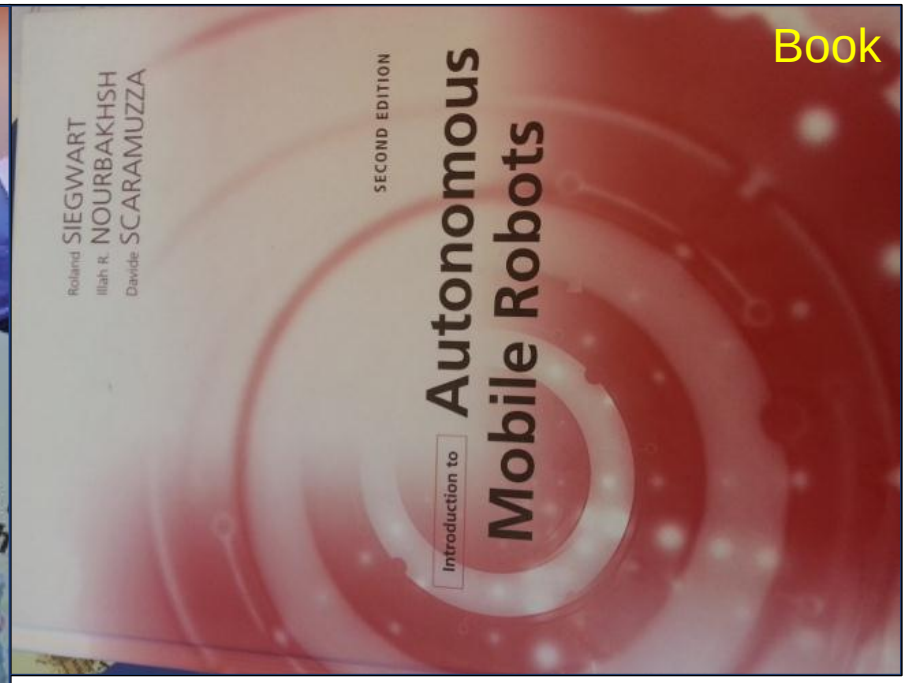
Perception | **Place Recognition**
Autonomous Mobile Robots

Margarita Chli – University of Edinburgh
Paul Furgale, Marco Hutter, Martin Rufli, Davide Scaramuzza, Roland Siegwart

Object Recognition | recognizing known objects

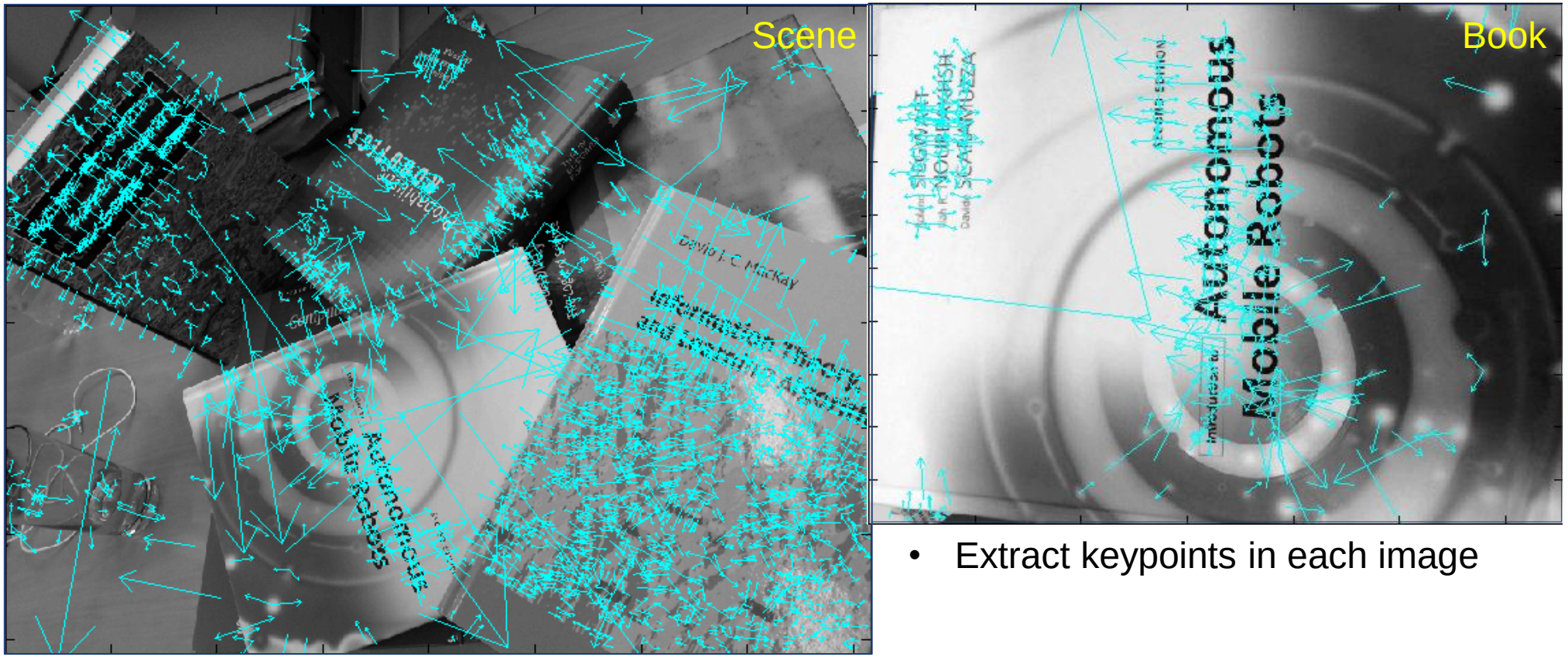
- **Q:** Is this Book present in the Scene?



Object Recognition | recognizing known objects

- **Q:** Is this Book present in the Scene?

Generated using D. Lowe's SIFT demo software:
http://www.cs.ubc.ca/~lowe/keypoints/

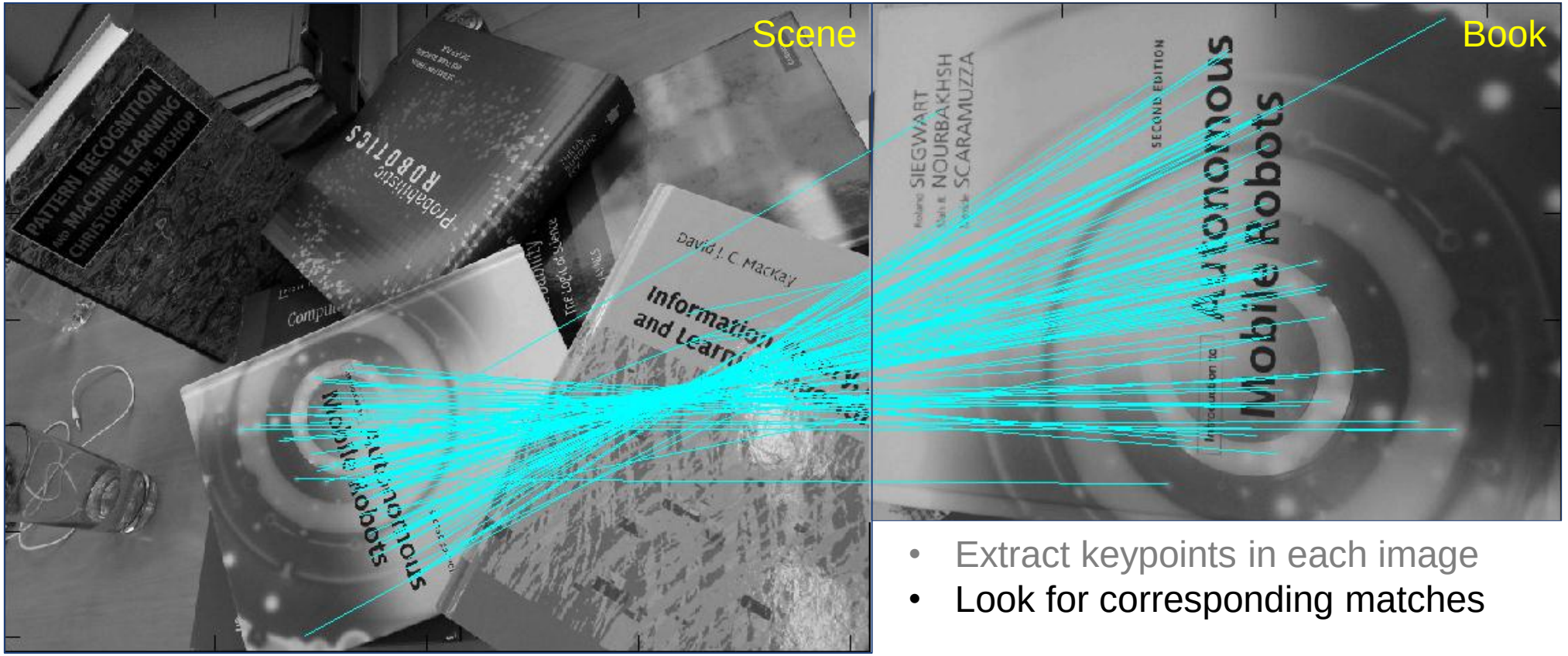


- Extract keypoints in each image

Object Recognition | recognizing known objects

- **Q:** Is this Book present in the Scene?

Generated using D. Lowe's SIFT demo software:
http://www.cs.ubc.ca/~lowe/keypoints/

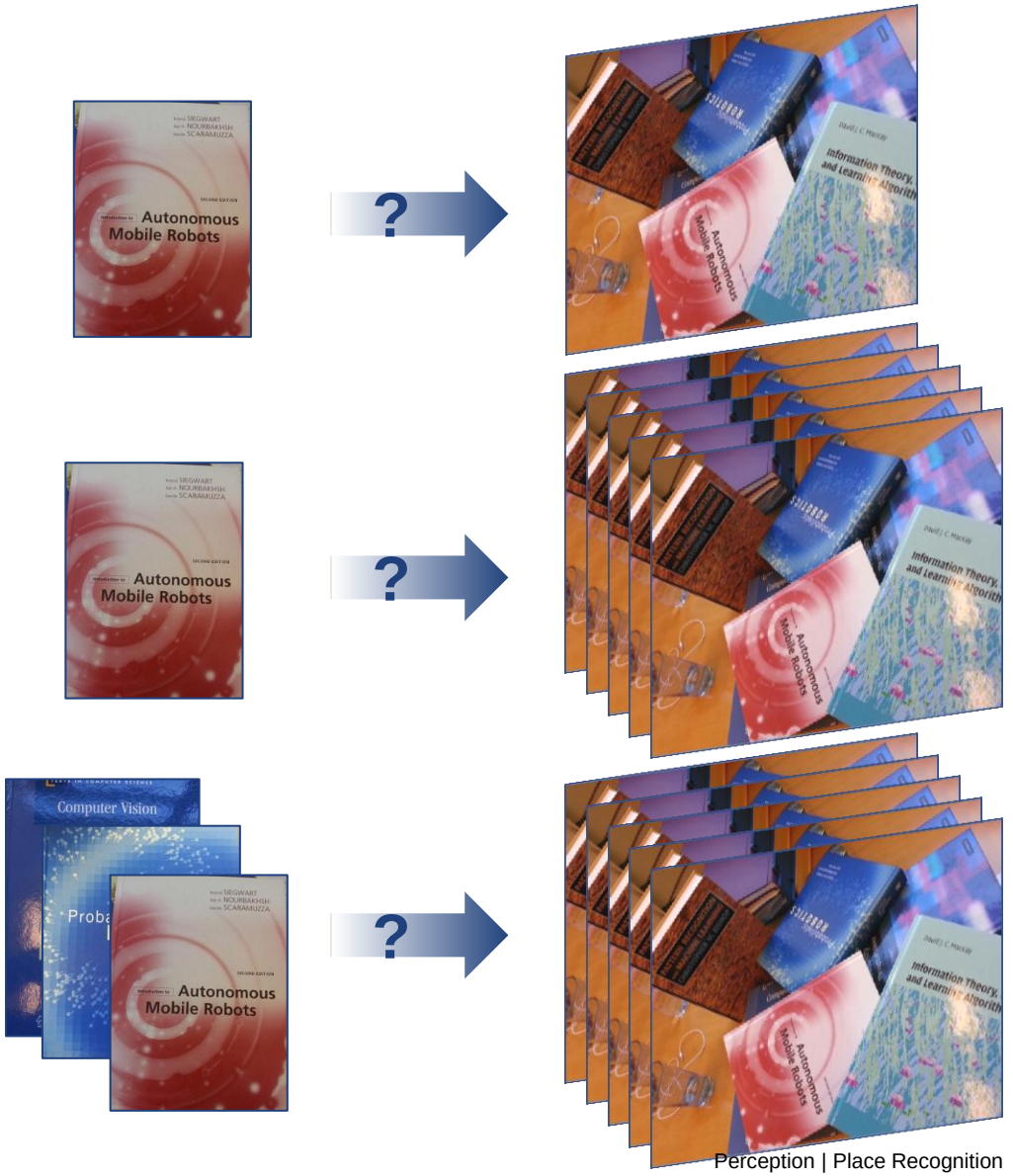


- Extract keypoints in each image
- Look for corresponding matches

- Most of the Book's keypoints appear in the Scene ⇨ **A:** the Book is present in the Scene

Object Recognition | taking this a step further…

- Find an object in an image
- Find an object in multiple images
- Find multiple objects in multiple images



Place Recognition | bag of words

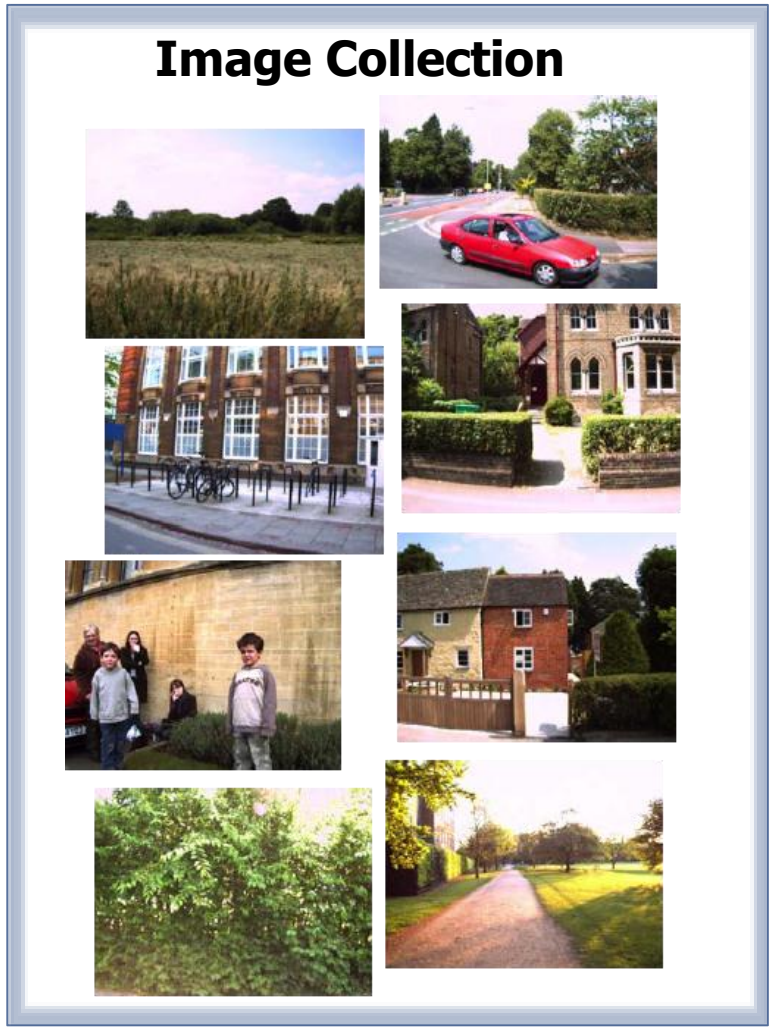
- Extension to scene/place recognition:
  - Is this image in my database?
  - Robot: Have I been to this place before?
    ⇨ '**loop closure**' problem, '**kidnapped robot**' problem

Place Recognition | bag of words

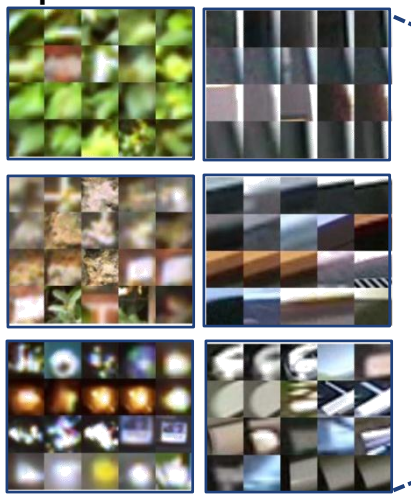
- Extension to scene/place recognition:
  - Is this image in my database?
  - Robot: Have I been to this place before?
    ⇨ '**loop closure**' problem, '**kidnapped robot**' problem

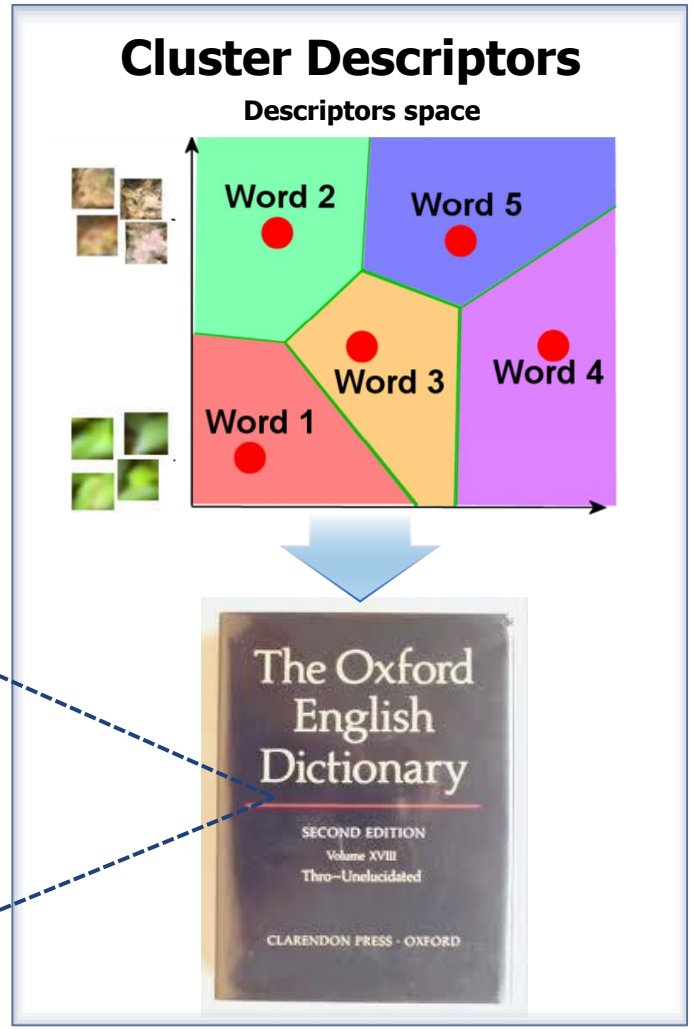
- Use analogies from text retrieval:
  - Visual Words
  - Vocabulary of Visual Words
  - "Bag of Words" approach

Place Recognition| building the visual vocabulary



Examples of Visual Words:



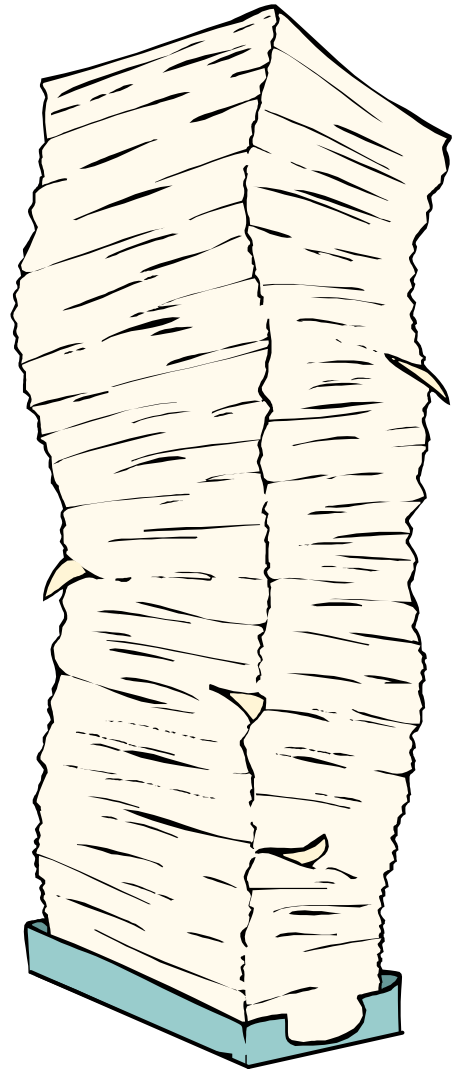


Images for this slide are courtesy of Mark Cummins

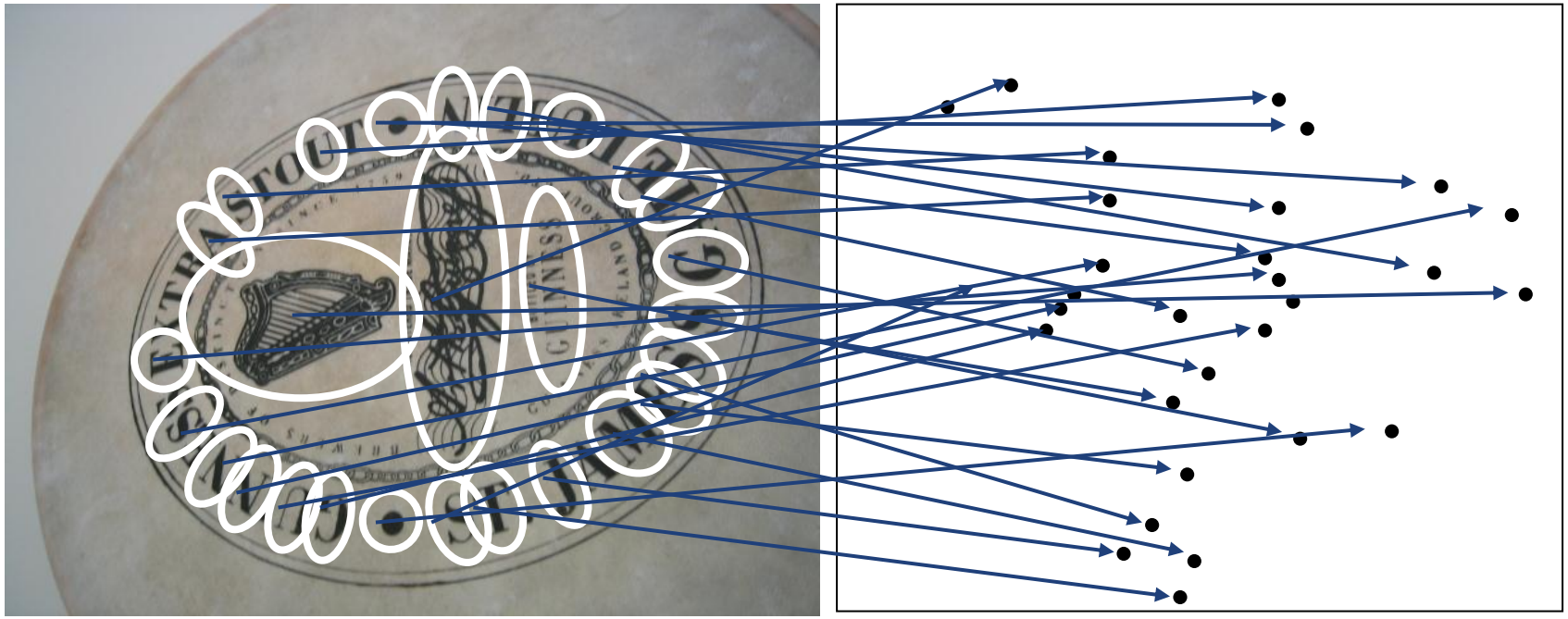
Vocabulary Tree | efficient place/object recognition

- We can describe a **scene as a collection of words** and look up in the database for **images with a similar collection of words**
  [Sivic and Zisserman, ICCV 2003]
- What if we need to find an object/scene in a database of millions of images?
  - Build **Vocabulary Tree** via hierarchical clustering
  - Use the **Inverted File system**: a way of efficient indexing

    (each node in the tree is associated with a list of images containing an instance of this node)
- [Nister and Stewénius, CVPR 2006]

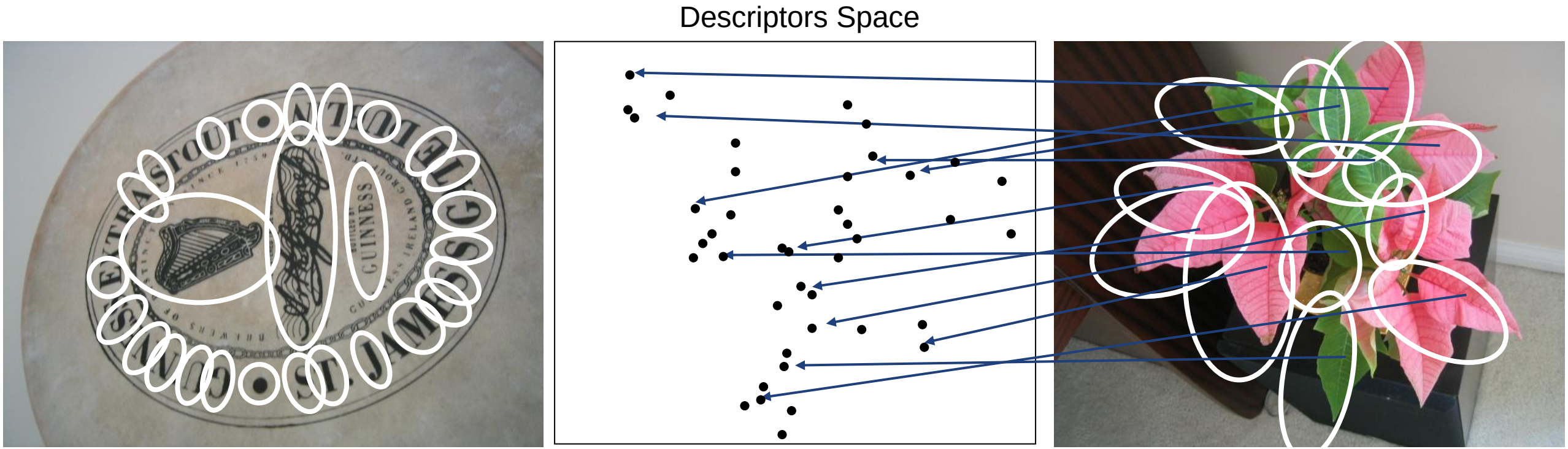


Vocabulary Tree | extract features



Slide based on D. Nister's slides

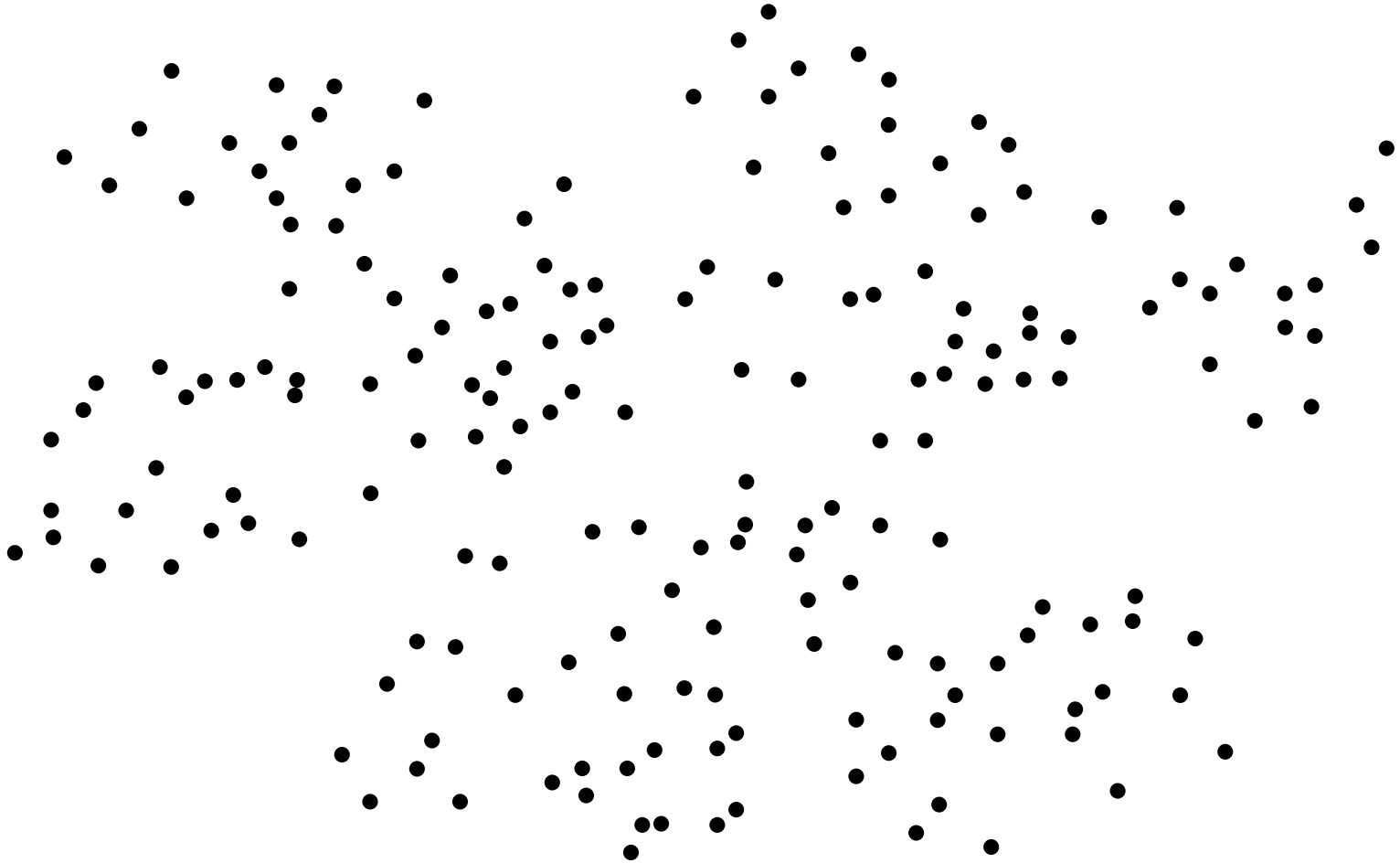
Vocabulary Tree | extract features



based on D. Nister's slides

Vocabulary Tree | hierarchical partitioning

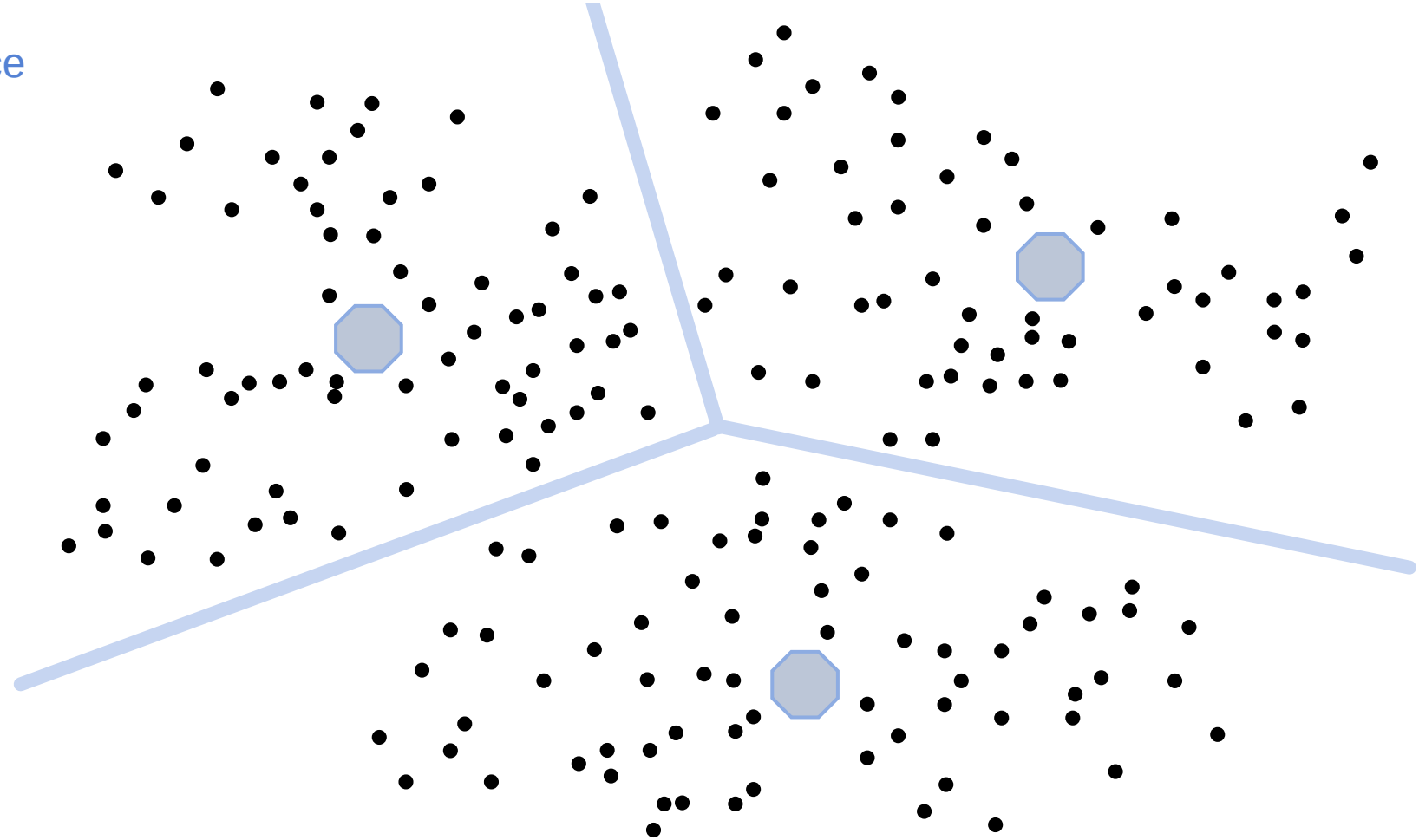
Descriptors space



based on D. Nister's slides

Vocabulary Tree | hierarchical partitioning

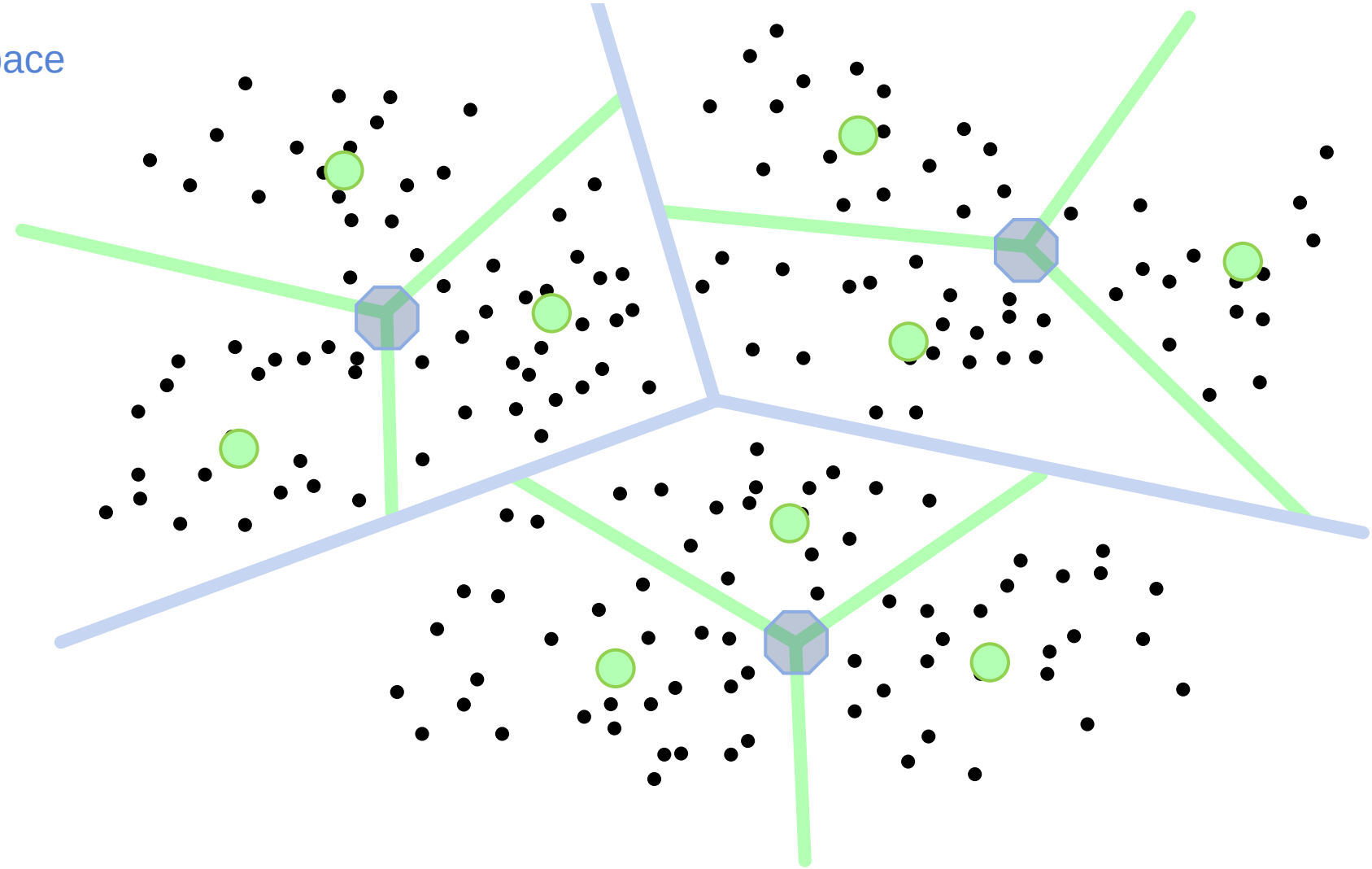
Descriptors space



based on D. Nister's slides

Vocabulary Tree | hierarchical partitioning

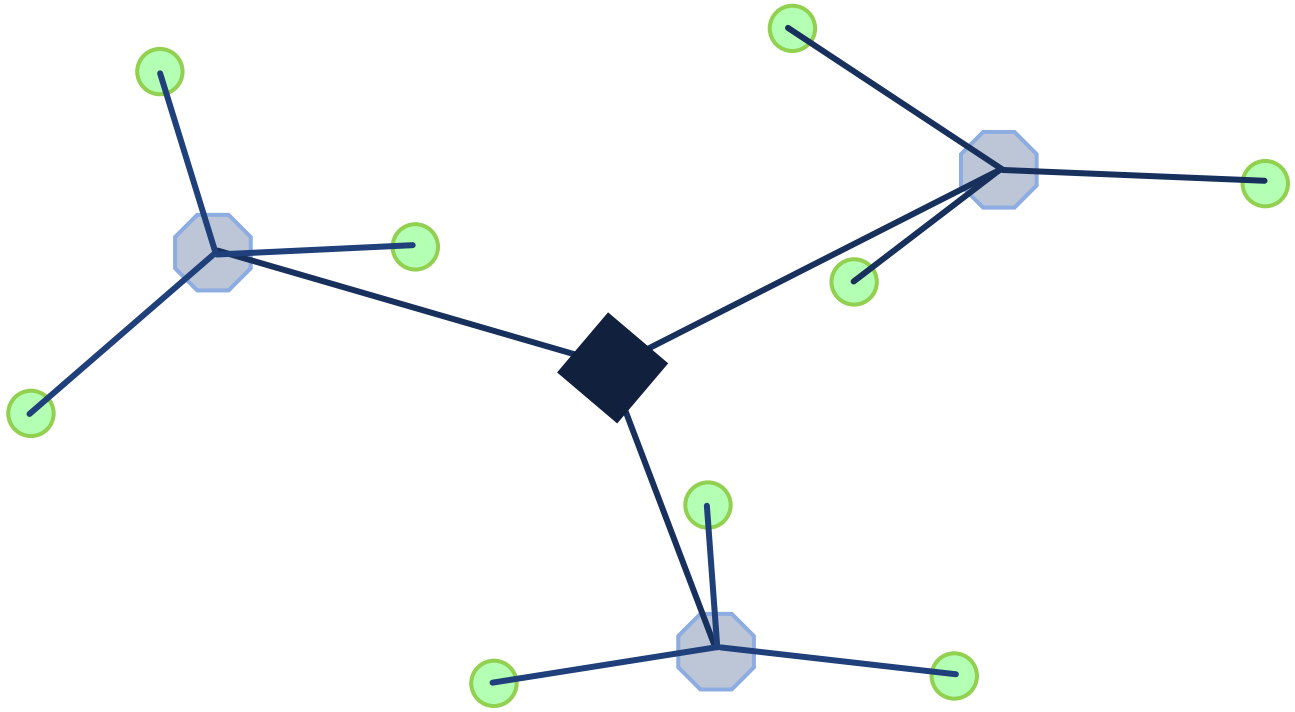
Descriptors space



based on D. Nister's slides

Vocabulary Tree | hierarchical partitioning

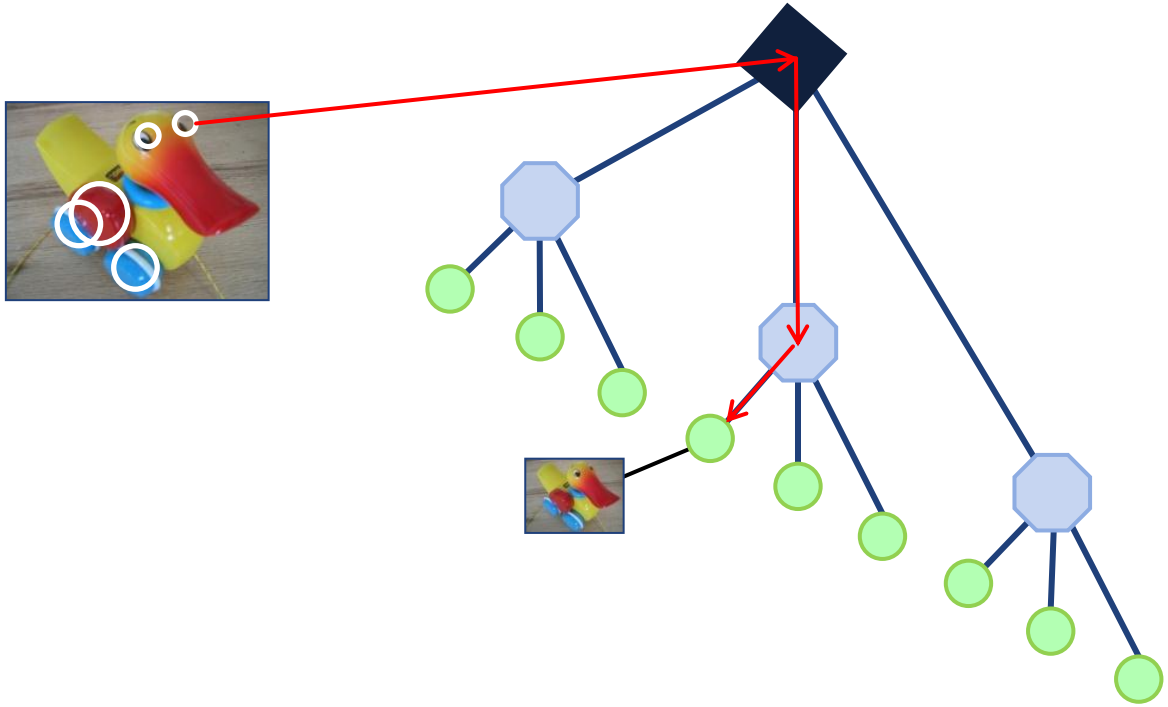
Descriptors space



based on D. Nister's slides

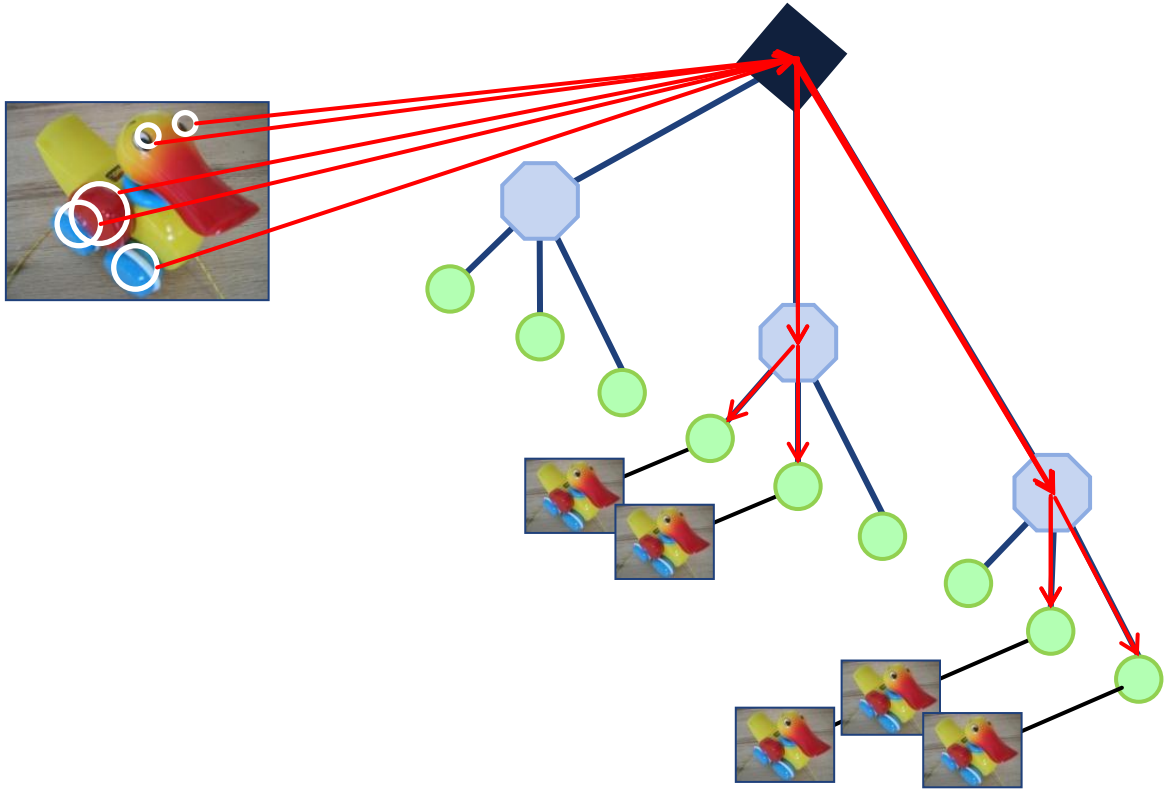
Vocabulary Tree | populating the tree

Model images



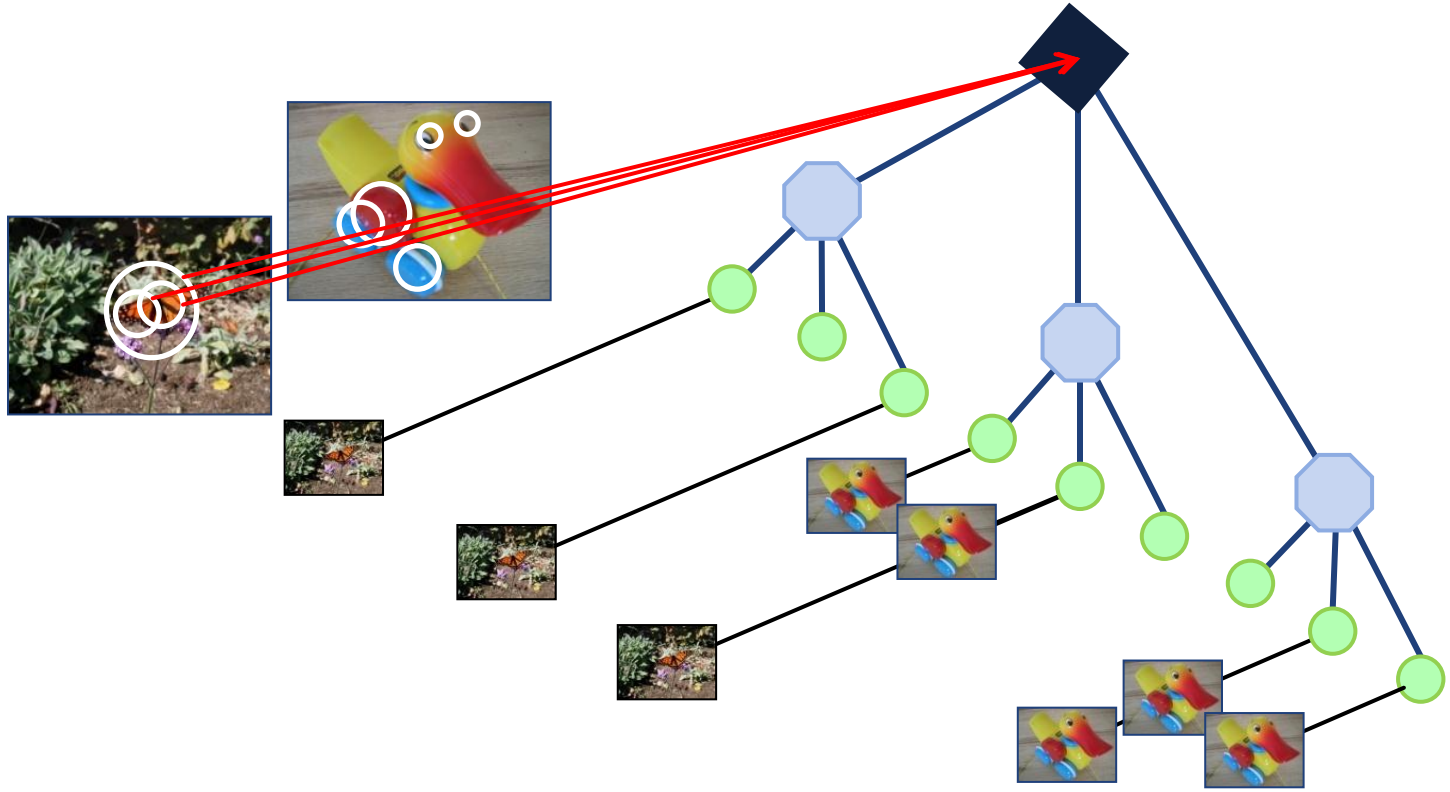
Vocabulary Tree | populating the tree

Model images



Vocabulary Tree | populating the tree

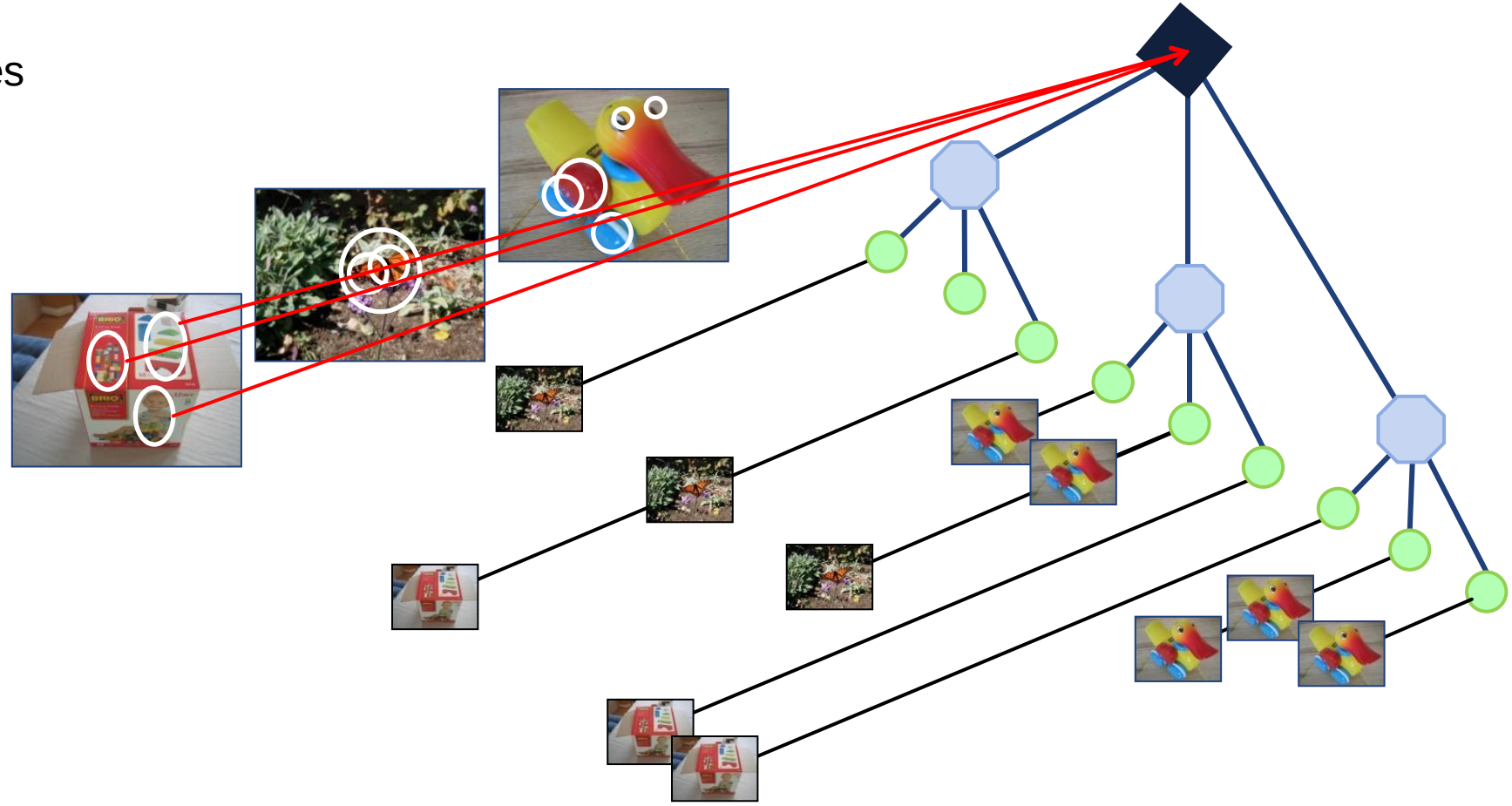
Model images



based on D. Nister's slides

Vocabulary Tree | populating the tree

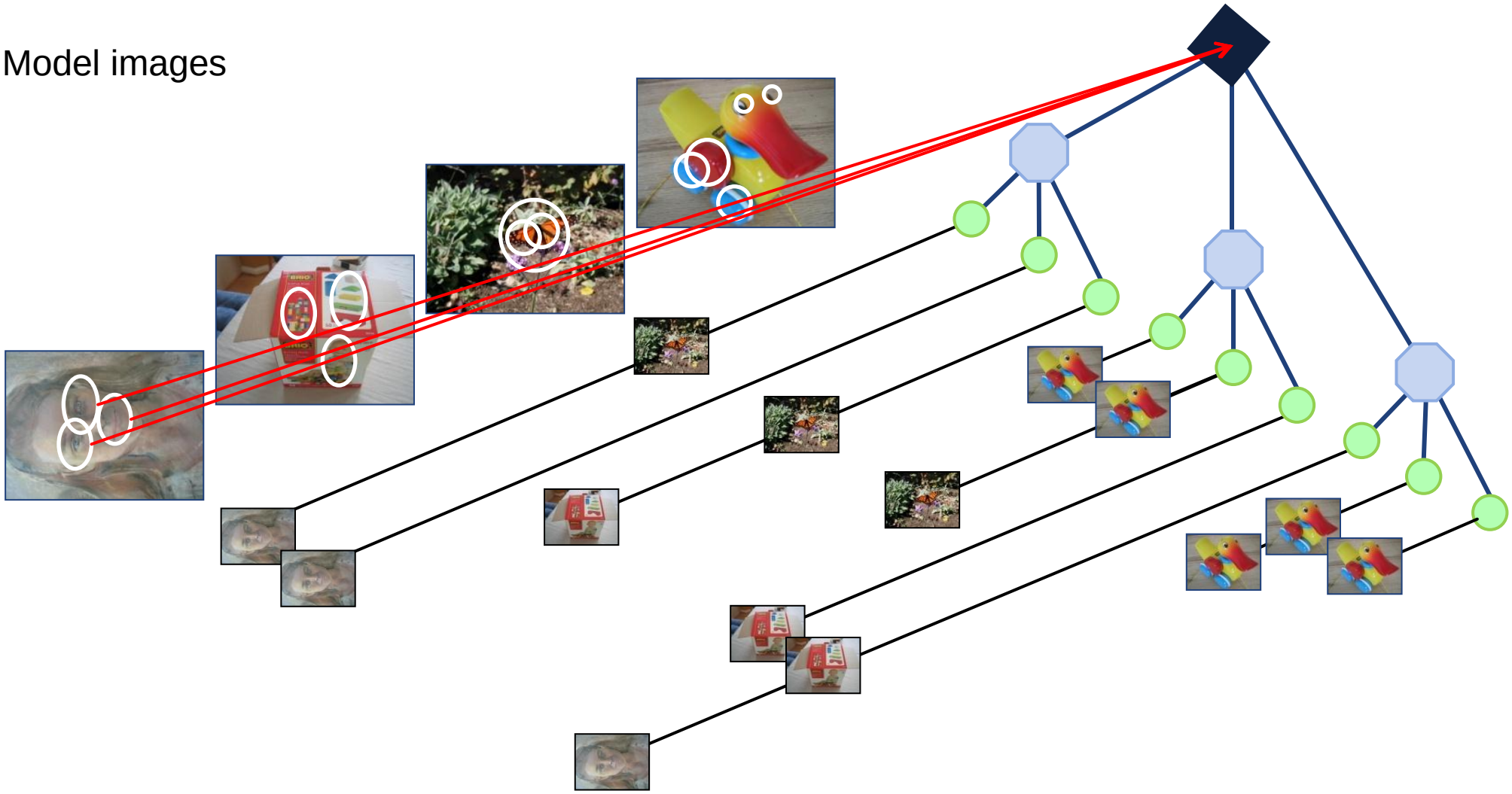
Model images



based on D. Nister's slides

Vocabulary Tree | populating the tree

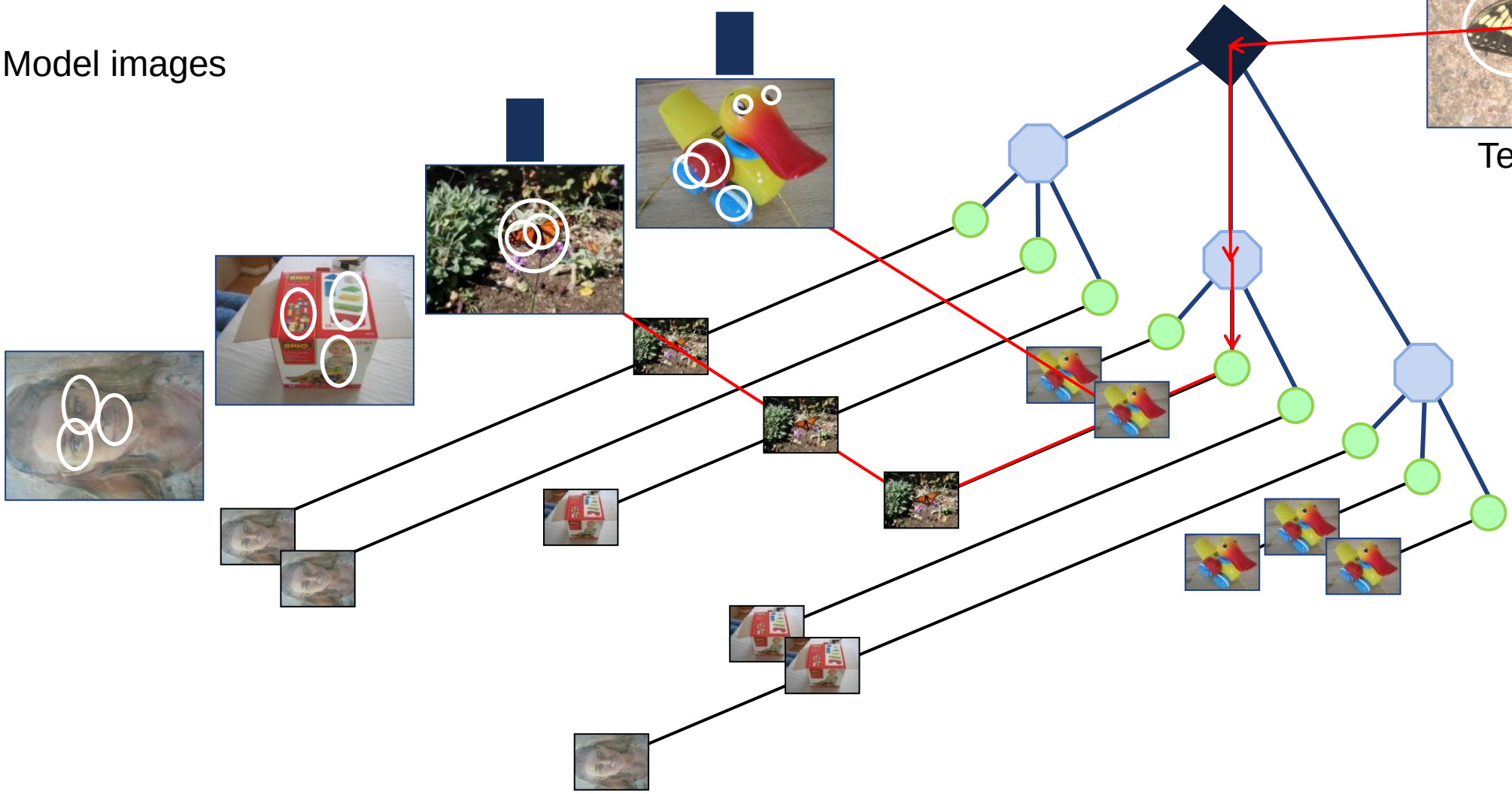
Model images

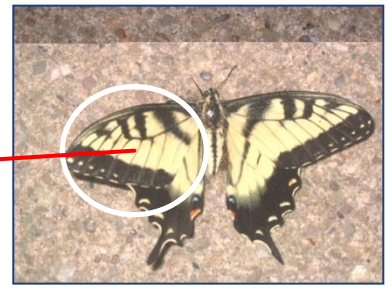


based on D. Nister's slides

Vocabulary Tree | look up a test image

Model images

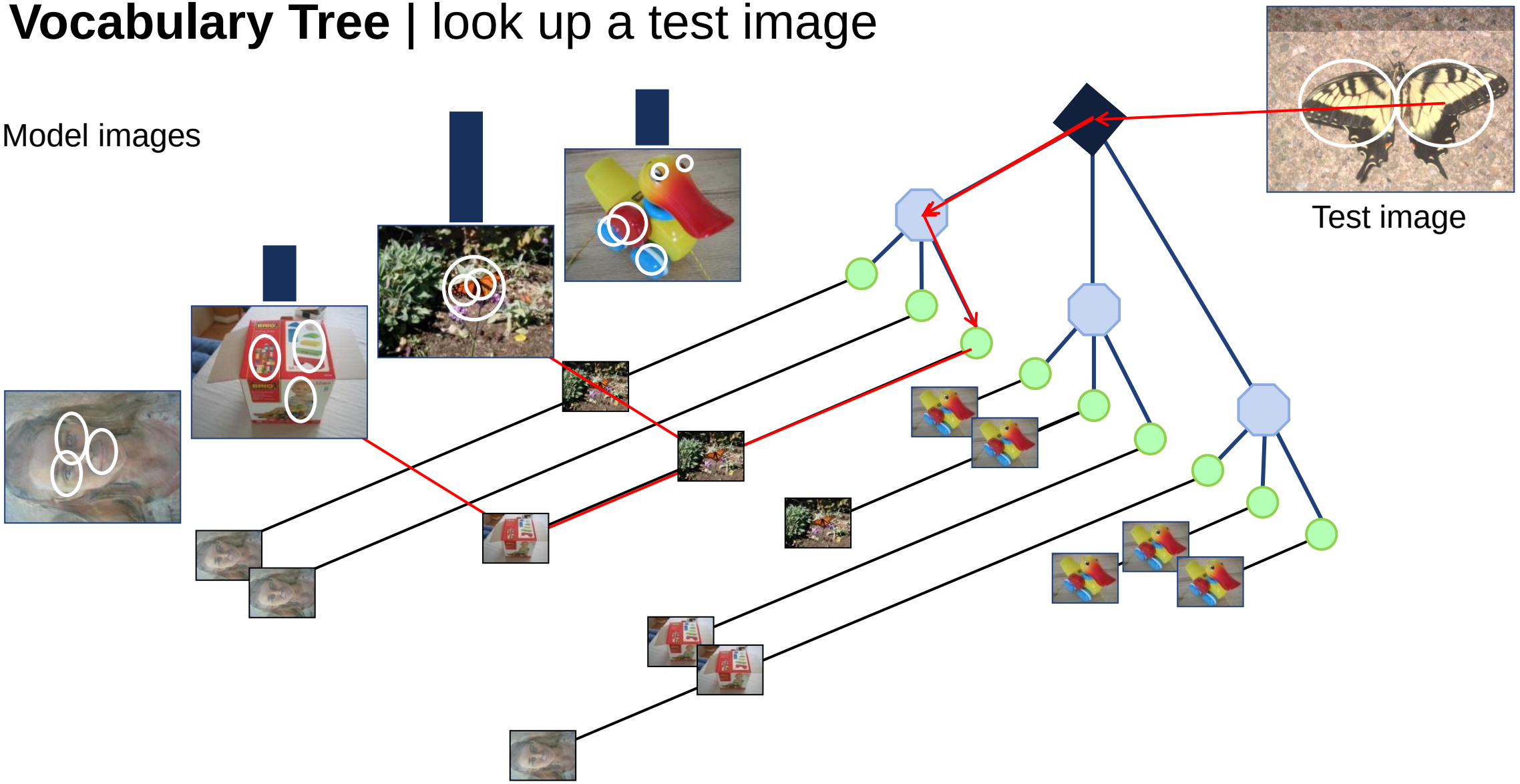




Test image

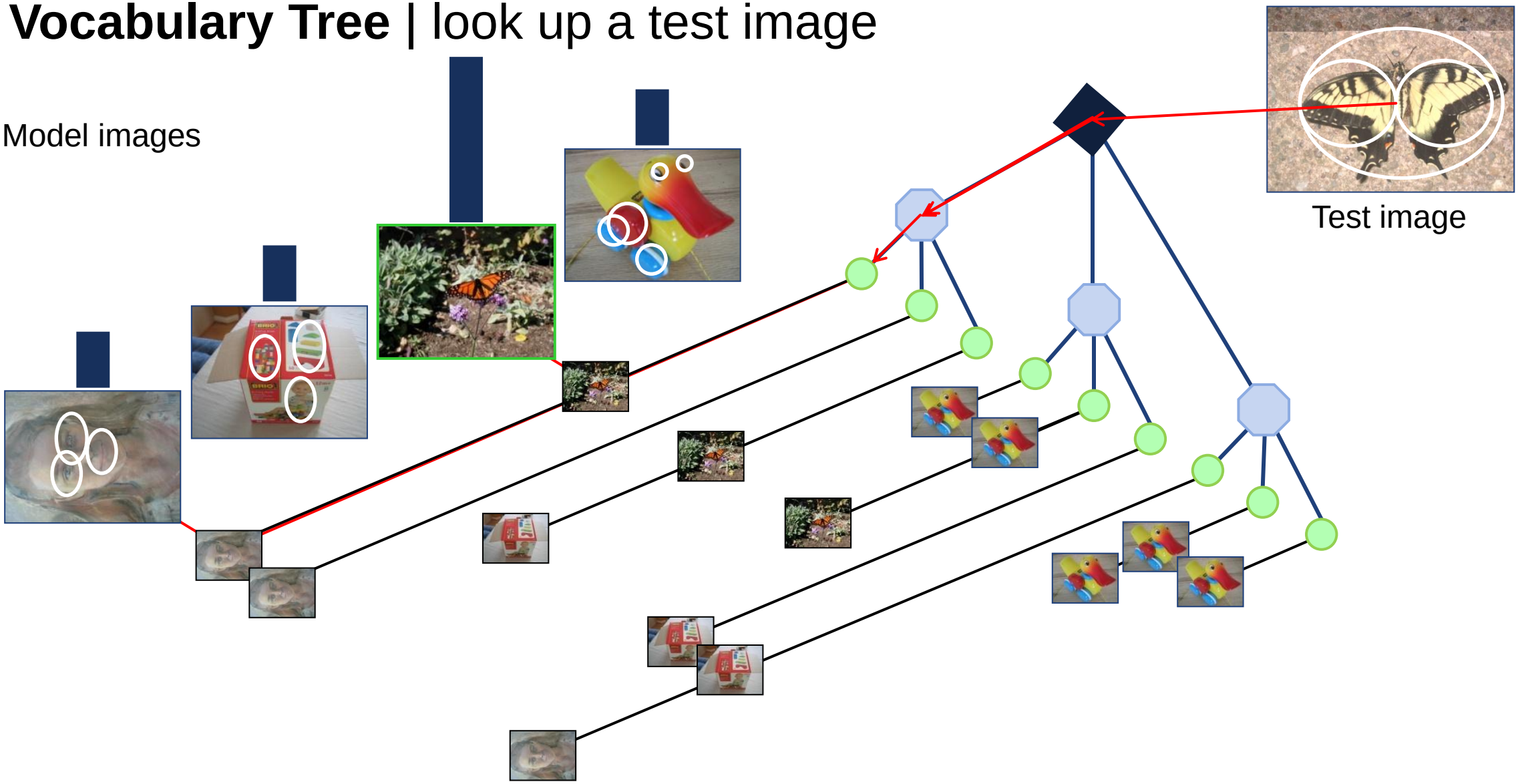
based on D. Nister's slides

Vocabulary Tree | look up a test image



based on D. Nister's slides

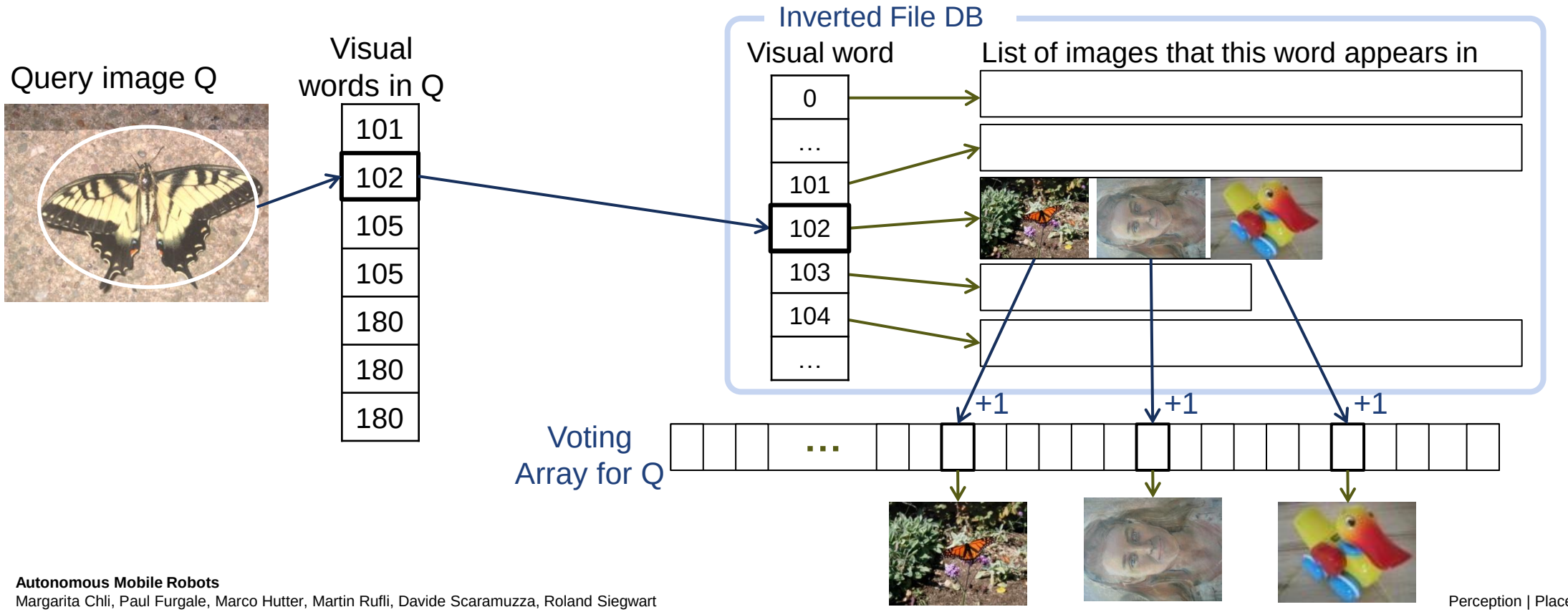
Vocabulary Tree | look up a test image



based on D. Nister's slides

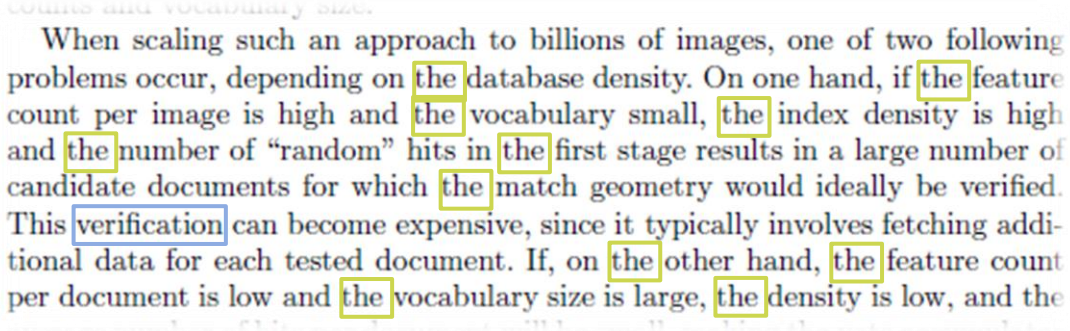
Vocabulary Tree | inverted file index

- Inverted File DB lists all possible visual words
- Each word points to a list of images where this word occurs
- Voting array: has as many cells as images in the DB – each word in query image, votes for an image



Vocabulary Tree | tf-idf

- **t**erm **f**requency-**i**nverse **d**ocument **f**requency: measures the importance of a word inside a document,(as part of a document DB)



When scaling such an approach to billions of images, one of two following problems occur, depending on the database density. On one hand, if the feature count per image is high and the vocabulary small, the index density is high and the number of "random" hits in the first stage results in a large number of candidate documents for which the match geometry would ideally be verified. This verification can become expensive, since it typically involves fetching additional data for each tested document. If, on the other hand, the feature count per document is low and the vocabulary size is large, the density is low, and the

- term frequency: frequency of word w_i in image j: $tf_{ij} = \frac{n_{i,j}}{\sum_k n_{k,j}}$
- inverse document frequency: $idf_i = \log \frac{|D|}{|\{d : w_i \in d\}|}$ ← No. all images (documents); ← No. all images containing w_i
- tf-idf of word w_i in image j is: $= tf_{ij} \cdot idf_i$
- Use it to weigh the importance of each word when voting for corresponding image

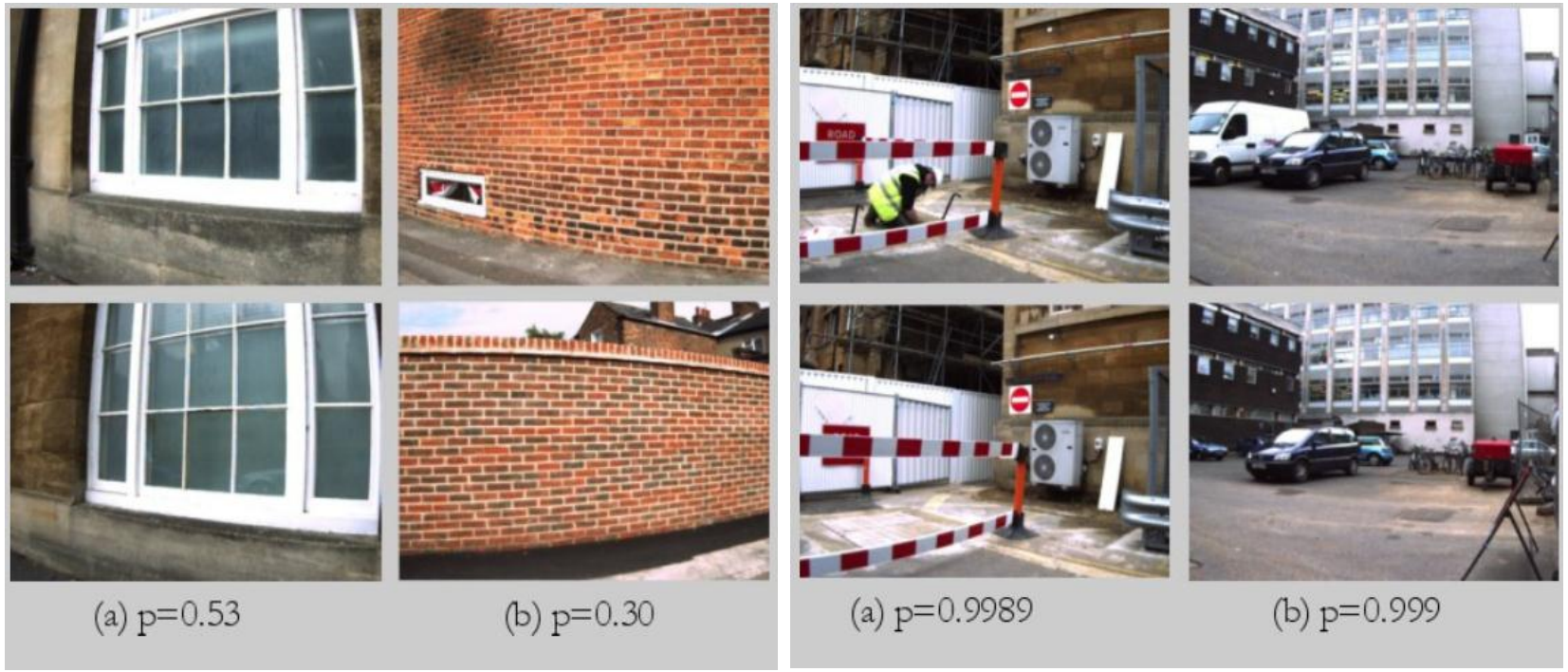
Place Recognition | FABMAP (2.0) [Cummins and Newman IJRR 2011]

- Use training images to build the Bag of Words database
- Probabilistic model of the world: the world is a **set of discrete places**
  - **Place** = a vector of occurrences of visual words in the local scene

- Captures the dependencies of words to distinguish the most characteristic structure of each scene
  (using the Chow-Liu tree)

- Very high performance
- Binaries available online

Images from robots.ox.ac.uk/~mjc/appearance_based_results.html



Place Recognition | robust performance

- Visual Vocabulary holds **appearance information**, but discards the spatial relationships between features
- Two images with the same features shuffled around in the image will be a 100% match when using only appearance information.

- If different arrangements of the same features are expected then one might use **geometric verification**
  - Test the *k* most similar images to the query image for geometric consistency (e.g. using RANSAC)
  - Further reading (out of scope of this course):
    - [Cummins and Newman, IJRR 2011]
    - [Stewénius et al, ECCV 2012]