Contents

- Bags of features
  - Textons for texture representation
  - Bags of words
  - ‘Dictionary learning’ (clustering)
  - Recognition
- Tracking
  - Meanshift and CAMshift
  - Other approaches

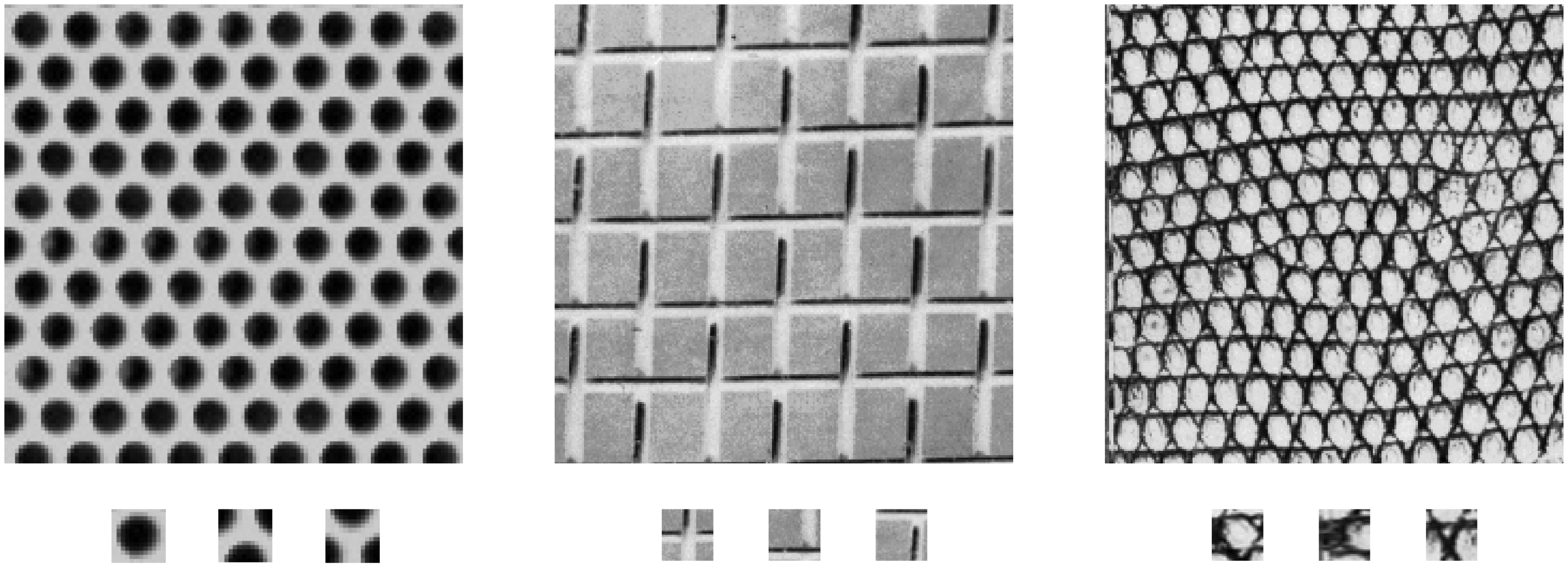
Bag of features models

Some slides adapted from Fei-Fei Li, Rob Fergus and Antonio Torralba

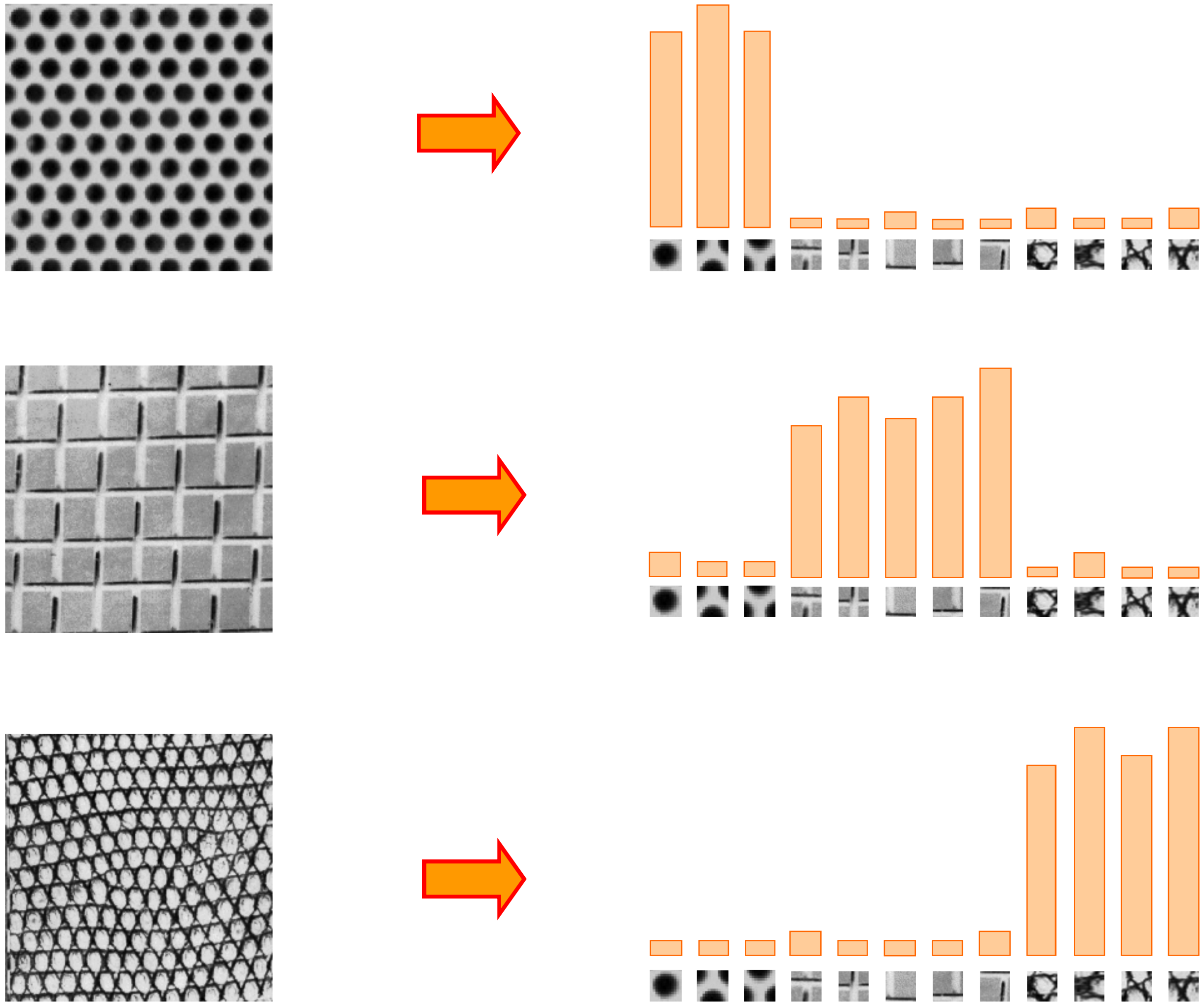
Origin 1: Texture recognition

Texture is characterized by the repetition of basic elements or *textons*.

Arguably, for stochastic textures, the identity of the textons matters more than their spatial arrangement.



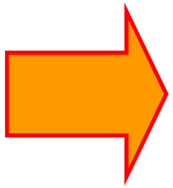
Origin 1: Texture recognition



Origin 2: Bag-of-words models

- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

Bags of features for object recognition

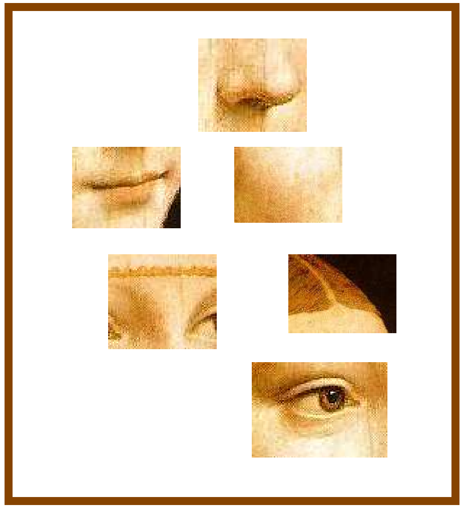


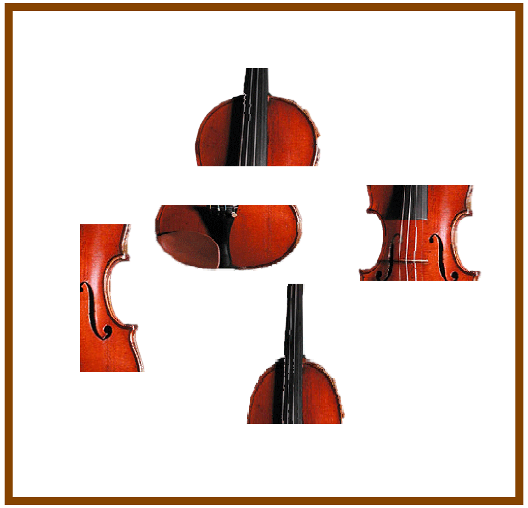
face, flowers, building

Quite effective for image-level classification

Bag of features: overview

1. Extract features





Bag of features: overview

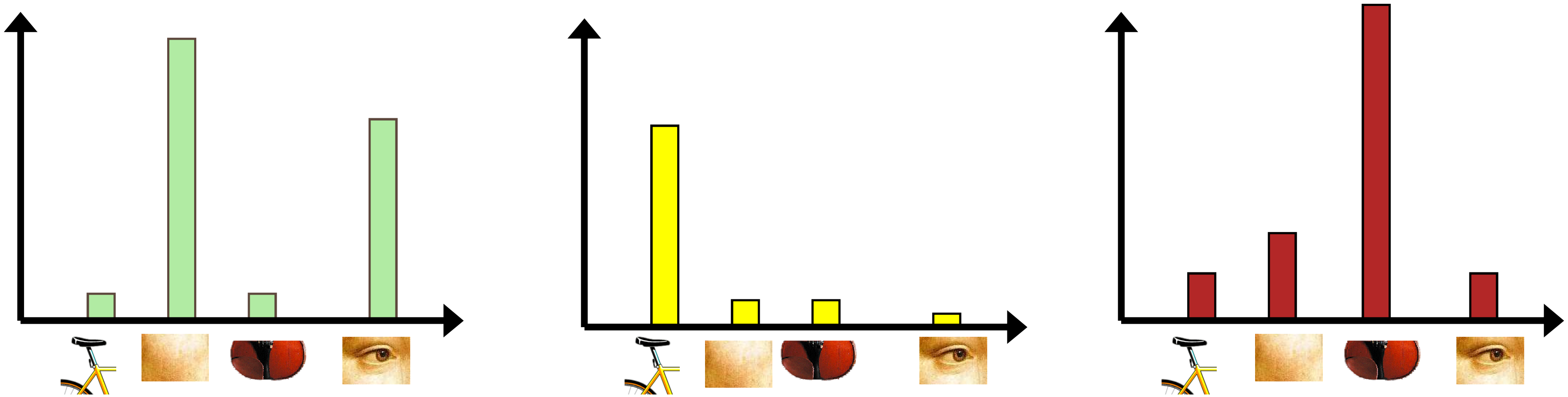
1. Extract features
2. Learn visual vocabulary ('dictionary' of visual 'words')

Bag of features: overview

1. Extract features
2. Learn visual vocabulary ('dictionary' of visual 'words')
3. Quantize features using visual vocabulary

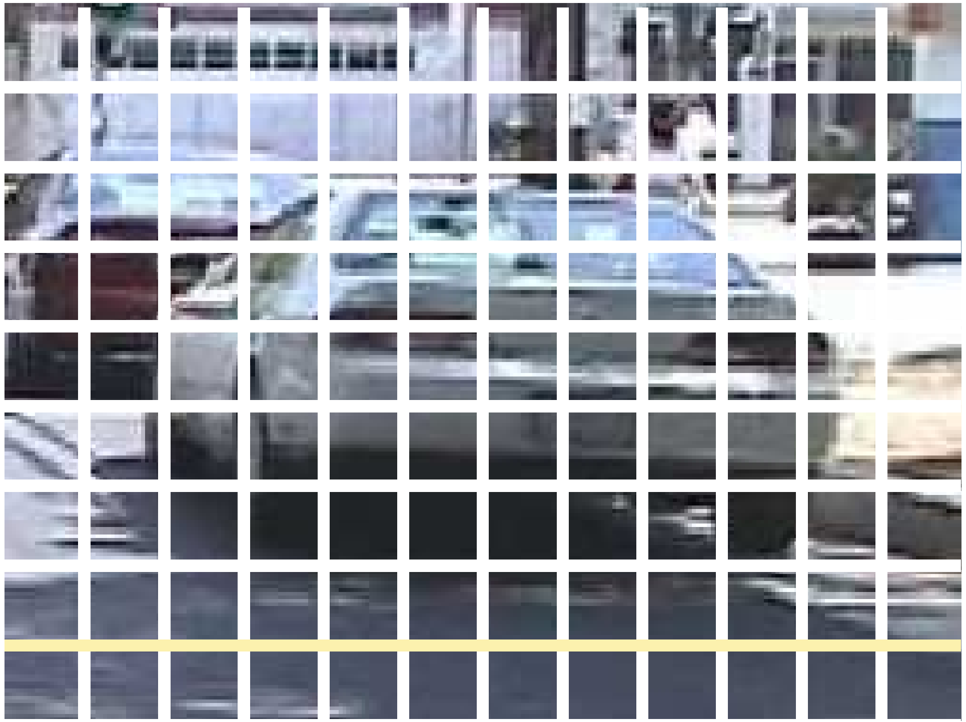
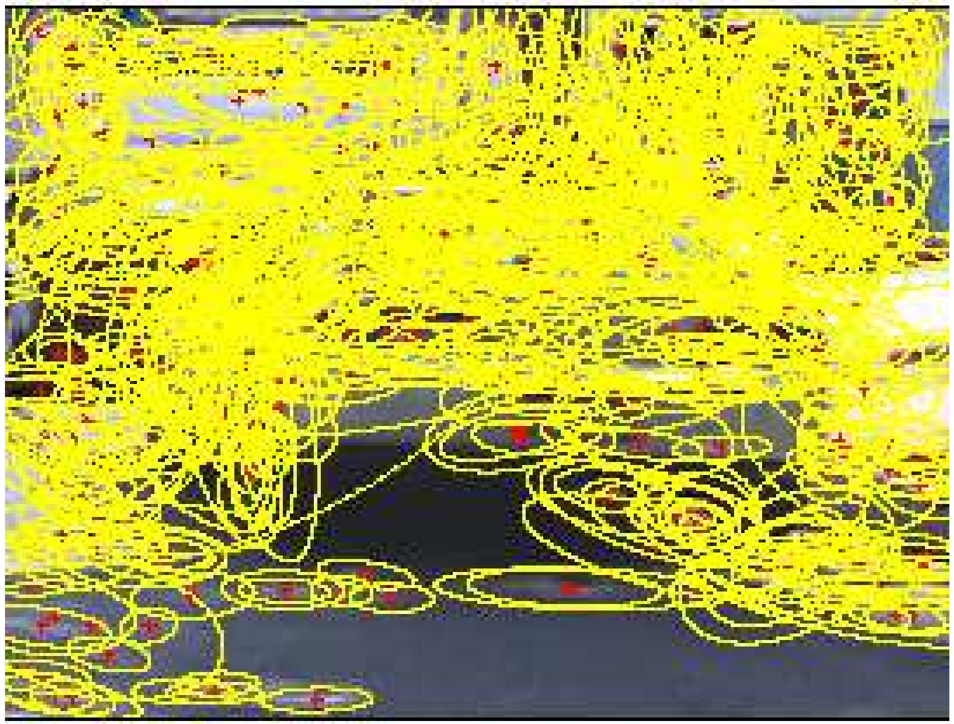
Bag of features: overview

1. Extract features
2. Learn visual vocabulary ('dictionary' of visual 'words')
3. Quantize features using visual vocabulary
4. Represent images by frequencies of visual words



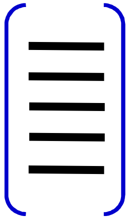
1. Feature extraction

- Extract features at locations on a regular grid, at random locations, at detected feature points, from superpixels…

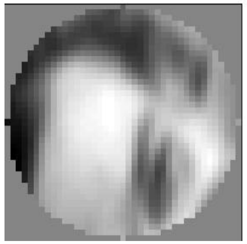
 

- Features: colour, texture, SIFT, SURF, etc.

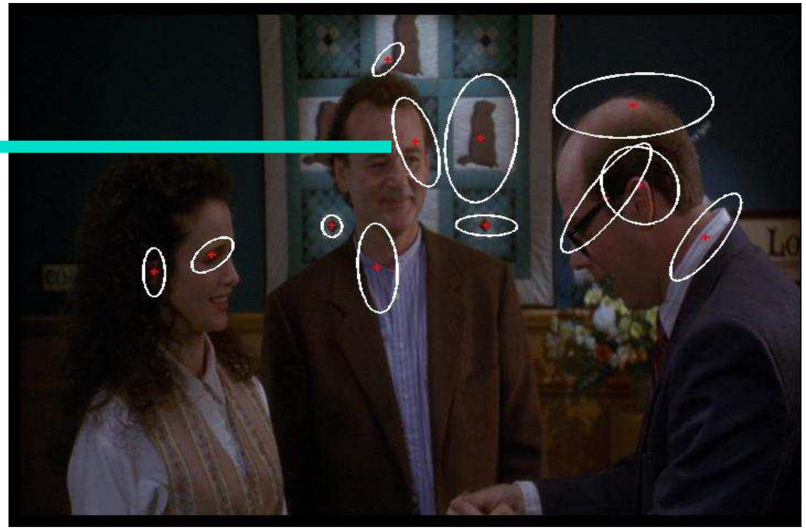
1. Feature extraction



SIFT descriptor

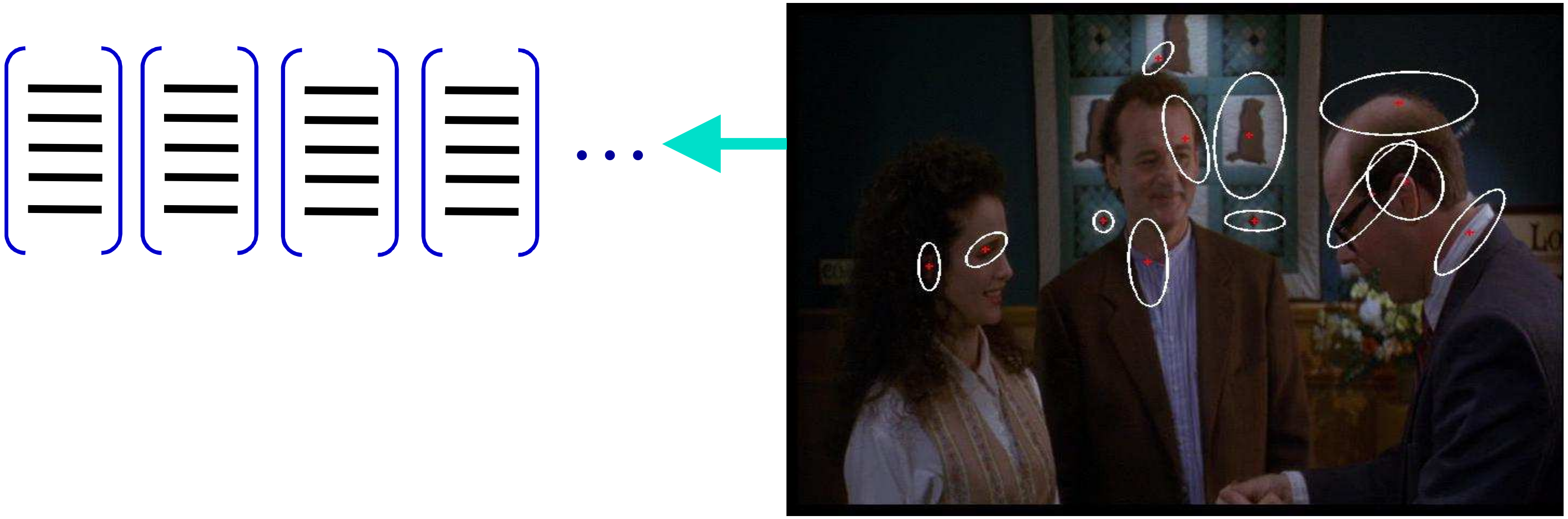


Normalized patch

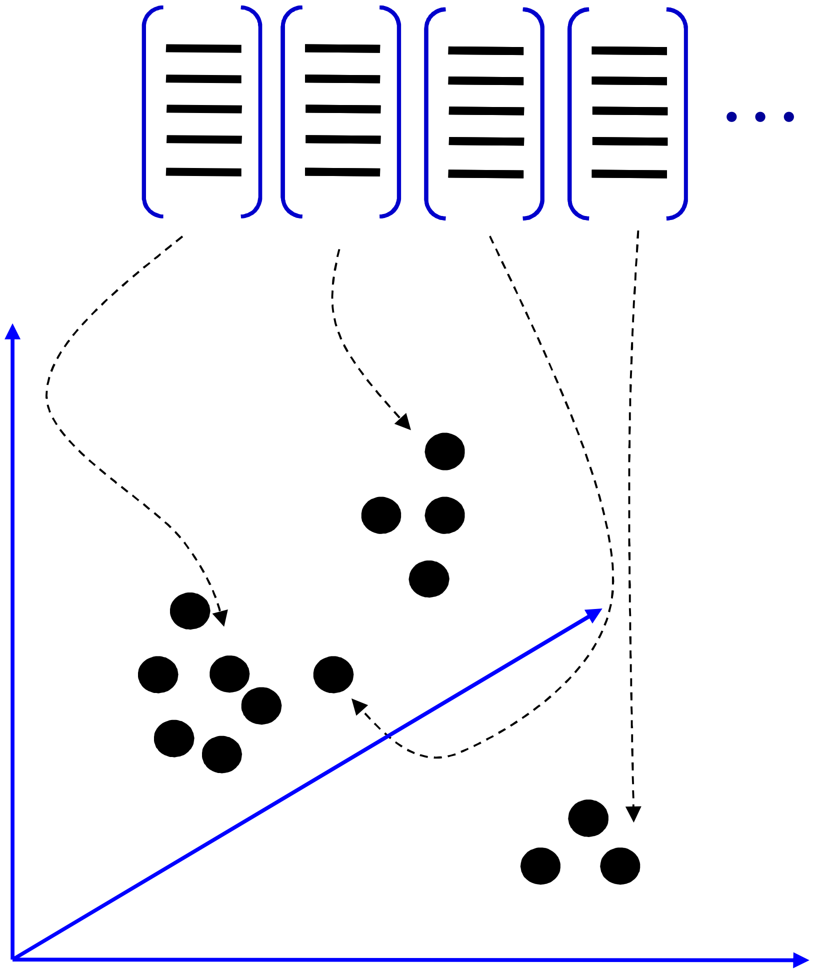


Patches (feature detection)

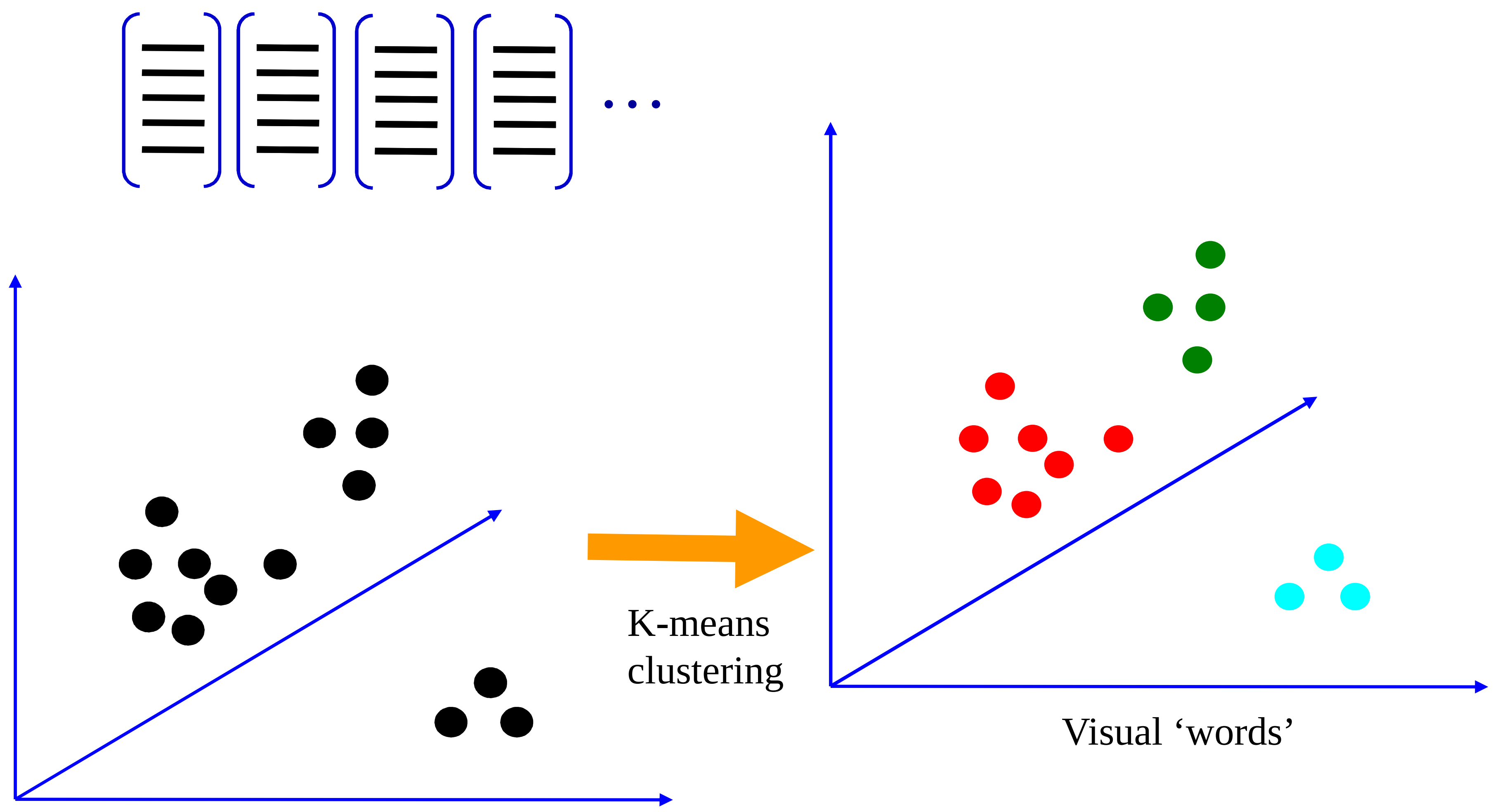
1. Feature extraction



2. Learning the visual vocabulary

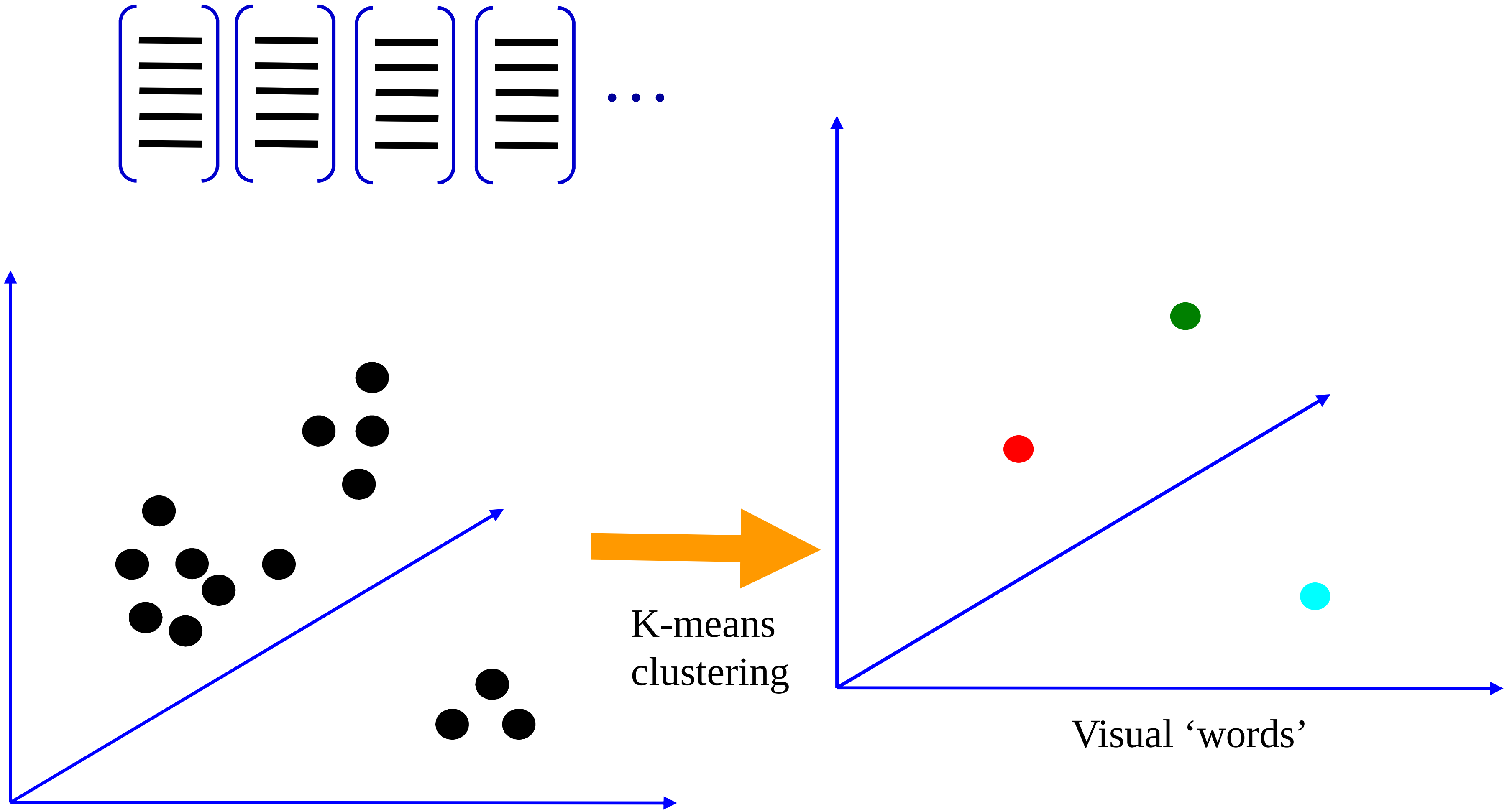


2. Learning the visual vocabulary



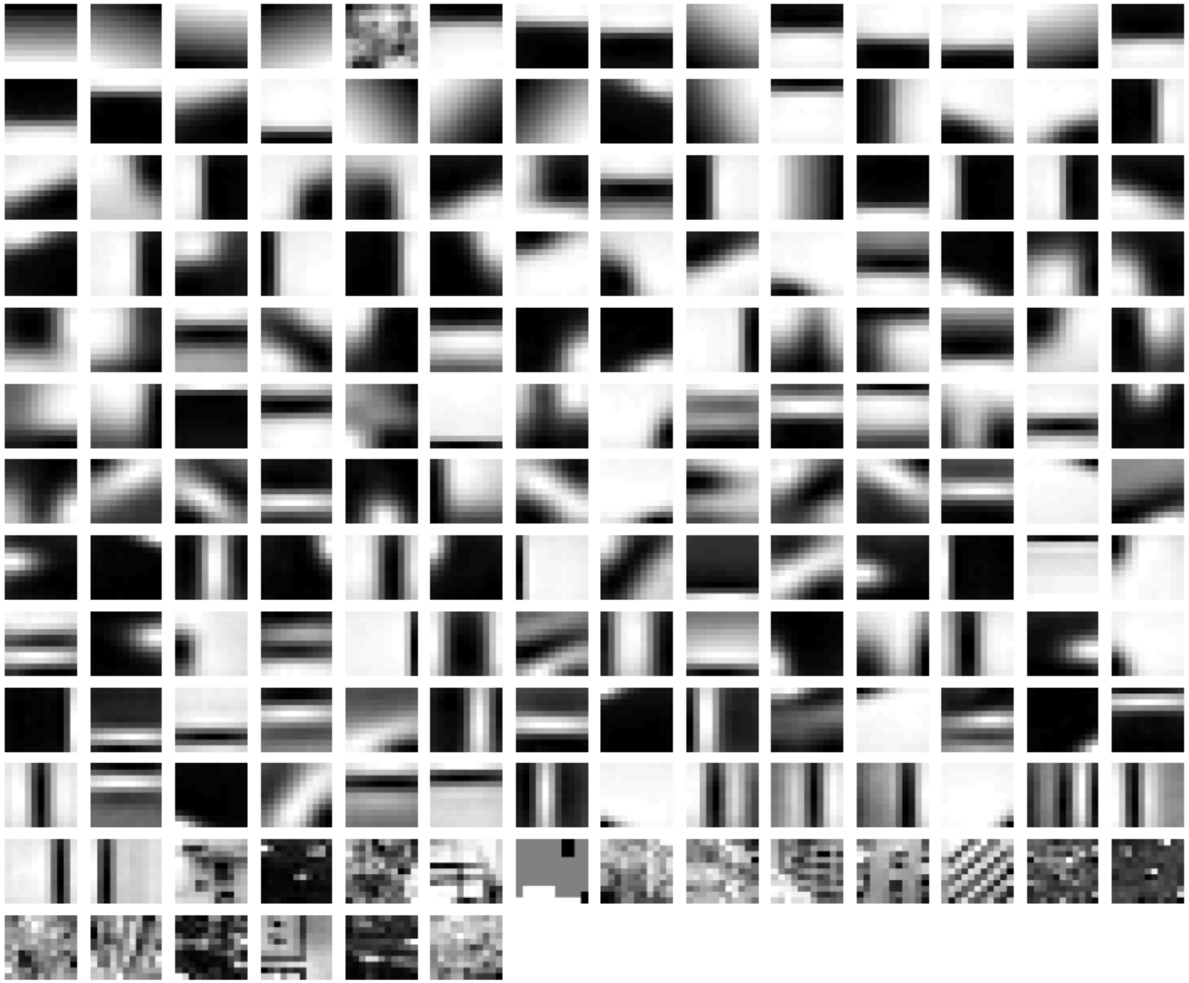
Based on slide by Josef Sivic

2. Learning the visual vocabulary



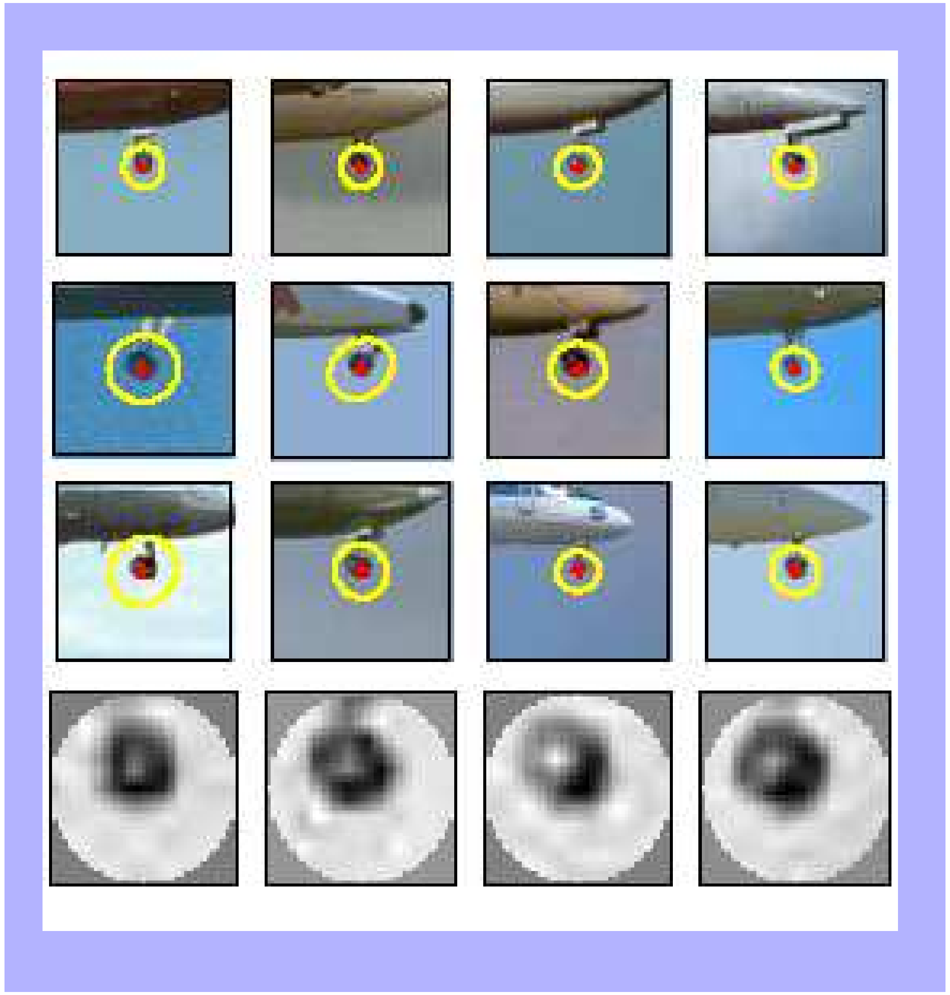
Based on slide by Josef Sivic

Example visual vocabulary



Fei-Fei et al. 2005

Image patch examples of visual words



Sivic et al. 2005

3. Image representation

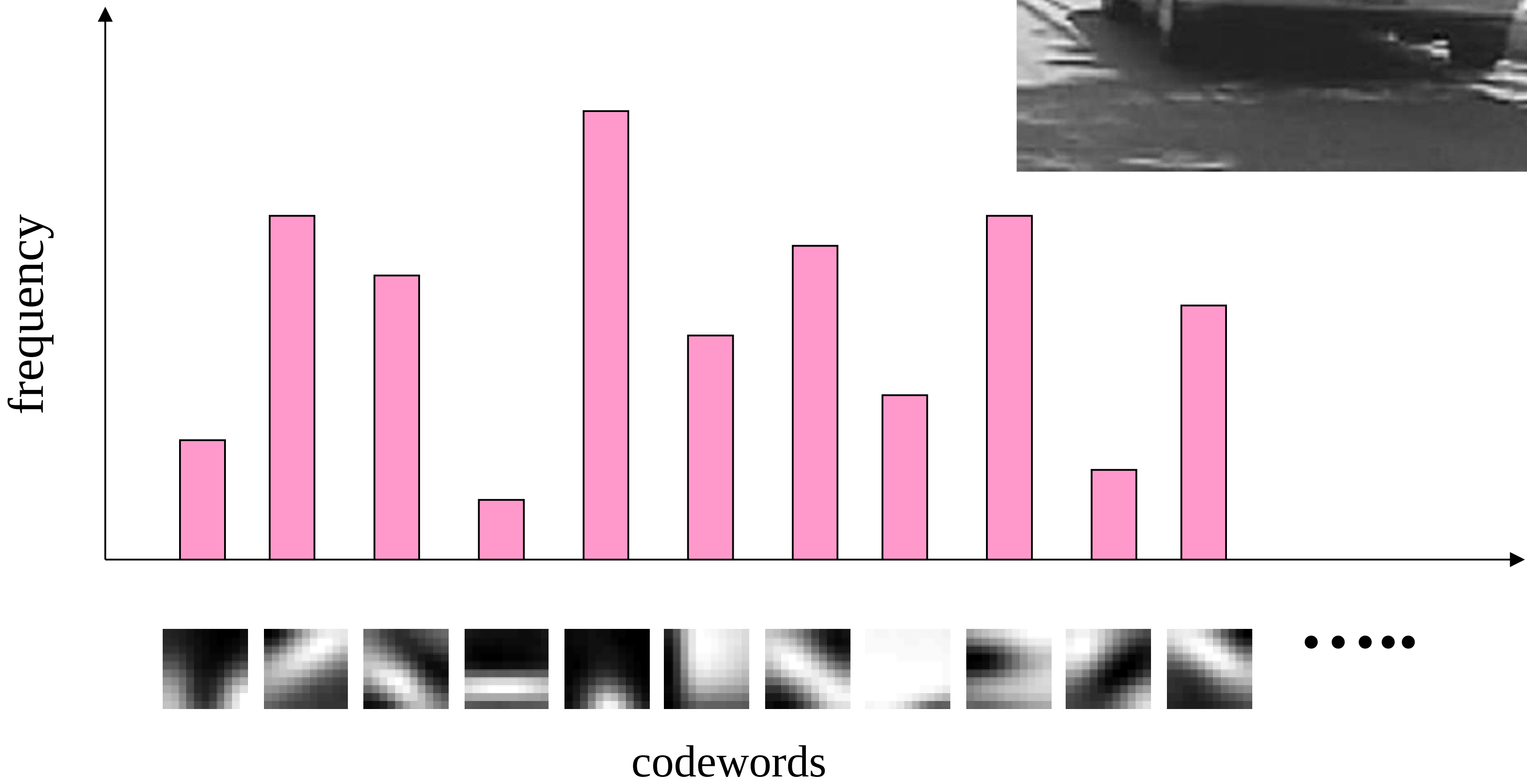
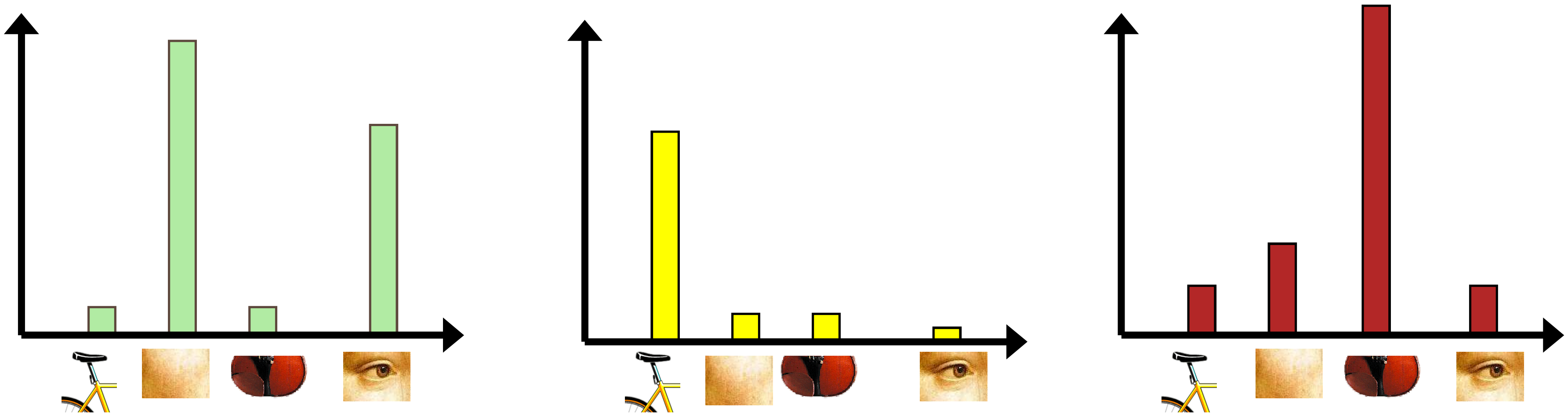


Image classification

Given a *bag-of-features* representation of each image:

- Train a classifier using the histograms as feature vectors
- Could involve defining a measure of histogram similarity



Tracking

- Following objects or features through an image sequence
- Estimating location (and orientation) in each frame
- Physical Constraints:
  - Inertia: motion cannot change abruptly

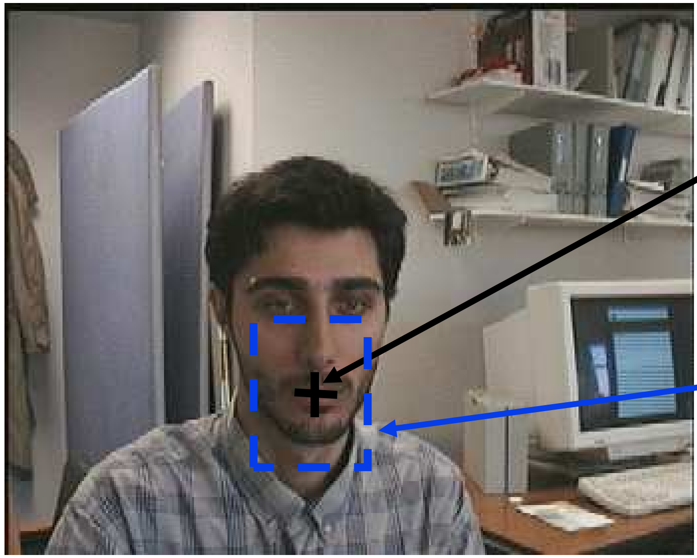
    ⇨No abrupt motion changes are observed *provided that* the frame rate is fast enough
  - If a 3D trajectory is smooth then its 2D projection is also smooth
- Useful Assumptions:
  - Location, speed and direction of motion do not change much between frames
  - Image motion is smooth

Targets to Track

- Local features (e.g. interest points, small objects)
- Contour fragments (e.g. partial object boundaries)
- Objects (possibly multiple parts, possibly deformable)

Tracking example

- Search in a local window
- Window position depends on previous estimate
- Extent of window depends on expected maximum speed of image motion

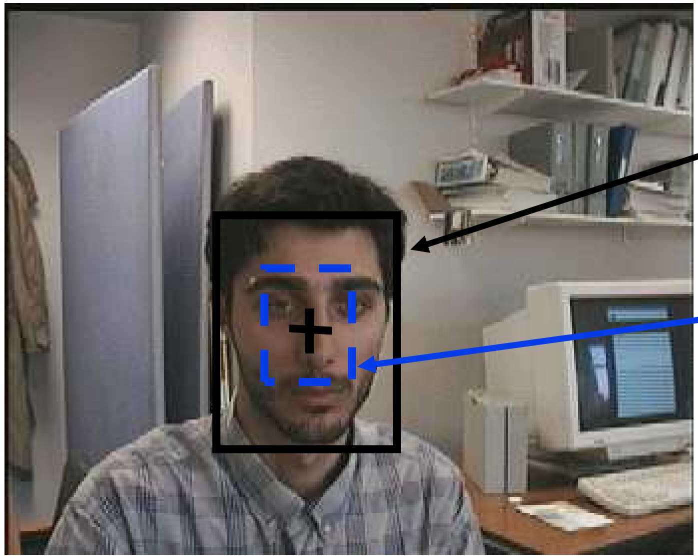


Estimate at frame t

Search window for frame t+1

Face tracking example

- Want to estimate location and scale (fixed aspect ratio)
- Search ranges depend on previous estimates

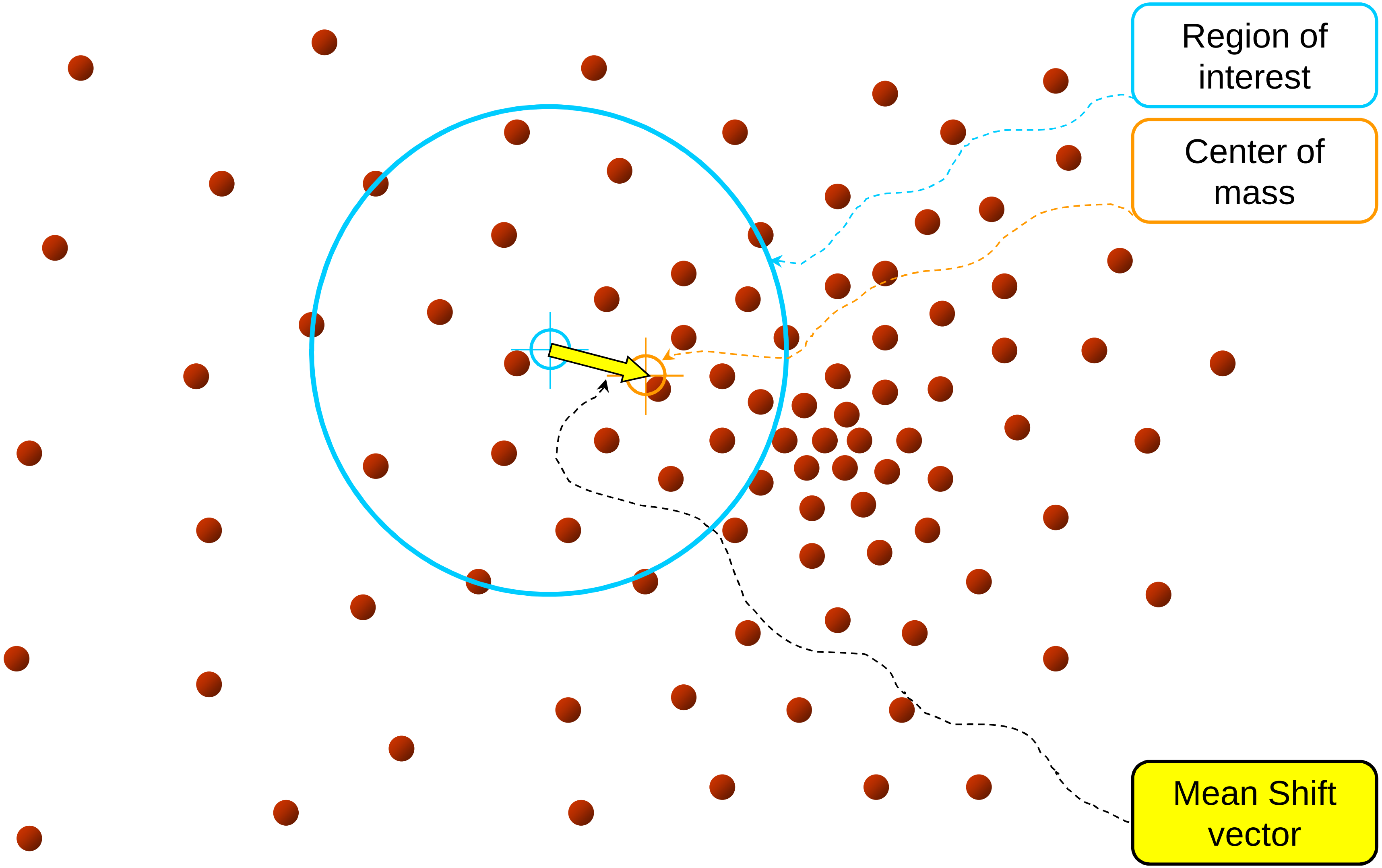


Box indicates estimated location and scale at time t

At t+1, search for object centred within this window and at a range of scales

Example: Head Tracking

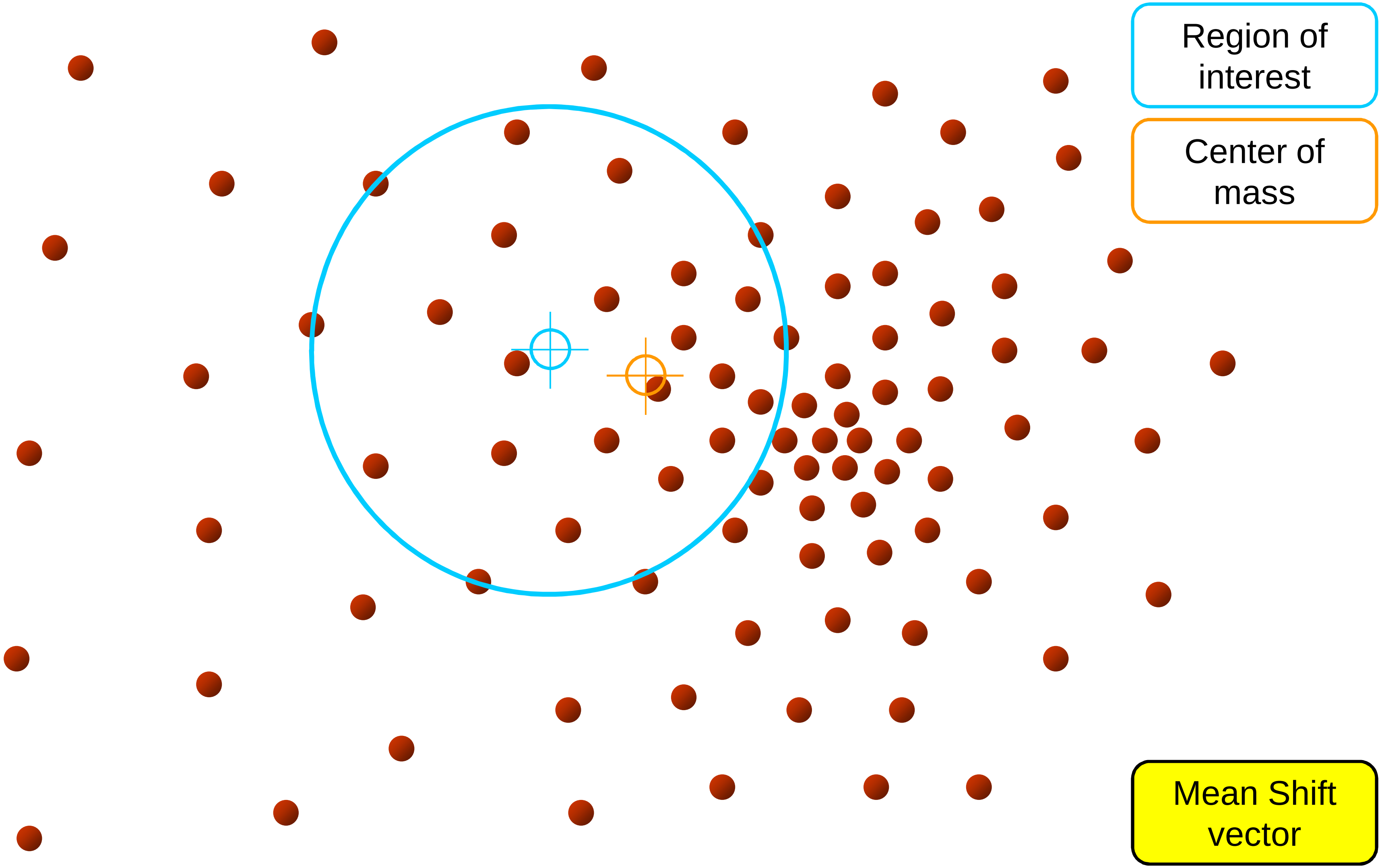
Mean Shift: Intuitive Description



Objective : Find the densest region

Distribution of identical billiard balls

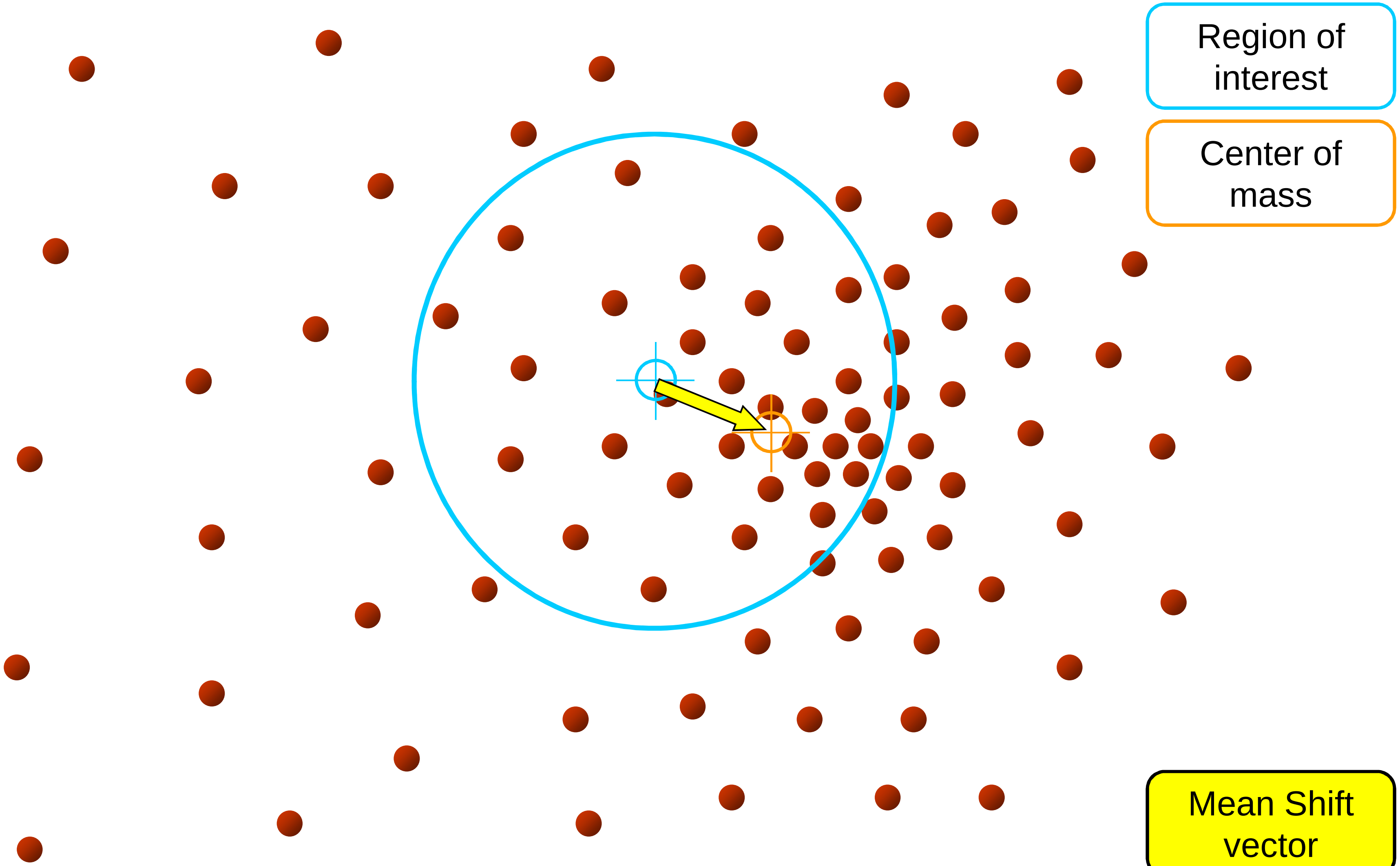
(Slide credit: Yaron Ukrainitz & Bernard Sarel)



Objective : Find the densest region

Distribution of identical billiard balls

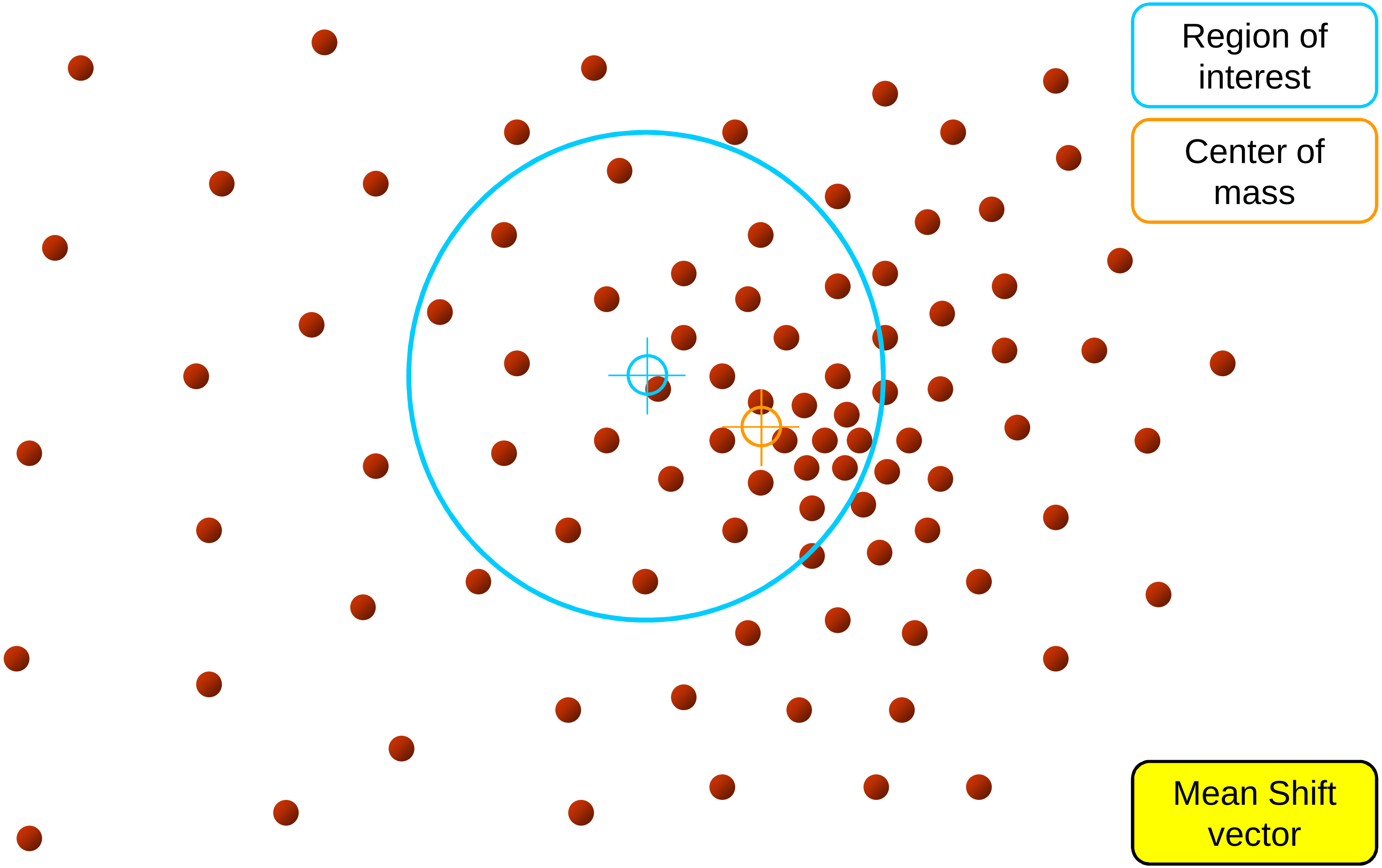
(Slide credit: Yaron Ukrainitz & Bernard Sarel)



Objective : Find the densest region

Distribution of identical billiard balls

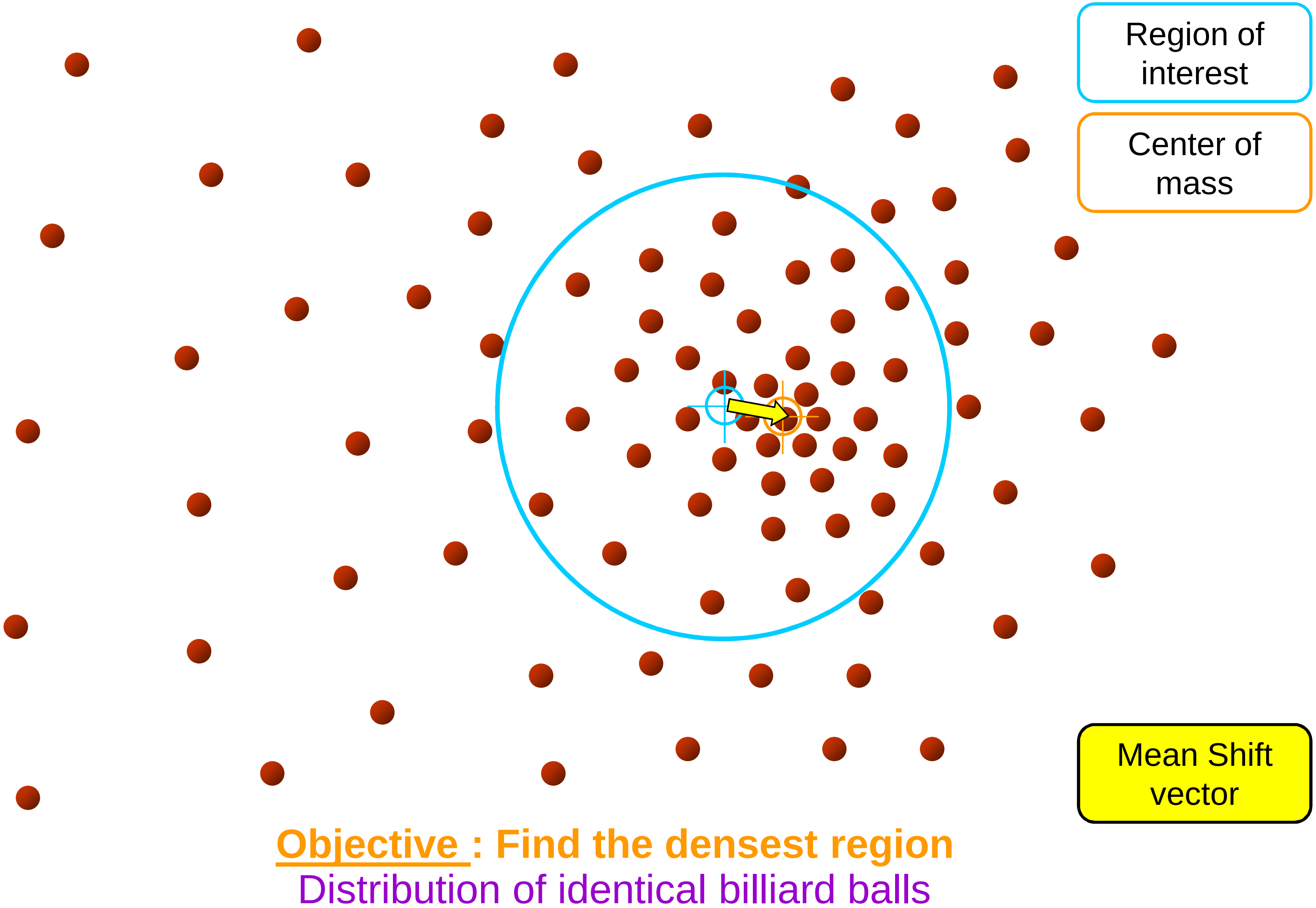
(Slide credit: Yaron Ukrainitz & Bernard Sarel)



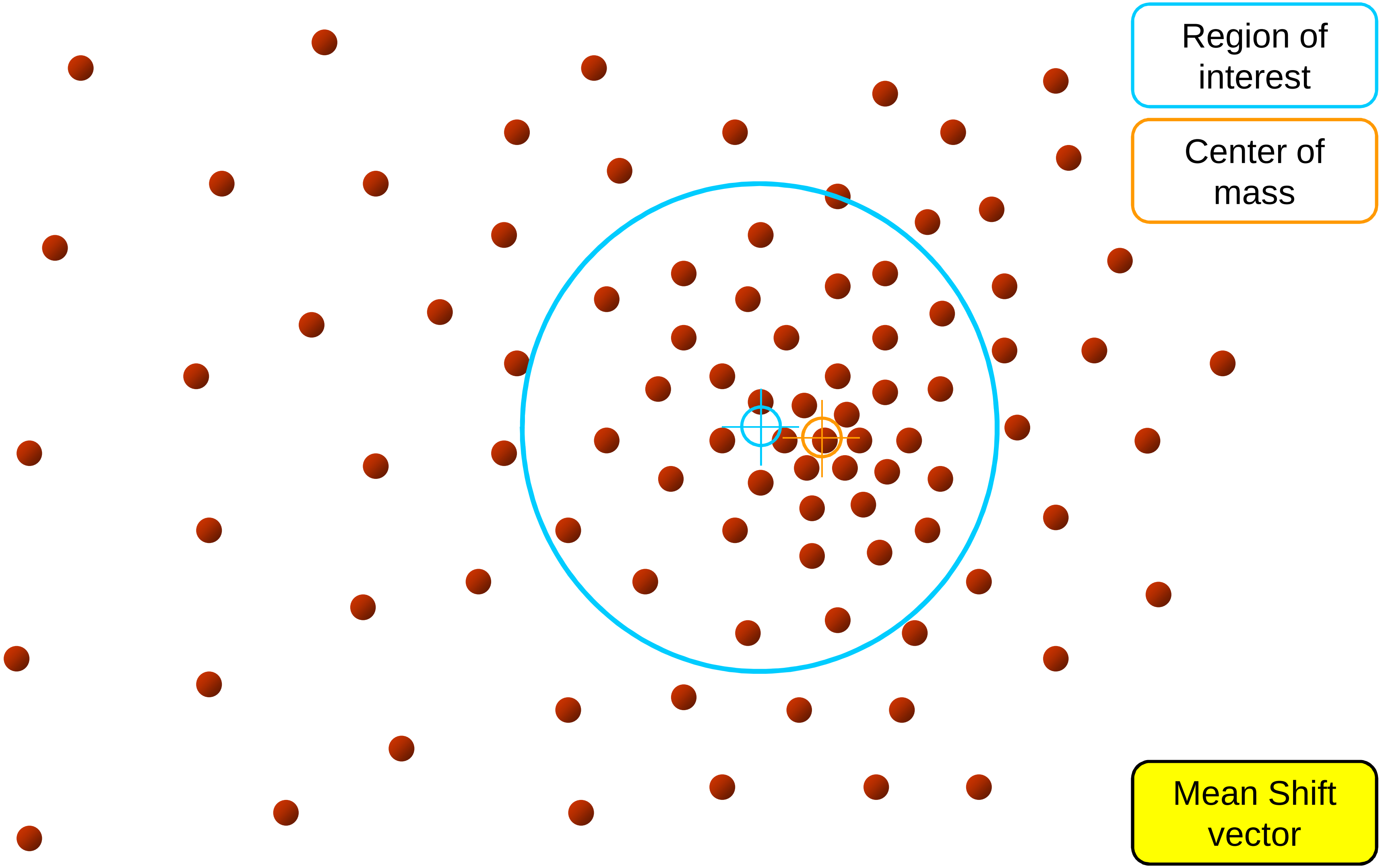
<u>Objective</u> : Find the densest region

Distribution of identical billiard balls

(Slide credit: Yaron Ukrainitz & Bernard Sarel)



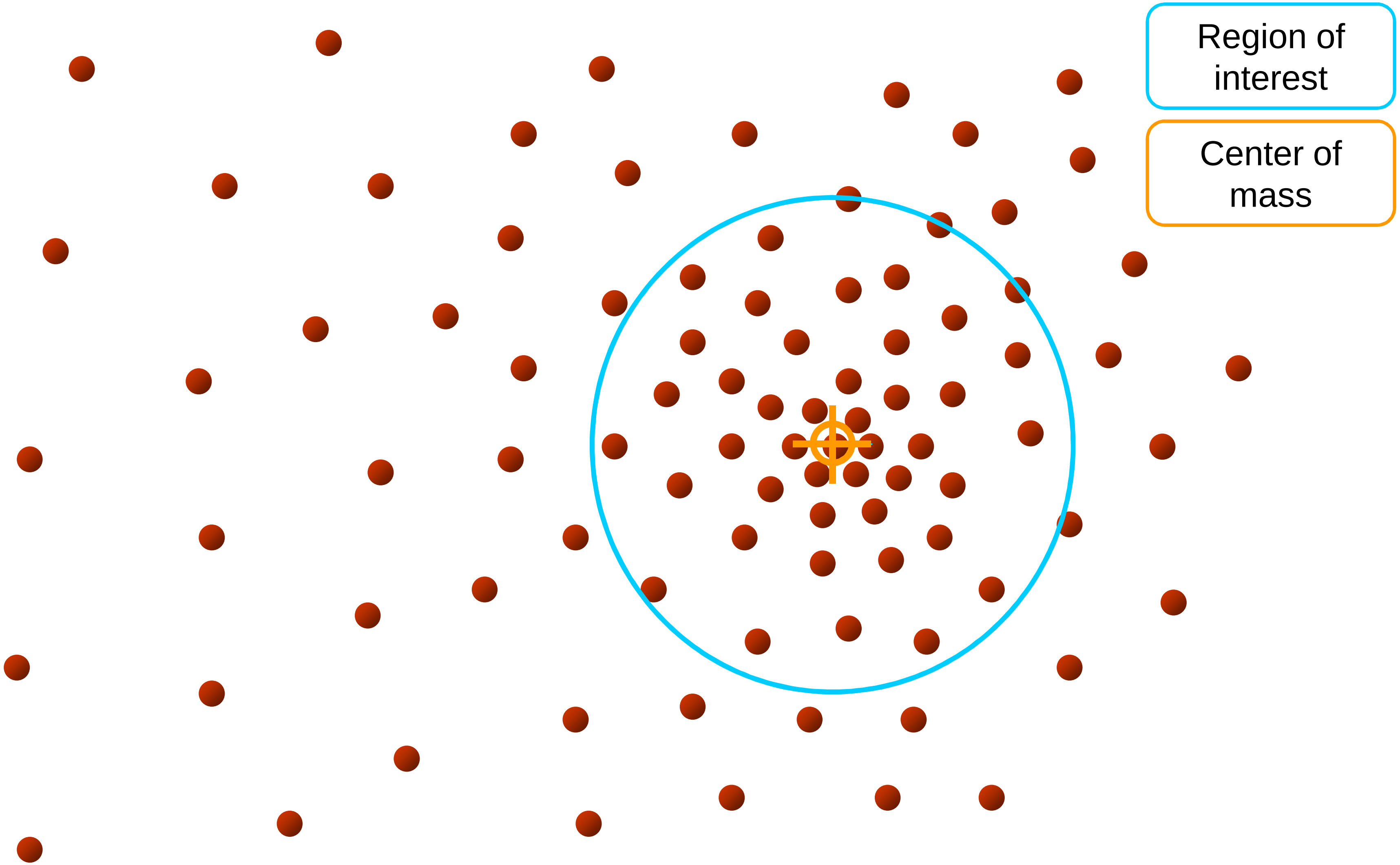
(Slide credit: Yaron Ukrainitz & Bernard Sarel)



Objective : Find the densest region

Distribution of identical billiard balls

(Slide credit: Yaron Ukrainitz & Bernard Sarel)



Objective : Find the densest region

Distribution of identical billiard balls

(Slide credit: Yaron Ukrainitz & Bernard Sarel)

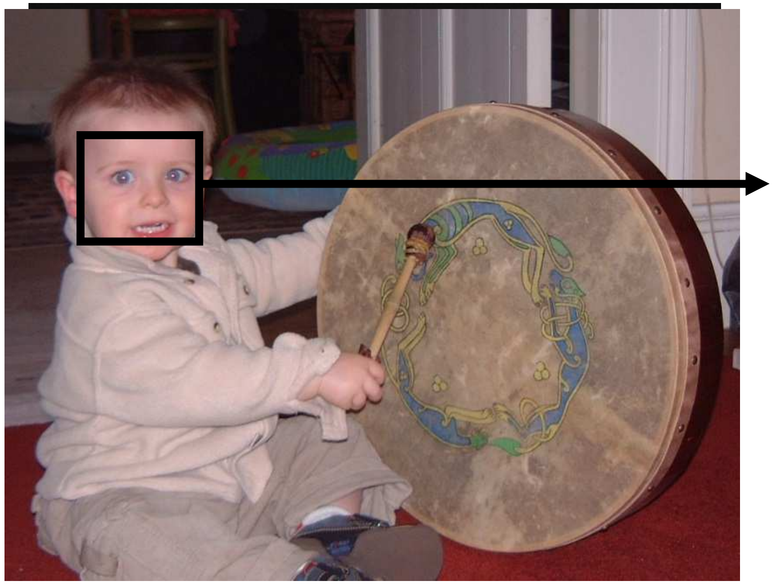
What is Mean Shift ?

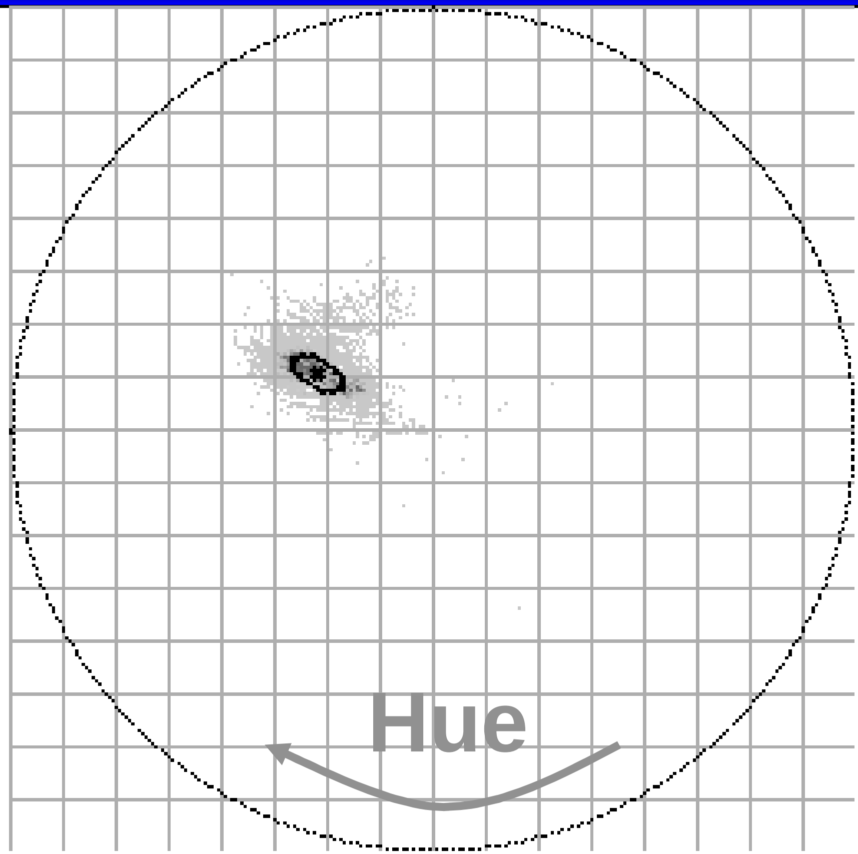
A tool for:
Finding modes in a set of data samples, manifesting an underlying probability distribution

The distribution could be in:

- **Color space**
- **Image space**
- **Scale space**
- **Actually any feature space you can conceive**
- **…**

Estimate histogram (model)

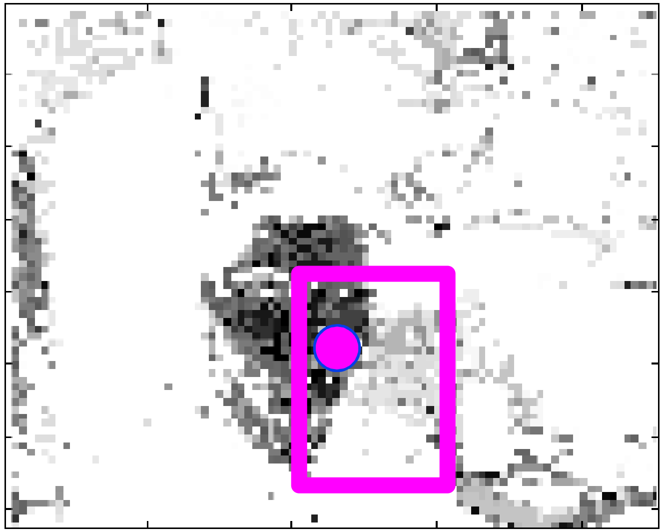




Mean-Shift Tracking using Colour

• Given each new video frame, compute a likelihood image from it by "looking up" each pixel in the histogram.
• Pixels form uniform grid of data points, each with a weight (pixel value) proportional to likelihood that the pixel is on the object we want to track.

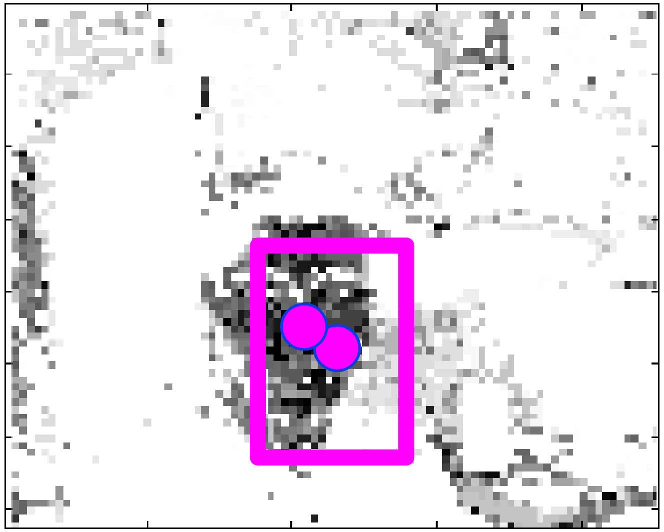
• Perform mean-shift using this weighted set of points.



Mean-Shift Tracking using Colour

• Given each new video frame, compute a likelihood image from it by “looking up” each pixel in the histogram.
• Pixels form uniform grid of data points, each with a weight (pixel value) proportional to likelihood that the pixel is on the object we want to track.

•Perform mean-shift using this weighted set of points.



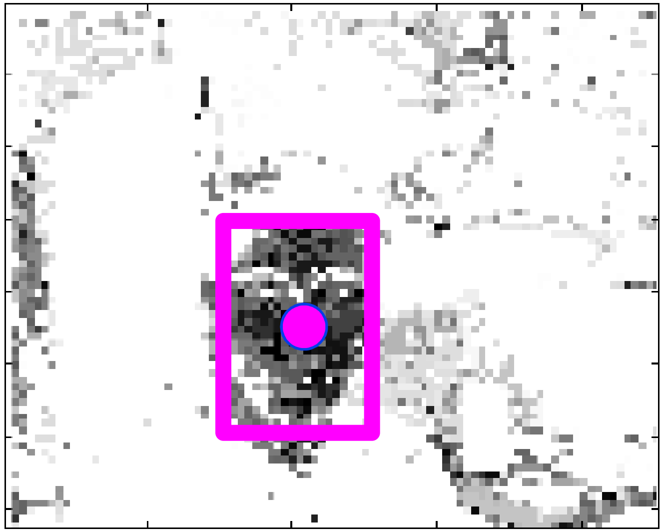
36

Mean-Shift Tracking using Colour

• Given each new video frame, compute a likelihood image from it by "looking up" each pixel in the histogram.

• Pixels form uniform grid of data points, each with a weight (pixel value) proportional to likelihood that the pixel is on the object we want to track.

•Perform mean-shift using this weighted set of points.



37

Mean-Shift Tracking using Colour

- This is fast and can work well when histogram is unimodal (i.e. object is “one” colour)
- Alternative is to build a histogram from search window, compute similarity function of image and model histograms, and perform mean-shift search to maximise this function.
- Can also adapt scale (by searching nearby scales).
- Can also use other features (e.g. texture)

Further reading:

Collins, R. “Mean-shift Blob Tracking through Scale Space”
IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2003

38

Tracking

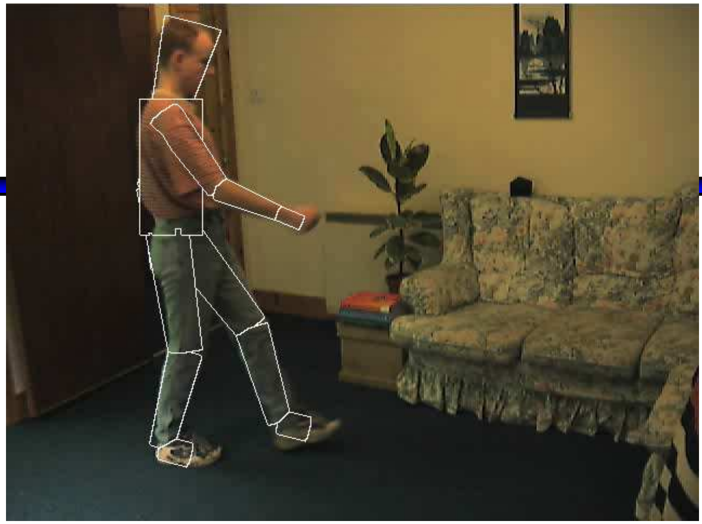
CAMShift in Matlab Computer Vision System Toolbox:

vision.HistogramBasedTracker System object

CAMShift uses a heuristic method to deal with scale.

G. Bradski, “Computer vision face tracking for use in a perceptual user interface” *Intel Technology Journal*, 2nd Quarter, 1998.

Further reading:
Y. Raja, S. J. McKenna, and S. Gong.
Tracking and segmenting people in varying lighting conditions using colour.
IEEE International Conference on Face & Gesture Recognition, 228-233, 1998.




MATRIX