Indexing local features

To generate candidate matches, find patches that have the most similar appearance (e.g., lowest SSD).

Simplest approach: compare them all, take the closest (or closest k, or within a thresholded distance).

In stereo case, may constrain by proximity if we make assumptions on max disparities.

Each patch / region has a descriptor, which is a point in some high-dimensional feature space (e.g., SIFT).
Indexing local features

- When we see close points in feature space, we have similar descriptors, which indicates similar local content.

Indexing local features

- With potentially thousands of features per image, and hundreds to millions of images to search, how to efficiently find those that are relevant to a new image?

Indexing local features: inverted file index

- For text documents, an efficient way to find all pages on which a word occurs is to use an index...
- We want to find all images in which a feature occurs.
- To use this idea, we'll need to map our features to "visual words".

Text retrieval vs. image search

- What makes the problems similar, different?

Visual words: main idea

- Extract some local features from a number of images ...

Visual words: main idea

- e.g., SIFT descriptor space: each point is 128-dimensional
Visual words: main idea

Each point is a local descriptor, e.g. SIFT vector.

Visual words

• Map high-dimensional descriptors to tokens/words by quantizing the feature space

• Quantize via clustering, let cluster centers be the prototype “words”

• Determine which word to assign to each new image region by finding the closest cluster center.

Visual words

• Example: each group of patches belongs to the same visual word

Figure from Sivic & Zisserman, ICCV 2003

Kristen Grauman
Visual words and textons

- First explored for texture and material representations
- Texton = cluster center of filter responses over collection of images
- Describe textures and materials based on distribution of prototypical texture elements.

Leung & Malik 1999; Varma & Zisserman, 2002

Recall: Texture representation example

- Visual words and textons
- First explored for texture and material representations
- Texton = cluster center of filter responses over collection of images
- Describe textures and materials based on distribution of prototypical texture elements.

Leung & Malik 1999; Varma & Zisserman, 2002

Visual vocabulary formation

Issues:
- Sampling strategy: where to extract features?
- Clustering / quantization algorithm
- Unsupervised vs. supervised
- What corpus provides features (universal vocabulary?)
- Vocabulary size, number of words

Inverted file index

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

Kristen Grauman

When will this give us a significant gain in efficiency?

- New query image is mapped to indices of database images that share a word.
- If a local image region is a visual word, how can we summarize an image (the document)?
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 what is seen. For a long time it was thought that the retinal image is projected onto the retina, transmitted point by point to visual centers in the brain; the cerebral cortex was a “black box” into which the visual image was projected. Through the movie screen, so to speak, upon which the visual image was transmitted, the visual experience is the output. In this system, each cell is responsible for a specific detail in the pattern of the retinal image.

Hubel and Wiesel have been able to demonstrate that the visual cortex is layered, and that the visual impulses along their path correspond to the various cell layers of the optical cortex. By following the visual impulses along their path, it is possible to determine that the specific functions of the visual cortex are distributed as follows:

• Visual cortex
  • Surplus, commerce, exports, imports, US, EU, bank, domestic, foreign, increase, trade, value
  • China, trade

For a long time it was thought that the retinal image, as it falls on the retina, undergoes a step in the visual process. Now we know that behind the origin of the visual perception in the brain there is a considerably more complicated course of events. By following the visual impulses along their path, it is possible to determine that the specific functions of the visual cortex are distributed as follows:

• Visual cortex
  • Surplus, commerce, exports, imports, US, EU, bank, domestic, foreign, increase, trade, value
  • China, trade

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 what is seen. For a long time it was thought that the retinal image is projected onto the retina, transmitted point by point to visual centers in the brain; the cerebral cortex was a “black box” into which the visual image was projected. Through the movie screen, so to speak, upon which the visual image was transmitted, the visual experience is the output. In this system, each cell is responsible for a specific detail in the pattern of the retinal image.

Hubel and Wiesel have been able to demonstrate that the visual cortex is layered, and that the visual impulses along their path correspond to the various cell layers of the optical cortex. By following the visual impulses along their path, it is possible to determine that the specific functions of the visual cortex are distributed as follows:

• Visual cortex
  • Surplus, commerce, exports, imports, US, EU, bank, domestic, foreign, increase, trade, value
  • China, trade

For a long time it was thought that the retinal image, as it falls on the retina, undergoes a step in the visual process. Now we know that behind the origin of the visual perception in the brain there is a considerably more complicated course of events. By following the visual impulses along their path, it is possible to determine that the specific functions of the visual cortex are distributed as follows:

• Visual cortex
  • Surplus, commerce, exports, imports, US, EU, bank, domestic, foreign, increase, trade, value
  • China, trade

3/30/2011
Video Google System
1. Collect all words within query region
2. Inverted file index to find relevant frames
3. Compare word counts
4. Spatial verification
Sivic & Zisserman, ICCV 2003

Sivic & Zisserman, ICCV 2003
• Demo online at:
  http://www.robots.ox.ac.uk/~vgg/Research/vgoglu/index.html

Scoring retrieval quality

Vocabulary Trees: hierarchical clustering for large vocabularies
• Tree construction:
Vocabulary Tree
• Training: Filling the tree

What is the computational advantage of the hierarchical representation bag of words, vs. a flat vocabulary?

Bags of words: pros and cons
+ flexible to geometry / deformations / viewpoint
+ compact summary of image content
+ provides vector representation for sets
+ very good results in practice
- basic model ignores geometry – must verify afterwards, or encode via features
- background and foreground mixed when bag covers whole image
- optimal vocabulary formation remains unclear
Summary

- **Matching local invariant features**: useful not only to provide matches for multi-view geometry, but also to find objects and scenes.
- **Bag of words** representation: quantize feature space to make discrete set of visual words
  - Summarize image by distribution of words
  - Index individual words
- **Inverted index**: pre-compute index to enable faster search at query time