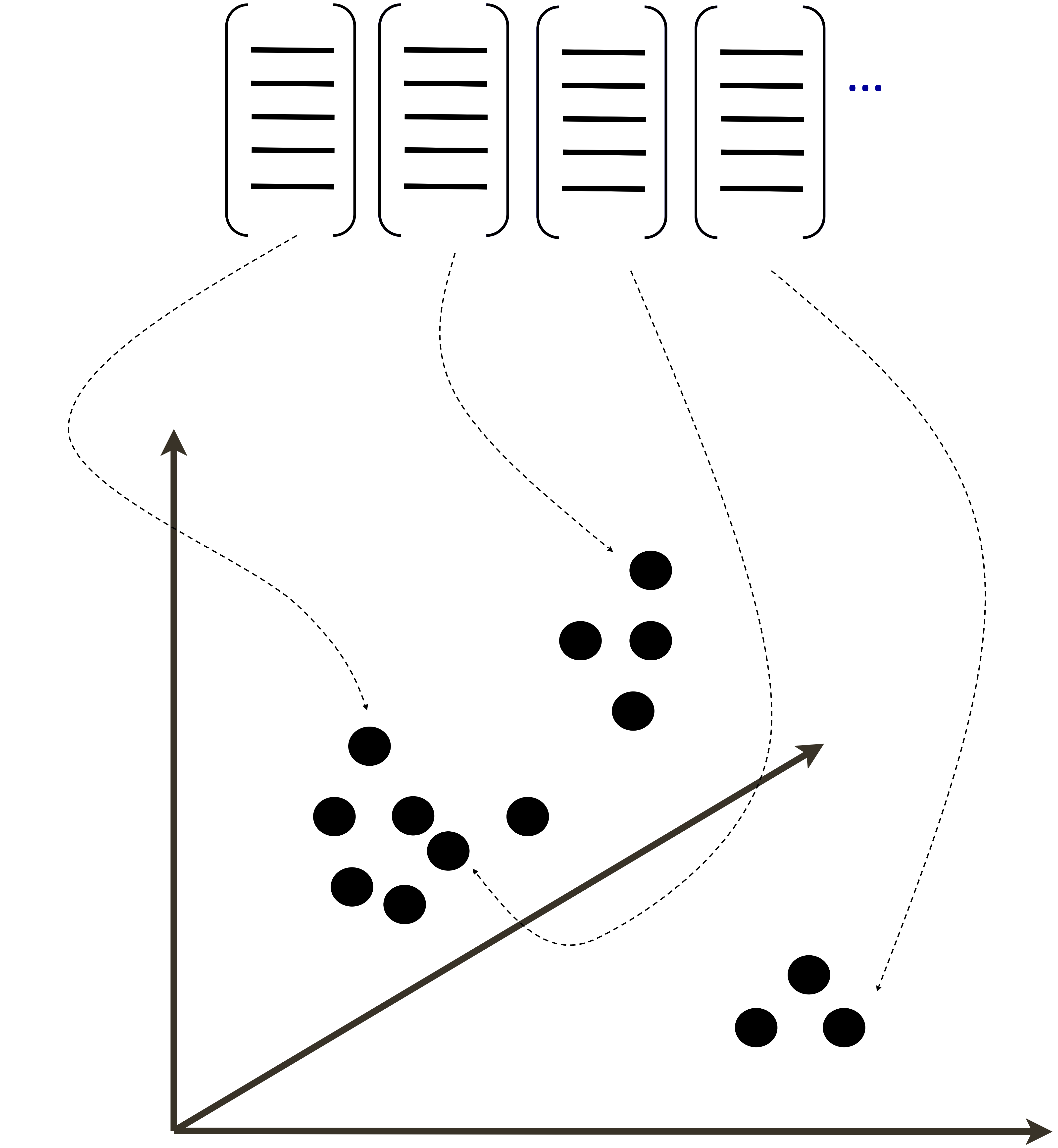
Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

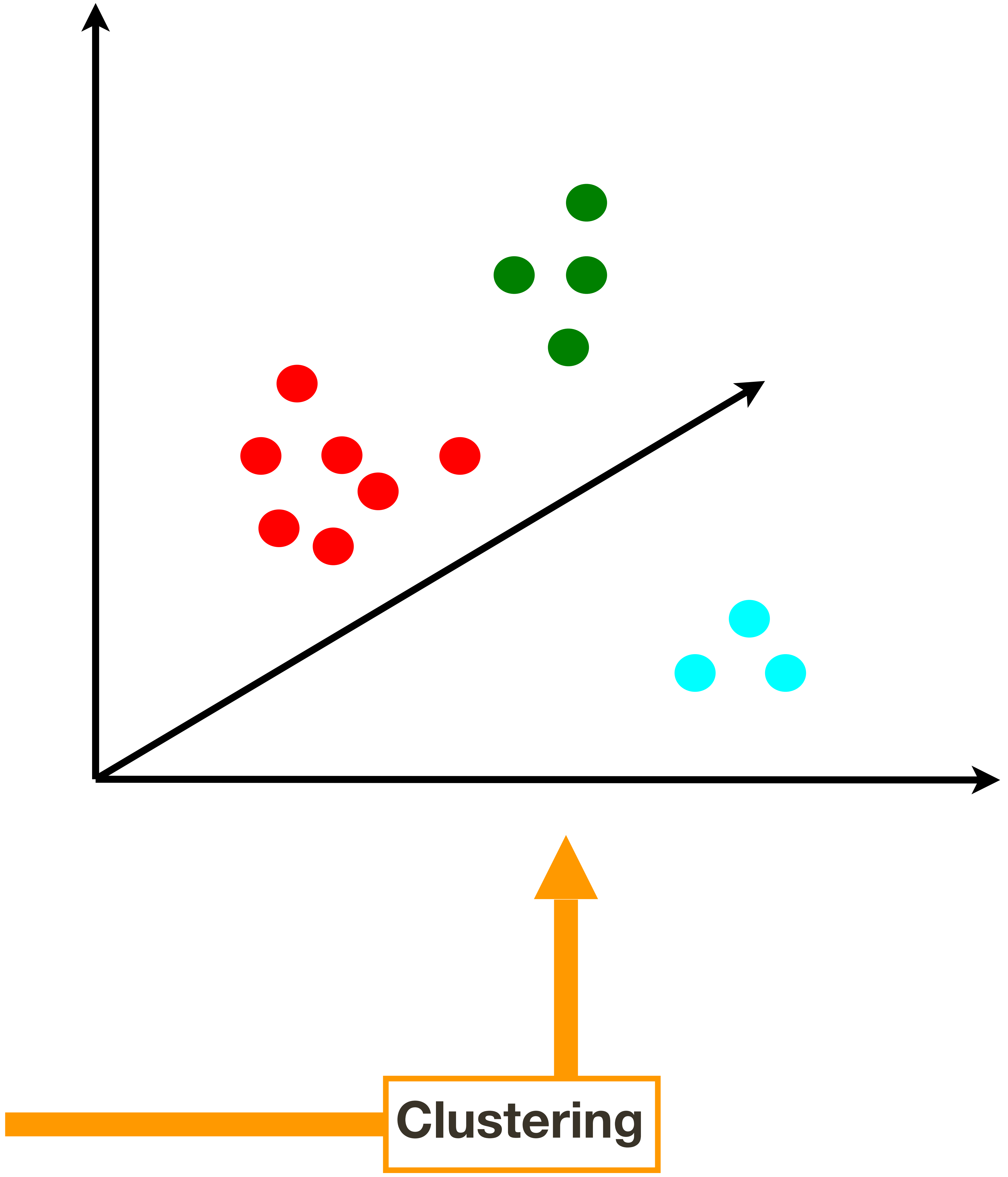
Creating **Dictionary**



Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

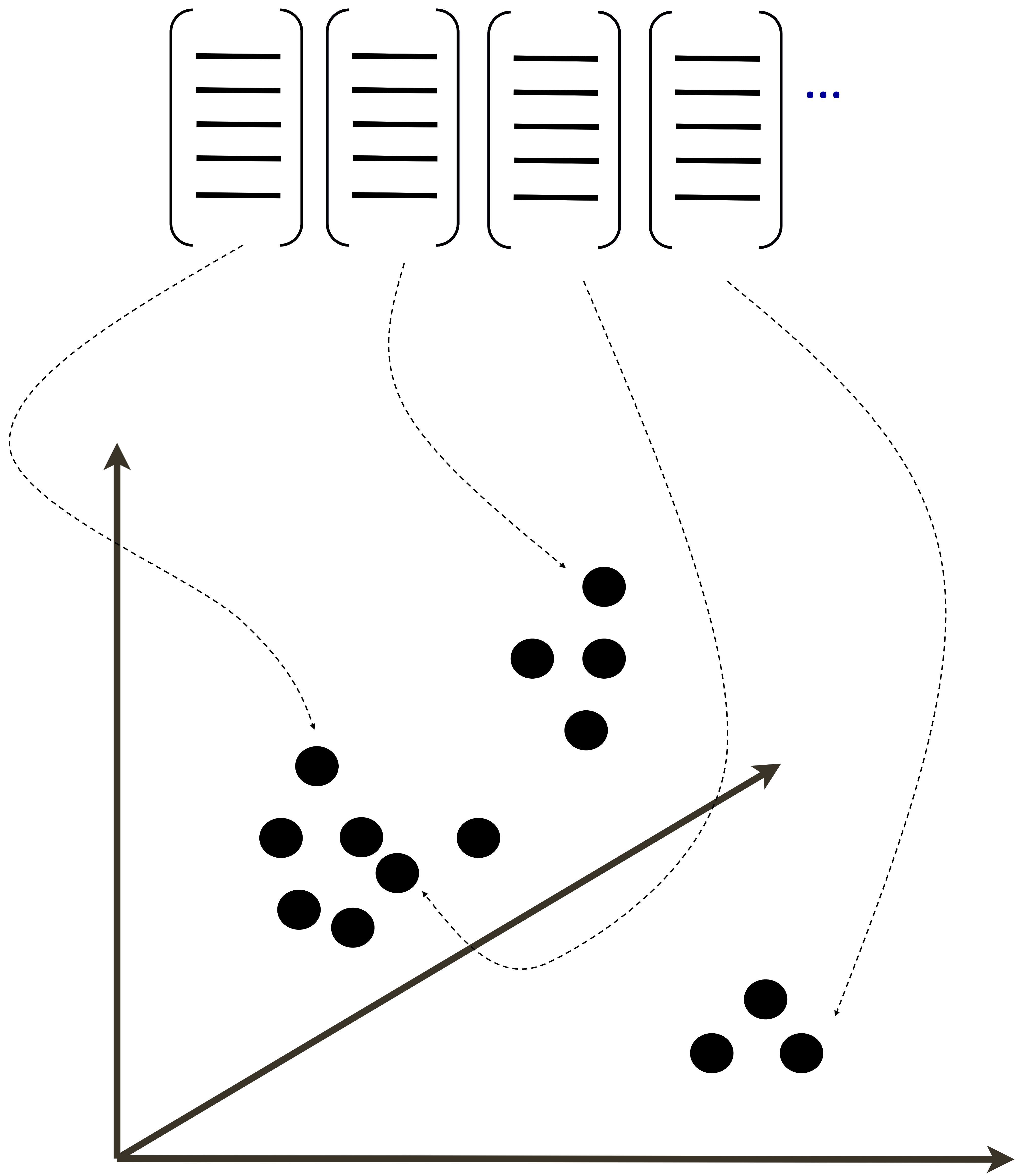
Creating **Dictionary**

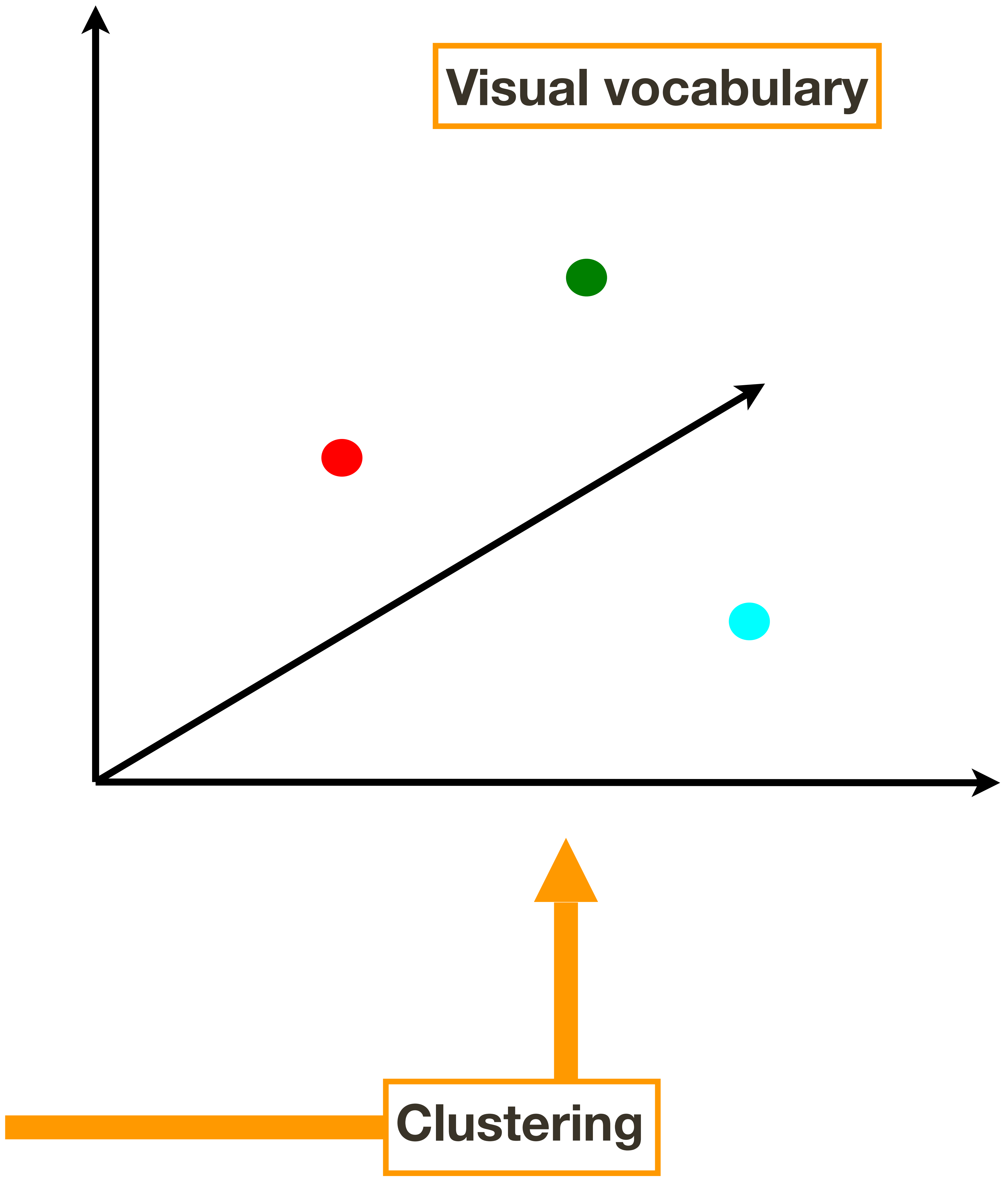




Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Creating **Dictionary**





Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

K-means clustering

K-means Clustering

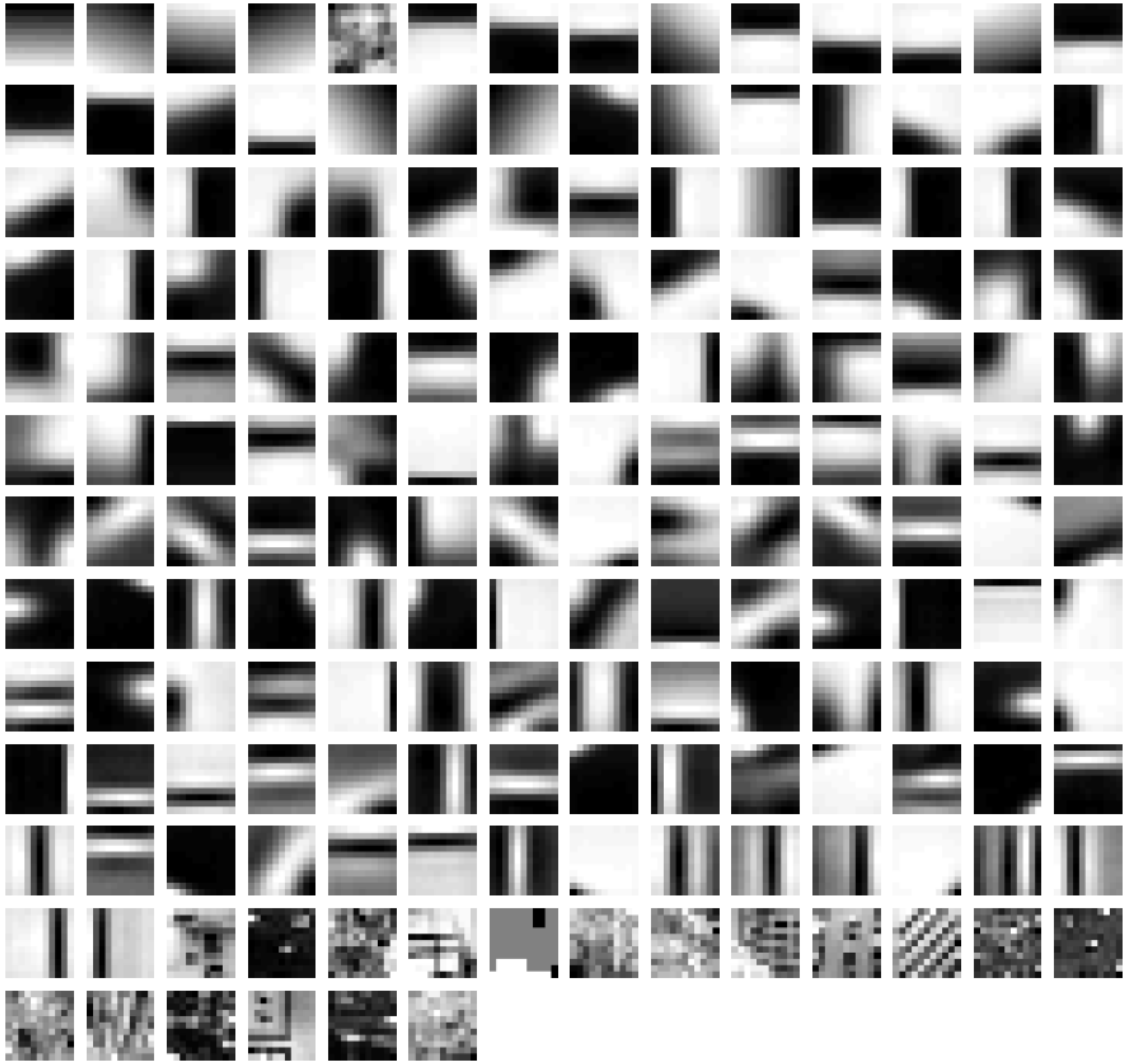
K-means is a clustering technique that iterates between

1. Assume the cluster centers are known. Assign each point to the closest cluster center.

2. Assume the assignment of points to clusters is known. Compute the best cluster center for each cluster (as the mean).

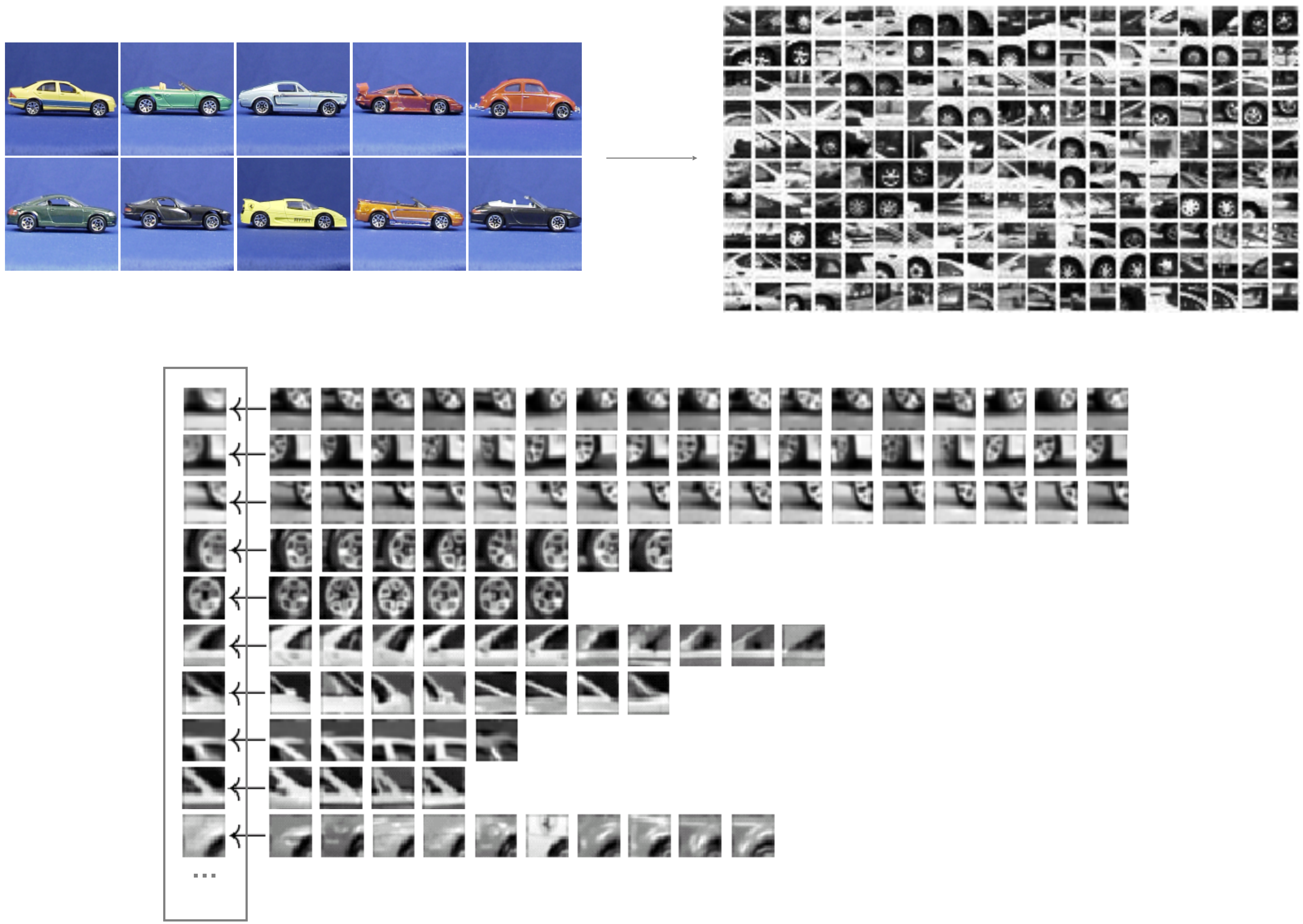
K-means clustering is initialization dependent and converges to a local minimum

Example **Visual Dictionary**



Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Example **Visual Dictionary**



Source: B. Leibe

Example **Visual Dictionary**

Source: B. Leibe

Standard **Bag-of-Words** Pipeline (for image classification)

Dictionary Learning:
Learn Visual Words using clustering

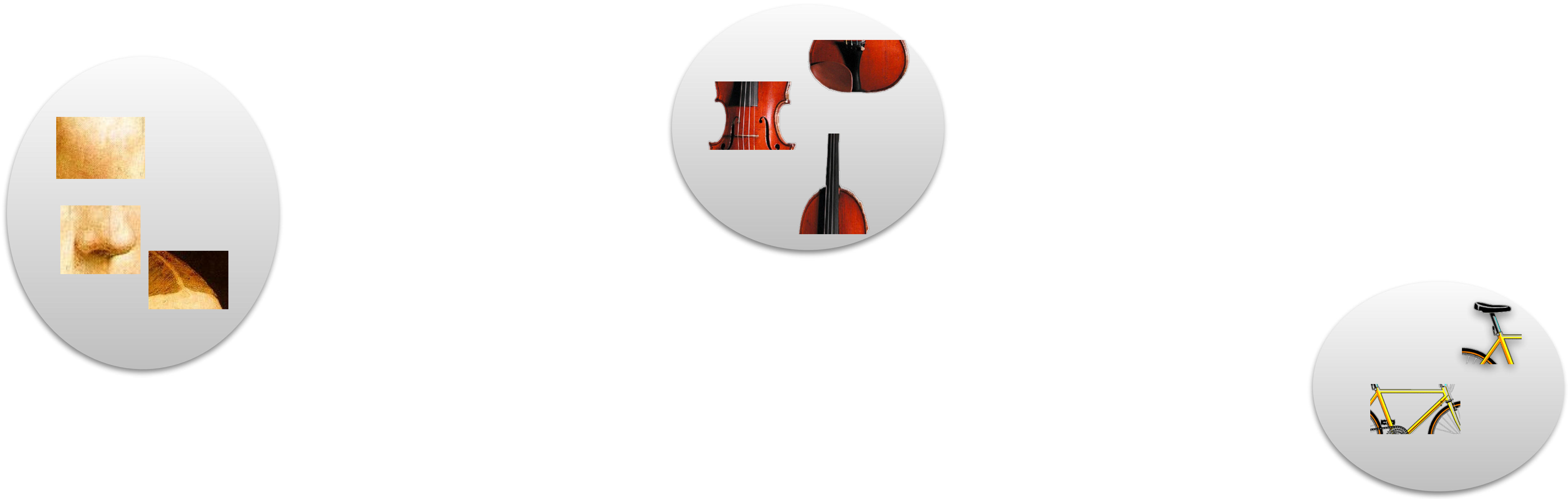
Encode:
build Bags-of-Words (BOW) vectors
for each image

Classify:
Train and test data using BOWs

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

2. **Encode:** build Bag-of-Words (BOW) vectors for each image

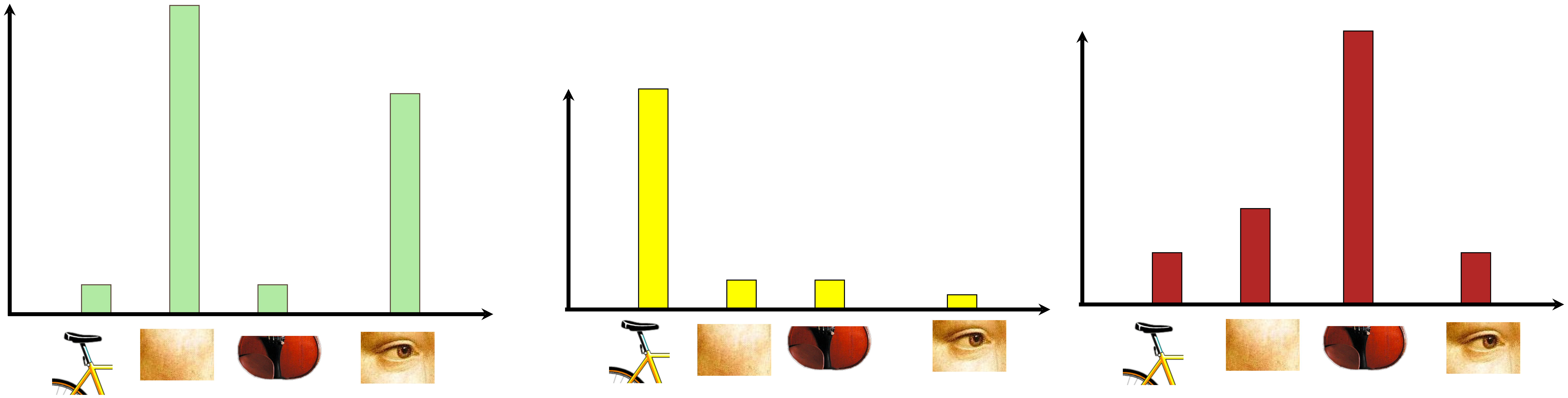
1. **Quantization**: image features gets associated to a visual word (nearest cluster center)



Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

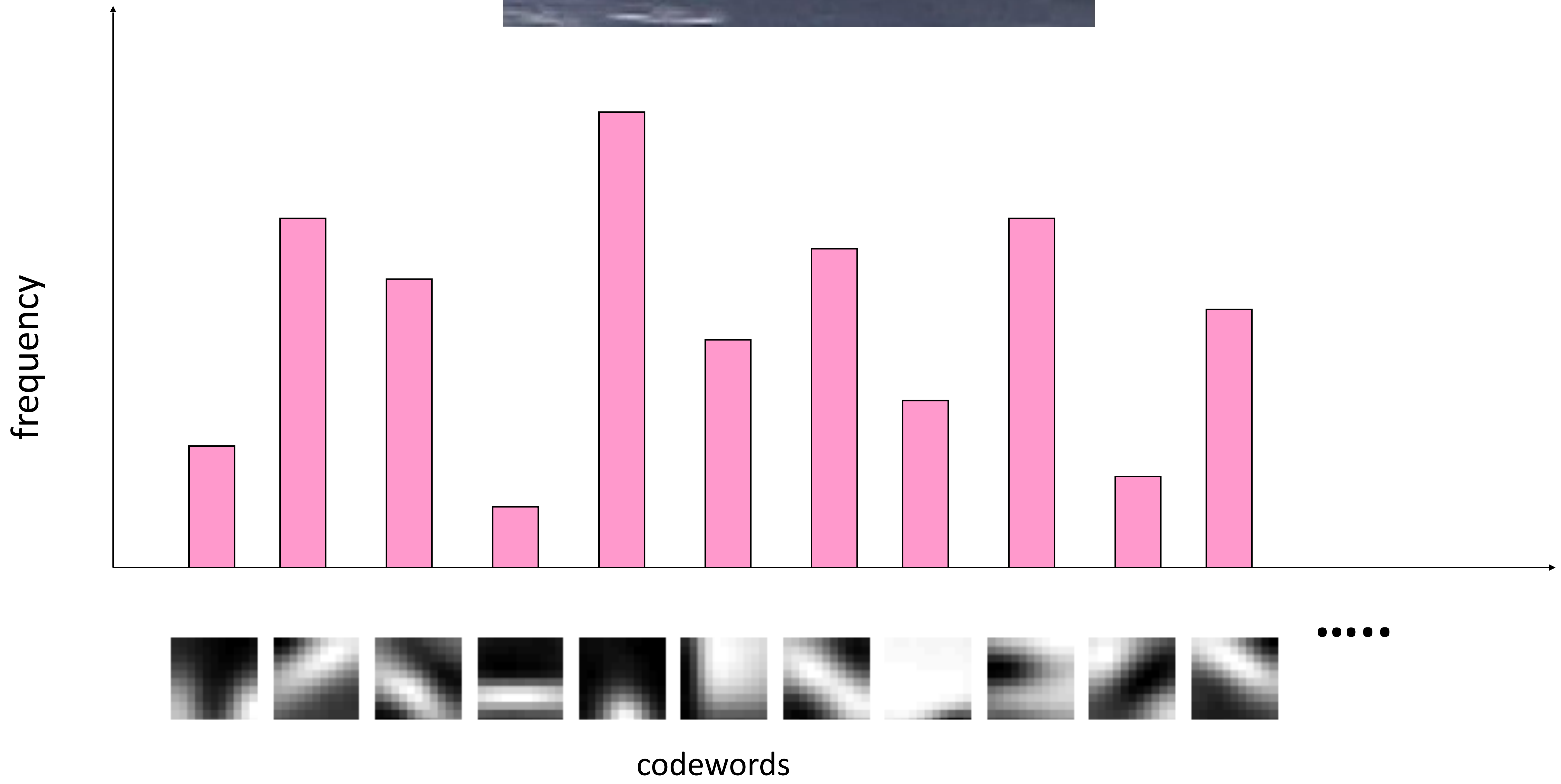
2. Encode: build Bag-of-Words (BOW) vectors for each image

2. **Histogram**: count the number of visual word occurrences



Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

2. **Encode:** build Bag-of-Words (BOW) vectors for each image



Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

Standard **Bag-of-Words** Pipeline (for image classification)

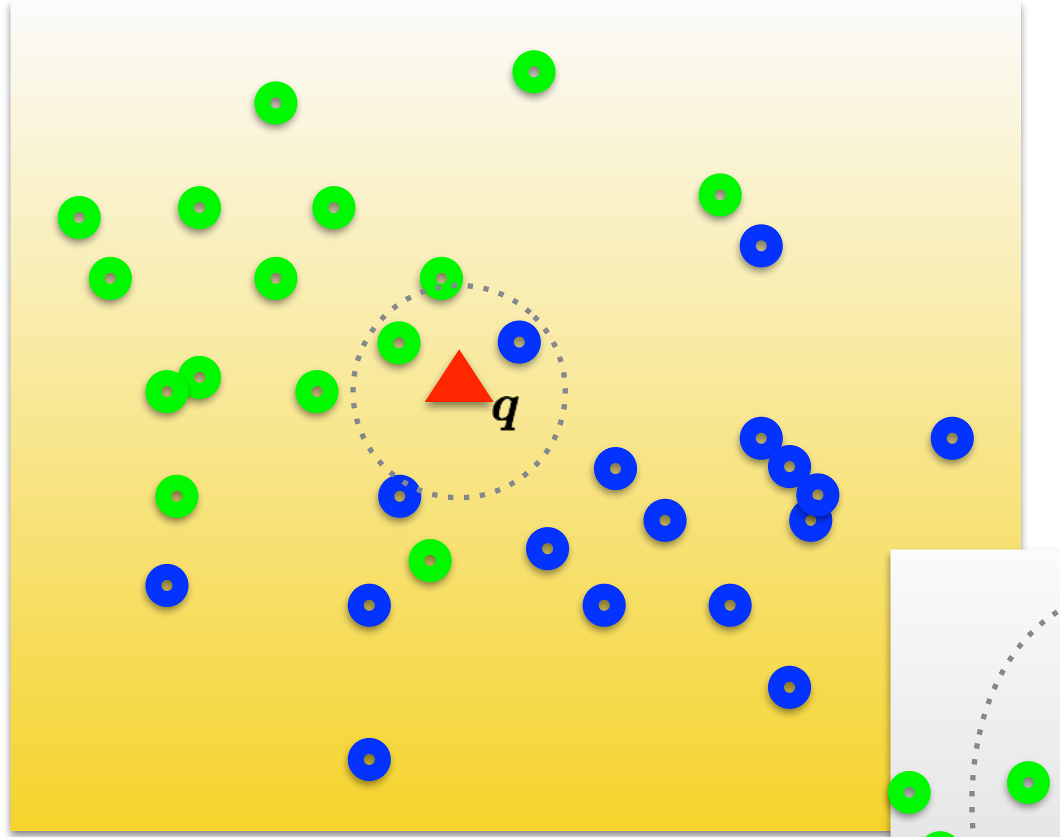
Dictionary Learning:
Learn Visual Words using clustering

Encode:
build Bags-of-Words (BOW) vectors
for each image

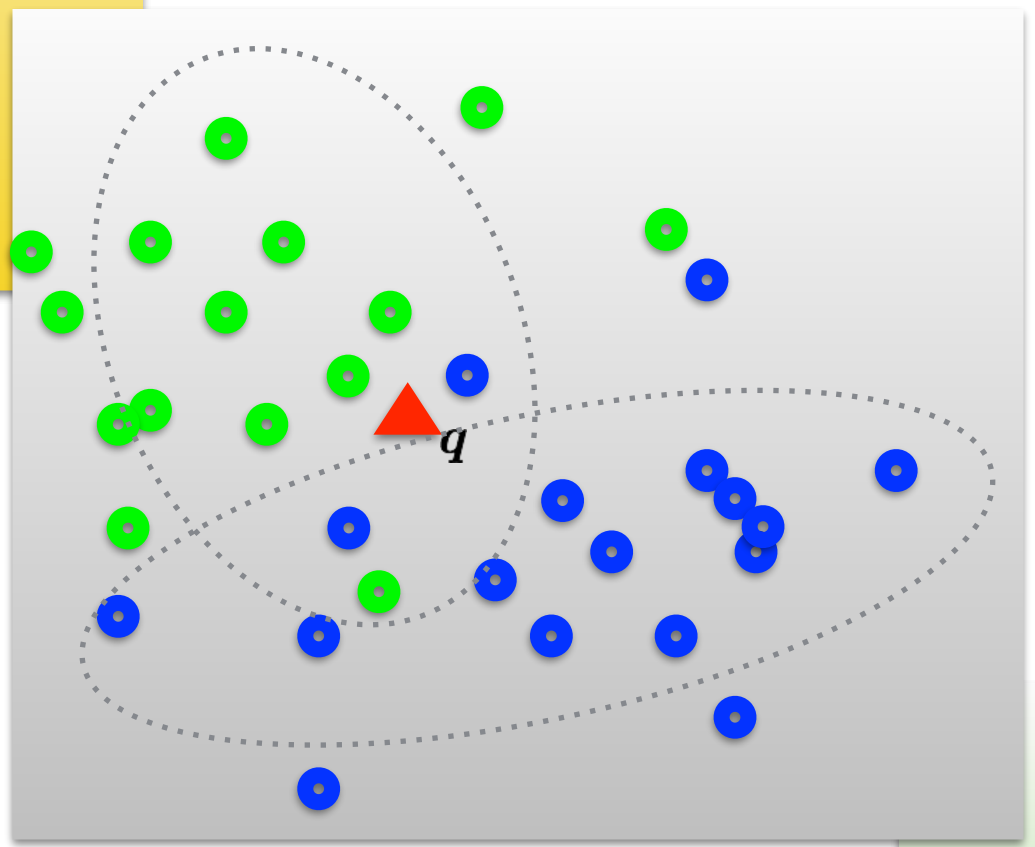
Classify:
Train and test data using BOWs

Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)

3. Classify: Train and text classifier using BOWs

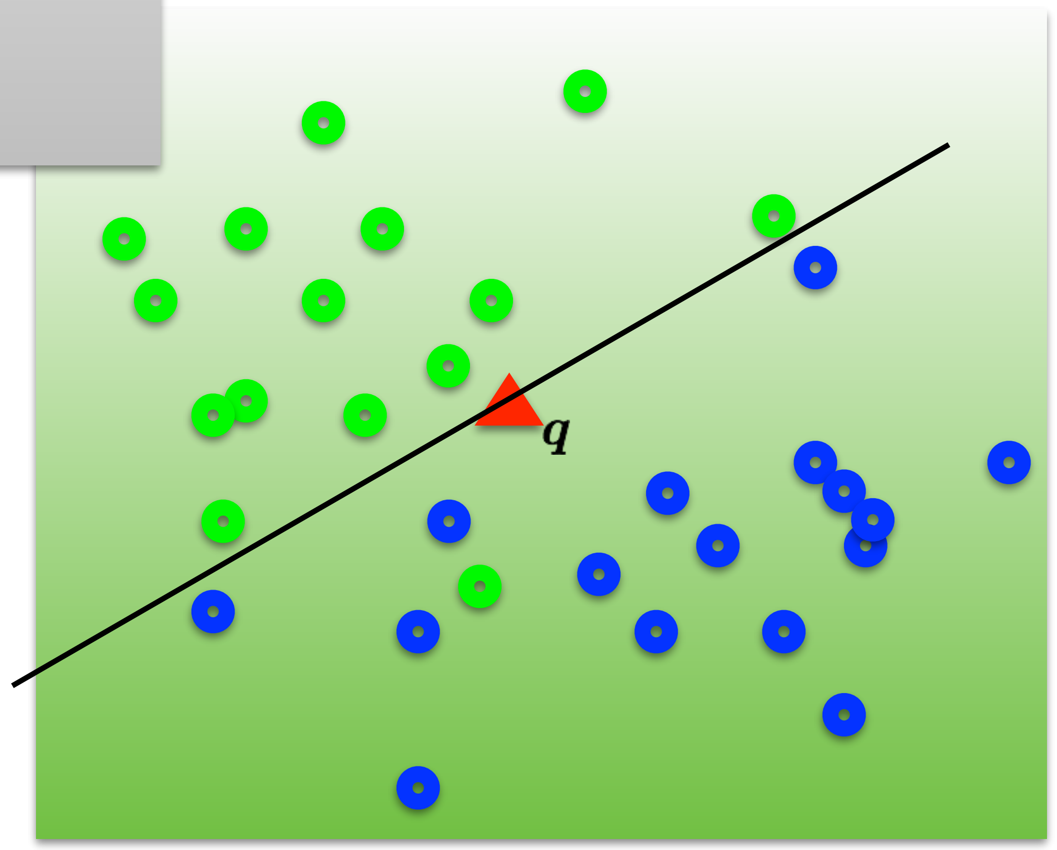


K nearest
neighbors



Naïve
Bayes

Support
Vector
Machine



Slide Credit: Ioannis (Yannis) Gkioulekas (CMU)