

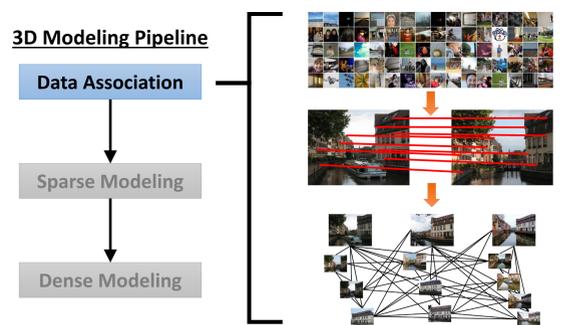


# Reconstructing the World\* in Six Days

\*(As Captured by the Yahoo 100 Million Image Dataset)

Jared Heinly, Johannes L. Schönberger, Enrique Dunn, Jan-Michael Frahm

## Motivation



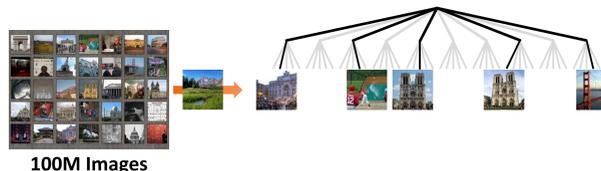
- We push 3D modeling from city-scale (~1M images) to world-scale datasets (~100M images)
- Data association is the biggest challenge at this scale

## Streaming Paradigm

- Tackle robustness, scalability, and completeness of data association
- Read images sequentially from disk
- Read each image only once
- Keep images in memory only as long as necessary



## Streaming Connected Component Discovery



### For Each Streamed Image:

- Retrieve  $k$  nearest neighbors using a bag-of-words representation
- Attempt registration to the set of  $k$  nearest neighbors



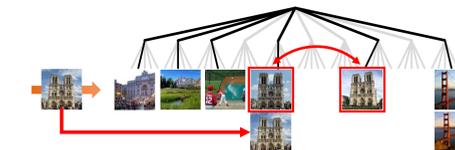
### If No Successful Registration:

- Create a new single-image cluster in the database



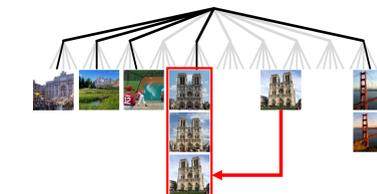
### If 1 Successful Registration:

- Add the image to the matching cluster



### If 2+ Successful Registrations:

- Add the image to the best-matching cluster
- Link clusters into a connected component
- Avoid matching streamed images to the same connected component twice

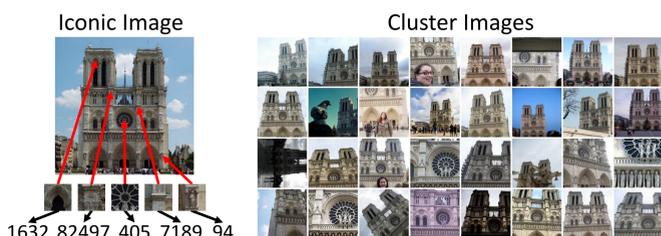


### If 2 Clusters Are Linked into a Component:

- Attempt direct registration between the clusters
- If successful, merge the clusters into a single representation

## Cluster Representation

- Use cluster images to create adaptive cluster representation

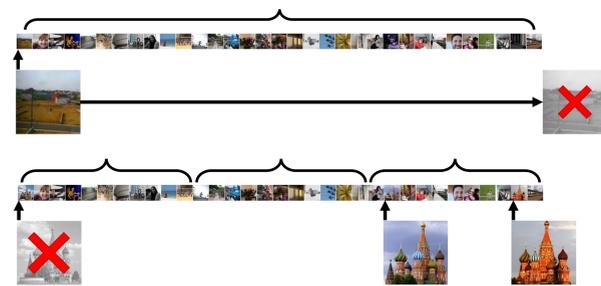


Iconic Image	Bag of Visual Words	Cluster Image	Registered Visual Words
	1632		1632
	82497		63917
	405		383
	7189		7189
	94		2219
	<b>63917</b>		<b>383</b>
	<b>383</b>		<b>2219</b>

## Cluster Discarding

- Some clusters are less important than others
- Discard clusters from memory that do not grow in size fast enough
- Discarding enables scalability to world-scale datasets

### Discard Rate

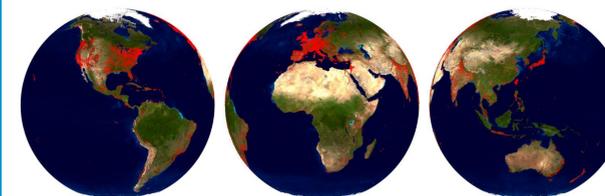


## Yahoo® Flickr® Dataset

100 Million Images  
14TB, 640x480 Resolution



1.5 Million Images Registered  
105 Hours



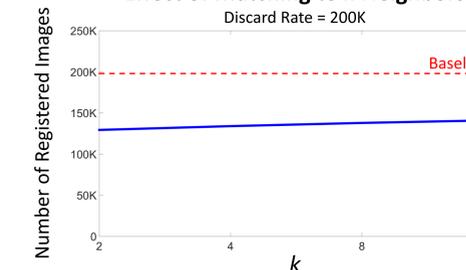
### Berlin, Germany (2.7M images)

	Frahm et al, 2010	Ours
Registered	4.6%	26%
Reconstructed	1.1%	8.7%
Data Association Time*	13.3 Hours	7.9 Hours

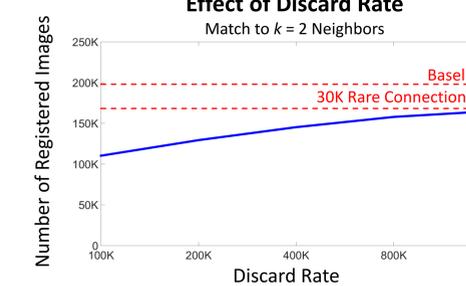
\*Equivalent Hardware Configuration

## Results

### Effect of Matching to $k$ Neighbors



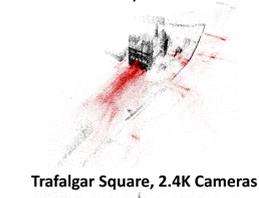
### Effect of Discard Rate



Berlin Cathedral, 26K Cameras



Notre Dame, 126K Cameras



Trafalgar Square, 2.4K Cameras

