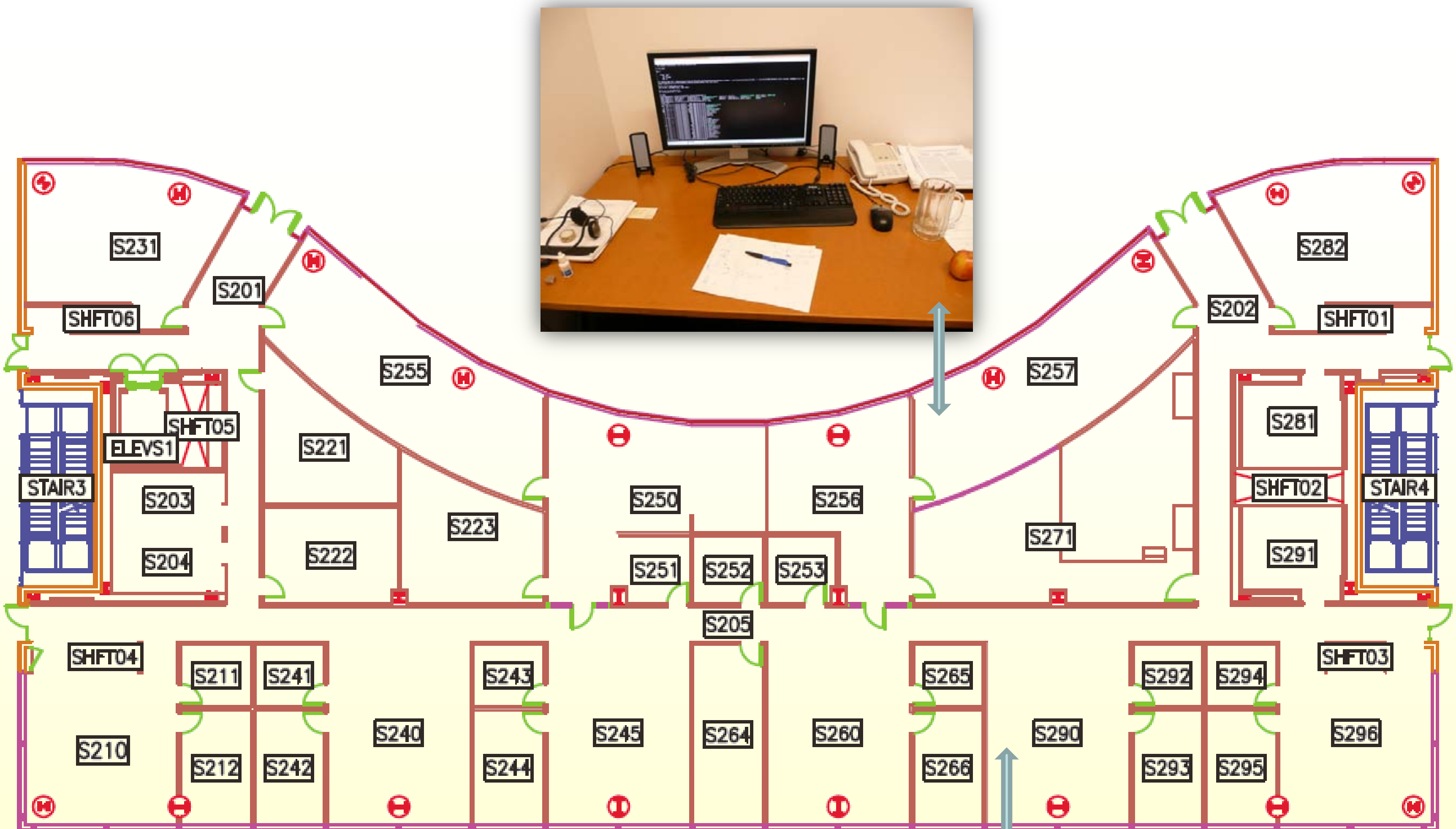
Image Webs: Computing and Exploiting Connectivity in Image Collections

STANFORD COMPUTER FORUM

Leonidas Guibas
Computer Science Dept.
Stanford University


S231
S201
SHFT06
S255
S221
SHFT05
ELEVS1
STAIR3
S203
S204
S222
S223
S250
S256
S251
S252
S253
S205
S257
S271
S202
S282
SHFT01
S281
SHFT02
STAIR4
S291
SHFT04
S211
S241
S210
S212
S242
S240
S243
S244
S245
S264
S260
S265
S266
S290
S292
S294
S293
S295
SHFT03
S296

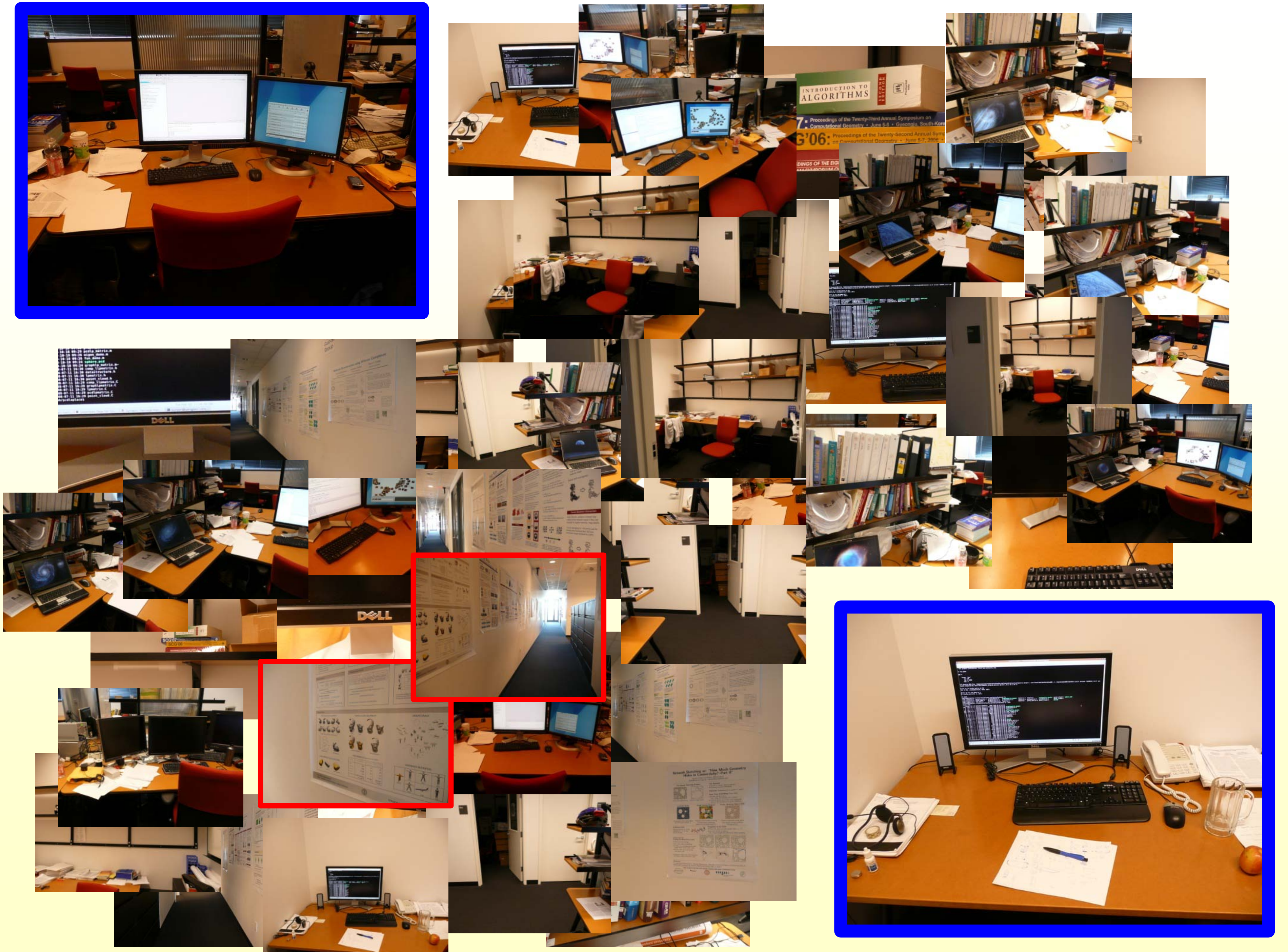
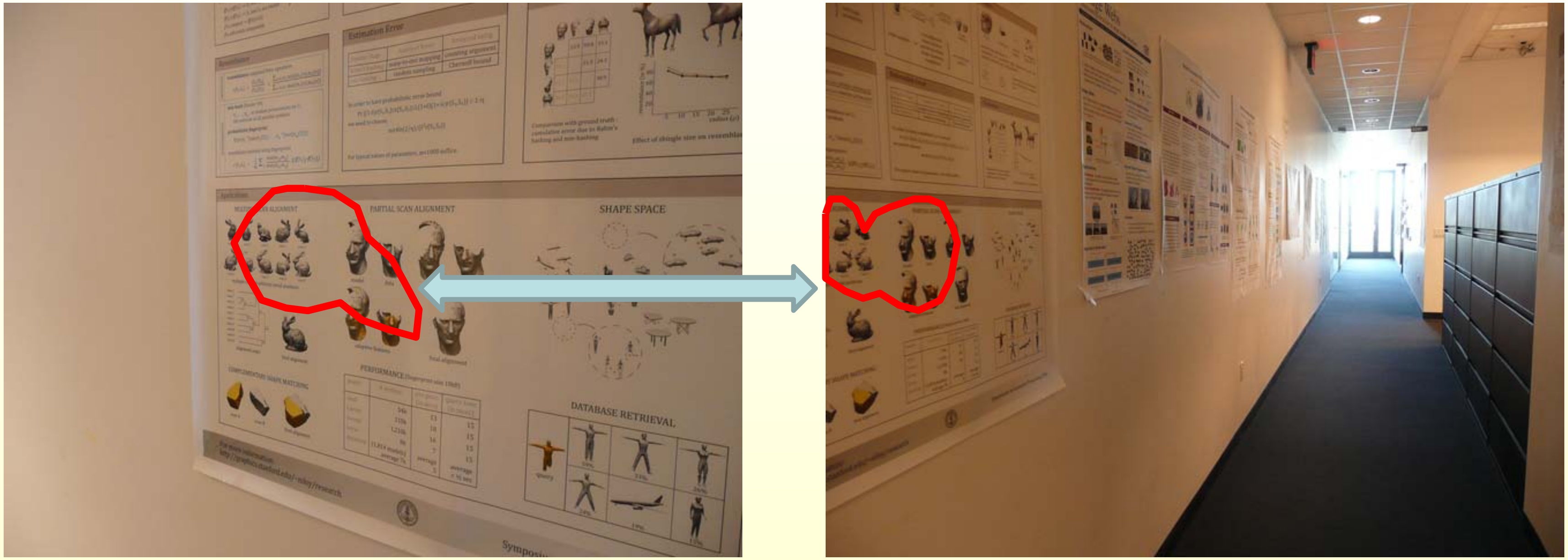

INTRODUCTION TO
ALGORITHMS

Image Match Links



Paths Through Image Collections

Path Homotopy Classes

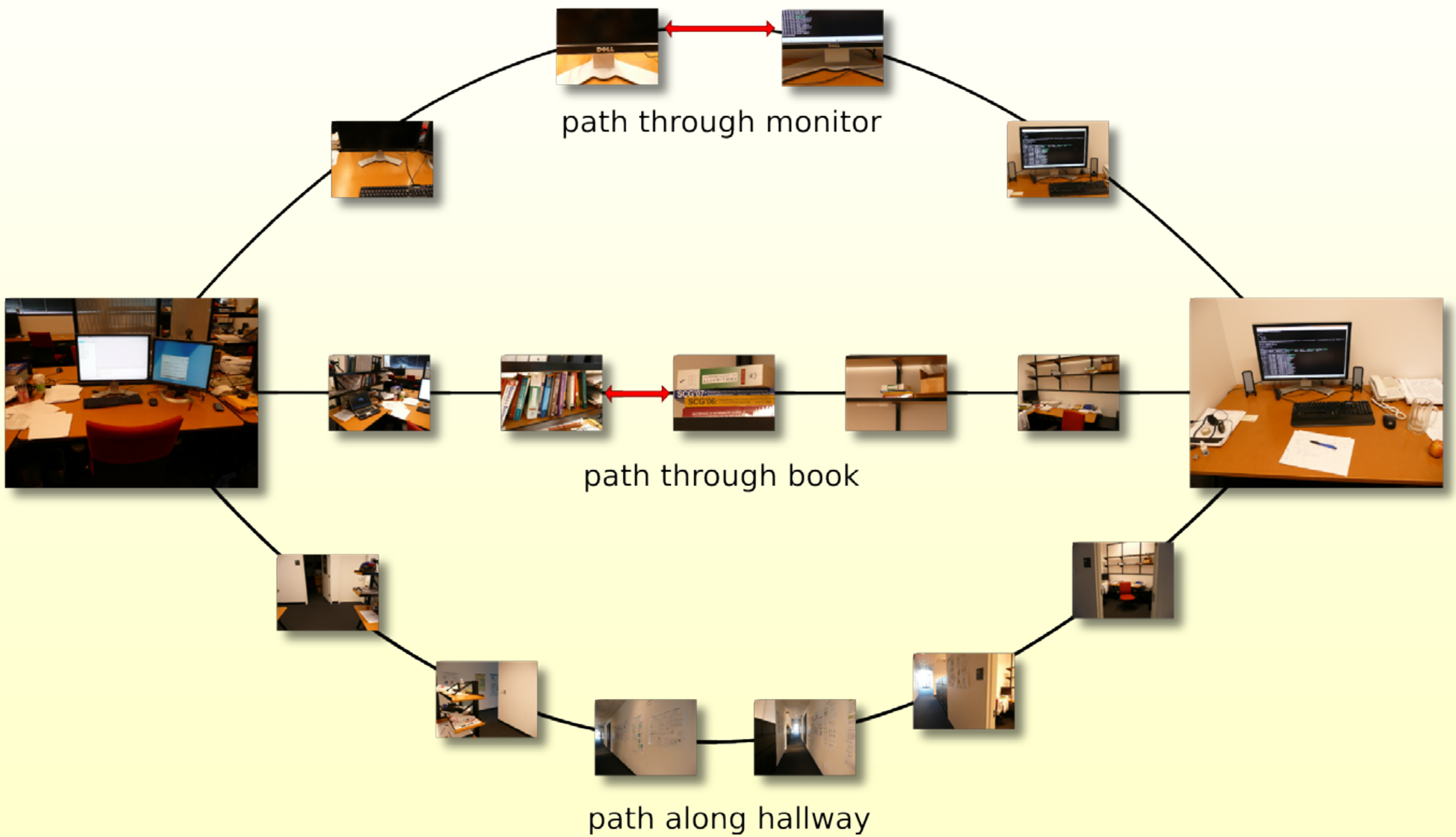
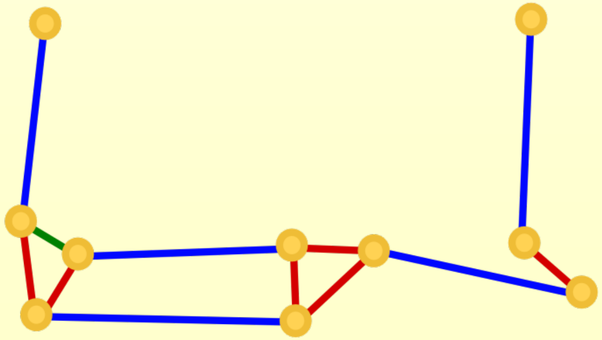
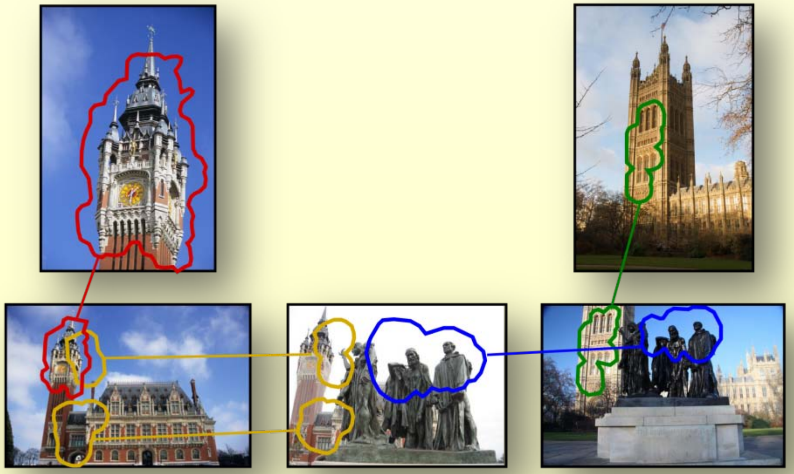


Image Webs

- The idea of Image Webs is to interlink images through a variety of link types, based on both content and image metadata (GPS, time, annotations)
- *The same way that the WWW of documents has proved useful, the hope is that interlinked webs of signals will also be valuable for propagating, extracting, and filtering information – and the web types two can cross-link and cross-fertilize*

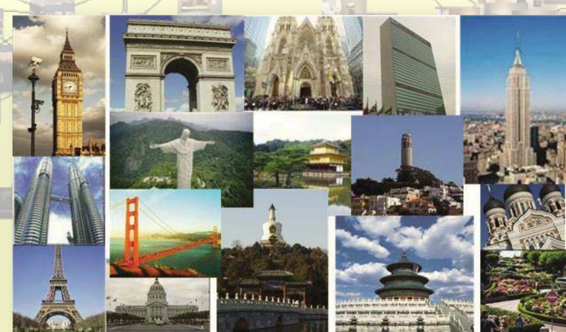




Global Connectivity

- Understand the local and global structure of image webs, aiming at a softer, more topological understanding
- Develop efficient Web construction algorithms
- Explore applications (image browsing, annotation transfer, social networks, etc.)

[Snavely et. al., Siggraph '06]

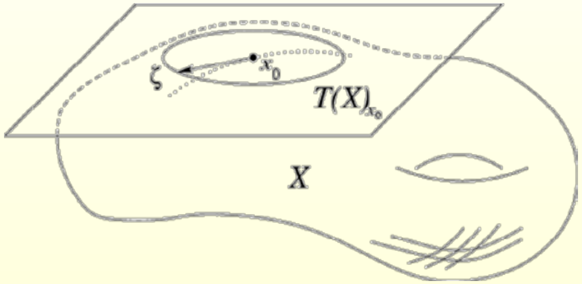


[Zheng et. al., CVPR 2009]

[Gammeter et al., ICCV 2009]

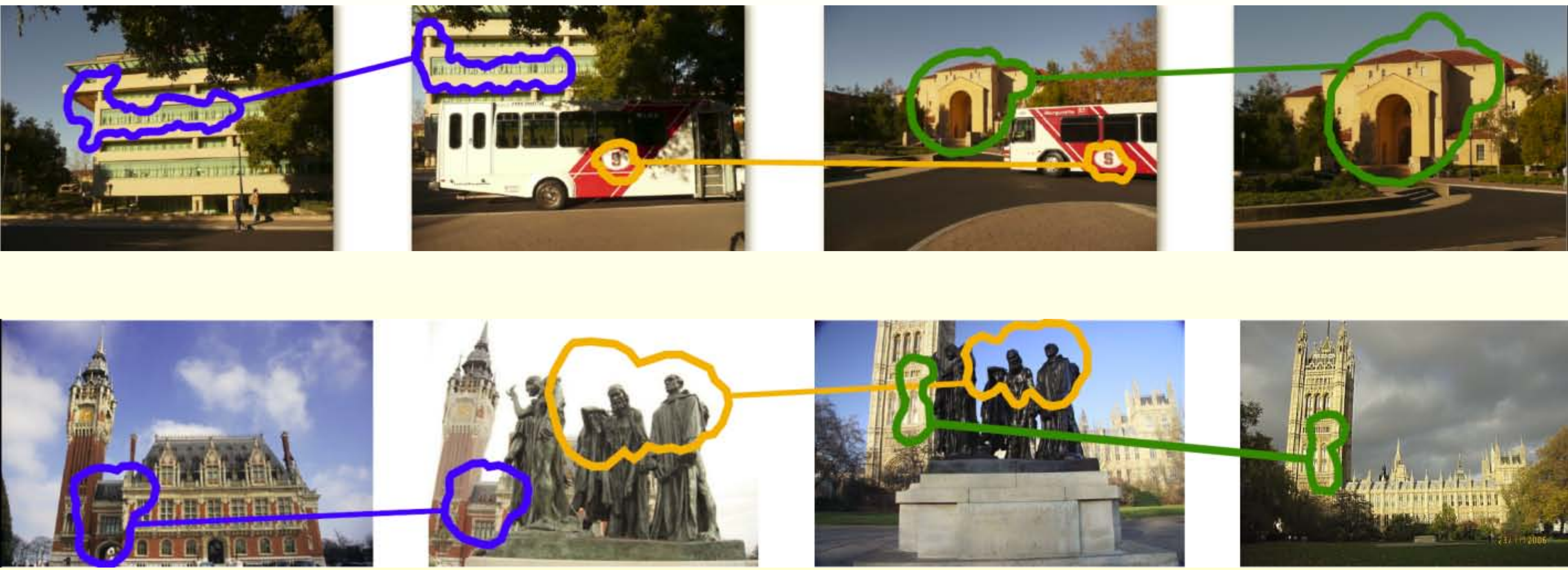
The Space of All Images

- If we frieze time, the local structure of the space of images is well understood: it that of a low dimensional manifold – the manifold of views
- This is also the local structure of an image web based on match links

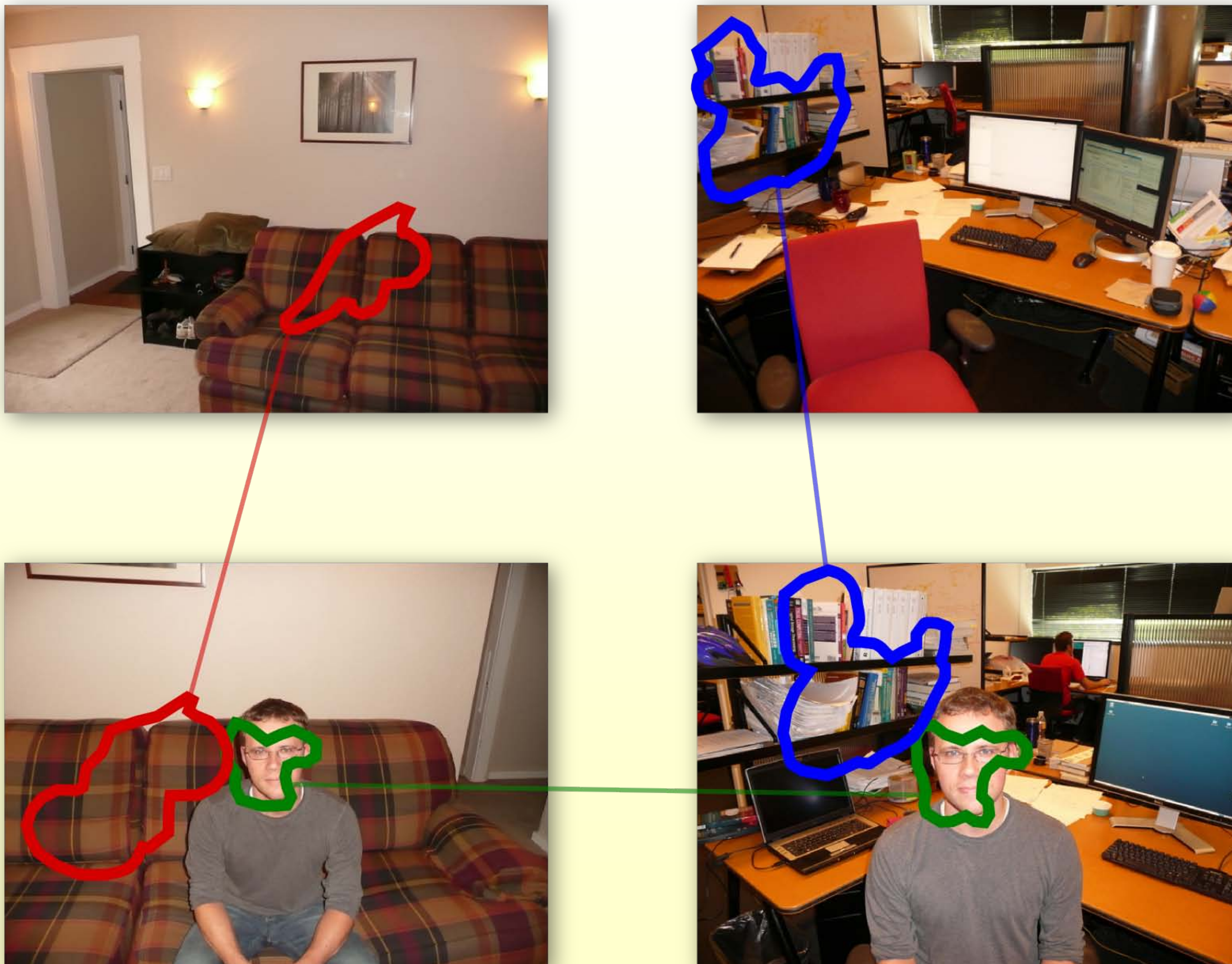


- But at larger scales the structure is more complex
  - because of moving objects
  - because of repeated similar objects
- For us this is exactly the structure that is of interest

Non-Local Links

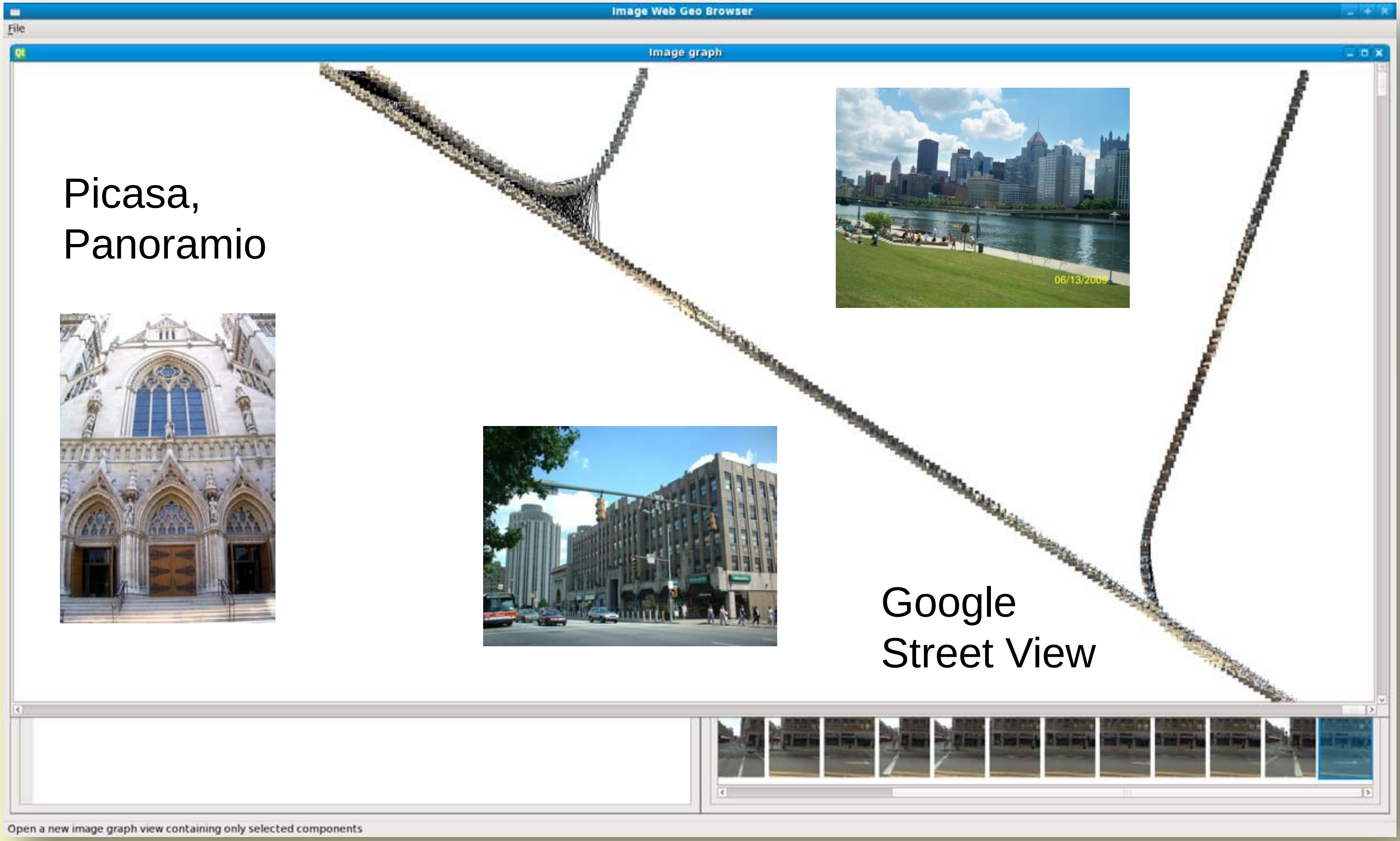


Proximity Through Mobility: Home to Office



Proximity Through Mobility on the Stanford Campus

Scaffold Webs: Getting Essential Connectivity



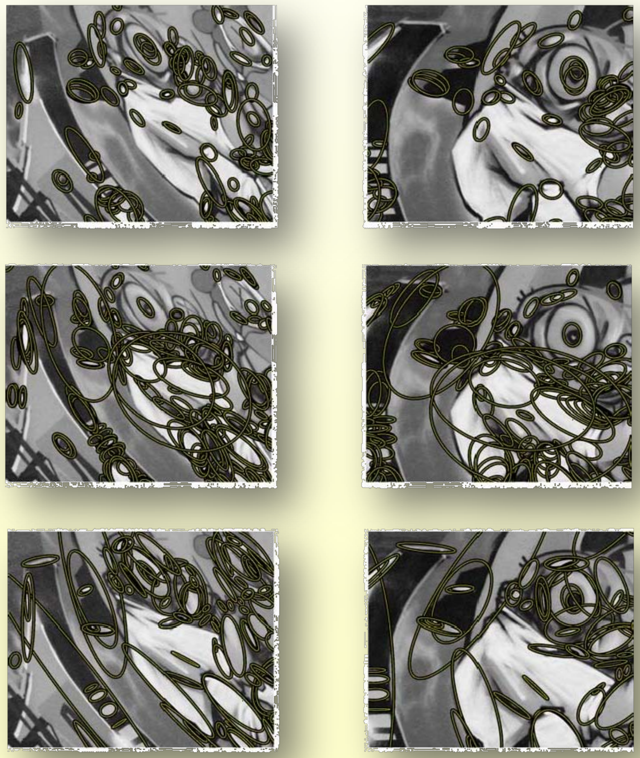
Getting Down to It: Building Image Webs

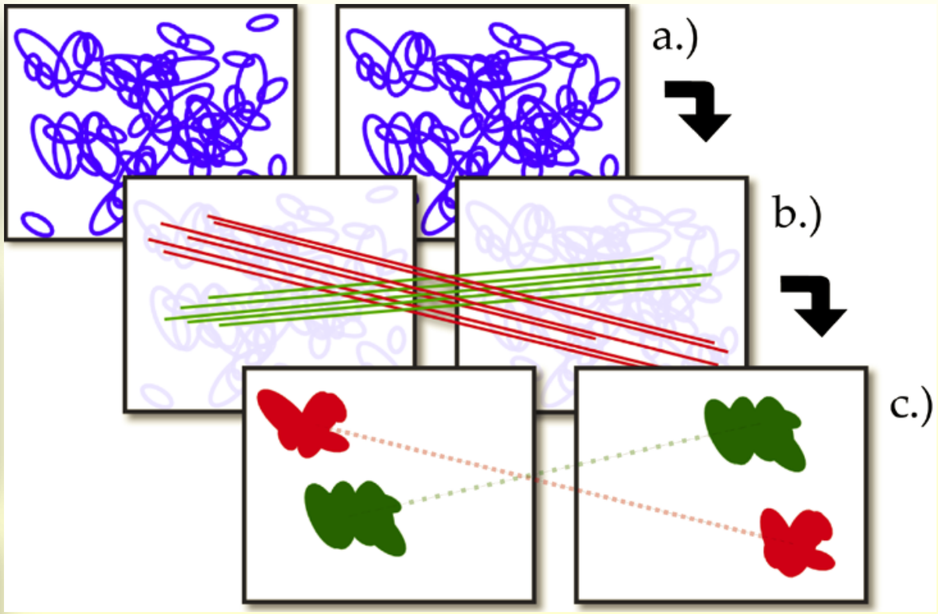
- Feature Extraction: interest points, associated with a region and summarized by a SIFT descriptor

Harris-Affine

Hessian-Affine

Maximally Stable Extremal Regions

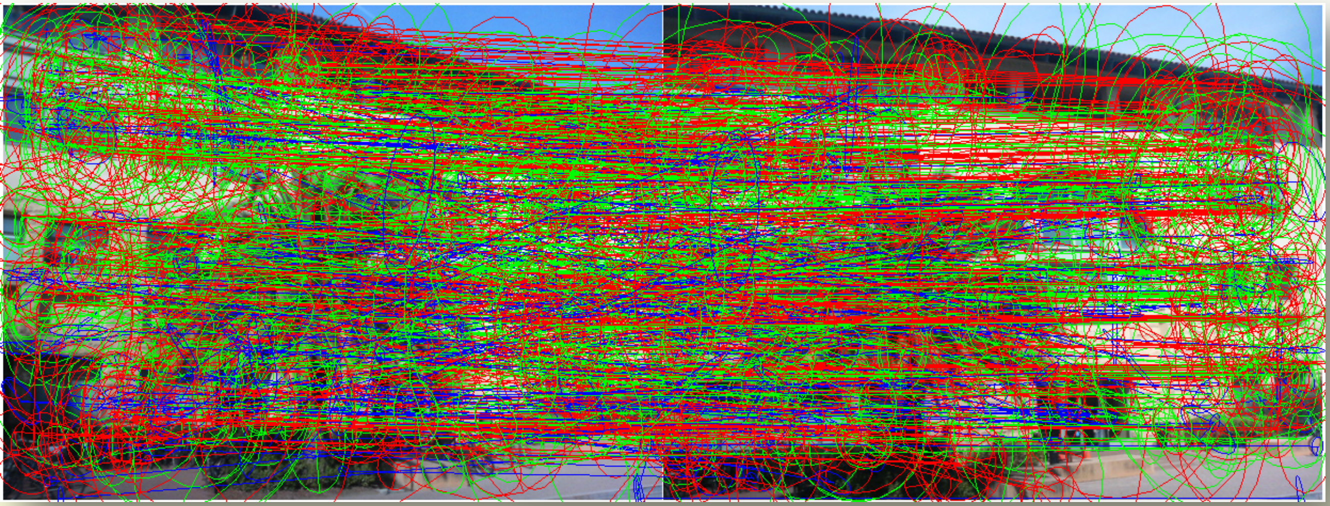




Geometric verification

K. Mikolajczyk, et al. A comparison of affine region detectors. IJCV, 2005

Getting Rid of False Feature Matches



raw feature matches

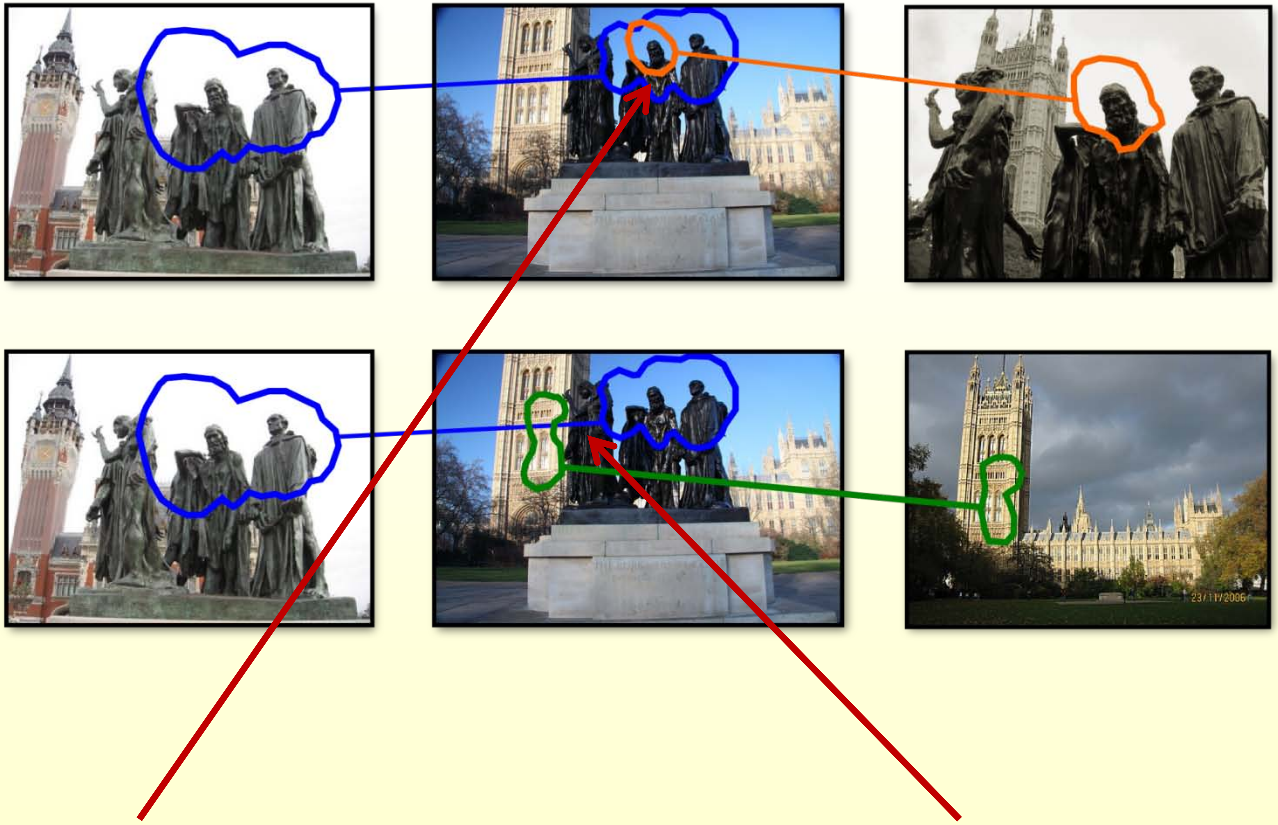
after geometric verification

Match Links

Match links connect image regions with coherent feature sets

Link aliasing: symmetries and repetitions can mislead …

More Links: Overlap and Pivot



Overlap link

Pivot link

Basic element of a Web is a pair (patch, image)

Links and Their Decorations

Link decoration:

- Match (M)-links (quality of match, transform attributes)
- Overlap (O)-links (degree of overlap)
- Pivot (P)-links (patch distance, visual attributes)

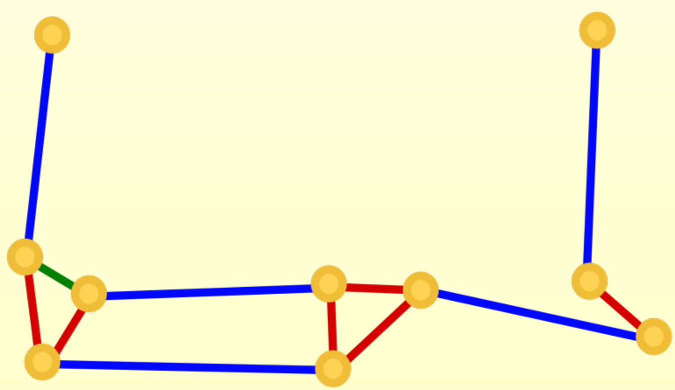


Image Webs Pipeline

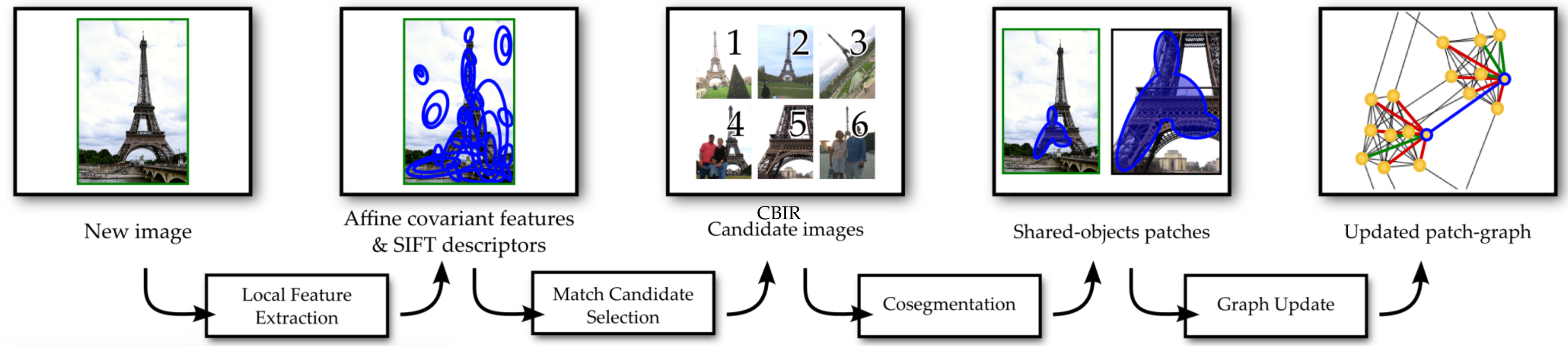


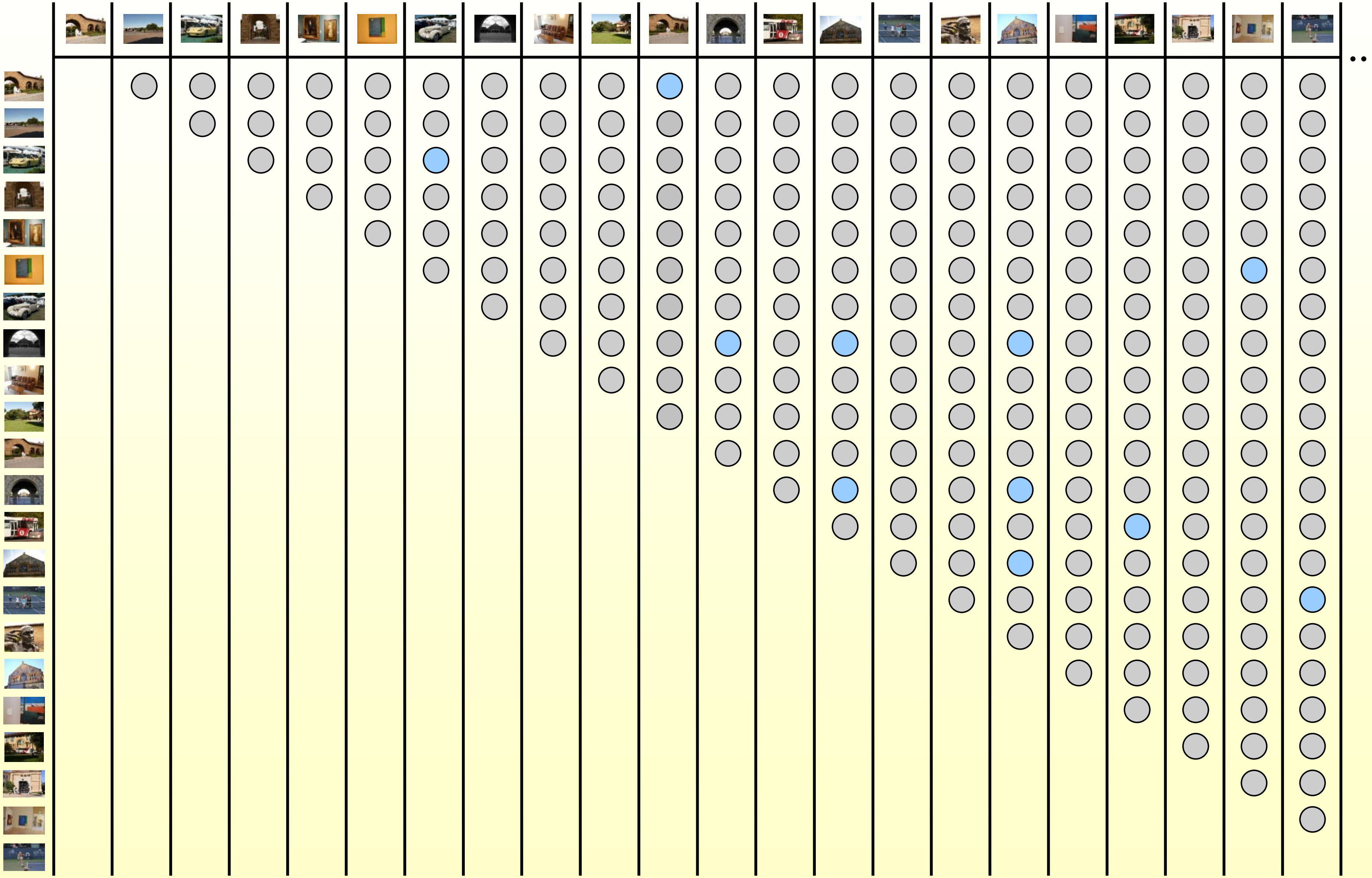
Image webs are complexes (graphs) on image patches, not images

Feature Matching

Geometric Verification

Co-Segmentation

Scalability



Distributed Computation

- Construction pipeline easily distributed on a computer cluster
  - A manager node issues feature extraction, CBIR, and cosegmentation jobs to worker nodes
  - Communication using the Internet Communication Engine (ICE) middleware and a shared file system

Gaining Efficiency: Pruning Pairs by CBIR Filtering

- Content-Based Image Retrieval (CBIR) via “Bag of Words” models:
  - cluster and quantize descriptors into vocabulary trees
  - use document information retrieval type indices

[Fei Fei, Fergus, Torralba]

- Used to retrieve “visually similar” images – in our case possible Web neighbors for which match links exist

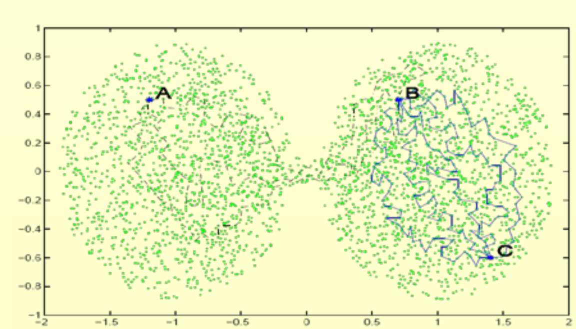
Computation Times (on a Cluster)

- Image matching steps (VGA image size)
  - Feature extraction (~ 4 sec per image)
  - CBIR indexing (~ 30 sec per image)
  - Cosegmentation operation (~ 1.5 sec per image pair)
- Image Web construction times*
  - Car (70 images ~ 1 minute)
  - Art museum (1200 images ~ 52 minutes)
  - Stanford campus (4200 images ~ 3 hours)

*just cosegmentation stage using up to 500 compute nodes

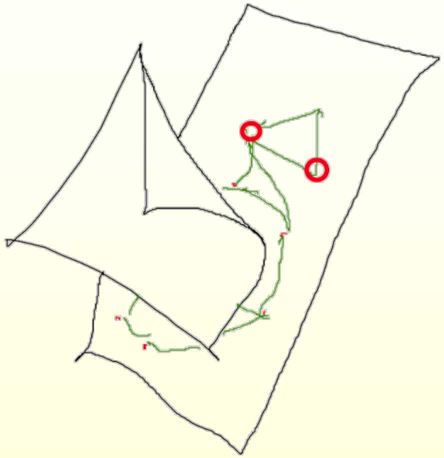
Scaling Up Web Construction

- We want to build Image Webs with millions of images -- and understand how they are connected
- We cannot afford to try cosegmentation on all image pairs
- CBIR is a useful filter, but …
- Vital connectivity information may reside in sparser areas of the Web



Getting an Unknown Graph to Reveal Itself …

- Testing for the presence of links is expensive
- Which images pairs should we try to connect?
- We seek a sparser graph which captures the connectivity of the unknown Web
  - On the one hand, the CBIR filter favors image pairs where links are likely to exist
  - But how can we tell is a particular link improves connectivity?
  - What should be our ultimate measure of Web utility?



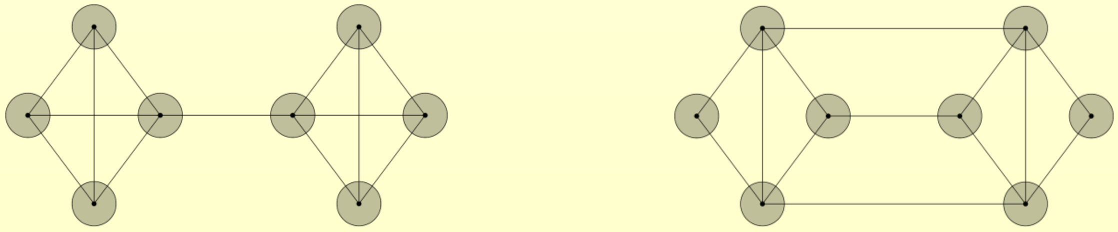
- Spectral graph theory and harmonic analysis to the rescue

Algebraic Connectivity Measures

- Connectivity of a graph based on heat diffusion notions
  - Second smallest eigenvalue of the graph Laplacian

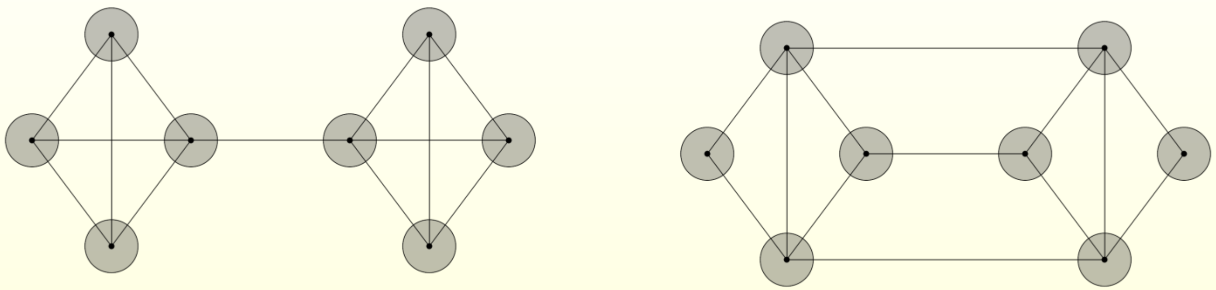
$$L_{i,j} = \begin{cases} d(i) & \text{if } i = j \\ -1 & \text{if } i \sim j \\ 0 & \text{otherwise} \end{cases}$$

  - Smallest eigenvalue of *L* is always 0 – and has a constant eigenvector
  - Multiplicity of 0: number of connected components



Algebraic Connectivity

- Connectivity Measure: Second smallest eigenvalue of the graph Laplacian



$\lambda_2 = 0.3542$ $\lambda_2 = 1.0968$

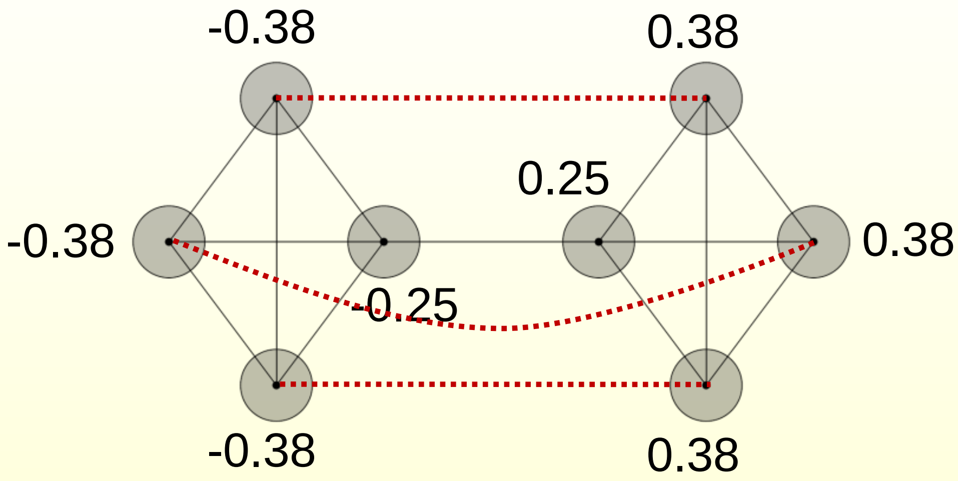
- Related to the diameter D of a graph with n nodes, random walk convergence, diffusion distances, and many other measures of graph connectivity
- The eigenvector corresponding to λ_2 is the Fiedler vector, and is often used to partition the graph

Building a “Good” Graph

- Objective:
  - Build a “well connected” graph in minimal time
- Difficulty:
  - Given a graph, finding the *k* extra edges which maximally increase algebraic connectivity is NP-hard
- Use a greedy strategy:
  - For every potential new connection, test its EdgeRank *R* – how much it will increase connectivity

Building a “Good” Graph

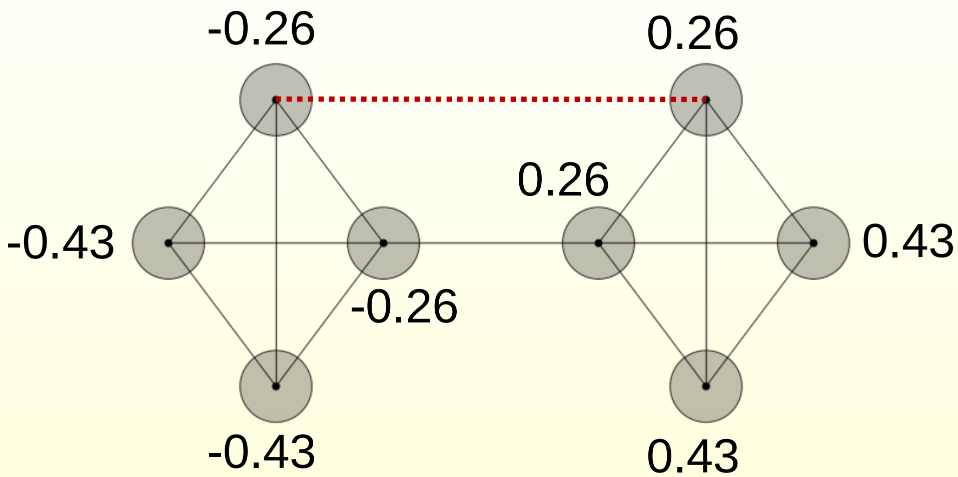
- Use a strategy from graph cuts



- Assign to each node its value in the Fiedler vector
- Add an edge *(i, j)* to maximize connectivity score:

Building a “Good” Graph

- Practical considerations



- Update the Fiedler vector after each new edge
- Can use the old estimate as a guess
- Use a *power iteration* to update the Fiedler vector

Building a "Good" Graph

- Power Iteration

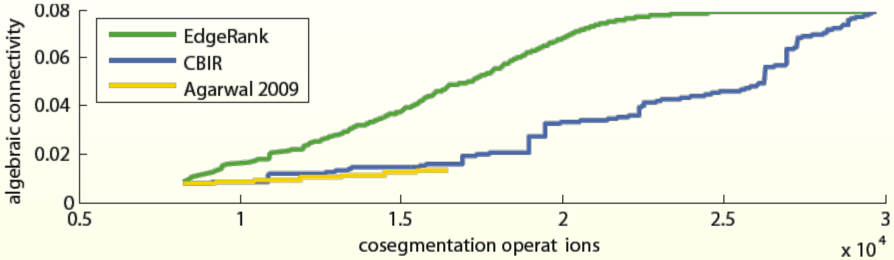
$$u_{i+1} = (2nI - L)u_i$$

$$u_{i+2}(j) = u_{i+1}(j) - \frac{1}{n}\sum_{k=1}^{n} u_{i+1}(k) \ \forall \ j$$

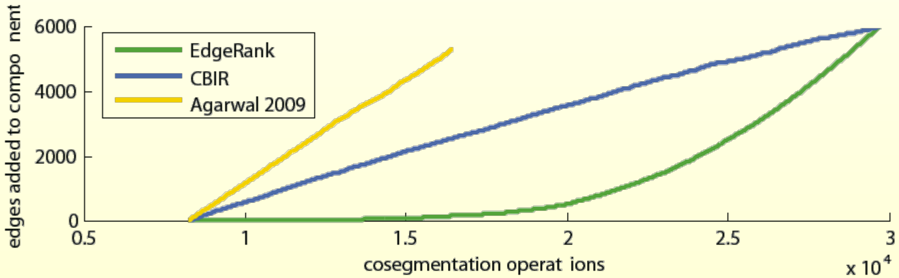
$$u_{i+3} = \frac{1}{\|u_{i+2}\|}u_{i+2}$$

  - Converges to the Fiedler vector
  - Convergence is fast if have a good estimate. We don't expect the Fiedler vector to change drastically
  - Small overhead: only 1 vector in memory

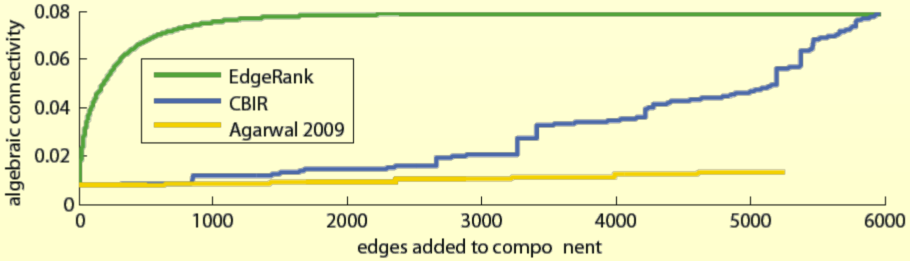
Algebraic Connectivity Speed-Ups



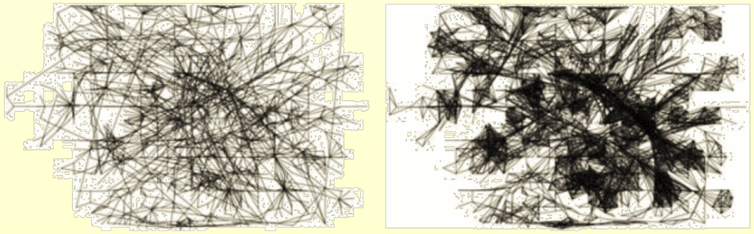
(a) Connectivity / Construction Time



(b) Edges / Construction Time



(c) Connectivity / Edges



(a) Edge Rank (b) Query Expansion

Agarwal 2009 = query expansion

Timings on Cluster with 500 Nodes

Collection Name (Source)	Images	Components (size > 1)	Largest Component	Construction Time (min) Phase 1	Phase 2	Total
Stanford (Flickr)	193,277	12,505	11,240	173	96	269
Pittsburgh (StreetView)	50,224	23	49,907	7.9	70	78
London (Panoramio)	17,925	902	4,617	7.7	5.9	14
Art Museum (created)	1,257	5	1,217	0.06	0.74	0.8

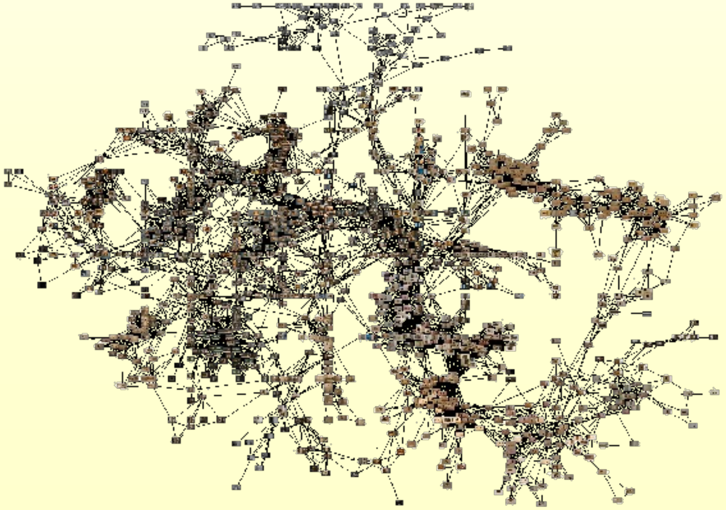
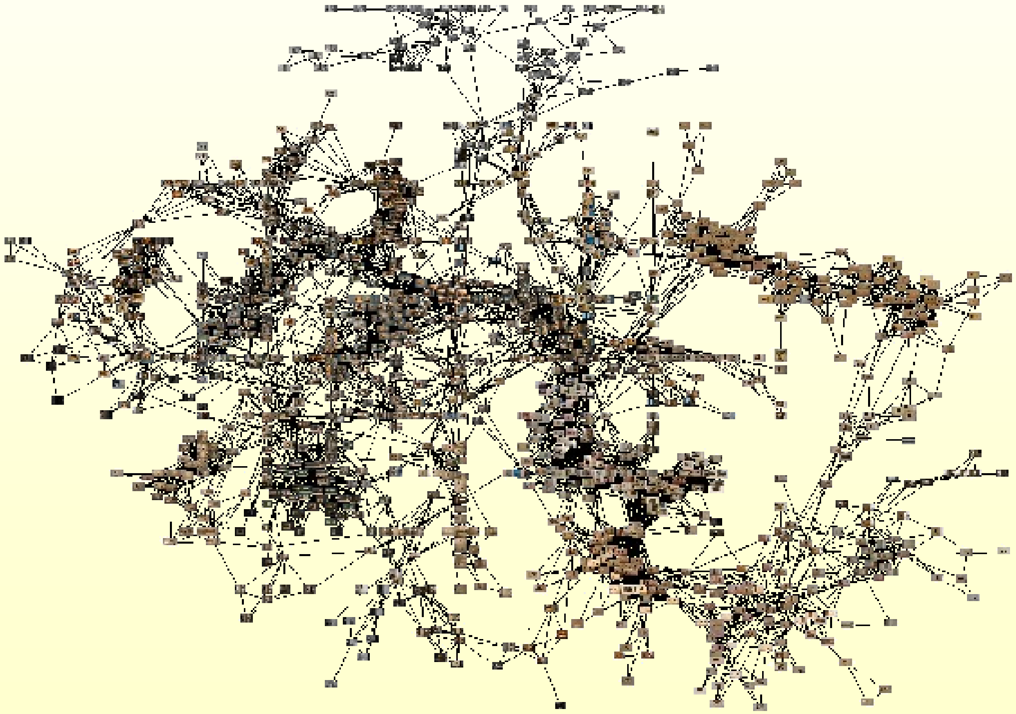
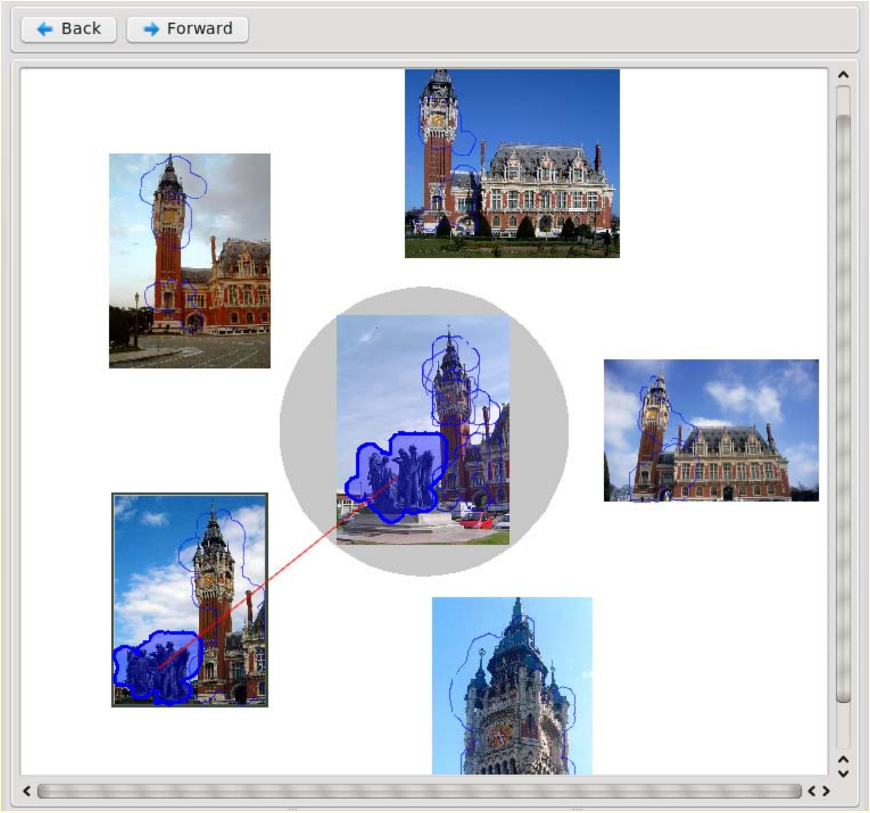


Image Web Applications

- Navigation through a space of images
- Enhanced image search
- Propagation of information through image links

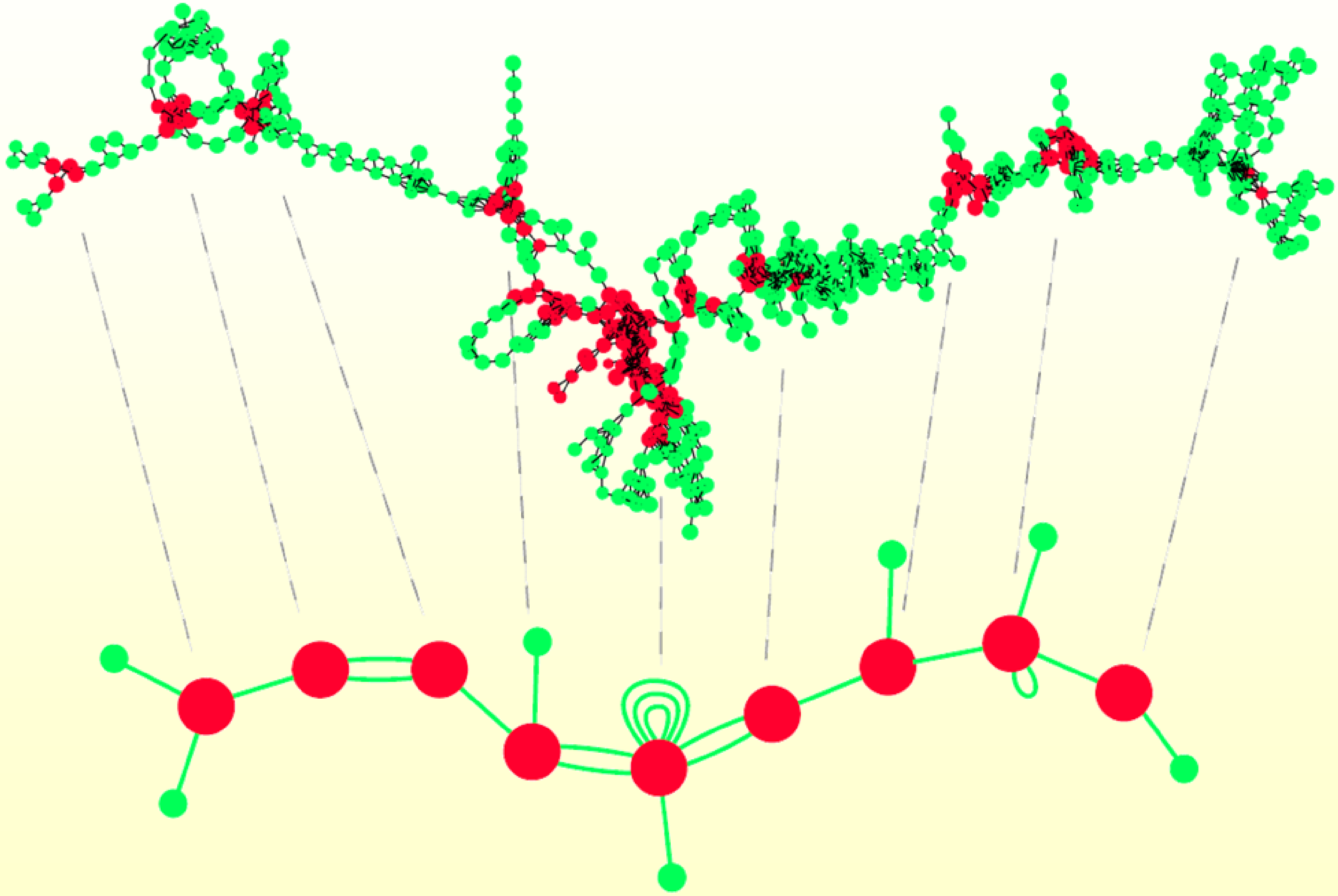


An Image Webs Browser



- How can we navigate through large Image Webs effectively?
- How do we mitigate the effects of wrong links?
- How do we extract "persistent" global structure

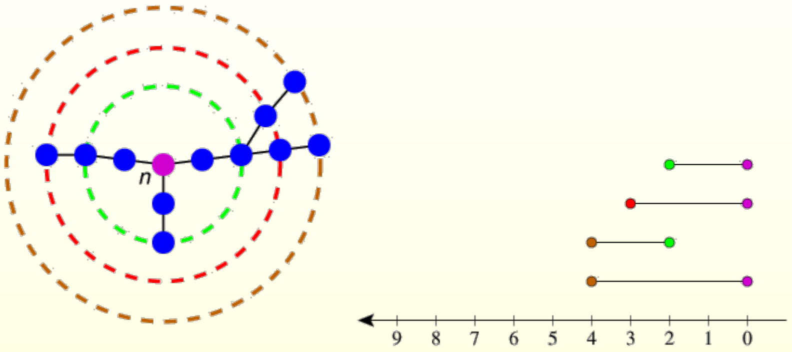
Computing a `Summary Graph’



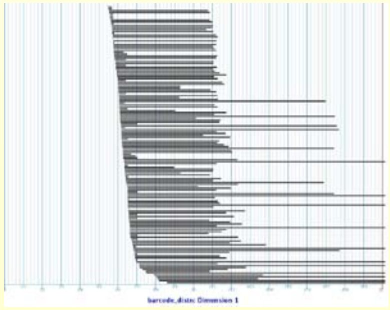
A global map makes navigation easy

Persistent Local Homology

- Image Webs are often stratified spaces because of the acquisition process – understanding the strata structure helps

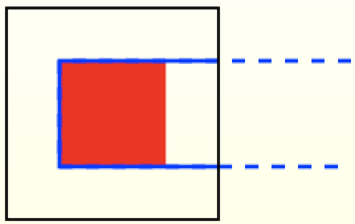


- Use some algebraic topology: image webs as combinatorial complexes
- Rips-Vietoris complex on images, based on distances coming from the links (affine maps)
- Exploit filtered complexes and persistence ideas

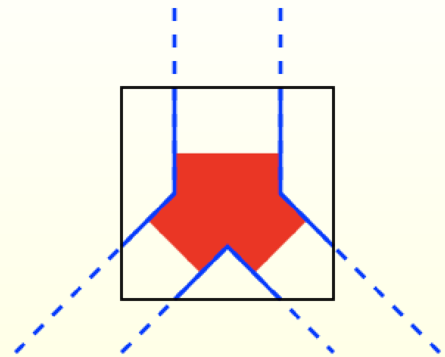


Persistent Local Homology

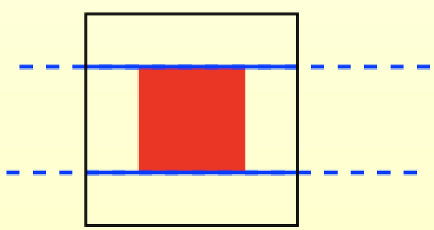
Different types of nodes in an Image Web:



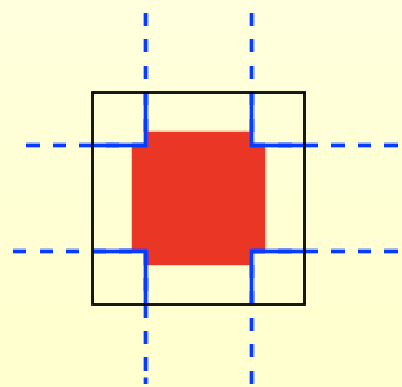
Branch 1



Branch 3

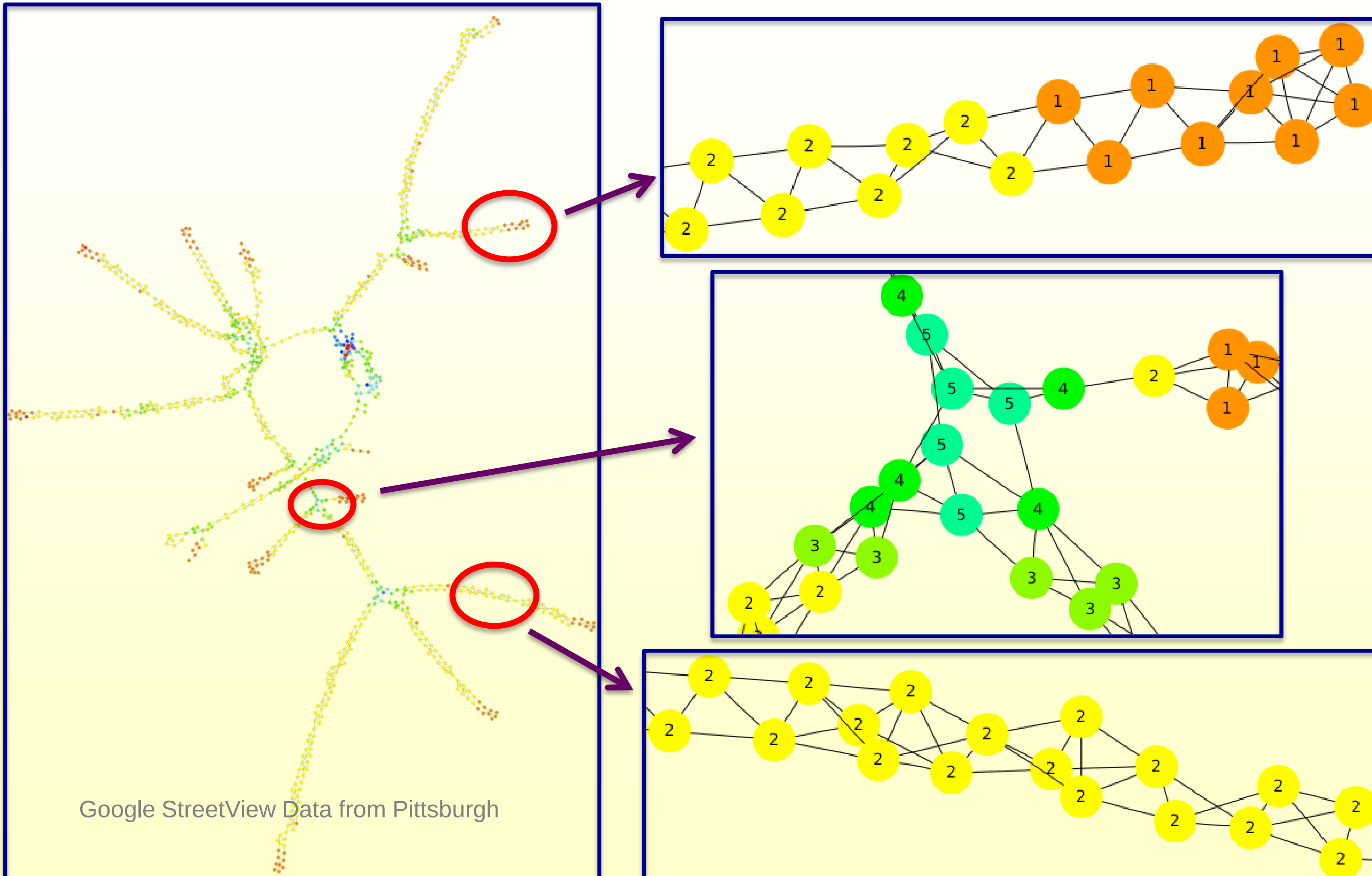


Branch 2

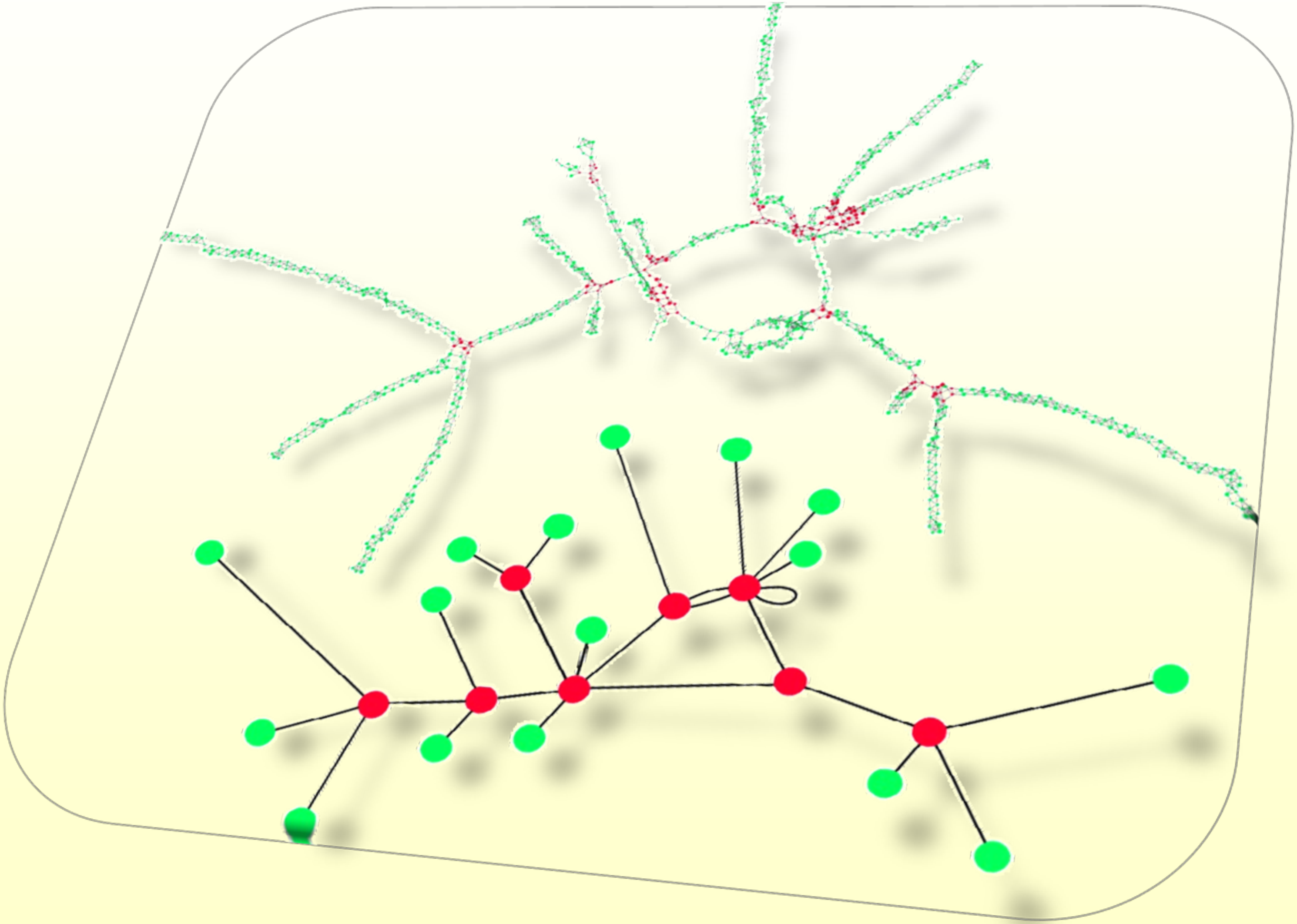


Branch 4

Persistent Local Homology



Summarizing Image Webs



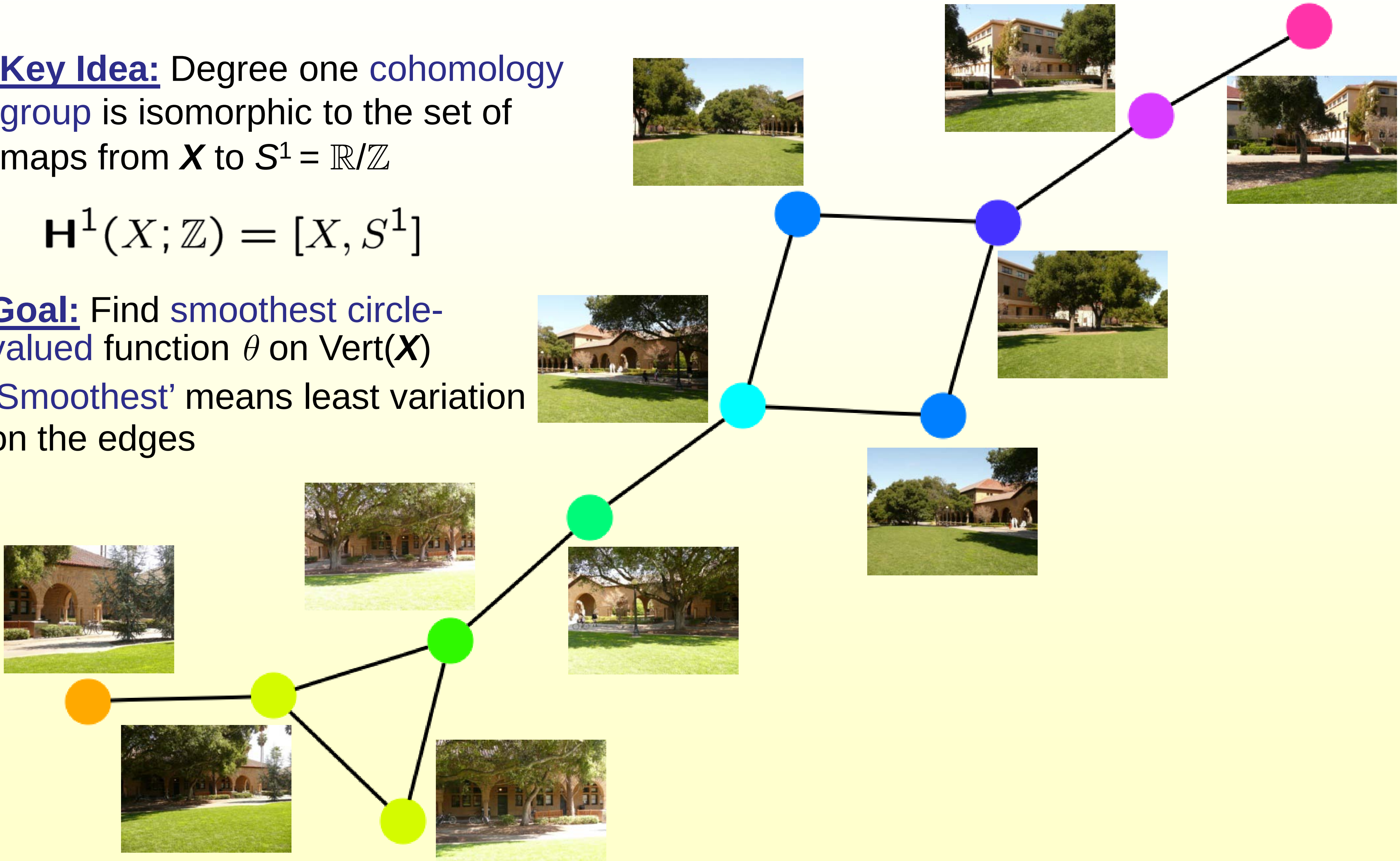
Parametrizing Edges/Loops

Key Idea: Degree one cohomology group is isomorphic to the set of maps from ***X*** to $S^1 = \mathbb{R}/\mathbb{Z}$

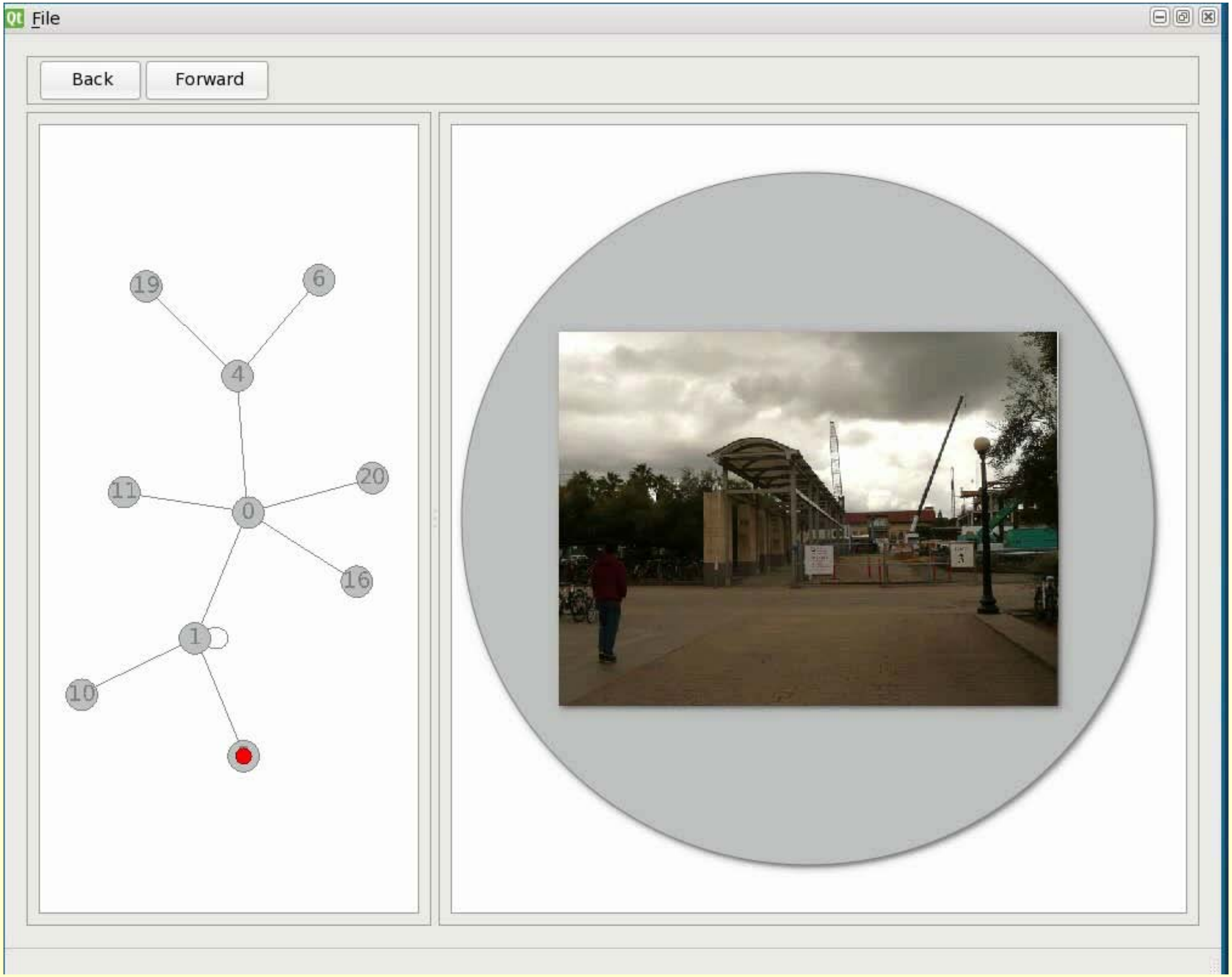
$$\mathbf{H}^1(X;\mathbb{Z}) = [X, S^1]$$

Goal: Find smoothest circle-valued function θ on Vert(***X***)

'Smoothest' means least variation on the edges

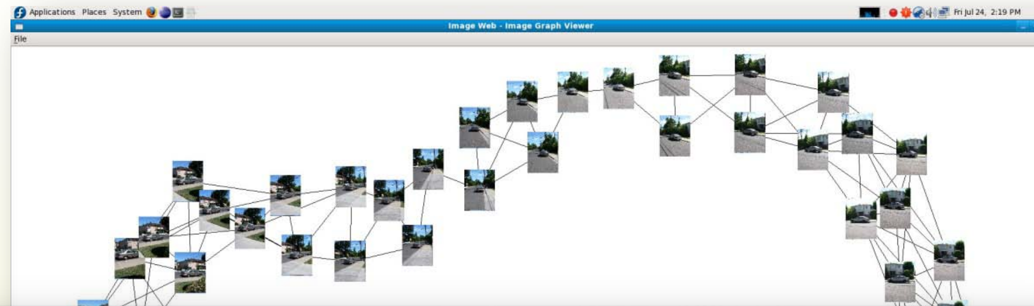


Web Navigation: Video



Information Transfer

- Object models as subwebs: focus and context
- Annotation transfer
- Linking people through their images: face webs
- Mobile webs: photo-guided navigation, collaborative exploration



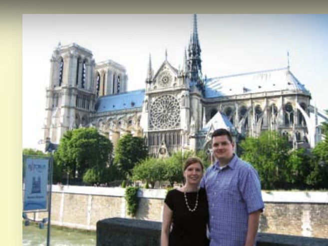
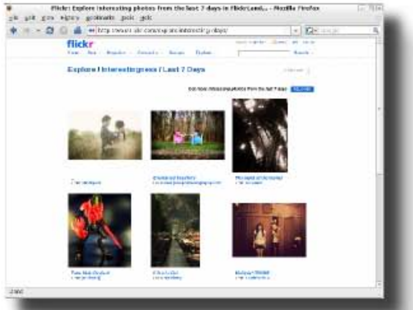


Image Annotation



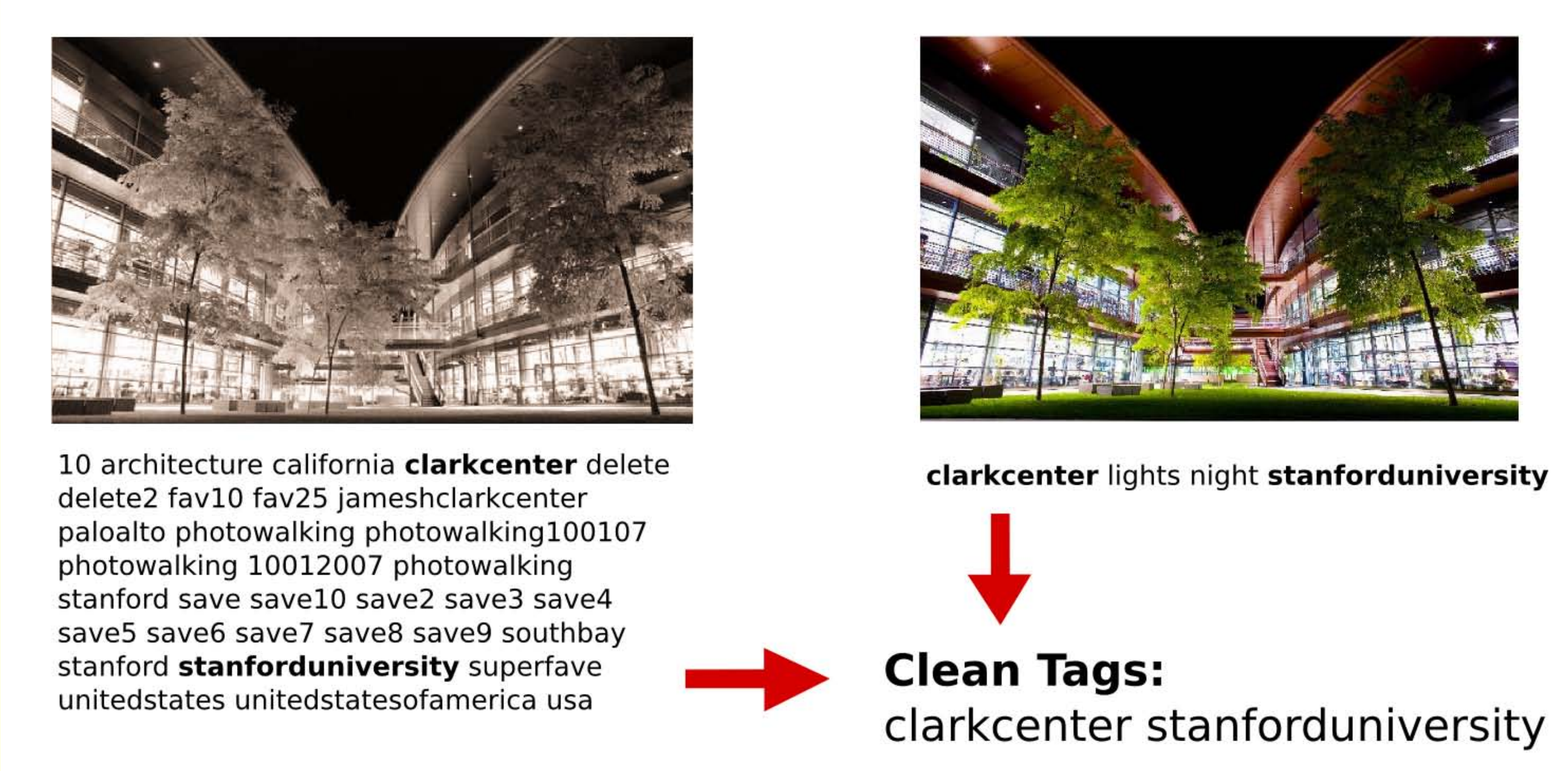
Flickr ~ 3.6 billion images
(2009)

stanford vacation
zurowskifamily

amerika catlovers city
goldengatebridge harbor may
monterey sanfrancisco
santacruz stanforduniversity usa

andreilinde dslr sony sonya700

Tag Cleaning by Simulated ESP Game on an Image Web



Auto-Tagging Experiment

- Dataset
  - Flickr search for “Stanford” -> 195,268 images
  - Image Web has 1,132,406 regions
  - Tag cleaning -> 525 clean tags

Auto-Tagging Experiment

- Tagging summary

	Provided Tags	Cleaned Tags	Suggested Tags
Number of Images	168,171	13,613	21,179
Percent of Dataset	86%	7.0%	11%

- Tagging examples

Successes

Image	Provided	Cleaned	Suggested
	church, stanford	stanford	stanford (1.01) stainedglass (0.04) university (0.03)
	<none>	<none>	stanford (0.189) california (0.173) university (0.073), stanforduniversity (0.041), hoovertower (0.001)
	stanford	stanford	stanford (1.05), university (0.15), tower (0.14), hoover (0.14), california (0.14), usa (0.13), 2009 (0.13)

Failures

Image	Provided	Cleaned	Suggested
	berkeley, stanford	berkeley, stanford	stanford (1.0357) berkeley (1.0298) university (0.1962) ca (0.1765)
	alcatraz, ghirardeli, goldengate, kipp, pier39, sandiego, sanfrancisco, stanford, streetsofsanfrancisco, students	alcatraz, goldengate, pier39, sanfrancisco, stanford	alcatraz (1.0019) goldengate (1.0019) pier39 (1.0019) sanfrancisco (1.0019) stanford (1.0019)

Auto Tagging Extensions

- Tag cleaning
  - Use graph neighborhood instead of direct matches to generate ESP game rounds

	Method 1: Direct match	Method 2: Neighborhood
Images that play a game	2868	3391
Images that “win” a game	2042	2813
Cleaned tags	130	242

- Tag suggestion
  - Estimate a relevance of a keyword to an image by the relative frequency with which it “wins” in ESP games with that image

Conclusion

- Interlinked images and other signals contain a wealth of information not apparent in any one image or signal alone
- Such signal webs form networks of maps; maps can be used to navigate as well as to transport information, so as to arrive at a global understanding of both the sensed environment and the acquisition process

- The value of these Webs is in the paths induced by the maps between images

Acknowledgements

- Joint work with


Kyle Heath


Natasha Gelfand


Maks Ovsjanikov


Dmitriy Morozov


Mridul Aanjenaya

and thanks to

NOKIA Connecting People

STANFORD COMPUTER SCIENCE

ONR Office of Naval Research

Google

AHPCRC Army High Performance Computing Research Center

NSF National Science Foundation WHERE DISCOVERIES BEGIN