Scene Chronology

{kmatzen,snavely}@cs.cornell.edu   http://bit.ly/cornell-chrono

Kevin Matzen
Noah Snavely

Cornell University

Problem Statement

We can now build planet-scale reconstructions [1]. Have we solved the 3D reconstruction problem?

NO! We often assume scenes are static, but even fixed physical structures can undergo dramatic visual changes on time scales from years down to days.

Cities are a great example! Billboards, signs, businesses, and street art are constantly changing. How can we **chronicle** these changing scenes at the granularity of weeks or days?

A Brief History of Times Square

Key challenges
- **Representation**: How do we represent dynamic 4D scenes?
- **Estimation**: How do we compute these 4D scenes from data and scale to millions of online photos?

4D Representations

Possible representations:
- 3D points augmented with temporal information
- Consistent segmentation of spatio-temporally coherent regions

We do both:
Step 1: Estimate a per-point time interval
Step 2: Cluster the points into coherent *plane-time cuboids*

Per-Point 4D Reconstruction

Semantic Spatio-Temporal Segmentation

Per-Point 4D Reconstruction

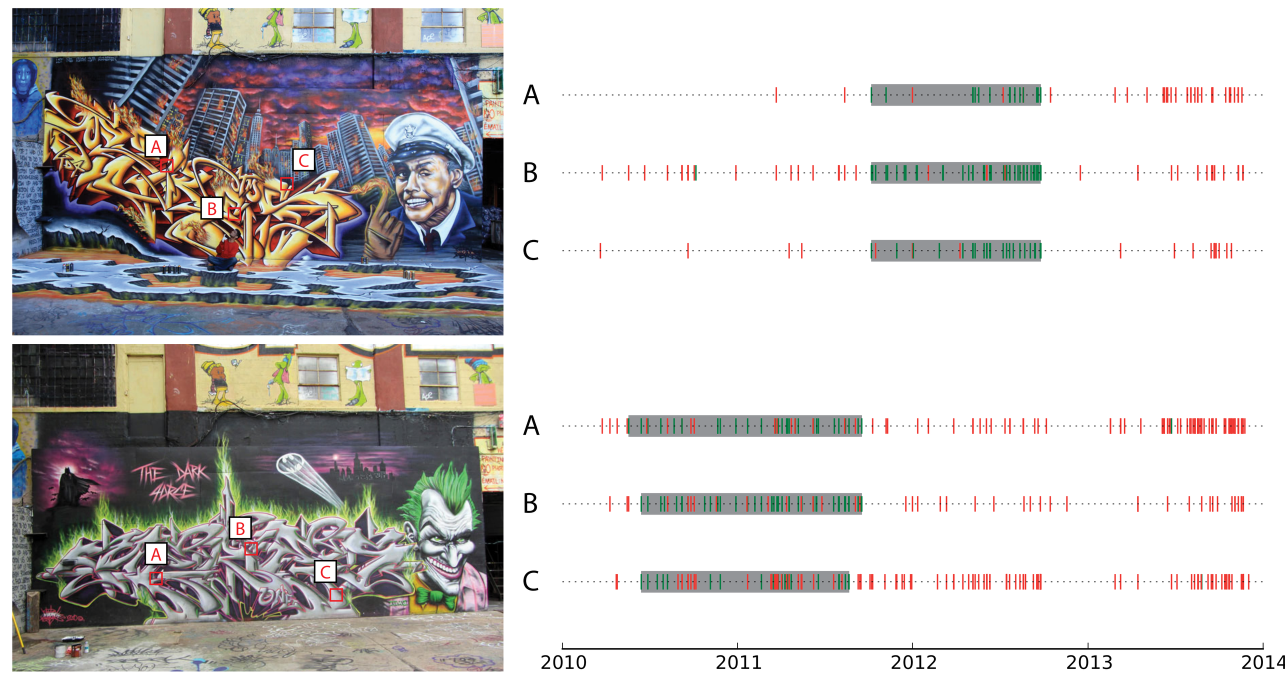
Observation: Standard tools such as Bundler + CMVS/PMVS work well to recover 3D content.

Problem: How can we recover per-patch temporal information?

Challenge: We want to use photo timestamps, but timestamps can be noisy and unreliable.

Solution: Use many positive and negative observations to robustly find a contiguous temporal interval of existence.

We find the interval with the maximal F_1 score to handle outlier observations.



Two co-located murals with observations and estimated intervals. Positive observations in green, negative observations in red, and estimated intervals in gray.

Semantic Spatio-Temporal Segmentation

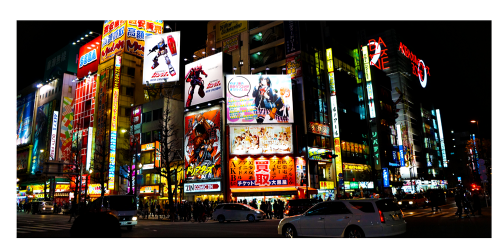
Problem: We have noisy per-point time intervals. How do we robustly segment these points into temporally coherent regions?

Two step approach:
(1) Build a spatio-temporal graph with point-to-point affinity defined by **distance**, **normal difference**, and **time interval overlap**.
(2) Use a **plane-time RANSAC** algorithm to find a good hypothesis, select the inliers as a cluster, and recursively apply to remaining connected components.

Datasets

Dataset	Photos Retrieved	Photos Registered	Patches	Positive Obs.	Negative Obs.
Times Square	1.2 million	246,000	13.6 million	5.3 billion	12.3 billion
Akihabara	48,000	12,700	12.2 million	635 million	821 million
5Pointz	171,000	13,900	1.7 million	132 million	254 million


Times Square - Iconic NYC theatre center.


Akihabara - Vibrant Tokyo electronics shopping district.


5Pointz - Living NYC graffiti art exhibit.

Visualization

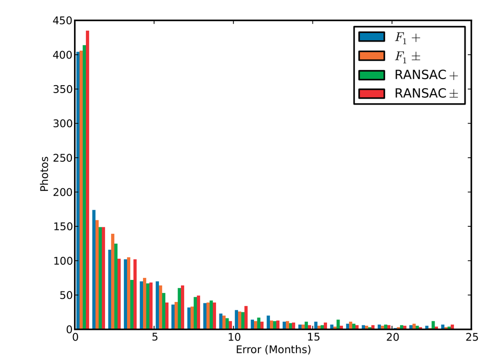
We can choose any point in recent history to visualize the scene. Please visit our website for more details.

Two snapshots of 5Pointz

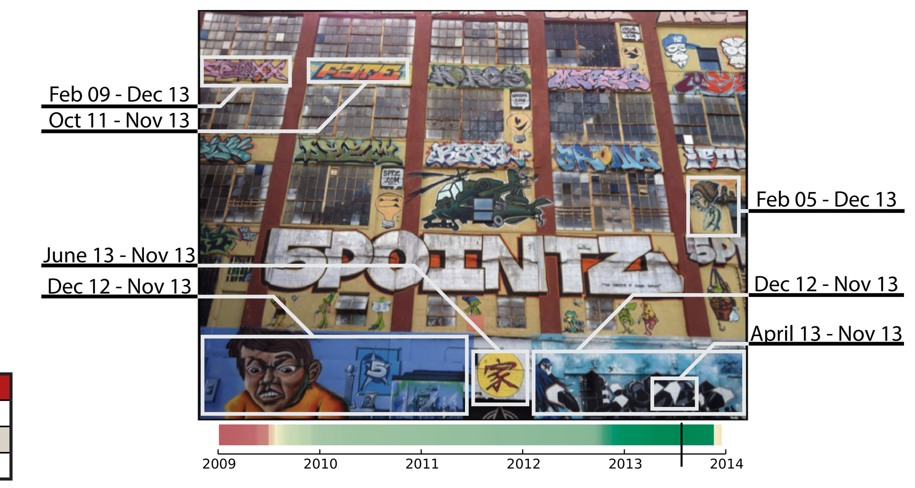
Changing of signs as Akihabara businesses come and go

Timestamp Prediction

Task: Register a novel photograph to our reconstruction and use observations to pinpoint date taken.



Percentile	F1+	F1 ±	RANSAC+	RANSAC ±
25	0.649	0.630	0.587	0.551
50	2.58	2.57	2.57	2.58
75	6.90	6.47	7.44	7.05



Left: Distribution of prediction error in months for 5Pointz. Right: An example photo with highlighted space-time components. Timeline is colored red for low probability and green for high probability with the mode marked.

Limitations

- Assume we can reconstruct the 3D scene
- Assumes enough redundancy to identify incorrect timestamps
- Some semantic elements are periodic
- Segmentation granularity dependent on thresholds

References

[1] Bryan Klingner, David Martin, James Roseborough. "Street View Motion-from Structure-from-Motion". ICCV 2013.
[2] Grant Schindler and Frank Dellaert. "Probabilistic Temporal Inference on Reconstructed 3D Scenes". CVPR 2010.

Acknowledgments

This work was funded in part by NSF grants IIS-1111534 and IIS-1149393, by support from the Intel Science and Technology Center for Visual Computing, and by support from Amazon Web Services in Education.