Recognizing or retrieving specific objects

Search photos on the web for particular places

Find these landmarks ...in these images and 1M more

[Josef Sivic]
Feature extraction

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

Interest point detector
- Csurka et al. 2004
- Fei-Fei and Perona, 2005
- Sivic et al. 2005
Learning the visual vocabulary

Clustering

Visual vocabulary

Clustering

Slide credit: Josef Sivic
Image representation

frequency

codewords
Comparing bags of words

- Rank frames by normalized scalar product between their (possibly weighted) occurrence counts---*nearest neighbor* search for similar images.

$$\text{sim}(d_j, q) = \frac{\langle d_j, q \rangle}{\|d_j\| \|q\|}$$

$$= \frac{\sum_{i=1}^{V} d_j(i) \ast q(i)}{\sqrt{\sum_{i=1}^{V} d_j(i)^2} \ast \sqrt{\sum_{i=1}^{V} q(i)^2}}$$

for vocabulary of $V$ words

[Kristen Grauman]
K-Nearest Neighbors classification

- For a new point, find the k closest points from training data
- Labels of the k points “vote” to classify

If query lands here, the 5 NN consist of 3 negatives and 2 positives, so we classify it as negative.
Inverted file index

- Database images are loaded into the index mapping words to image numbers.
Inverted file index

When will this give us a significant gain in efficiency?

- New query image is mapped to indices of database images that share a word.
TF-IDF weighting

- **Term frequency** – **inverse document frequency**
- Describe frame by frequency of each word within it, downweight words that appear often in the database
- (Standard weighting for text retrieval)

\[
t_i = \frac{n_{id}}{n_d} \log \frac{N}{n_i}
\]

- Total number of documents in database
- Number of occurrences of word \(i\) in document \(d\)
- Number of words in document \(d\)
- Number of documents word \(i\) occurs in, in whole database
Hierarchical clustering for large vocabularies

- Tree construction:

[Nister & Stewenius, CVPR’06]
Vocabulary tree

• Training: Filling the tree

[Nister & Stewenius, CVPR’06]
Vocabulary tree

- Training: Filling the tree

[Nister & Stewenius, CVPR’06]
RANSAC verification
Scoring retrieval quality

Database size: 10 images
Relevant (total): 5 images

Query

\[
\text{precision} = \frac{\text{Number of relevant}}{\text{Number of returned}}
\]

\[
\text{recall} = \frac{\text{Number of relevant}}{\text{Number of total relevant}}
\]

Results (ordered):

[Ondrej Chum]
Support Vector Machines
Linear classifiers

• Find linear function to separate positive and negative examples

\[ x_i \text{ negative: } x_i \cdot w + b < 0 \]

\[ x_i \text{ positive: } x_i \cdot w + b \geq 0 \]

Which line is best?
Support Vector Machines (SVMs)

- Discriminative classifier based on optimal separating line (for 2d case)
- Maximize the margin between the positive and negative training examples
Support vector machines

- Want line that maximizes the margin.

$$\mathbf{x}_i \text{ positive } (y_i = 1) : \mathbf{x}_i \cdot \mathbf{w} + b \geq 1$$

$$\mathbf{x}_i \text{ negative } (y_i = -1) : \mathbf{x}_i \cdot \mathbf{w} + b \leq -1$$

For support vectors, $$\mathbf{x}_i \cdot \mathbf{w} + b = \pm 1$$
Support vector machines

- Want line that maximizes the margin.

For support vectors, \( \mathbf{x}_i \cdot \mathbf{w} + b = \pm 1 \)

Distance between point and line:
\[
\frac{\mathbf{w}^T \mathbf{x} + b}{\|\mathbf{w}\|} = \frac{\pm 1}{\|\mathbf{w}\|}
\]

For support vectors:
\[
M = \left| \frac{1}{\|\mathbf{w}\|} - \frac{-1}{\|\mathbf{w}\|} \right| = \frac{2}{\|\mathbf{w}\|}
\]

\[x_i \text{ positive } (y_i = 1) : \mathbf{x}_i \cdot \mathbf{w} + b \geq 1\]

\[x_i \text{ negative } (y_i = -1) : \mathbf{x}_i \cdot \mathbf{w} + b \leq -1\]
Support vector machines

- Want line that maximizes the margin.

For support vectors, $x_i \cdot w + b = \pm 1$

Distance between point and line:
$$\frac{|x_i \cdot w + b|}{||w||}$$

Therefore, the margin is
$$\frac{2}{||w||}.$$
Finding the maximum margin line

1. Maximize margin \( \frac{2}{\|w\|} \)

2. Correctly classify all training data points:

   \( x_i \) positive \((y_i = 1)\) : \( x_i \cdot w + b \geq 1 \)

   \( x_i \) negative \((y_i = -1)\) : \( x_i \cdot w + b \leq -1 \)

Quadratic optimization problem (over all training data):

Minimize \( \frac{1}{2} w^T w \)

Subject to \( y_i(w \cdot x_i + b) \geq 1 \)

C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery, 1998
Finding the maximum margin line

Constrained optimization problem:

Minimize  \[ \frac{1}{2} \mathbf{w}^T \mathbf{w} \]

Subject to  \[ y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \text{ for all training points } i \]

Lagrangian formulation for unconstrained optimization:

\[
\min_{\mathbf{w}, b} \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_i \alpha_i \left[ y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \right]
\]

Condition for optimum:

\[
\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_i \alpha_i y_i \mathbf{x}_i = 0
\]

\[
\frac{\partial L}{\partial b} = \sum_i \alpha_i y_i = 0
\]
Learning the weights

• Primal problem

\[
\min_{w,b} \frac{1}{2} w^T w - \sum_i \alpha_i [y_i (w \cdot x_i + b) - 1]
\]
Learning the weights

• Primal problem

\[
\min_{w,b} \frac{1}{2} w^\top w - \sum_i \alpha_i \left[ y_i (w \cdot x_i + b) - 1 \right]
\]

\[
w = \sum_i \alpha_i y_i x_i\]

\[
\sum_i \alpha_i y_i = 0
\]
Learning the weights

• Primal problem

\[
\min_{w,b} \frac{1}{2} w^\top w - \sum_i \alpha_i [y_i(w \cdot x_i + b) - 1]
\]

• Consider the dual problem

\[
\begin{align*}
    w &= \sum_i \alpha_i y_i x_i \\
    \sum_i \alpha_i y_i &= 0
\end{align*}
\]

\[
\min_\alpha \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \sum_i \alpha_i
\]
Learning the weights

- **Primal problem**

\[
\min_{w, b} \frac{1}{2} w^T w - \sum_i \alpha_i [y_i (w \cdot x_i + b) - 1]
\]

- **Consider the dual problem**

\[
w = \sum_i \alpha_i y_i x_i \quad \sum_i \alpha_i y_i = 0
\]

- **Take derivatives with respect to \( \alpha_i \) and set to 0**

\[
\min_{\alpha} \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \sum_i \alpha_i
\]
Learning the weights

- **Primal problem**

\[
\min_{\mathbf{w}, b} \frac{1}{2} \mathbf{w}^{\top} \mathbf{w} - \sum_{i} \alpha_i \left[ y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \right]
\]

- **Consider the dual problem**

\[
\mathbf{w} = \sum_{i} \alpha_i y_i \mathbf{x}_i
\]

\[
\sum_{i} \alpha_i y_i = 0
\]

\[
\max_{\alpha} \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) - \sum_{i} \alpha_i
\]

- **Take derivatives with respect to** \( \alpha_i \) **and set to 0**

Note: only dot products of data points appear here
Finding the maximum margin line

- Solution: \( w = \sum \alpha_i y_i x_i \)

  - Most of the \( \alpha_i \) will turn out to be 0.
  - The non-zero \( \alpha_i \) correspond to support vectors.
Finding the maximum margin line

• Solution: \( w = \sum_{i} \alpha_i y_i x_i \)

\[ b = y_i - w \cdot x_i \] (for any support vector)

\[ w \cdot x + b = \sum_{i} \alpha_i y_i x_i \cdot x + b \]

• Classification function:

\[ f(x) = \text{sign}(w \cdot x + b) = \text{sign}(\sum_{i} \alpha_i y_i x_i \cdot x + b) \]

If \( f(x) < 0 \), classify as negative.
If \( f(x) > 0 \), classify as positive.

Note: only dot products of data points appear here
Generic category recognition: basic framework

• Build or train object model
  – Choose a representation
  – Learn or fit parameters of model or classifier

• Generate candidates in new image

• Score the candidates
Window-based models
Building an object model

Given the representation, train a binary classifier.

No, not a car.

Yes, car.
Window-based models
Generating and scoring candidates

Car and non-car Classifier

Slide: Kristen Grauman
Window-based object detection: recap

**Training:**
1. Obtain training data
2. Define features
3. Define classifier

**Given new image:**
1. Slide window
2. Score by classifier

---

[Diagram showing the process of window-based object detection with training examples and feature extraction]
HoG descriptor

Dalal and Triggs, CVPR 2005
Person detection with HoGs and linear SVMs

- Map each grid cell in the input window to a histogram counting the gradients per orientation.
- Train a linear SVM using training set of pedestrian or non-pedestrian windows.

Dalal and Triggs, CVPR 2005
Person detection with HoGs and linear SVMs

Dalal and Triggs, CVPR 2005