Urban-scale quantitative visual analysis

Josef Sivic

with D. Crandall, C. Doersch, A. Efros, A. Gupta, S. Lee, N. Maissonneuve, S. Singh

CityLabs@INRIA

INRIA – Willow Project

Département d'Informatique, Ecole Normale Supérieure, Paris

What Makes Paris Look like Paris?

[Doersch, Singh, Gupta, Sivic, Efros SIGGRAPH'12]

One of these is from Paris

Raise your hand if …

…this is Paris

Raise your hand if …

Humans are good but
what about machine?

Can machine recognize visual elements characterizing an urban area?

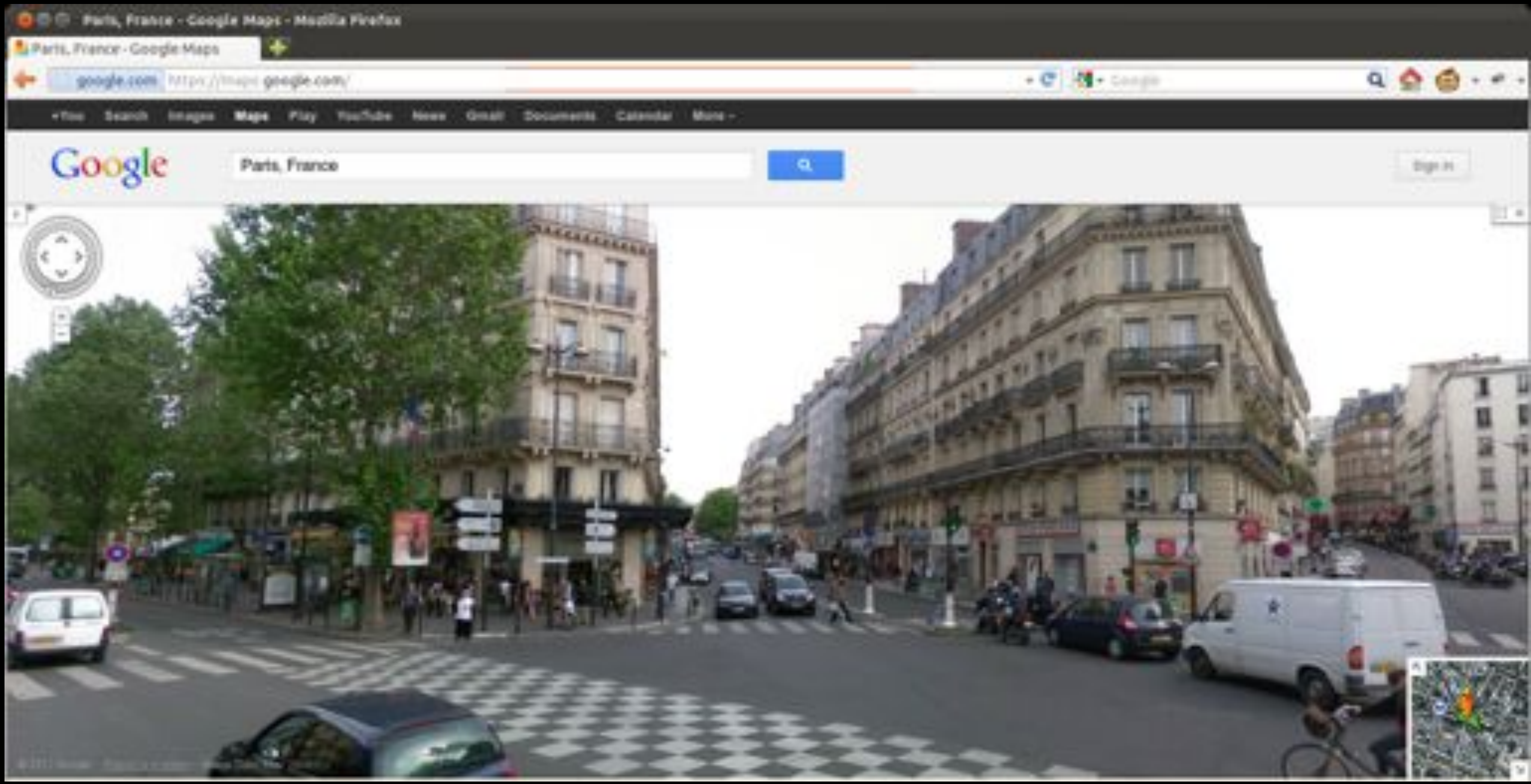
Our Goal:

Given a large geo-tagged image dataset,
*we automatically discover **visual elements***
that characterize a geographic location

Our Hypothesis

- The visual elements that capture Paris:
  - Frequent: Occur often in Paris
  - Discriminative: Are not found outside Paris

Map based imagery provides a comprehensive visual record of a city




Paris


Courbevoie
Nanterre
Pantin
Bobigny
Montreuil

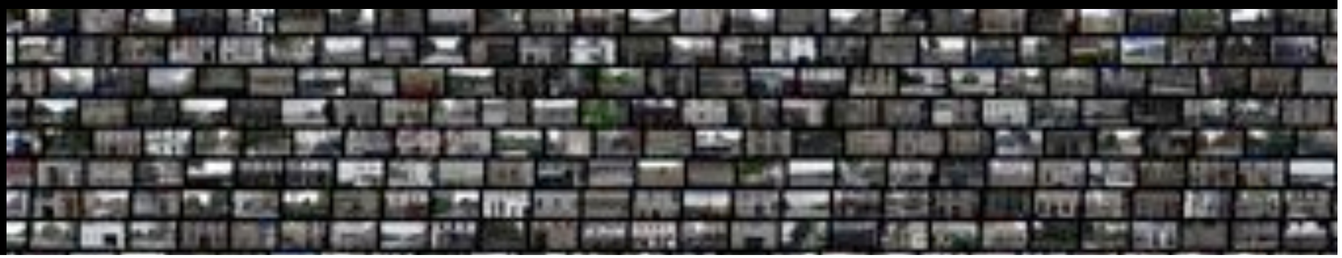

London
Prague
Paris
Milan
Barcelona
New York
Boston
San Francisco
Philadelphia
Mexico City
Tokyo
Sao Paulo
Positive Set
Negative Set

Hundreds of Millions of Patches

Scientific challenges

1. **Difficult learning problem**: How do you represent and automatically learn vocabulary of architectural elements characteristic for a city?

2. **Efficiency**: need to search through large amount of visual data (hundreds of millions of data points)



Problem formulation

Weakly supervised machine learning: [Bach and Harchaoui'08, Xu et al.'04]

Given a set of inputs x_i and supervisory meta-data y_i , $i = 1,\ldots,N$

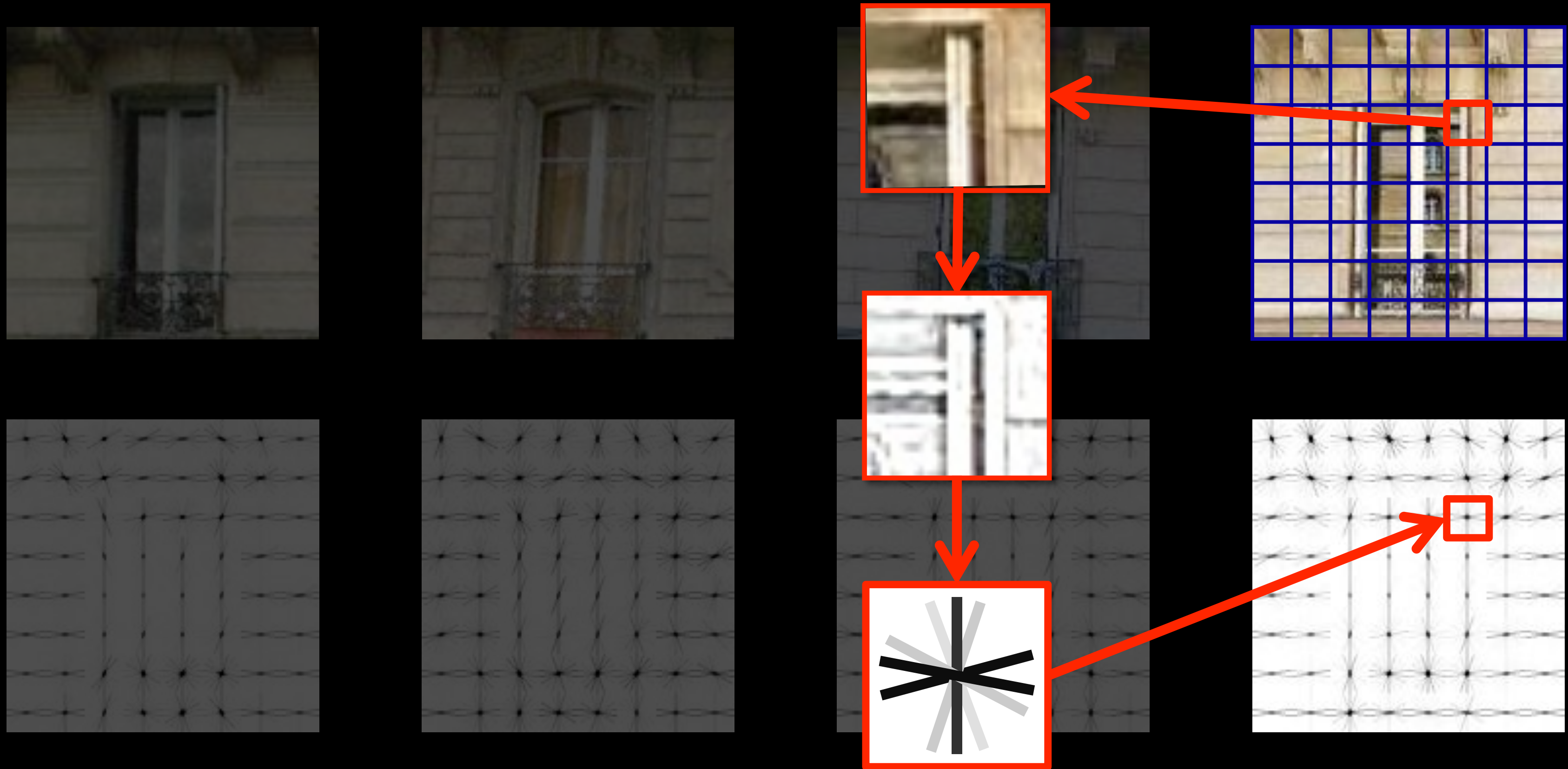
learn vocabulary $\hat{z}_i = f(x_i)$ by solving

$$\min_{f,z} \underbrace{\sum_{i=1}^{N} \ell(z_i, f(x_i))}_{\text{Discriminative loss on data}} + \underbrace{\Omega(f)}_{\text{Regularization}}$$

$$\underbrace{s.t. \quad g(z) = y}_{\text{Supervision from available meta-data}}$$

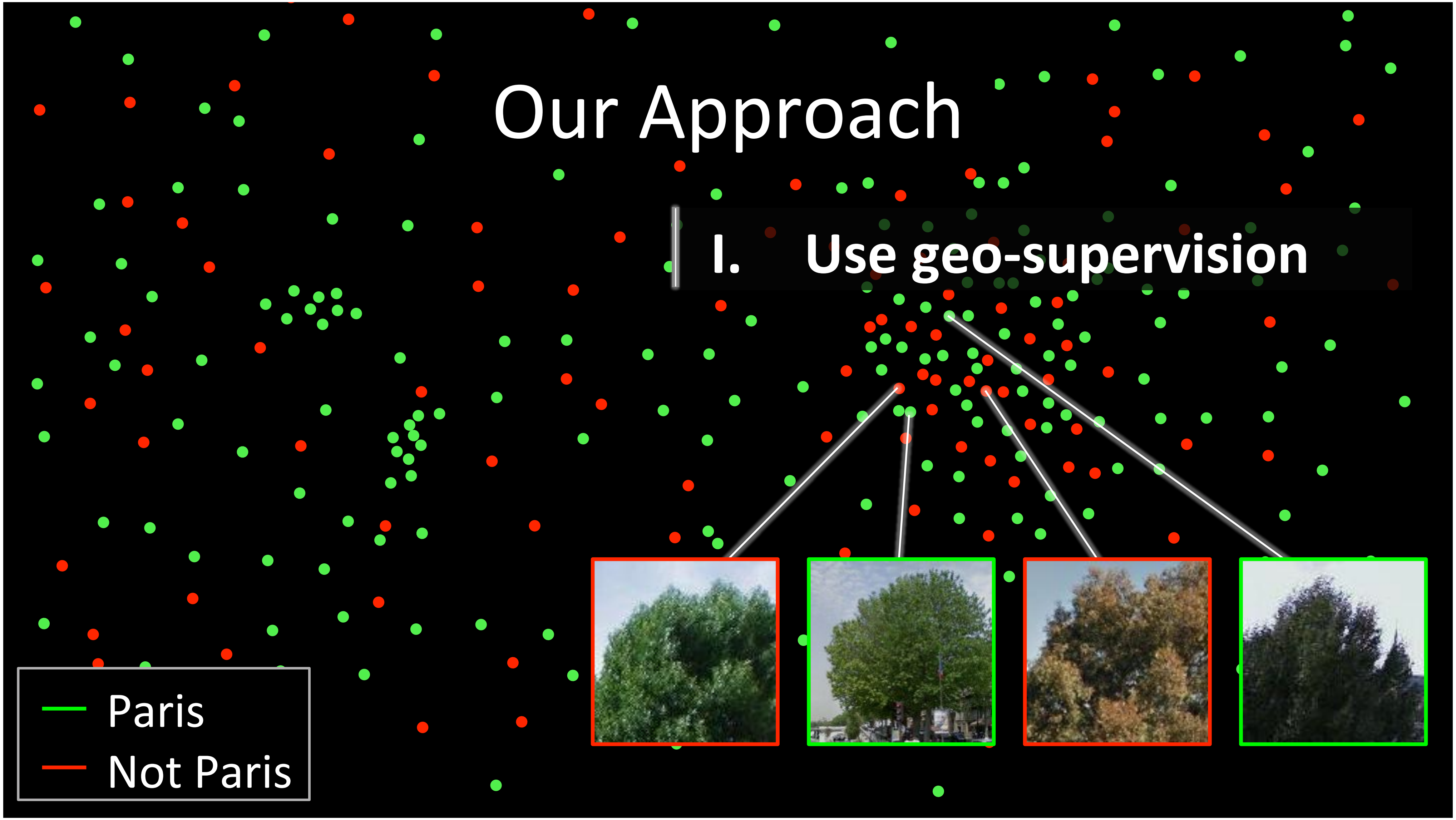
Input {x_i}: Millions of image patches extracted from street-view images from different cities
Supervisory meta-data {y_i}: geo-tags for each image

Representing Visual Patterns



[Dalal and Triggs, “Histogram of Oriented Gradients”]

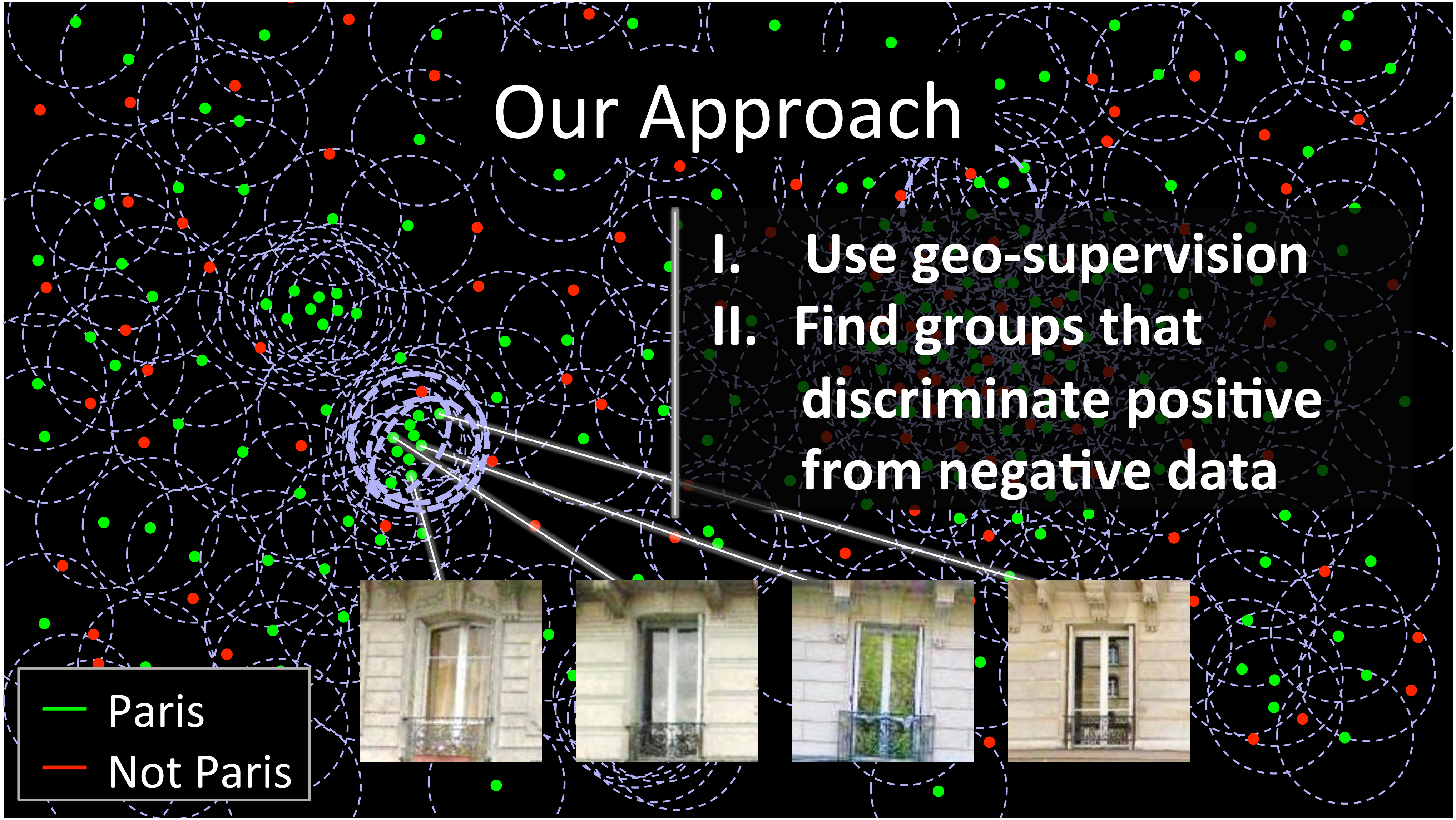
Our Approach


Our Approach
I. Use geo-supervision
Paris
Not Paris

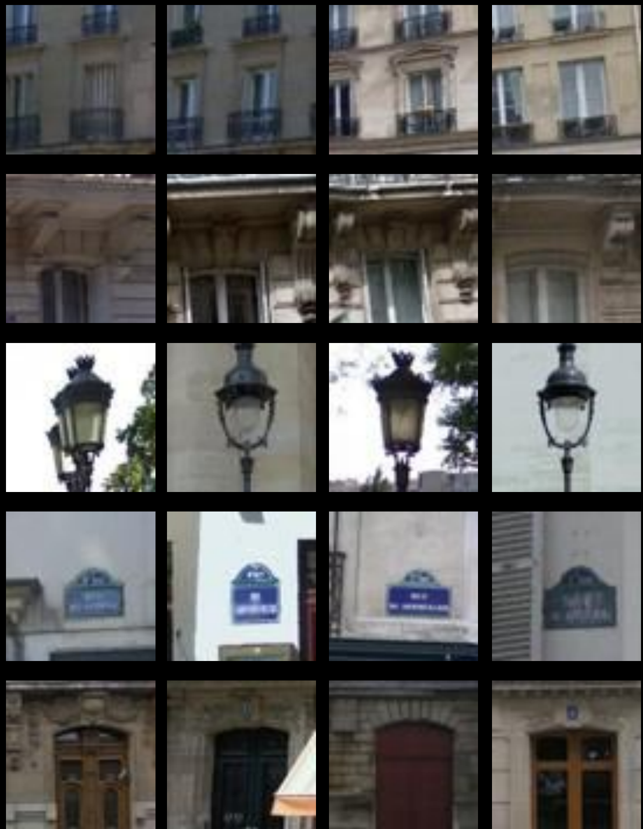
Our Approach

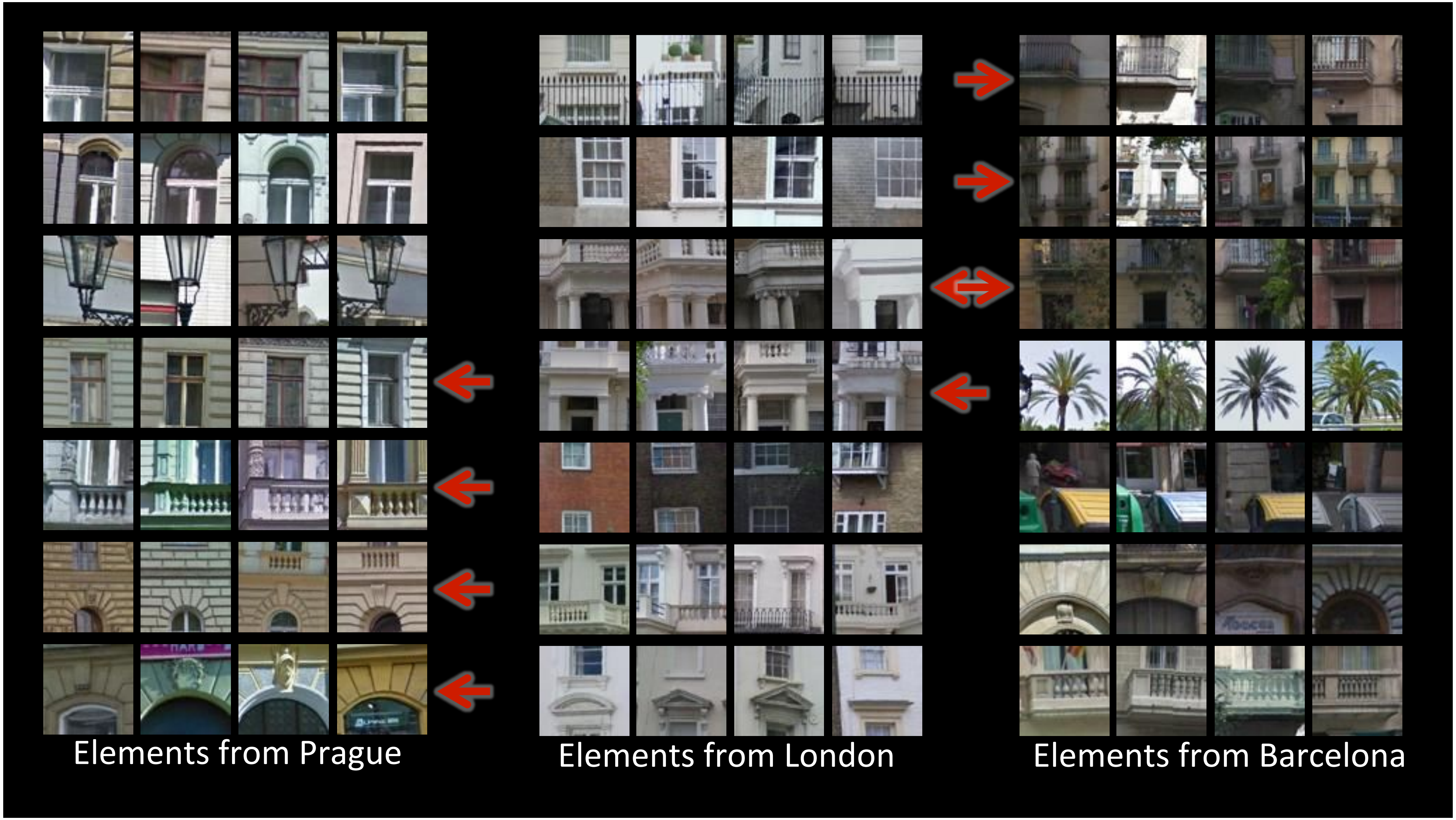
I. Use geo-supervision

II. Find groups that discriminate positive from negative data

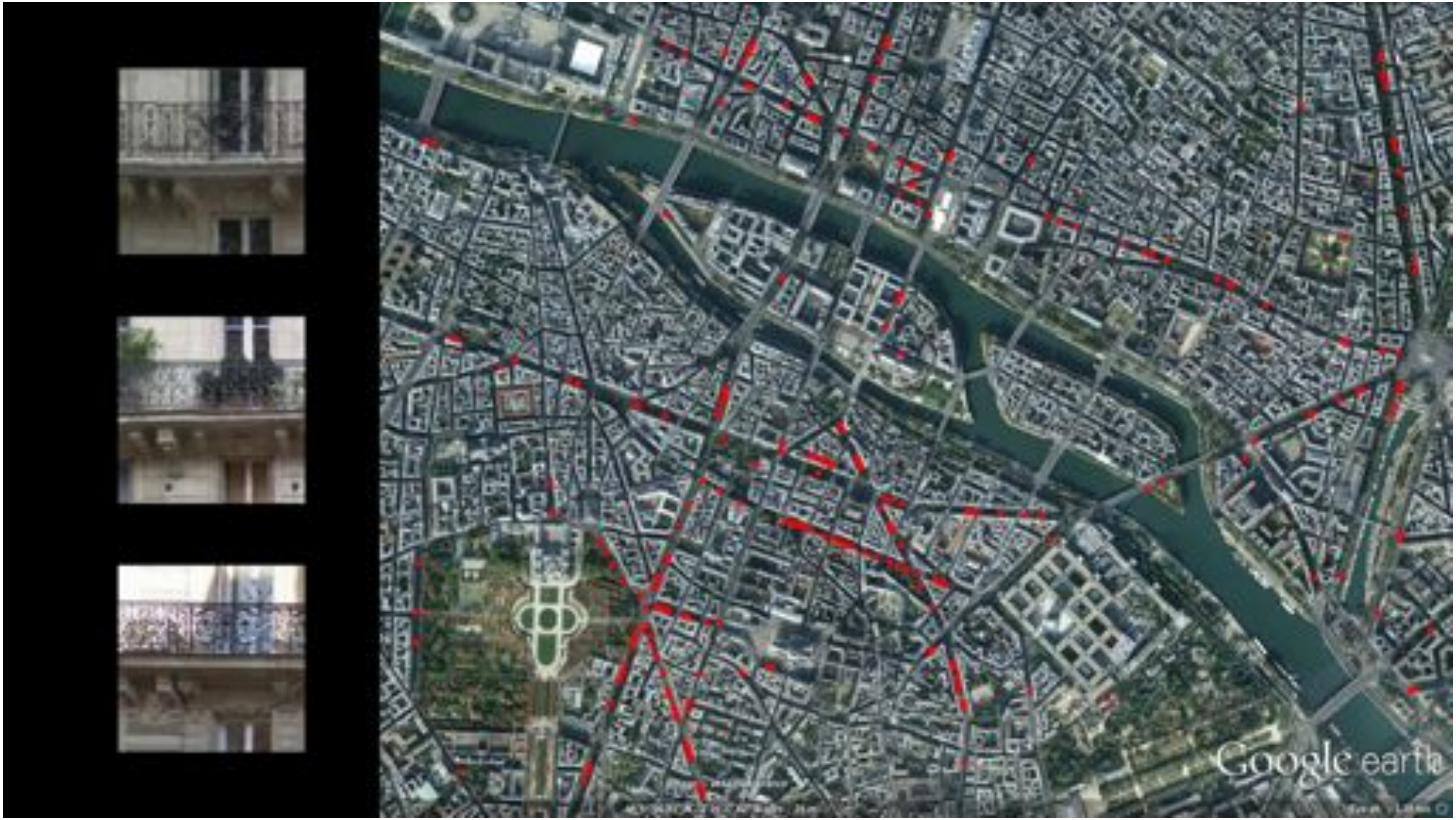


Paris: A Few Top Elements

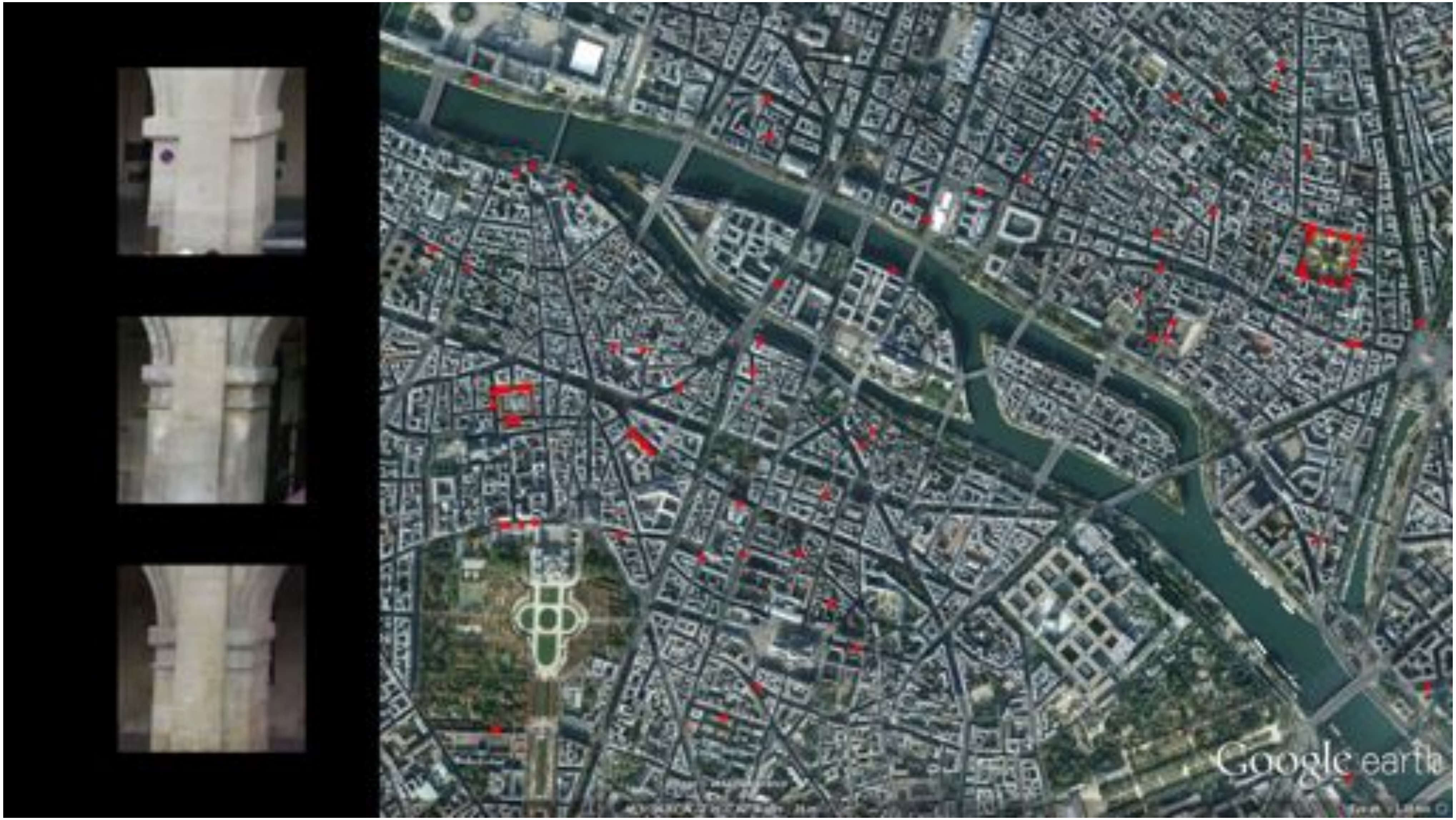


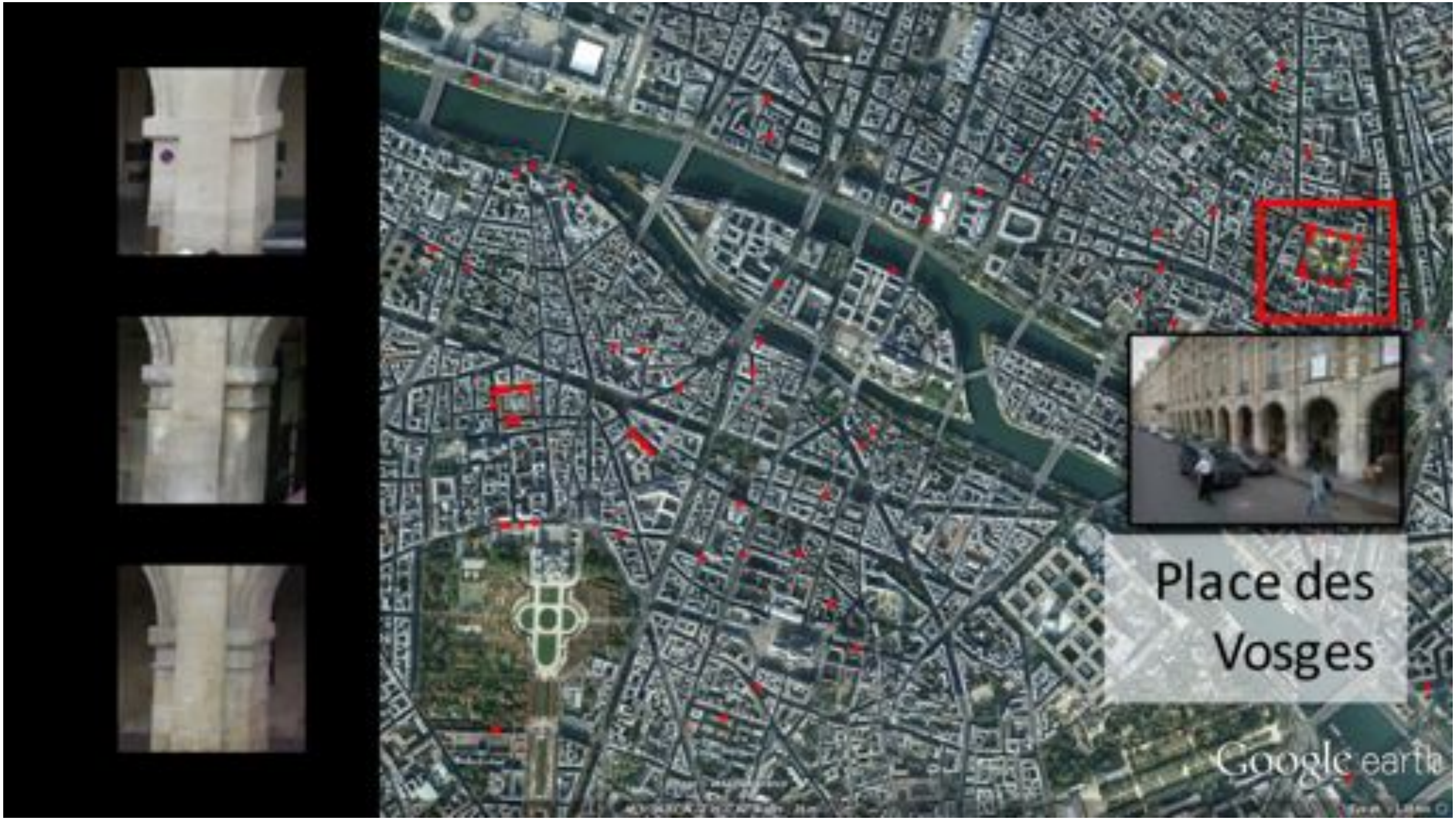

Elements from Prague
Elements from London
Elements from Barcelona


Google earth


Google earth


Google earth


Google earth


Place des
Vosges
Google earth


Place des
Vosges
Marché
St. Germain
Google earth


London
Prague
Paris
Milan
Barcelona
Google earth


London
Prague
Paris
Milan
Barcelona
Google earth


London
Prague
Paris
Milan
Barcelona
Google earth


Louvre
/Opera
Marais
Latin Quarter

Analyze architecture style over time

Geo-located Google street-view images

+

Cadastre maps

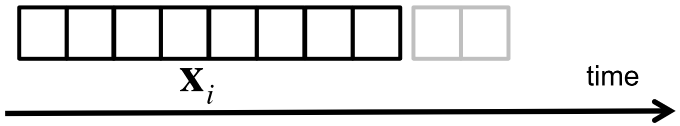
[Lee, Maissonneuve, Crandall, Efros, Sivic, ICCP 2015]

Cadastre map of Paris: 128k buildings with meta-data

•

Find temporal trends in architecture style

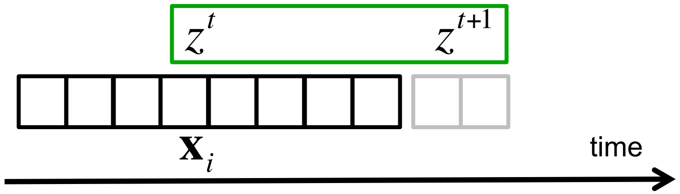
Problem: how to learn sequences of visual elements?



-1800 | 1801-1850 | 1851-1914 | 1915-1919

Find temporal trends in architecture style

Problem: how to learn sequences of visual elements?



$$\min_{f,z} \sum_{i=1}^{N} \ell(z_i^t, f(\mathbf{x_i})) \quad + \quad \Omega(f)$$

$$s.t. \quad g(z^t) = y$$

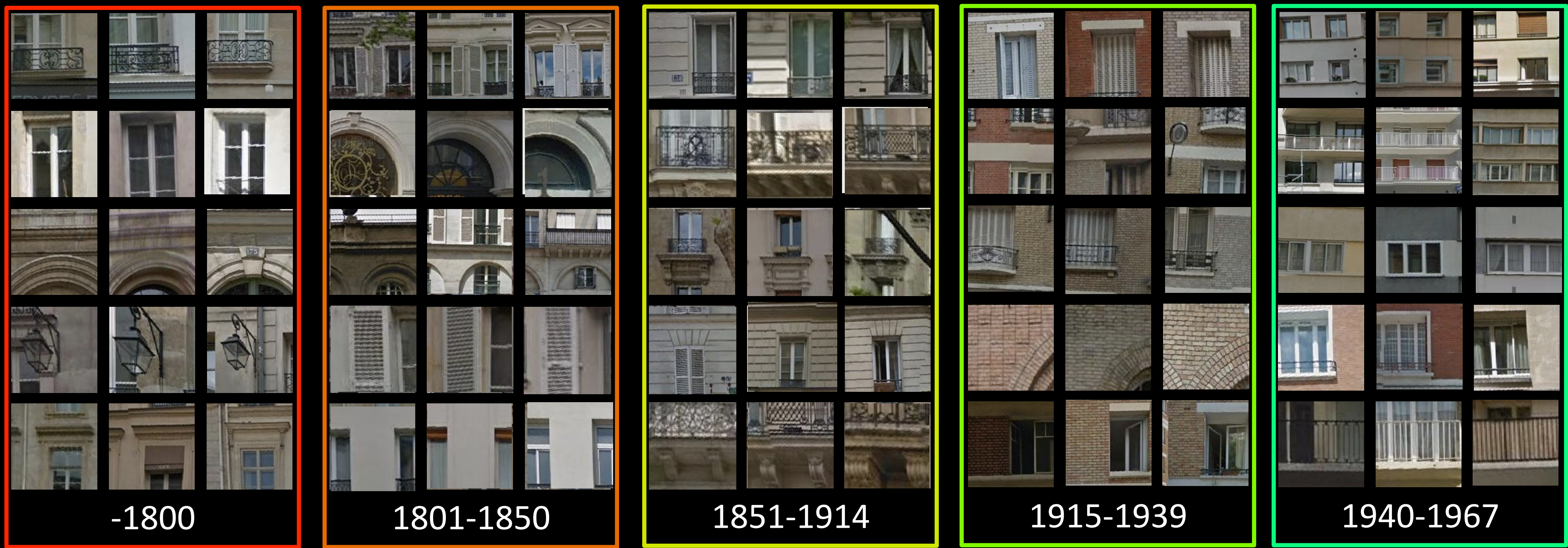
$$h(z^t) = z^{t+1}$$

Capture "**trends"** by an additional constraint on vocabulary labels **learnt** from data

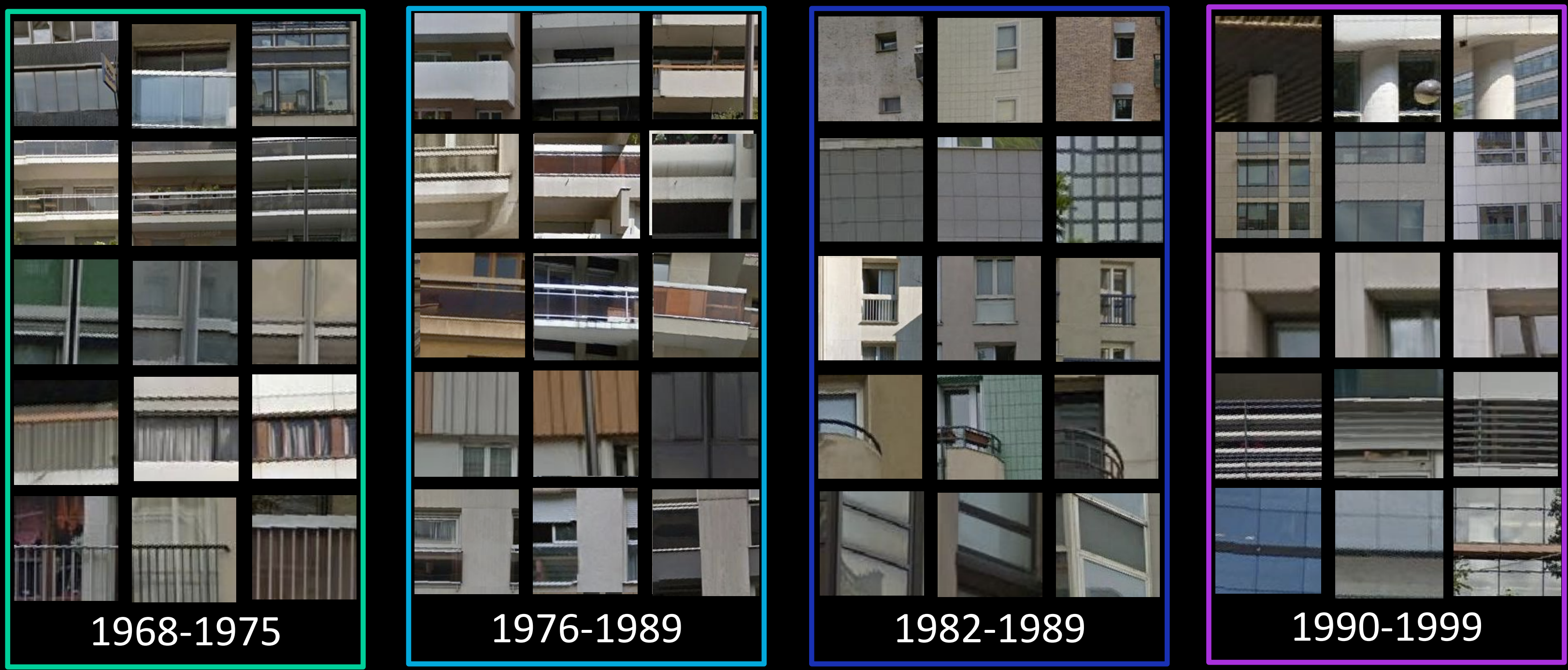
Scientific challenges:

- What is the **appropriate form** of temporal constraints?
- Can we learn **jointly f**, **z** and **h** from billions of inputs?

Visual elements specific for a time-period



Visual elements specific for a time-period



Evolution of architectural elements

-1800 | 1801-1850 | 1851-1914 | 1915-1919

[Lee, Maissonneuve, Crandall, Efros, Sivic, ICCP 2015]

Evolution of architectural elements

-1800 | 1801-1850 | 1851-1914

-1800 | 1801-1850 | 1851-1914

1851-1914 | 1915-1939

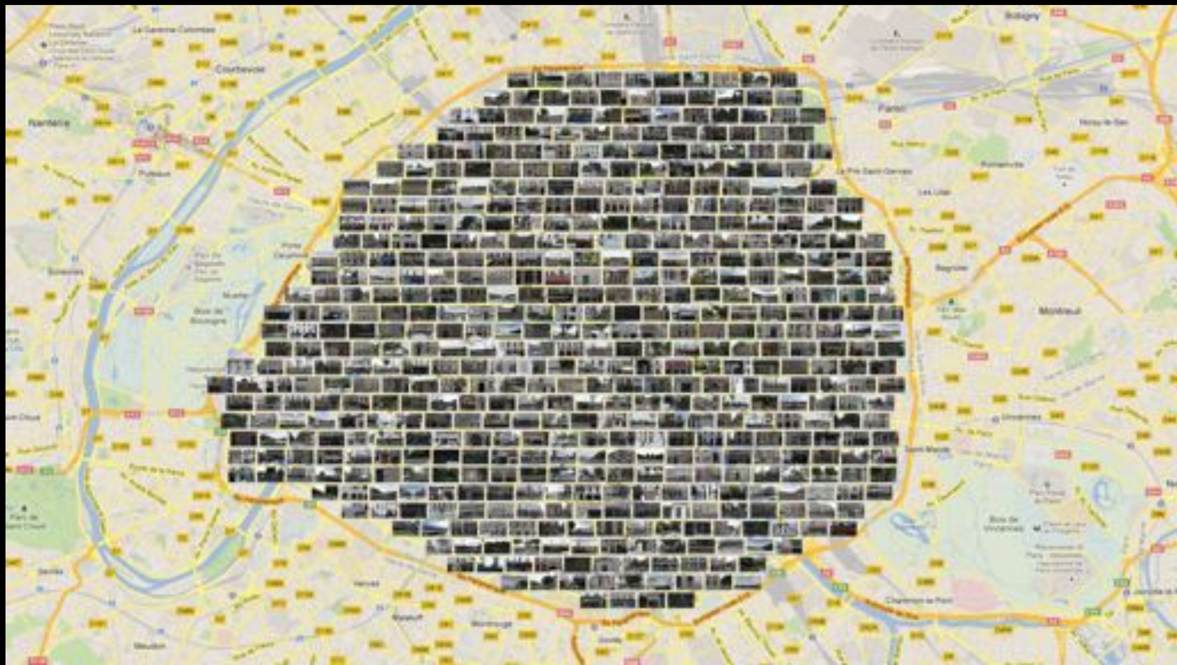
-1800 | 1801-1850 | 1851-1914 | 1915-1919

1801-1850 | 1801-1850 | 1851-1914 | 1915-1919

1940-1967 | 1968-1975 | 1976-1981 | 1982-1989 | 1990-1999

So far: analysis of Street-view imagery

- Static 2D record of the city at present time



What next?

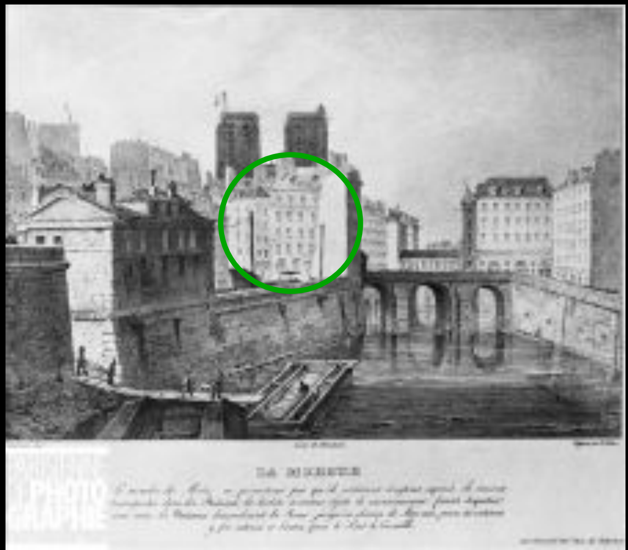
I. Historical urban visual record

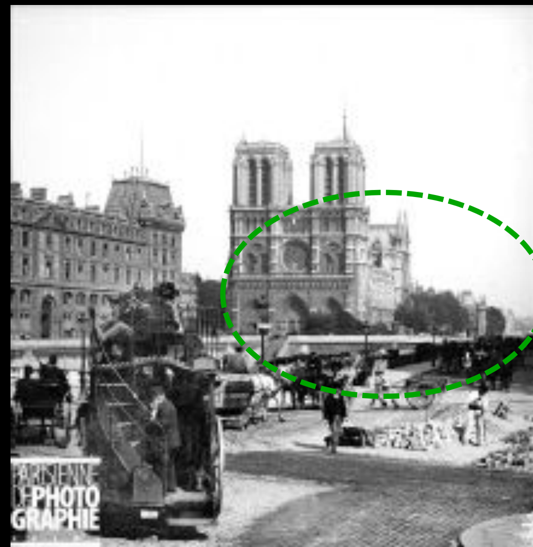
II. 3D urban visual record

III. Dynamic urban visual record

I. Historical urban visual record

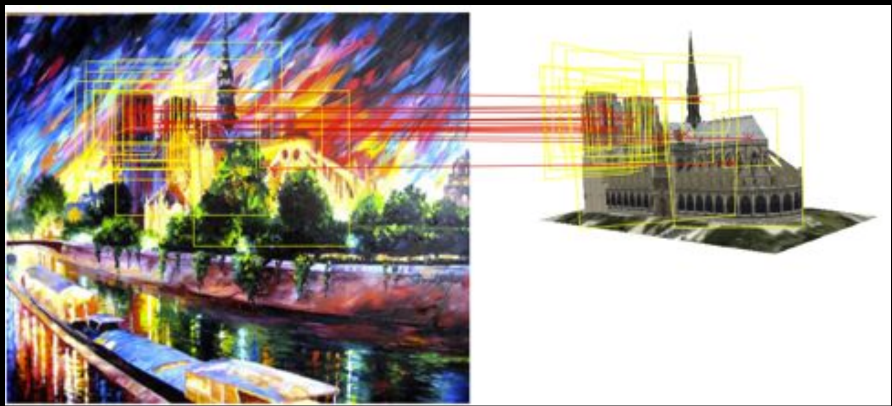
Quantify evolution of a particular place over time





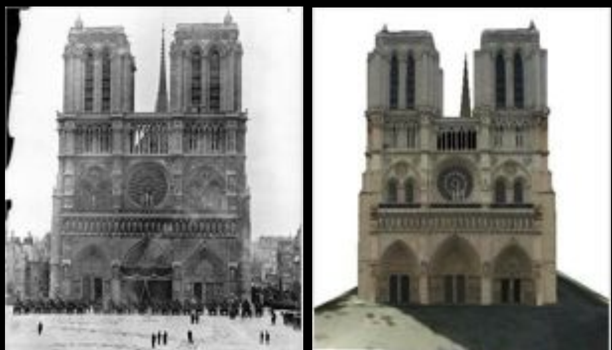
Applications: new ways to access archives for archeology, history, or architecture…

Example: painting to 3D model alignment

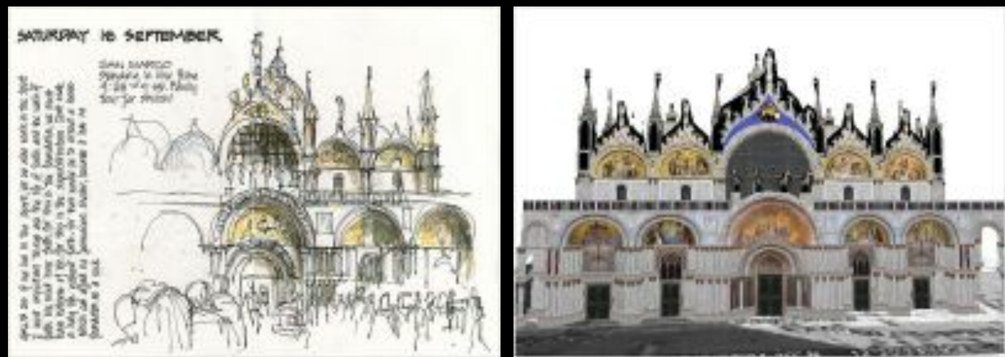


Painting

Painting



Historical photograph



Sketch

[Aubry, Russell, Sivic, *Painting-to-3D model alignment via discriminative visual elements*, Transactions on Graphics, 2014]


Google earth

II. 3D urban visual record

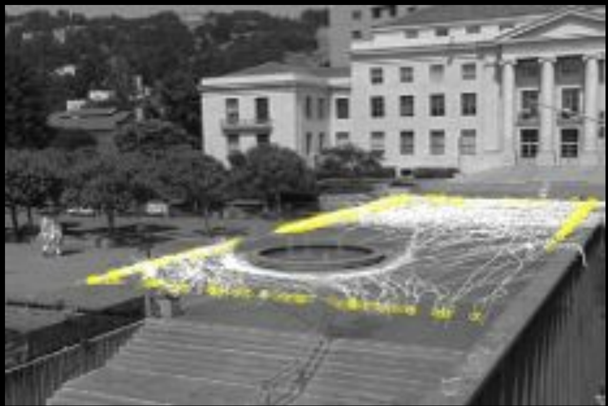
Goal: detailed semantic 3D reconstruction for simulation of urban environments (e.g. noise, pollution, energy consumption).
[ANR project SEMAPOLIS, 2013-2016]

Towards semantic 3D reconstruction

Goal: detailed semantic 3D reconstruction for simulation of urban environments (e.g. noise, pollution, energy consumption).
[ANR project SEMAPOLIS, 2013-2016]

III. Dynamic urban visual record

Cameras around us



Public space cameras

Car cameras

Citizen cameras

Towards large-scale temporal analysis

Extract statistics of human behaviors across a city over time

crossing street

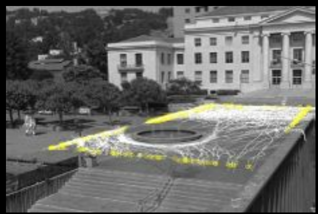
"bicycle accident"

riding bicycle

Applications: new ways to optimize road safety, urban planning or commerce in cities

Summary

- Multiple data sources and archives provide a comprehensive visual record of cities.

- Goal: develop visual quantitative analysis tools to:
    - understand, simulate and optimize urban environments