Mobile Visual Search

Bernd Girod
Stanford University
bgirod@stanford.edu
Mobile Visual Search
Visual Bookmarks for Movies and TV Shows
Mobile Visual Search Applications

- Wine Labels
- Museum Guide
- Ads/Catalogs
- Comparison Shopping
- Movie Posters
- Landmarks
- Real Estate
- Media covers (CD, DVD, books)
Outline

• Computer vision: “Bag of Words” matching
• Feature compression
• Phone-based vs. server-based processing
• Demos (during the break)
Computing Visual Words

\[ \sum d_x \]
\[ \sum d_y \]
\[ \sum |d_{x1}| \]
\[ \sum |d_{y1}| \]

Oriented Patch

Blob Response
“Bag of Words” Matching & Geometric Consistency Check
Vocabulary Tree

Features
Architecture A: Send Image

Image

Wireless Network

Information

20 kbps → 20 sec

Camera  Client  Server  Feature Extraction  Feature Matching
Architecture B: Send Features

Features → Wireless Network

Camera
Feature Extraction
Feature Coding
Client

Server
Feature Matching
Architecture C: Features on Phone

Wireless Network

Features

Information

Camera

Feature Extraction

Feature Matching

Client
GPS-Aided Landmark Recognition
Transform Coding of SURF/SIFT Descriptors

- Original Descriptor → Orthonormal Transform → Quantizer → Entropy Coder
- Channel
- Client
- Server
- Entropy Decoder → Inverse Quantizer → Inverse Transform → Decoded Descriptor

[Chandrasekhar et al., VCIP 2009]
CHoG: Compressed Histogram of Gradients

- Patch
- Gradients
- Gradient distributions for each bin
  - Spatial binning
  - Histogram compression
  - CHoG Descriptor

Gradient distributions for each bin.
Location Histogram Coding

Feature Locations \((x,y)\) → Spatial Binning → Context-based Arithmetic Coding

Quantize

Refinement Bits

[Tsai et al., MobiMedia 2009]
Compressed Feature Vector

Size (bits)  52  59  84  1024  1088

SIFT  Location x,y  1088 bits

CHoG  Location x,y  ~ 84 bits

Compressed x,y → CHoG  ~ 59 bits

[Tsai et al., MobiMedia 2009]
Nokia N95
330 MHz ARM
64 MB RAM

Timing Analysis

Uplink: 20 kbps
0.5 sec RTT

Execution Time (sec)

Server Delay

Upload Image
40 kByte

“Send Features”

“Send Image”

Upload Features
2.2 kByte

Extract Features
Timing Analysis

Nokia N95
330 MHz ARM
64 MB RAM

Uplink: 60 kbps
0.5 sec RTT

Execution Time (sec)

“Send Features”
“Send Image”

Server Delay
Upload 2.2 kByte
Extract Features

Upload Image
40 kByte
**Timing Analysis**

Nokia N95
330 MHz ARM
64 MB RAM

- **Uplink:** > 1 Mbps
- **WLAN:** 0.1 sec RTT

Execution Time (sec)

- **Server Delay**
- **Extract Features**

"Send Features" to "Send Image"
Mobile Augmented Reality

Send Query Frame

Extract Features

Query VocTree

Check Geometry

Send ID and Geometry

Compensate Camera Pose

Display ID and Draw Boundary

John Mayer Inside Wants Out

Track Camera Pose

Low Motion

High Motion

Time
Mobile Augmented Reality: Books
Mobile Augmented Reality: CDs
Concluding Remarks

• Mobile Visual Search is ready for prime-time:
  – Bag of Words approach: > 95% recall for > 1M database
  – Vocabulary trees for fast retrieval (1 sec for > 1M database)

• Feature compression is a key problem
  – Sending a JPEG image over wireless link can be rather slow
    ➔ send salient image features to server
  – Small database: send compressed database features to phone
  – CHoG outperforms SIFT and SURF in rate-constrained recognition performance: ~60 bits incl. location
  – Feature extraction ~1 sec on 330 MHz smartphone
  – MPEG is considering standardization