Clues from the Beaten Path: Location Estimation with Bursty Sequences of Tourist Photos

Chao-Yeh Chen and Kristen Grauman
The University of Texas at Austin

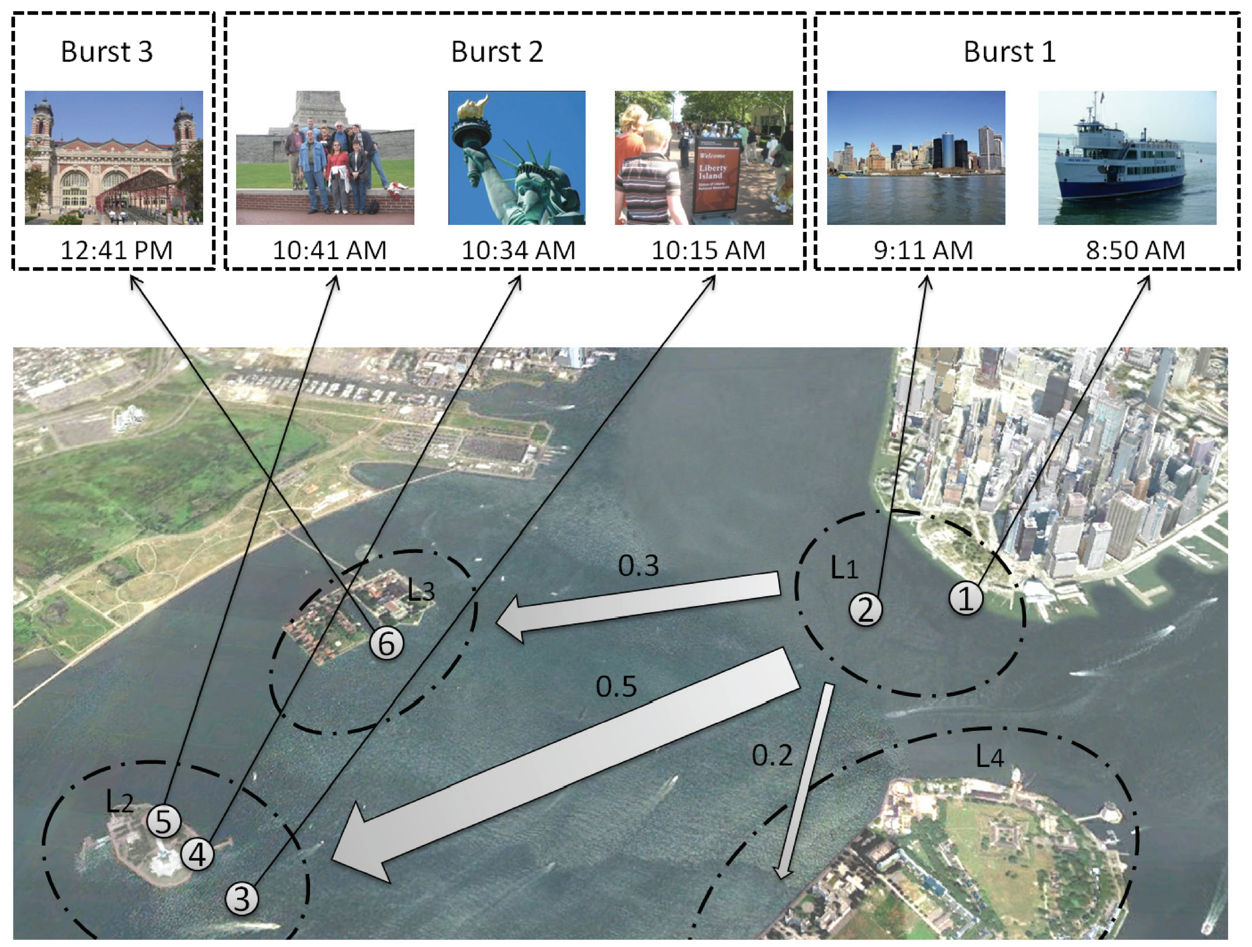
Problem

Where did I take these images?

Photos of distinctive "landmarks" are recognizable with current methods, but when traveling, people also take pictures of themselves, food, cars, storefronts,…

Main Idea

Hypothesis: tourists often take similar paths through a city
- Exploit these patterns to recognize *sequence* of photos
- Novel set-to-set likelihood treats each "burst" as observation



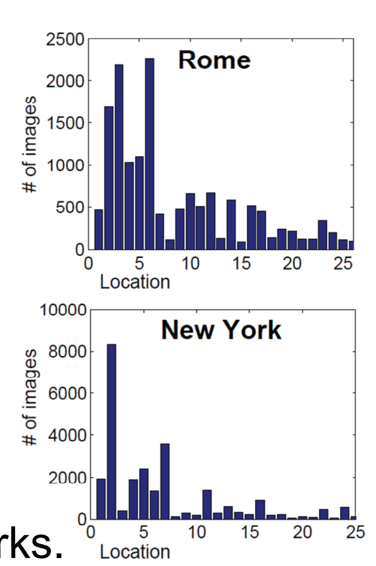
Previous work

- Nearly all work recognizes photos individually [Philbin et al. 2007, Schindler et al. 2007, Hays et al. 2009,…]
- Exploit learned *physical travel constraints* across the globe [Kalogerakis et al. 2009]
- Predict labels with *context of local window* of nearest frames [Li et al. 2009]

Data

Properties of Flickr datasets

Dataset	Rome	New York
# Defined Locations	26	25
Ave Location Size	0.2 mi^2	3 mi^2
# Train\Test Images	32942\22660	28950\28250
# Train\Test Users	604\470	665\877
Avg # photos per test seq	52 (std 119)	37 (std 71)
Avg time period of test seq	3.77 days	3.33 days



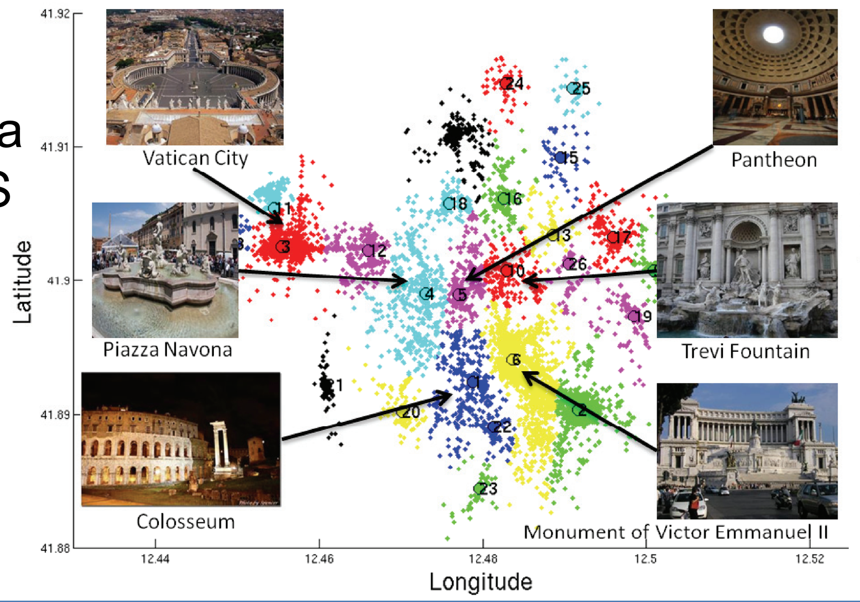
Less than 50% of images contain distinctive landmarks.

Approach: Training stage

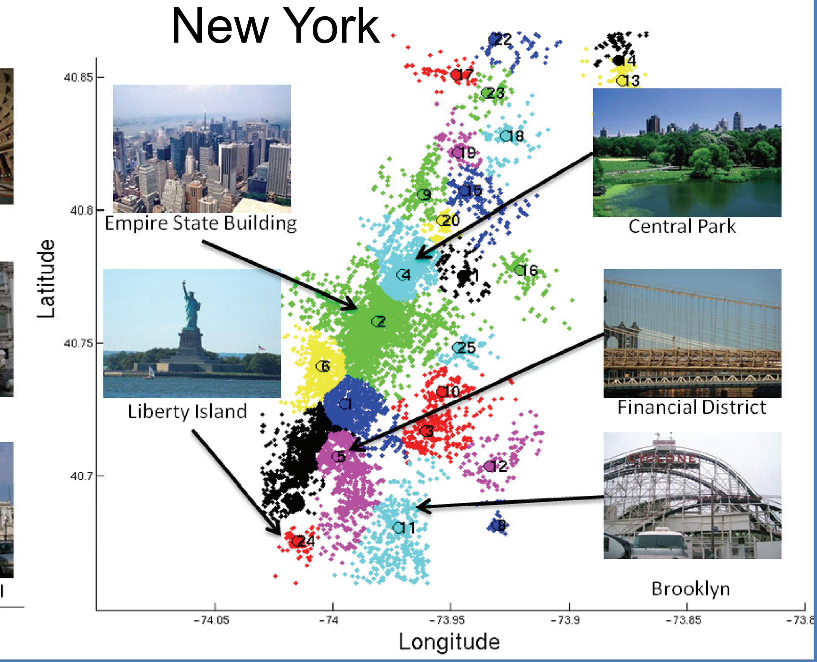
Discovering a city's locations

Define locations via mean-shift on GPS coordinates of training images

Rome



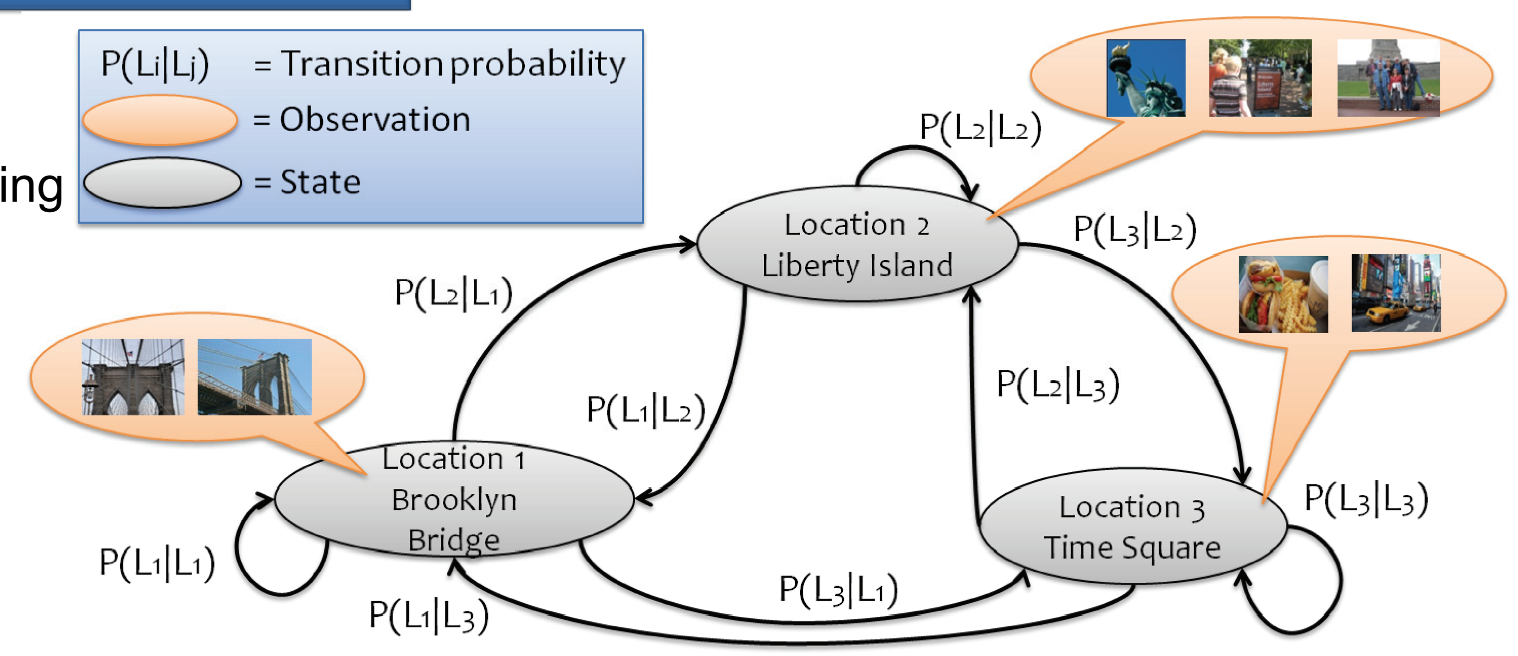
New York



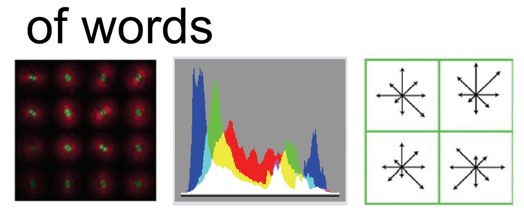
Learning the HMM & feature extraction

Learn *initial state* priors and *state transition* probabilities from traveling image sequences

P(Li|Lj) = Transition probability
= Observation
= State



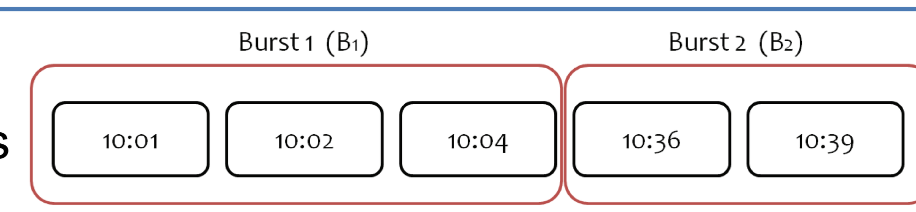
Visual features: Gist, color histogram, bag of words



Approach: Testing stage

Grouping photos into bursts

Mean shift on the timestamps to compute bursts

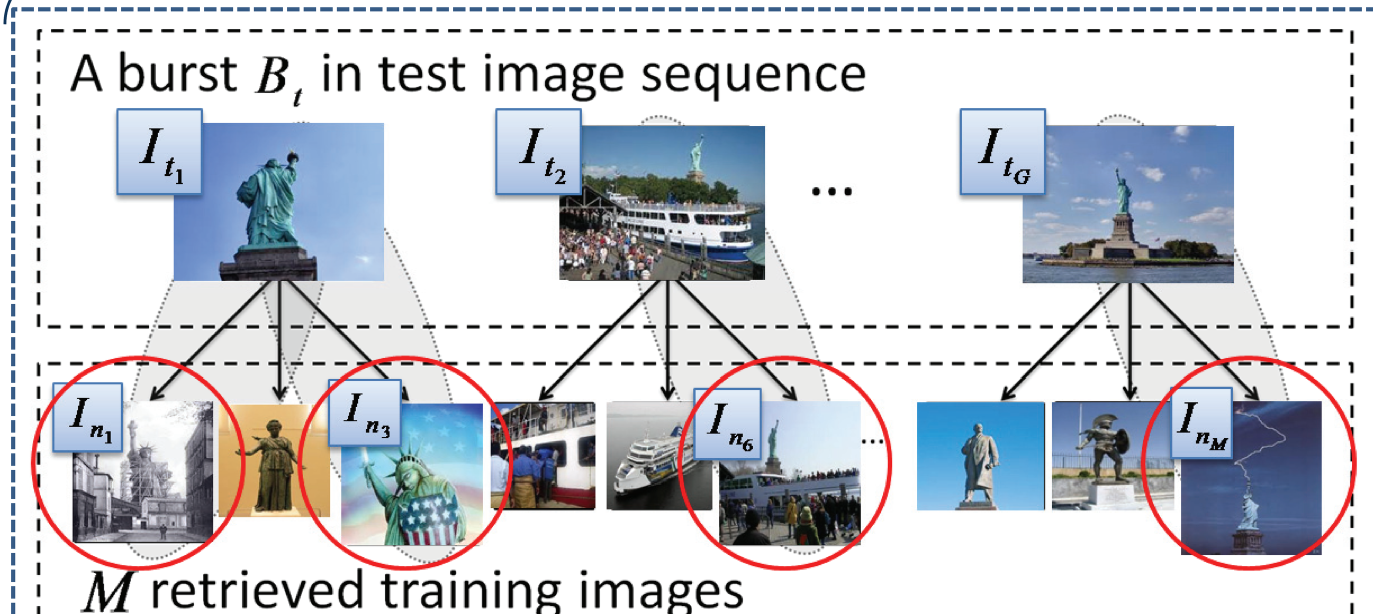


Location estimation via HMM inference

We define a novel set-to-set matching-based observation likelihood for bursts, to reduce the effect of *non-distinctive* and *non-informative* images.

The distance $D(I_t, I_m)$ comes from the visual feature similarity between two images

$$P(L_t = i | I_{t_1}, \ldots, I_{t_G}) \propto \left(\sum_{m \in M_i} \omega(I_m) \right) + \lambda_c, \quad \omega(I_m) = \frac{\exp(-\gamma D(I_{t_*}, I_m))}{\sum_{l=1}^{M} \exp(-\gamma D(I_{t_*}, I_{n_l}))},$$



To compute $\sum \omega(I_m)$,
- Given a burst B_t containing G images, retrieve K neighbors for each test image
- Images inside red circles are from L_i,

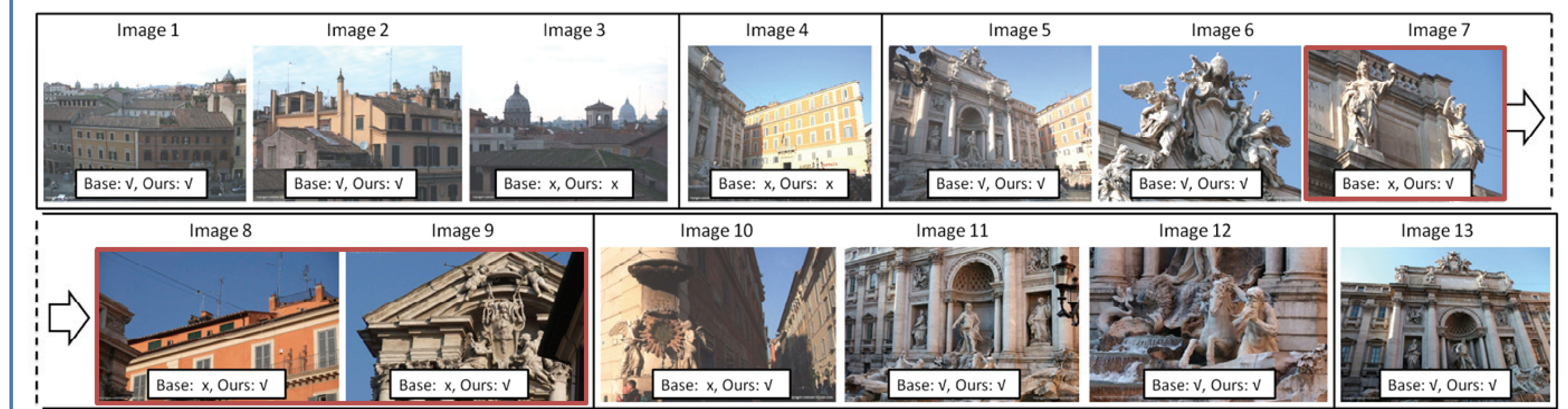
➔ $M_i = \{I_{n1}, I_{n3}, I_{n6}, I_{nM}\}$.

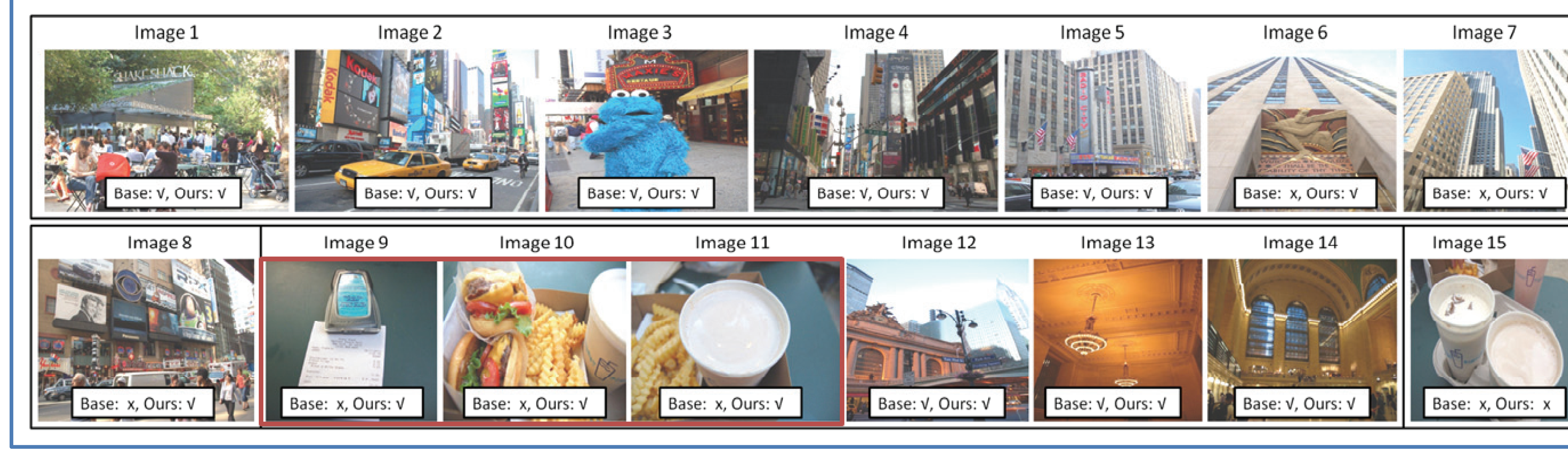
Using the initial state priors and state transition probabilities, we infer the most likely sequence of location labels using the Forward Backward algorithm.

Results

Advantage of proposed Burst-HMM

Baseline: img-to-img HMM





Location estimation accuracy

	NN	Img-HMM	Int-HMM	Burst Only	Burst-HMM (Ours)
Avg/seq	0.1502	0.1608	0.1728	0.1764	**0.2036**
Overall	0.1592	0.1660	0.1771	0.2617	**0.2782**

(a) Rome dataset

	NN	Img-HMM	Int-HMM	Burst Only	Burst-HMM (Ours)
Avg/seq	0.2323	0.2124	0.2330	0.2099	**0.3021**
Overall	0.2302	0.2070	0.2304	0.2055	**0.3143**

(b) New York City dataset

NN: nearest neighbors [Hays et al. 2009]
Img-HMM: simple img-to-img HMM
Int-HMM: same, but with time intervals [Kalogerakis et al. 2009]
Burst Only: our model, but without travel priors

Travel guides' beaten paths

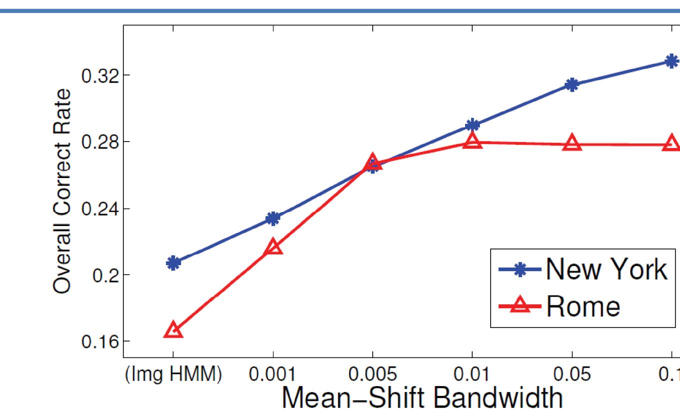
7 itineraries for spending "3 days in NYC" from 7 popular travel guides

	Rand. Walk	Rand. Walk (TS)	Guidebook
Route Prob.	$6.3 \cdot 10^{-12}$	$4.2 \cdot 10^{-11}$	$2.0 \cdot 10^{-4}$

Our learned model agrees with suggested routes!

Impact of burst density

The Img-HMM is a special case where the bandwidth is fixed at 0.



Conclusion

- Travel patterns (beaten paths) strengthen within-city location recognition.
- Event related burst and set-to-set likelihood better handle spectrum of photos from real tourists, relative to several existing approaches.