



# Part 1: Bag-of-words models

by Li Fei-Fei (UIUC)

**Object**



**Bag of  
'words'**



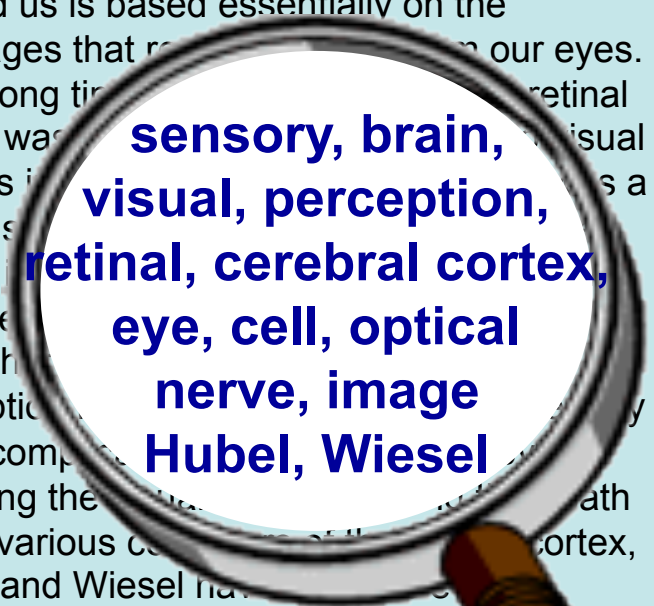
# Analogy to documents

Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based essentially on the messages that reach our eyes.

For a long time, the retinal image was considered as a simple picture. As a result, the visual centers in the brain were thought of as a movie screen on which the retinal image is projected.

It was not until Hubel and Wiesel discovered that the visual system knows the difference between a perceptible edge and a perceptible spot that the visual system was found to be more complex than a simple movie screen.

Following the discovery of Hubel and Wiesel, it was found that the message about the image falling on the retina undergoes a complex analysis in a system of nerve cells stored in columns. In this system each cell has its specific function and is responsible for a specific detail in the pattern of the retinal image.



**sensory, brain,  
visual, perception,  
retinal, cerebral cortex,  
eye, cell, optical  
nerve, image  
Hubel, Wiesel**

China is forecasting a trade surplus of \$90bn (£51bn) to \$100bn this year, a threefold increase on 2004's \$32bn. The Commerce Ministry said the surplus would be created by a predicted 30% increase in exports to \$750bn, compared with \$580bn in 2004.

The increase in exports will be partly due to a 30% increase in exports to \$750bn, compared with \$580bn in 2004. This will be partly due to a 30% increase in exports to \$750bn, compared with \$580bn in 2004.

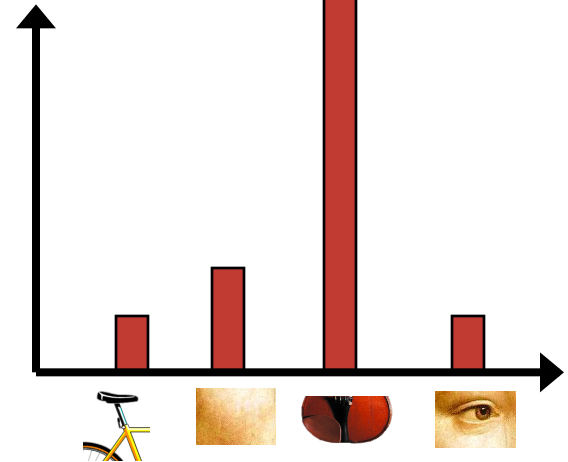
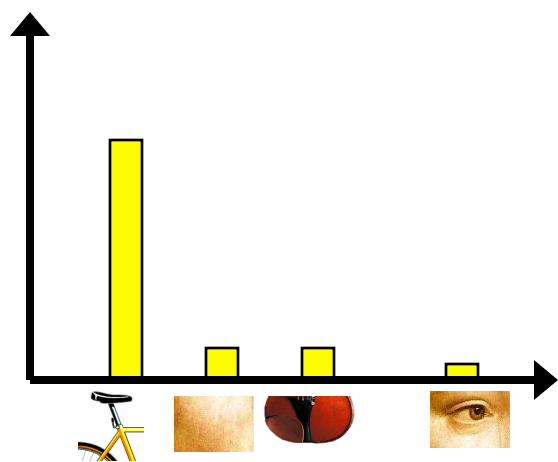
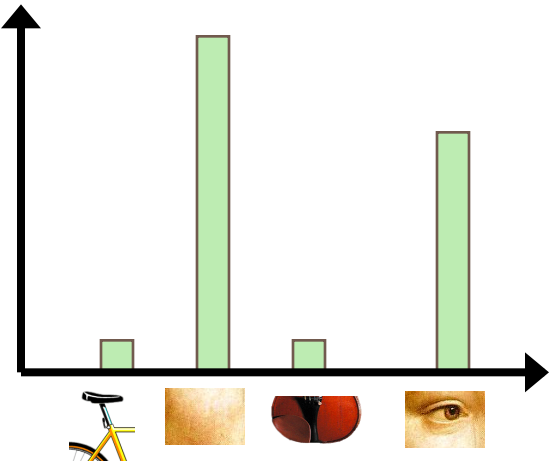
The increase in exports will be partly due to a 30% increase in exports to \$750bn, compared with \$580bn in 2004. This will be partly due to a 30% increase in exports to \$750bn, compared with \$580bn in 2004.

The increase in exports will be partly due to a 30% increase in exports to \$750bn, compared with \$580bn in 2004. This will be partly due to a 30% increase in exports to \$750bn, compared with \$580bn in 2004.

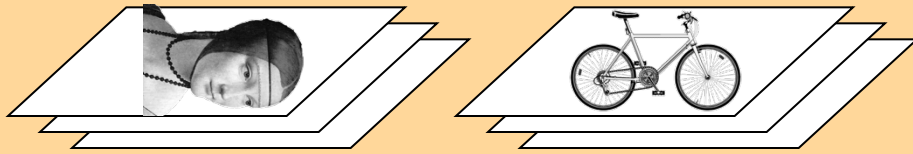


**China, trade,  
surplus, commerce,  
exports, imports, US,  
yuan, bank, domestic,  
foreign, increase,  
trade, value**

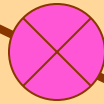
China's trade surplus is expected to reach \$100bn this year, a threefold increase on 2004's \$32bn. The Commerce Ministry said the surplus would be created by a predicted 30% increase in exports to \$750bn, compared with \$580bn in 2004. This will be partly due to a 30% increase in exports to \$750bn, compared with \$580bn in 2004.



# Representation



1. feature detection & representation



2. codewords dictionary

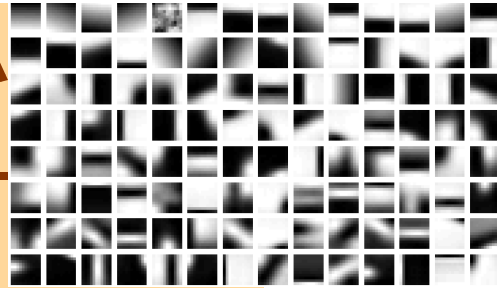
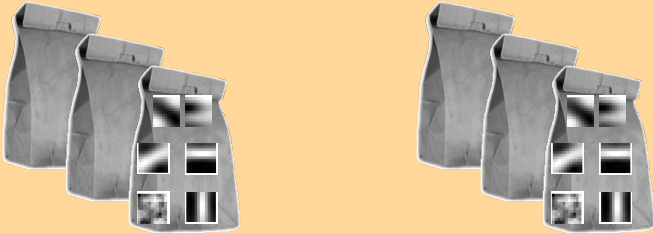
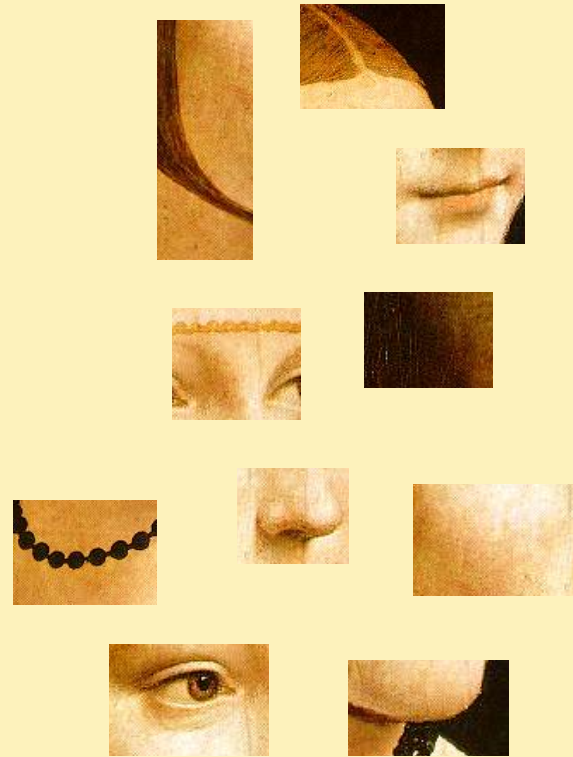


image representation

3.



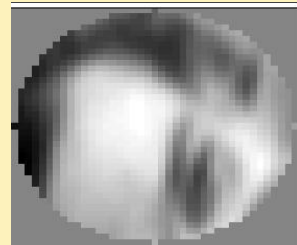
# 1. Feature detection and representation



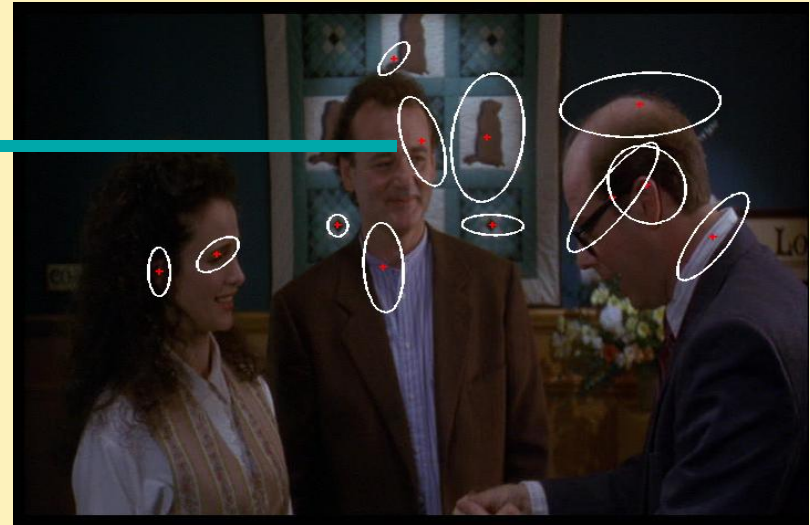
# 1. Feature detection and representation



**Compute  
SIFT  
descriptor**  
[Lowe'99]



**Normalize  
patch**



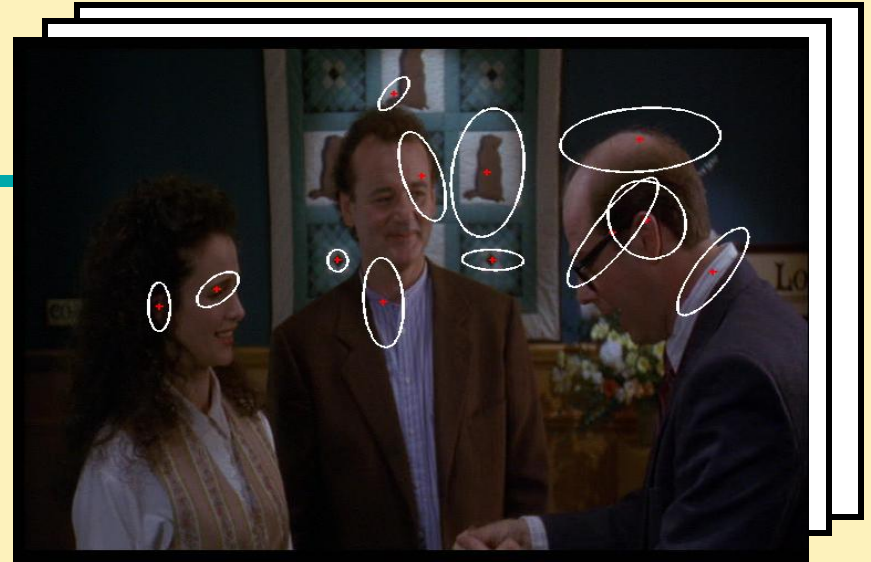
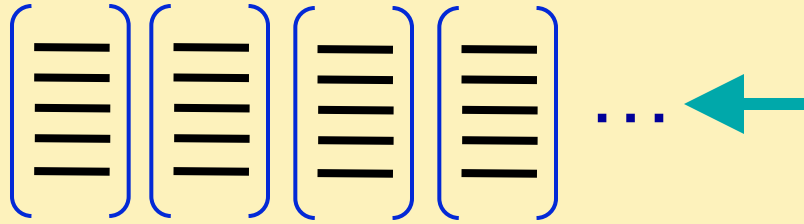
**Detect patches**

[Mikojaczyk and Schmid '02]

[Matas et al. '02]

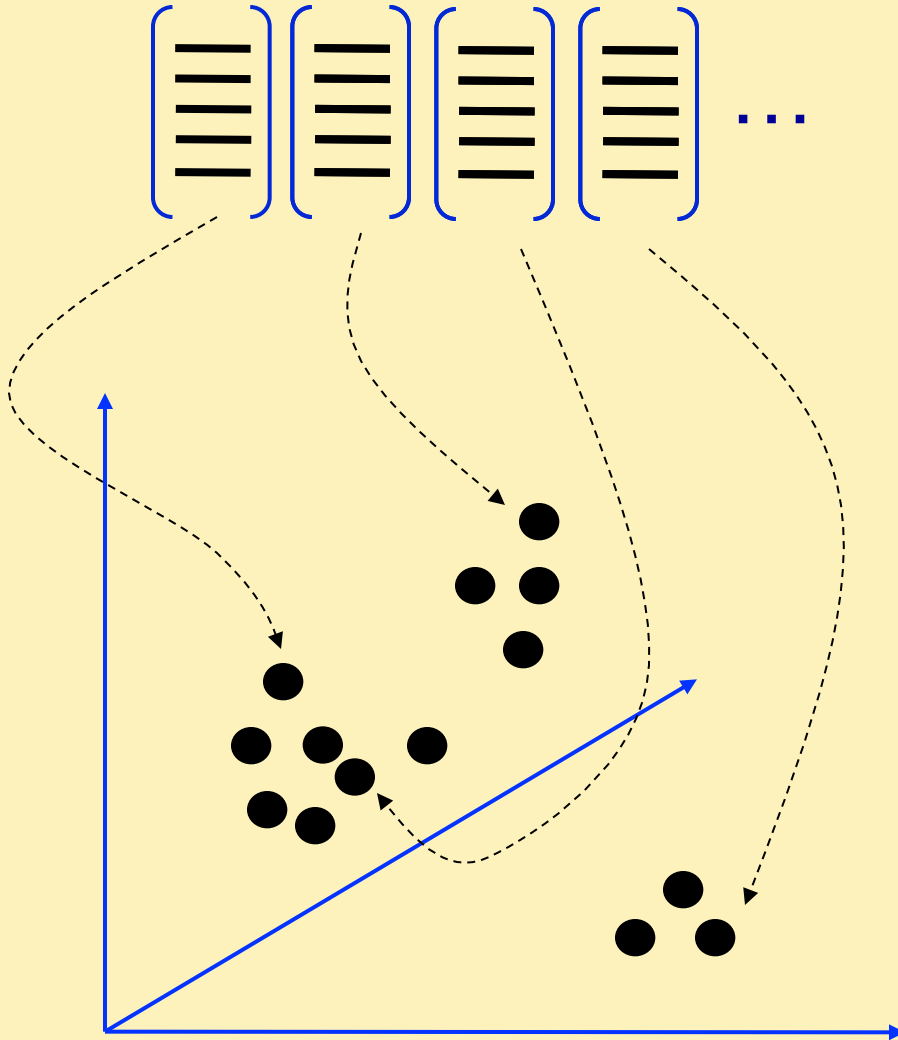
[Sivic et al. '03]

# 1. Feature detection and representation

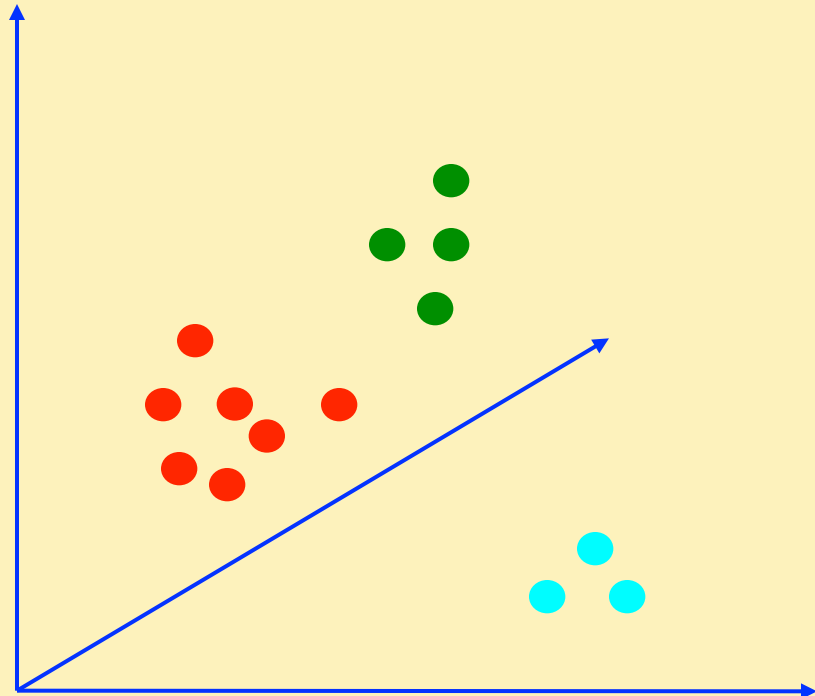
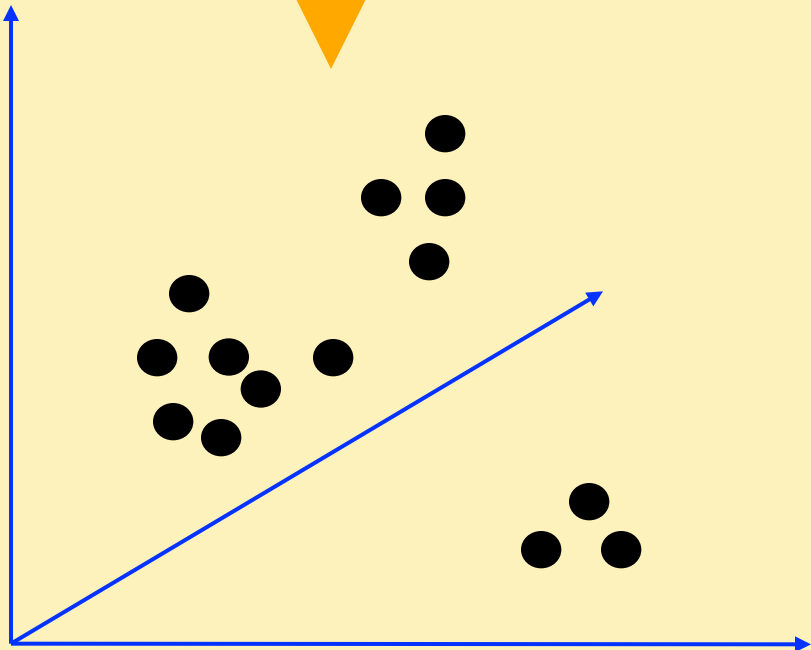
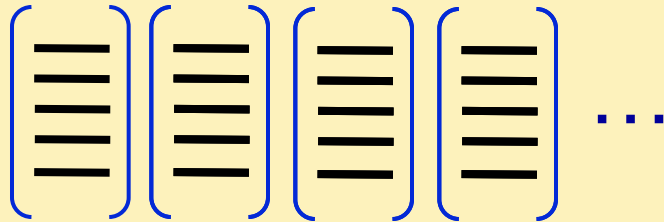




## 2. Codewords dictionary formation

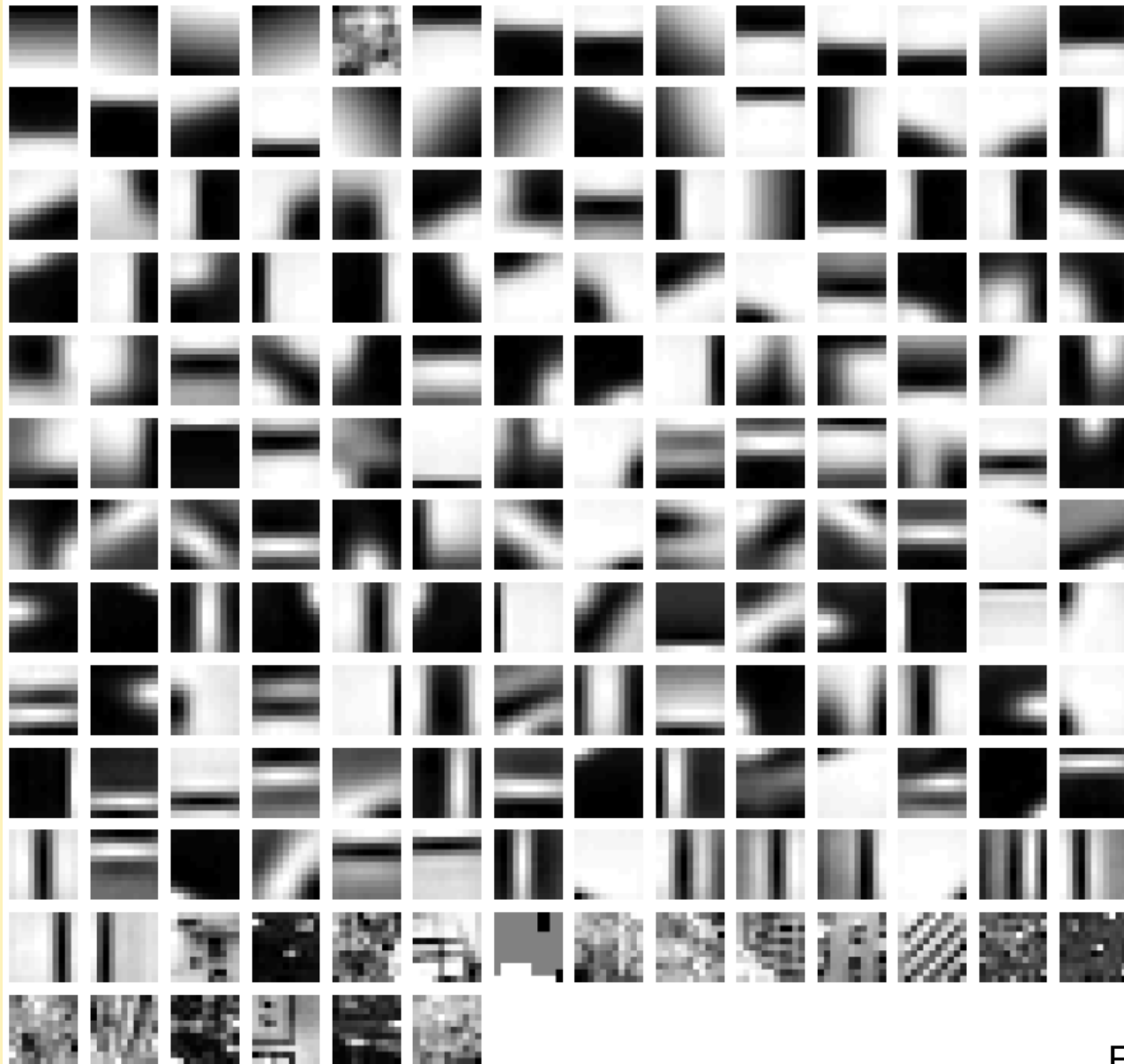


## 2. Codewords dictionary formation



Vector quantization

## 2. Codewords dictionary formation



# What Is a Good Clustering?

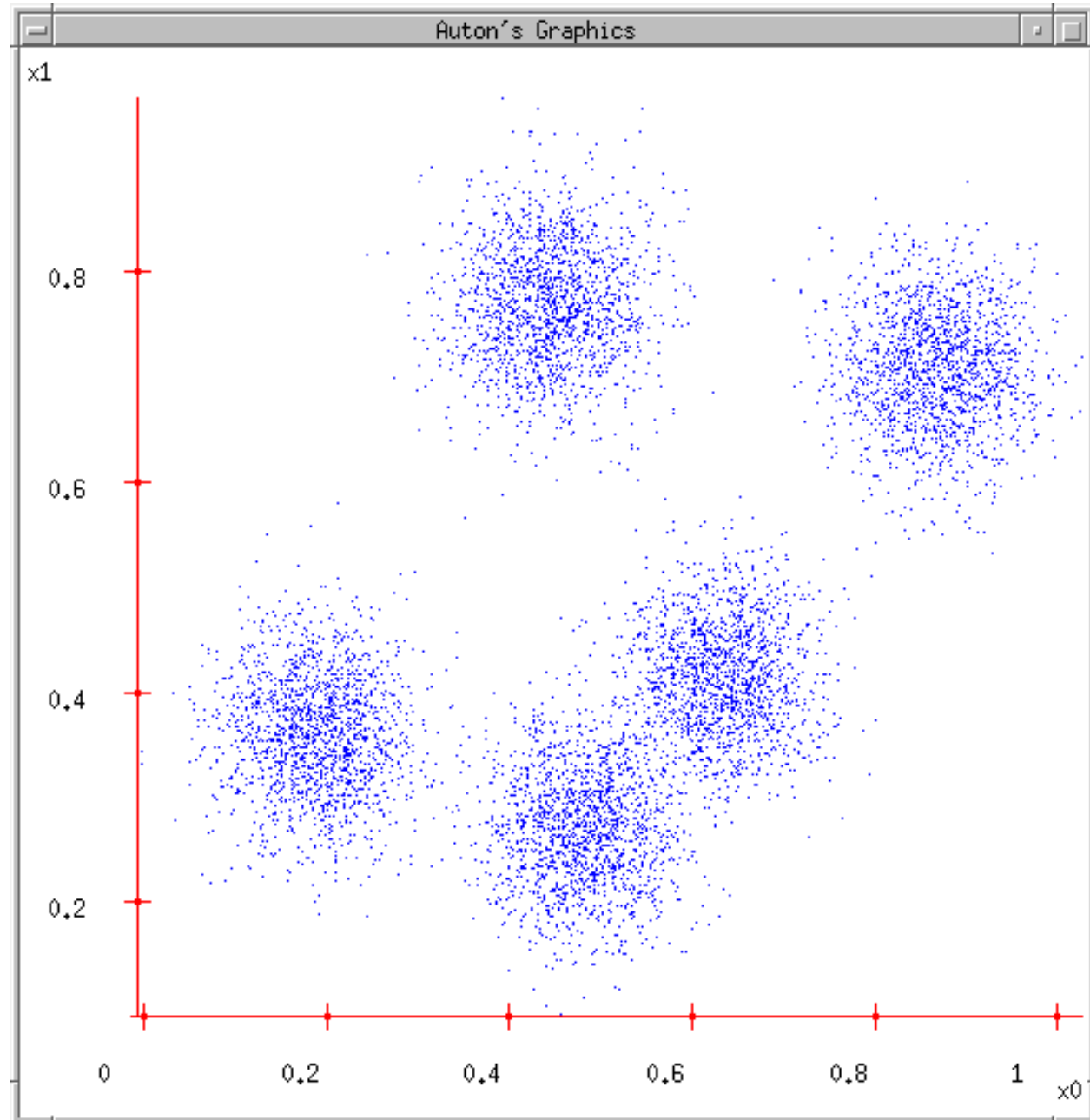
- A good clustering method will produce clusters with
  - High intra-class similarity
  - Low inter-class similarity
- Precise definition of clustering quality is difficult
  - Application-dependent
  - Ultimately subjective

# *K-Means* Clustering

- Given  $k$ , the *k-means* algorithm consists of four steps:
  - Select initial centroids at random.
  - Assign each object to the cluster with the nearest centroid.
  - Compute each centroid as the mean of the objects assigned to it.
  - Repeat previous 2 steps until no change.

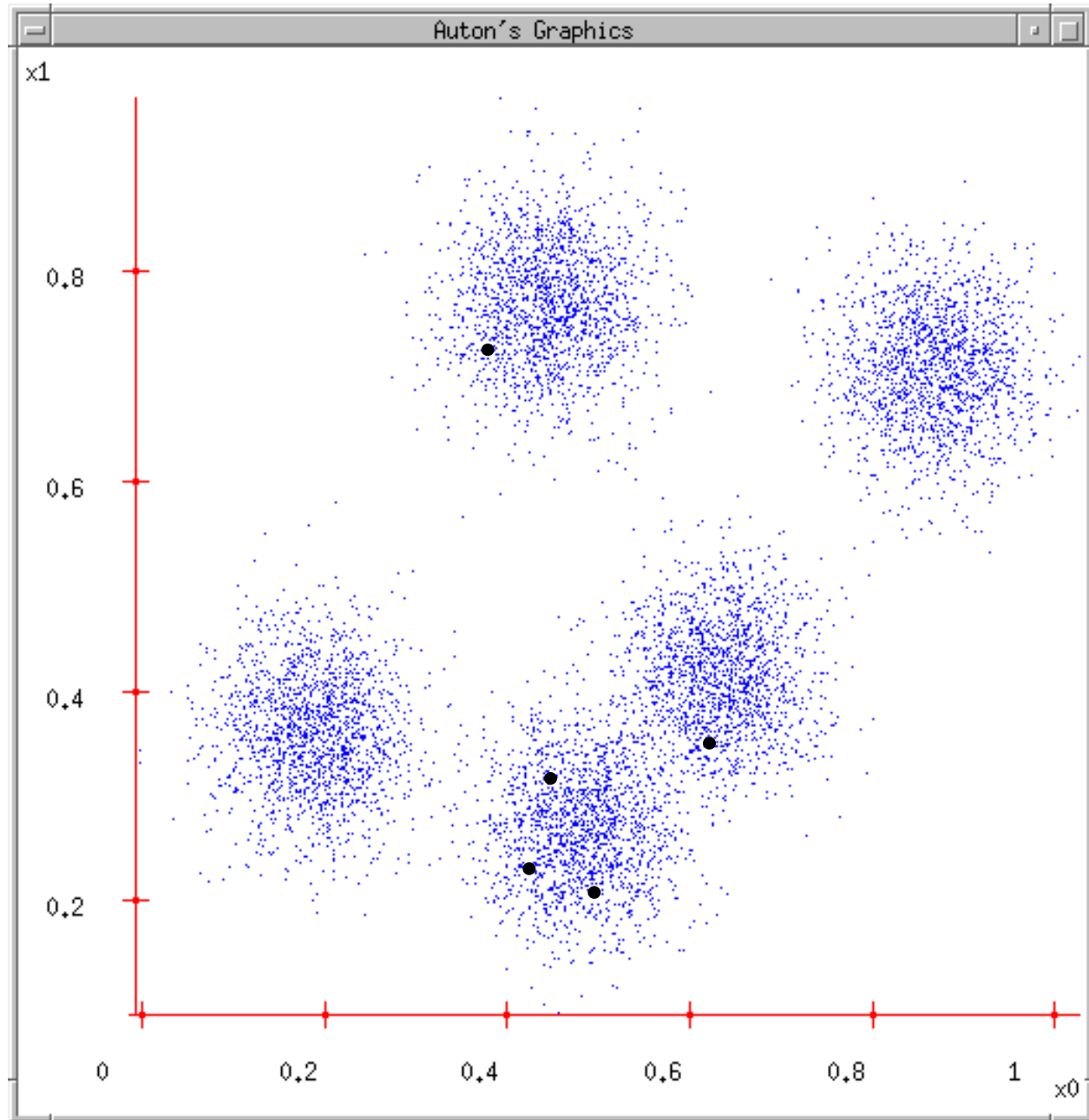
# K-means

1. Ask user how many clusters they'd like.  
(e.g.  $k=5$ )



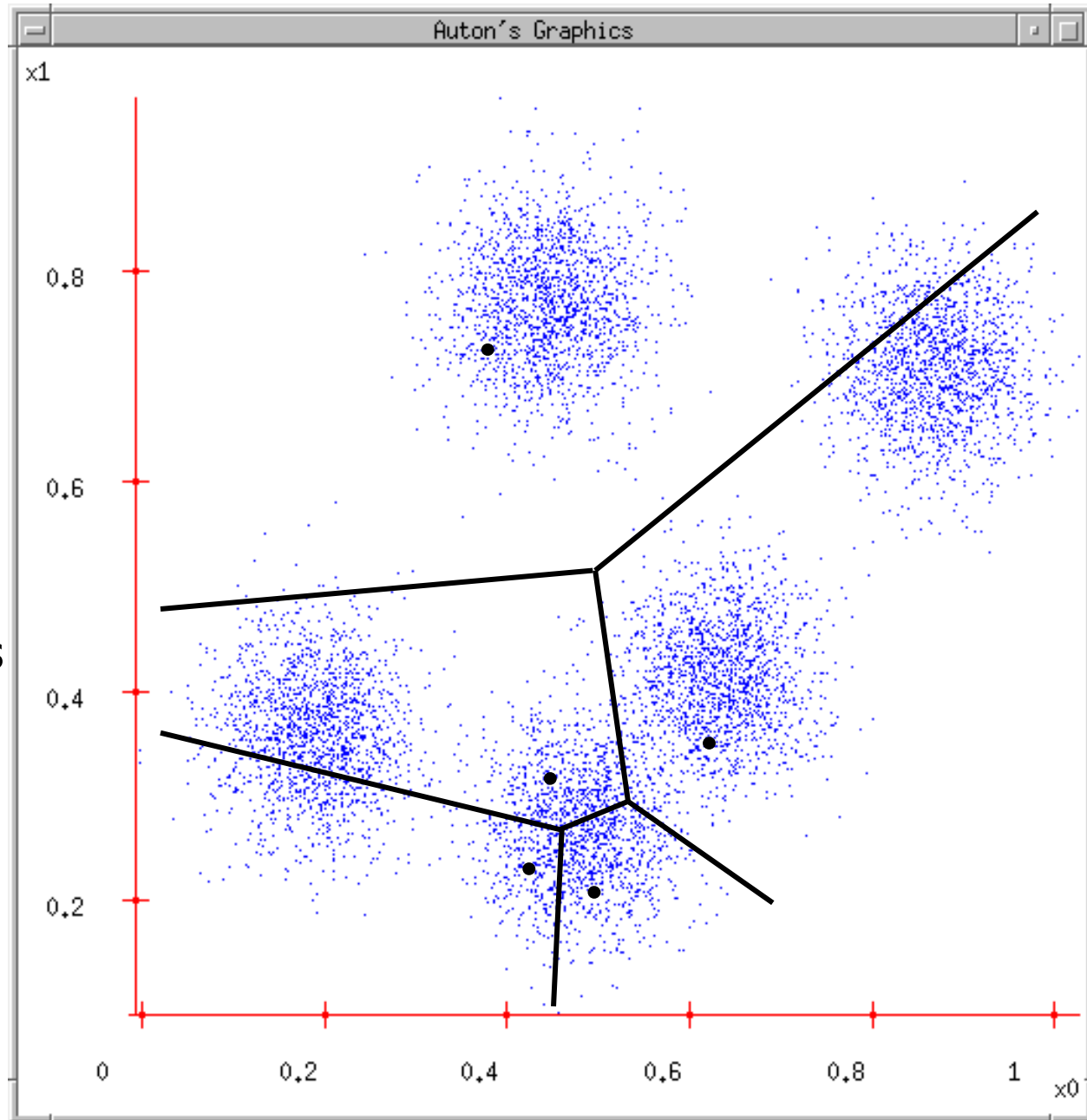
# K-means

1. Ask user how many clusters they'd like.  
(*e.g. k=5*)
2. Randomly guess k cluster Center locations



# K-means

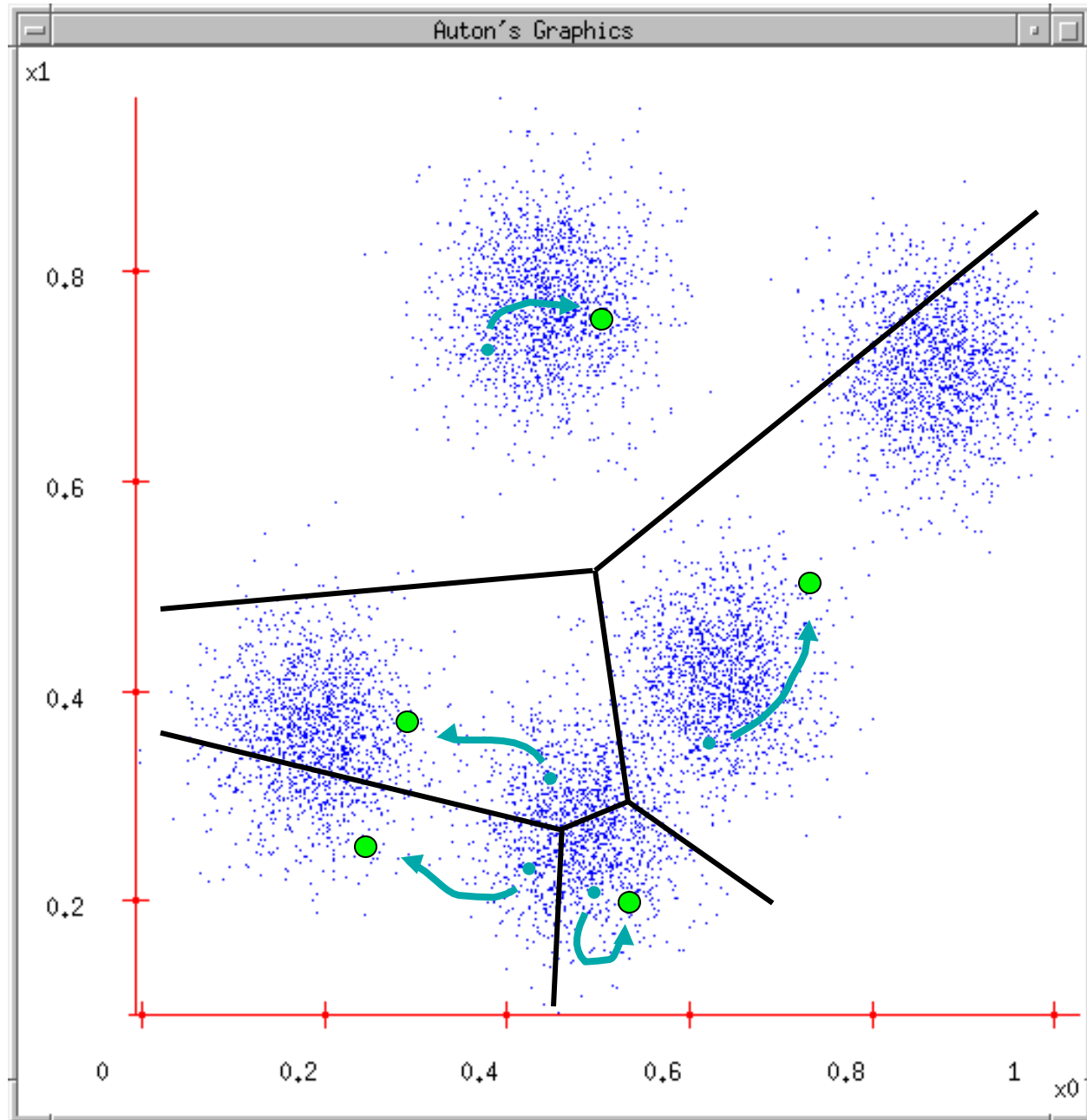
1. Ask user how many clusters they'd like.  
(e.g.  $k=5$ )
2. Randomly guess  $k$  cluster Center locations
3. Each datapoint finds out which Center it's closest to. (Thus each Center "owns" a set of datapoints)





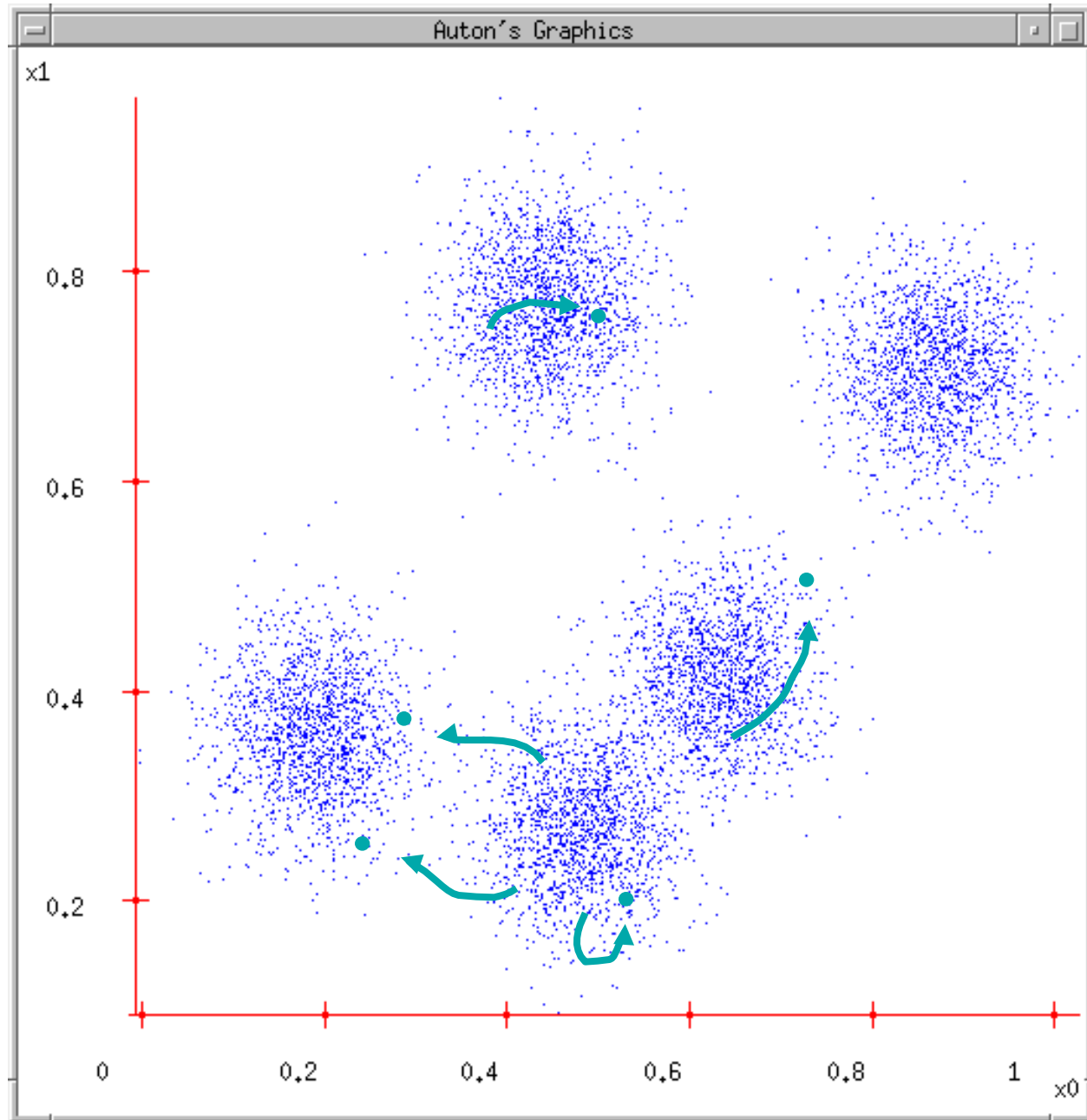
# K-means

1. Ask user how many clusters they'd like.  
(e.g.  $k=5$ )
2. Randomly guess  $k$  cluster Center locations
3. Each datapoint finds out which Center it's closest to.
4. Each Center finds the centroid of the points it owns



# K-means

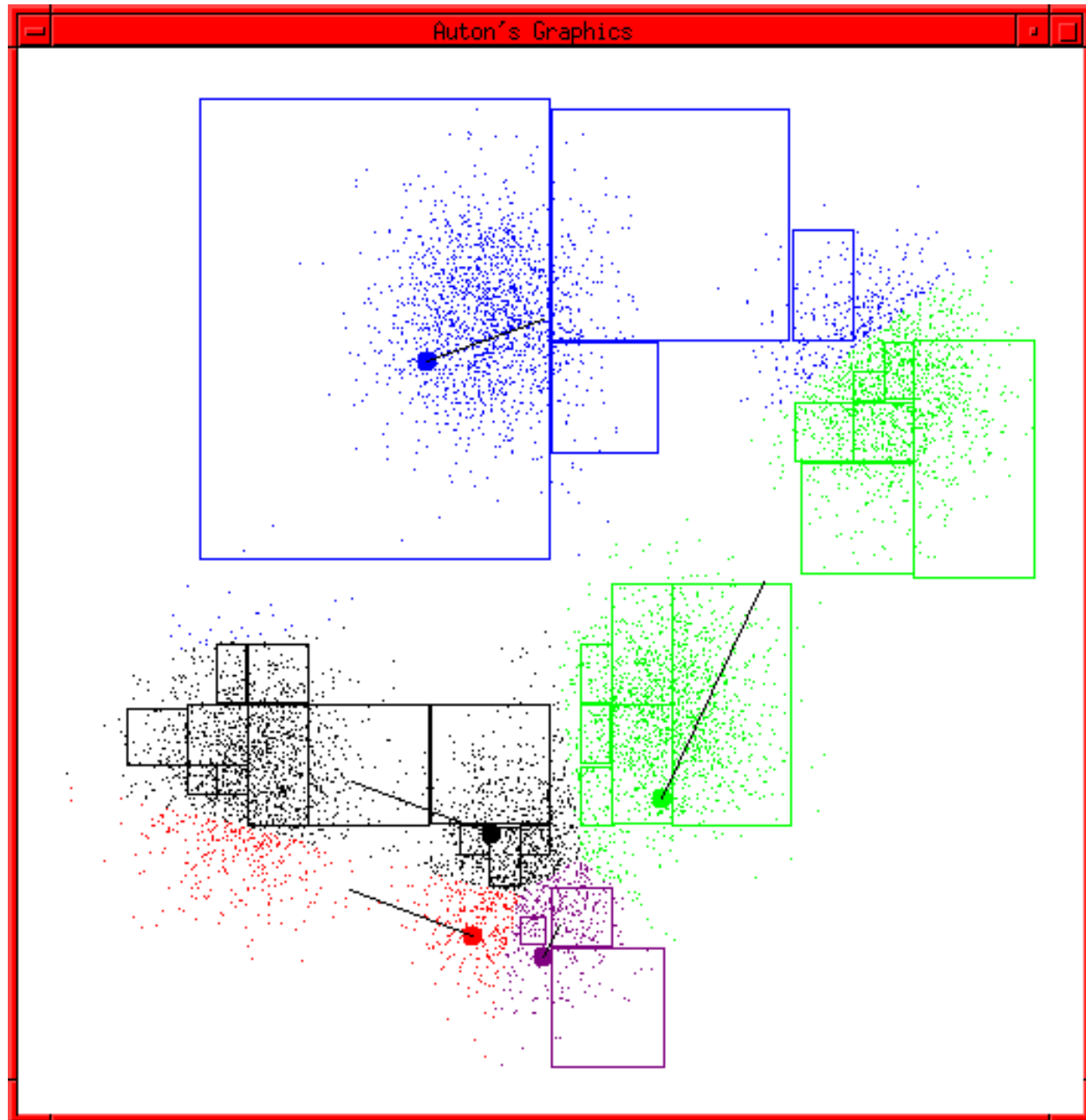
1. Ask user how many clusters they'd like.  
(e.g.  $k=5$ )
2. Randomly guess  $k$  cluster Center locations
3. Each datapoint finds out which Center it's closest to.
4. Each Center finds the centroid of the points it owns...
5. ...and jumps there
6. ...Repeat until terminated!



# K-means Start

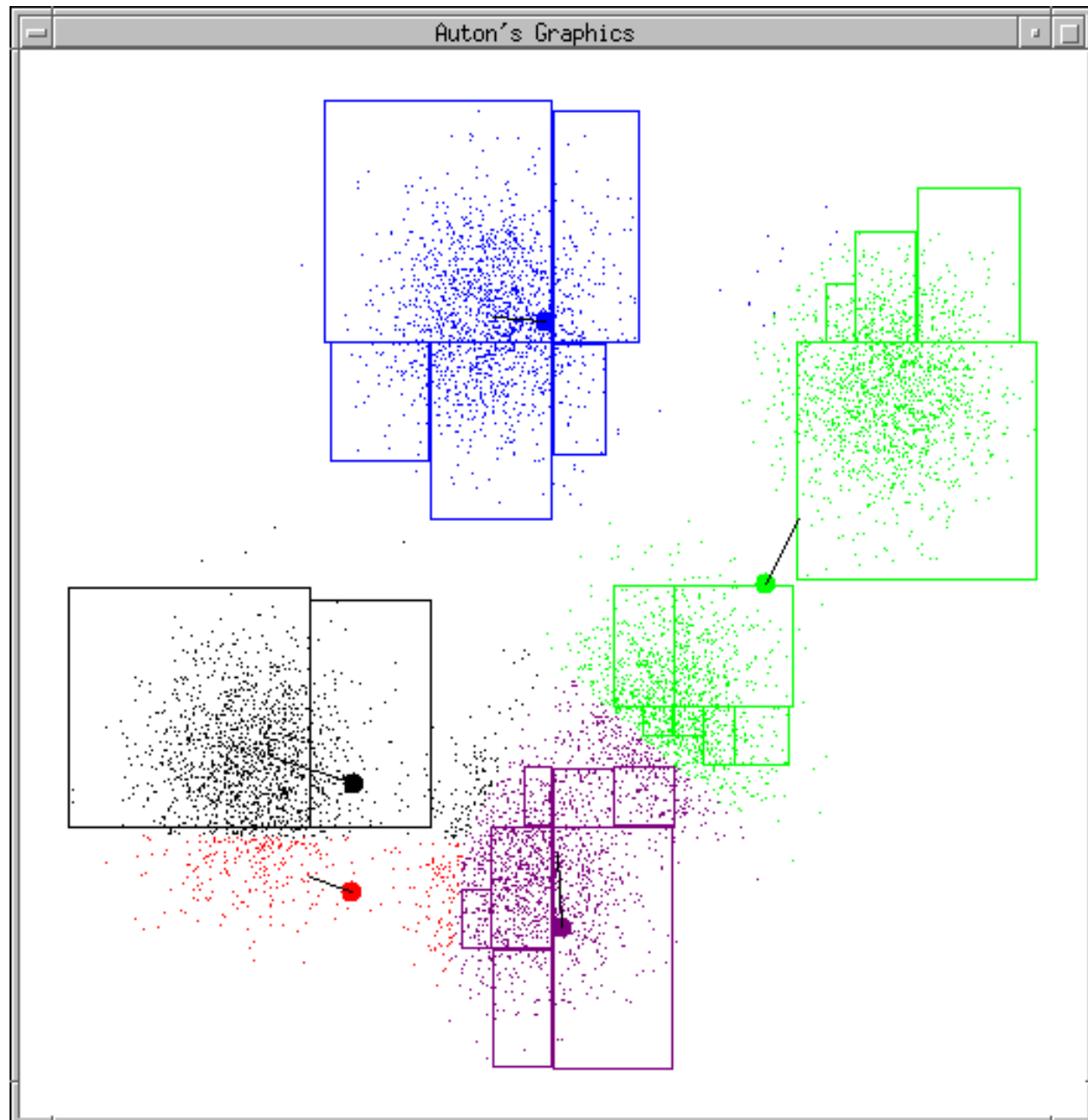
Example generated by  
Dan Pelleg's super-  
duper fast K-means  
system:

*Dan Pelleg and Andrew  
Moore. Accelerating Exact  
k-means Algorithms with  
Geometric Reasoning.  
Proc. Conference on  
Knowledge Discovery in  
Databases 1999, (KDD99)  
(available on  
[www.autonlab.org/pap.html](http://www.autonlab.org/pap.html))*



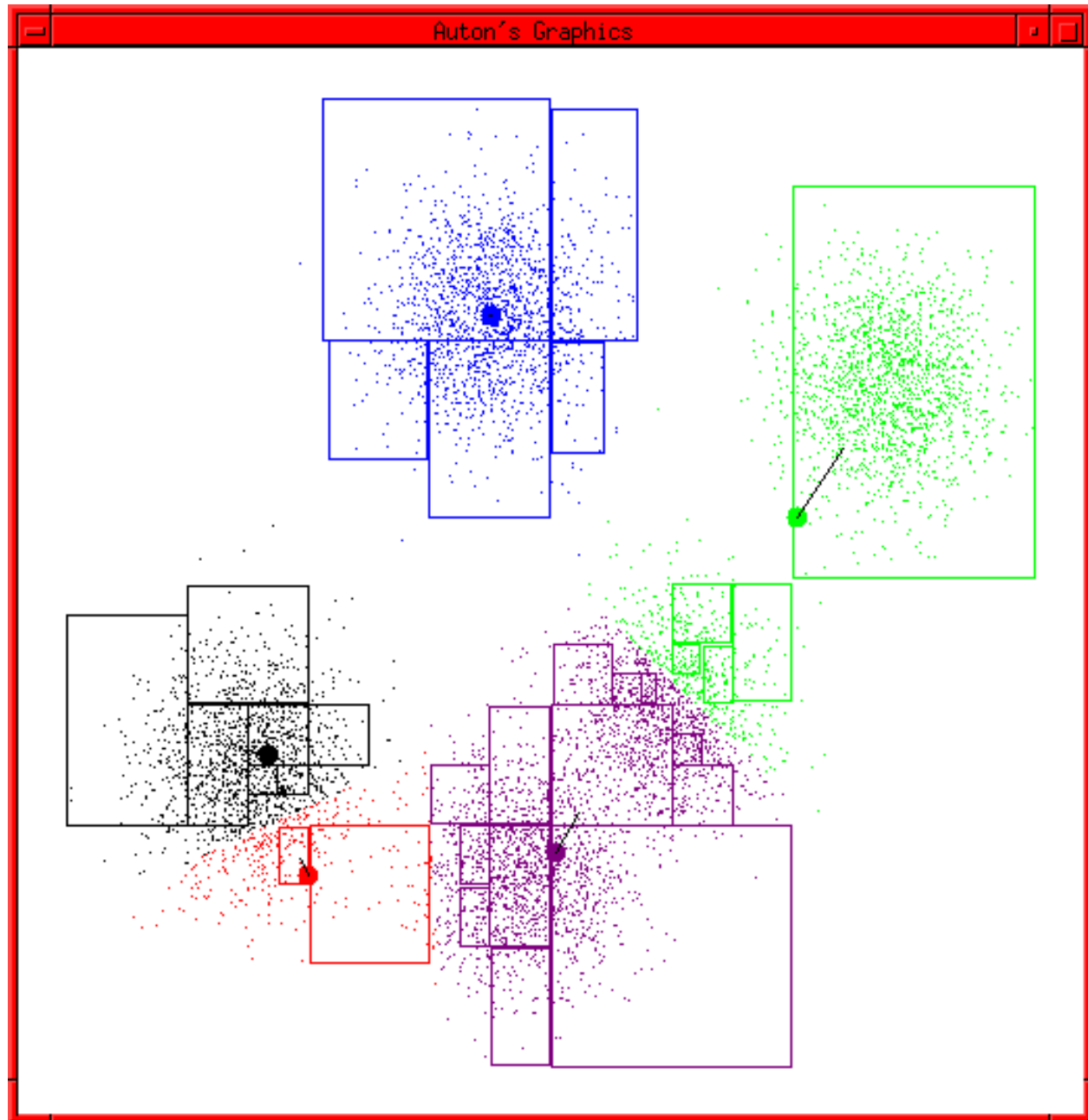
# K-means continues

...



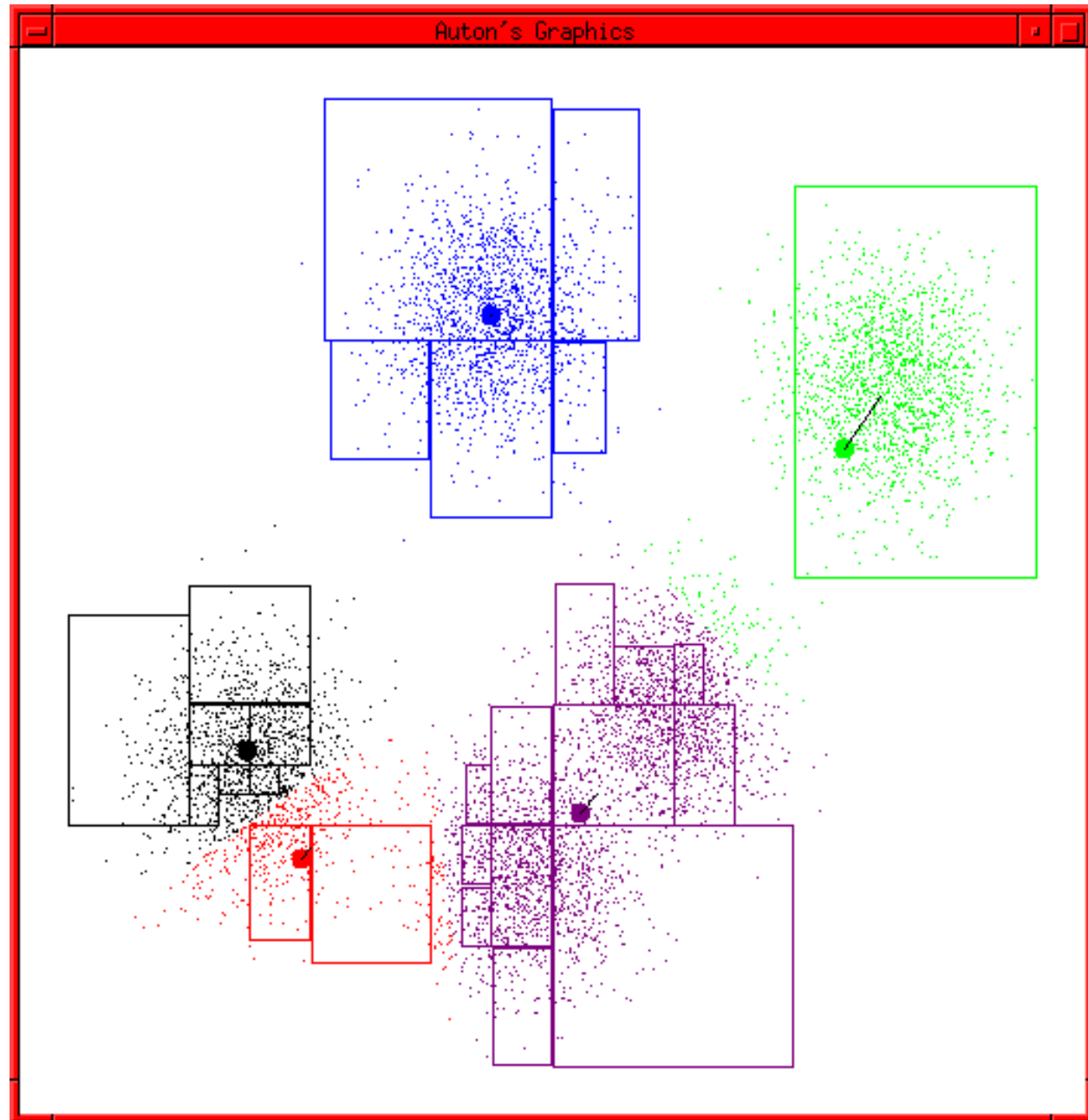
# K-means continues

...



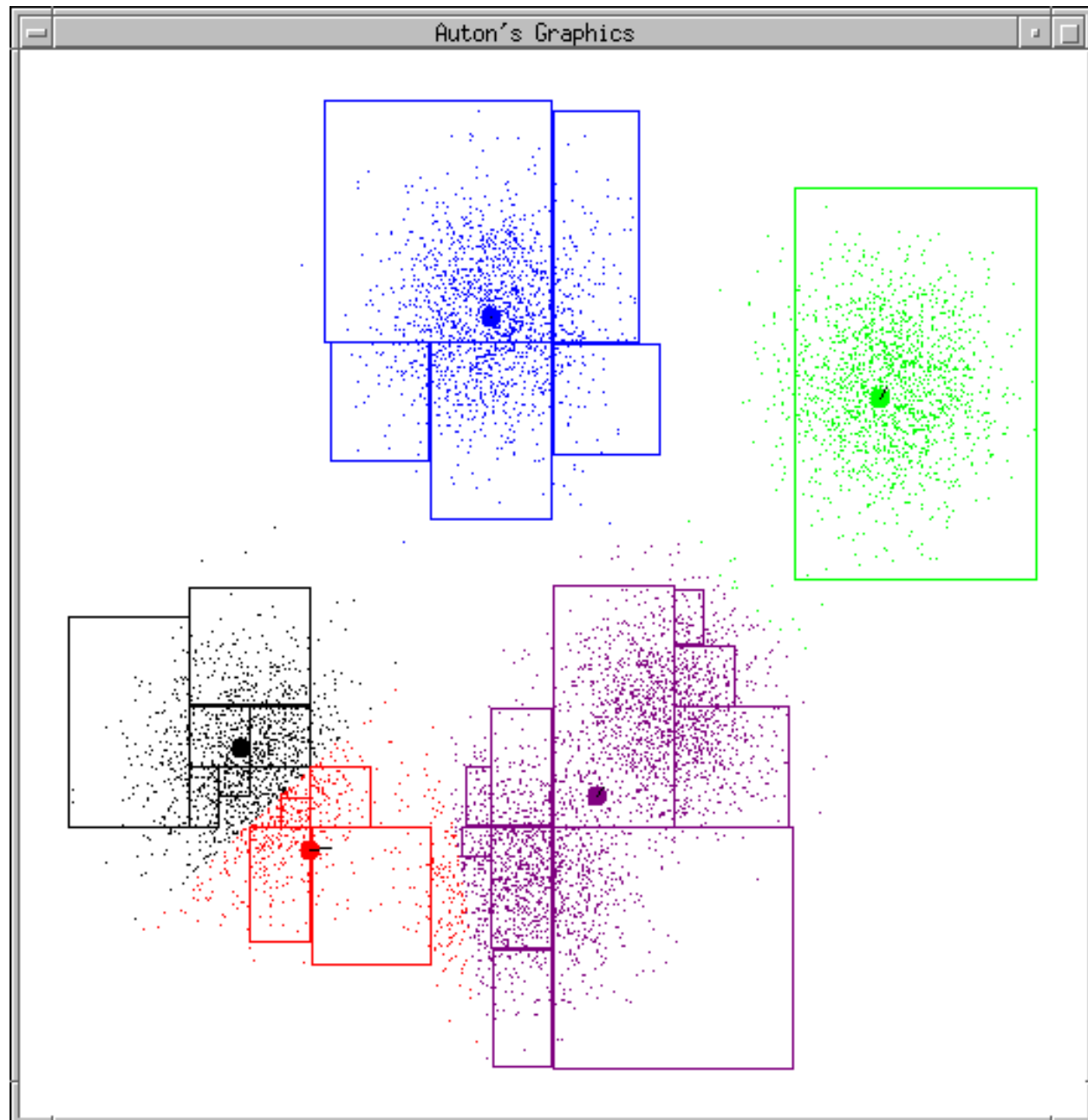
# K-means continues

...



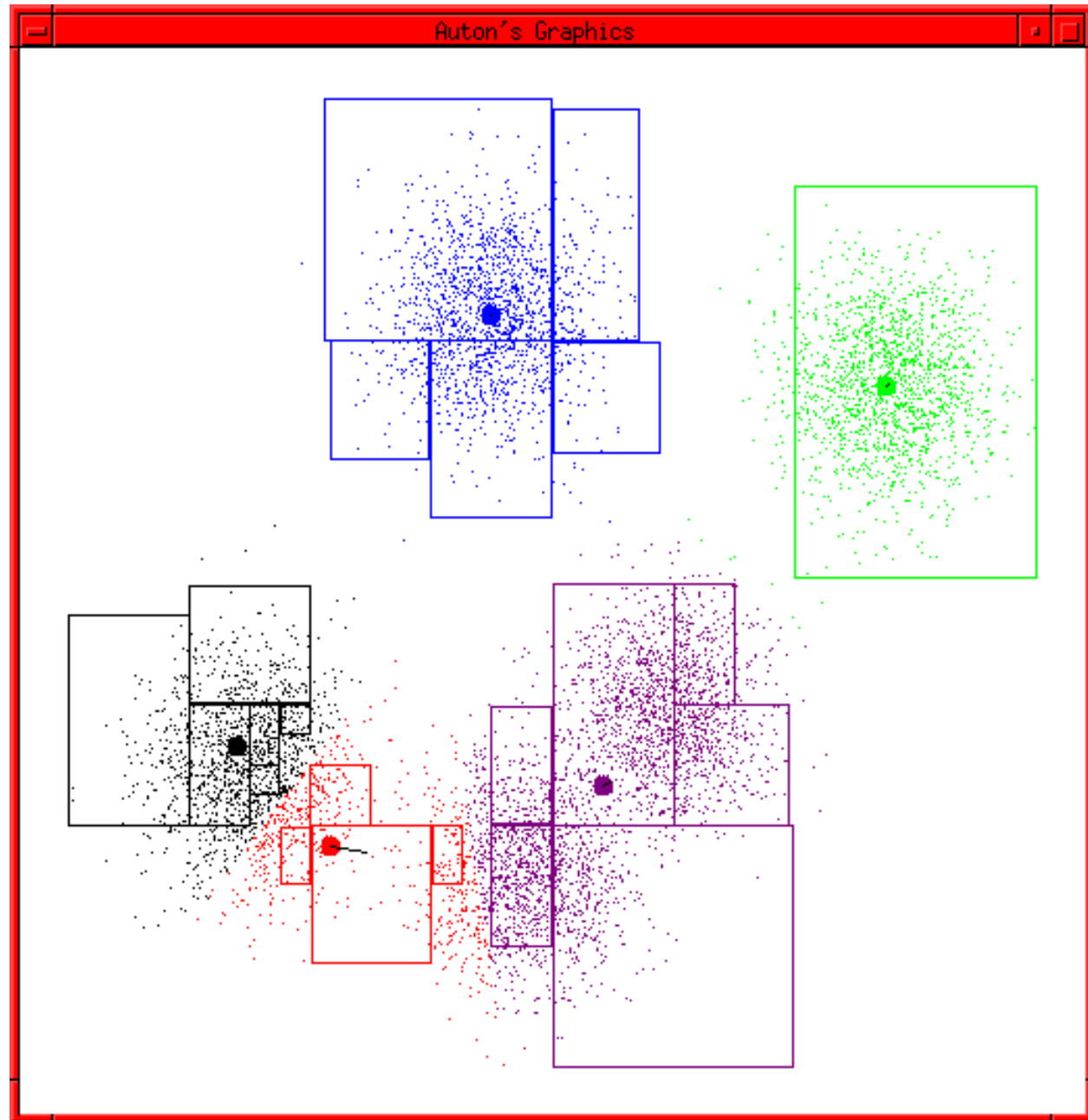
# K-means continues

...



# K-means continues

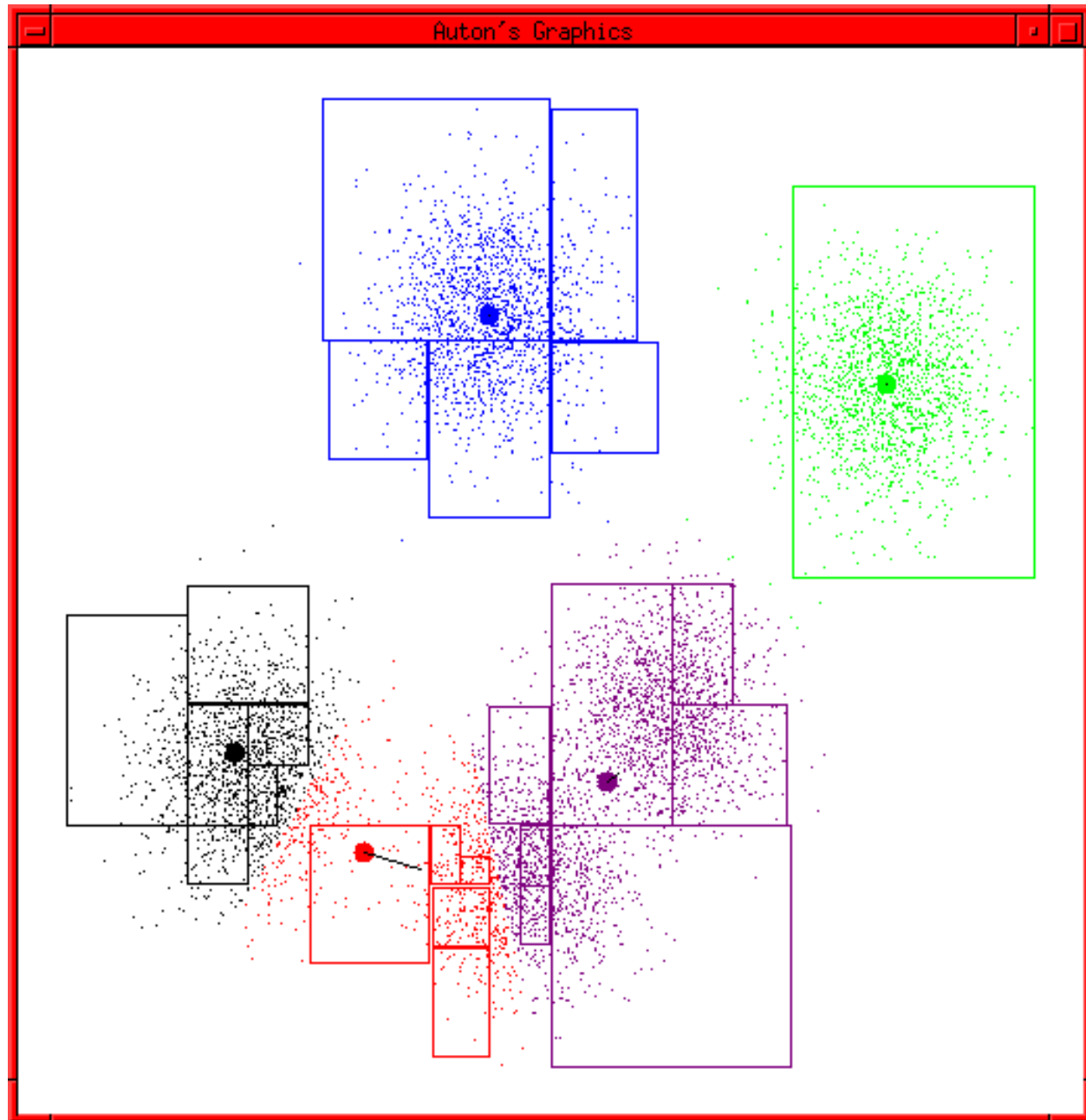
...





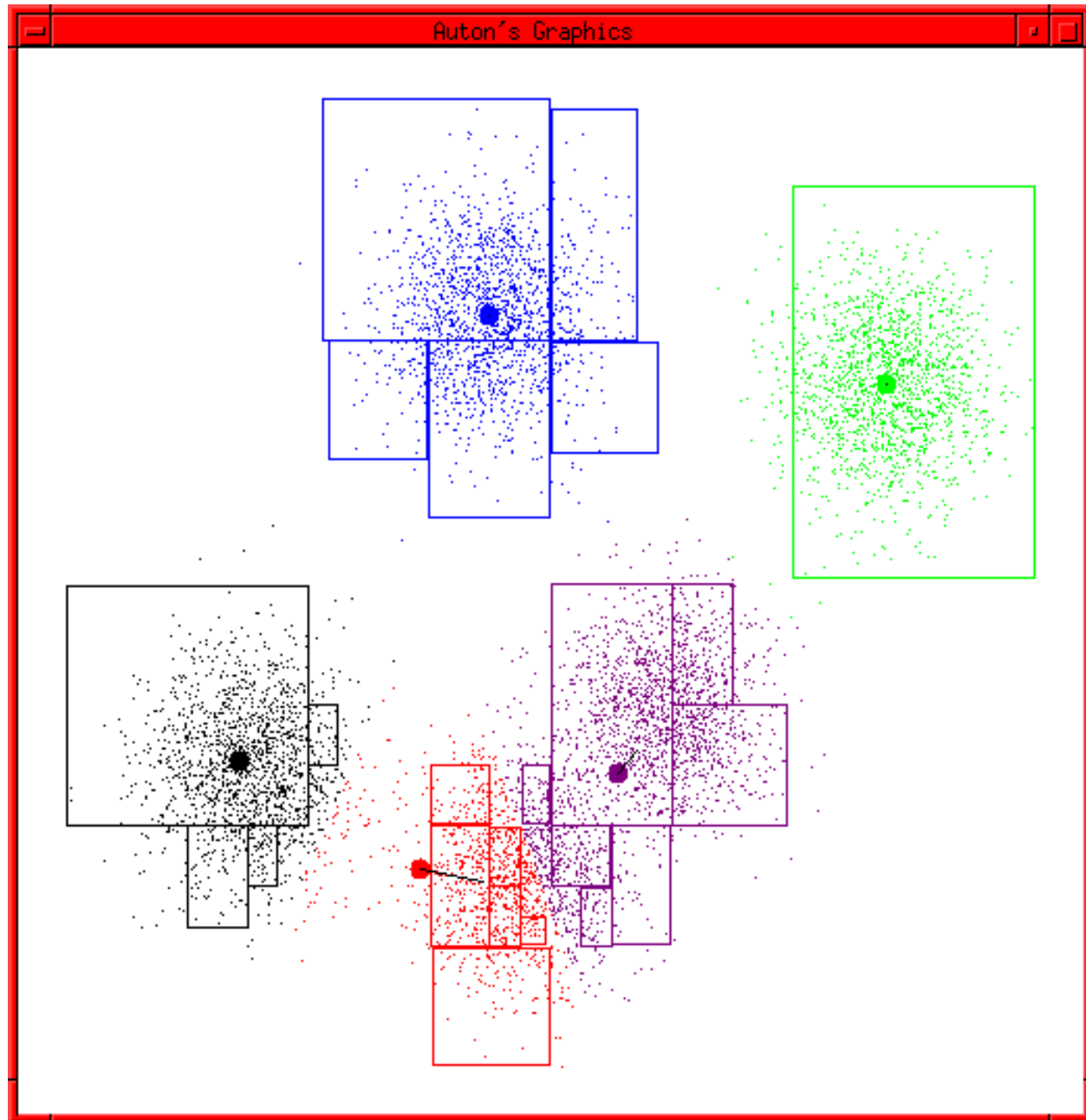
# K-means continues

...



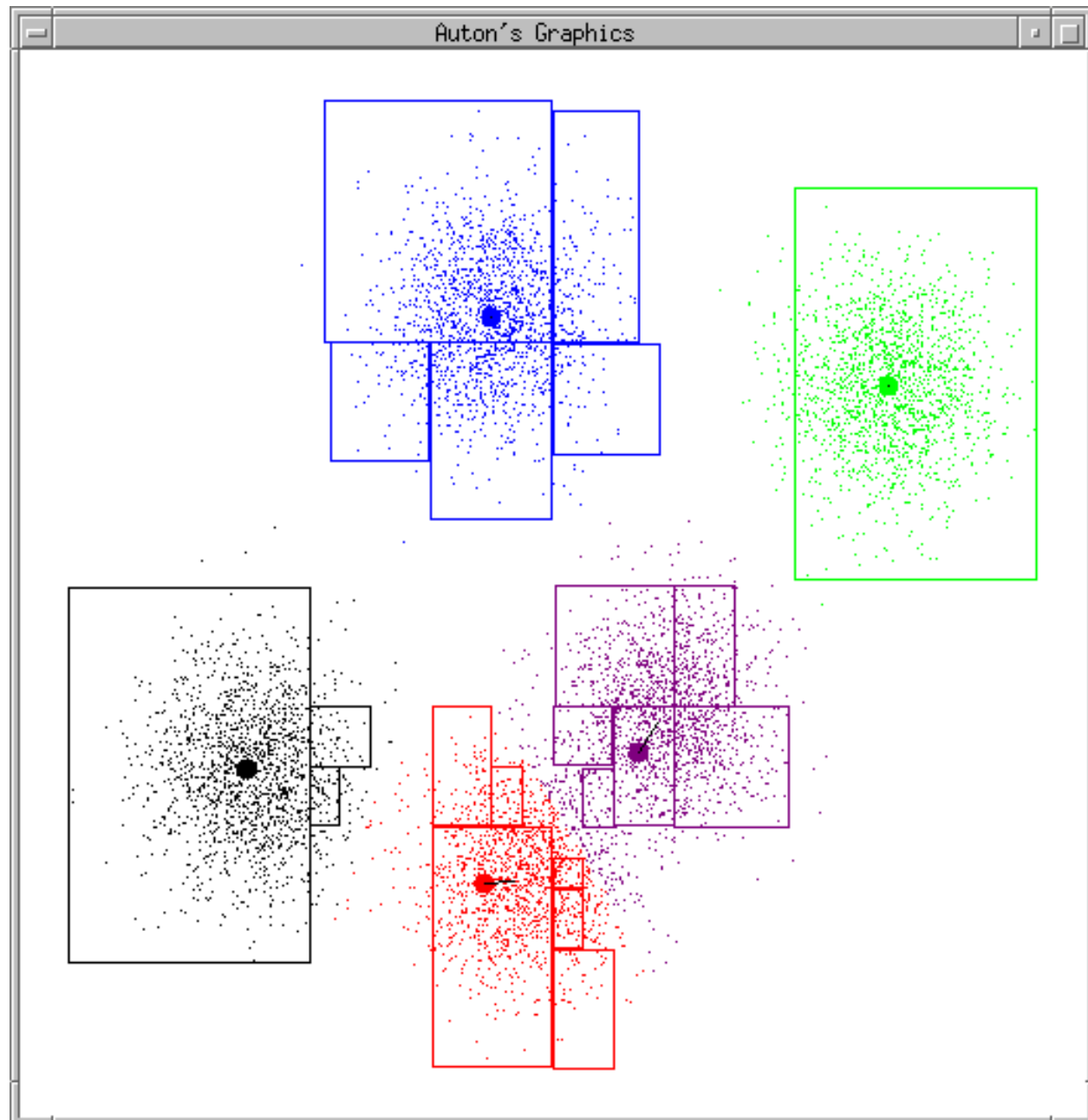
# K-means continues

...

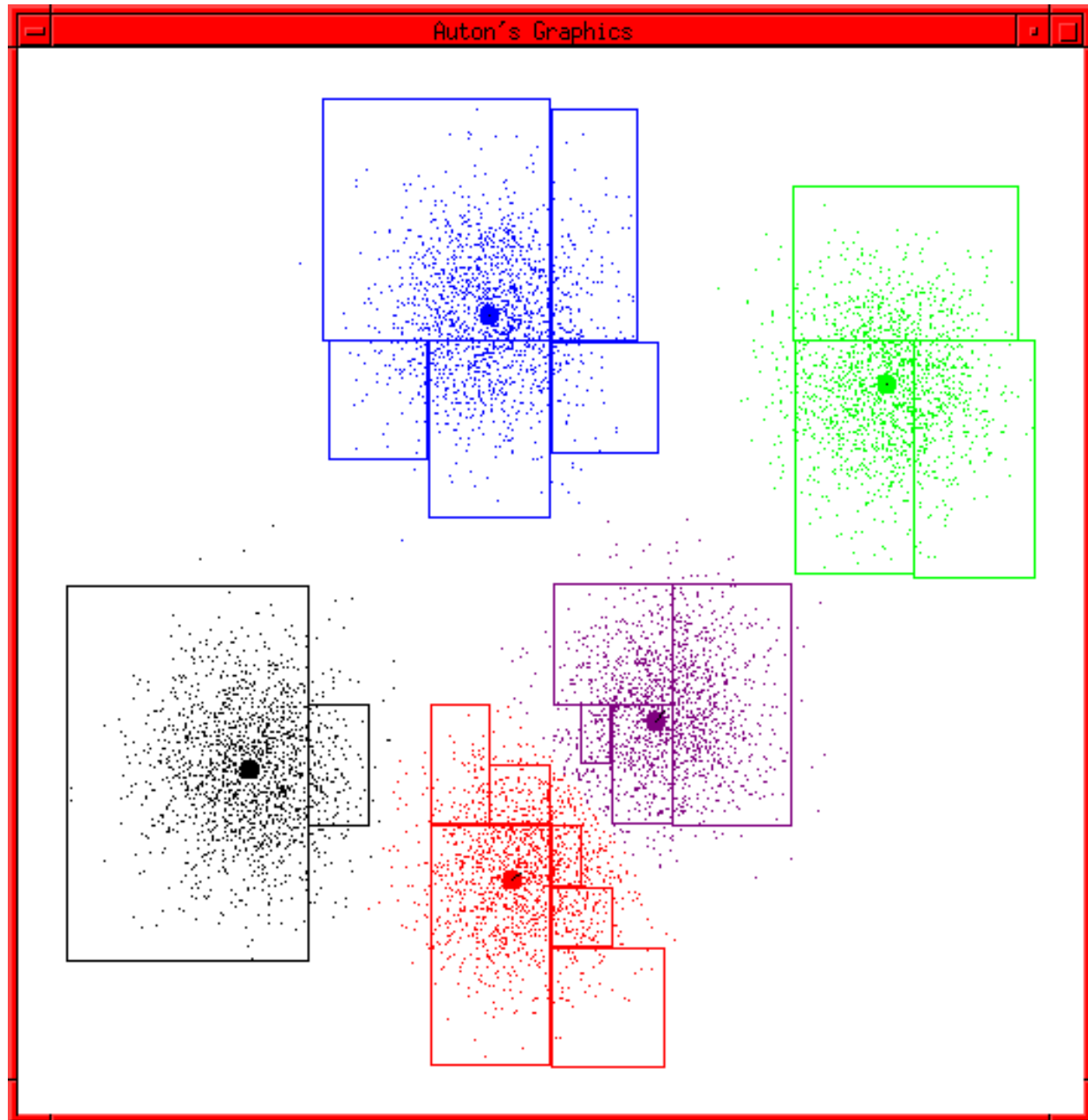


# K-means continues

...



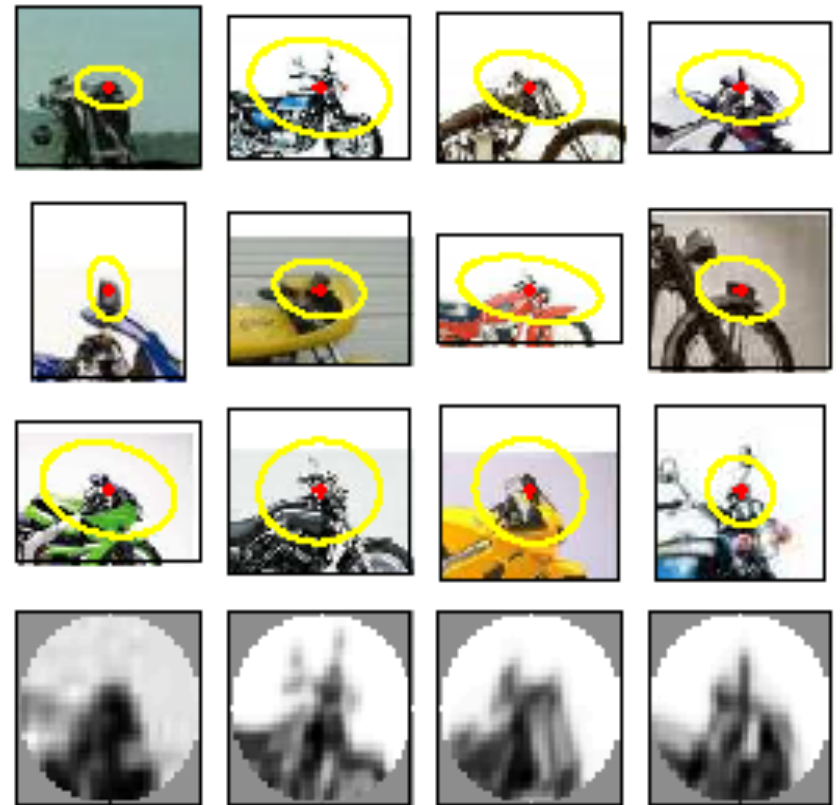
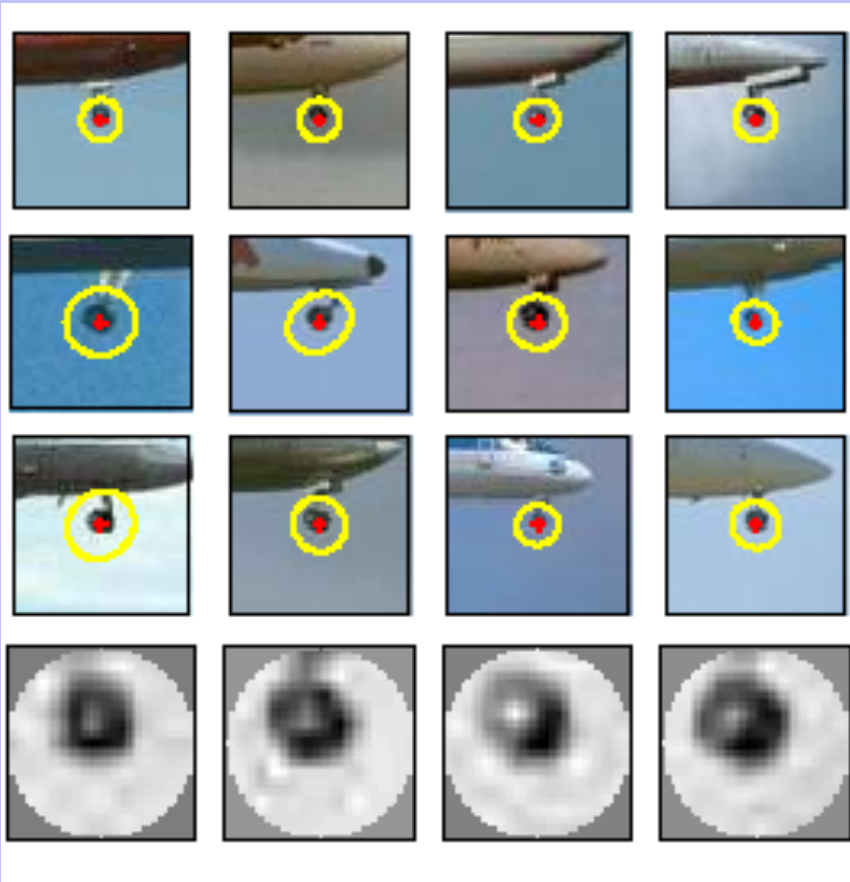
# K-means terminates



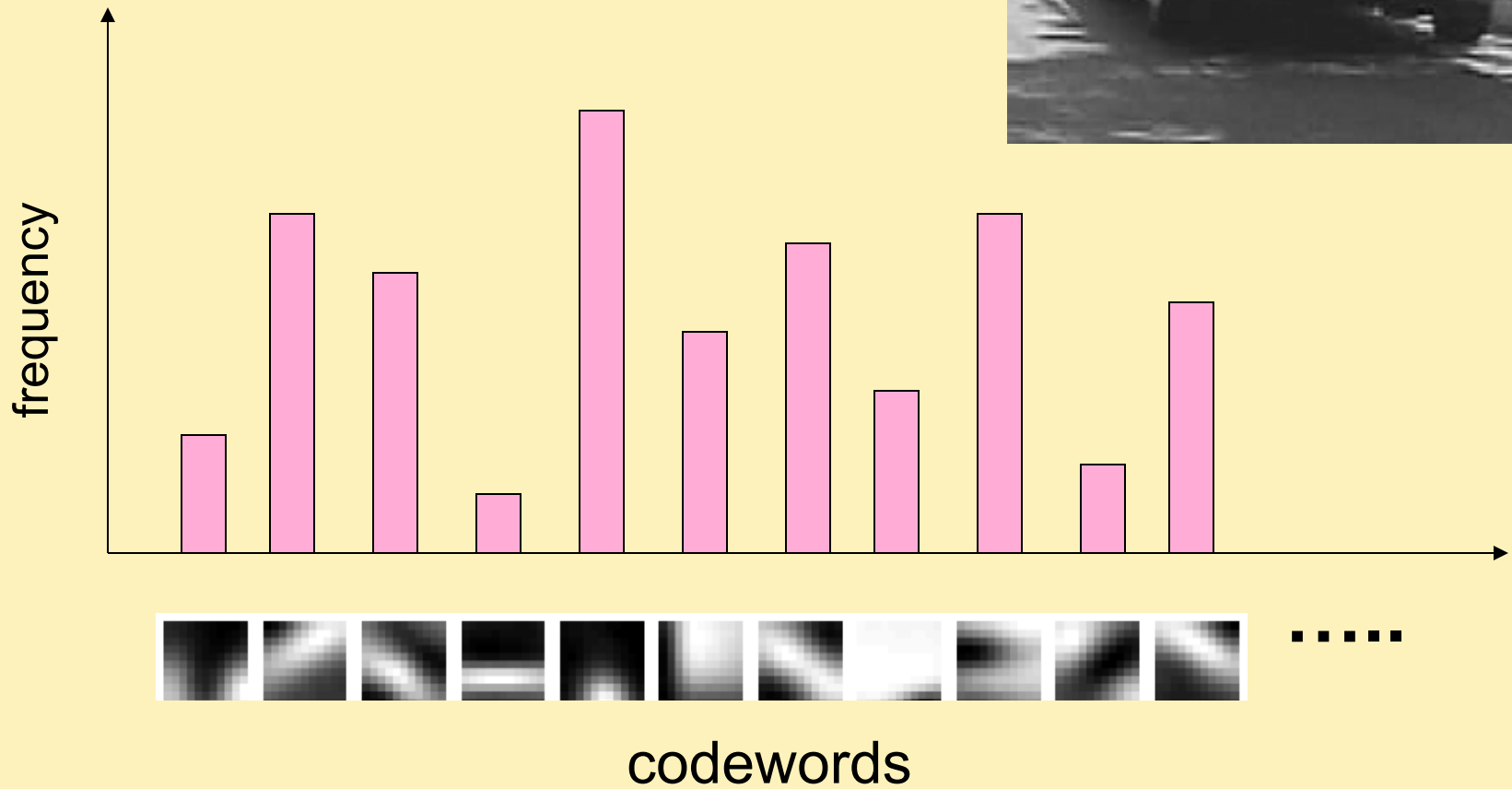
# K-means Questions

- What is it trying to optimize?
- Are we sure it will terminate?
- Are we sure it will find an optimal clustering?
- How should we start it?
- How could we automatically choose the number of centers?

# Image patch examples of codewords

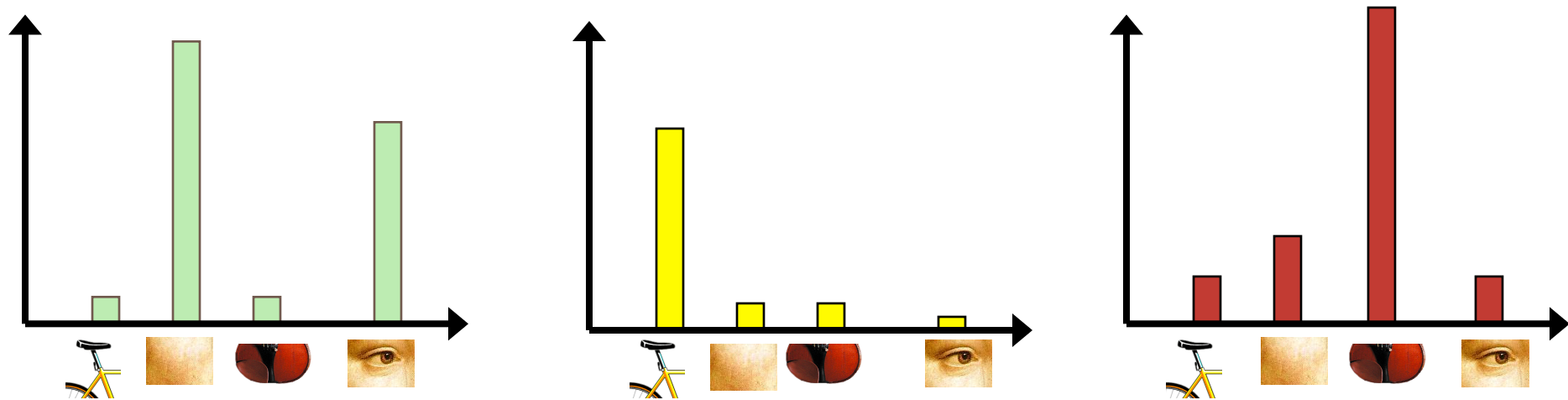


# 3. Image representation



# Image classification

- Given the bag-of-features representations of images from different classes, how do we learn a model for distinguishing them?





# Discriminative methods

- Learn a decision rule (classifier) assigning bag-of-features representations of images to different classes

