Scalable Recognition with a Vocabulary Tree

Juan David Adarve
Malik Goumidi

MLSR Presentation
10/12/2010
PLAN

- Problem description
- Building the vocabulary tree
- Scoring
- Results
- Conclusion
Problem Description

- To create a recognition scheme scaling well with large image databases
Constraints

- Scalability
- Retrieval speed
- Training speed
- Size of the data structure (compactness)
- Rotation, scale, lighting conditions
Related work

• Evolution:
  • 400 frames to 35000
  • 10-15 minutes training vs offline unsupervised training
  • New indexing way: offline vs on-the-fly

• Size of structure:
  • Pyramid of histograms doubling the number of bins along each axis at each level
PLAN

• Problem description

• Building the vocabulary tree

• Scoring

• Results

• Conclusion
Building the Vocabulary Tree

- Training set of data:
  - Build a hierarchical quantization of visual words
  - Thanks to hierarchical K-means clustering
- Visual word = A feature vector extracted from the image (own version of MESR and SIFT)
- $K =$ the branch factor of the tree
- $L =$ depth of the tree
Extracting features (visual words)
Hierarchical K-means
Organizing them in a tree
Searching in the tree

- Propagation down the tree of a descriptor vector (visual word)
  - Comparison of the vector with $k$ children (at each level)
  - Choice of the closest one
  - For a tree with $L$ levels: $kL$ dot products (efficient if $k$ is not too large)
- The path found can be encoded by an integer and used for scoring
PLAN

• Problem description

• Building the vocabulary tree

• Scoring

• Results

• Conclusion
TF.IDF

- TF: Term Frequency
- IDF: Inverse Document Frequency

\[ q_i = \hat{n}_i \cdot \hat{w}_i \]

- \( q_i \): number of descriptor vectors of the query image
- \( n_i \): number of descriptor vectors of the query image
- \( w_i \): the weight of those descriptors
- \( N \): the number of images in the database
- \( N_i \): the number of images in the database with at least one descriptor vector path through node \( i \).
Relevance of an image:
  → The normalized difference between the query and database vectors

\[ s(q, d) = \left\| \frac{d}{\|d\|} - \frac{q}{\|q\|} \right\| \]
Scoring

- The same idea for retrieving a text containing specific words:
  - One Text
  - One Vocabulary tree

- Search one word of the text in the vocabulary tree
  - TF: Times the word occurs in a text
  - IDF: $\ln(\text{Number of text} / \text{Number of text where the term occurs})$
PLAN

• Problem description

• Building the vocabulary tree

• Scoring

• Results

• Conclusion
Results

The image set with ground truth contains 6376 images in groups of four that belong together.

Number of leaf nodes more important than branch factor.
PLAN

- Problem description
- Building the vocabulary tree
- Scoring
- Results
- Conclusion
Conclusion

- A very powerful recognition tool
- Scalability with huge databases
- Better results with large vocabulary
- Use of K-means and TF-IDF to optimize the tree