Visual Vocabularies

Visual Recognition and Search Presentation
by Joseph Cooper

Classification Task

- Assign a query image into one of several classes
  - Yes/No (detection)
  - One of several images (retrieval)
  - One of several types (recognition)
- If we can compress the query down to a fixed-length vector, we can leverage many machine learning techniques for classification (e.g., SVM).

Classify with Pixels

- Histogram of entire image
- Classify each pixel

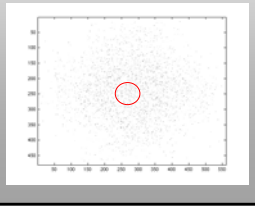
Fatis & Olsen 2003

Lastra et al. 2007

Use Spatial Information

- Scale/pad/crop all images to common size
  (treat each pixel as a dimension in the feature vector)
- Fourier transform
  (select specific fixed frequencies of interest)

Too much spatial information!
Loses all invariance.



Central Idea

- Develop a canonical set of features—"the vocabulary"
  - Define a basis set for a new space
  - Partition the space into classes
- Describe new images in terms of the vocabulary
- Classify images (or parts of images) according to where they fall in the vocabulary space

Steps

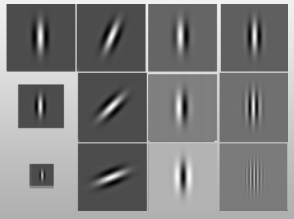
- Define a way of acquiring image features
- Partition the space defined by features
- Assign a value to each partition
  - (or combinations of partitions)
- Classify new images by the partitions in which they fit

Training Set Issues

- Supervision level
- Type of classification
- Recognition abstraction
  - A specific motorcycle
  - All motorcycles
  - All things with two wheels
  - All vehicles
  - All blue things

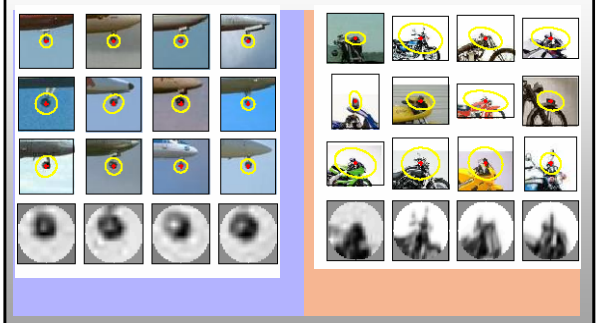
Determine Feature Transform

- Filter banks
- Textons
- Interest operators
- SIFT
- Histograms
- Patches
- Wavelet transforms
- Higher order info (spatial relationships between features)
- Etc.



Find features invariant to the right things, but distinct in the right ways.

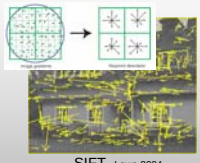
Sparse patches



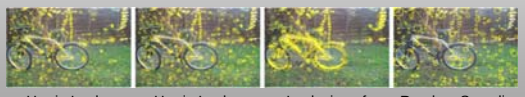
Sivic et al. 2005

Sparse vs. dense sampling

- Sparse sampling does not provide enough samples



SIFT Lowe 2004

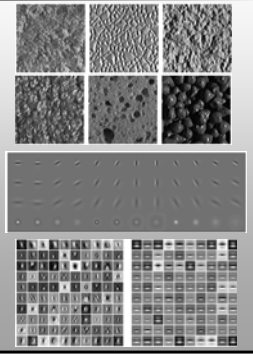


Harris-Laplace | Harris-Laplace 0 threshold | Laplacian of Gaussian | Random Sampling

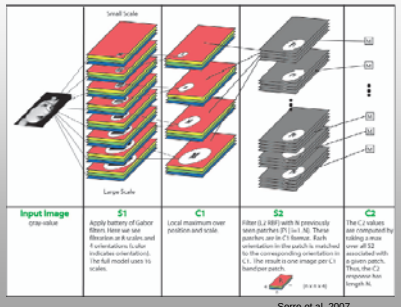
Nowak et al. 2006

Dense Encoding
Visual words = textons

- Previous use of local feature quantization:
- *Texton* = cluster center of filter responses over collection of images [Leung & Malik, 1999; Varma & Zisserman 2002]
- Represent texture or material with histogram of texton occurrences (or prototypes of whatever feature type employed)



Dense patches



Serre et al. 2007

What do we do with a Set of Features? Create a Vocabulary

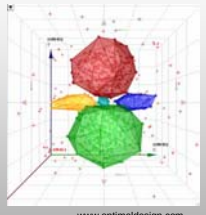
- A given image may have many or few "interesting points"
- Data are not ordered
- Collapse the feature set into a vector representation
  - K-means
  - Hierarchical k-means
  - Regular lattice
- Memory/Processing considerations

Partitioning methods

- Flat vs. hierarchical
  - Nearest neighbor
  - Hash
  - Trees
  - Forests
- Dynamic vs. static
  - Adapt to recent needs

K-means

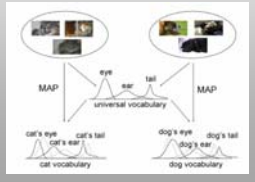
- Build up probability of features belonging to a given class
- Slow

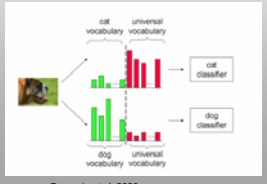


www.optimaldesign.com

"Universal Vocabulary"

- Generic codebook from random images
- Single vocabulary for all classes from relevant images

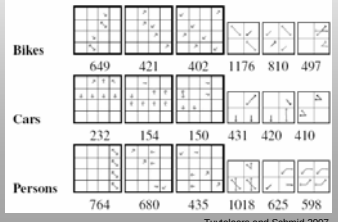




Peronnin et al. 2006

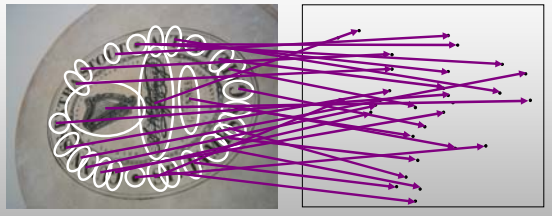
Regular Lattice

- Discretize feature space
- Characterize each feature from the set as a point in a very high dimensional space
- Use sparse representation of clustered points



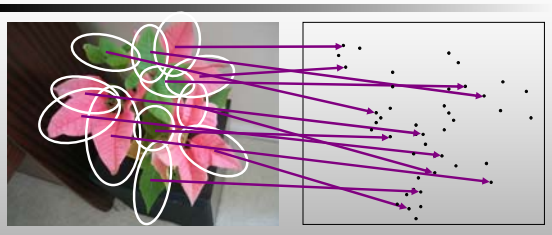
Tuytelaars and Schmid 2007

Vocabulary tree example



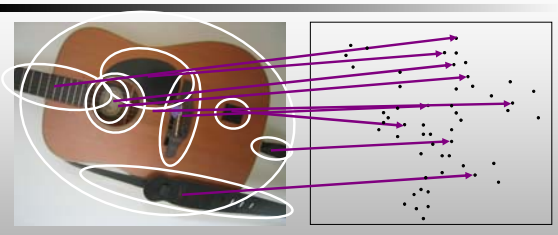
Nister et al. 2006

Extract features from training set



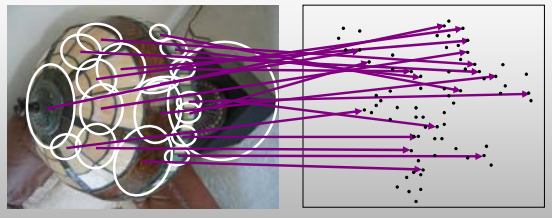
Nister et al. 2006

Extract features from training set

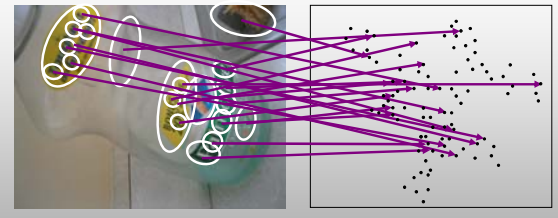


Nister et al. 2006

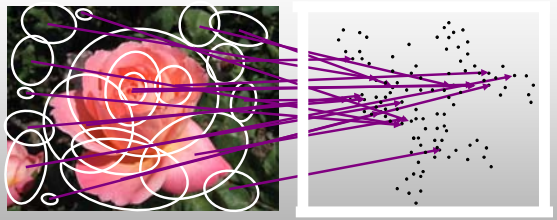
Extract features from training set



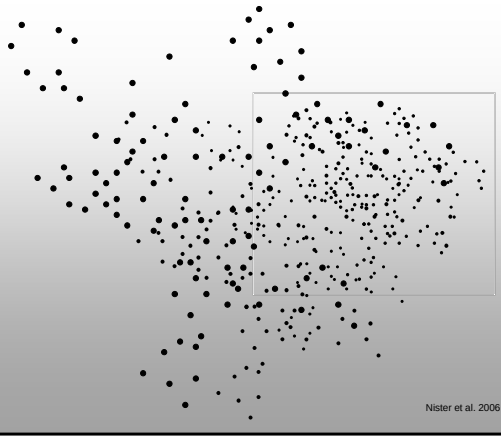
Nister et al. 2006



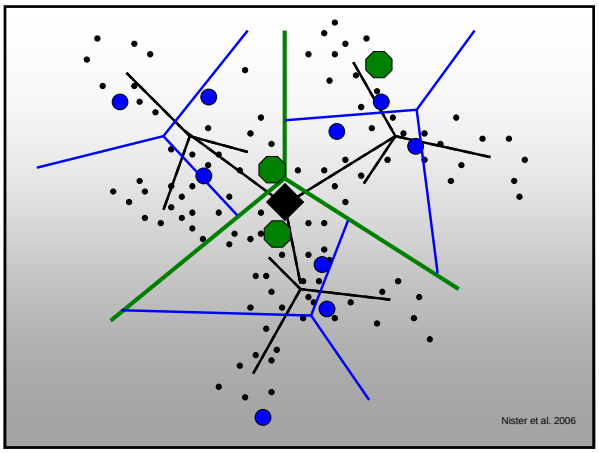
Nister et al. 2006

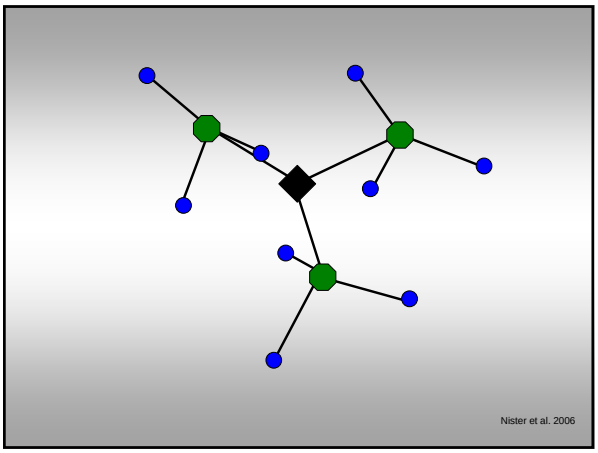


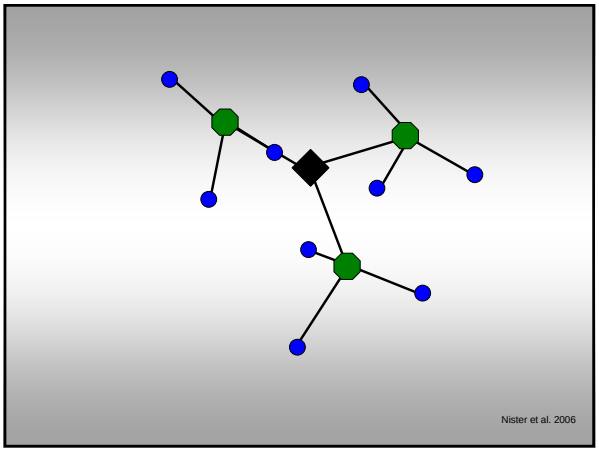
Nister et al. 2006

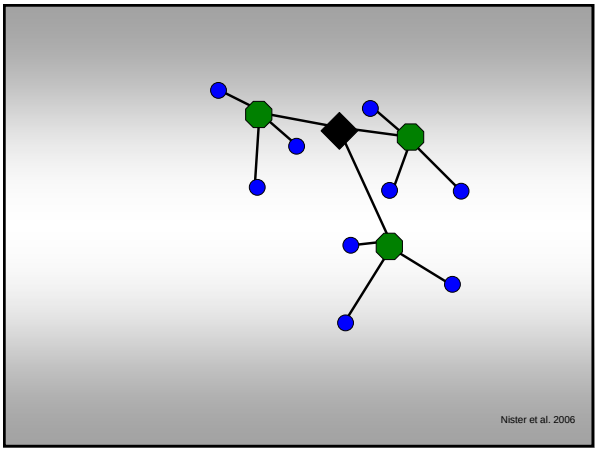


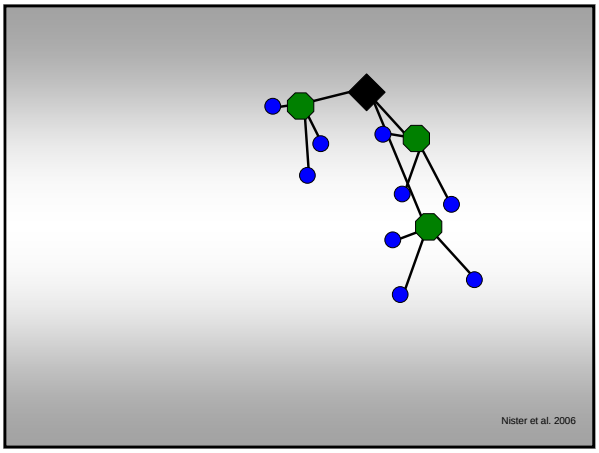
Nister et al. 2006

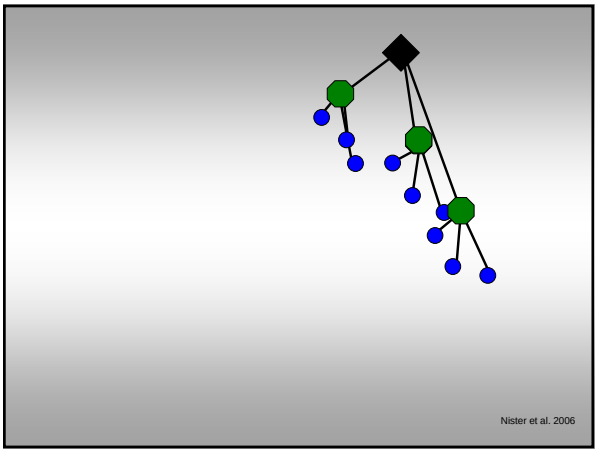

Nister et al. 2006

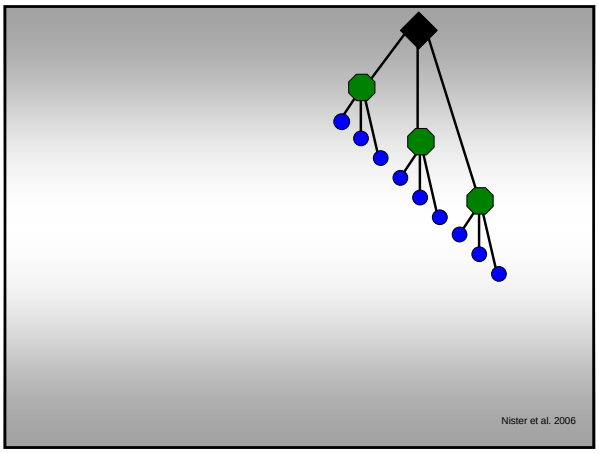

Nister et al. 2006

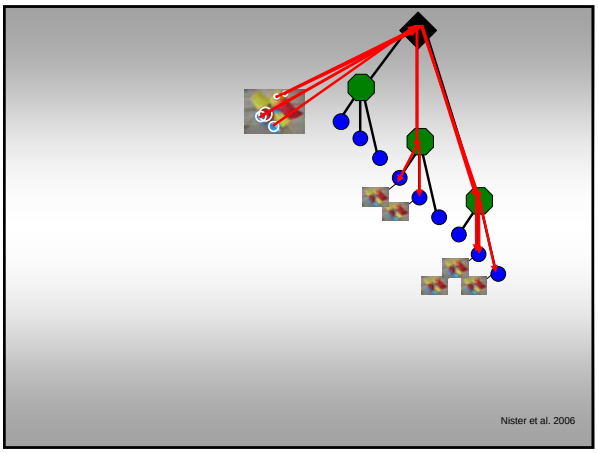

Nister et al. 2006

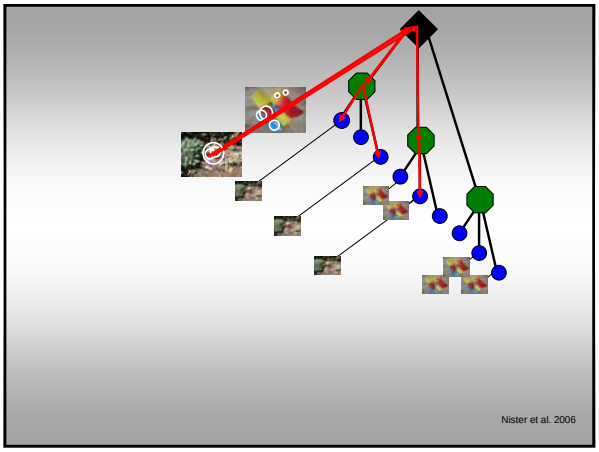

Nister et al. 2006

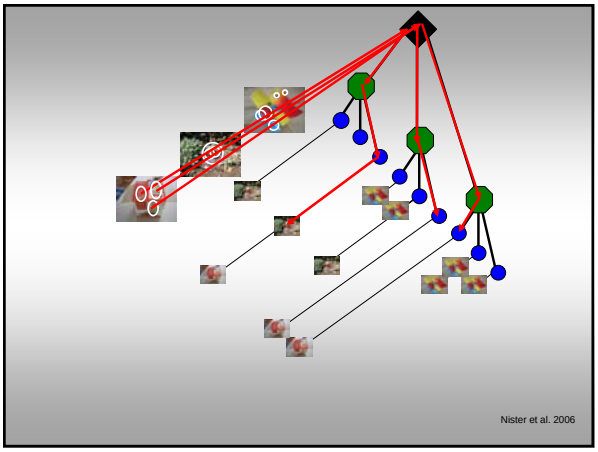

Nister et al. 2006

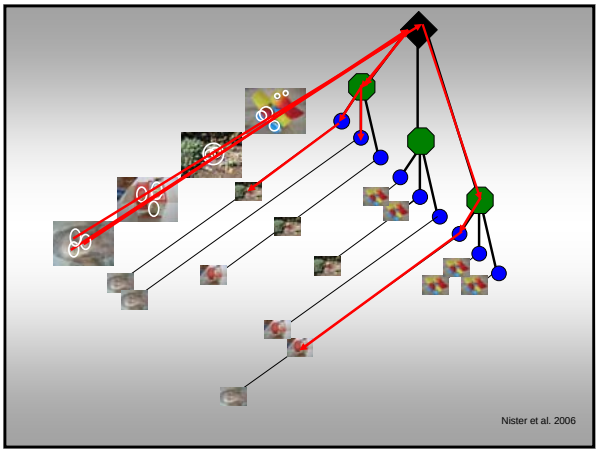

Nister et al. 2006

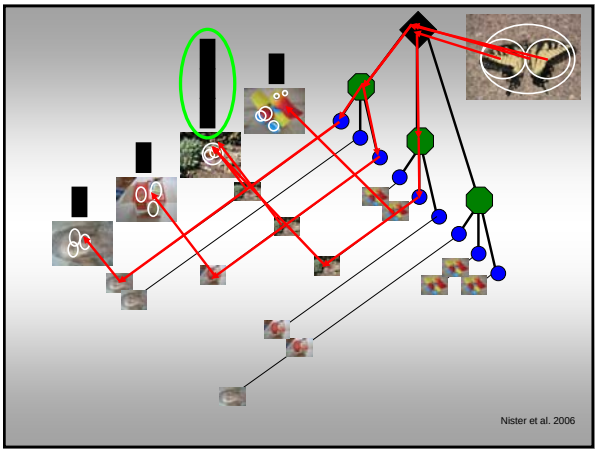

Nister et al. 2006


Nister et al. 2006


Nister et al. 2006


Nister et al. 2006


Nister et al. 2006

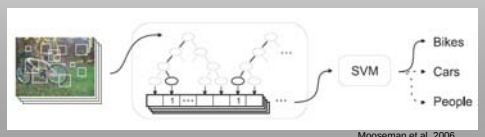

Nister et al. 2006

Select Intelligent Partitions

- What partitioning is informative?
  - Randomly choose features
  - Cluster features
  - Look for discriminating features (EM) (Use labels to inform decision)
- Multiple vocabularies

Use Labels to Build Classifier

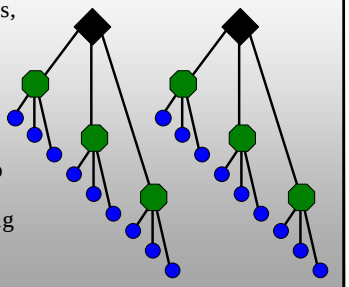
- Select words that are strong indicators of a specific class or subset of classes
- Voting scheme (histogram)
- ML classifier
  - Naïve Bayes
  - SVM



Mooseman et al. 2006

Randomized Forests

- Instead of K-means, use decision tree features selected semi-randomly
- Each leaf node is a "word" not a vote
- Individual trees do not need to be as complex, improving speed

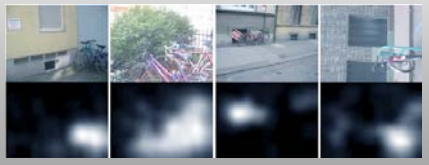


Use Vocabulary to Classify

- Extract features from new image
- Find matching prototype features in pre-computed vocabulary
- Determine match
  - Specific image (inverted index)
  - Specific class
    - Which one of many
    - Present/not present

Discussion

- Nearly all methods discard spatial information
- Nearly all methods discard color information
- Why not label hierarchically



Discussion

- What is the space actually covered by a given vocabulary/classification system?
- How easy are they to break?
  - How do you build the worst possible dataset?
- How do these methods compare to human vision?
- Is it possible to use parametric functions describing image parts as words?