Populating/Updating GIS: Automatically?

Toni Schenk
Department of Civil and Environmental Engineering and Geodetic Science
The Ohio State University
schenk.2@osu.edu

Motivation

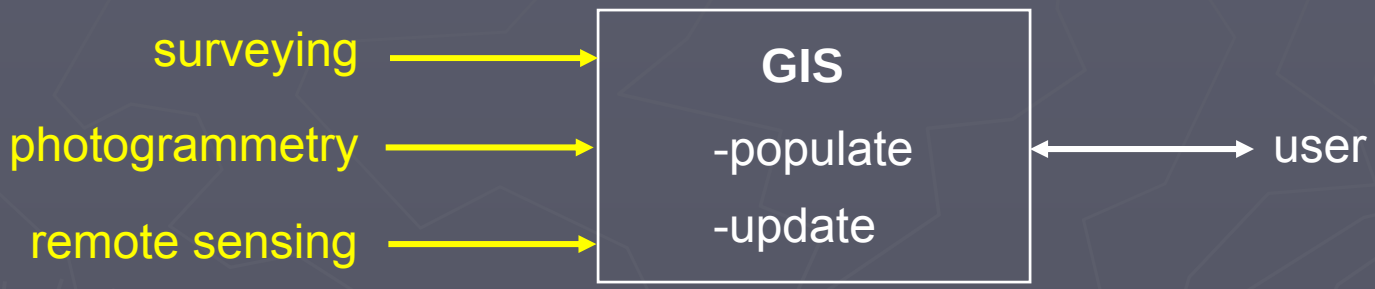
- Usefulness of GIS depends greatly on its information content. Information should be
  - rich (satisfy current and future needs)
  - accurate
  - up-to-date
- Keeping information up-to-date may be biggest challenge
  - how to detect changes
  - how to update changes

Motivation

- Populating and updating GIS manually is
  - slow
  - expensive
- Ever increasing amount of data poses challenge to extract GIS information
- Can the processes of creating and updating GIS be automated?

Your new GIS update is ready to be installed. Do you want to install it now?

GIS Players



data → **information** → **knowledge**

Geospatial Information providers

Geospatial Information managers

Data

- sensor output, result of data acquisition
- raw signals or preprocessed
- data is not an end product, nor does it directly provide answers to applications
- examples: pixels, laser points
- ➔ data needs to be (further) processed

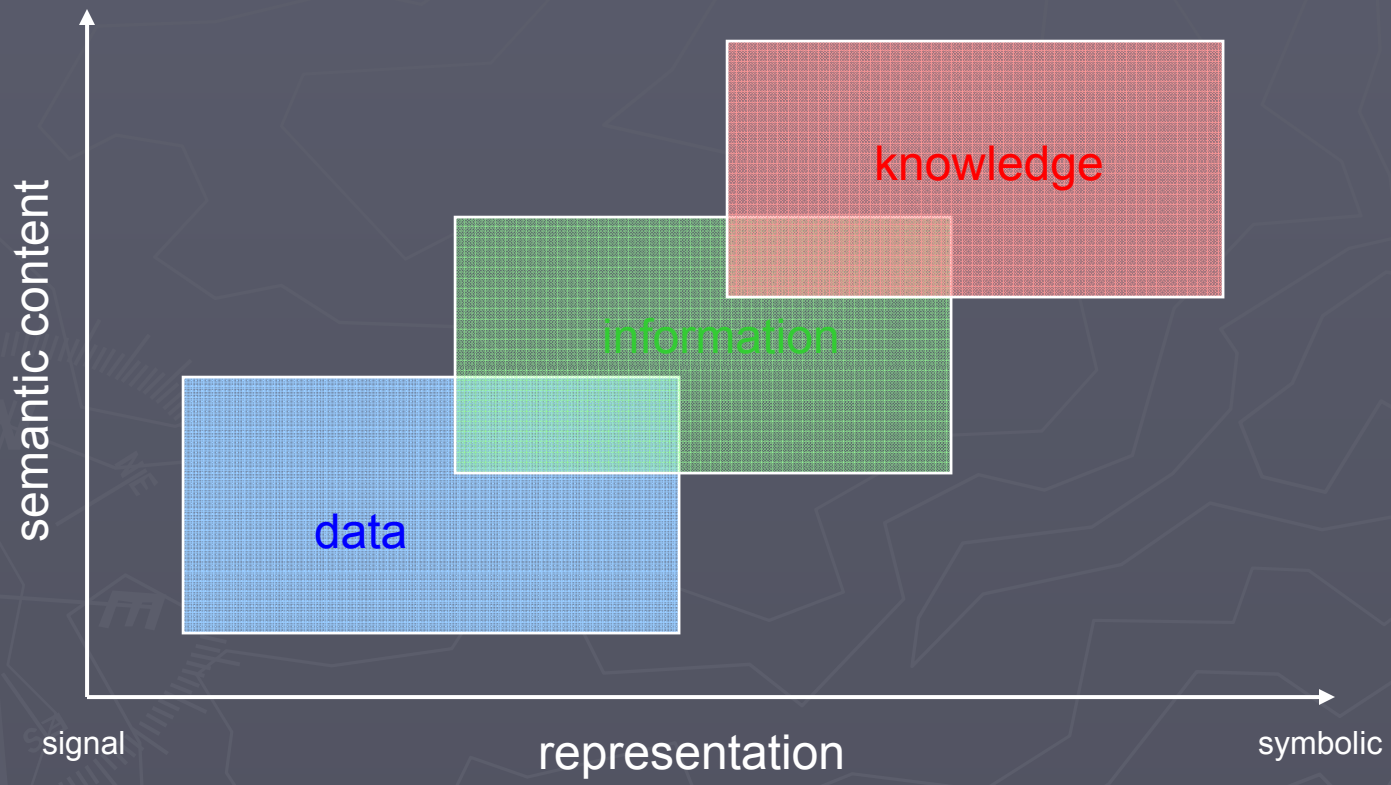
Information

- more specific than data to answer questions
- information can be retrieved from data → information implicitly in data
- extracting information from data makes it explicit → feature extraction
- features are information primitives

Knowledge

- elusive concept, ill-defined → means different things to different people
- to 'deal' with knowledge it must be represented → knowledge and its representation are closely related
- knowledge: facts, procedures, heuristics that can be used to make inferences

Data, Information, Knowledge



Data, Information and GIS

- GIS contains explicit geospatial information about objects, e.g. a polygon is labeled with ‘building’, the vertices may be considered data
- The object is extracted from raw data, e.g. images
- Volume of raw data >> volume of GIS information

Example

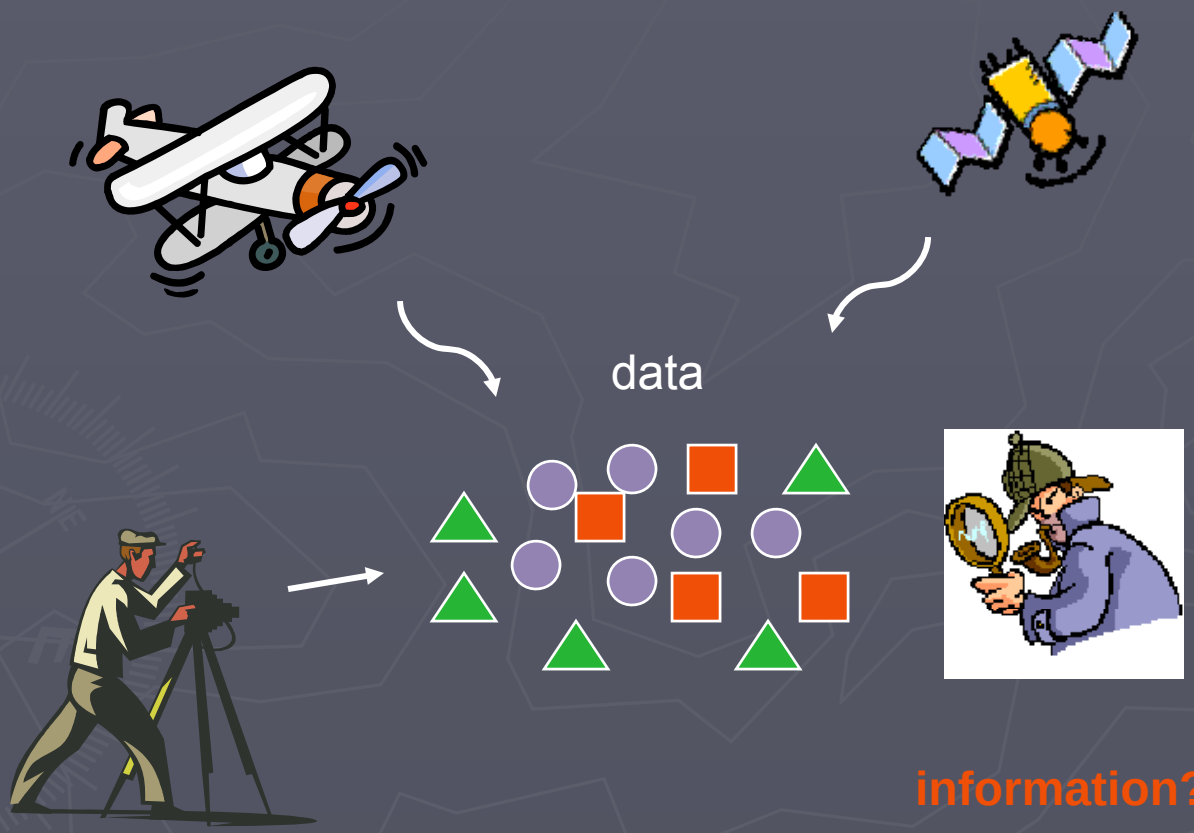
- GIS of a rural county:
  - 10,000 centerline segments
  - 80,000 parcels
  - 13,000 sewer line segments
  - 130,000 spot elevation points
  - 12,000 water line segments

180 MB

- Input data: aerial images
  - GSD 0.1 m
  - County 25 x 25 miles ➔0.16 10e+12 pixels ➔ 3 x 2.5 x 0.16 10e+12 MB

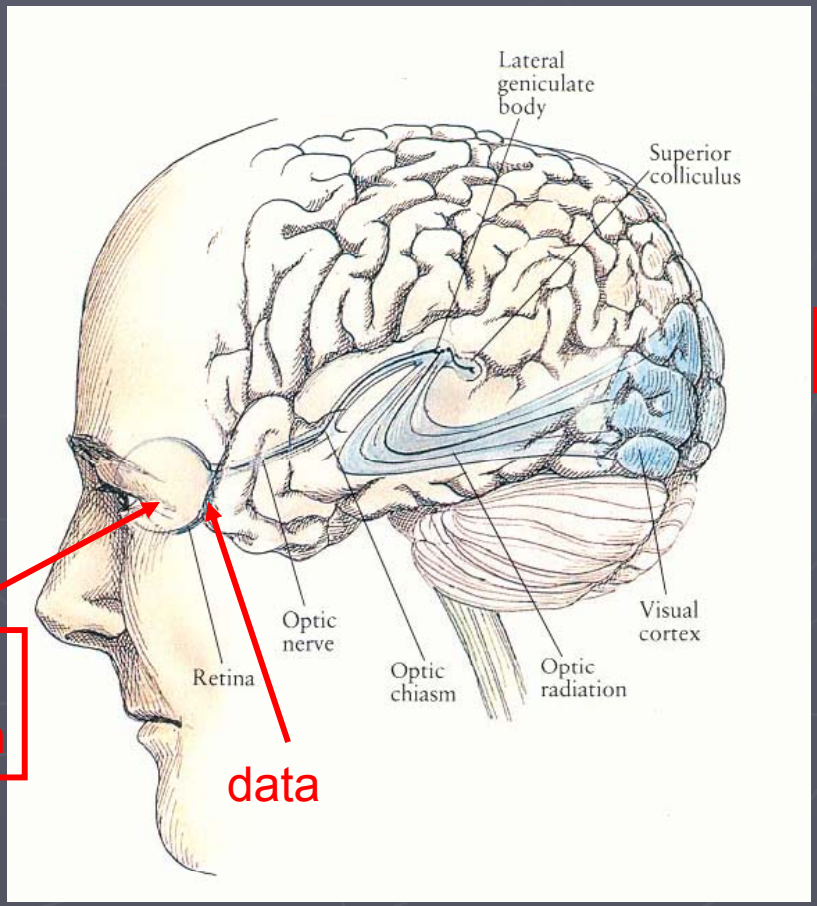
1.2 TB

Data Deluge



information?

From Data to Information and Knowledge in Human Vision



scene

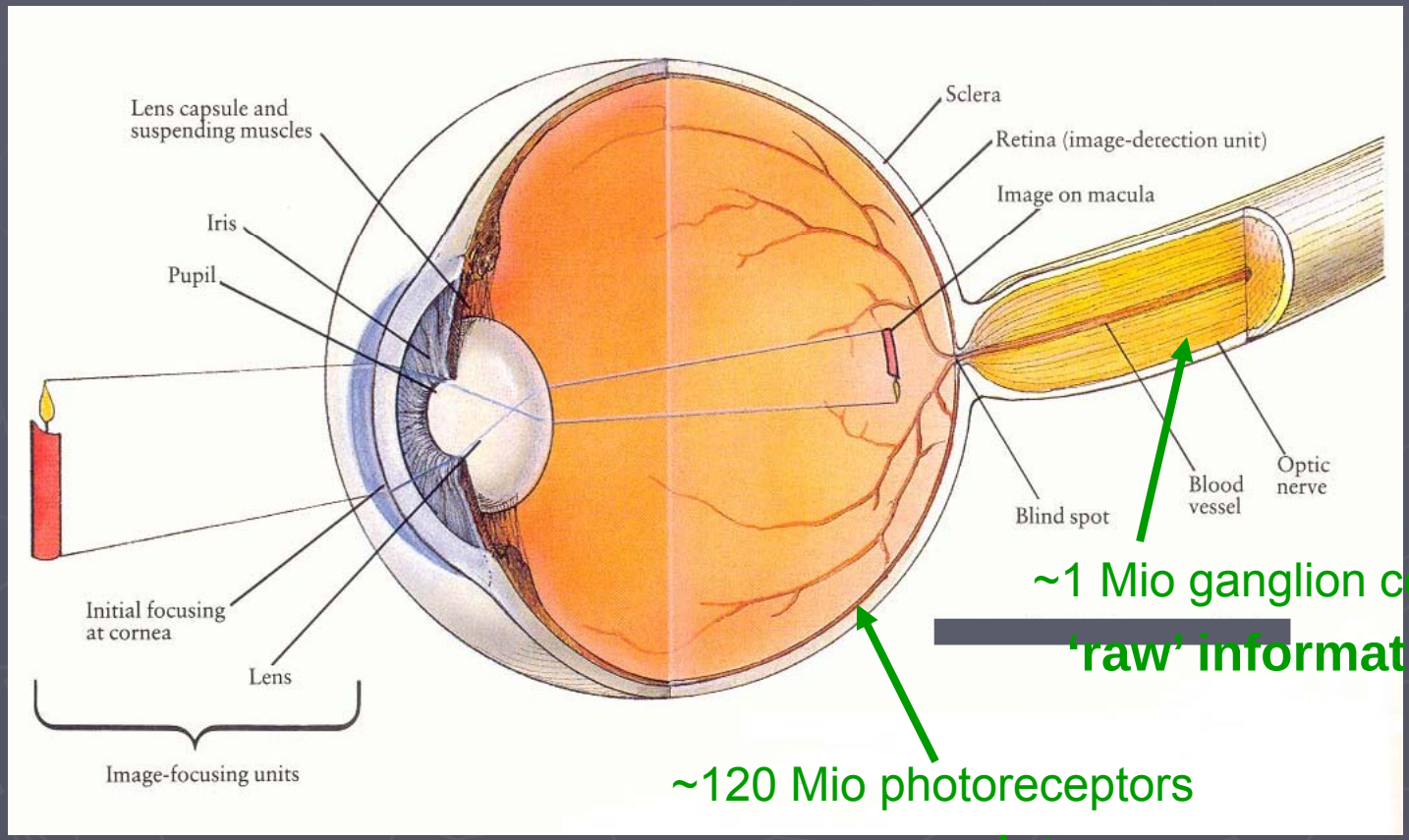
data acquisition

knowledge

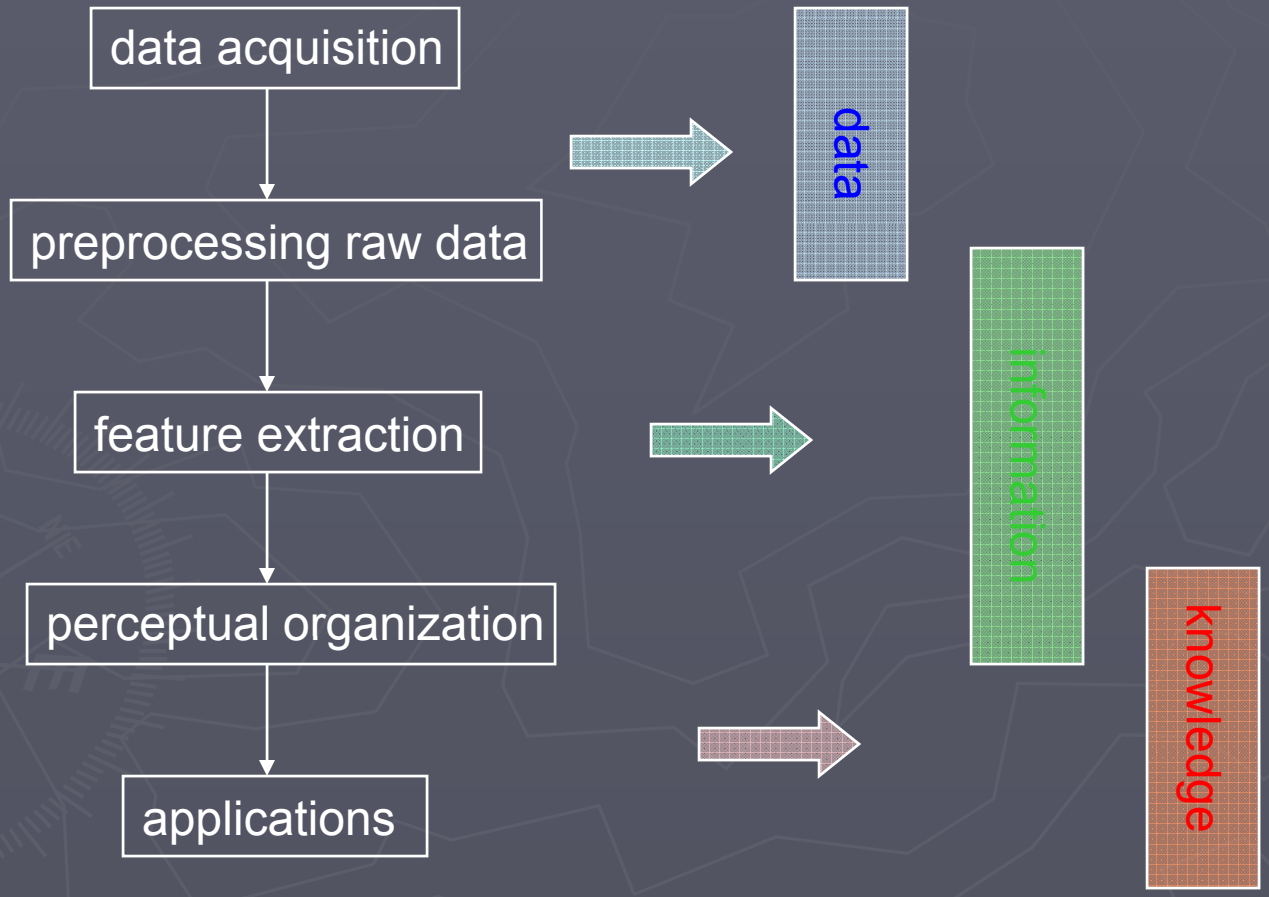
perception

information

Image Formation on Retina



Paradigm

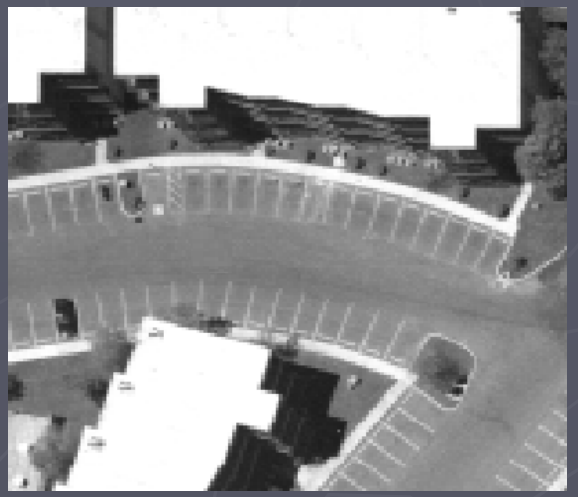

data acquisition
preprocessing raw data
feature extraction
perceptual organization
applications
data
information
knowledge

Representational Issue

image ‘seen’ by computer

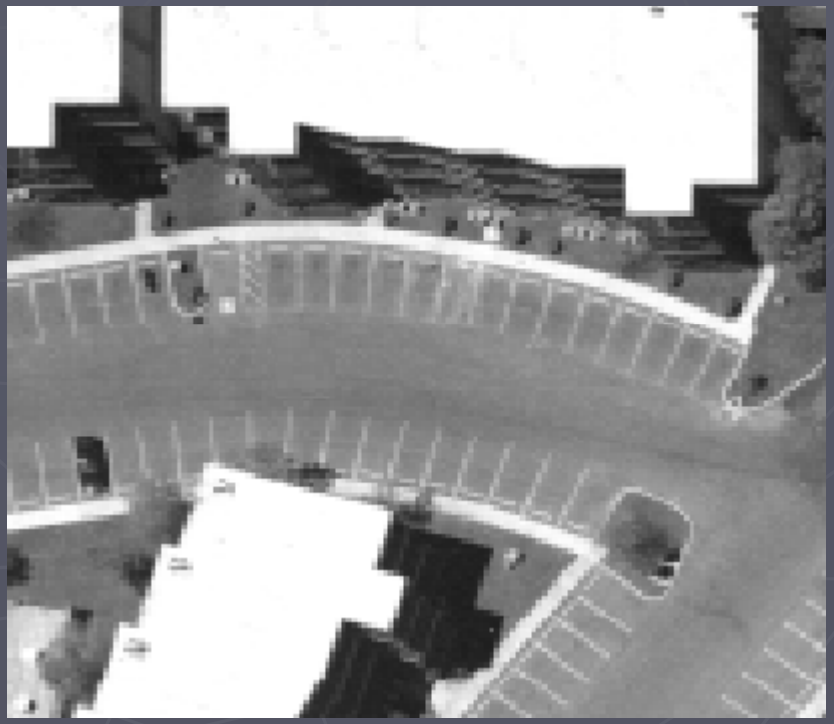
11	12	18	15	25	23	26	35	39
13	14	13	16	27	28	23	29	35
10	11	14	26	29	24	25	31	33
8	10	15	27	29	28	24	33	43
7	11	16	21	31	29	24	38	41
17	13	15	25	28	31	34	39	37
12	14	16	22	25	30	35	36	38
13	15	17	23	27	29	28	29	31
15	16	12	24	29	22	25	26	36
12	14	20	19	25	26	28	31	40

image seen by human



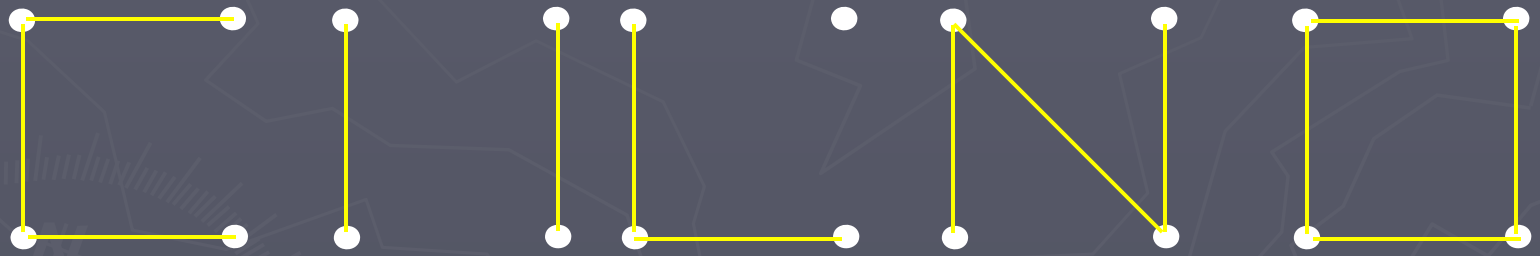
Spatial Extent: The Peephole

- 3 x 3



Connect-the-dots Metaphor

connect-the-dots
algorithm

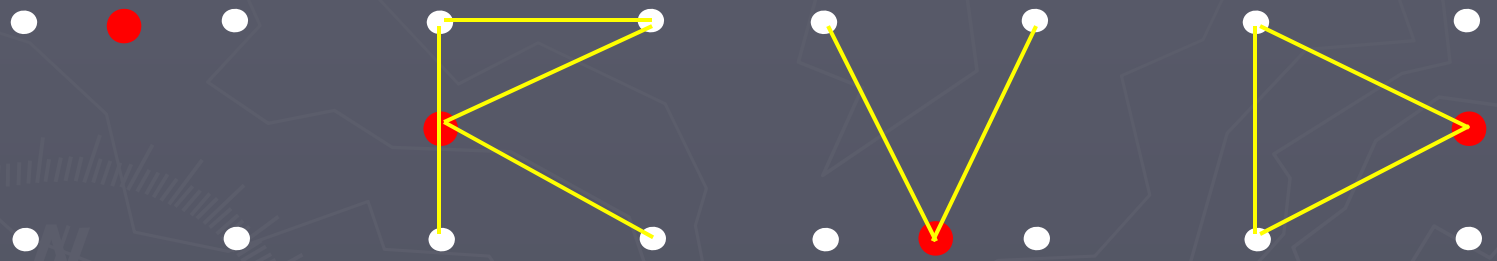


Meaning of structure?

‘C’ ‘I’ ‘I’ ‘L’ ‘N’ ‘O’

‘Intelligent Data Acquisition’

add one new point: where should it be added to yield maximum information



information gain

0 4-6 1 1?

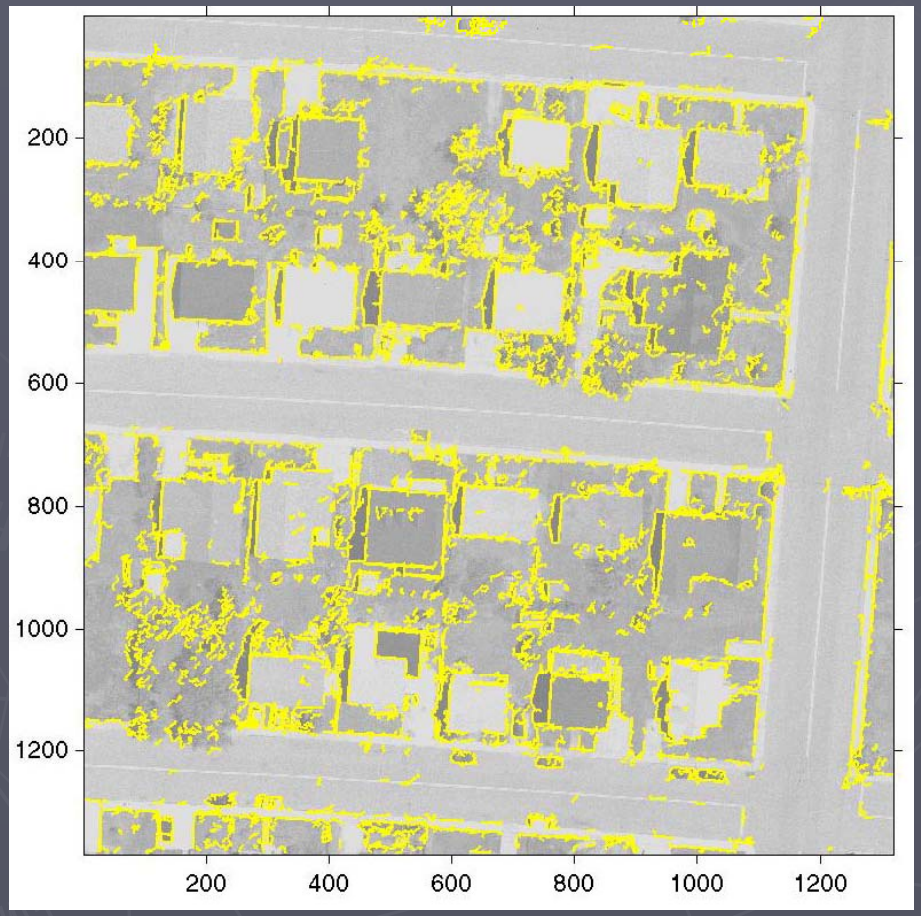
Example Image

Ocean City, 6146

1368 x 1320 pixels

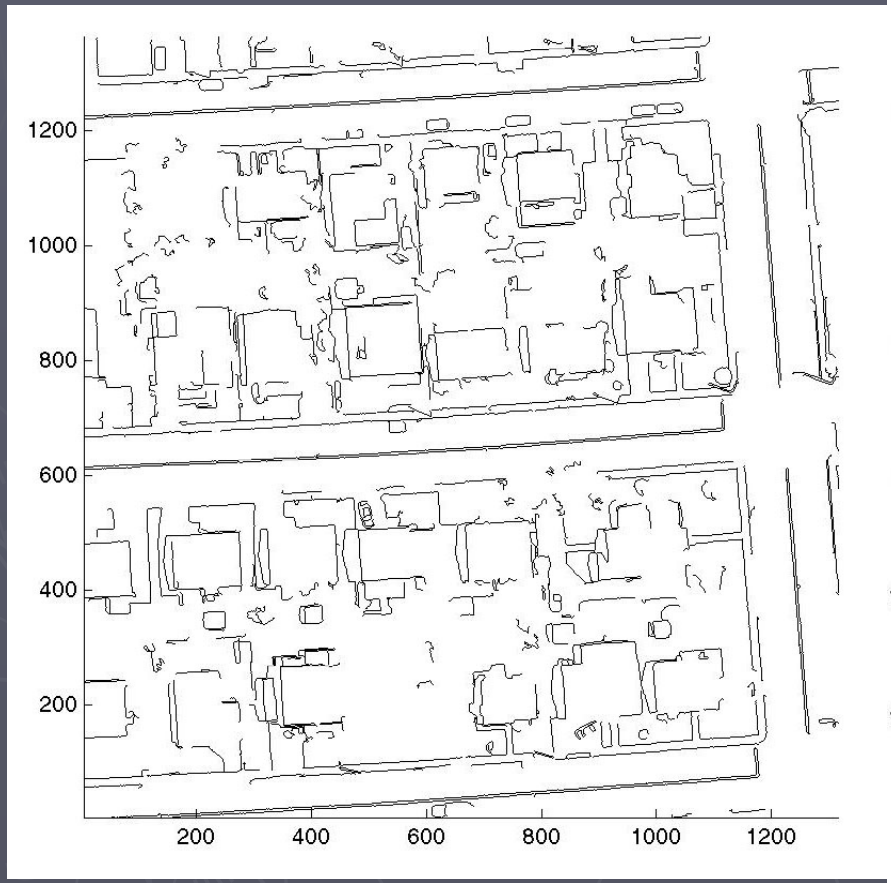
pixelsize ~ 15 micron

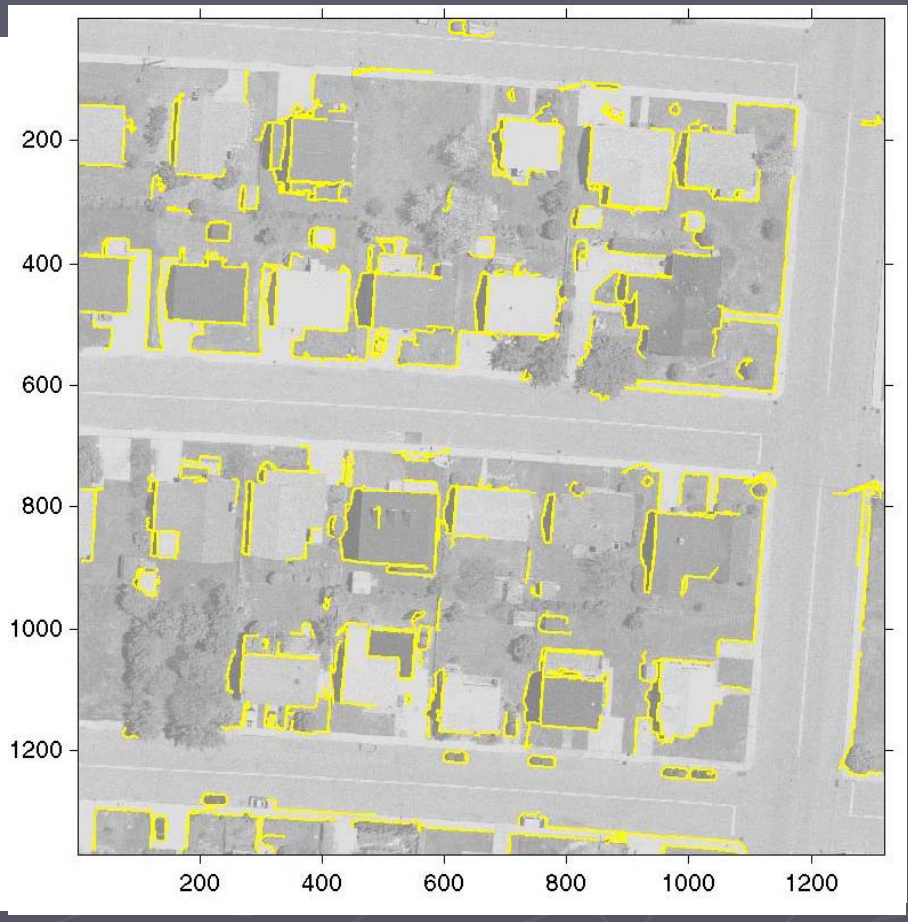
Examples of Edge Images



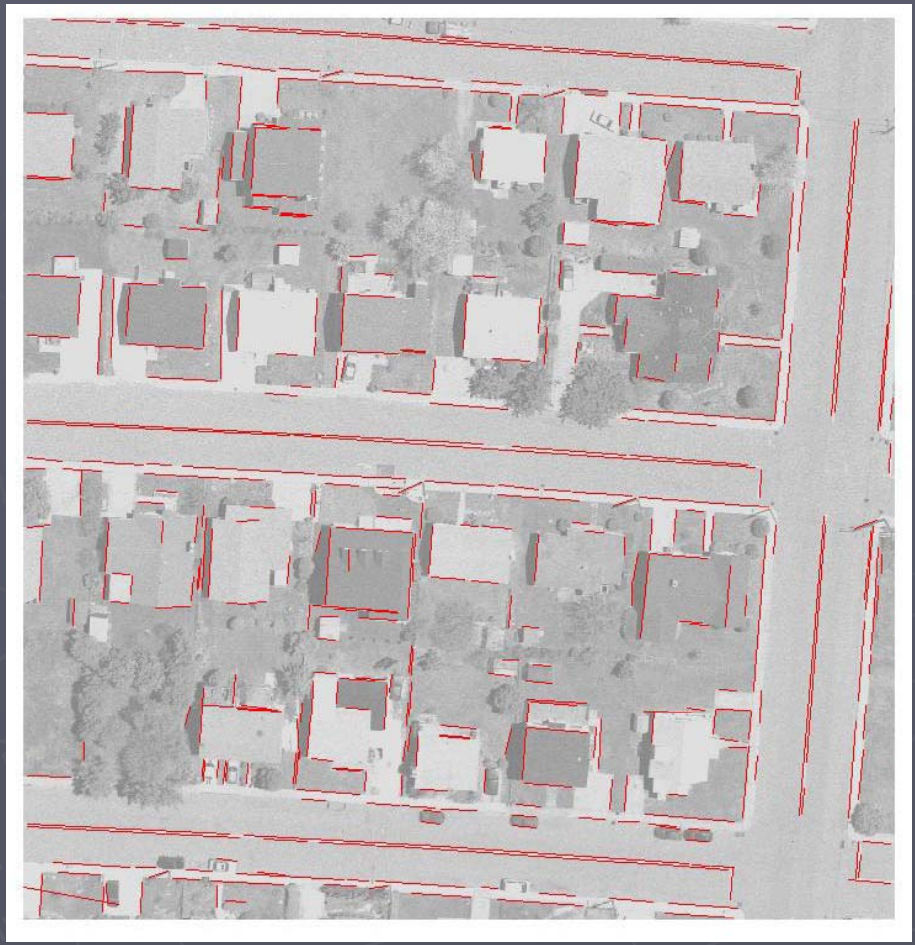
Canny edge detector

Examples of Edge Images



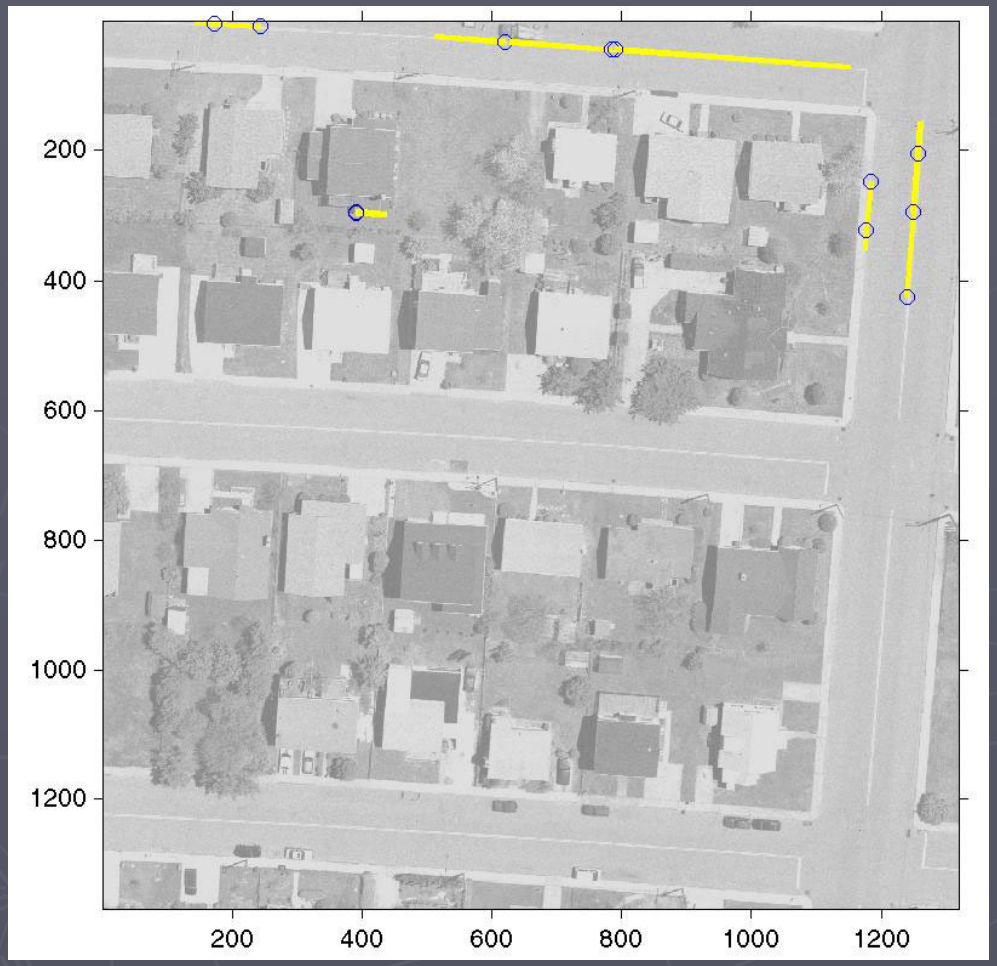


Segmented Edges



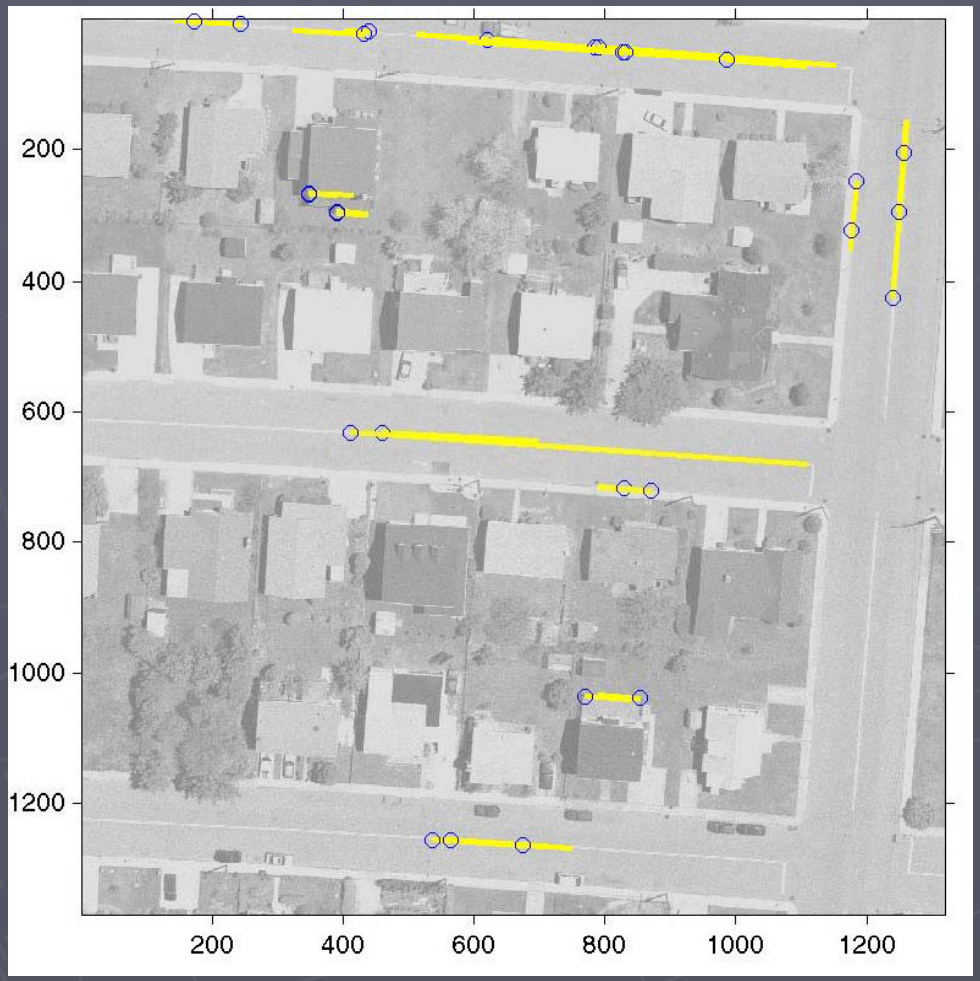
straight line segments
length > 25 pixels

Connecting Edge Segments



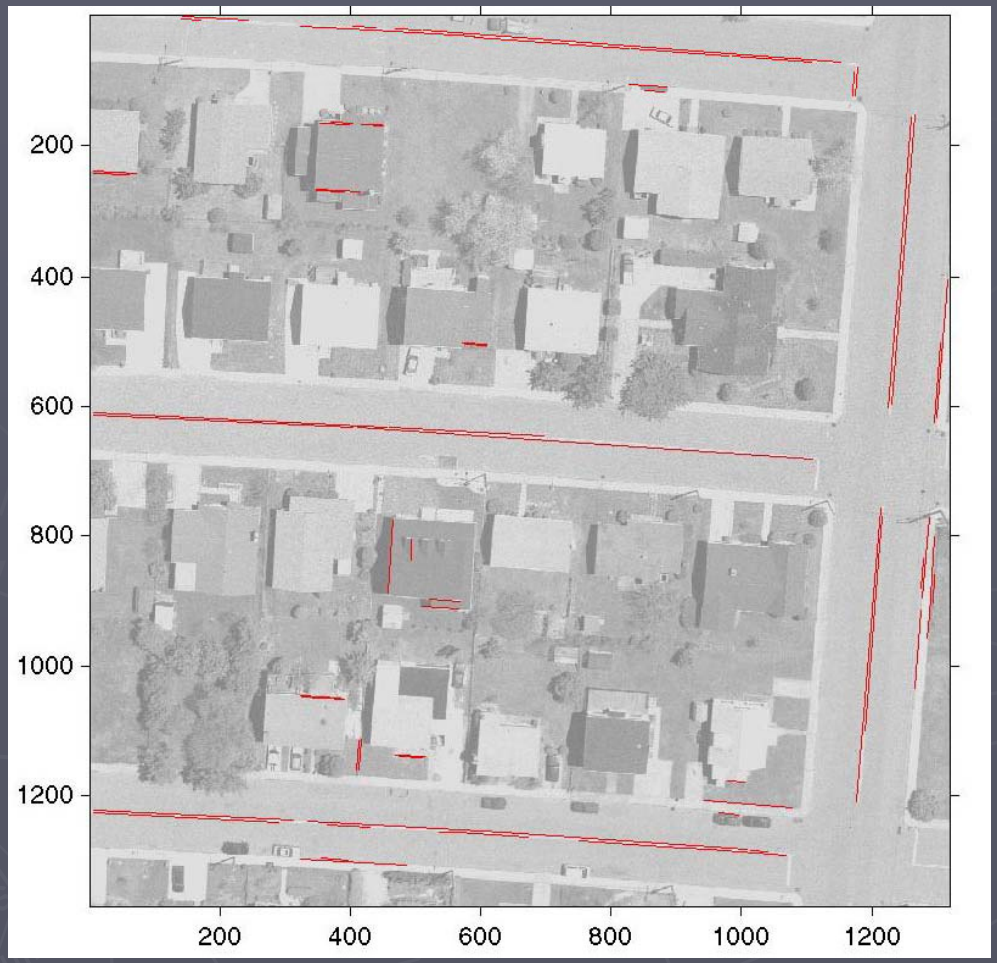
connects edge segments with same line properties and small gap

Connecting Edge Segments



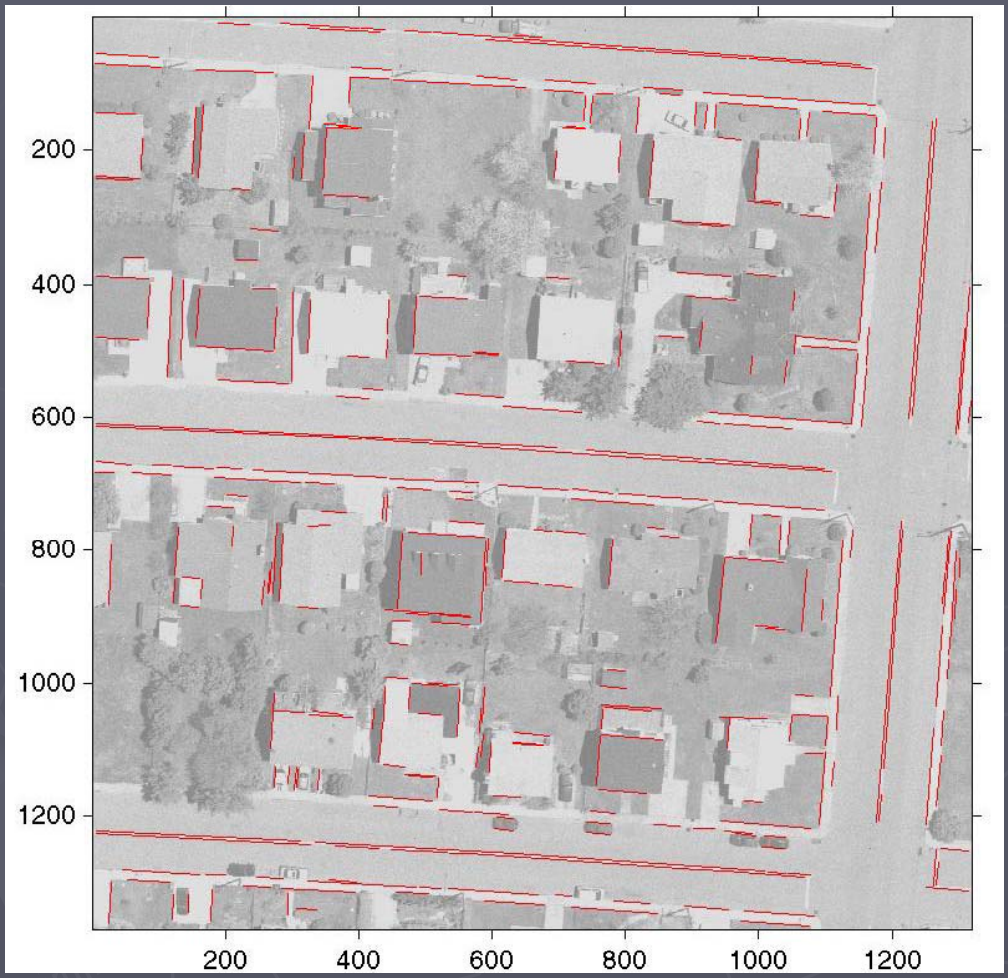
connects edges
with larger gaps

Parallel Edges



distance between parallel lines < 10 pixels

Parallel Edges

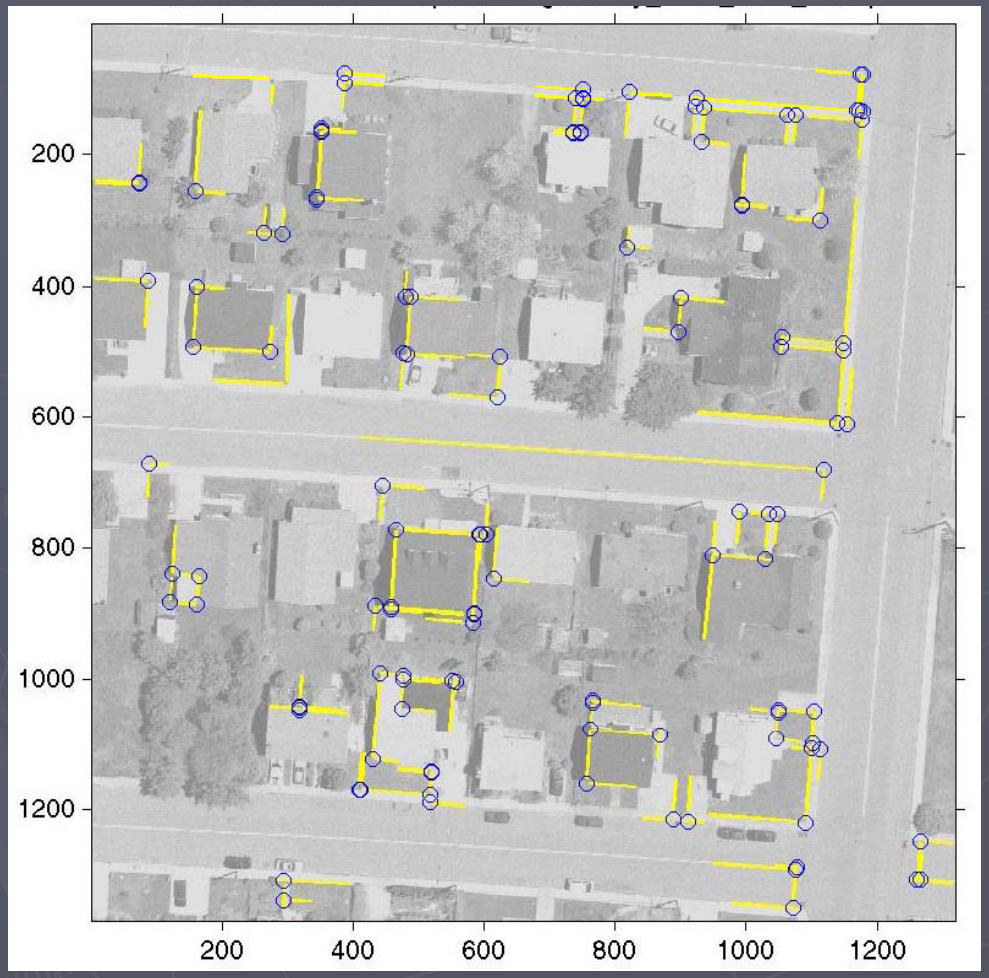


distance between
parallel edges < 200 pixels

Perpendicular Edges

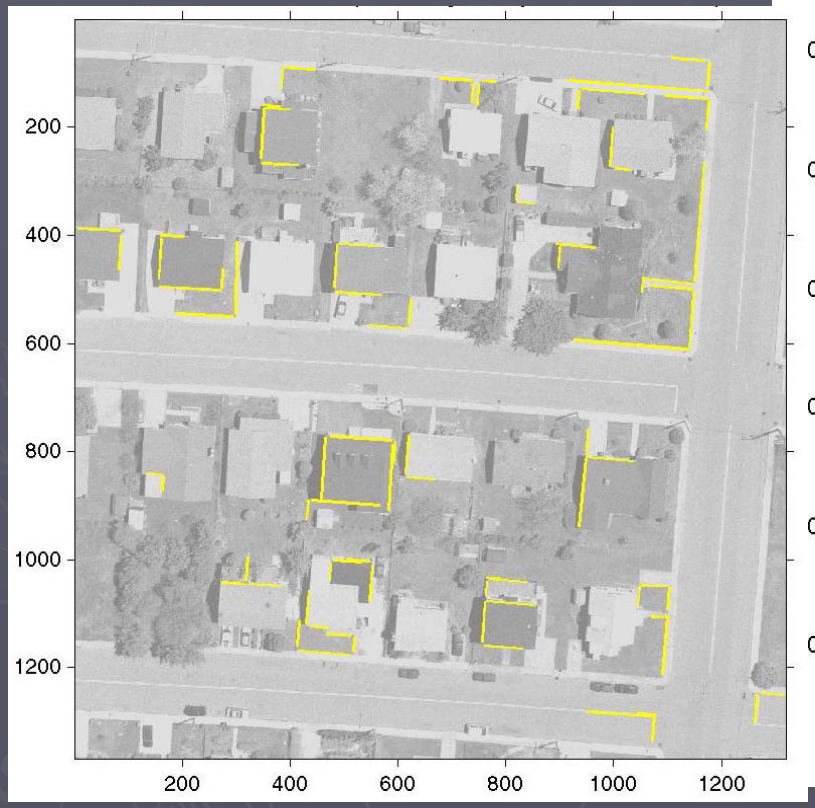
edges are grouped to perpendicular pairs, considering proximity

Computed Corner Points

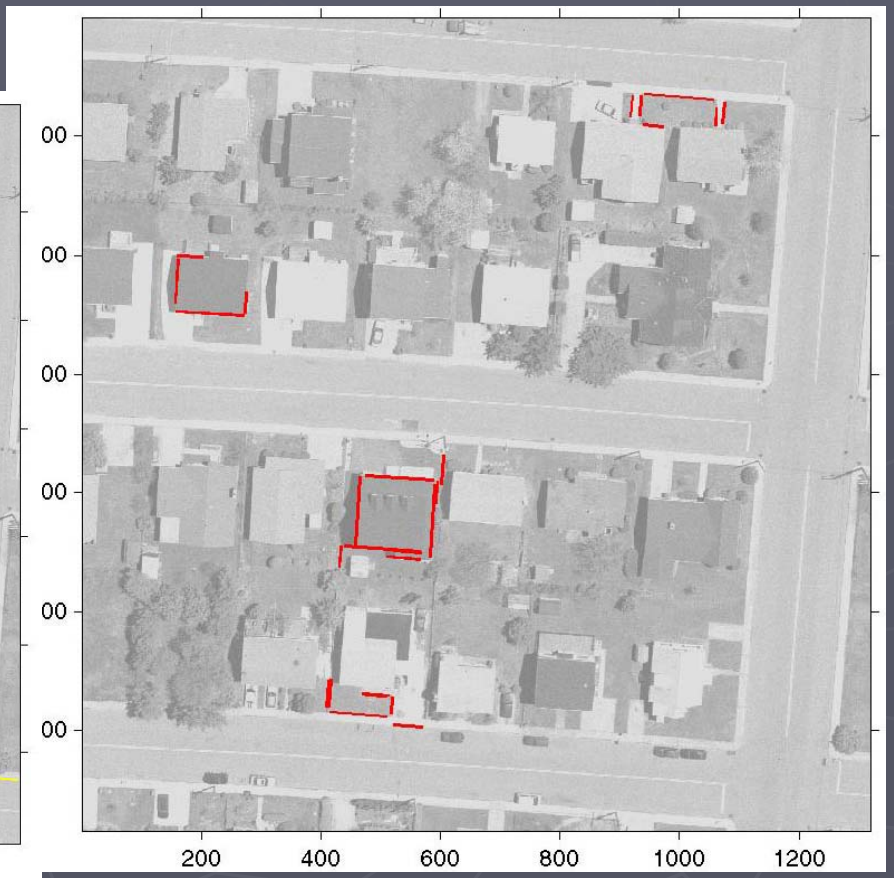


corners are computed
from perpendicular edges

Clustering Perpendicular Edges

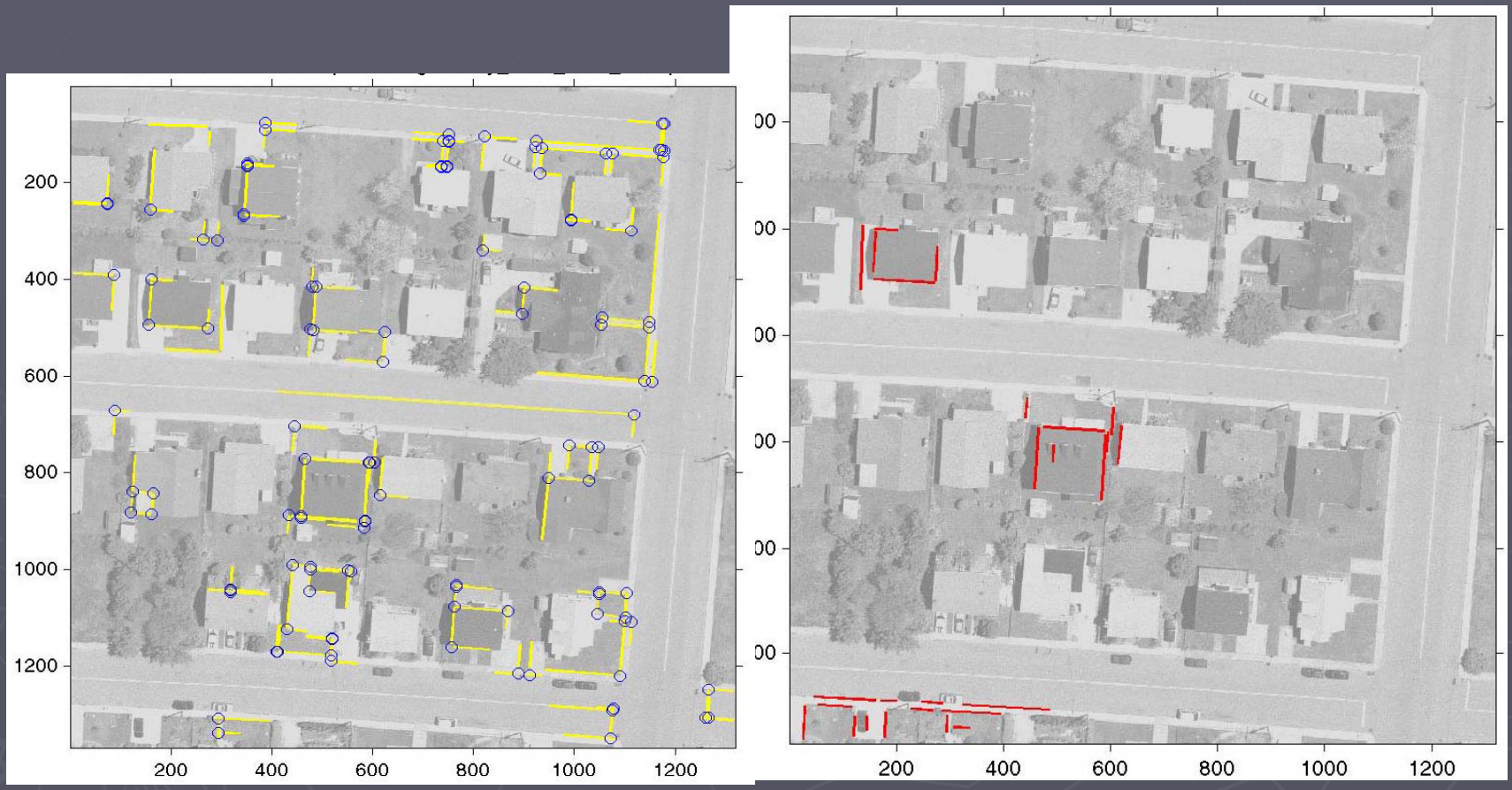


all perpendicular edges



example of 4 clusters

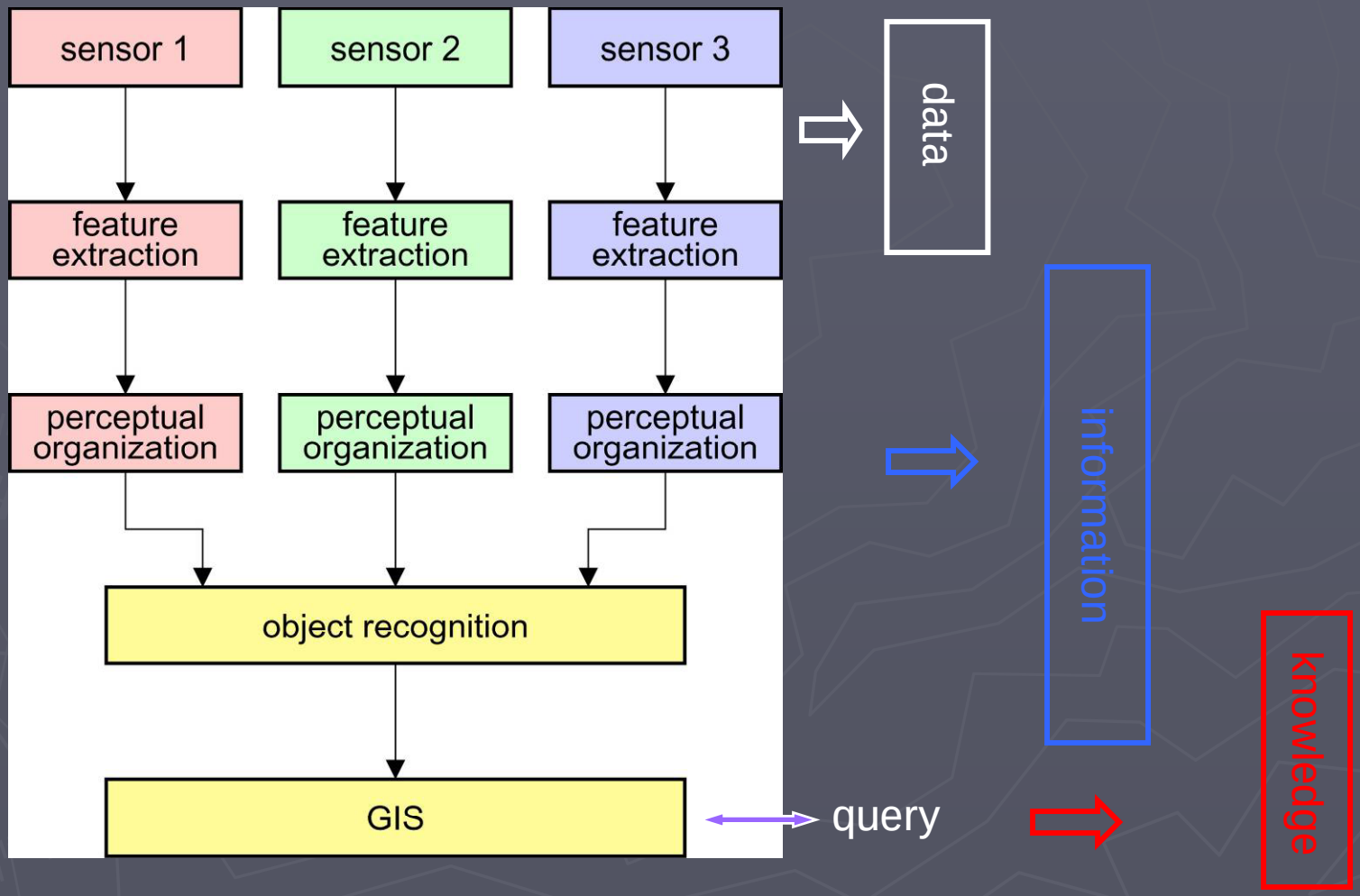
Clustering Perpendicular Edges



all perpendicular edges

example of 3 clusters

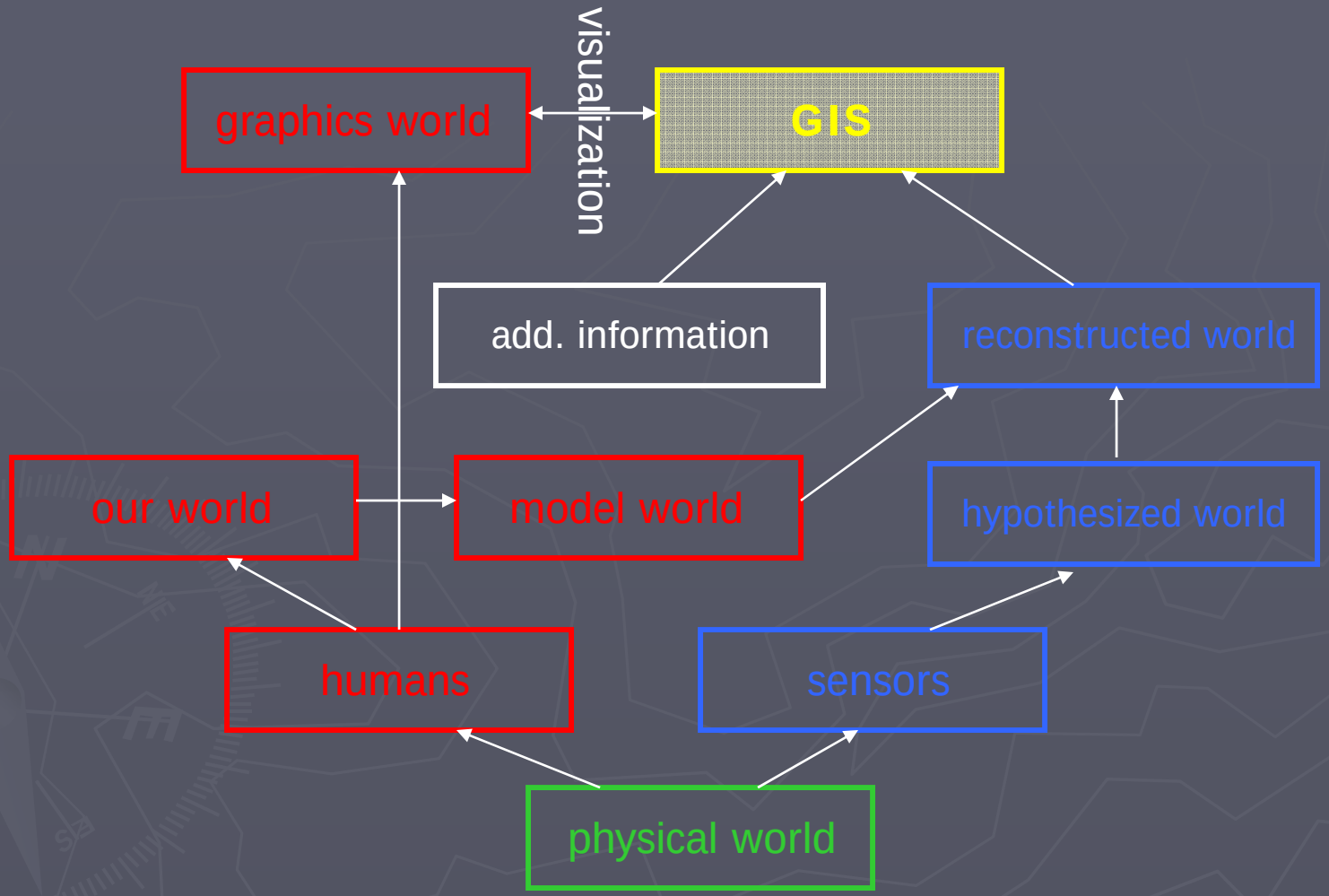
From Data to Information and Knowledge in a GIS Environment



Information and Knowledge Generating Processes

processes	results
data acquisition	data
feature extraction	information primitives
perceptual organization	information chunks
object recognition, fusion (GIS generation)	geometric and semantic information
GIS queries	knowledge

A World of Models



Concluding Remarks

- data < information < knowledge
- bottleneck in generating and updating GIS lies in information extraction from data and NOT in data collection
- end products consist of information + knowledge
- from data to information by extracting features and their perceptual organization

Concluding Remarks

- Over lifetime of GIS:
  - cost of updating >> original creation
- Automate update
- Aspects of updating
  - Detect changes
  - Implement changes

Concluding Remarks

- Additional updating considerations
  - Update attributes of objects
  - Precision
  - Upgrading from 2-D to 3-D
  - Missing information

thank you…