Wrocław University of Technology

Localization of multiple near-duplicate fragments of images

Mariusz Paradowski[1,2], Andrzej Śluzek[2,3]

[1]Wrocław University of Technology, Wrocław, Poland

[2]Nanyang Technological University, Singapore

[3]Nicolaus Copernicus University, Toruń, Poland

Outline

- Image, sub-image and image fragment matching
- Matching local features
- Geometric approach
  - Affine transformation
  - Histogram of elementary transformations
  - Matching examples
- Topological approach
  - Topological constraint
  - Topological graph
  - Matching examples
- Results
- Summary

Image and sub-image matching

- Given a query image localize images containing the given query
- Mark the exact location of the query within the image
- Query image (or a query ROI) becomes a model
- Combination of local keypoints and global geometry
- Application of model based techniques is possible (e.g. RANSAC)

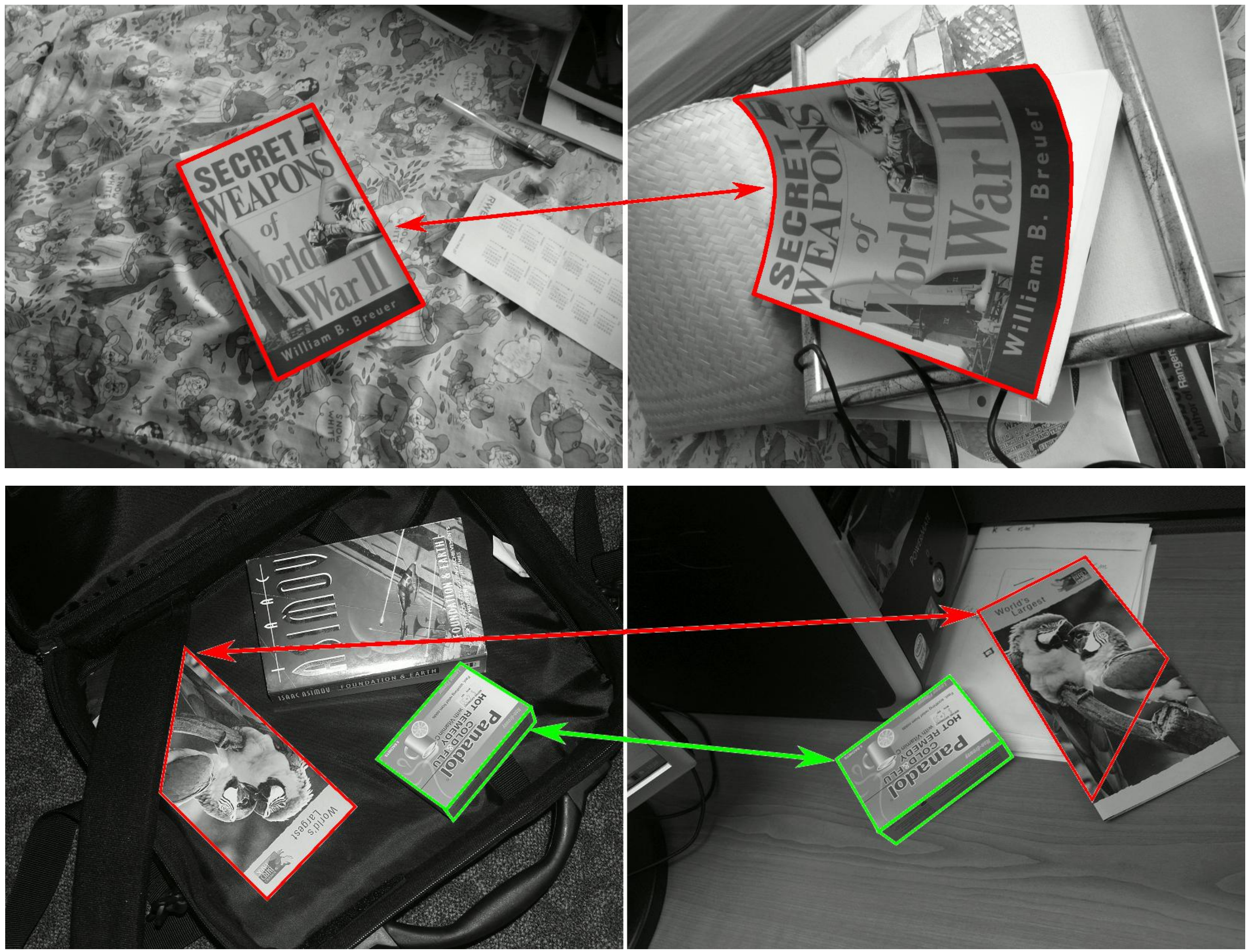
Image fragment matching

- Capturing similar fragments of images instead of whole (query) images
- Unknown number of similar image fragments
- Completely random content
- Inability to define a model of a fragment
- Inability to apply typical machine learning approaches

Problem definition

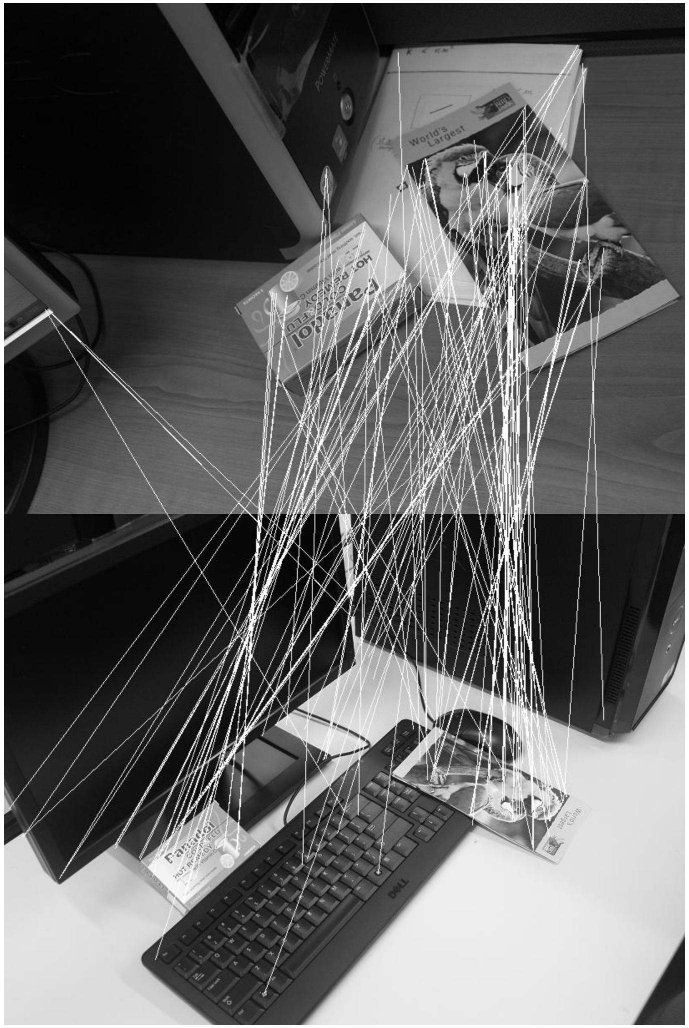
Given two **random images** (i.e. without **any *a priori* knowledge** about their content) find **near-duplicate fragments** of these images.

Problem definition - example



Matching of local features

- An image is described by hundreds of keypoints (e.g. SIFT)
- A keypoint describes a fragment of image
- Keypoints from two images are paired (e.g. mutual nearest neighbors)
- Only a small fraction of keypoint pairs is correct
- RANSAC approach fails due to model limitations
- **Correct keypoint pairs have to be found and grouped into objects**



The geometrical approach - background

- Global geometry is utilized to verify keypoint matching
- Localization of near-duplicate planar surfaces
- Application of affine geometry
  - Estimation of 6 parameters
  - Utilization of keypoint centers (stable but slower)
  - Utilization of ellipse parameters (faster but less stable)
  - Homography is locally approximated by affine transformation
- Making the data meaningful
- Decomposed transformations are probabilistic events
- Probability density function – histogram
- Some resemblance to Parzen window
  - Non-parametric approach
  - Iteration over the data

The geometrical approach - outline

- **Input:** Image I, Image J
- Calculate local keypoints (e.g. Harris-Affine + SIFT)
- Find nearest keypoint pairs between images
- Build many triangle pairs from keypoint pairs
- Calculate affine transformations
- **Decompose affine transformations**
- **Build histogram of affine transformations**
- **Find histogram peaks**
- Create fragments out of triangles found in peaks
- **Output:** Related image fragments

Affine transformation

Affine transformation:

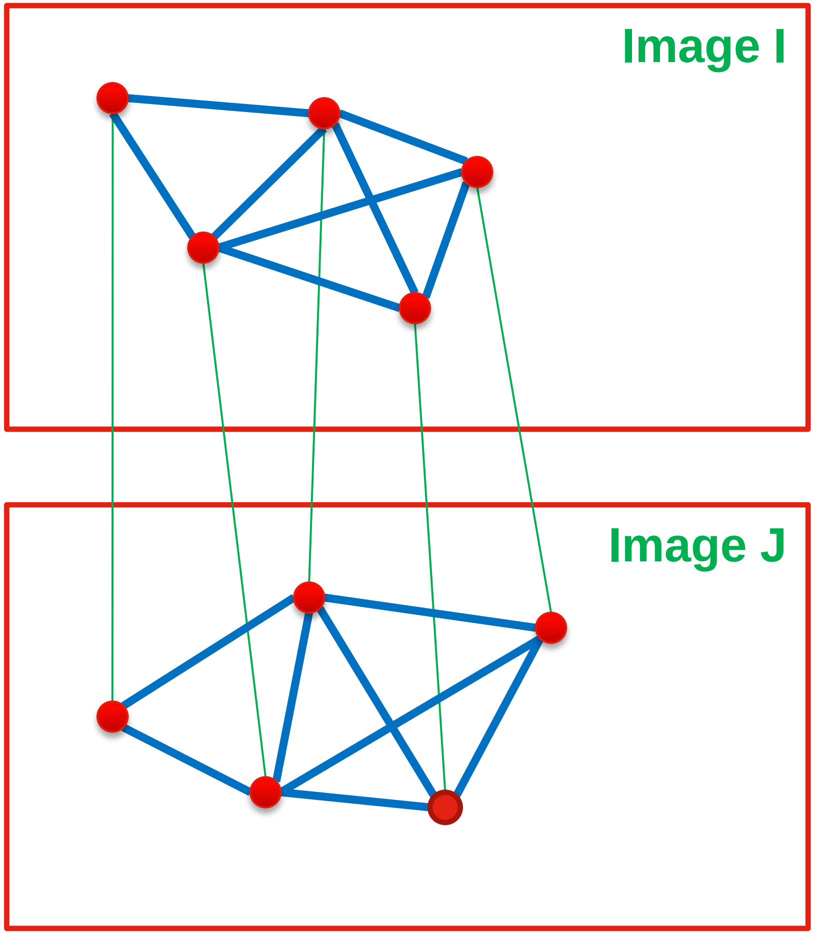
$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{A} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} A & B & C \\ D & E & F \\ 0 & 0 & 1 \end{bmatrix}$$

Reconstruction of affine matrix from 3 points:

$$\begin{bmatrix} A \\ B \\ C \\ D \\ E \\ F \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ v_1 \\ v_2 \\ v_3 \end{bmatrix} \begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 \\ x_2 & y_2 & 1 & 0 & 0 & 0 \\ x_3 & y_3 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_1 & y_1 & 1 \\ 0 & 0 & 0 & x_2 & y_2 & 1 \\ 0 & 0 & 0 & x_3 & y_3 & 1 \end{bmatrix}^{-1}$$

Matching triangles

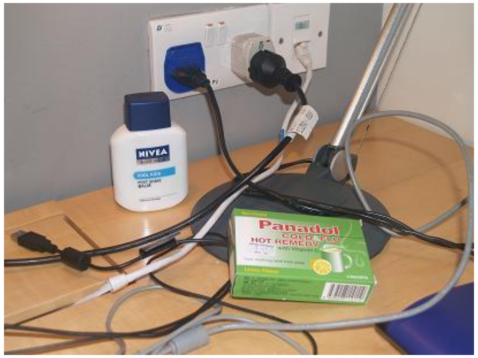
- Pairs of triangles are the elementary geometrical structure
- Single affine transformation from a pair of triangles
- $O(n^3)$ computational complexity (n – number of keypoints)
- $O(nm^2)$ computational complexity (m – size of keypoint neighborhood)
- Number of triangles: 100.000 - 1.000.000
- Triangles can be replaced by ellipse pairs
  - O(nm) complexity
  - Faster approach
  - Less precise due to imprecise ellipses

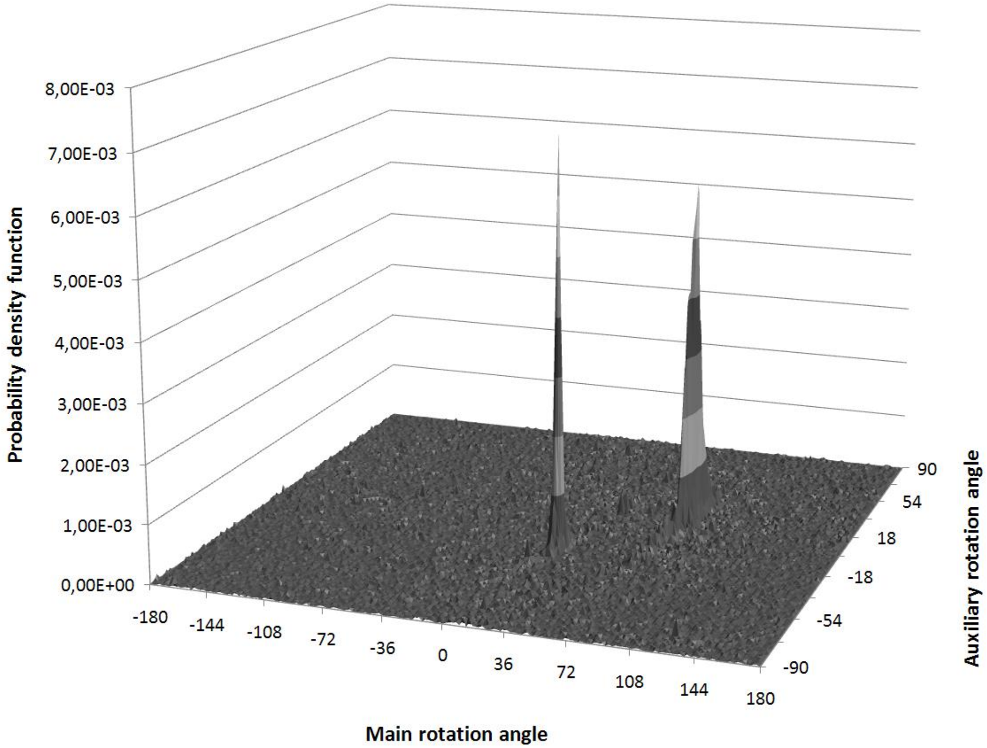


Transformation decomposition

- Affine transformation has 6 degrees of freedom
- Parameters of the transformation are entangled
- Interpretation of transformation parameters is difficult
- Transformation decomposition allows to get the underlying meaning
- Recreation of elementary transformations makes reading the affine transformation easier
- SVD decomposition
  - 2 rotations (main rotation, auxiliary rotation)
  - 2 translations (OX, OY)
  - 2 scales (OX, OY)
- 3D decomposition
  - 3 rotations (OZ, OX, OZ)
  - 2 translations (OX, OY)
  - 1 scale (OZ)

Histogram of transformations

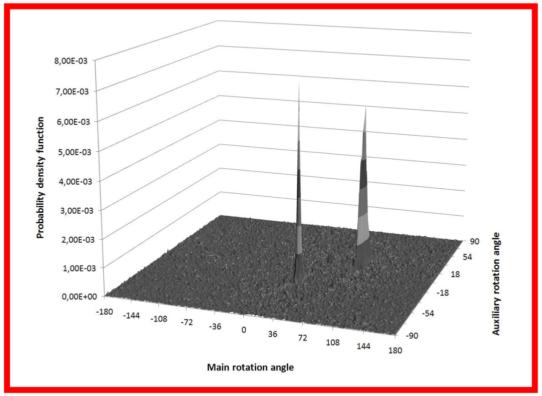




Joint probability of elementary geometric transformations $P(s_x, s_y, t_x, t_y, \alpha, \beta)$

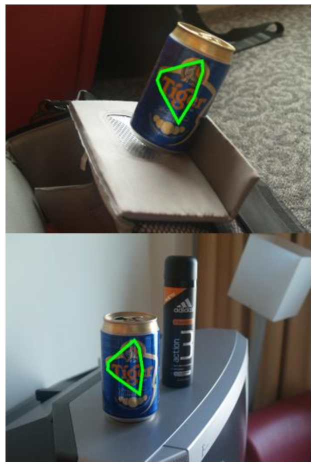
From the histogram to near-duplicate fragments

- Histogram of elementary transformations is huge
- Hash-tables for efficient storage
  - Memory complexity $O(nm^2)$
  - Computational complexity $O(nm^2)$
- Histogram peaks represent near-duplicate fragments
- Non-duplicate keypoint pairs are a noise
- Threshold based histogram analysis
- Threshold t = 10 (fully sufficient)
- Merging triangles from neighboring bins
- Construction of convex hulls



Some matching examples

Road sign



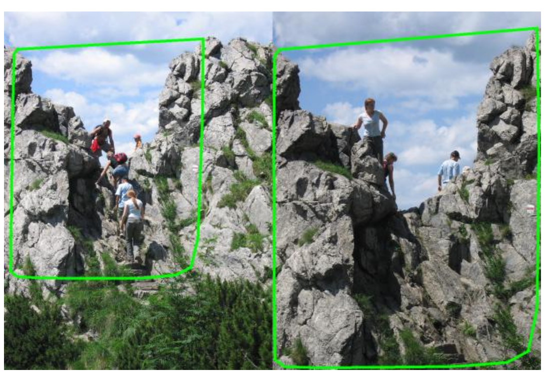
Can (non-planar)

Box and a bottle

Background clutter

Repeated fragments

Different side of a tower



Zoom and details change

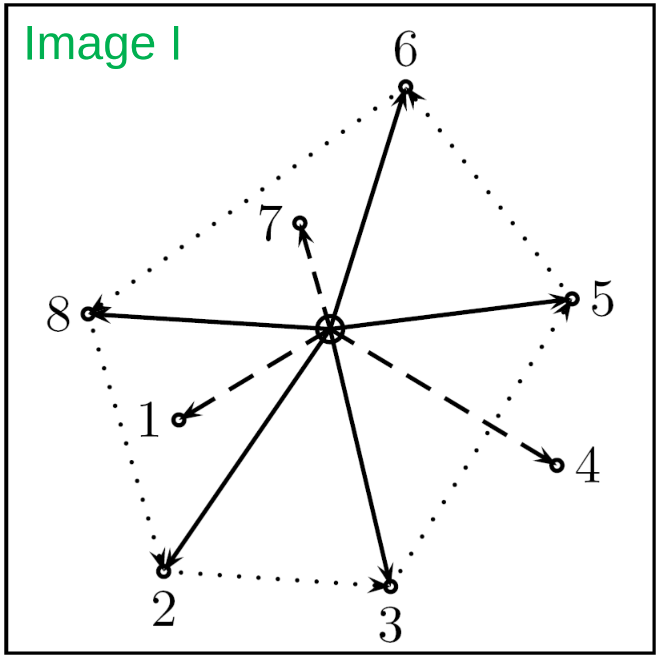
The topological approach - background

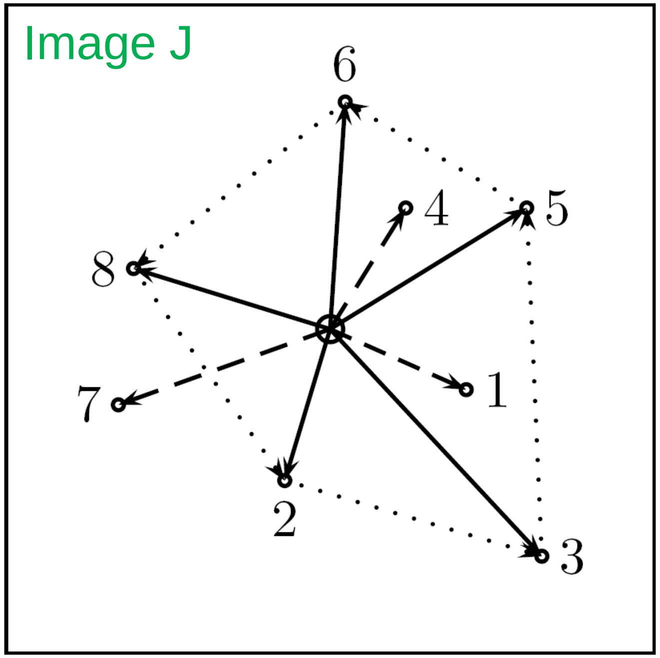
- Geometrical approach is strict and sound but not flexible
- Non-linear geometric models become complex to estimate
- Application of image topology instead of geometry (*Forget about numbers, focus on relations*)
- Verification of semi-local image topology
  - Keypoint pairs
  - Spatial neighborhood of keypoints
  - Relations between neighboring pairs
  - Constraints on neighboring pairs

The topological approach - outline

- **Input:** Image I, Image J
- Calculate local keypoints (e.g. HarAff, SIFT, MSER,SURF)
- Find nearest keypoint pairs between images
- For each pair determine its spatial neighbors
- **Verify topological constraint for each pair**
- **Filter out all pairs not following the constraint**
- **Connect pairs according to topological data**
- **Build image fragments out of pair connected components**
- **Output:** Related image fragments

The topological invariant - order of angles

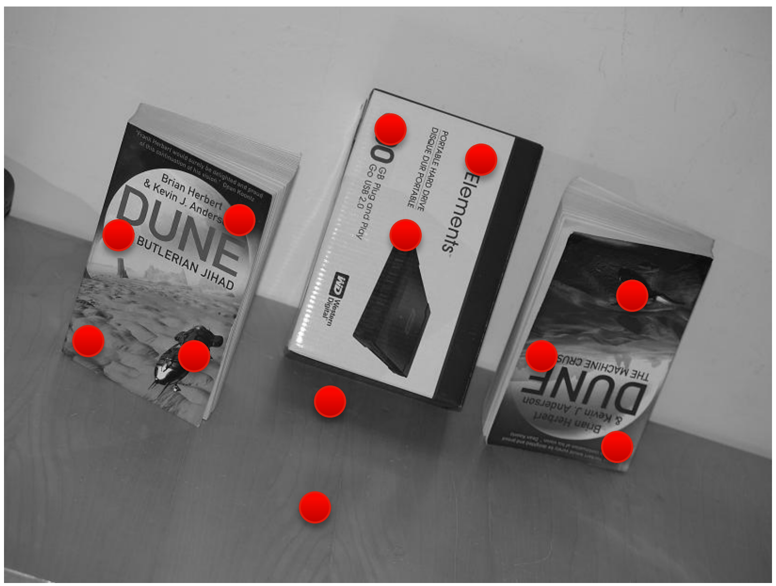


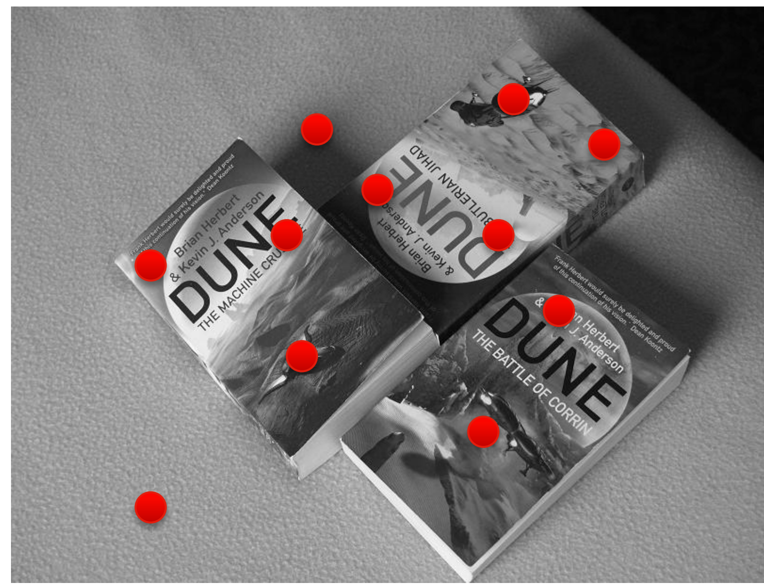


Order of angles of vectors to neighboring points has to be consistent.

Largest subset of consistent neighbors is found.

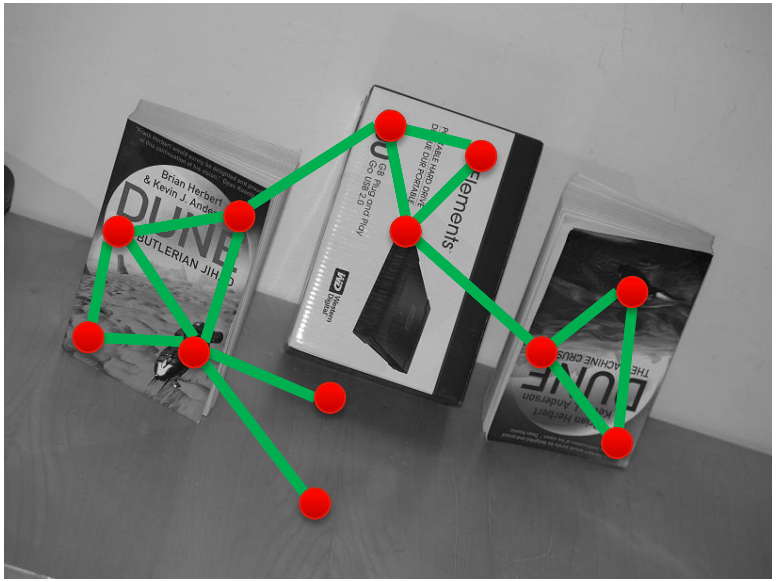
Construction of the topological graph - keypoint pairs

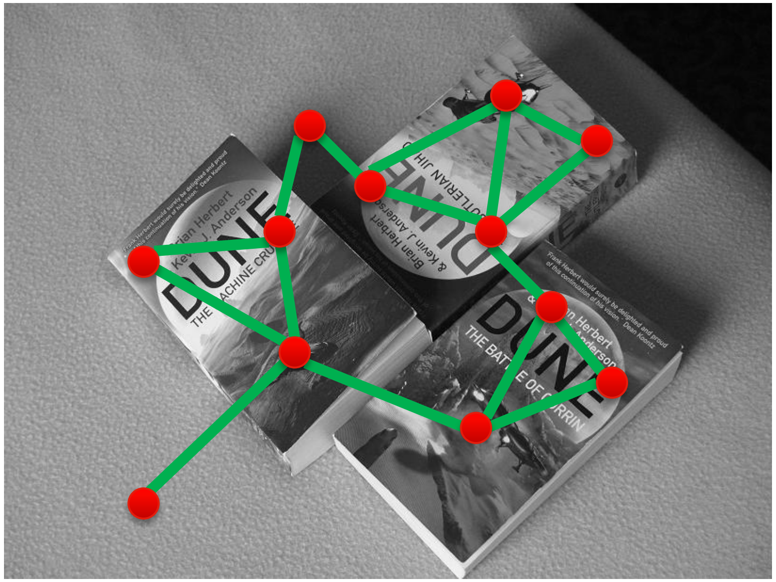




Keypoints are detected and paired

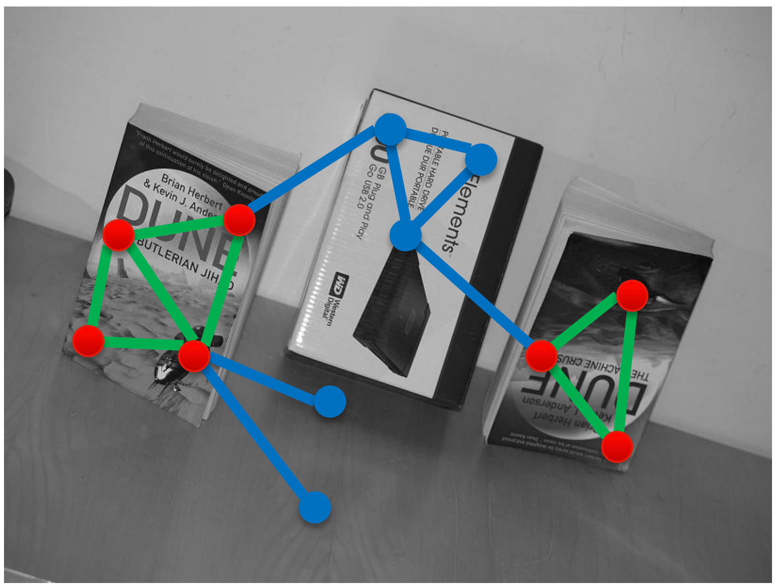
Construction of the topological graph - spatial neighborhood

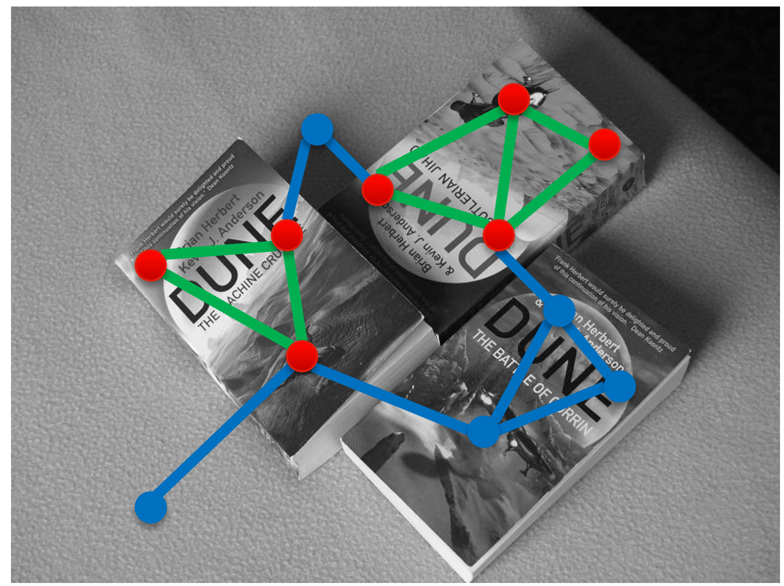




Spatial neighbors are found.
Each keypoint pair gets a set of neighboring pairs.

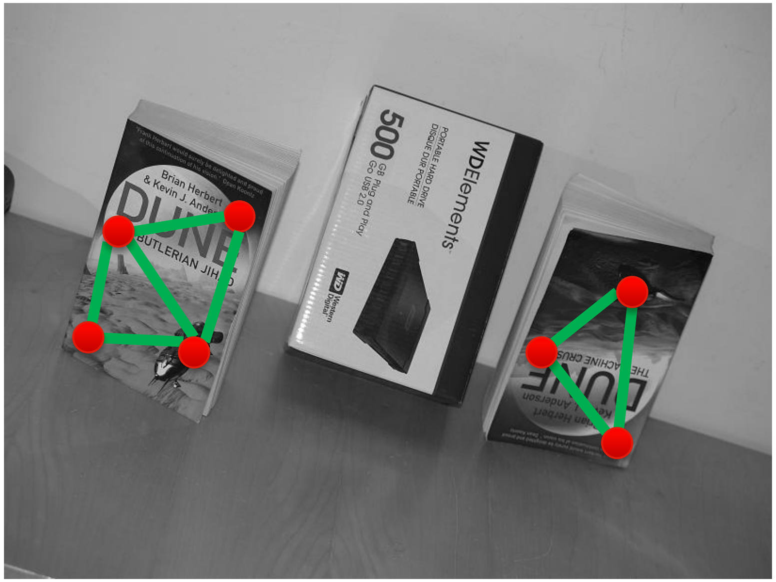
Construction of the topological graph - topological filtering

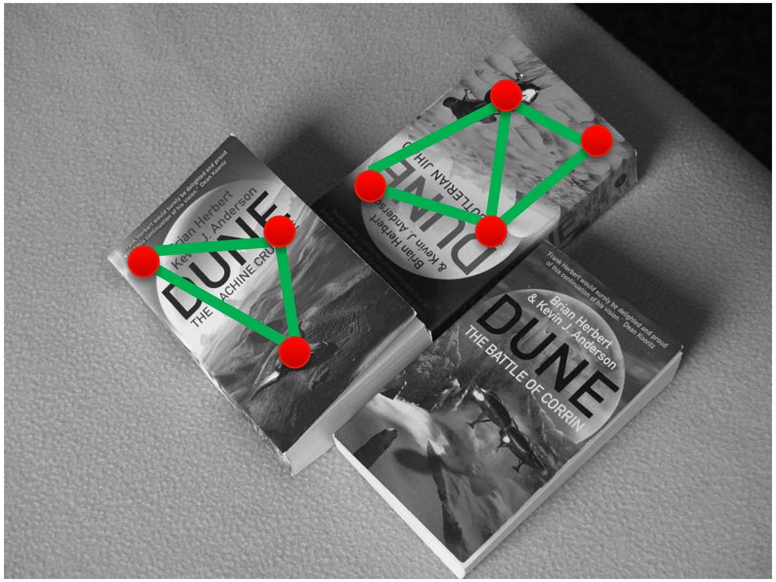




Topological constraints are verified.
For each keypoint pair consistency is checked.
Inconsistent keypoint pairs are marked for removal.

The topological graph - filtered and connected keypoints

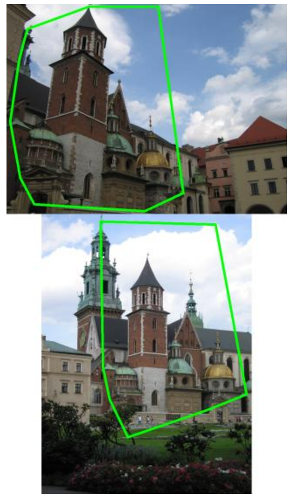




Nodes and edges are removed.
Near-duplicate fragments are found using graph connected component search.

Some matching examples

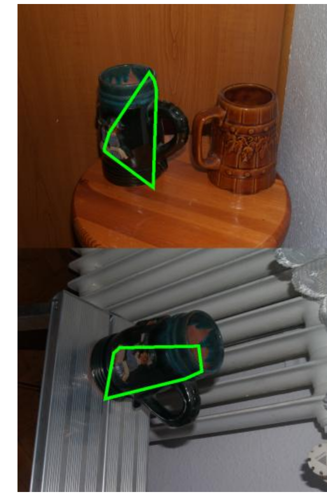
Two objects



Viewpoint change

Deformed book

Deformed bag



Cup (non-planar)

Different camera position

Different camera position

Evaluation - fragment matching dataset

Quality measure	HarAff SIFT	HarAff GLOH	SURF	MSER SIFT
Geometry				
Precision [area]	0.96	0.96	0.90	0.95
Recall [area]	0.64	0.50	0.49	0.53
Precision [object]	0.97	0.97	0.98	0.94
Recall [object]	0.81	0.71	0.61	0.68
Topology				
Precision [area]	0.64	0.62	0.50	0.71
Recall [area]	0.79	0.74	0.70	0.63
Precision [object]	0.98	0.97	0.97	0.98
Recall [object]	0.92	0.88	0.79	0.78

Results on Oxford5k database

- Query images with ROI (model based approach)
- Our approaches are taken „as is”
- Query image ROI is not used
- Ranking by
  - detected region size
  - detection threshold (only geometrical)
- Retrieval quality
  - Precision is high
  - Recall is the biggest challenge
  - Mean averaged precision is comparable to state of the art

Retrieval method	mAP
Bag of words	0.618
Bag of words + spatial	0.645
WGC, no prior	0.383
WGC + HE + prior	0.547
HE + WGC + weights + MA	0.615
S4E12	0.789
S4E12 + QE	0.901
Geometrical approach	**0.628**
Topological approach	**0.715**

Possible applications

- Image retrieval (e.g. Oxford5k database)
- Visual objects formation (clustering)
- Matching based annotation (classification)
- Human face identification
- Vision-based vehicle navigation

Summary

- Image fragment matching
- Local approach – matching keypoints (e.g. HarAff, SURF, MSER, SIFT)
- Image geometry and topology for validation
  - Non-parametric affine-based geometric approach
  - Local topological constraint and the topological graph
- Very few requirements
  - **No query image region of interest**
  - **No object/fragment model definition**
  - **No training routines, no training datasets**
  - **All presented examples are done with default parameters**
- Very high precision, lower recall
- Resistance to high background clutter
- On-line, camera-based fragment matching

Thank you !

Any questions?