VK Multimedia Information Systems

Mathias Lux, mlux@itec.uni-klu.ac.at

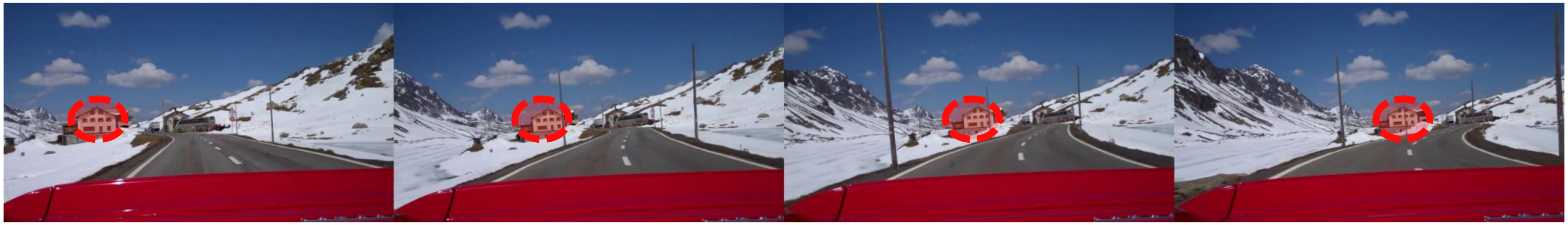
Dienstags, 16.oo Uhr c.t., E.1.42

Agenda

- Local features
- Bag of visual words
- Clustering

Local Features

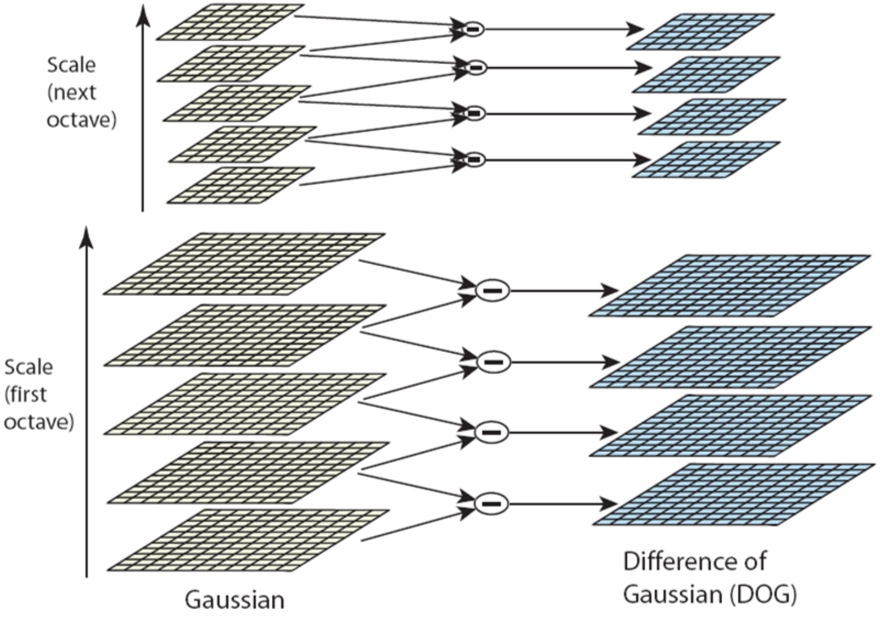
- Capture points of interest
  - Example: SIFT, SURF, …
  - Instead of global description
- Cp. Ferrari driving video
  - House moves over different frames



Feature Extraction

Scale space extrema detection

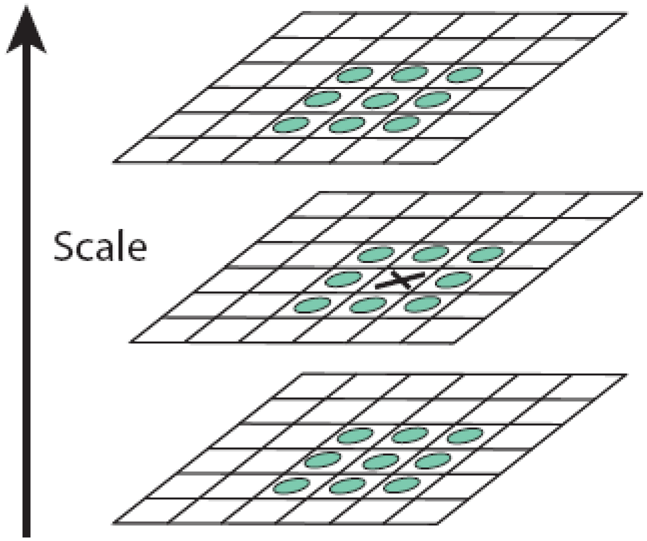
- Interest point identification
  - Difference of Gaussians
    - Use Gaussian blurred images at different octaves (resolutions)
    - Compute differences of adjacent blurred images pixel wise



Feature Extraction

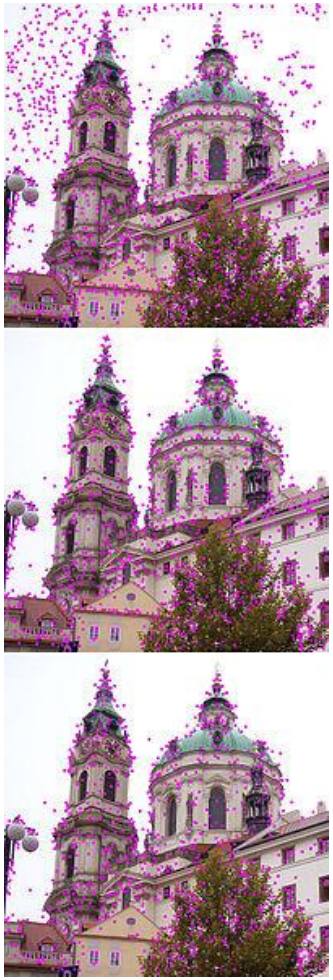
Scale space extrema detection

- Compare each pixel
  - 8 direct neighbours
  - 2x9 neighbours in different scales
- Find minima and maxima
- Which are considered candidate interest points



Feature Extraction

- Scale space extrema detection produces too many candidate interest points
- I.e. SIFT reduces by
  - – discarding low-contrast keypoints
  - – eliminating edge responses



src. Wikipedia http://en.wikipedia.org/wiki/File:Sift_keypoints_filtering.jpg

Feature Extraction

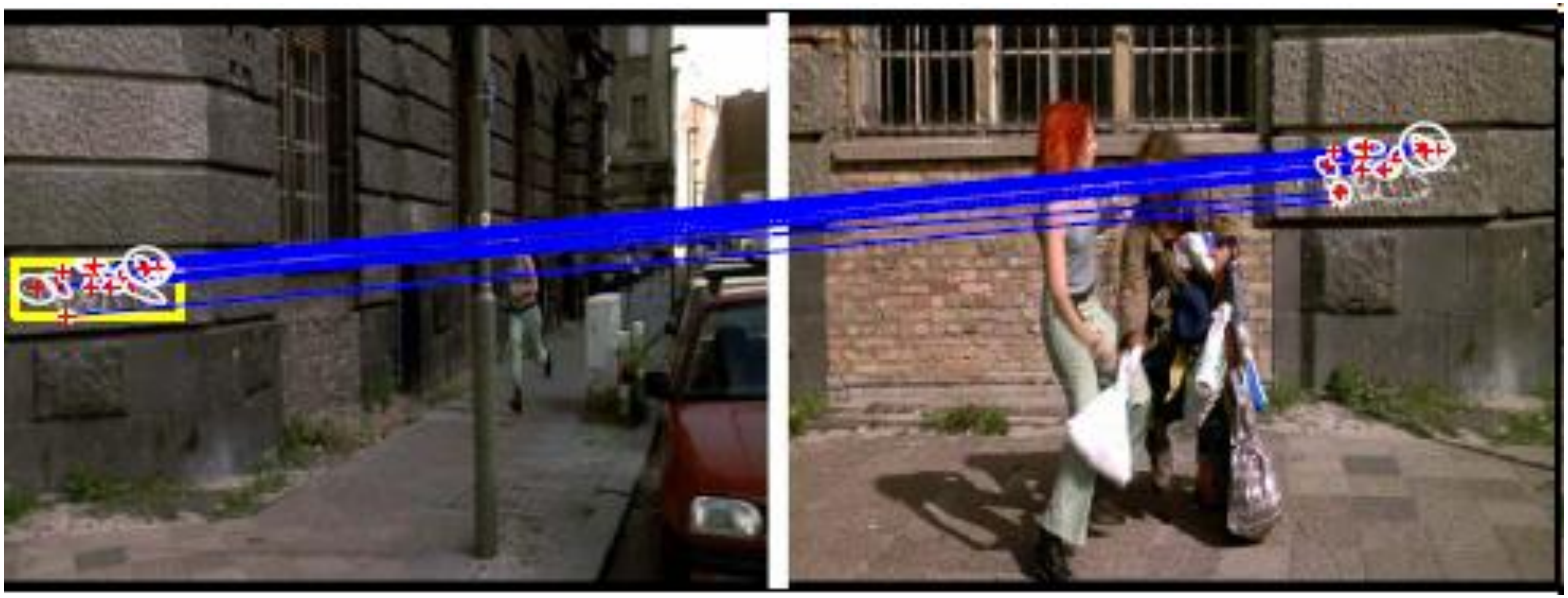
- Orientation assignment
  - based on local image gradient directions
  - achieves invariance against rotation
- Extraction
  - gradient magnitude at every scale
  - for all neighbouring pixels
  - gradient histogram with 36 bins
  - peaks are interpreted as main directions

Keypoint Descriptor

- Extracted from
  - scale of the keypoint
  - a 16x16 pixel neighborhood
  - gradient and orientation histograms
- Descriptor has 128 dimensions

Local Feature Matching

- Descriptors matching with L1, L2



Src. Sivic & Zisserman: Video Google: A Text Retrieval Approach to Object Matching in Videos, ICCV 2003, IEEE

Use Cases

- Image Stitching
  - creating panoramas from multiple images.
- 3D scene reconstruction
  - cp. Microsoft Photosynth
  - see http://photosynth.net/

Local Features

- Scale Invariant Feature Transform: SIFT
  - Lowe, David G. (1999). "Object recognition from local scale-invariant features". Proceedings of the ICCV 1999, pp. 1150–1157
- Speeded Up Robust Features: SURF
  - Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool, "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346--359, 2008
- Performance
  - Mikolajczyk, K.; Schmid, C. (2005). "A performance evaluation of local descriptors". IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (10): 1615–1630
- In detail lecture book
  - Kristen Grauman and Bastian Leibe: Visual Object Recognition, Morgan Claypool, Synthesis, 2011

Local Features

- Process can be adapted to specific needs
  - interest point / blob detection
    - Laplacian of Gaussian (LoG)
    - Difference of Gaussians (DoG)
    - Maximally stable extremal regions (MSER)
    - etc.
  - feature point description
    - SIFT, SURF, GLOH, HOG, LESH, …

Local Features in Java

- Java SIFT (ImageJ Plugin)
  - http://fly.mpi-cbg.de/~saalfeld/Projects/javasift.html
- jopensurf
  - http://code.google.com/p/jopensurf/
- MSER
  - Lire, net.semanticmetadata.lire.imageanalysis.mser.MSER
- OpenIMAJ
  - extensive library: http://www.openimaj.org/

Local Features in Applications

- OpenCV
  - – platform independent
  - – based on C
  - – build with cmake

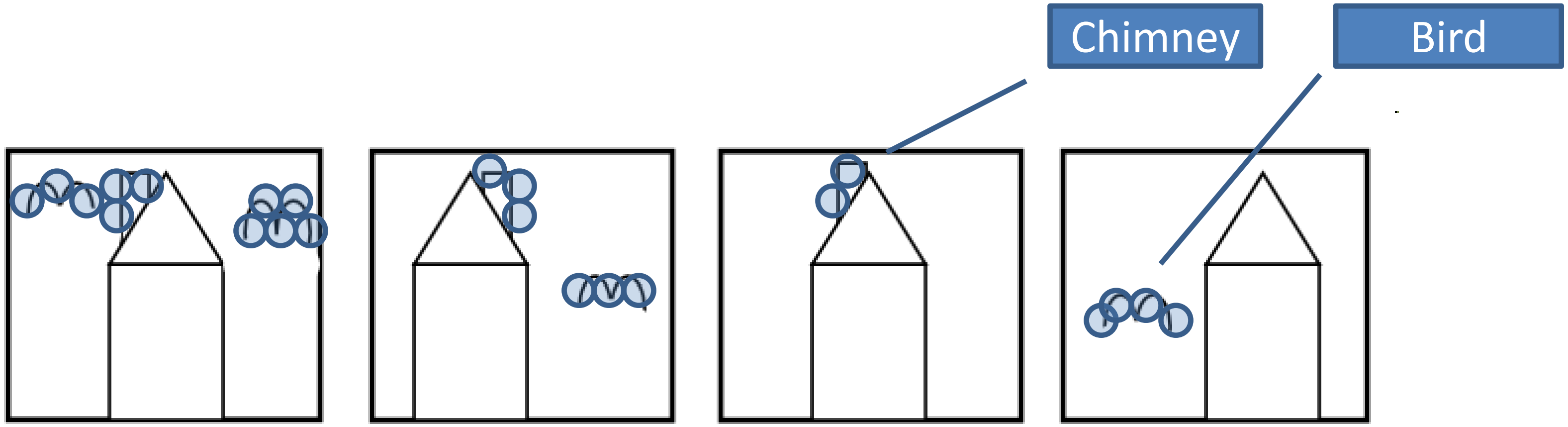
- http://opencv.willowgarage.com/wiki/

Bag of Visual Words

- Local features are computationally expensive
  - many features per frame / image
  - pair wise distance computation leads to a huge number of distance function calls
  - e.g. *n* features vs. *m* features -> _m*n_ distance function calls.

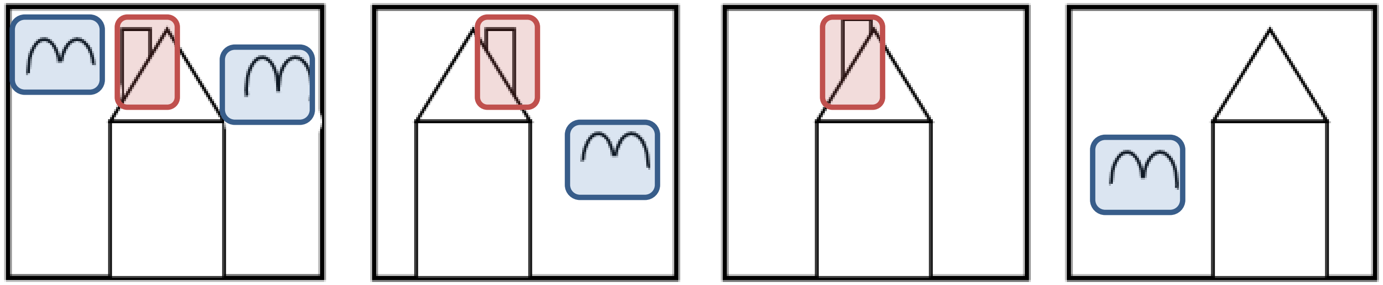
Bag of Visual Words

- Group similar local features
- Assign identifier to such a group



Bag of Visual Words

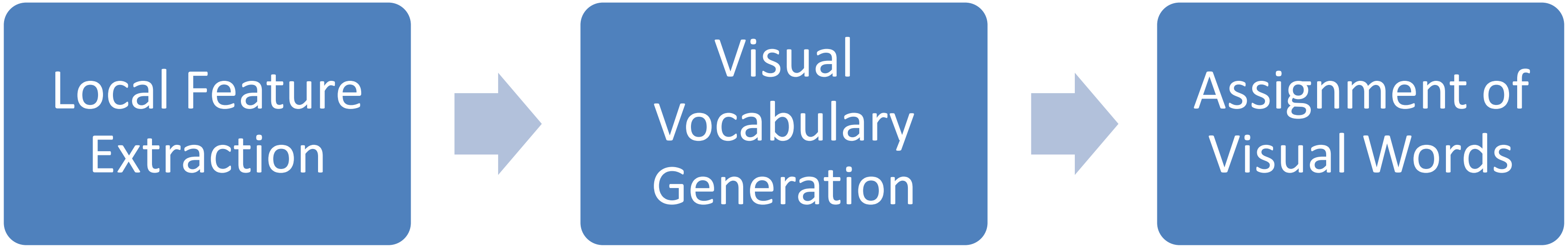
- Tag images containing features of group
  - {bird, bird, chimney}, {bird, chimney}, {chimney}, {bird}



Bag of visual words

- Groups are created unsupervised
  - not named, no semantic entities
  - model created is called visual vocabulary or codebook
- Group labels are called visual words
  - just a number, not a concept

BoVW Pipeline Overview



Local Feature Extraction

Visual Vocabulary Generation

Assignment of Visual Words

Local Feature Extraction

- Extract SIFT / SURF features
  - k_i >> 1 features for image I_i
  - the bigger the image the more features

Visual Vocabulary Generation

- Select representative sample
- Cluster the union set of features
  - to a pre-selected number of clusters

- Example: 1M images
  - Select 50,000 randomly
  - Cluster features of the 50k images

Assignment of Visual Words

- For each image I in the corpus
  - For each feature of I
    - Find the best matching cluster (center)
    - Assign visual word to the image

Best practice

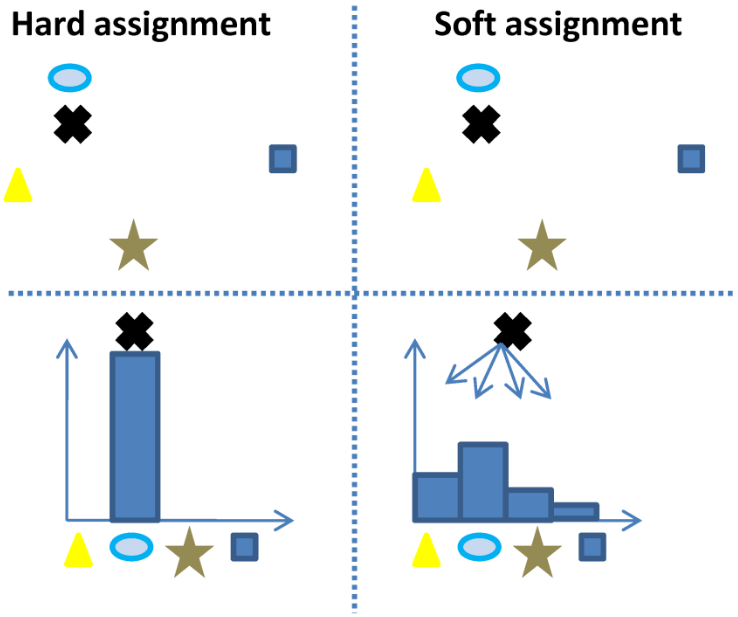
- Representative sample of documents
  - – random sampling
  - – up to a manageable number of features
- Vocabulary generation
  - – parallel or distributed implementation
  - – re-generate when necessary
- Assignment based on medians / medoids
  - – employ good index structure (e.g. hashing)

Example: SURF

- Simplicity data set
  - 1000 images, 10 categories, 100 images each
- SURF features (jopensurf)
  - 98 ms / image for extraction
- Vocabulary creation
  - 400 images,
  - with ~ 92.000 features (depends on sampling)
  - 10.000 clusters, ~ 2 minutes processing time

Fuzzyness

- fuzzy instead of binary assignments
  - one feature can express multiple visual words
  - based on a fuzzy membership function
  - also called “soft assignments”



Alternative Clustering Approach

- Fuzzy C-Means
  - add a feature to more than one cluster
  - adds robustness in terms of vocabulary size

Weighting

- TF works
- IDF not so well
- Distribution?