Large-scale 3D Modeling from Crowdsourced Data

Jan-Michael Frahm, UNC
Enrique Dunn, SIT
Marc Pollefeys, ETHZ, Microsoft
Jared Heinly, URCV
Johannes Schönberger, ETHZ
Large-scale Modeling Pipeline

Jared Heinly

URCV
urcventures.com
Large-scale 3D Modeling from Crowdsourced Data
Outline

• Challenges and opportunities of crowd-sourced data
• Image retrieval and indexing
• Connected component discovery
• Reconstruction of connected components
• Design of a real pipeline
• Lessons learned
Challenges and Opportunities of Crowd-sourced Data
What is crowd-sourced data?

• User-driven
What is crowd-sourced data?

• User-driven
What is crowd-sourced data?

- User-driven
What is crowd-sourced data?

• User-driven
What is crowd-sourced data?

- User-driven
What is crowd-sourced data?

• User-driven
What is crowd-sourced data?

- Diverse
What is crowd-sourced data?

• Diverse
Opportunities

• Diversity – Space

Li, Snavley, Huttenlocher, Fua, ECCV 2012
Opportunities

• Diversity – Time

Matzen, Snavley, *ECCV 2014*
Opportunities

• Real-time
Opportunities

• Real-time
Opportunities

• Representation of interests of camera/internet-enabled humanity
Challenges

• Diversity – Viewpoint
Challenges

• Diversity – Illumination
Challenges

- Diversity – Weather
Challenges

• Diversity – Quality
Challenges

- Diversity – Appearance
Challenges

• Diversity – Ambiguity
Challenges

- Sparsity – Space
Challenges

• Sparsity – Time
Challenges

Daily Number of Photos Shared on Select Platforms, Global, 2005 – 2015

Meeker, KPCB Internet Trends 2016
Where can we get the data?

flickr  Mapillary

twitter  Google
What are some sample datasets?

- Yahoo Flickr Creative Commons 100M
  - 14 TB
  - 640x480 resolution
  - Half have geotags
What are some sample datasets?

- Yahoo Flickr Creative Commons 100M
What are some sample datasets?

• Cornell BigSFM
  • Datasets with 1 – 16K images
Outline

• Challenges and opportunities of crowd-sourced data
• Image retrieval and indexing
  • Connected component discovery
  • Reconstruction of connected components
• Design of a real pipeline
• Lessons learned
Image Indexing and Retrieval
3D Modeling Pipeline

- **Data Association**
  - **Biggest Challenge**
  - **Performs at Current Scale**
- **Sparse Modeling**
- **Dense Modeling**
What is image indexing?
What is image indexing?
What is image retrieval?
Methods of Image Indexing

- Global descriptor

- Local descriptors
Global Descriptors

- Tiny image
Global Descriptors

• Tiny image

[0.12, 0.85, 0.37, ...]  [0.16, 0.82, 0.41, ...]  [0.09, 0.88, 0.36, ...]  [0.11, 0.79, 0.33, ...]
Global Descriptors

• Tiny image

[0.12, 0.85, 0.37, ...] → [0.43, 0.51, 0.34, ...]
[0.64, 0.30, 0.82, ...]
[0.16, 0.82, 0.41, ...]
[0.09, 0.88, 0.36, ...]
[0.94, 0.05, 0.57, ...]
[0.48, 0.24, 0.60, ...]
[0.11, 0.79, 0.33, ...]
...
Global Descriptors

• Tiny image

[0.12, 0.85, 0.37, ...]  
[0.43, 0.51, 0.34, ...]  
[0.64, 0.30, 0.82, ...]  
[0.16, 0.82, 0.41, ...]  
[0.09, 0.88, 0.36, ...]  
[0.94, 0.05, 0.57, ...]  
[0.48, 0.24, 0.60, ...]  
[0.11, 0.79, 0.33, ...]  
...
Global Descriptors

• GIST

Oliva and Torralba, *IJCV 2001*
Global Descriptors

• GIST

[0.12, 0.85, 0.37, 0.40, ...]

Oliva and Torralba, *IJCV 2001*
Global Descriptors

• Search methods
  • Brute force

[0.43, 0.51, 0.34, ...]
[0.64, 0.30, 0.82, ...]
[0.16, 0.82, 0.41, ...]
[0.09, 0.88, 0.36, ...]
[0.94, 0.05, 0.57, ...]
[0.48, 0.24, 0.60, ...]
[0.11, 0.79, 0.33, ...]
Global Descriptors

- Search methods
  - kd-tree
Global Descriptors

• Search methods
  • Hashing

Large-scale 3D Modeling from Crowdsourced Data
Local Descriptors
Local Descriptors

• 2D Features
  • Real-valued descriptors
    • SIFT
    • SURF
  • Binary descriptors
    • BRIEF
    • BRISK
    • FREAK

[0.03, 0.15, 0.03, 0.92, ...]

[0.13, 0.55, 0.47, 0.02, ...]

[0.70, 0.05, 0.32, 0.01, ...]

[0, 1, 1, 0, 1, 0, 0, 1, ...]

[1, 0, 1, 0, 0, 1, 1, 0, 1, ...]

[1, 1, 1, 0, 1, 1, 0, 1, 0, ...]
Local Descriptors

- [0.03, 0.15, 0.03, 0.92, ...]
- [0.13, 0.55, 0.47, 0.02, ...]
- [0.70, 0.05, 0.32, 0.01, ...]
- [0.84, 0.26, 0.72, 0.98, ...]
- [0.08, 0.11, 0.57, 0.36, ...]
Visual Vocabulary

The quick brown fox jumped...
Visual Vocabulary

[0.81, 0.41, ...] → 6142

[0.95, 0.01, ...] → 708

[0.23, 0.74, ...] → 97503

[0.37, 0.08, ...] → 3818

[0.88, 0.61, ...] → 282734

[0.50, 0.92, ...] → 68185
Visual Vocabulary

Large-scale 3D Modeling from Crowdsourced Data
Vocabulary Tree

[0.03, 0.15, 0.03, ...]

Hierarchical K-Means

0 1 2 3 4 5 6 7
Bag of Words

[183, 976, 4385, 97121, 11802]

[976, 1018, 4385, 11802, 187049]
Bag of Words

* Dot Product
Binary Vocabulary

• Hierarchical K-Means not as effective
  • Trzcinski, Lepetit, Fua, *Pattern Recognition Letters 2012*

• Vocabulary construction
  • K-Medoids clustering

• Vocabulary retrieval
  • LSH (Locality Sensitive Hashing)
Image Indexing and Retrieval
Outline

• Challenges and opportunities of crowd-sourced data
• Image retrieval and indexing
• Connected component discovery
  • Reconstruction of connected components
• Design of a real pipeline
• Lessons learned
Connected-Component Discovery
Connected Component Discovery
Connected Component Discovery
Connected Component Discovery

• Discovery strategies
  • Brute force
  • Image retrieval
  • Clustering
  • Streaming
Component Discovery: Brute Force

• Naive
Component Discovery: Brute Force

• Preemptive matching
  • Wu, 3DV 2013

• Determine candidate image pairs to verify
Component Discovery: Brute Force

• Two-view geometry classification
  • Schönberger, CVPR 2015, GCPR 2015

• Determine candidate image pairs to verify
  • Image pair descriptor
Component Discovery: Brute Force

• Two-view geometry classification
  • Schönberger, CVPR 2015, GCPR 2015
• Determine candidate image pairs to verify
  • Image pair descriptor

Quantization

Prediction

Random forest classifier
Component Discovery: Image Retrieval

• K-nearest neighbor retrieval

![Example images](image1.png) → ![Example images](image2.png)

![Example images](image3.png) → ![Example images](image4.png)

![Example images](image5.png) → ![Example images](image6.png)
Component Discovery: Image Retrieval

- MatchMiner
  - Lou, Snavely, Gehrke, *ECCV 2012*
  - Prioritize retrieval to maximize length of transitive links
Component Discovery: Image Retrieval

• MatchMiner
  • Lou, Snavely, Gehrke, *ECCV 2012*
  • Prioritize retrieval to maximize length of transitive links

![Diagram of image retrieval process with nodes and links indicating query image, verified true matches, candidate pairs with varying similarity levels.](image-url)
Component Discovery: Image Retrieval

• VocMatch
  • Havlena and Schindler, *ECCV 2014*
  • Treat visual words as direct feature matches
Component Discovery: Image Clustering

• Building Rome on a Cloudless Day
  • Frahm et al, *ECCV 2010*
  • K-medoids clustering of GIST descriptors
Component Discovery: Image Clustering

• Building Rome on a Cloudless Day
  • Frahm et al, *ECCV 2010*
  • K-medoids clustering of GIST descriptors
  • Iconic image selection
Component Discovery: Streaming

- Building the World in Six Days
  - Heinly, Schönberger, Dunn, Frahm, *CVPR 2015*
  - Combine image retrieval and iconic image selection
Large-Scale Performance

- **2006**
  - 24 PCs
  - 2 weeks
  - [Snavely et al.]
  - 400s

- **2008**
  - 24 PCs
  - 36 hours
  - [Snavely et al.]
  - 40s

- **2008**
  - 1 PC
  - 24 hours
  - [Li et al.]
  - 2s

- **2009**
  - 64 PCs
  - 24 hours
  - [Agarwal et al.]
  - 0.5s

- **2010**
  - 1 PC
  - 24 hours
  - [Frahm et al.]
  - 0.03s

- **2015**
  - 1 PC
  - 6 days
  - [Heinly et al.]
  - 0.005s

Time Spent per Processed Image

- 10^3
- 10^2
- 10^1
- 10^0
- 10^{-1}
- 10^{-2}
- 10^{-3}
Streaming Paradigm

• Read images sequentially from disk
• Read each image only once
• Keep images in memory only as long as necessary
Challenges of Streaming

• Limited window in which to perform association
• No control over image order
Streaming Data Association

100M Images
Streaming Data Association

100M Images

Large-scale 3D Modeling from Crowdsourced Data
Image Registration

[Images of different scenes and corresponding arrows, with emojis indicating success and failure]

Large-scale 3D Modeling from Crowdsourced Data
Streaming Data Association

100M Images
Streaming Data Association

100M Images
Streaming Data Association

100M Images
Streaming Data Association

100M Images
Cluster Formation

100M Images

Iconic Image
Cluster Representation

Iconic Image

Bag of Visual Words

1632
82497
405
7189
94
Cluster Representation

Iconic Image

Bag of Visual Words

1632
82497
405
7189
94
Cluster Representation

- Iconic Image
- Bag of Visual Words:
  - 1632
  - 82497
  - 405
  - 7189
  - 94
- Cluster Image
- Registered Visual Words:
  - 1632
  - 63917
  - 383
  - 7189
  - 2219
Cluster Representation

Iconic Image

Bag of Visual Words

Cluster Image

Registered Visual Words

1632
82497
405
7189
94

1632
63917
383
7189
2219
Cluster Representation

Iconic Image

Bag of Visual Words

1632
82497
405
7189
94
63917
383
2219

Cluster Image

Registered Visual Words

1632
63917
383
7189
2219
Streaming Data Association

100M Images
Streaming Data Association

100M Images
Streaming Data Association

100M Images
Streaming Data Association

100M Images
Streaming Data Association

100M Images
Streaming Data Association

100M Images
Streaming Data Association

100M Images
A problem has been detected and Windows has been shut down to prevent damage to your computer.

**DRIVER_IRQL_NOT_LESS_OR_EQUAL**

If this is the first time you’ve seen this Stop error screen, restart your computer. If this screen appears again, follow these steps:

Check to make sure any new hardware or software is properly installed. If this is a new installation, ask your hardware or software manufacturer for any windows updates you might need.

If problems continue, disable or remove any newly installed hardware or software. Disable BIOS memory options such as caching or shadowing. If you need to use Safe Mode to remove or disable components, restart your computer, press F8 to select Advanced Startup Options, and then select Safe Mode.

**** ABCD.SYS – Address F73120AE base at C0000000, DateStamp 36B072A3

Kernel1 Debugger Using: COM2 (Port 0x2F8, Baud Rate 19200)

Beginning dump of physical memory

Physical memory dump complete. Contact your system administrator or technical support group.
Streaming Data Association

100M Images
Streaming Data Association

100M Images
Cluster Discarding

Discard Rate

Large-scale 3D Modeling from Crowdsourced Data 101
Cluster Discarding
Cluster Discarding
Cluster Discarding
Recoverable Cluster Size

What is the minimal recoverable cluster size?

\[ w_1 \quad w_2 \quad \cdots \quad w_k \]

\[ c_r = \text{Recovered cluster size} \]
\[ c = \text{Cluster size} \]
\[ c_r = c - k \]

Recoverable as long as \( c > k \)

One missed image per discard window
Memory Consumption

What is the upper bound of clusters in memory?

\[ w_1 \]

\[ w \]

\[ \frac{w}{k} \]

\[ 2 \]

\[ 2 \]

\[ 2 \]

\[ 2 \]

\[ 2 \]

\[ \frac{w}{2} \]

\[ \frac{w}{2} \]

\[ \frac{w}{2} \]

\[ 1 \]

\[ 1 \]

\[ 1 \]

\[ 1 \]

\[ 1 \]

\[ 1 \]

\[ 1 \]

\[ 1 \]

\[ 1 \]
Memory Consumption

What is the upper bound of clusters in memory?

\[ B = \text{Upper bound of clusters} \]

\[ B \leq \sum_{i=1}^{k} \frac{w}{i} \leq w \log(k) \]

Number of clusters per discard window

Number of discard windows
Memory Consumption

Number of Clusters in Memory

Number of Processed Images (Millions)

Number of Clusters (Thousands)
Streaming Data Association

100M Images
Streaming Data Association

100M Images
Streaming Data Association

100M Images
Image Registration

[Diagram showing various images and reactions]
Duplicate Structure Disambiguation

- Heinly, Dunn, Frahm, *ECCV 2014*
Duplicate Structure Disambiguation

• Heinly, Dunn, Frahm, *ECCV 2014*
• Automatic model splitting and rearranging

[GitHub Link](https://github.com/jheinly/sfm_duplicate_structure_correction)
Connected Components

• Discovery strategies
  • Brute force
  • Image retrieval
  • Clustering
  • Streaming
Outline

• Challenges and opportunities of crowd-sourced data
• Image retrieval and indexing
• Connected component discovery
• Reconstruction of connected components
• Design of a real pipeline
• Lessons learned
Reconstruction of Connected Components
Reconstruction of Components

- Large-scale structure from motion
Reconstruction Strategies

• Brute force
Reconstruction Strategies

• Global solvers
  • Discrete-Continuous Optimization for Large-Scale SfM
  • Crandall, Owens, Snavely, Huttenlocher, *CVPR 2011*
Reconstruction Strategies

• Global solvers
  • 1DSfM – Global translation solver
  • Wilson, Snavely, *ECCV 2014*
Reconstruction Strategies

• Skeletal graphs
  • Snavely, Seitz, Szeliski, *CVPR 2008*
Reconstruction Strategies

• Iconic images
  • GIST-based clustering
  • Frahm et al, *ECCV 2010*
Reconstruction Strategies

• Iconic images
  • Streaming-based clusters
  • Heinly et al, CVPR 2015
Reconstruction Strategies

• Iconic images
  • Streaming-based clusters
  • Heinly et al, *CVPR 2015*
Outline

• Challenges and opportunities of crowd-sourced data
• Image retrieval and indexing
• Connected component discovery
• Reconstruction of connected components
• Design of a real pipeline
  • Lessons learned
Design of a Real Pipeline
Overview

• Reconstructing the World in Six Days, *CVPR 2015*
• Streaming connected component discovery
• 1 PC
• Multi-core, multi-GPU implementation
• Open-source

github.com/jheinly/streaming_connected_component_discovery
Overview

Large-scale 3D Modeling from Crowdsourced Data
Prepare Images

Disk → File Reader → Buffer → JPEG Decode & Resize → Buffer → SIFT → Buffer → Visual Words → Buffer

Voc-Tree Query KNN → Geometric Verification → Update Iconic Images → Merge Clusters → Discard Clusters → Buffer → Save SIFT → Disk
Update Clusters and Components

Large-scale 3D Modeling from Crowdsourced Data
Save Results

Large-scale 3D Modeling from Crowdsourced Data
Batch Processing and Buffering

1. Disk
2. File Reader
3. Buffer
4. JPEG Decode & Resize
5. Buffer
6. SIFT
7. Buffer
8. Visual Words
9. Buffer
10. Voc-Tree Query
11. KNN
12. Geometric Verification
13. Update Iconic Images
14. Merge Clusters
15. Discard Clusters
16. Buffer
17. Save SIFT
18. Disk

Large-scale 3D Modeling from Crowdsourced Data
File Reader

Disk → File Reader → Buffer → JPEG Decode & Resize → Buffer → SIFT → Buffer → Visual Words → Buffer

Voc-Tree Query KNN → Geometric Verification → Update Iconic Images → Merge Clusters → Discard Clusters → Buffer → Save SIFT → Disk
File Reader

• Decouple JPEG decode from file I/O
• Multiple hard drives
• One thread per drive
• Enforce a balanced load

• 1024x768 images
• 4 hard drives
• 120 Hz
JPEG Decode

- `libjpeg-turbo`
- Resize by power of 2 during decode
- OpenCV resize to final resolution

- 4 threads
- 177 Hz
SIFT

Large-scale 3D Modeling from Crowdsourced Data
SIFT

• SiftGPU by Changchang Wu
• CUDA-enabled, multi-process support

• 8 threads, 8 GPUs
• GTX 295
• 138 Hz
Visual Words

Disk → File Reader → Buffer → JPEG Decode & Resize → Buffer → SIFT → Buffer → Visual Words → Buffer

Voc-Tree
Query
KNN

Geometric Verification

Update Iconic Images

Merge Clusters

Discard Clusters

Save SIFT

Disk

Large-scale 3D Modeling from Crowdsourced Data
Visual Words

• Optimized VocabTree2 by Noah Snavely
  • SSE, multithreading, and faster data structures
• 1 million visual word vocabulary

• 4 threads
• 434 Hz
Vocabulary Tree Query KNN

1. Disk
2. File Reader
3. Buffer
4. JPEG Decode & Resize
5. Buffer
6. SIFT
7. Buffer
8. Visual Words
9. Buffer
10. Voc-Tree Query KNN
11. Geometric Verification
12. Update Iconic Images
13. Merge Clusters
14. Discard Clusters
15. Buffer
16. Save SIFT
17. Disk
Vocabulary Tree Query KNN

- Optimized VocabTree2 by Noah Snavely
  - SSE, multithreading, and faster data structures
- 1 million visual word vocabulary
- 25 nearest neighbors
- 16 threads
- 4,475 Hz
Geometric Verification
Geometric Verification

- ARRSAC, Raguram, Frahm, Pollefeys, ECCV 2008
- GPU-enabled SIFT matching
- Essential matrix estimation
  - Focal length from EXIF
- 2 verification attempts per image
- Save verification results in SQLite database

- 16 threads, 8 GPUs
- 261 Hz
Update Clusters and Components

Disk → File Reader → Buffer → JPEG Decode & Resize → Buffer → SIFT → Buffer → Visual Words → Buffer

Voc-Tree → Query → KNN → Geometric Verification → Update Iconic Images → Merge Clusters → Discard Clusters → Buffer → Save SIFT → Disk
Update Clusters and Components

- Update vocabulary tree inverted index
  - Optimized VocabTree2 by Noah Snavely

- 16 threads
- 14,485 Hz
Save SIFT

Disk → File Reader → Buffer → JPEG Decode & Resize → Buffer → SIFT → Buffer → Visual Words → Buffer

Voc-Tree Query KNN → Geometric Verification → Update Iconic Images → Merge Clusters → Discard Clusters → Buffer → Save SIFT → Disk
Save SIFT

- Save SIFT features to SSD
  - 2D keypoints used in SfM
- 3 threads
- 186 Hz
Overview

Large-scale 3D Modeling from Crowdsourced Data
Reconstruction

- Reconstruction of iconic and connecting images
- Register remaining images from clusters
- COLMAP by Johannes Schönberger
Notre Dame Cathedral, Paris, France

126K Cameras

Large-scale 3D Modeling from Crowdsourced Data
Berlin Cathedral, Berlin, Germany

26K Cameras
## Results

- Complexity

<table>
<thead>
<tr>
<th></th>
<th>Our Method</th>
<th>Previous Works</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Computation</strong></td>
<td>$O(N)$</td>
<td>Ranges from $O(N)$ to $O(N^2)$</td>
</tr>
<tr>
<td></td>
<td>2 registration attempts per image</td>
<td>Multiple registration attempts</td>
</tr>
<tr>
<td></td>
<td></td>
<td>per image</td>
</tr>
<tr>
<td><strong>Memory</strong></td>
<td>$O(\log(N))$</td>
<td>$O(N)$</td>
</tr>
<tr>
<td></td>
<td>Close to $O(1)$ in practice</td>
<td></td>
</tr>
</tbody>
</table>
## Results

### Berlin, Germany (2.7M images)

<table>
<thead>
<tr>
<th></th>
<th>Frahm et al, 2010</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registered</td>
<td>4.6%</td>
<td>26%</td>
</tr>
<tr>
<td>Reconstructed</td>
<td>1.1%</td>
<td>8.7%</td>
</tr>
<tr>
<td>Data Association Time*</td>
<td>13.3 Hours</td>
<td>7.9 Hours</td>
</tr>
</tbody>
</table>

*Identical Hardware Configuration
## Results

### Berlin, Germany (2.7M images)

<table>
<thead>
<tr>
<th>Data Association Strategy</th>
<th>Match Attempts</th>
<th>Registered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frahm et al, 2010 GIST</td>
<td>1</td>
<td>4.6%</td>
</tr>
<tr>
<td>Streaming GIST</td>
<td>2</td>
<td>8.9%</td>
</tr>
<tr>
<td>Streaming Bag-of-Words</td>
<td>1</td>
<td>25%</td>
</tr>
<tr>
<td>Streaming Bag-of-Words</td>
<td>2</td>
<td>26%</td>
</tr>
</tbody>
</table>
Results – Effect of Parameters

• 2.2M geo-tagged subset of Yahoo 100M dataset

• Baseline
  • Ground-truth connectivity
Results

Effect of Discard Rate
Match to $k = 2$ Neighbors

Number of Registered Images vs. Discard Rate

Recoverable Baseline
Yahoo® Flickr® 100M Dataset

- 14 TB
- 640x480 resolution
Results

• 100 million images
• 1.5 million images registered
• 6 days
  • 4.4 days streaming
  • 0.9 days SfM
• 1 PC
Results

Sacre Coeur, Paris, France
Results

Brandenburg Gate, Berlin, Germany
Results

Buckingham Palace, London, England
Results

Golden Gate Bridge, San Francisco, California
Results

Eiffel Tower, Paris, France
Results
Results
Results
Results
Results
Results
Outline

• Challenges and opportunities of crowd-sourced data
• Image retrieval and indexing
• Connected component discovery
• Reconstruction of connected components
• Design of a real pipeline
• Lessons learned
Lessons Learned

• 1-in-a-million bugs hurt
• Engineering matters
• Ask questions and dig into code
• Expect and handle data of all forms
• Create visualizations whenever possible
Questions?