Multi-resolution image recognition

Jean-Baptiste Boin
Roland Angst
David Chen
Bernd Girod
Outline

• Scale distribution
• Presentation of two different approaches and experiments
• Analysis of previous results
Motivation

- Typical image retrieval applications: similar resolution in database images and queries
- Performance drops when the resolutions are very different (high-res database image vs. low-res query)
- OK for some applications (product recognition), not ideal for others (large painting recognition)
Scale distribution – derivation

- Goal: Find the average distribution of scales for an “ideal” feature detector
- Hypotheses: continuous representation of the scales + a few assumptions on the feature detectors
Scale distribution – derivation

\[ \rho(s) \propto 1/s^3 \]

\[ F(s) = \begin{cases} 
0 & \text{if } s \leq s_0 \\
1 - s_0^2/s^2 & \text{if } s \geq s_0 
\end{cases} \]

- \( F(2s_0) = 0.75 \) – Qualitative justification

\{s_0 \leq s \leq 2s_0\} contains \( N-N/4 = 3N/4 \) features
Dataset used for the experiments

- Images extracted from a public art repository (Web Gallery of Art): more than 30,000 images
- We keep cropped regions of fixed size (17,146 images at resolution 1024x768)
- We generate queries of half the size of the database images (512x384) by rotating/scaling/translating

(zoom factor defined before downsampling)
Scale distribution – experiment

Queries (2x) → SURF features → Match + RANSAC → all extracted features

Reference images → SURF features → Match + RANSAC → matching features
Scale distribution – results

Effect of discretized scale detector
Scale distribution – results

- Power law: a bit of a stretch, but gives a rough idea of the behavior
Baseline (single REVV)

- Aggregate 250 SURF features with coarsest scale
Tile based approach

• Each DB image is represented by 5 tiles

Reference images (5 tiles per orig. image)

Queries

h
w

h
w

Multi-resolution image recognition
Scale based aggregation – Idea

- Main conclusion from previous analysis: most features have a scale in the interval $[s_0, 2s_0]$
  - 75% in theory
  - ~70% in practice (depends on size of query)
- Aggregate features according to scale domains and position
Scale based aggregation – Idea

Scale based aggregation – Idea

Multi-resolution image recognition
Scale based aggregation – Idea

query

\[ s > s_1 \]

\[ s_2 < s < s_1 \]

\[ s_3 < s < s_2 \]

...
Scale based aggregation – Experiment

• Database side: case of limited scale variation, we only consider 2 levels

\[
\begin{array}{c|c}
\text{s > } 2s_0 & \text{} \\
\hline
\text{s < } 2s_0 & \text{s < } 2s_0 \\
\end{array}
\]

• Query side: we only have 2 bins (s > 2s_0 and s < 2s_0)
Multi-resolution image recognition

Scale based aggregation – Experiment

Queries

Reference images

SURF

REVV \( (s \leq s_0) \)

REVV \( (s \geq s_0) \)

REVV \( (s \leq s_0) \)

REVV \( (s \leq s_0) \)

REVV \( (s \leq s_0) \)

REVV \( (s \leq s_0) \)
Scale based aggregation

- How do we merge the two lists?
  1. By cheating: we take the “best” rank in each list
  2. By using the correlation scores to re-rank the results
  3. By using a linear combination of the best correlation score for each image
Multi-scale experiments

- Zoom = 2x (query represents ~25% of original image)
Multi-scale experiments

- Zoom = 1.5x (query represents ~44% of original image)
Multi-scale experiments

- Zoom = 1x (query represents ~100% of original image)
Analysis of results

• Current problem of our approach: hard-binned scale (assumes good reproducibility of scale extraction)

• Justification of the good results obtained in the tiling approach: REVV and surface overlap
Scale reproducibility – experiment

Queries (2x) → SURF features → Match + RANSAC → Reference images

log scale of matching features in both images
Scale reproducibility – results

[Graph showing data points with a trend line labeled 'additive noise']

Multi-resolution image recognition
Scale reproducibility – results

![Graph showing additive noise](image)

**Multi-resolution image recognition**
REVVV and surface overlap – experiment

Queries

Reference images

SURF features → REVVV

Upper-left tile, Upper-right tile, Lower-left tile, Lower-right tile, Downsamp. tile

log rank (for each type of tile)
REVV and surface overlap – results

Precision at rank 10 for each type of tile

Strongly correlated with area overlap

50% overlap: ~90% precision at rank 10

25% overlap: ~70% precision at rank 10
Conclusion of multi-resolution exploration

- Considerable unsolved issues
  - Scale reproducibility (try other values of thresholds)
  - Increased cost of running 2 queries, but no real gain in non-optimal conditions
- The simpler (tile-based) approach is “too good”
  - Shows the robustness of REVV
- Hybrid approach?
Conclusion of multi-resolution exploration

- Hybrid approach
Conclusion of multi-resolution exploration

• Hybrid approach

\[ s > 2s_0 \quad \text{query} \]

\[ s > 2s_0 \quad s > 2s_0 \]

\[ s < 2s_0 \quad s < 2s_0 \]

\[ s < 2s_0 \quad s < 2s_0 \]

\[ 2s_0 \]

\[ s_0 \]

→ drawback: doubles the storage requirement
→ possibility to reduce computation amount