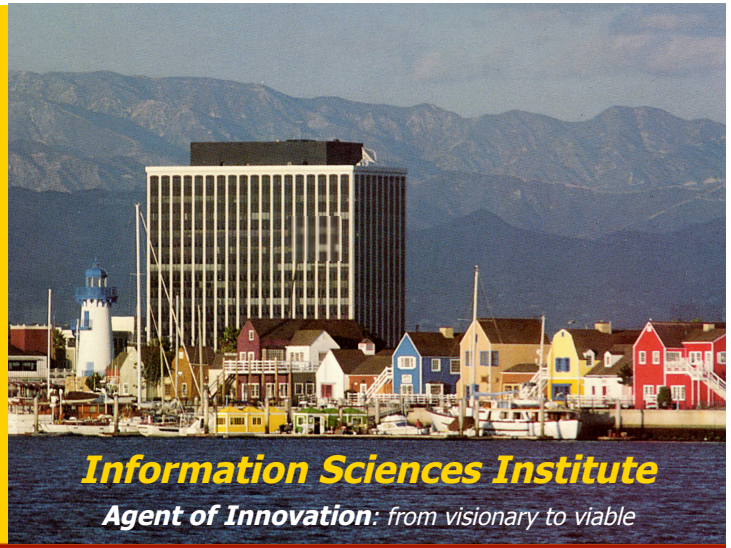
USC Viterbi
School of Engineering



A General Approach to Discovering, Registering, and Extracting Features from Raster Maps

Craig Knoblock
University of Southern California
&
Geosemble Technologies

Joint work with Ching-Chien Chen, Yao-Yi Chiang, Aman Goel, Matthew Michelson, and Cyrus Shahabi

USC

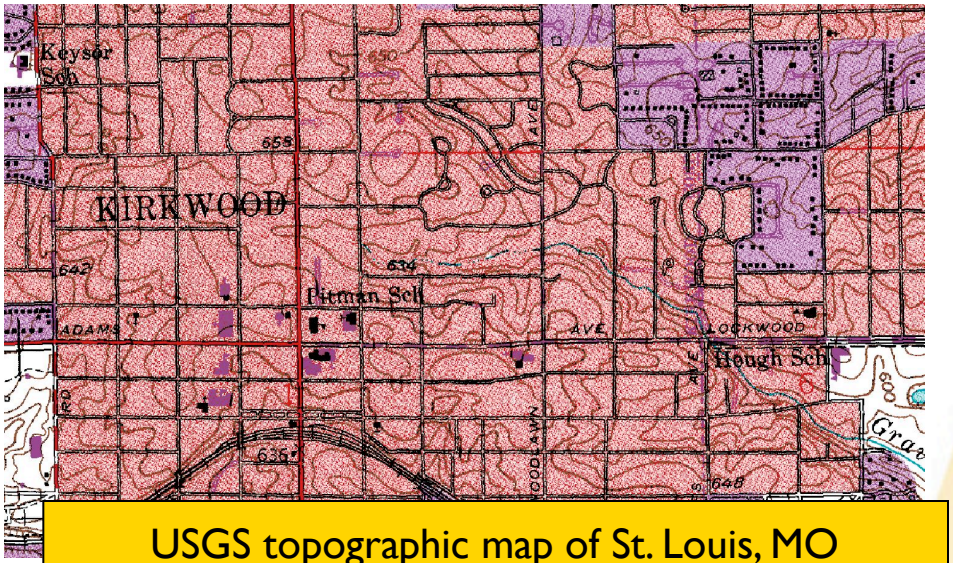
USC Viterbi School of Engineering

Introduction

ISI Information Sciences Institute

- **Raster maps are a rich source of geospatial data:**
  - Easily accessible
  - Many different types of information
  - Often contains information that cannot be found elsewhere

Travel map of Tehran, Iran



USGS topographic map of St. Louis, MO

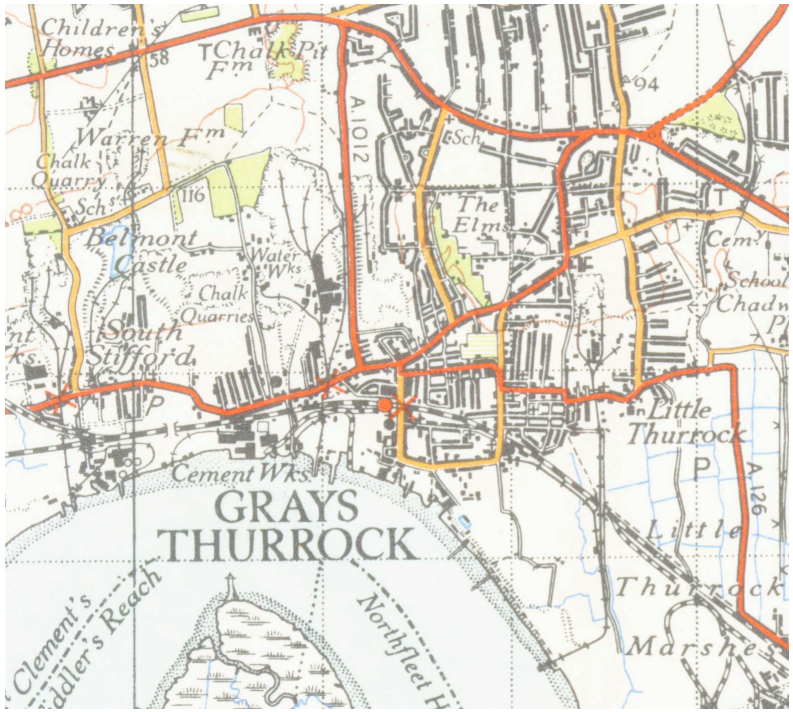
USC

USC Viterbi School of Engineering

Challenges

ISI Information Sciences Institute

- **Maps have lots of useful information, but…**
  - They have overlapping features
  - There is limited access to the meta-data
  - Often only available in raster format
- **How do we find, register, and extract and recognize the features in a raster map**



USC

USC Viterbi School of Engineering

Outline

ISI Information Sciences Institute

- Map Discovery
- Automatic Extraction of Features
- Feature Extraction from Noisy Maps
- Automatic Registration of Maps
- Next Steps
- Related Work and Discussion

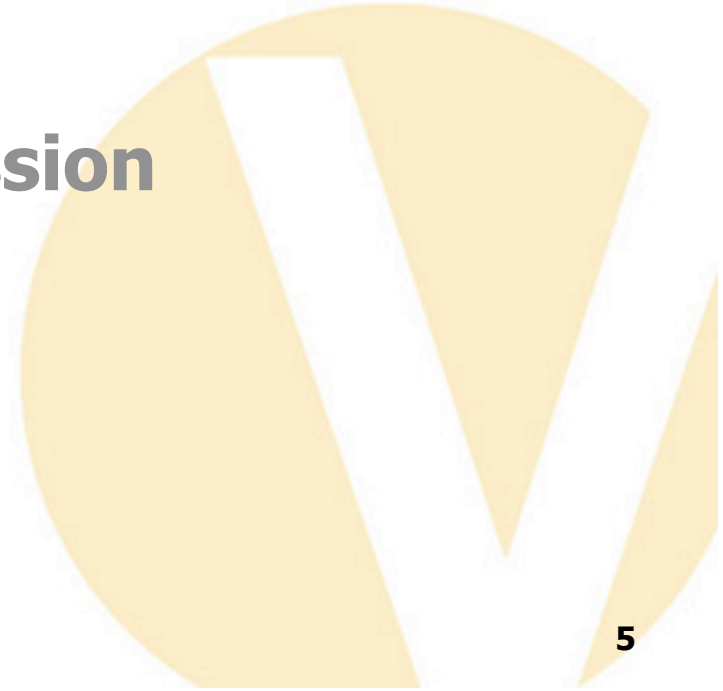
USC

USC Viterbi School of Engineering

Outline

ISI Information Sciences Institute

- **Map Discovery**
- Automatic Extraction of Features
- Feature Extraction from Noisy Maps
- Automatic Registration of Maps
- Next Steps
- Related Work and Discussion



USC

USC Viterbi School of Engineering

Map Discovery

ISI Information Sciences Institute

- **Collect candidate maps from the Web**
  - —Standalone maps
    - *Found using an image search engine*
  - —Maps embedded in PDF documents
    - *Found using a general search engine and then extracting the images*
- **Classify the images**
  - —Extract features from the images
  - —Identify similar images using Content Based Image Retrieval (CBIR)
  - —Classify the image using k-Nearest Neighbor

USC

USC Viterbi School of Engineering

Identifying Maps

ISI Information Sciences Institute



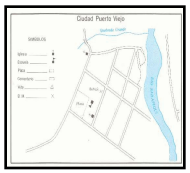

Image Server

Map Image Repository

Non-Map Image Repository

Our approach :

- Extract features
- Find similar images
- Classify image

Map Server

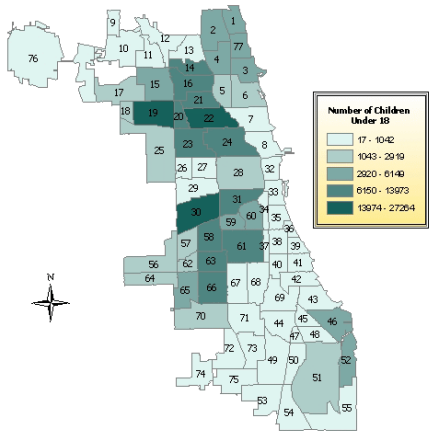

USC

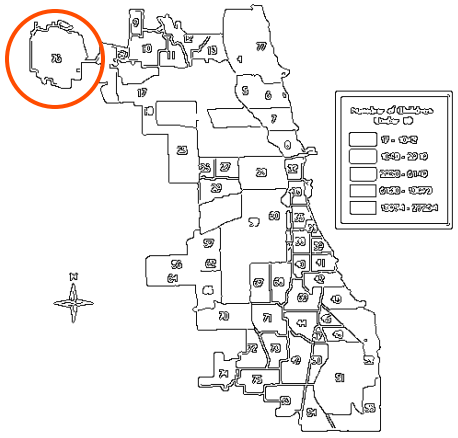
USC Viterbi School of Engineering

Extract Features

ISI Information Sciences Institute

- **Water-filling features**
  - Zhou, X.S. et al. - Water-filling: A novel way for image structure feature extraction, 1999, Intl. conference on Image Processing
  - Works well on images with strong edges





    - Works on standard Canny edge maps of original images
      - Color invariant

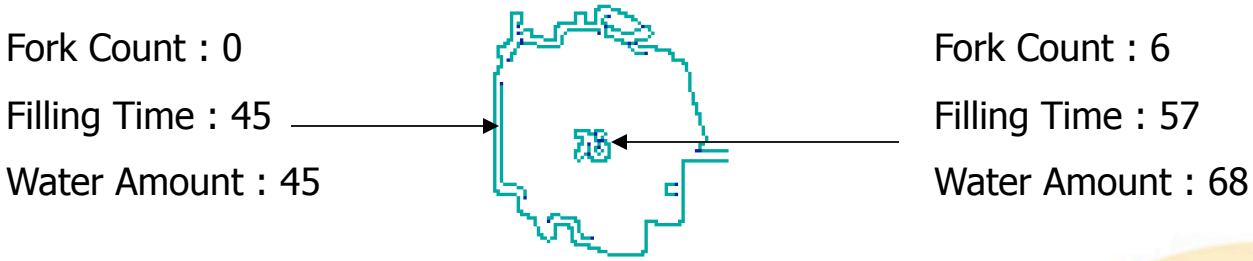
USC

USC Viterbi School of Engineering

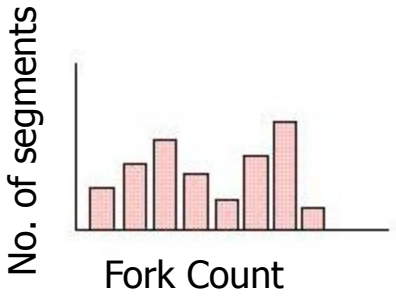
Water-Filling Features

Information Sciences Institute

- Features computed for each segment



- Normalized histogram - size invariant



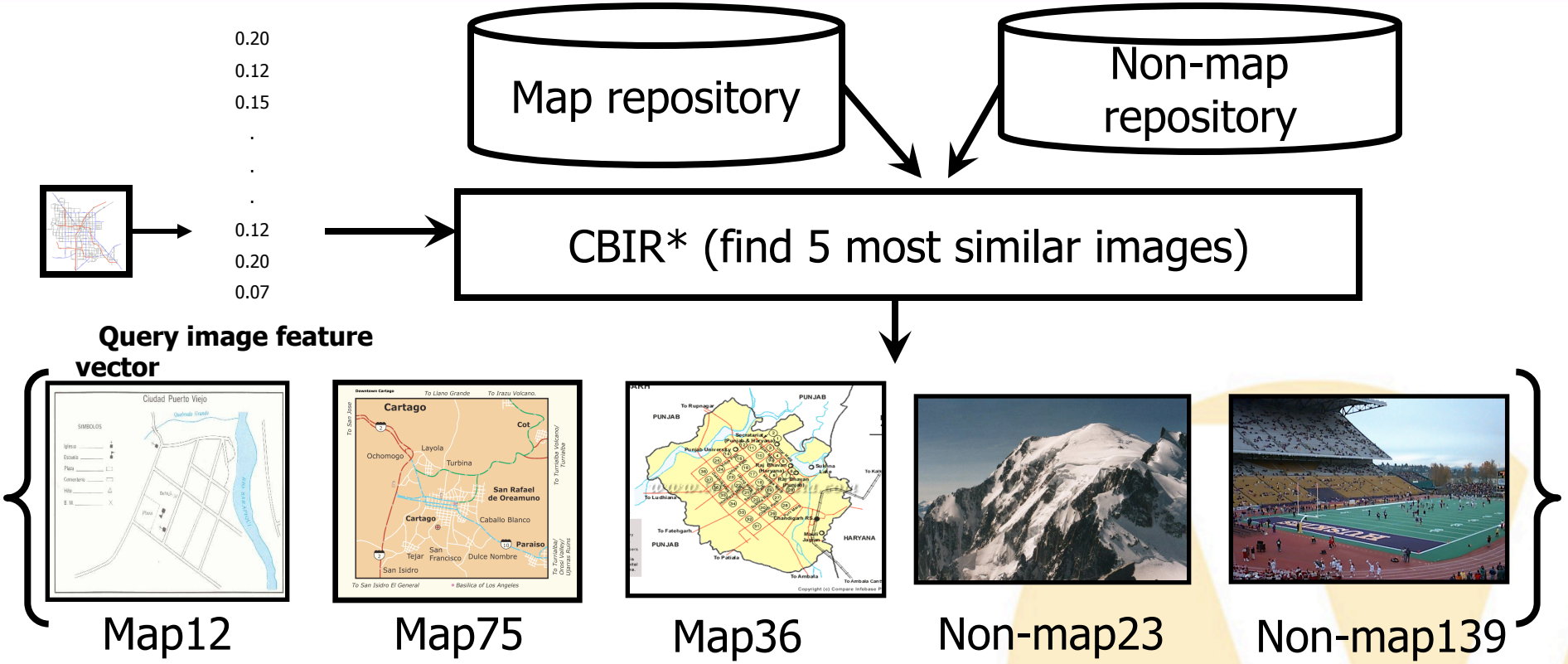
- 3 features x 8 buckets = 24 element feature vector

USC

USC Viterbi School of Engineering

Content-Based Image Retrieval (CBIR)

ISI Information Sciences Institute



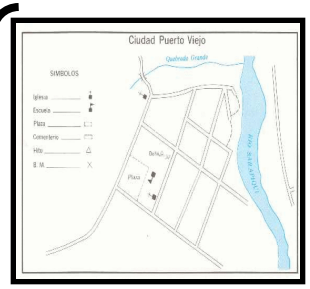
- Built on top of Lire system

* In our experiment we used 9 similar images

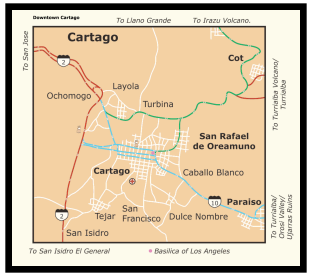
USC Viterbi School of Engineering

k - Nearest neighbor classification

Information Sciences Institute



Map12

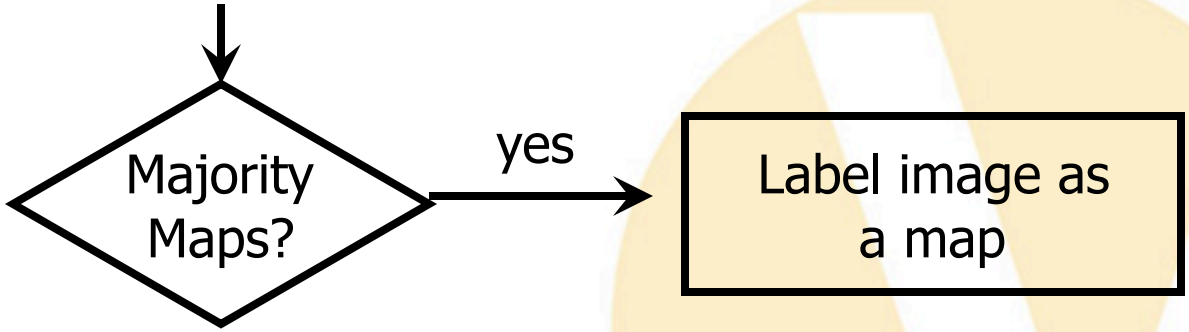


Map75

Map36

Non-map23

Non-map139



USC

USC Viterbi School of Engineering

Results

ISI Information Sciences Institute

Precision	Recall	F1-measure
77.39%	71.20%	74.17%

USC

Results are average over 10 runs

USC Viterbi School of Engineering

Outline

ISI Information Sciences Institute

- Map Discovery
- **Automatic Extraction of Features**
- Feature Extraction from Noisy Maps
- Automatic Registration of Maps
- Next Steps
- Related Work and Discussion

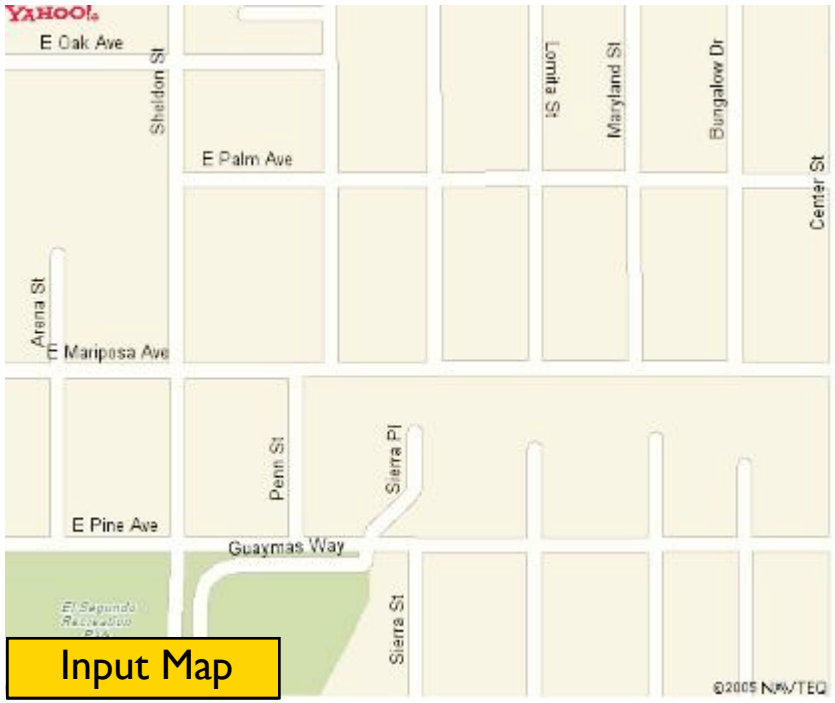
USC

USC Viterbi School of Engineering

Background Removal

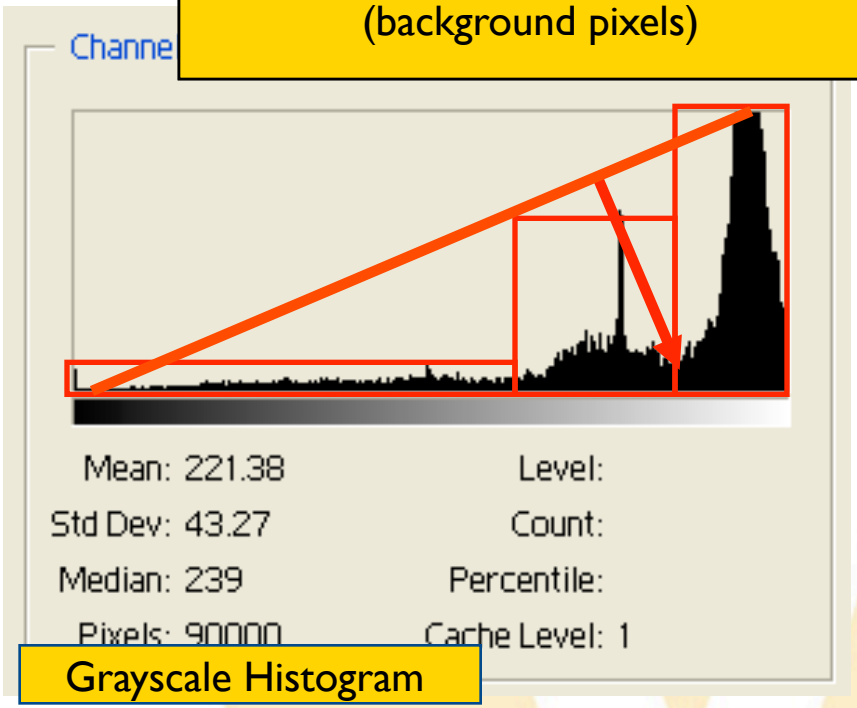
ISI Information Sciences Institute

- **Use the Triangle method (Zack, 1977) to locate clusters in the grayscale histogram**
- **Remove the background clusters**



Input Map

Remove the dominate cluster (background pixels)



Grayscale Histogram

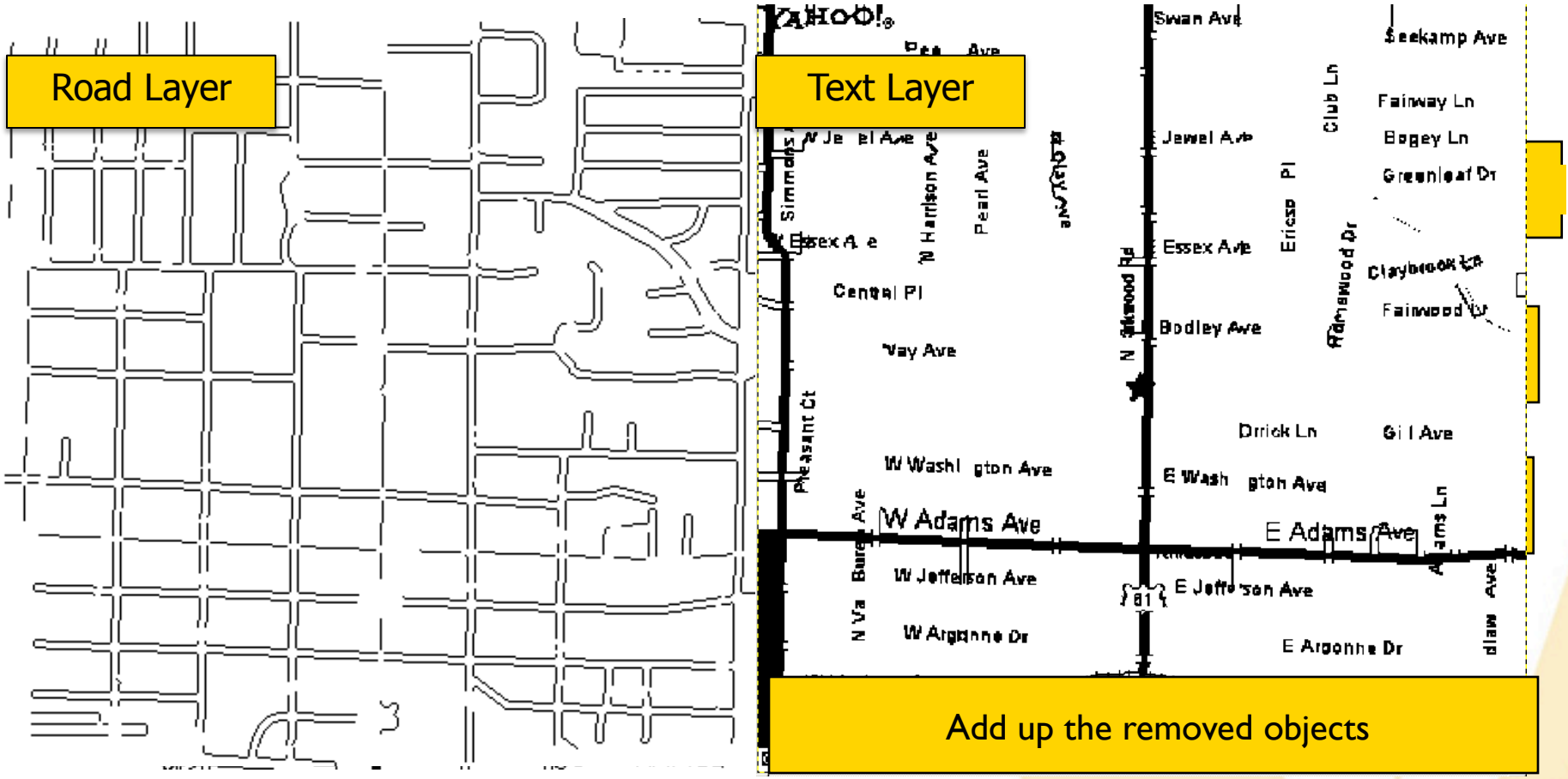
USC

USC Viterbi School of Engineering

Text/Graphics Separation

ISI Information Sciences Institute

- **Separate linear structures from text (Cao and Tan, 02)**



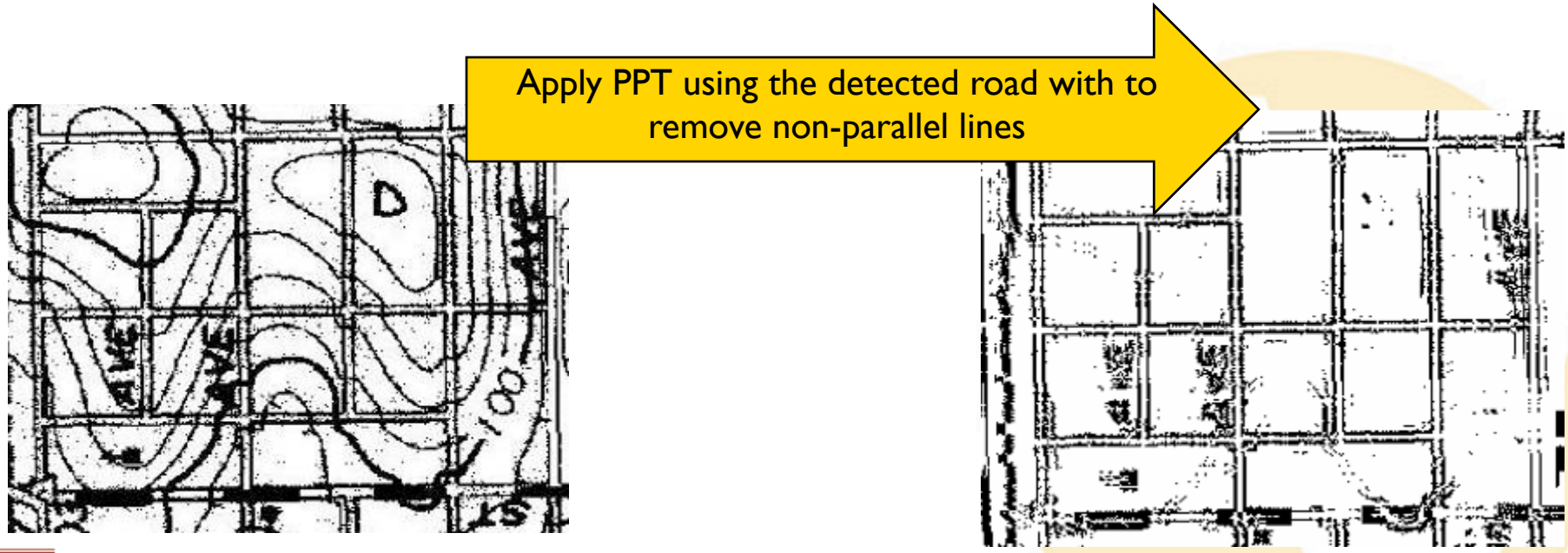
USC

USC Viterbi School of Engineering

Road Format and Road Width Detection

ISI Information Sciences Institute

- **Apply parallel-pattern tracing (PPT) iteratively on different sizes of road width**
- **If it is a double-line road layer, the actual road width**
  - Has the maximum percentage of parallel pattern pixels
  - The percentage is larger than a threshold

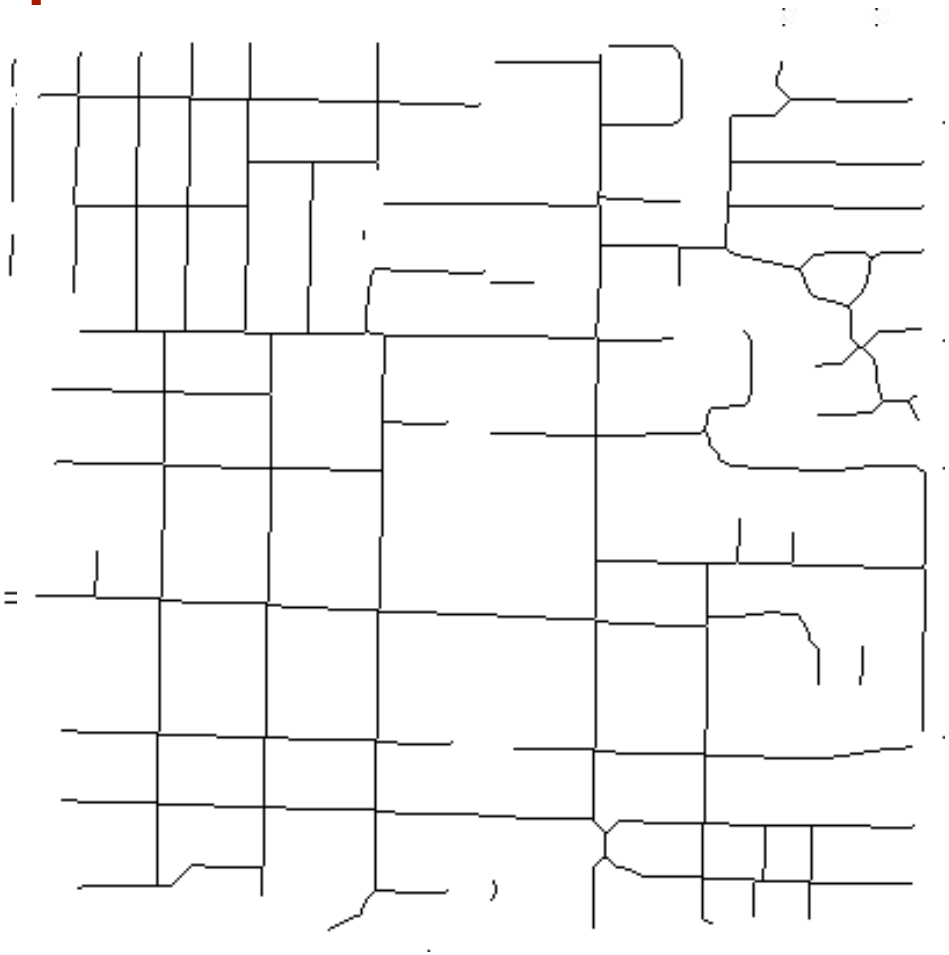


USC

USC Viterbi School of Engineering

Road Topology Extraction

- **Use morphological operations to reconnect broken lines and generate one-pixel width roads**



Morphological Operations: Use the detected road format and road width to determine the number of iterations

Dilation

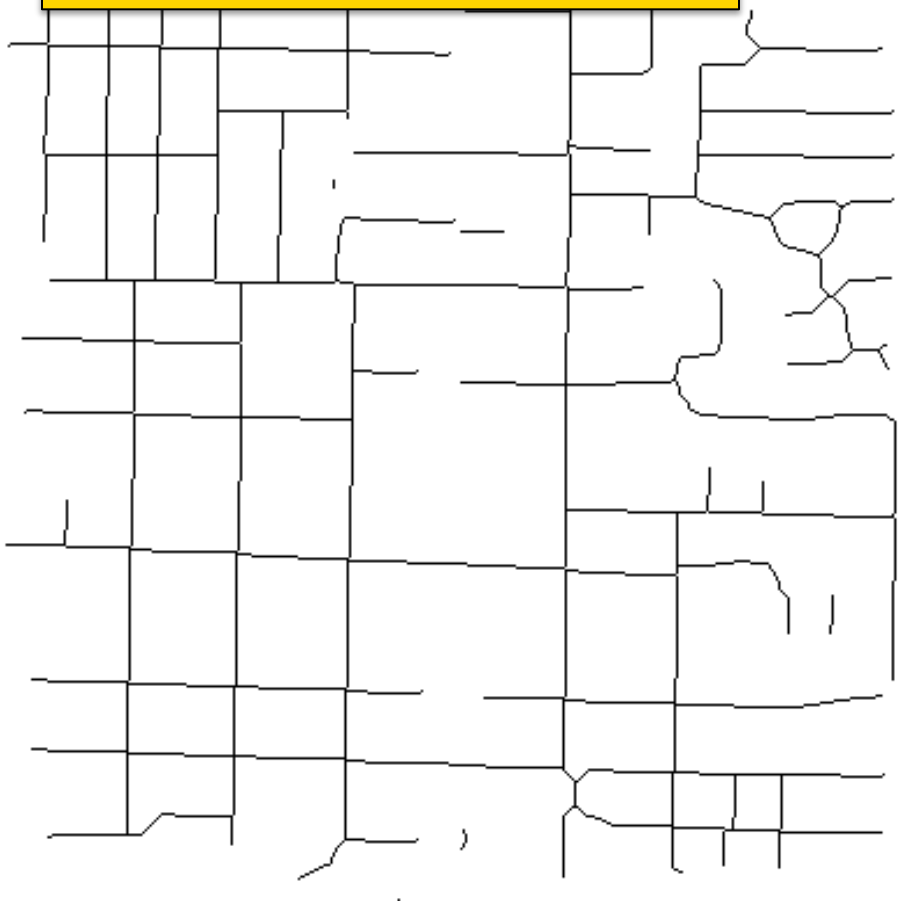
Erosion

Thinning

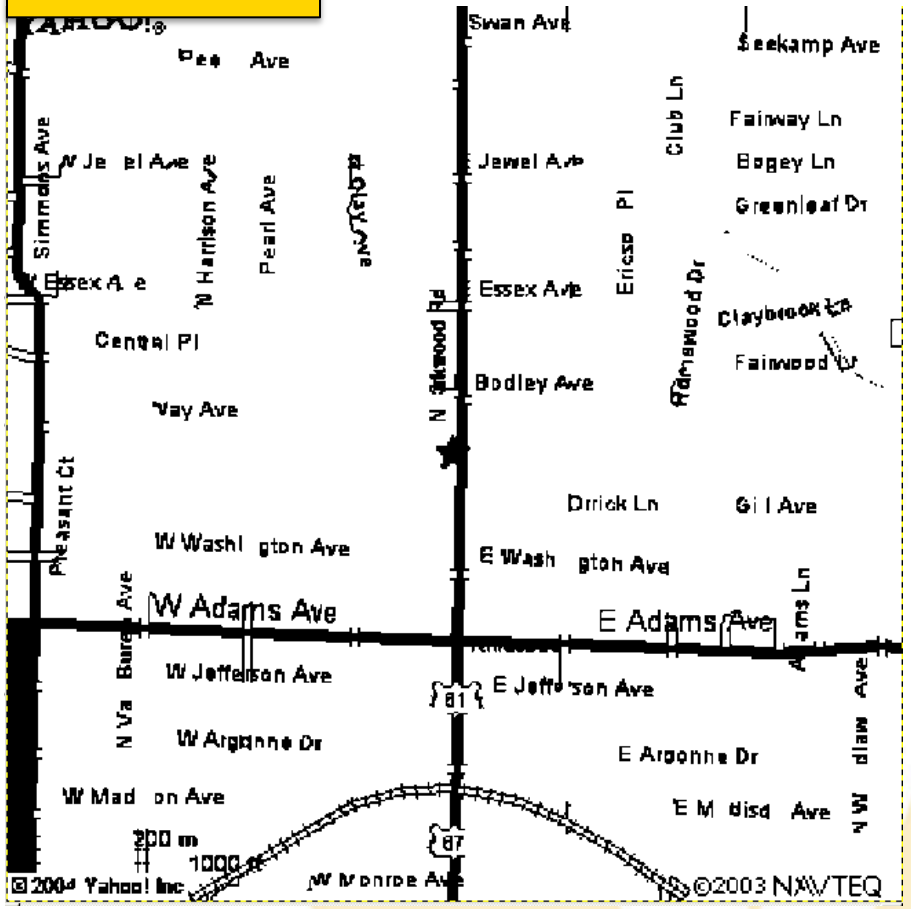
USC

Extracted Road and Text Layers

Road Layer (road topology)



Text Layer



USC Viterbi School of Engineering

Outline

ISI Information Sciences Institute

- Map Discovery
- Automatic Extraction of Features
- **Feature Extraction from Noisy Maps**
- Automatic Registration of Maps
- Next Steps
- Related Work and Discussion

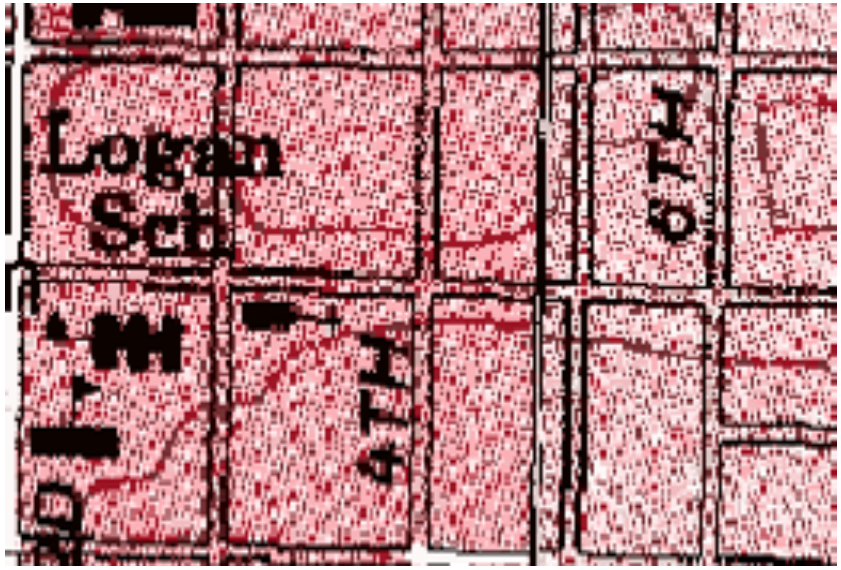
USC

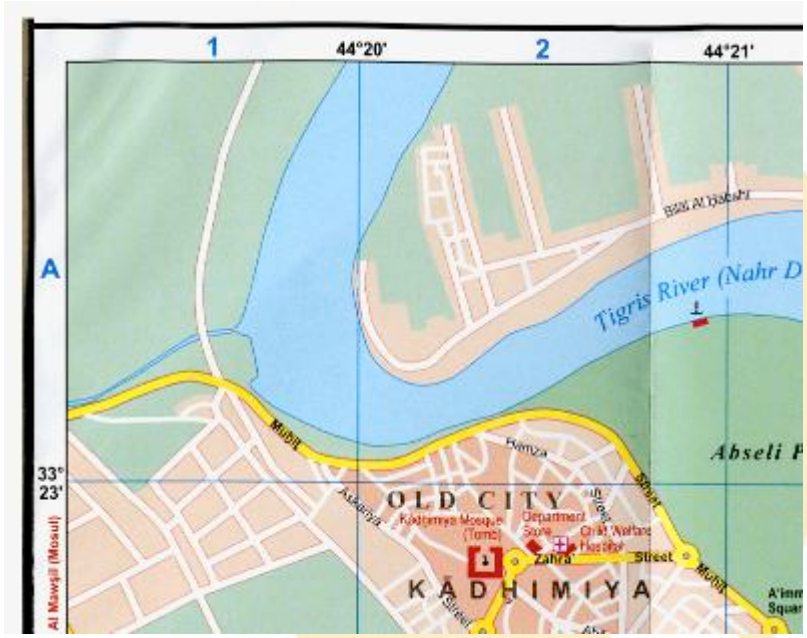
USC Viterbi School of Engineering

Supervised Map Decomposition

ISI Information Sciences Institute

- **What if we cannot automatically remove the background from raster maps?**
  - Raster maps usually contain noise from scanning and compression process





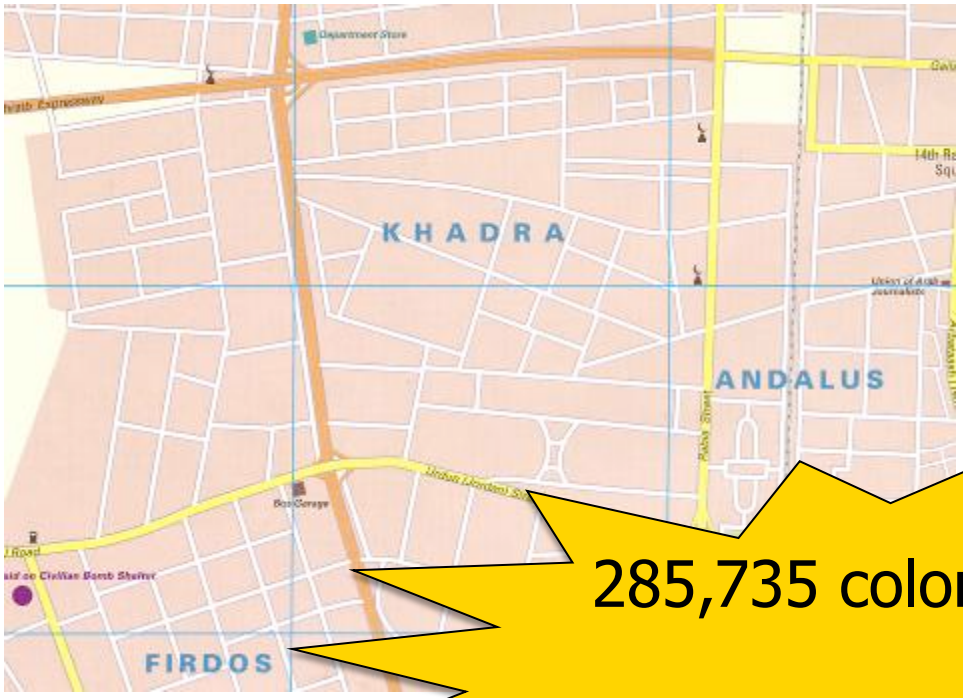
USC

USC Viterbi School of Engineering

Difficulties

ISI Information Sciences Institute

- **Raster maps contain numerous colors**
  - Manually examining each color for extracting features is laborious



285,735 colors

Grayscale histogram

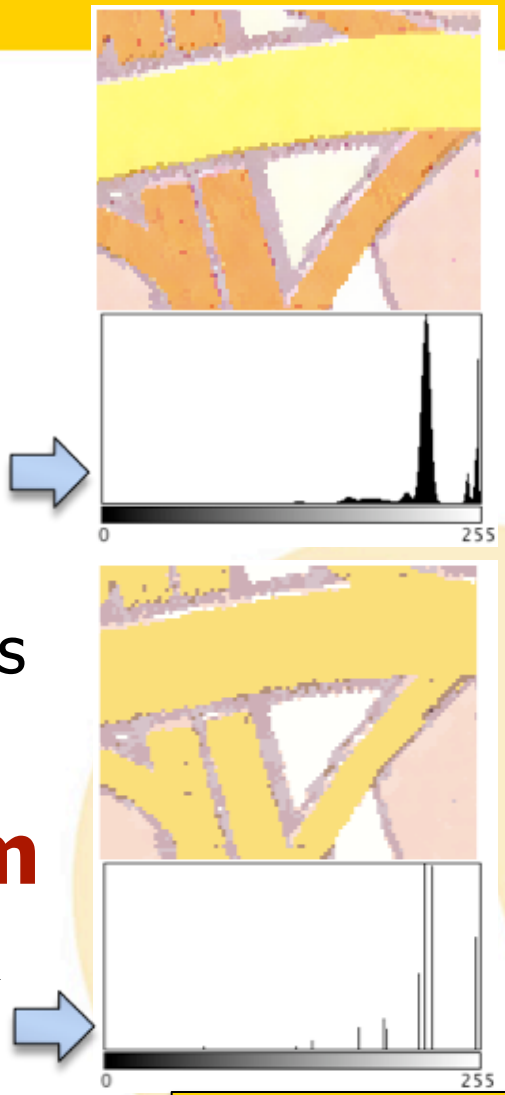
USC

USC Viterbi School of Engineering

Color Segmentations

Information Sciences Institute

- **The Mean-shift algorithm**
  - —Consider distance in the RGB color space and in the image space
  - —Preserve object edges
  - —From 285,735 to 155,299 colors
- **The K-means algorithm**
  - —Limit the number of colors to K
  - —From 155,299 to 10 colors (K=10)



Grayscale histogram

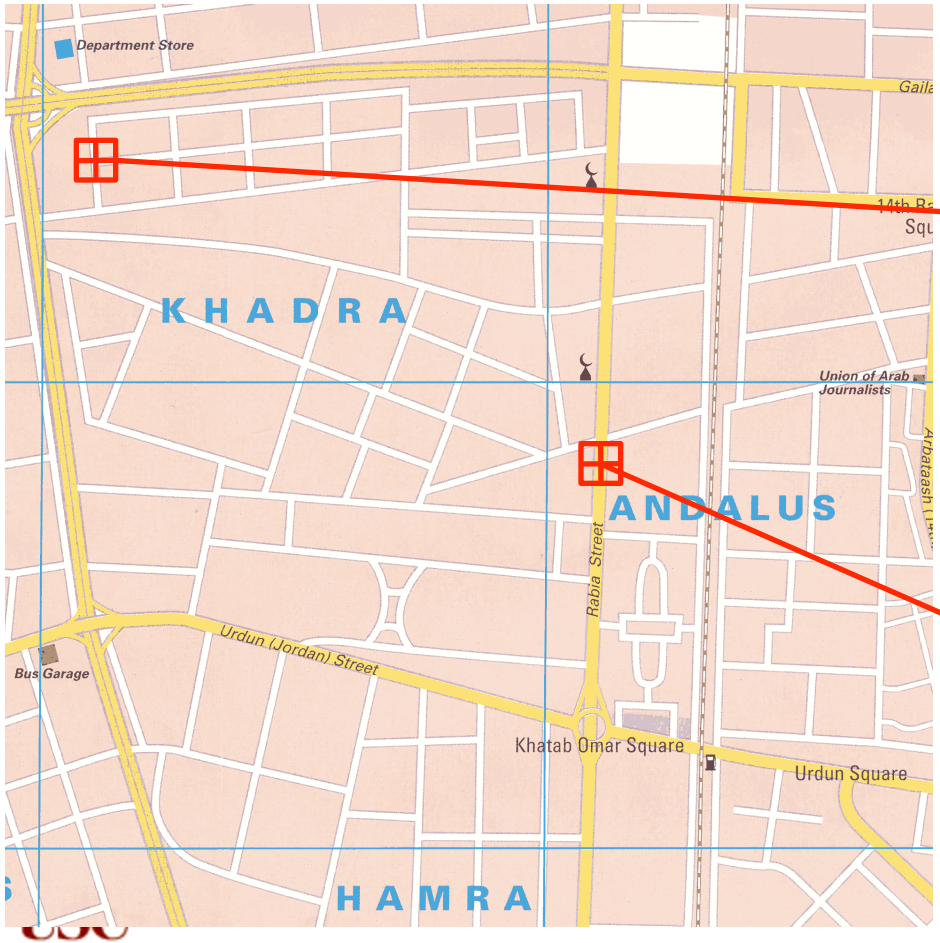
USC

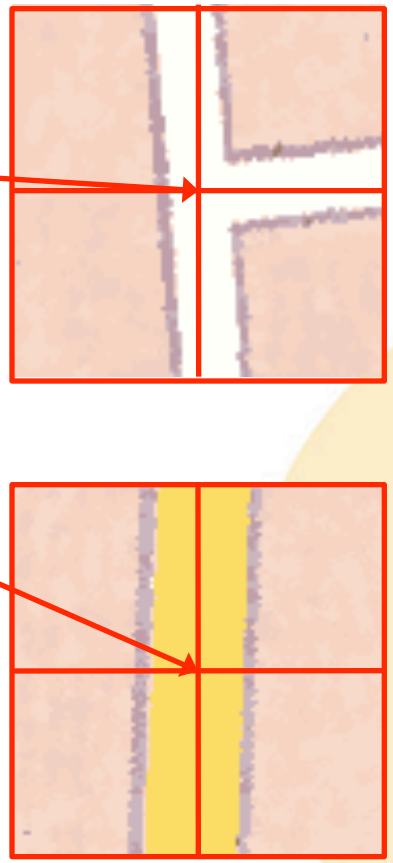
USC Viterbi School of Engineering

User Labeling

ISI Information Sciences Institute

- **To extract the road layer, the user needs to provide a user label for each road color (at most K colors)**





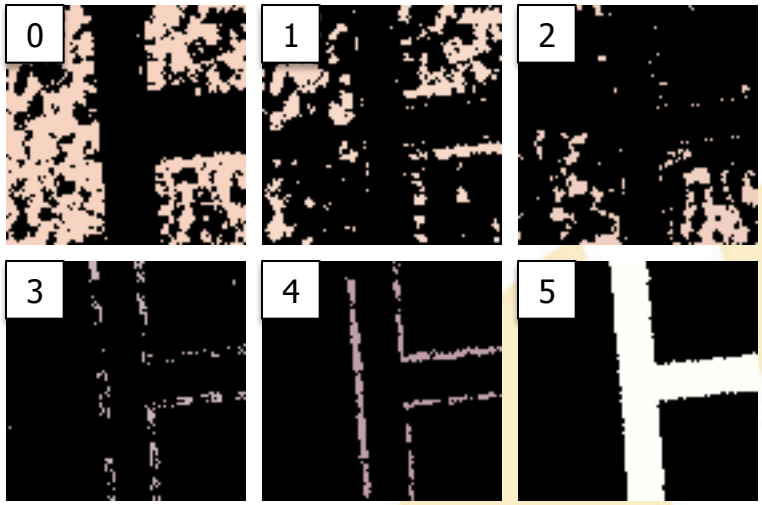
User label should be (approximately) centered at a road intersection or at the center of a road line

USC Viterbi School of Engineering

Label Decomposition

ISI Information Sciences Institute

- **Decompose each user label into color images so that every color image contains only one color**



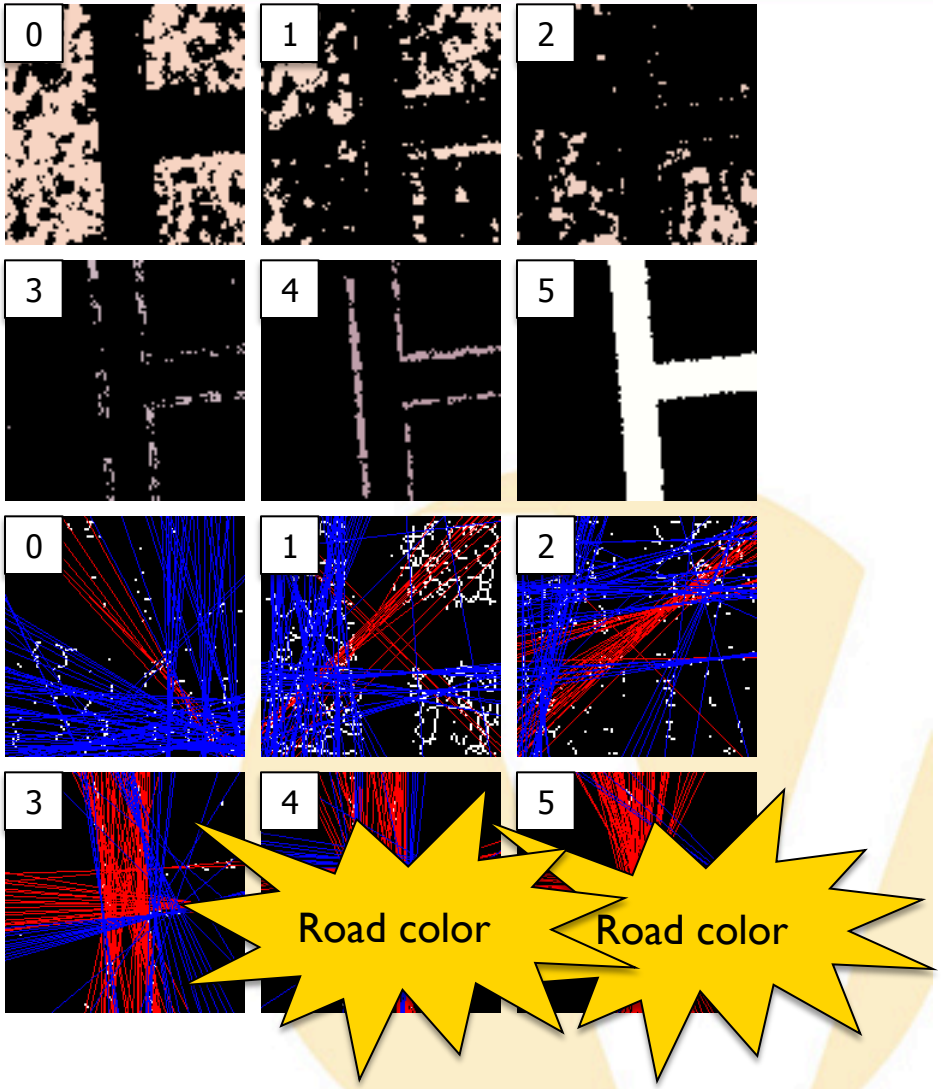
(background is shown in black)

USC

Hough-Line Approach to Identify Road Color

- **Detect Hough lines**
- **The center of the user label is the center of a road line**
  - The Hough lines that are far away from the image center are NOT constructed by road pixels
- **Identify road colors using**
  - The average distance between the Hough lines to the image center

Red Hough lines are within 5 pixels to the image center

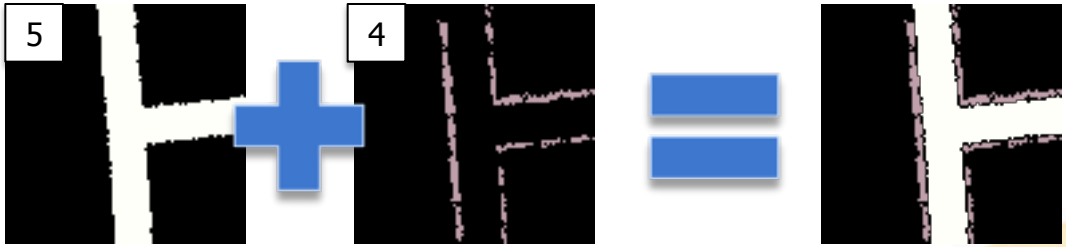


USC Viterbi School of Engineering

Initial Road Template

ISI Information Sciences Institute

- **Generate an initial road template using the images of identified road colors from the Hough line approach**



(background is shown in black)

(road pixels are shown in red, background is shown in black)

USC

USC Viterbi School of Engineering

Road Topology Extraction using Identified Road Colors

ISI Information Sciences Institute

- Identify a set of road colors from each user label
- Use the identified road colors to extract road pixels
- Apply morphological operations to remove solid areas and reconnect lines

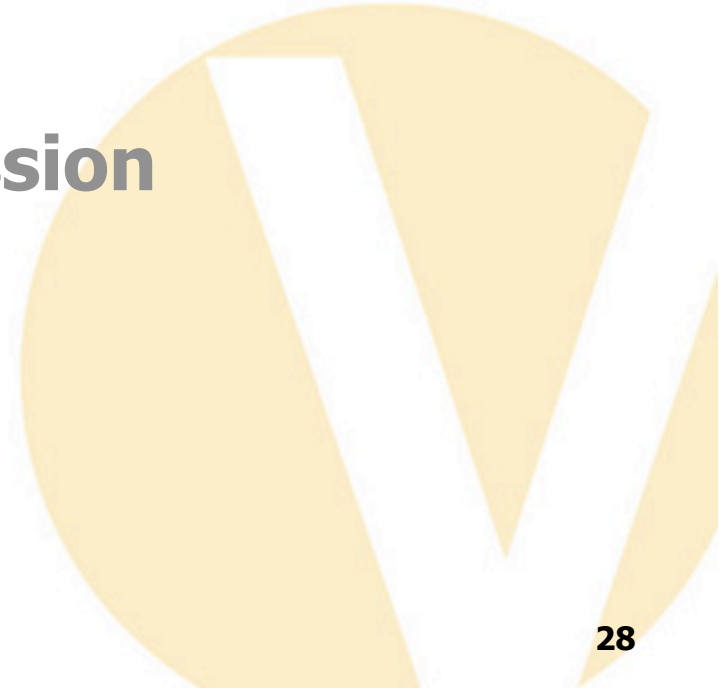
USC

USC Viterbi School of Engineering

Outline

ISI Information Sciences Institute

- Map Discovery
- Automatic Extraction of Features
- Feature Extraction from Noisy Maps
- **Automatic Registration of Maps**
- Next Steps
- Related Work and Discussion



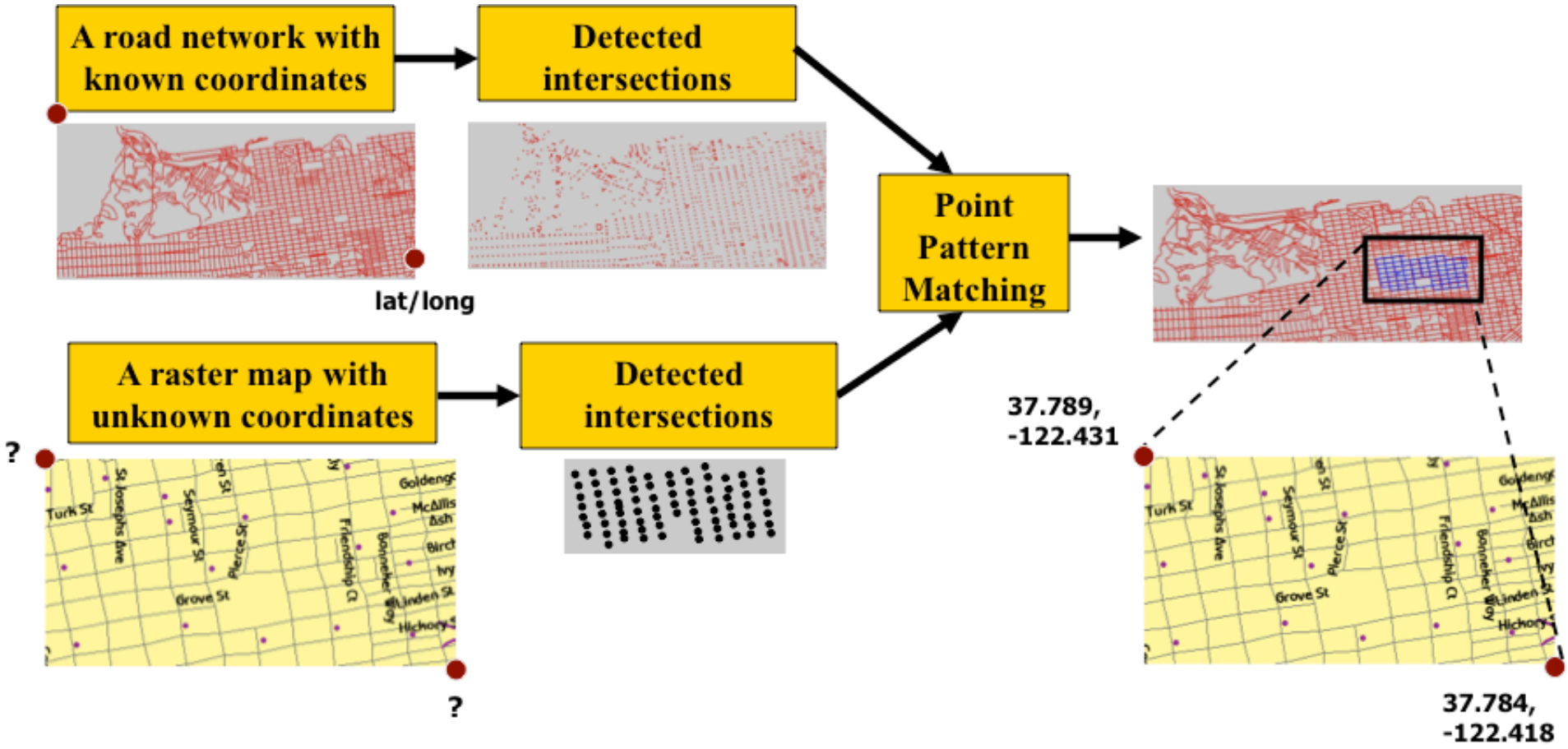
USC

USC Viterbi School of Engineering

Automatic Map Registration

ISI Information Sciences Institute

- **Exploit the pattern of intersections found on a map and compare to a road vector dataset**



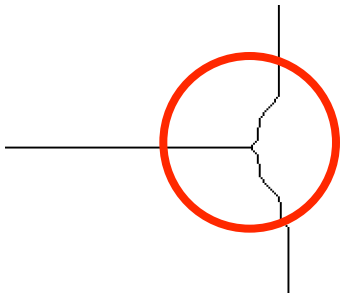
USC Viterbi School of Engineering

Road-Intersection Template Extraction

ISI Information Sciences Institute

- **Road-intersection template**
  - road intersection position
  - road connectivity
  - road orientation

- **Road lines are distorted by the thinning operator**
- **The extracted road-intersection templates will not be accurate**



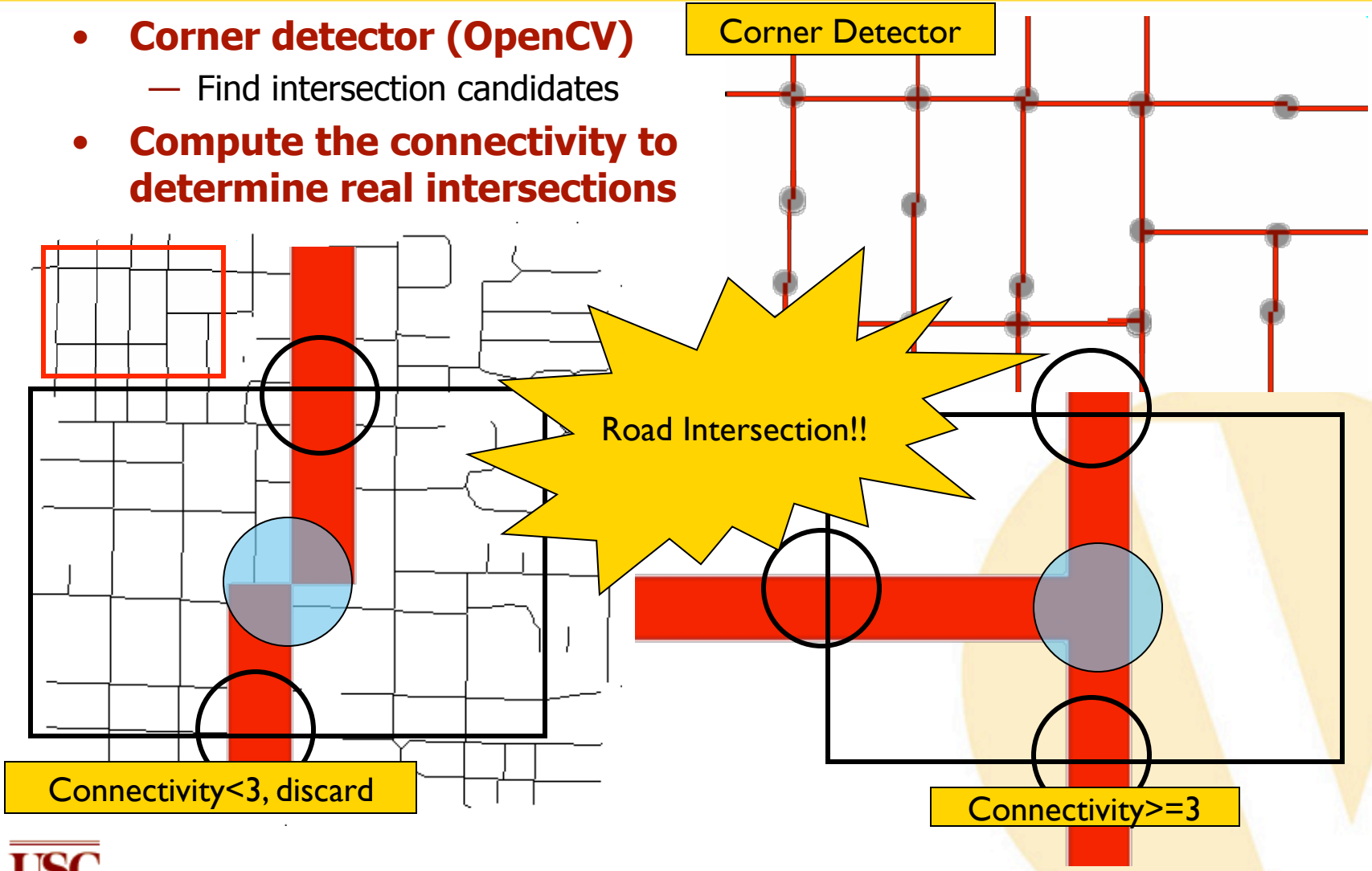
USC

USC Viterbi School of Engineering

Road-Intersection Position Detection

ISI Information Sciences Institute

- **Corner detector (OpenCV)**
  - Find intersection candidates
- **Compute the connectivity to determine real intersections**

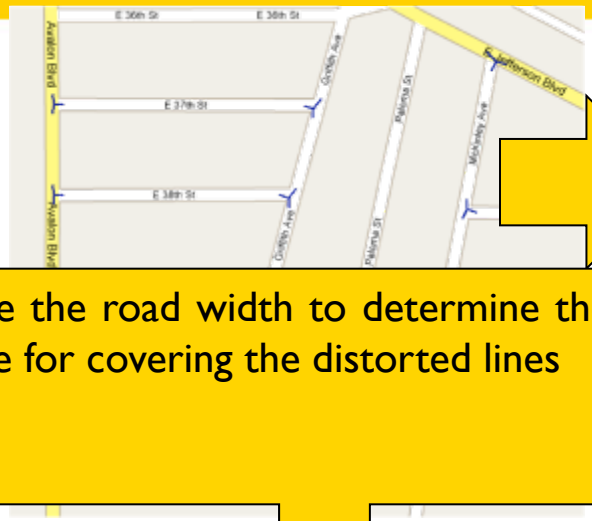


USC

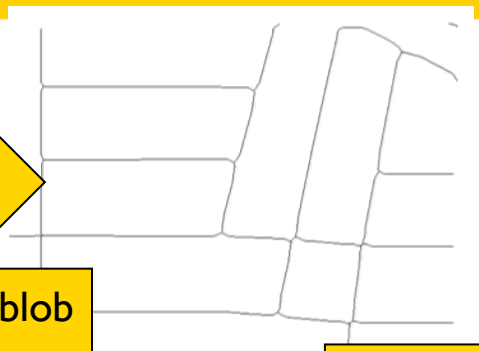
USC Viterbi School of Engineering

Distortion Correction

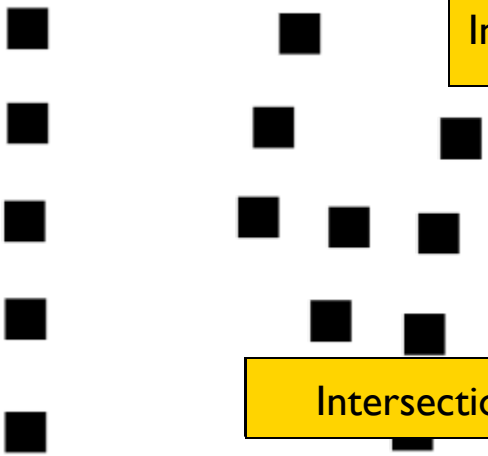
ISI Information Sciences Institute



Use the road width to determine the blob size for covering the distorted lines

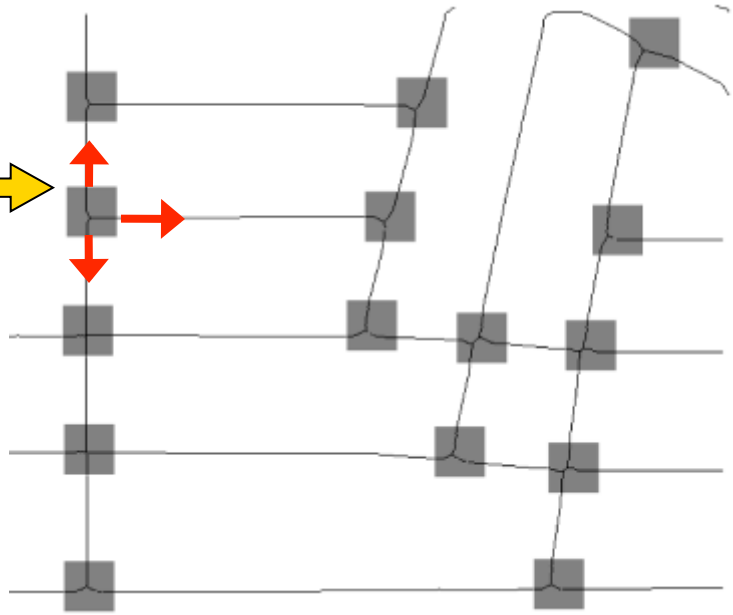


The thinned lines



Intersection Positions

Intersect the images



USC Viterbi School of Engineering

Accurate Road-Intersection Templates

ISI Information Sciences Institute

With distortion

Avoid distortion

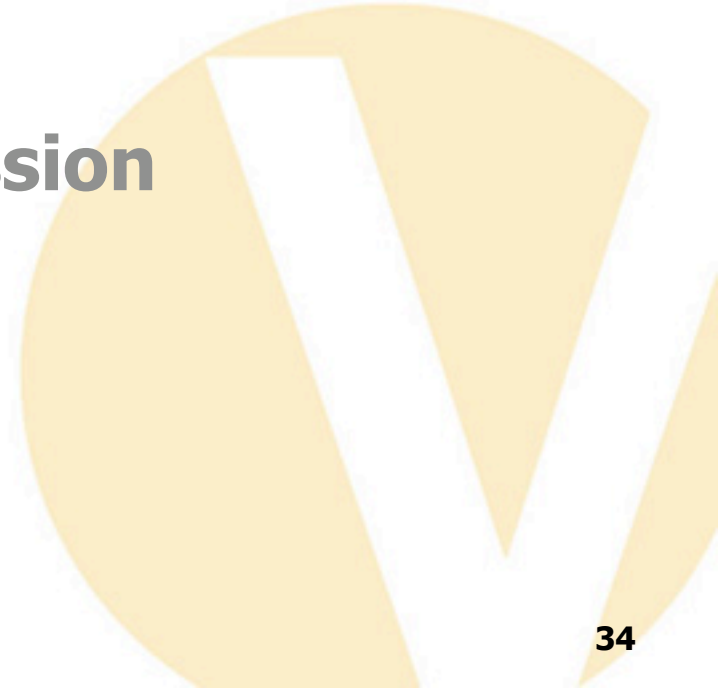
USC

USC Viterbi School of Engineering

Outline

ISI Information Sciences Institute

- Map Discovery
- Automatic Extraction of Features
- Feature Extraction from Noisy Maps
- Automatic Registration of Maps
- **Next Steps**
- Related Work and Discussion



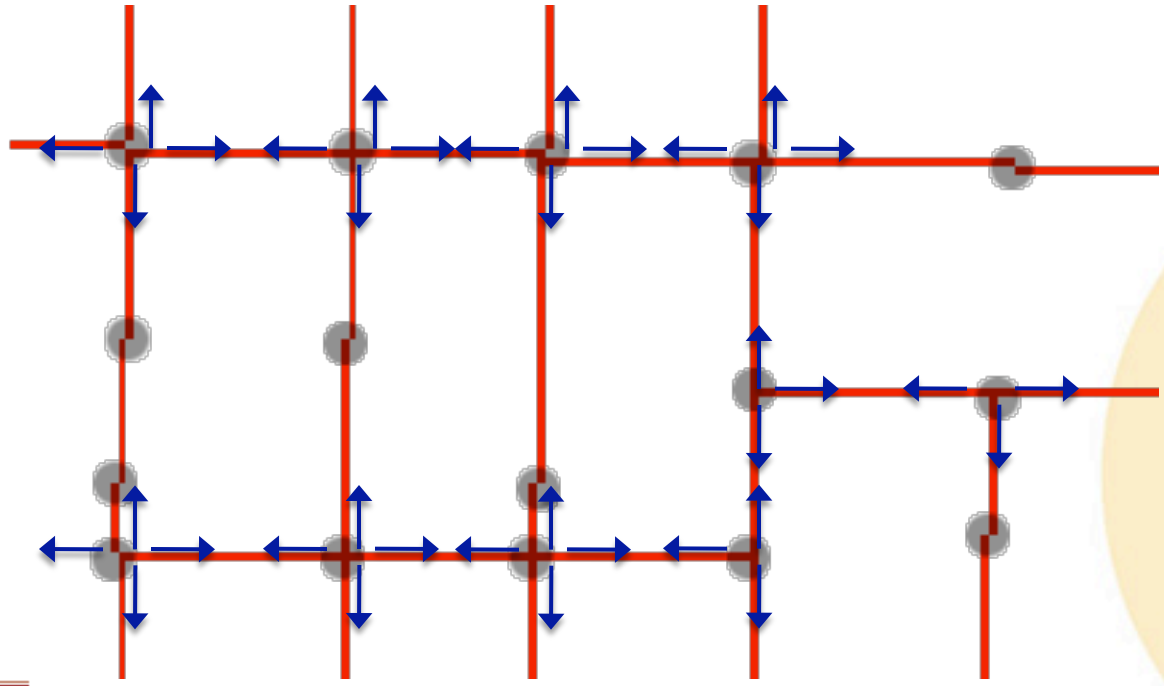
USC

USC Viterbi School of Engineering

Next Steps: Road Vectorization

ISI Information Sciences Institute

- **Start from the extracted road intersections to connect the salient points and produce the road vector**

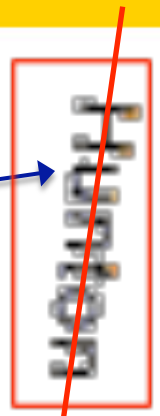


USC

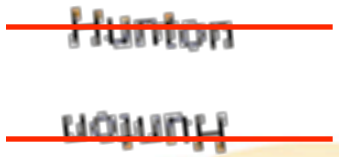
USC Viterbi School of Engineering

Next Steps: Text Recognition

ISI Information Sciences Institute



Rotate each string image according to its central axle



Optical character recognition

- **Generalize OCR techniques to apply to maps**
  - Identify individual characters regardless of orientation
  - Exploit background knowledge to improve accuracy

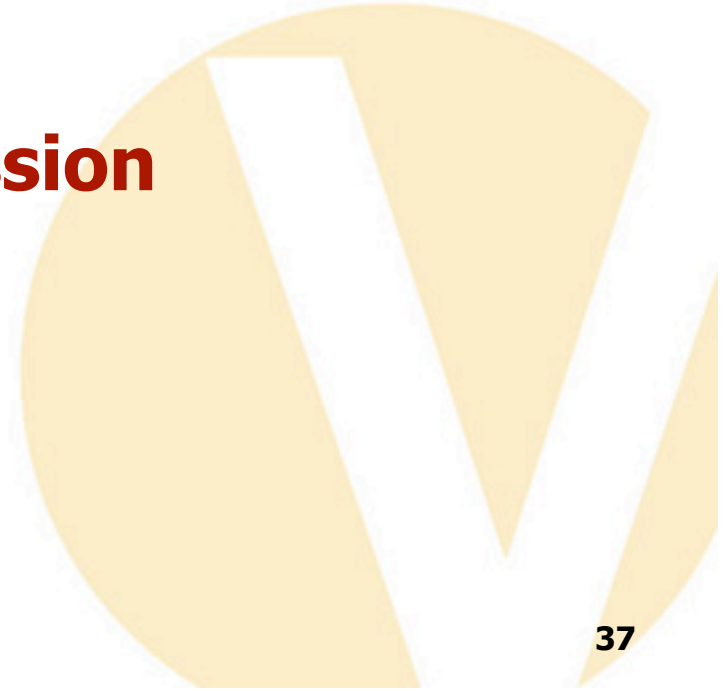
USC

USC Viterbi School of Engineering

Outline

ISI Information Sciences Institute

- Map Discovery
- Automatic Extraction of Features
- Feature Extraction from Noisy Maps
- Automatic Registration of Maps
- Next Steps
- **Related Work and Discussion**



USC

USC Viterbi School of Engineering

Related Work

ISI Information Sciences Institute

- **Map Feature Extraction Using Map Specification (Samet and Soffer, 94, 96; Myers et al., 96)**
  - — Require huge amount of prior information and training
- **Text/Graphics Separation and Text Recognition (Bixler, 00; Li et al.,00; Cao and Tan 02; Vela, 03; Pouderoux, 07)**
  - — Require fixed pre-processing steps, e.g., binarization with fixed threshold
- **Supervised Graphics Extraction (Khotanzad and Zink, 03; Salvatore and Guitton, 04; Chen et al. 06)**
  - — Laborious training tasks, e.g., labeling all combination of line and background pixels
- **Road Extraction and Vectorization (Bin, 98; Habib et al., 99;Itonaga et al., 03)**
  - — Require lots of parameter tunings, e.g., road width
- **Map, Vectors, and Imagery Conflation (Chen et al., 06; Chen et al., 08;Wu et al., 07)**
  - — Exploit for determining feature locations

USC

USC Viterbi School of Engineering

Discussion

ISI Information Sciences Institute

- **Presented a general approach to discovering, registering, extracting features from maps**
- **Contributions**
  - —Ability to identify maps
  - —Ability to extract road and text layers
  - —Automatic recognition of road intersection
  - —Algorithms to automatically determine the geocoordinates
- **Applications**
  - —Annotating imagery
  - —Creating and updating maps
  - —Constructing gazetteers

USC