### Problem
- **3D models obtained from Structure-from-Motion**, every point has 2 SIFT descriptors [4]
- Nearest 2D-to-3D correspondences from query image to 3D model for pose estimation
- Feature matching: Effective (fast) and Efficient (many images registered)

### Direct vs Indirect Matching
- **Direct**: Effective but slow
  - Example: aprox. k-d tree-based search.
- **Indirect**: Efficient but not as effective
  - Example: Image retrieval-based method from Ischirna et al. [1]

### 3D point representations
- Different possibilities to represent 3D points by their descriptors inside visual words:
  - Use all descriptors for each 3D point
  - Compute mean/median of list of descriptors, assign it to all visual words (vws) activated by any of the points descriptors
  - Assign descriptors of point to visual words, compute mean/median if more than one descriptor of same 3D point assigned to same vw
  - Integer mean per vw: Round entries of means to nearest integer values

### Influence of the vocabulary
- **Generic vocabulary** from clustering SIFT descriptors from unrelated set of images [5]
  - Specific vocabularies offer significant improvement (slightly faster)
  - Experimented with 10k, 100k and 1M visual words, best results for 100k

### Related Work
- Geo-referenced version of Dubrovnik model, containing positions of query images
- Compute the average camera position over 10 repetitions for each query image
- For our methods: Report distance of average camera position to ground truth

### Localization accuracy
- **Ground truth**: Vienna images with SIFT features

### Dataset
- **Processed**: 3D SIFT descriptors (RANSAC vocabulary)
  - Feature matching with RANSAC: nearest neighbors are transformed to 3D coordinates

### Prioritized Search
- 2D-to-3D correspondences from linear search (through visual words) and SIFT ratio test
- **First search** by visual words, stop if N correspondences are found

### Pose Estimation using RANSAC
- **RANSAC** will take many samples if inlier-ratio is low (usually the case for rejected images)
- Assumes an inlier-ratio of 0.9/1.6 for RANSAC vs. 0.9 for GPU-based method by Ischirna et al. [1]

### Comparison with state-of-the-art
- **PF**: Prioritized feature matching in 3D points to 2D features proposed by Li et al. [3]
- **PF-F2P**: Match features against points if PF fails [5]
- Vocabulary tree-based methods: Retrieve the 10 top ranked images, perform pairwise matching and pose estimation
- **GPU-based method** by Ischirna et al. [1]: only available for the Vienna dataset

### Source code available at http://www.graphics.rwth-aachen.de/localization