Automatically and Accurately Conflating Road Vector Data, Street Maps and Orthoimagery

Ching-Chien Chen

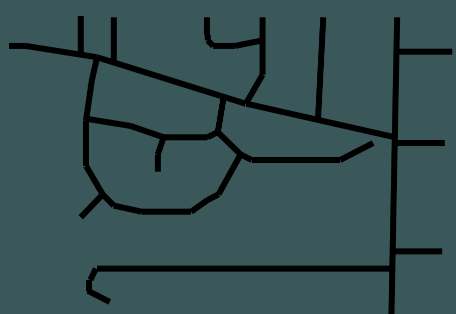
Ph.D. Dissertation
March 2005

Outline

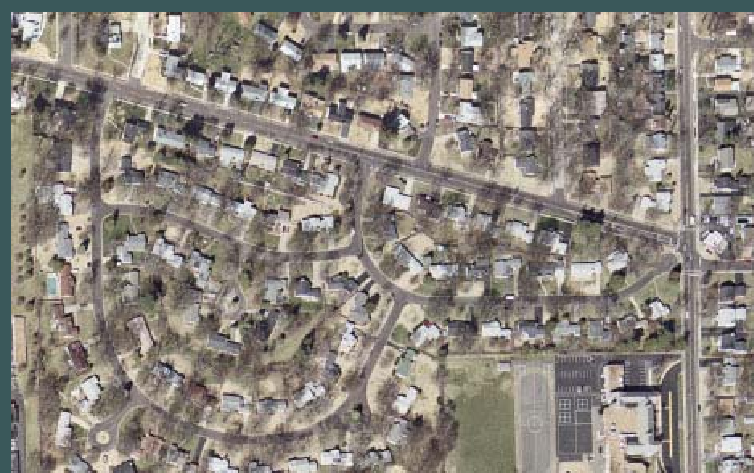
- Introduction & Motivation
- Our approach: AMS-conflation
  - Vector and imagery conflation (pre-qualifying research)
  - Map and imagery conflation
    - Finding control points in the imagery and in the maps
    - Geospatial point pattern matching (GeoPPM)
    - Image and map conflation using rubber-sheeting
- Experimental Results
- Related Work
- Conclusion and Future Work

Introduction

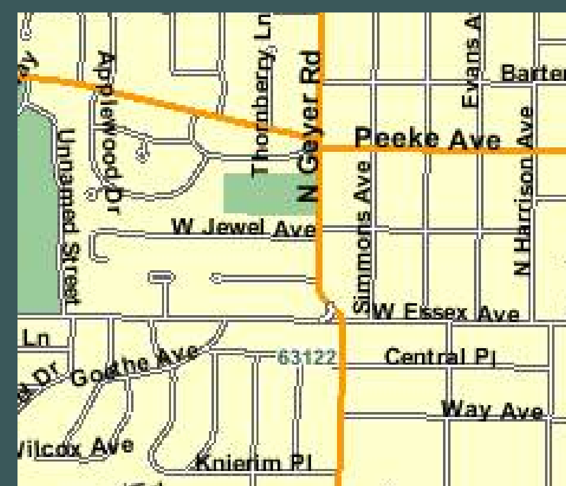
- Geospatial data sources have become widely available
- Automatically and accurately integrating and aligning two spatial datasets is a challenging problem



Road network
(in vector format)



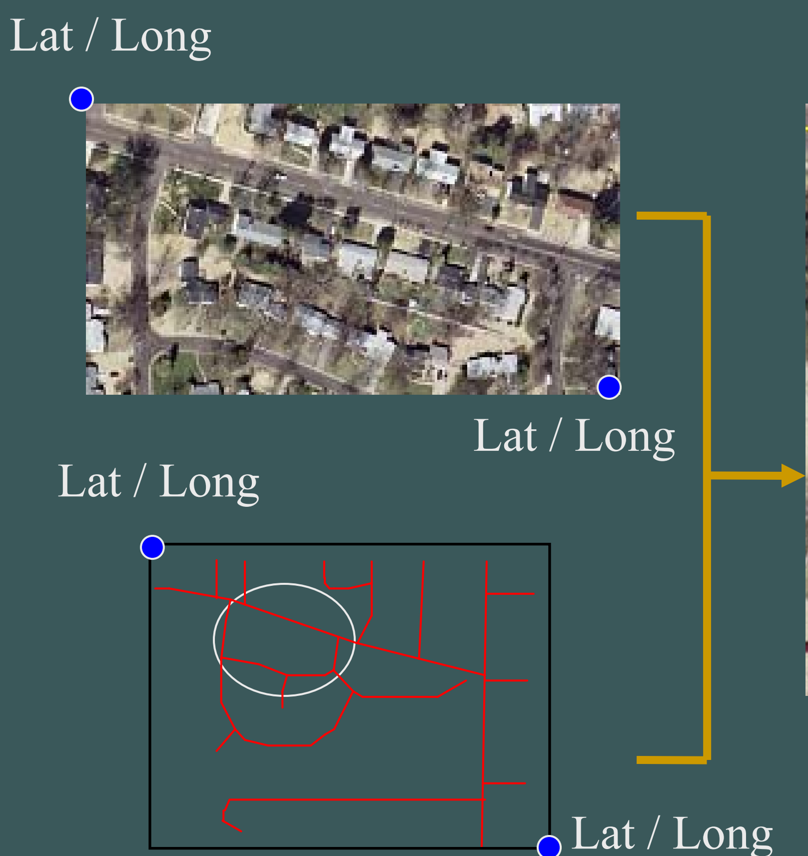
Orthoimagery
(in raster format)



Street maps
(in raster format)

Motivation :
Vector and Imagery Integration

- Challenges
  - Different projections, accuracy levels, resolutions result in spatial inconsistencies

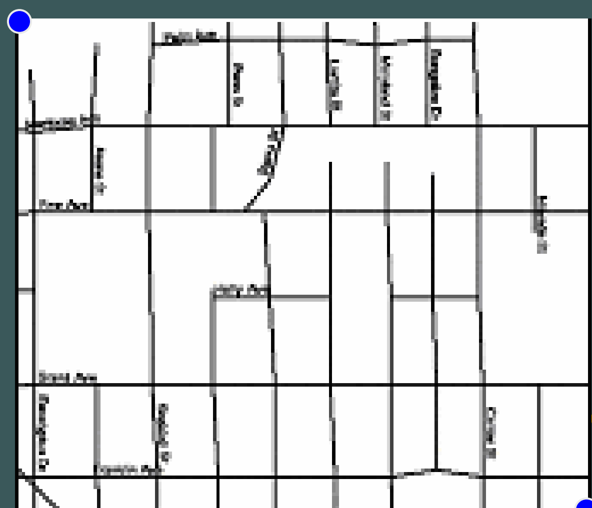


Motivation :
Map and Imagery Integration

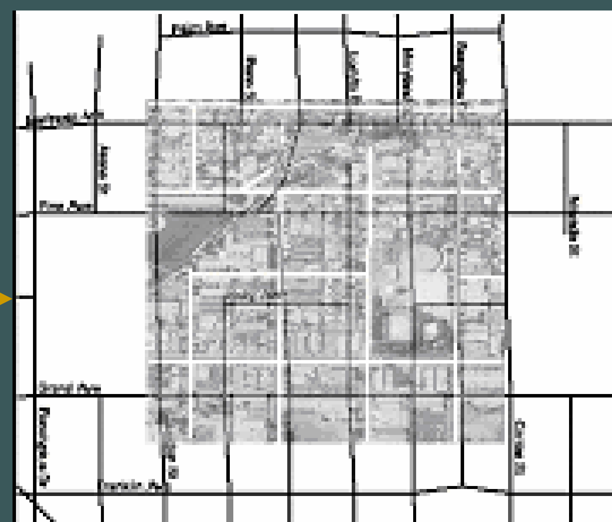
Lat / Long

Lat / Long

Lat / Long



Lat / Long



- Integration Challenge
  - Different geographic projections and accuracy levels

Motivation :
Map and Imagery Integration

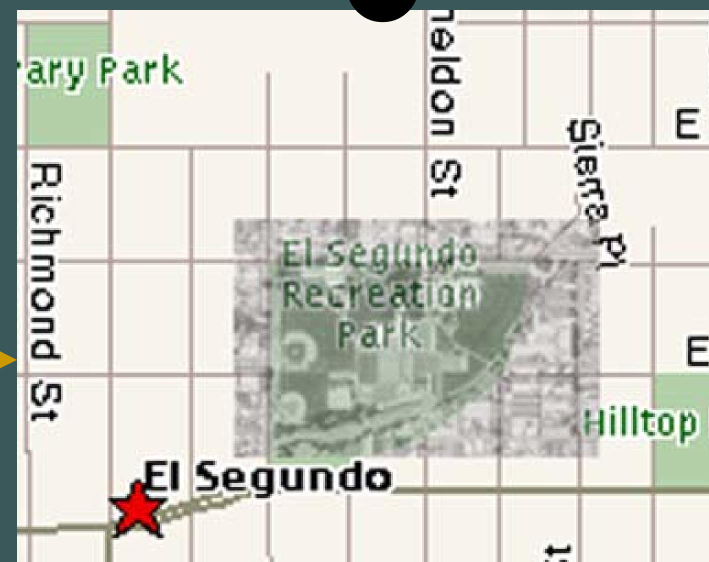
?

Lat / Long

Lat / Long

?

?



- Another Integration Challenge
  - Some online maps are not geo-referenced

Motivation

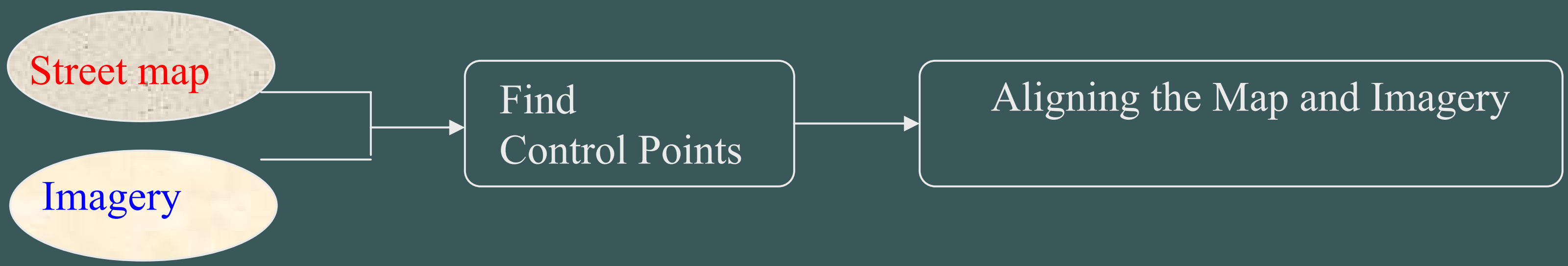
- Traditionally, the problems of vector-imagery and map-imagery alignment have been in the domain of <u>GIS</u> and <u>Computer Vision</u>
- In GIS literature
  - The alignments were previously performed <u>manually</u>
    - Commercial products: *ESRI MapMerger* ; *Able R2V*; *Intergraph I/RASC*
- In Computer Vision literature
  - The alignments were performed automatically based on image processing techniques
    - Often required <u>significant CPU time</u>

Outline

- Introduction & Motivation
- Our approach: AMS-conflation
  - Vector and imagery conflation (pre-qualifying research)
  - Map and imagery conflation
    - Finding control points in the imagery and in the maps
    - Geospatial point pattern matching (GeoPPM)
    - Image and map conflation using rubber-sheeting
- Experimental Results
- Related Work
- Conclusion and Future Work

Aligning Geospatial Data Using Conflation Technique

- Conflation: Compiling two geo-spatial datasets by establishing the correspondence between the matched entities and transforming other objects accordingly
- Requires identifying matched entities, named <u>control points</u>, on both datasets

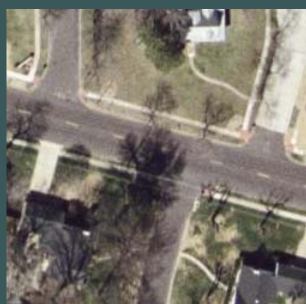


Our Approach: AMS-Conflation

Automatic Multi-Source Conflation

- Thesis statement: By exploiting multiple sources of geospatial information, we can achieve automatic and accurate conflation of road vector data, street maps and orthoimagery.
  - Automatically exploiting information from each of the sources to be integrated to generate accurate control point pairs
  - Exploited geospatial information from one data source can help the processing of the other source

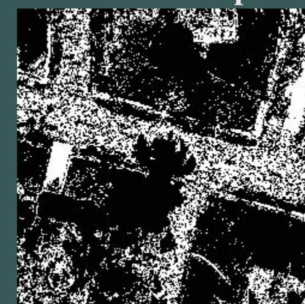
AMS-Conflation :
Exploit Inferred Information from the Data Source

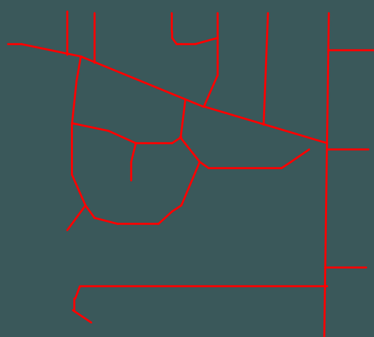


Inferred information from the data source

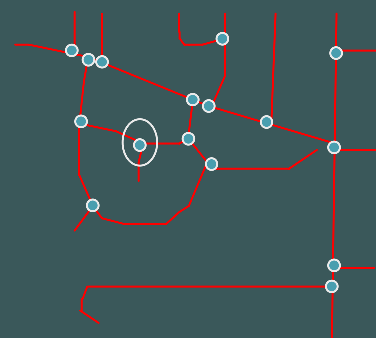
Detected edges

Classified road pixels

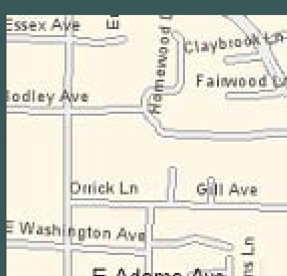




Inferred information from the data source

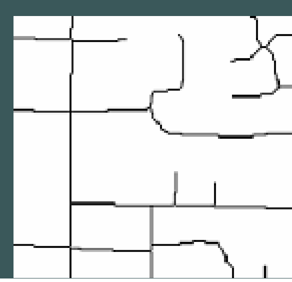


Road Intersections
Road Directions

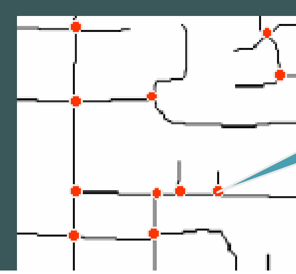


Inferred information from the data source

Detected edges



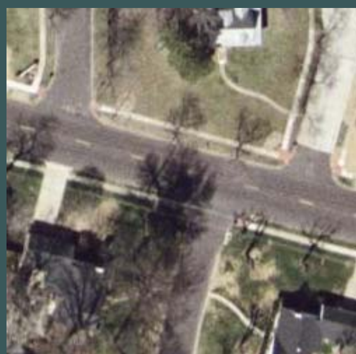
Detected intersections by corner detector



- Degree: 3
- Directions: 1^0, 90^0, 180^0

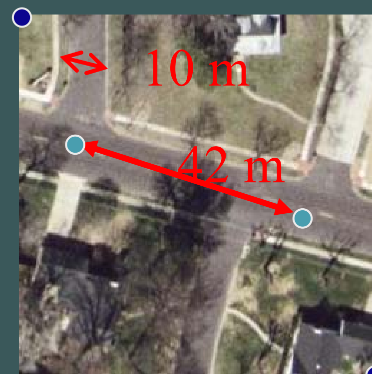
AMS-Conflation :

Exploit Metadata about the Data Source



Metadata about the data source

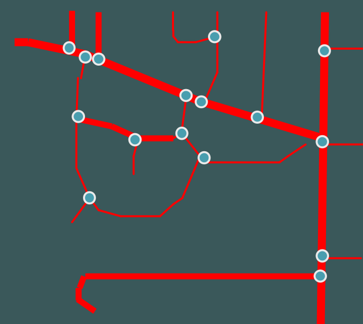
Long: -90.43
Lat: 38.595



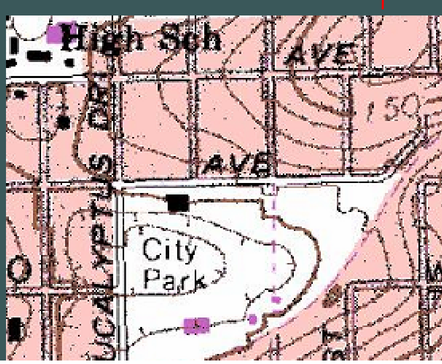
Long: -90.42
Lat: 38.594

Geo-coordinates
Resolution

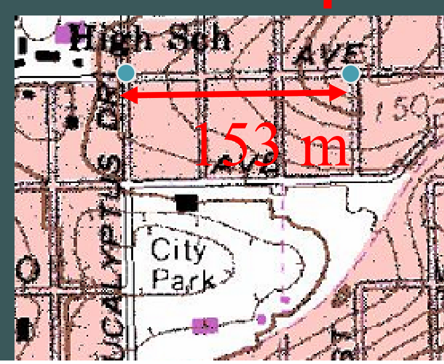
Metadata about the data source



Road widths



Metadata about the data source

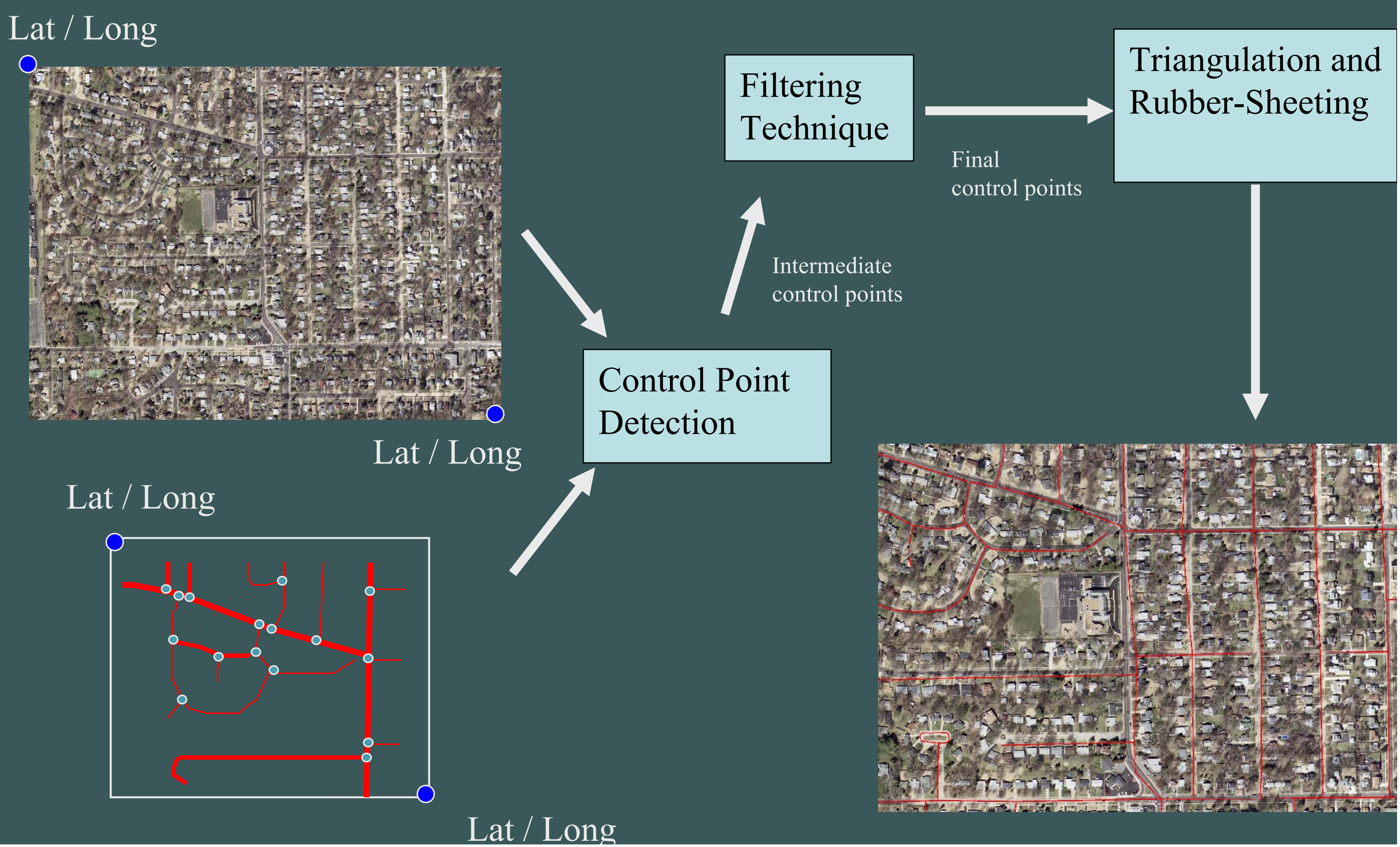


Resolution
(or map scale)

AMS-Conflation :

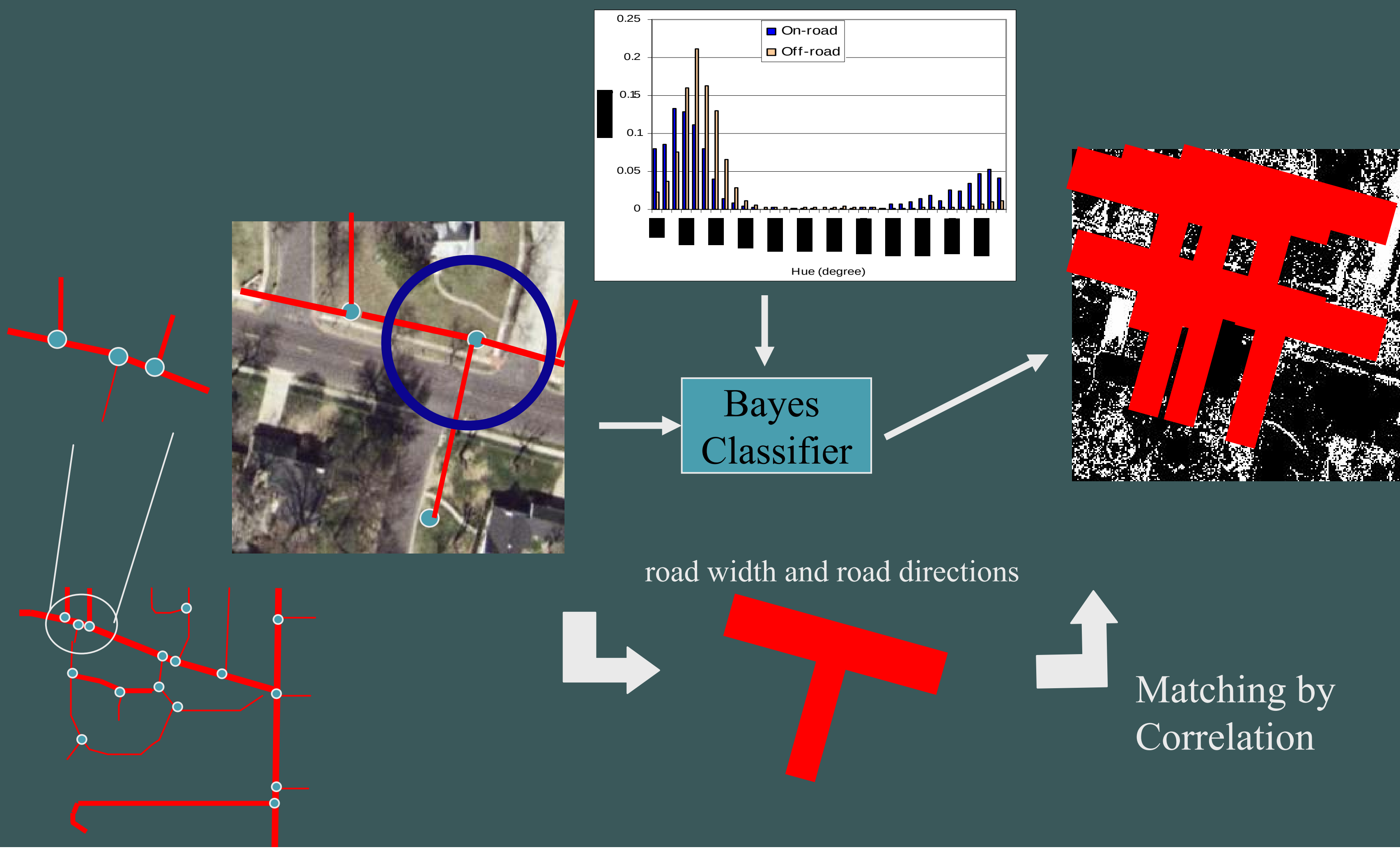
Exploit Peripheral Datasets to the Data Source

AMS-Conflation to Align Vector and Imagery



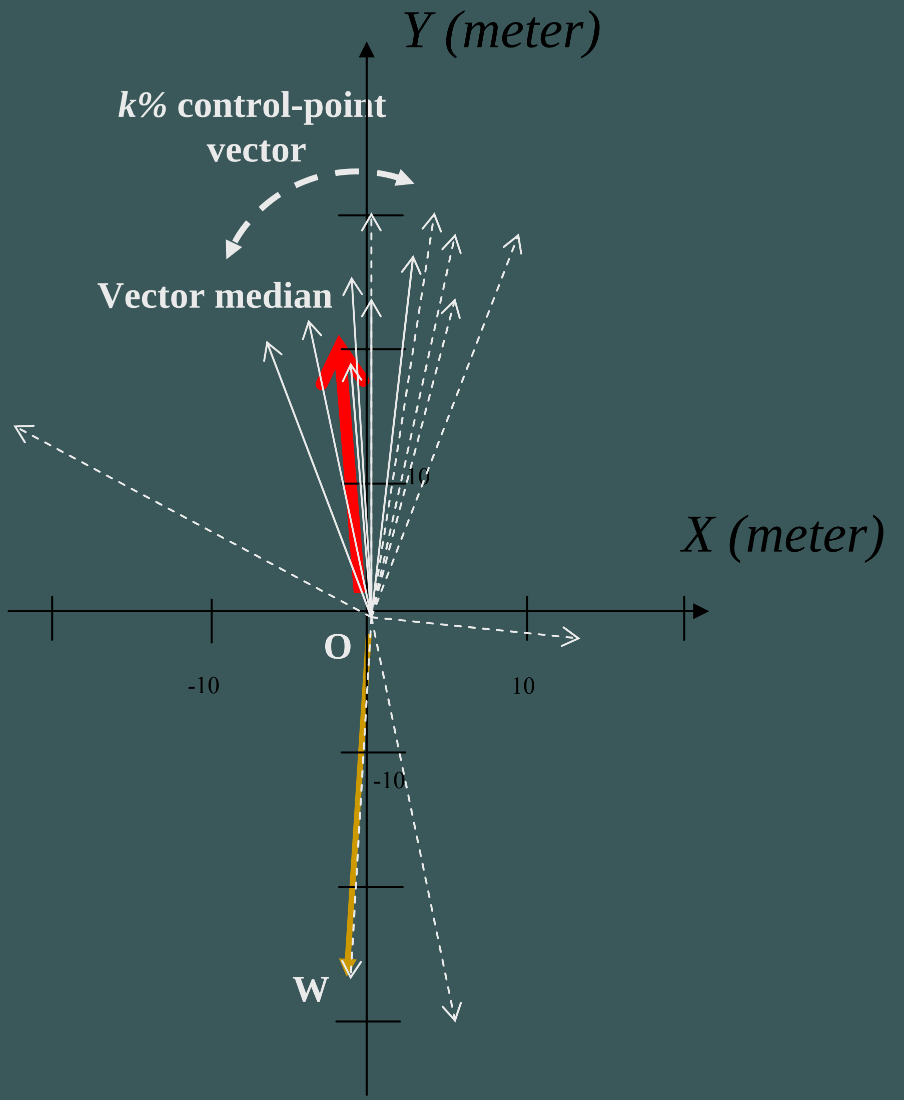
Aligning Vector and Imagery:

Finding Control Point Pairs Using Localized Template Matching (LTM)



Aligning Vector and Imagery:

Filtering Control Point Pairs Using Vector Median Filter (VMF)

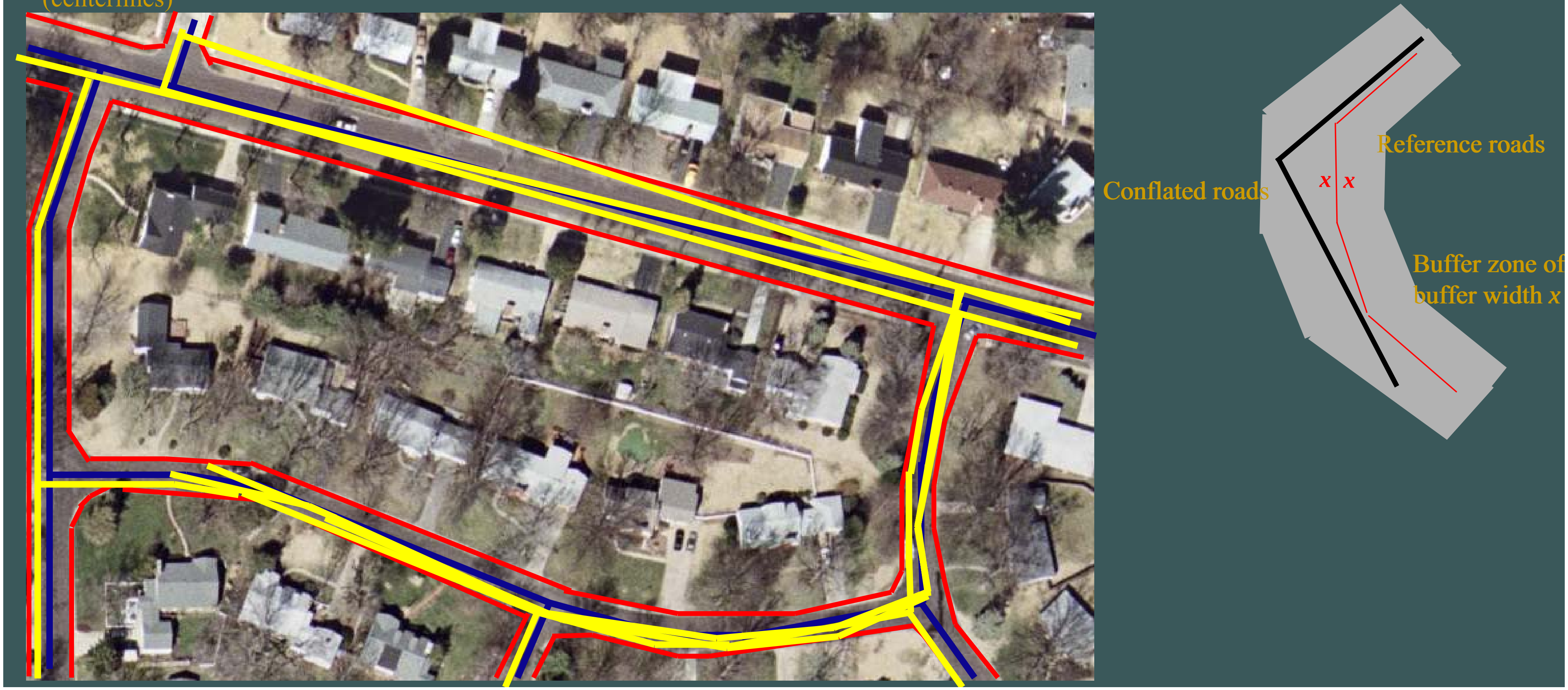


Evaluation

Using road-buffer method

- Red lines: Reference roads (roadsides)
- Blue lines: Reference roads (centerlines)

- Completeness : the percentage of the reference roads for which we generated conflated lines
  - (Length of matched reference roads)/(Length of reference roads)
- Correctness : the percentage of correctly conflated lines with respect to the total conflated lines
  - (Length of matched conflated lines)/(Total length of conflated lines)
- Positional Accuracy : the percentage of conflated roads within x meters to the reference roads

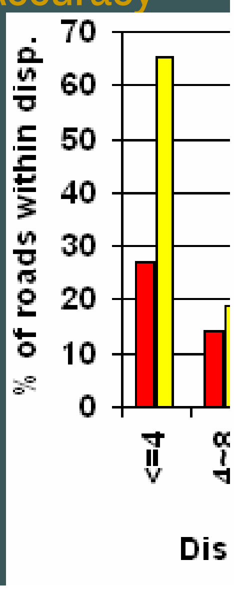


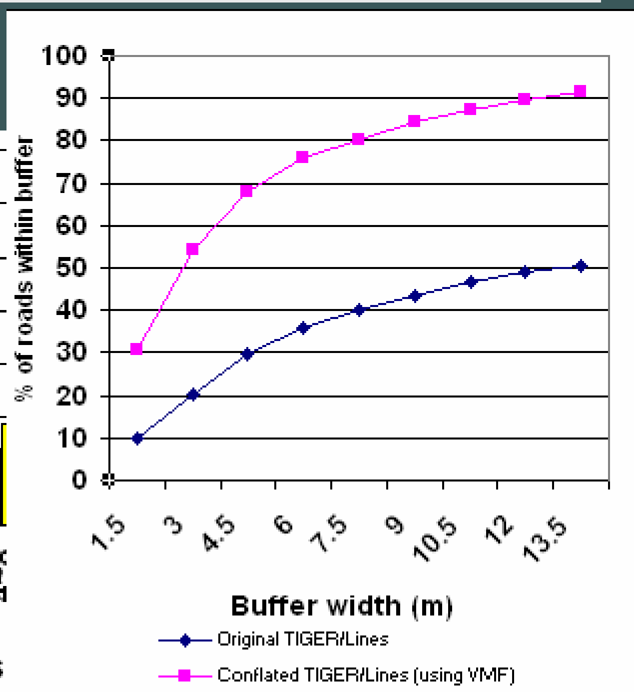
Results:
One of Our Four Test Areas

- For the other test areas, we align different road vector data (MO-DOT, NAVSTREETS and TIGER/Lines) with the imagery

	Original TIGER/Lines	Conflated TIGER/Lines
Completeness	37.9%	84.7%
Correctness	31.3%	88.49%

Positional Accuracy

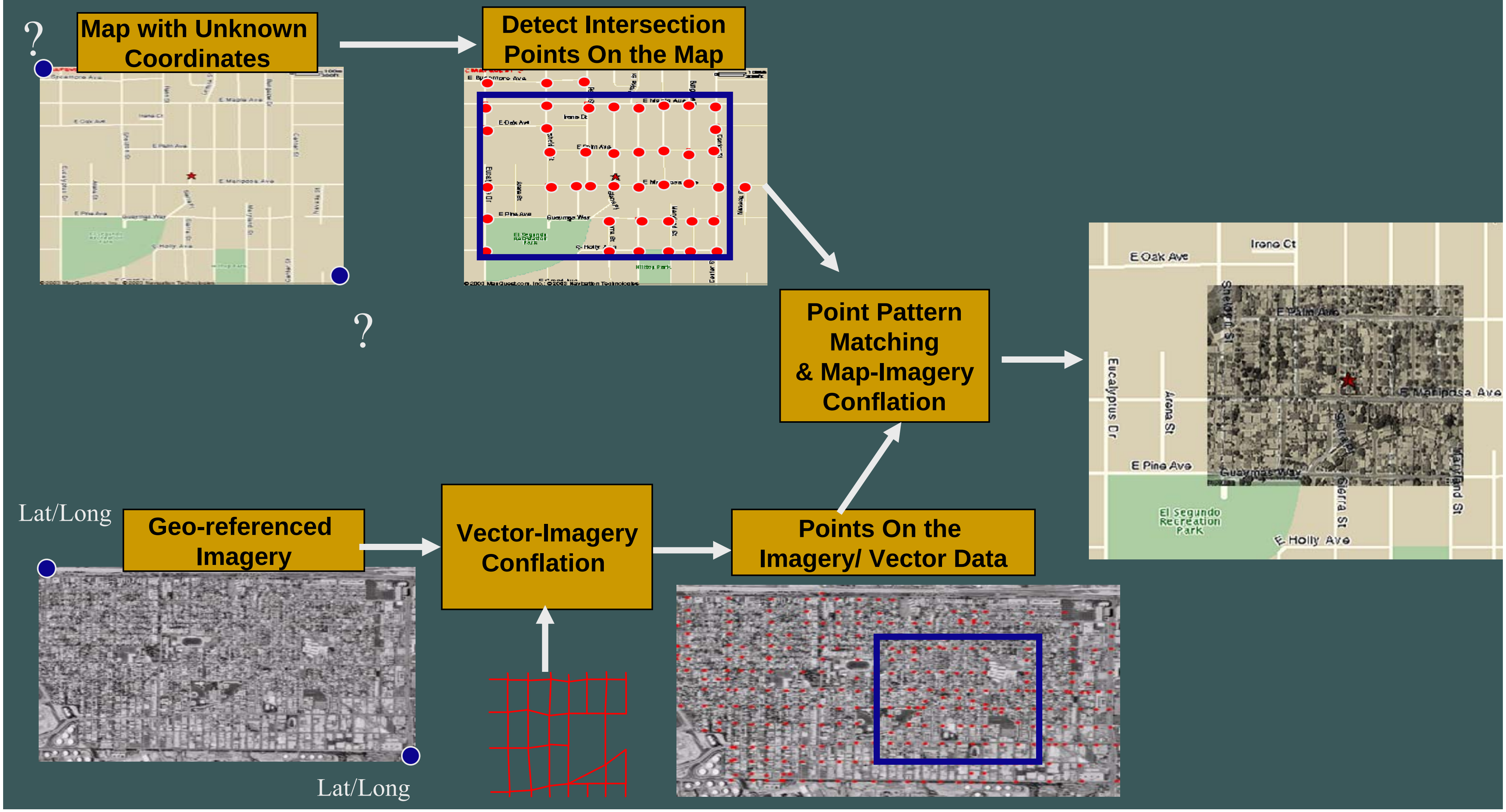




Yellow Lines: Conflated TIGER/Lines

Red Lines: Original TIGER/Lines

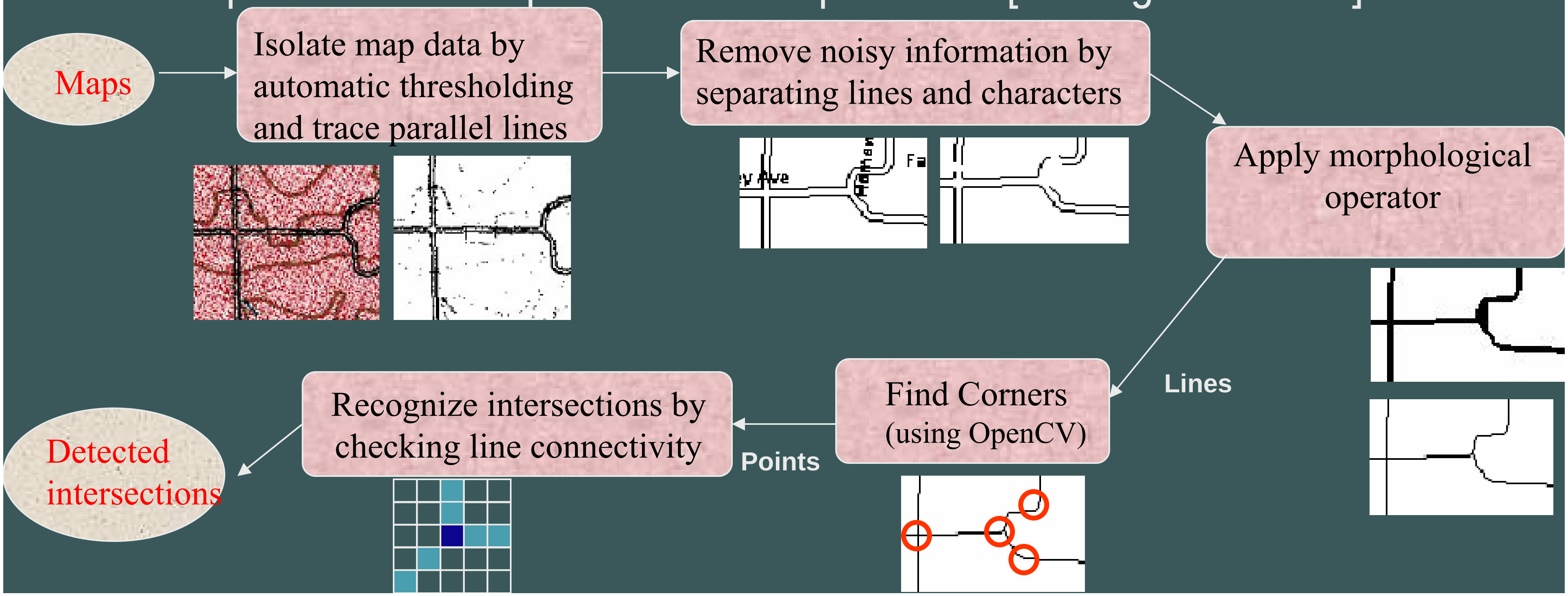
AMS-Conflation to Align Maps and Imagery



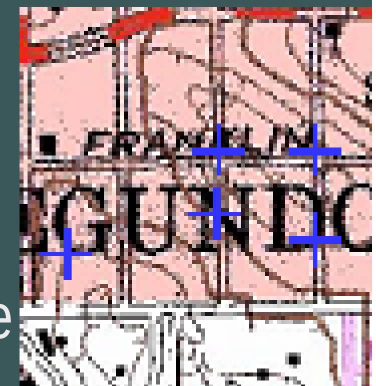
Finding Intersection Points on Maps

- Difficult to identify intersection points automatically and accurately
  - Varying thickness of lines
  - Single-line map v.s. double-line map
  - Noisy information: symbols and alphanumeric characters

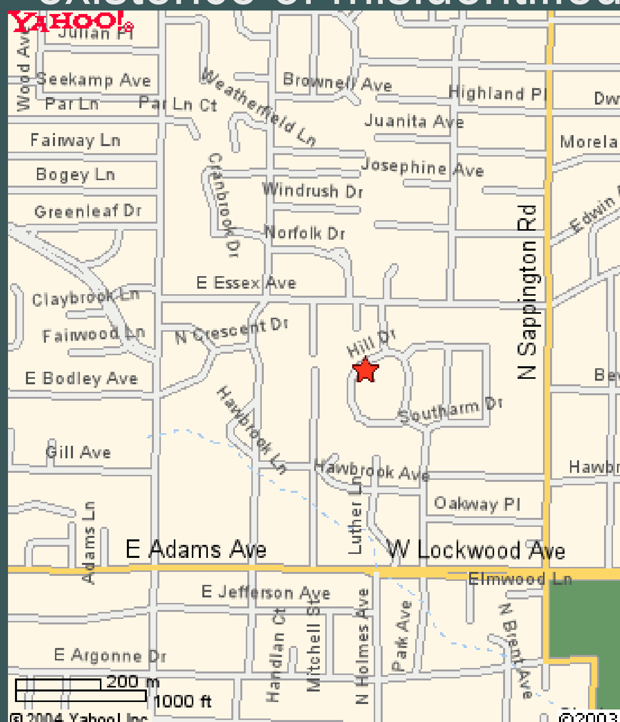
- We proposed a technique to detect intersections in [acm-gis'04]
- Our primitive technique is further improved in [Chiang et al.'05 ?]



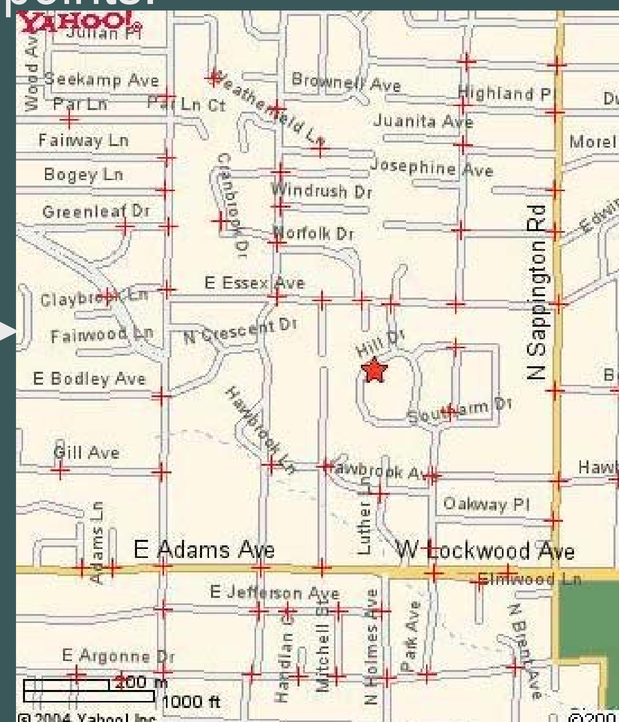
Finding Intersection Points on Maps



- Some <u>noisy points</u> will be detected as intersection points.
- Our geo-spatial point matching algorithm can tolerate the existence of misidentified intersection points.



Identify Intersections



Point Pattern Matching

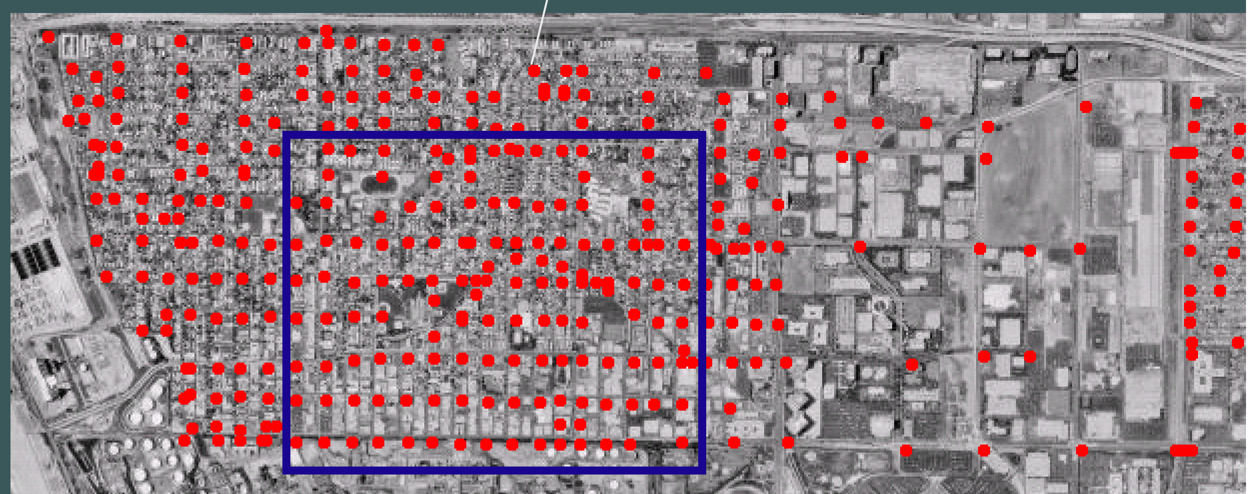
- Find the mapping between these points
  - Why ? To generate a set of control point pairs
- How to solve the point sets matching problem :
  - A geometric point sets matching problem
  - Find the <u>transformation *T*</u> between the layout (with relative distances) of the two point sets

80 points

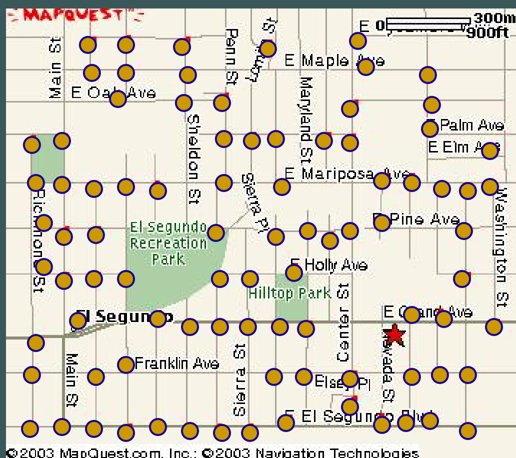
Example: (x,y) = (83,22)

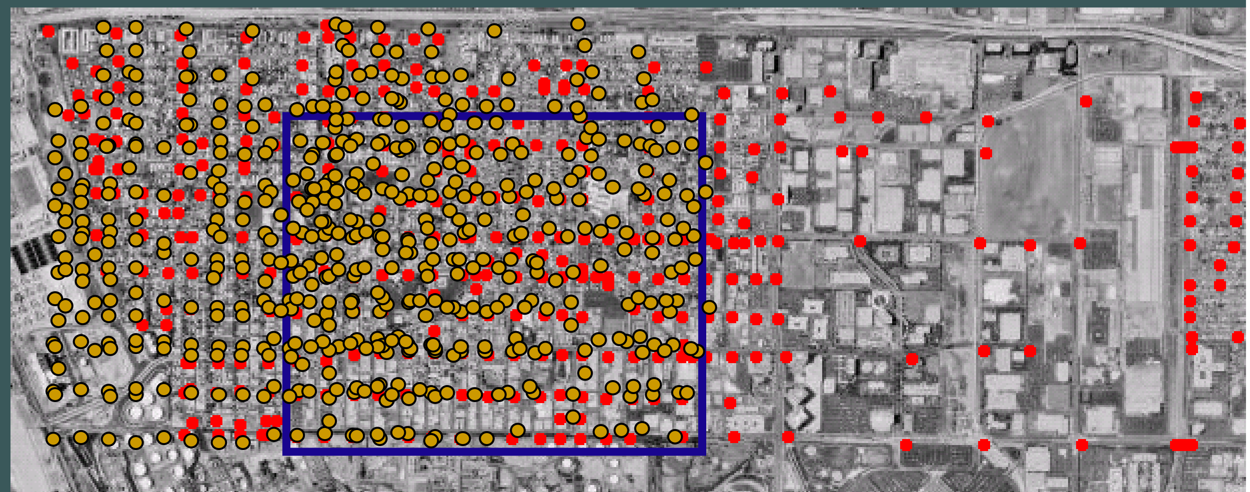
400 points

Example: (lon,lat) = (-118.407088,33.92993)

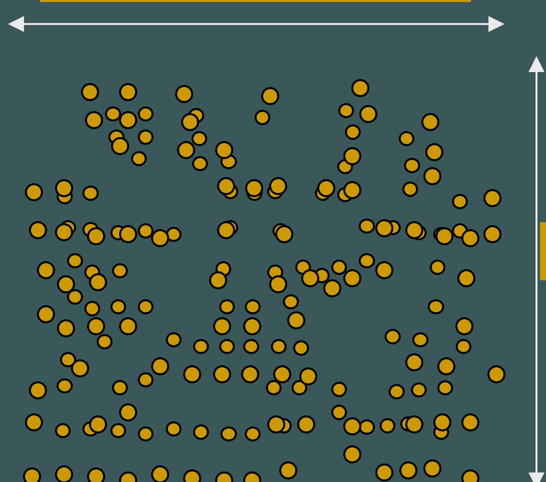


Point Pattern Matching: Finding the Transformation





Scaling in West-East direction



Scaling in N n direction

- Transformation = Scaling + Translation
  - Transforms most points on map to points on imagery
  - Find matching point pairs to solve this transformation

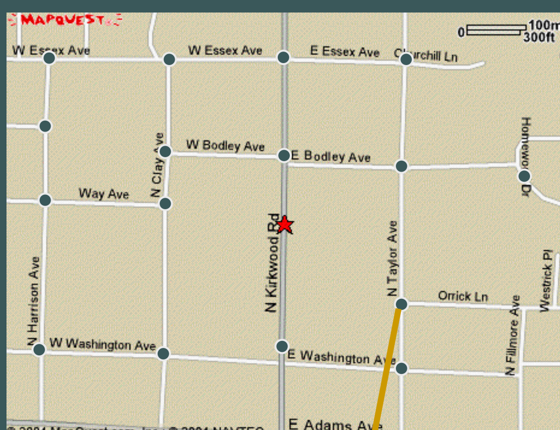
Point Pattern Matching: A Brute-Force Algorithm

- Iterate all point pair in M, and for each chosen point pair in M examining all point pairs in S
  - Time-consuming : $O(m^3\, n^2\, \mathit{log}\, n)$
    - Can we improve it by randomization ? Not always !
      - Noisy points on maps
      - Some missing points on imagery

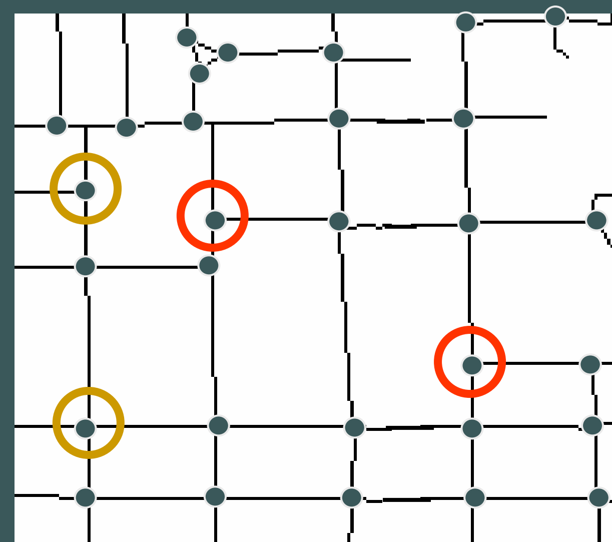
Geospatial Point Pattern Matching (GeoPPM):

Exploit Geometric Info. Associated with Each Intersection

- Intersection degree: the number of intersected roads
- Directions of Intersected road segments



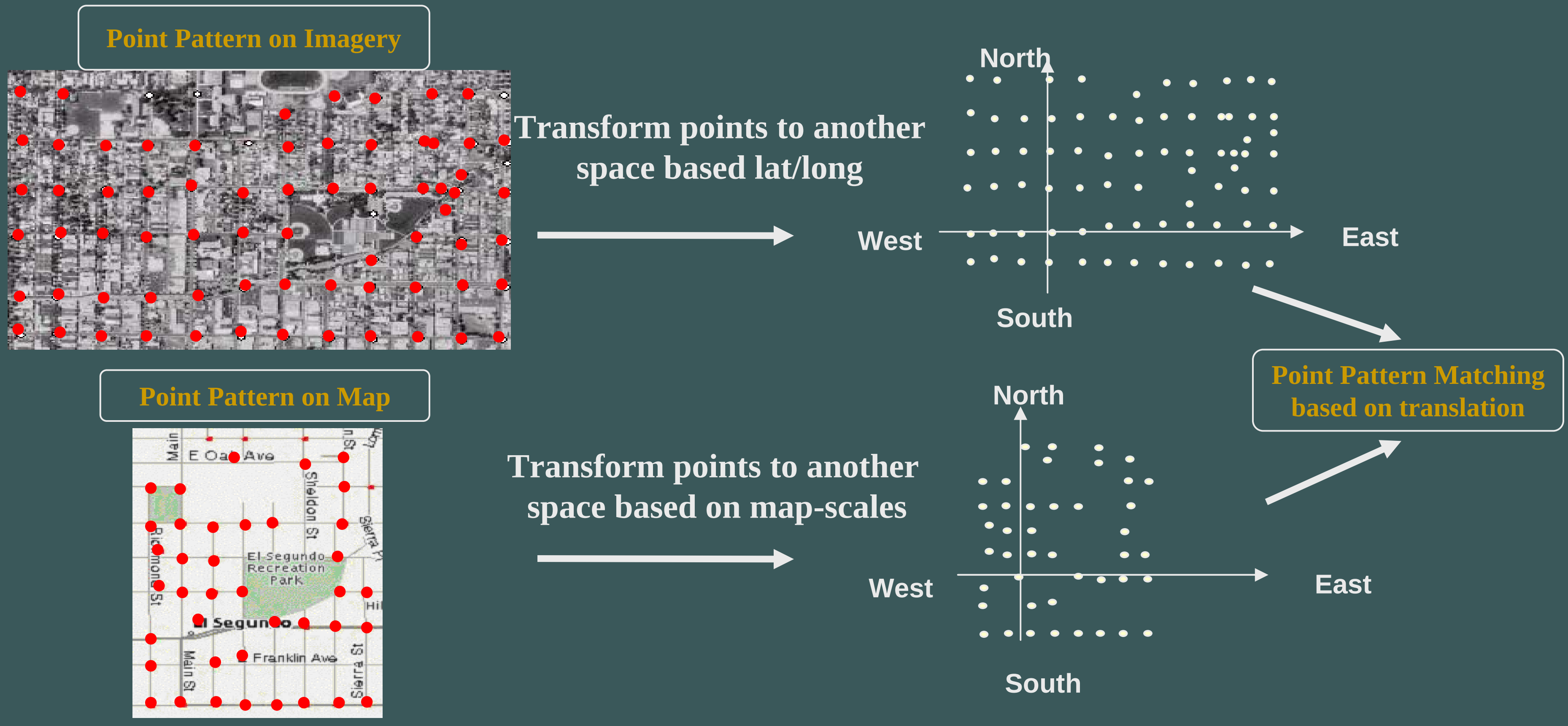
Degree:3;
Directions:0, 90, 270



Geospatial Point Pattern Matching (GeoPPM):

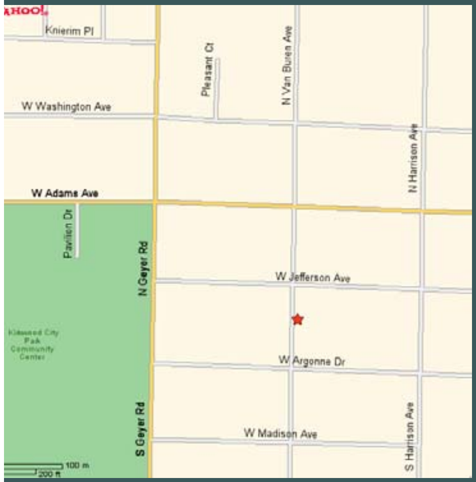
Exploit Map Scale

- We need to consider <u>translation</u> only $O(m^3\ n^2\ log\ n) \rightarrow O(m^2\ n\ log\ n)$



Geospatial Point Pattern Matching (GeoPPM):
For Map with Unknown Map Scale

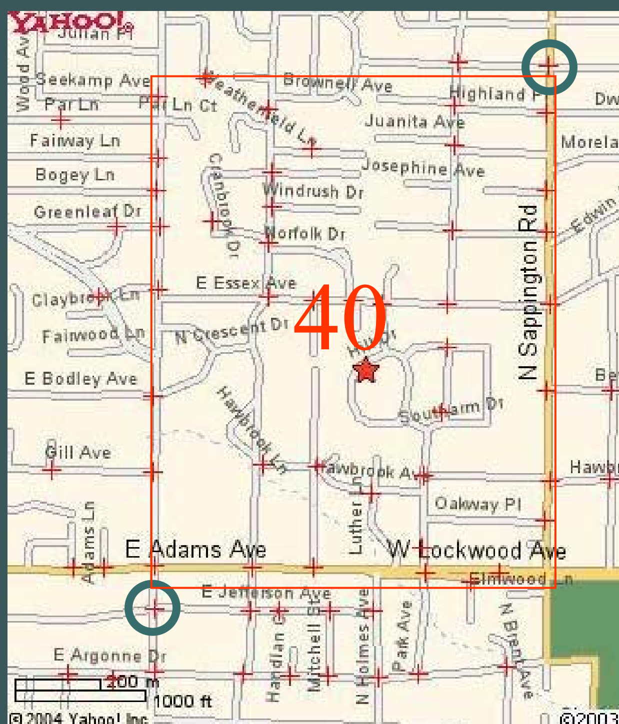
- Exploiting Point Density and Localized Distribution of Points
  - Assumption: we focus on medium to high resolution maps
    - We are conflating maps with high resolution imagery !



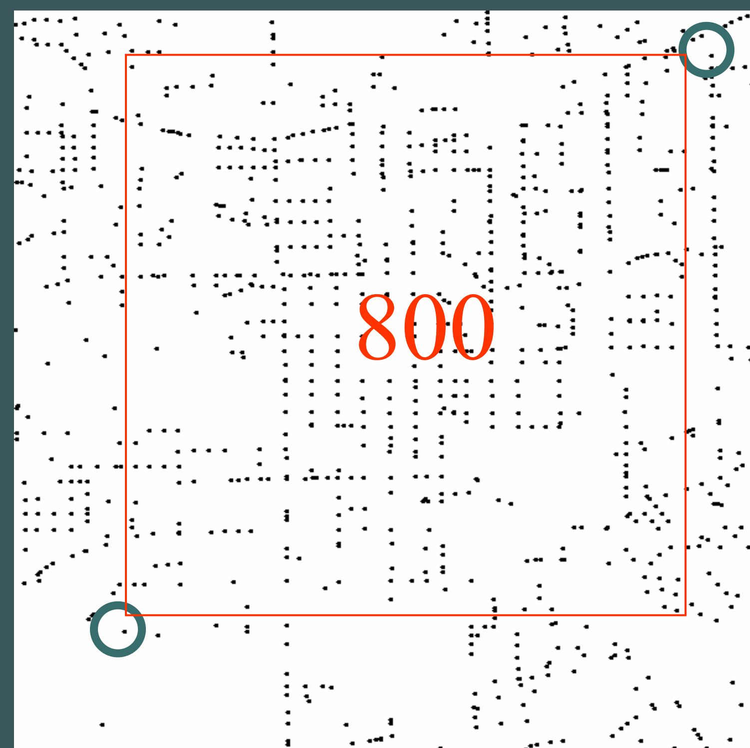
Level 1: 1.2 m/pixel **Level 2:** 4.25 m/pixel **Level 3:** 14.08 m/pixel **Level 4:** 35 m/pixel

Coarse level map: map with smaller map-scale (low resolution)

Geospatial Point Pattern Matching (GeoPPM):
Exploit Point Density



55 points



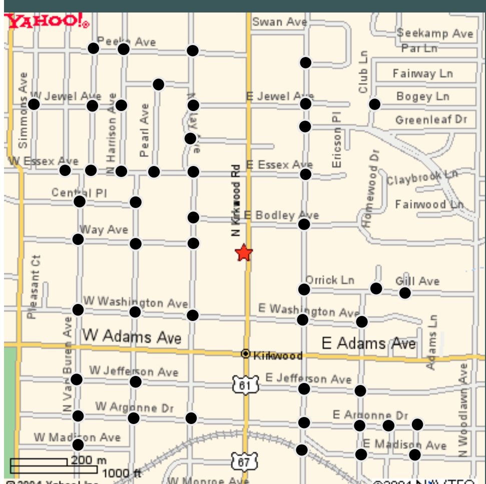
1059 points

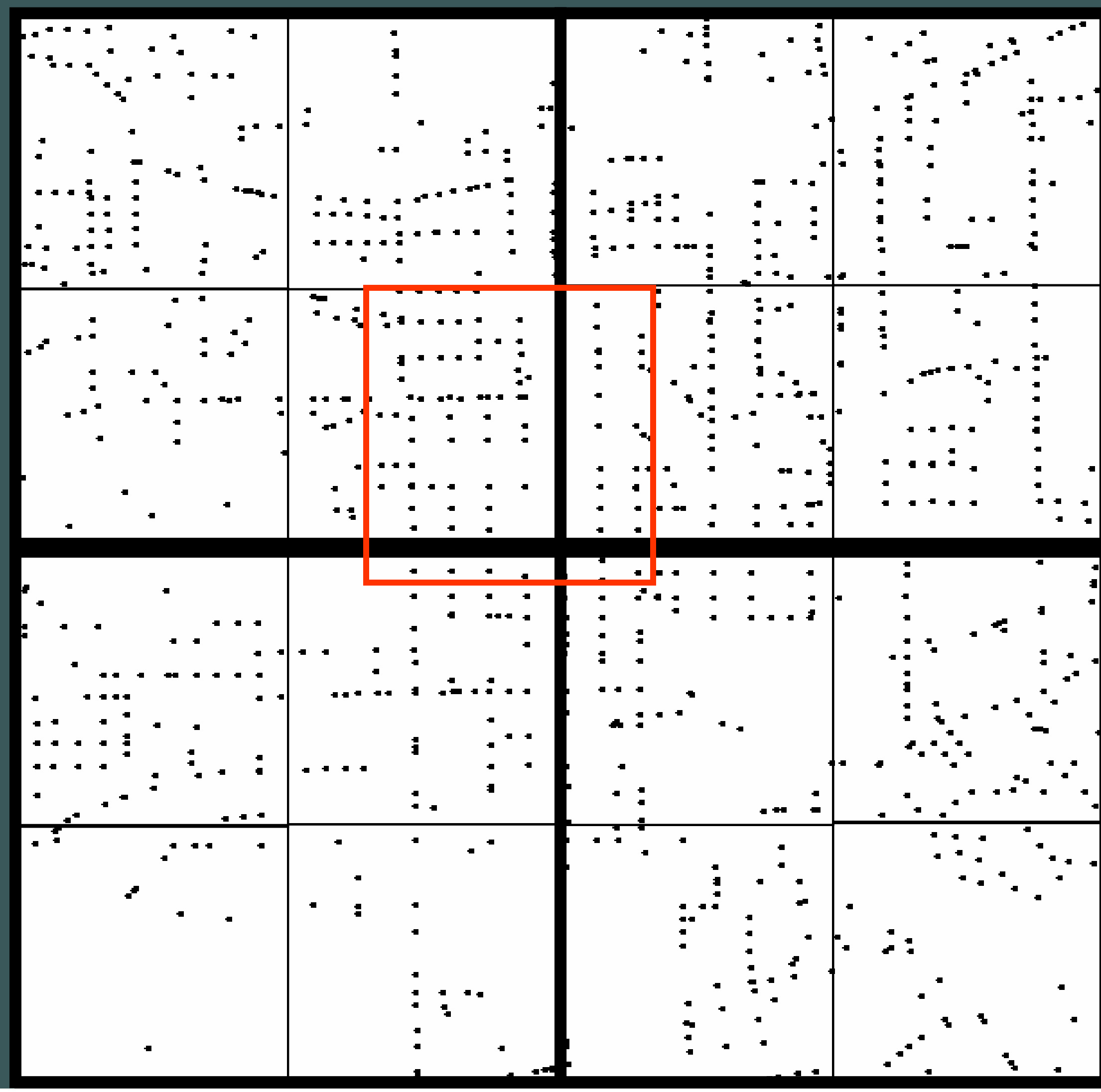
Geospatial Point Pattern Matching (GeoPPM):

Exploit Localized Distribution of Points

- The points are in a cluster !

57 detected map points

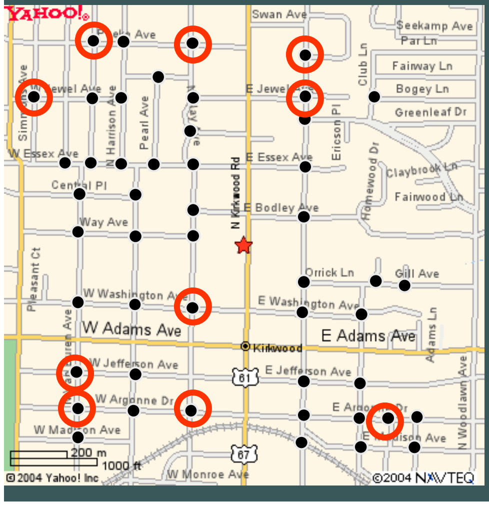


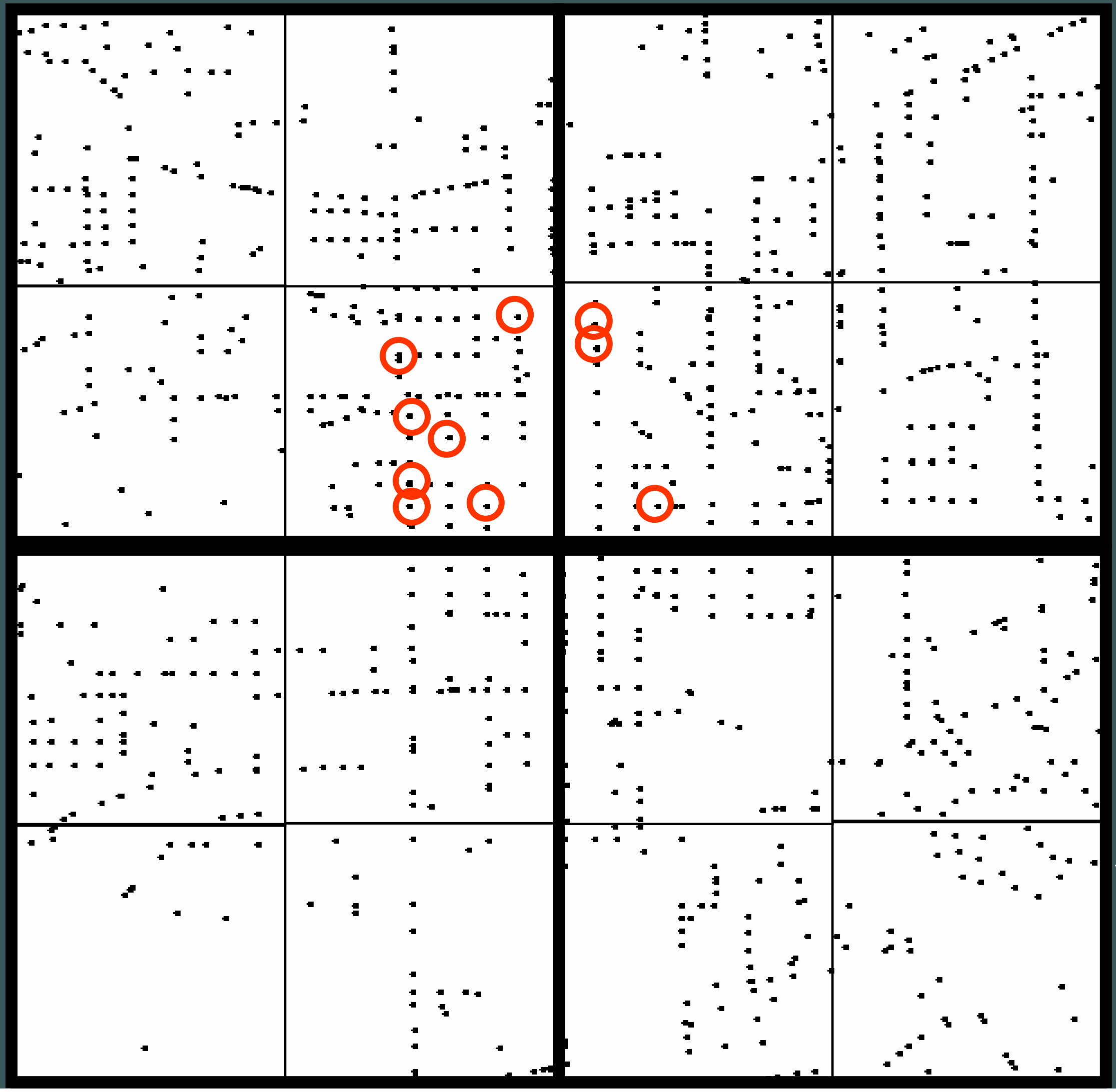


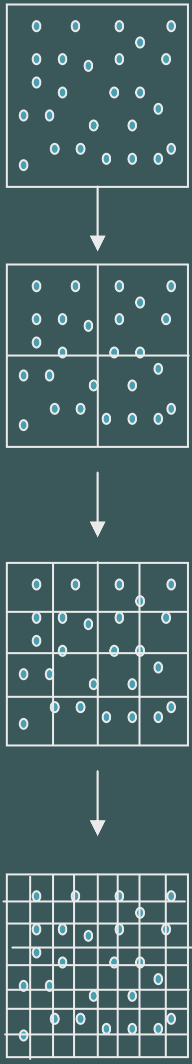
1059 points

Geospatial Point Pattern Matching (GeoPPM):

Exploit Localized Distribution of Points Using HiGrid Structure







Level 0
(1 cell)

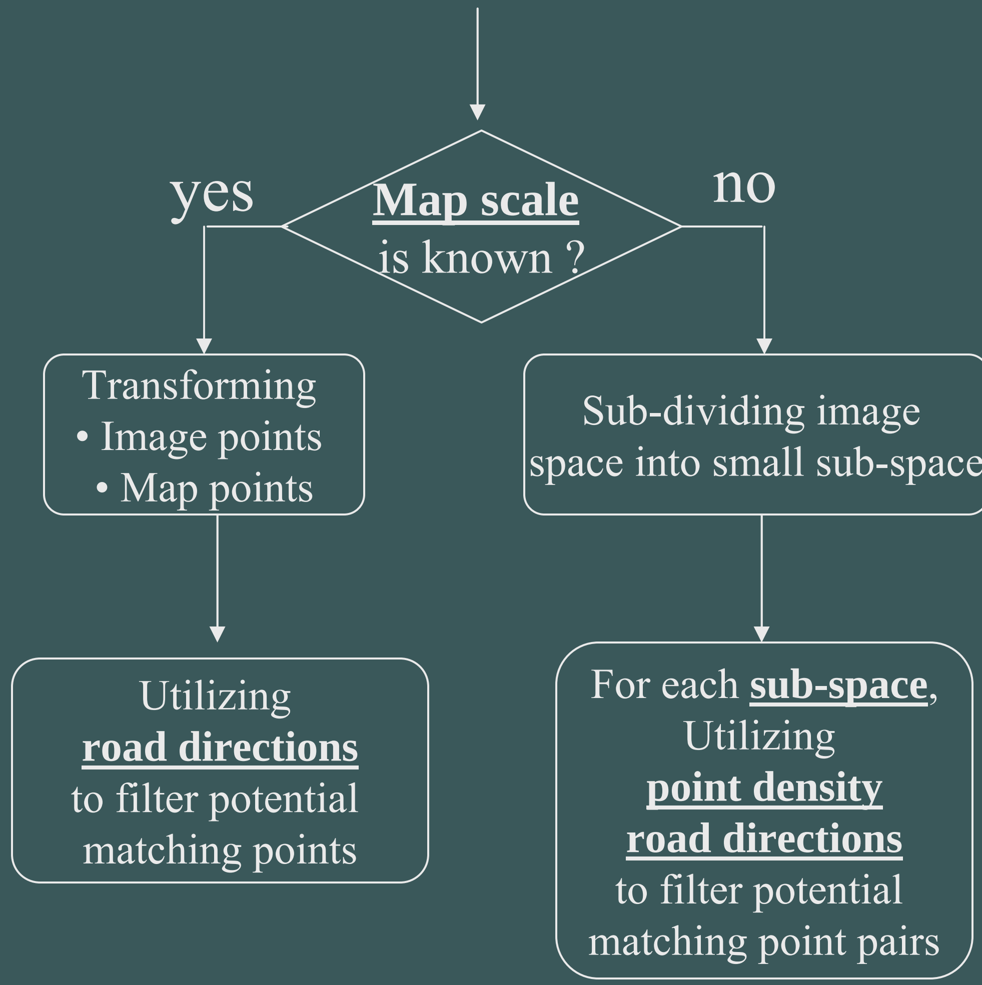
Level 2
(4 cells)

Level 3
(16 cells)

Level 4
(64 cells)

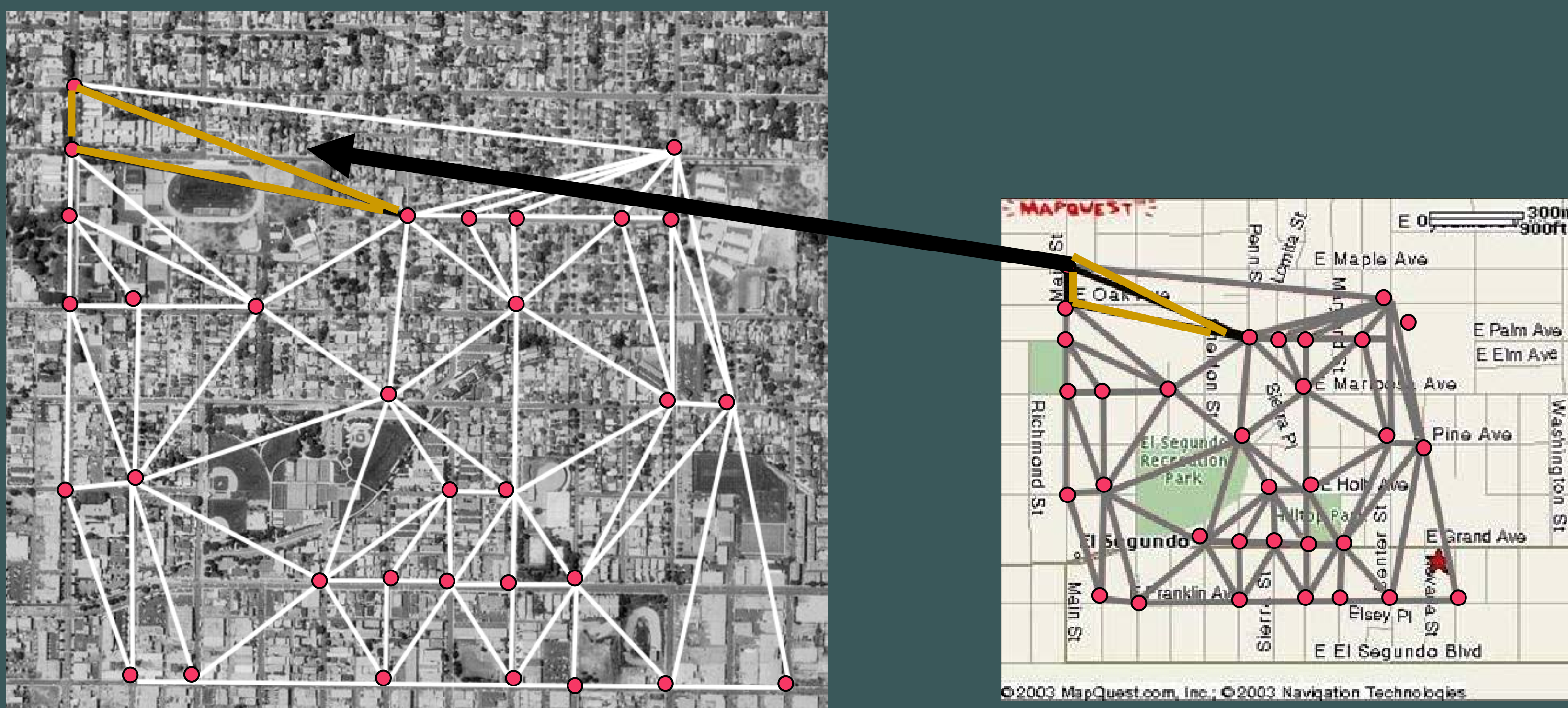
Geospatial Point Pattern Matching (GeoPPM)

- Current GeoPPM implementation
  - Utilizing these exploited information simultaneously to prune search space
    - Road directions
    - Map scale
    - Point density
    - Localized distribution of points



Aligning Maps and Imagery

- Using matched point pattern to align maps with imagery by <u>Delaunary triangulation and rubber-sheeting [Saalfeld’88]</u>
  - Space partition to build influence regions: Delaunary triangulation
  - Warping maps’ pixels within each triangle to the corresponding pixels on imagery : based on Delaunary triangles and rubber-sheeting



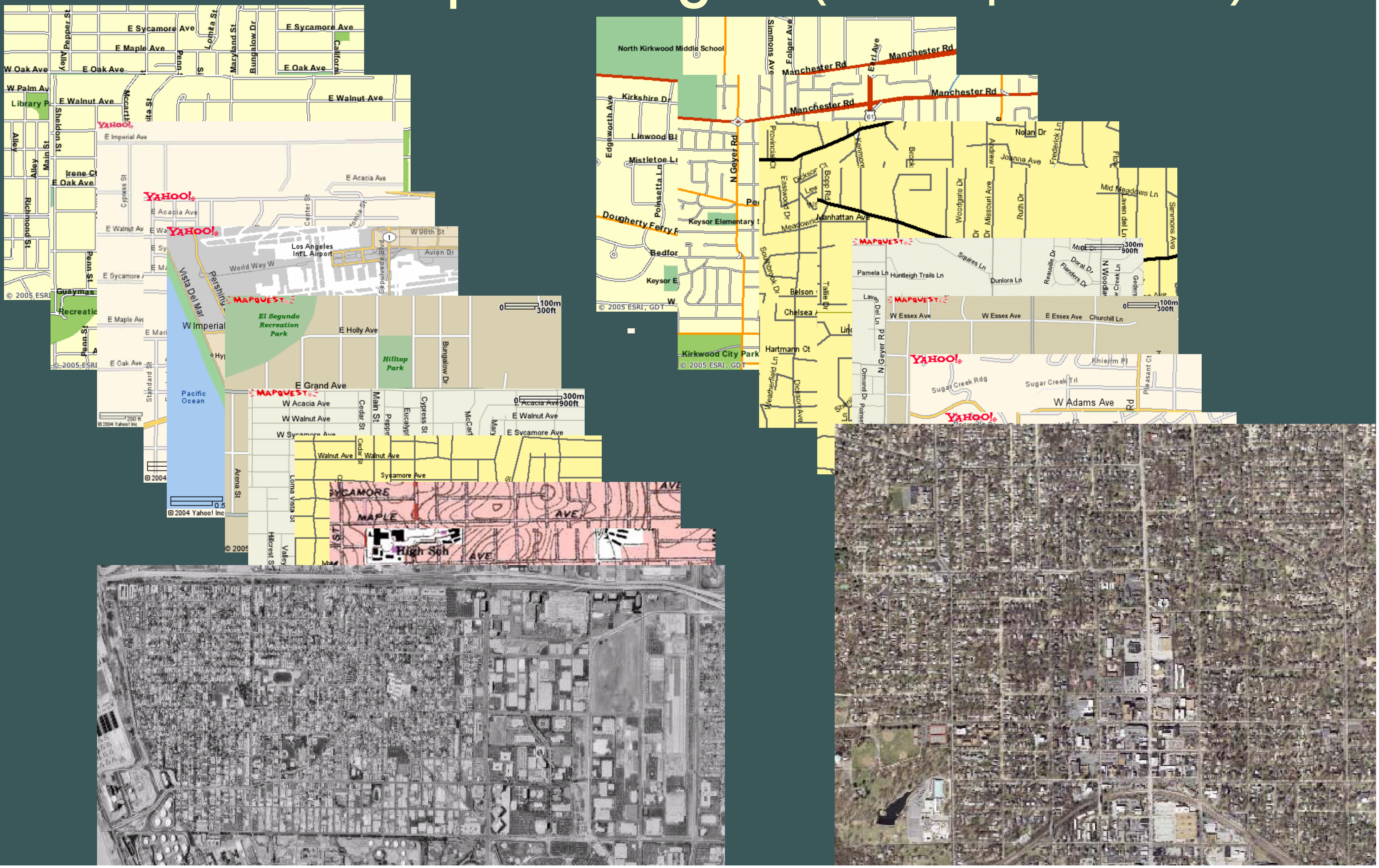
Outline

- Introduction & Motivation
- Our approach: AMS-conflation
  - Vector and imagery conflation (pre-qualifying research)
  - Map and imagery conflation
    - Finding control points in the imagery and in the maps
    - Geospatial point pattern matching (GeoPPM)
    - Image and map conflation using rubber-sheeting
- Experimental Results
- Related Work
- Conclusion and Future Work

Experimental Setup: Test Data Sets

	Test area 1	Test area 2
Imagery	Geo-referenced USGS black-white orthoimagery (1 meter/pixel)	Geo-referenced USGS color orthoimagery (0.3 meter/pixel)
Maps	5 maps for each map service: ESRI, MapQuest, Yahoo, TIGER, USGS topographic maps	5 maps for each map service: ESRI, MapQuest, Yahoo, TIGER, USGS topographic maps
Vector Data	TIGER/Lines	MO-DOT
Area covered	Partial area of City of El Segundo, CA (total road length is about 84.32km)	Partial area of St.Louis, MO (total road length is about 364.28km)

Experimental Setup:
Some Sample Images (50 maps in total)

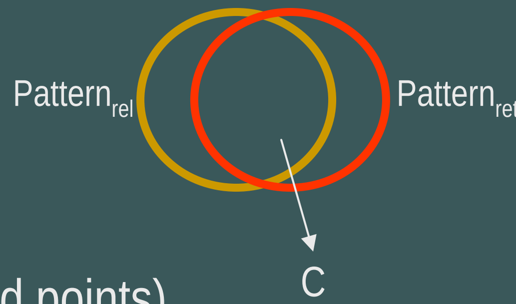


Results


MAPQUEST
E Imperial Ave
E Acacia Ave
E Walnut Ave
E Sycamore Ave
E Maple Ave
E Oak Ave
E Palm Ave
E Elm Ave
E Mariposa Ave
E Pine Ave
California St
Washington St
Indiana Ct
Illinois Ct
Sepulveda Blvd
Illinois St
Center
Bung
Mary

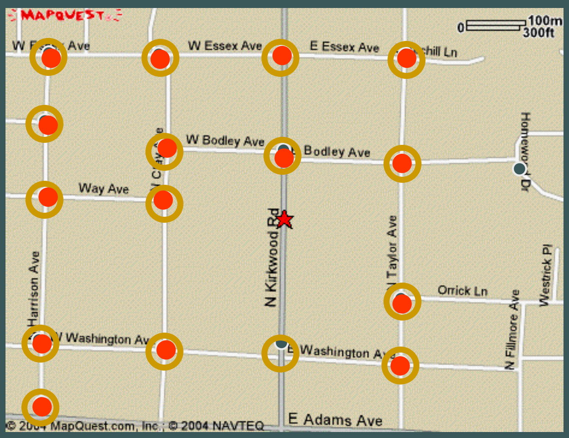

W Jewel Ave
Taylor Est
Pearl Ave
W Mermod Pl
ood Rd
N Taylor Ave
Ericson Pl
Homewood Dr
E Bodley Ave
Orrick Ln
E Washington Ave

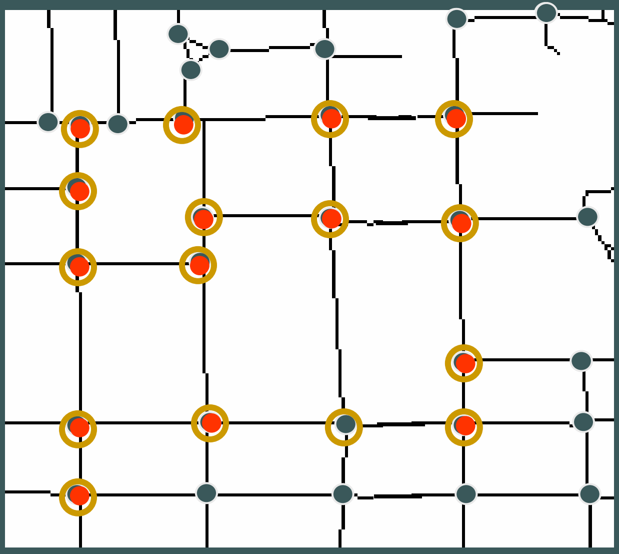
Evaluation:
The performance of GeoPPM



- Definition:
  - Pattern$_{rel}$: The relevant point pattern (and there are *x* matched points)
  - Pattern$_{ret}$: The retrieved point pattern by GeoPPM (and there are *y* matched points)
  - Let set C = Pattern$_{rel}$ ∩ Pattern$_{ret}$ (and there are *z* matched points)
  - Precision = z / y ; Recall = z / x

x =16
y =15
z =15
Precision=100%
Recall= 93.75%





Evaluation:

The performance of GeoPPM in Precision/Recall

	ESRI map	MapQuest map	Yahoo map	TIGER map	Topographic map
Precision	96.0%	95.2%	94.0%	84.2%	93.9%
Recall	80.2%	84.8%	88.3%	75.6%	80.94%

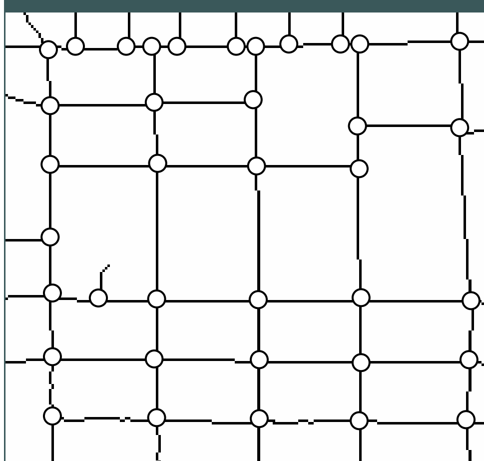
	Test data set 1 (El Segundo, CA)	Test data set 2 (St. Louis, MO)
Precision	91.9%	93.4%
Recall	84.6%	77.4%

	Precision	Recall
Res $\leq$ 2m/pixel (38%)	87.4%	78.2%
2 < Res $\leq$ 4 (18%)	92.9%	84.0%
4 < Res < 7(33%)	96.4%	88.6%
Res > 7 (13%)	91.6%	77.1%

Evaluation: The performance of GeoPPM

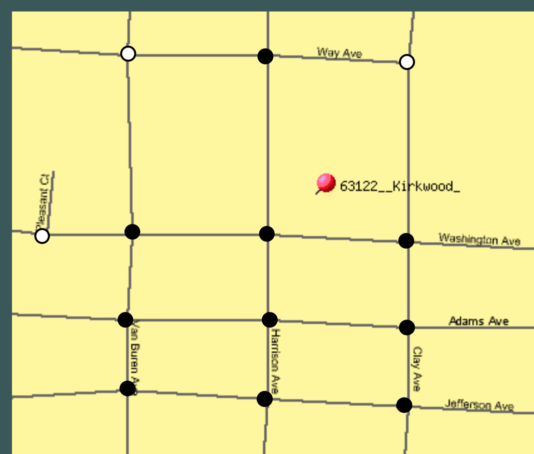
- One of our 50 tested maps where the intersection point set is not accurately aligned with the corresponding point pattern on the image
  - 13 points TIGER map (resolution: 1.85m/pixel)
  - 13 matched points in relevant point pattern
  - 10 matched points in the retrieved pattern

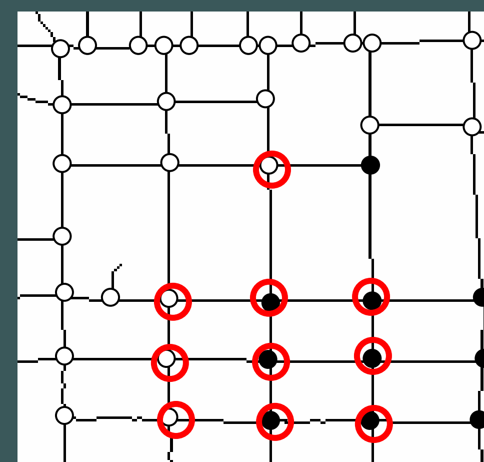
Map points



Imagery points

GeoPPM

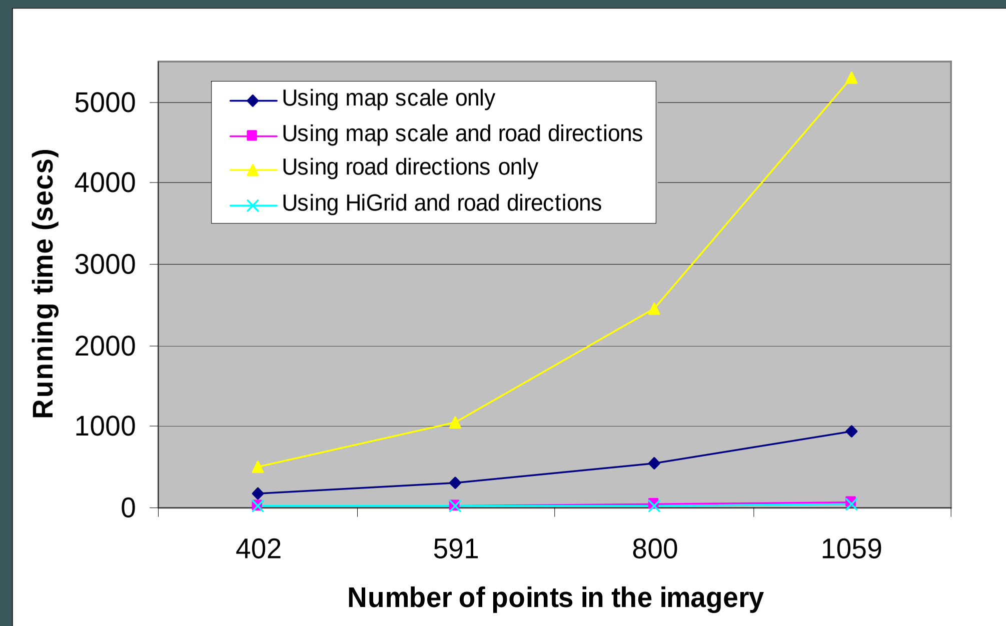




Evaluation:
The running time of GeoPPM

- Platform: Windows 2000; CPU Xeon 1.8GHz with 1GMB memory
- Test on a Yahoo map with 57 points with varying number of image points

57 map points

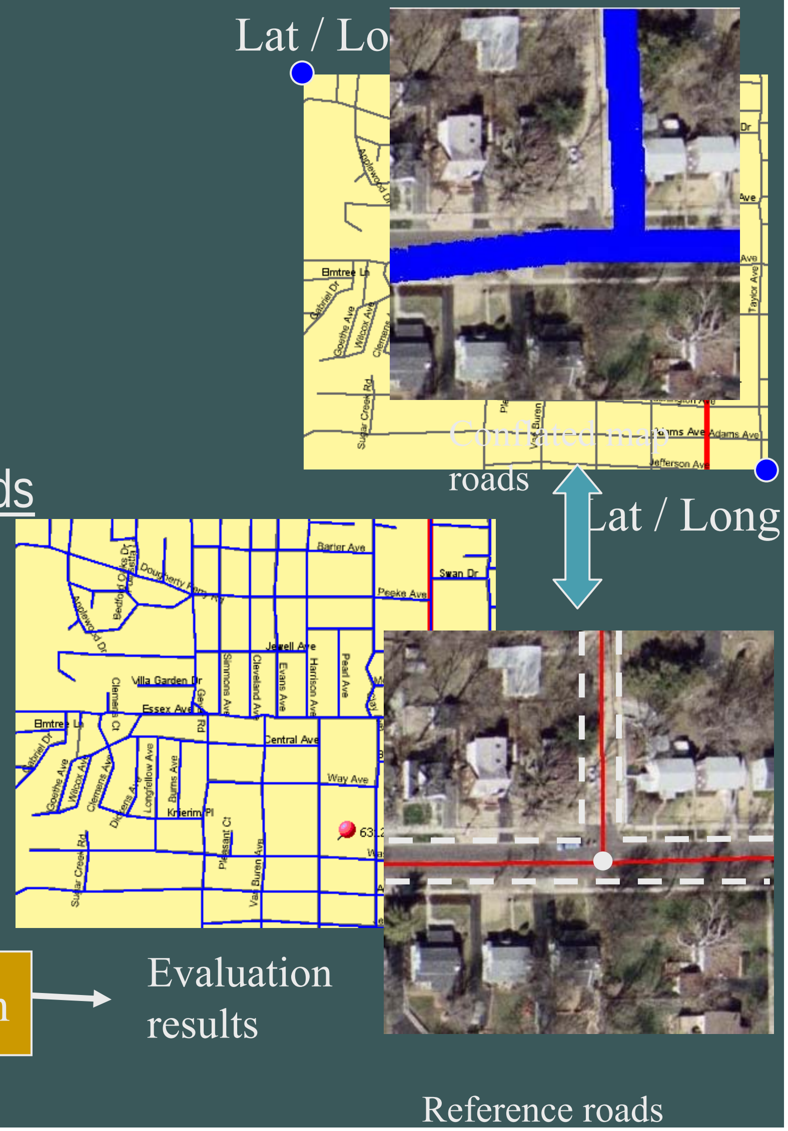


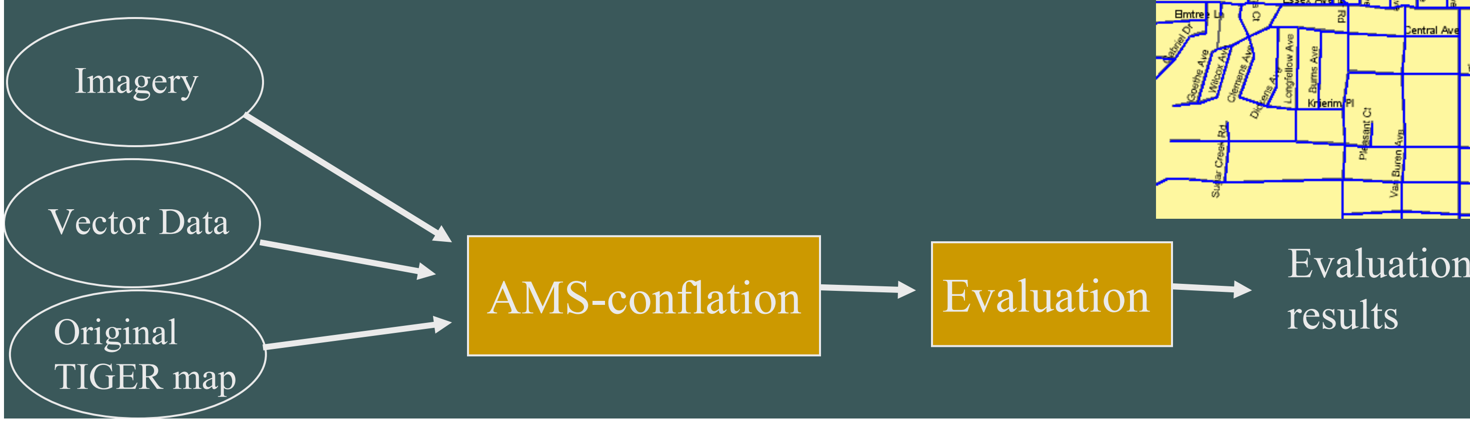
	Brute force algorithm	Using map scale only	Using map scale and road directions	Using road directions	Using HiGrid and road directions
402 imagery points	5 hours 58 minutes	171 seconds	16 seconds	503 seconds	11 seconds
591 imagery points	N/A	317 seconds	26 seconds	1049 seconds	17 seconds
800 imagery points	N/A	540 seconds	42 seconds	2449 seconds	26 seconds
1059 imagery points	N/A	934 seconds	70 seconds	5298 seconds	38 seconds

Evaluation:

The performance of overall map-imagery conflation

- The <u>conflated map roads</u> v.s.
  the <u>corresponding roads</u> in the imagery
  - Use TIGER maps for evaluation
    - TIGER maps are georeferenced
    - Roads on TIGER maps can be “vectorized” (or represented) by TIGER/Lines vector data
  - Compare conflated map roads with <u>reference roads</u> (manually plotted roads)
    - Completeness/ Correctness /Positional Accuracy

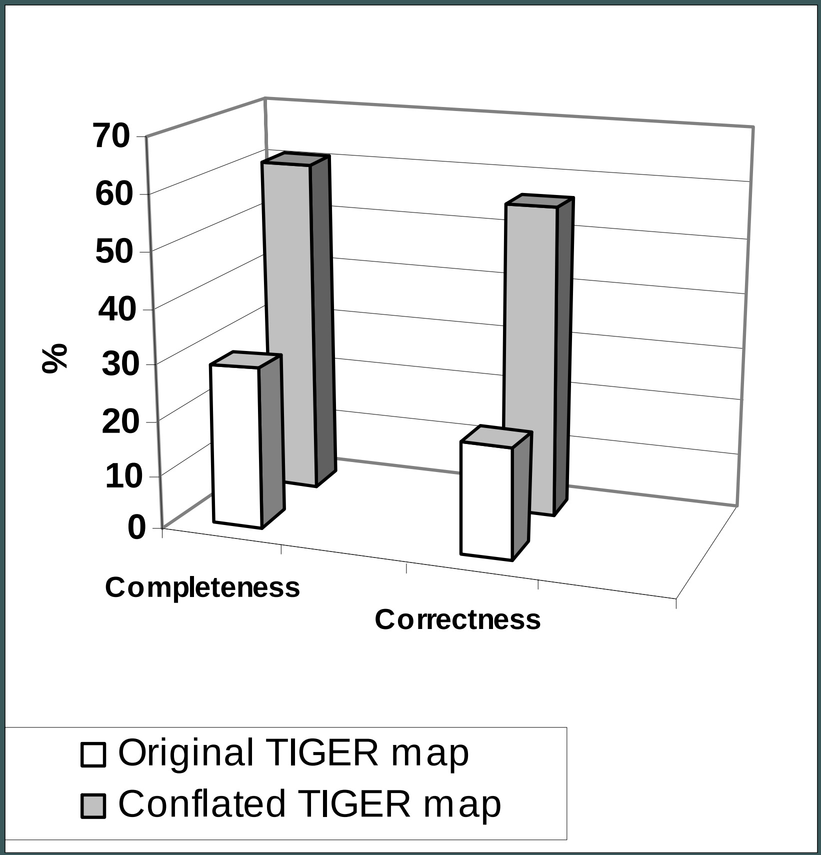




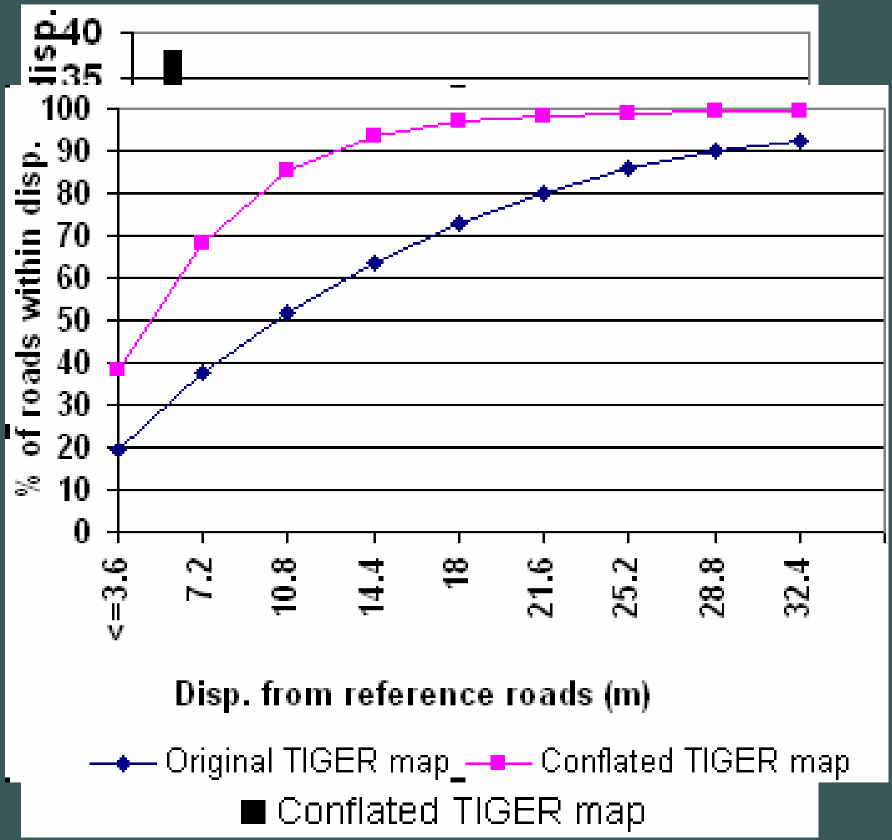
Evaluation:

The performance of overall map-imagery conflation

Completeness/Correctness



Positional Accuracy



Outline

- Introduction & Motivation
- Our approach: AMS-conflation
  - Vector and imagery conflation (pre-qualifying research)
  - Map and imagery conflation
    - Finding control points in the imagery and in the maps
    - Geospatial point pattern matching (GeoPPM)
    - Image and map conflation using rubber-sheeting
- Experimental Results
- Related Work
- Conclusion and Future Work

Related Work

- Vector to vector conflation based on corresponding features identified from both vector datasets (in GIS domain)
  - [Walter et al. 99]: Matching features (e.g. intersection points or polygons) at geometry level
  - [Cobb et al. 98]: Matching features both at spatial/non-spatial level
- Vector to imagery conflation
  - Utilizing matched polygons [Hild et al. 98]
  - Utilizing matched lines [Filin et al. 00]
  - Utilizing matched junction-points [Flavie et al. 00]
  - All above solutions
    - Require lots of CPU time
    - Utilize vector data only for verifying detected features not for extracting features

Related Work

- Raster to raster conflation:
  - To the best of our knowledge, there is no research addressing the problem of automatic conflation of maps and imagery
  - Related work of imagery-imagery conflation
    - Sato et al. [Sato 01] proposed an edge detection process was used to determine a set of features that can be used to conflate two image data sets
      - Their work requires that the coordinates of both image data sets be known
    - Dare et al. [Dare 00] proposed multiple feature extraction and matching techniques
      - Need to manually select some initial control points
    - Seedahmed et al. [Seedahmed 02] proposed an approach extract features from imagery by Moravec feature detector and obtain transformation parameters by investigating the strongest clusters in the parameter space
      - Require lots of CPU time
    - Commercial products: Able R2V and Intergraph I/RASC
      - Need to manually select all control points

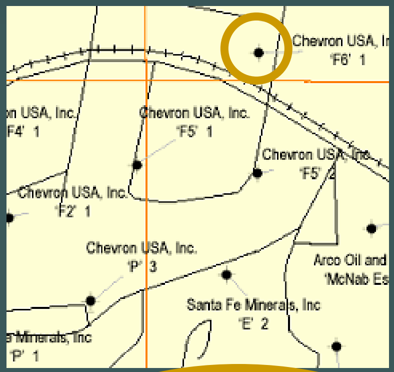
Conclusion

- Our contributions : AMS-Conflation
  - Automatic Vector to Imagery Conflation
    - Vector to Black-White Imagery Alignment [sstd'03]
    - Vector to Color (High-resolution) Imagery Alignment [stdbm'04] [GeoInformatica'05 ?]
  - Automatic Map to Imagery Conflation [ng2i'03a] [acm-gis'04] [Transactions in GIS ?]
  - Applications
    - Vector-imagery and map-imagery conflation web services
    - Building Finder: A System to Automatically Annotate Buildings in Satellite Imagery [ng2i'03b]

Future Work

- Improvements of AMS-Conflation
  - Vector-Imagery conflation: Devising an automatic approach to decide whether new training for road classification is needed for a given image
  - Map-Imagery conflation: Utilizing extracted text from maps
- Generalization of AMS-Conflation to deal with other geospatial datasets
  - Point to map conflation
  - Elevation data conflation

Future Work



- Point to map conflation: Integration of gepspatial point data with maps

apinumber	STATUS...	OPERATOR	LEASE	WELLNO_	SEC	map	RNG	TWN	LATITUDE	LONGITUDE
11102789	004	Wasibi Oil Co	Well No.	B	34	207	19W	05N	34.491765000000001	-118.856876
11106431	014	Arizona Oil Company	Well No.	1	13	207	20W	04N	34.431190999999998	-118.92027899999
11106042	006	Texaco E & P Inc	Standrad-Arundell	1	17	207	19W	04N	34.429599000000003	-118.885361
11106013	006	Jahn's Oil Company	Goodenough	1	18	207	19W	04N	34.42944	-118.918792
11106004	006	C. W. Colgrove	Bursin	46-18	18	207	19W	04N	34.424782999999998	-118.912588

- Elevation data conflation: Integration of low resolution elevation data (e.g., USGS DEM) with higher resolution elevation data (e.g., contour lines) by matching highest/lowest points

- Commercialization of AMS-Conflation
  - GeoSemble Tech.: http://www.geosemble.com/