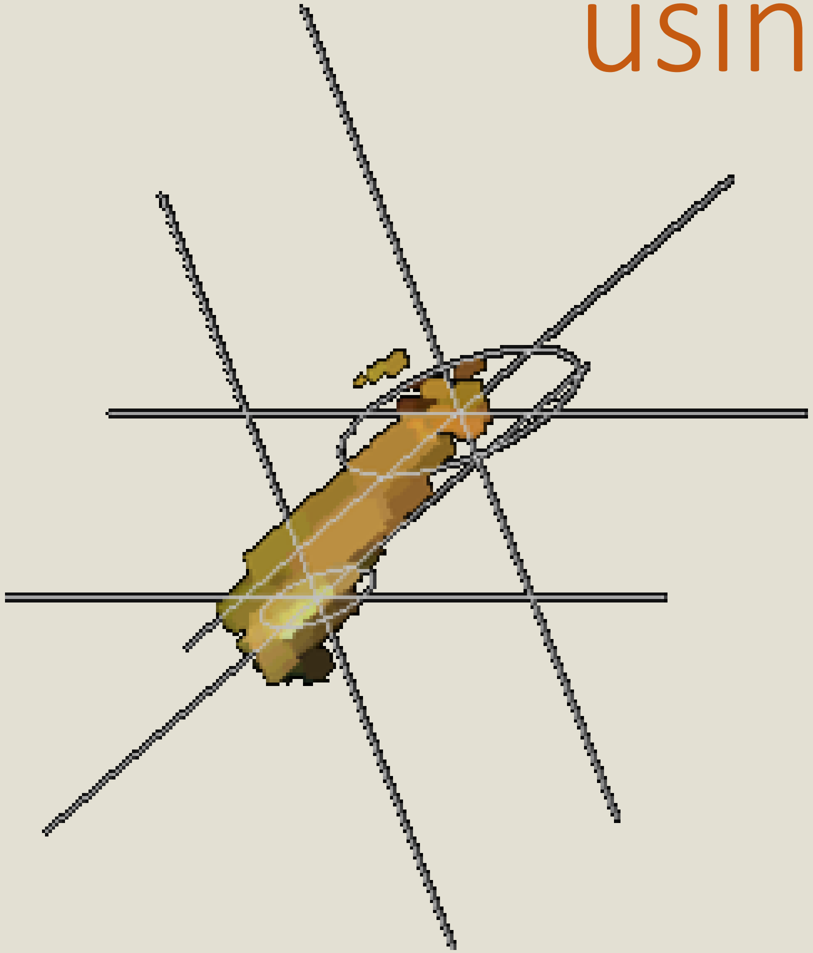
Representing Navigation Landmarks using Terrain Spatiograms



Damian M. Lyons

Robotics & Computer Vision Lab

Fordham University, NY USA

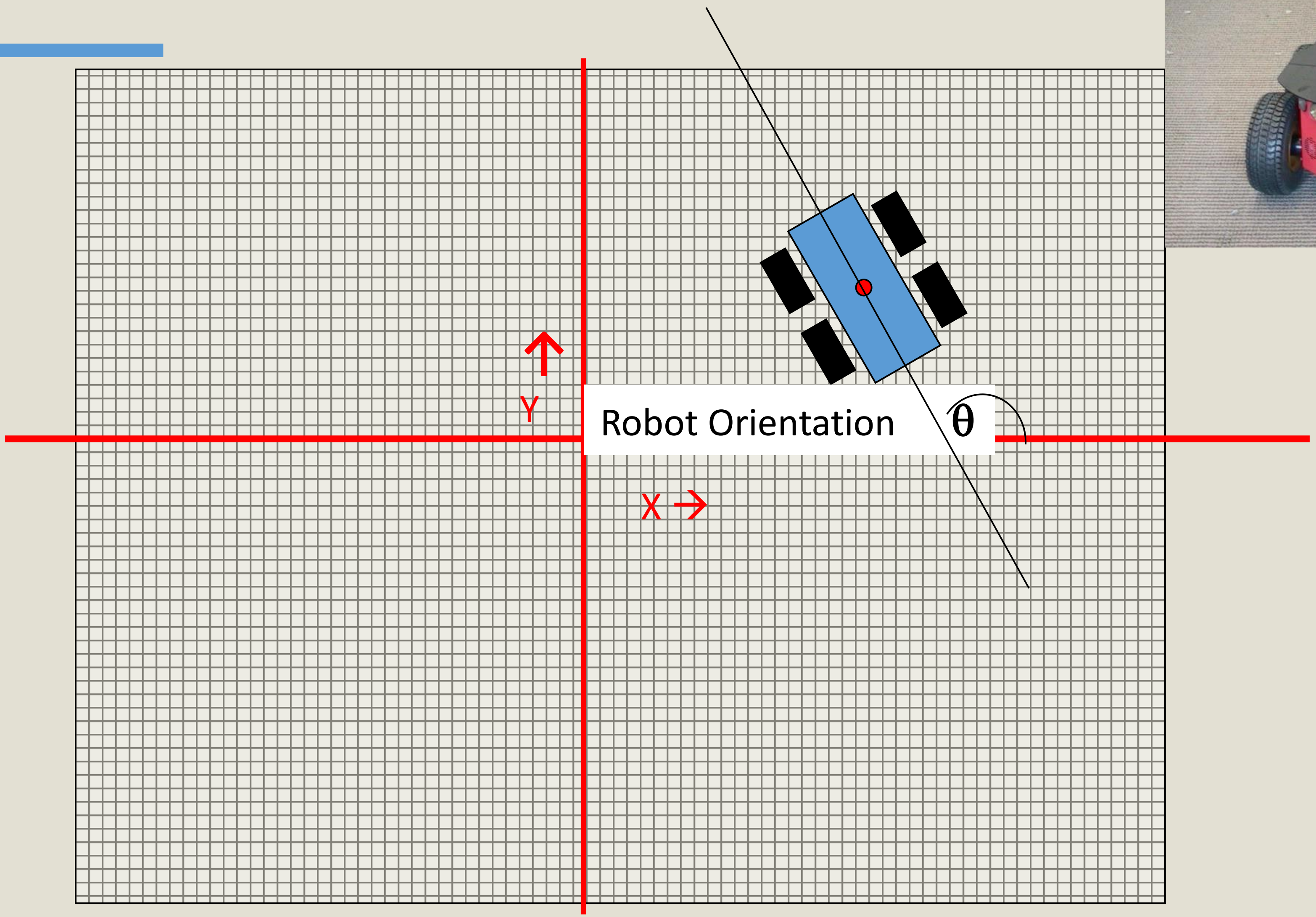
Overview

- Robots and Sensors, especially RGB-D Sensors
- The wayfinding problem for autonomous robots, and the role of landmarks
- Approaches to representing landmarks
- Image Histograms and Terrain Spatiograms
- Handling occluded landmarks
- Automatically selecting landmarks
- Comparing the performance of terrain spatiograms with some other approaches.
- Conclusion

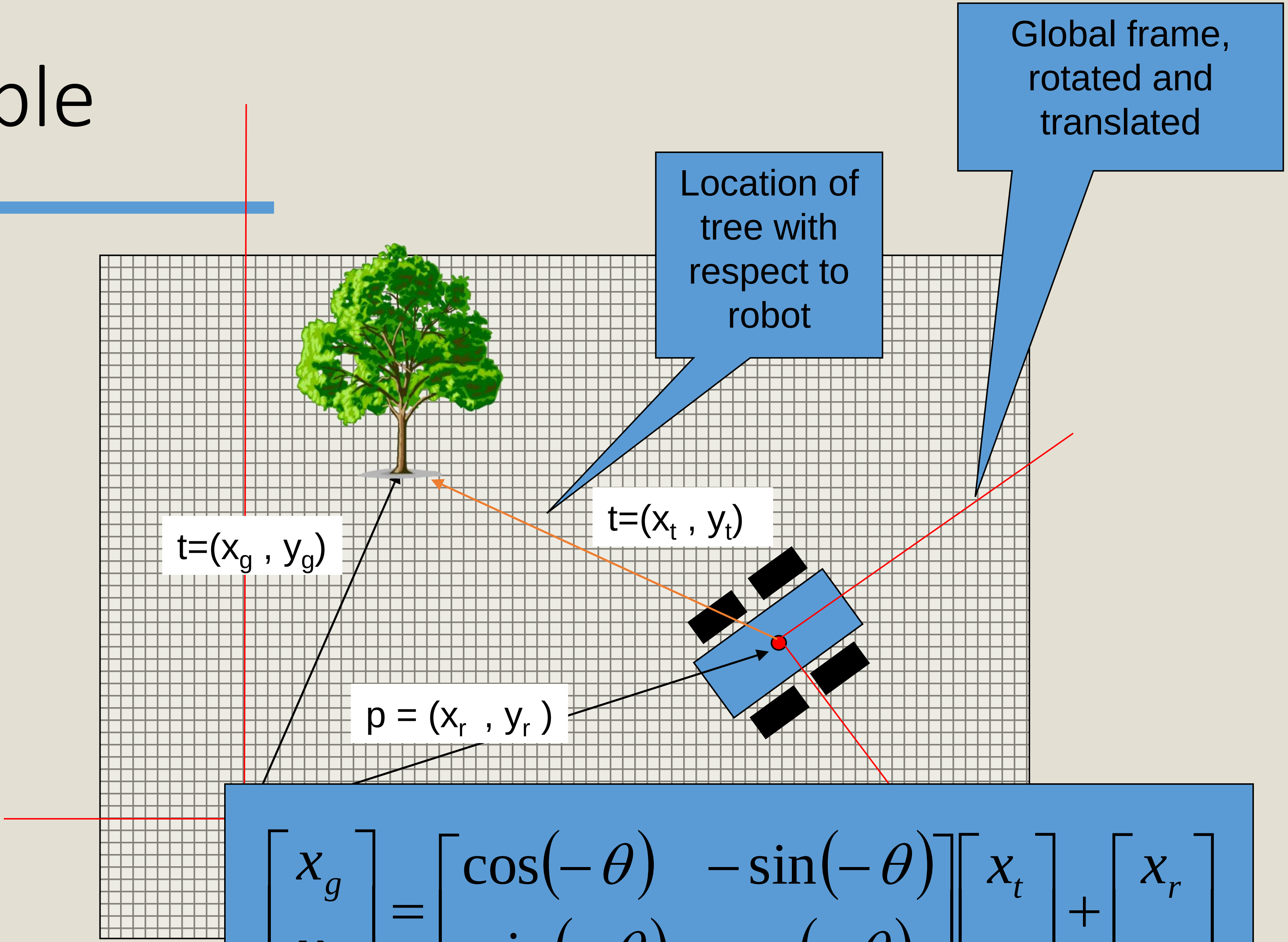
Overview

- Robots and Sensors, especially RGB-D Sensors
- The wayfinding problem for autonomous robots, and the role of landmarks
- Approaches to representing landmarks
- Image Histograms and Terrain Spatiograms
- Handling occluded landmarks
- Automatically selecting landmarks
- Comparing the performance of terrain spatiograms with some other approaches.
- Conclusions

The State of 2D Robot (x,y,θ)



Example

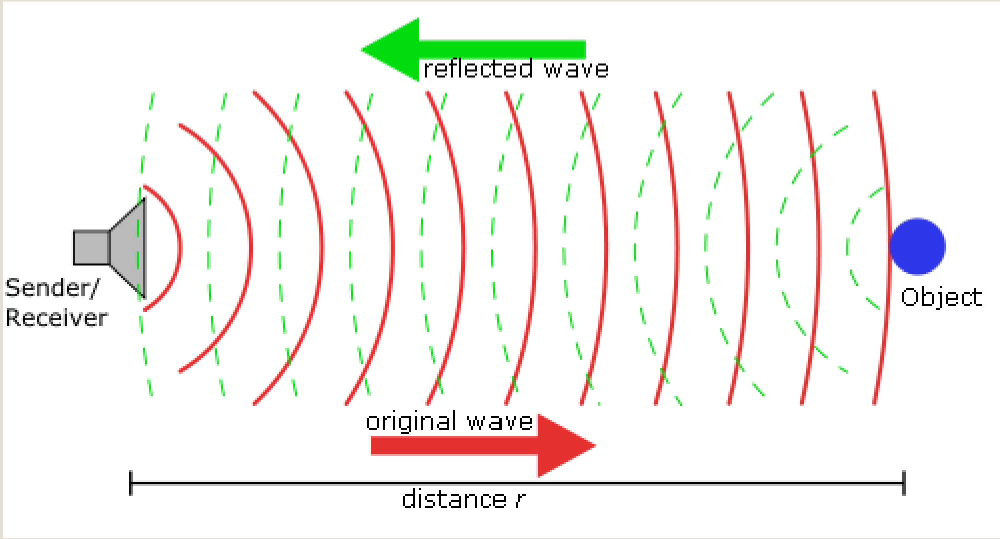


$$\begin{bmatrix} x_g \\ y_g \end{bmatrix} = \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} + \begin{bmatrix} x_r \\ y_r \end{bmatrix}$$

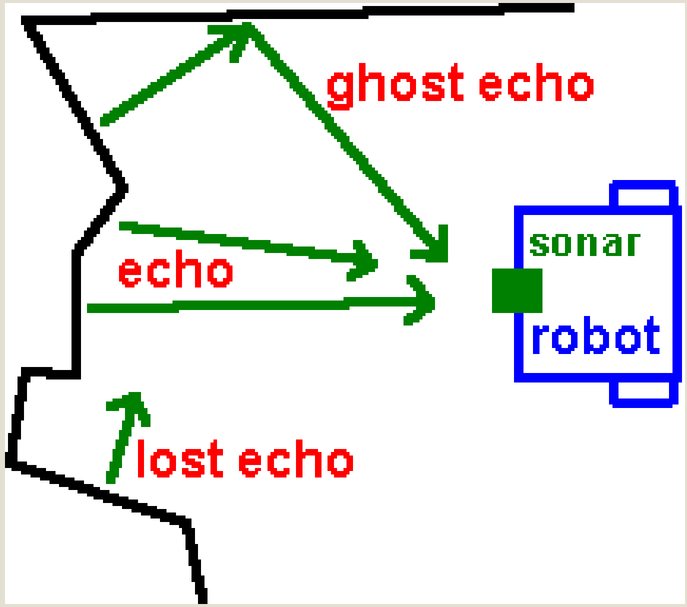
Sensors: Sonar

- Sonar: **SO**und **N**avigation **A**nd **R**anging

Sonar is a method of finding the distance to an object by measuring the time it takes for a pulse of sound (usually ultrasound) to make the round trip back to the transmitter after bouncing off the object (Time of Flight Measurement - TOF).
At sea level, in air, sound travels at about 344 metres per second (1130 feet per second). In practical terms this means 2.5 cm is covered in about 74 microseconds.



Sonar Issues



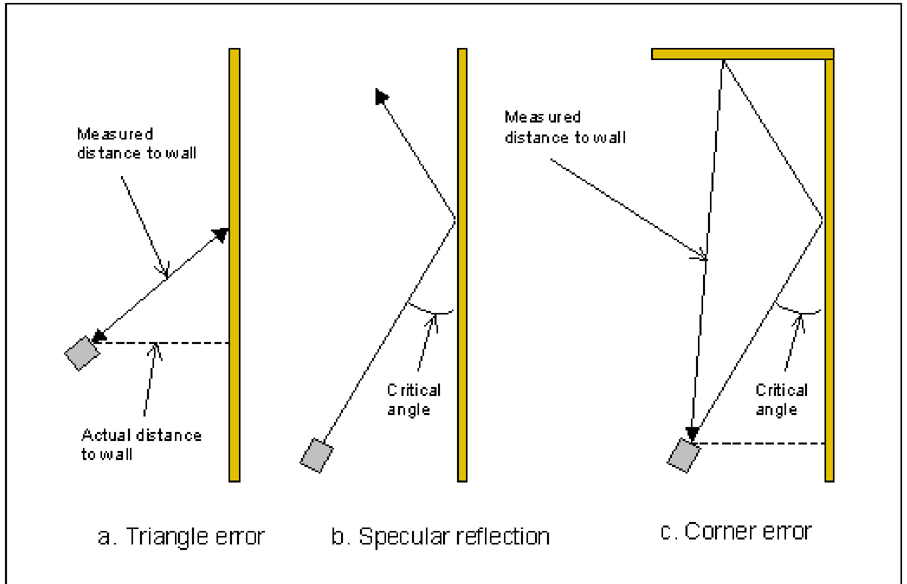


Fig 2. Sources of error

Sonar works best when the sensor is parallel to the target.

Sensors: Vision

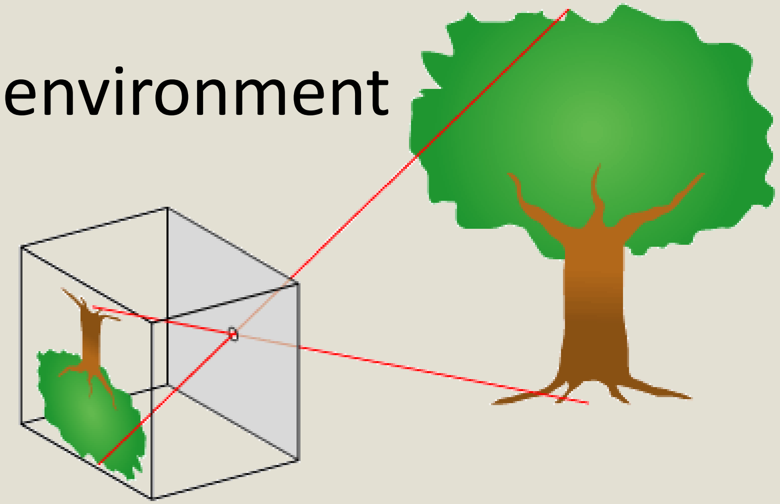
- Vision is a passive approach to sensing
- In theory, provides a lot of information about the environment
- In practice, can be difficult to interpret

Digital camera

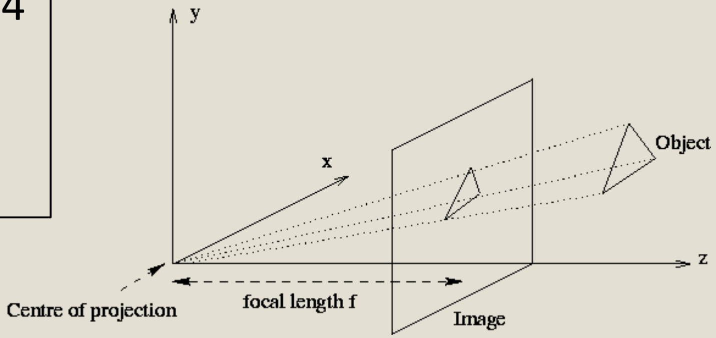
1/3” Sony CCD Camera
(analog camera)

Video Frame Grabber

Produces images of size
640 x 480 x 3 bytes
at video frame rate



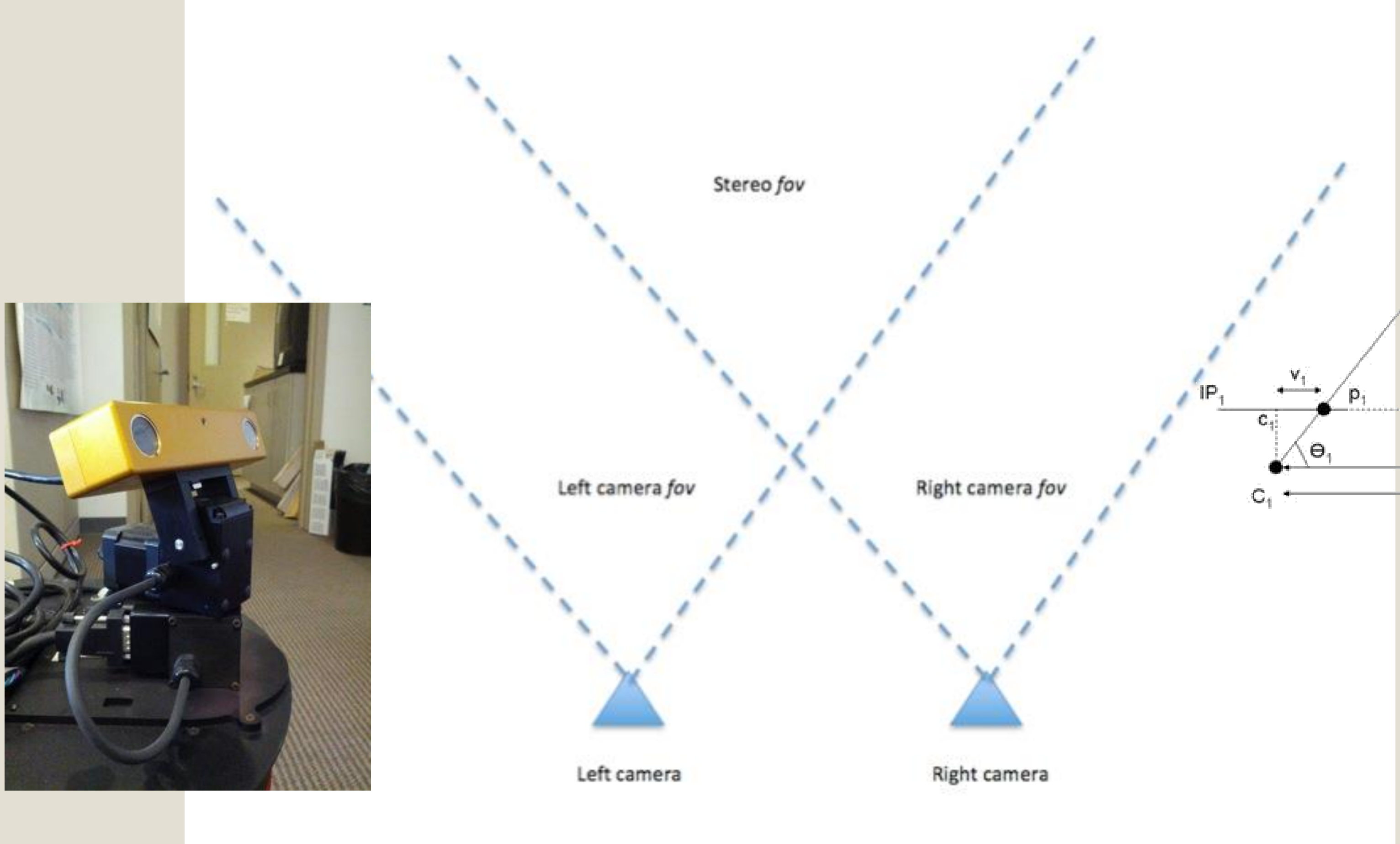
	0	1	2	3	4
0	(r,g,b)	(r,g,b)	(r,g,b)	……	
1	(r,g,b)	(r,g,b)	…….		
……….					

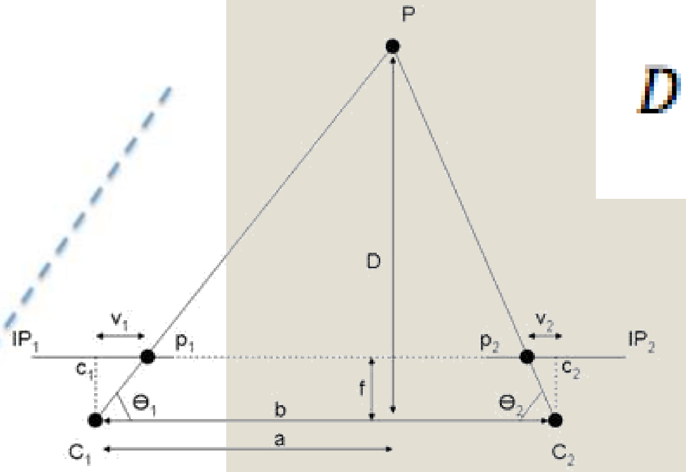


Issues With Vision

- Each point on the image (u,v) corresponds to a point in the scene (x,y,z)
- But that requires mapping 2D to 3D .. which is an <u>underconstrained</u> mapping; information is lost.
- Lighting changes dramatically alter an image (but are not changes in the elements of the scene)
- Objects may be occluded, hard to recognize, hard to separate from other objects, etc.

Sensors: Stereovision - Image(RGB) + Depth!





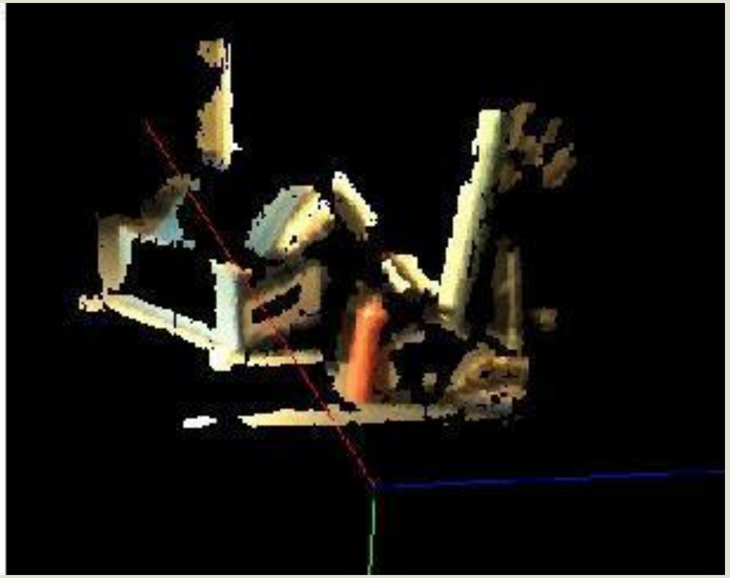
$$D = \frac{f \times b}{d}$$

Color Image I[u][v]=(r,g,b)
Depth Image D[u][v]=d
u=0..639, v=0..479

Example Stereo Information

(a)
Stereo camera image

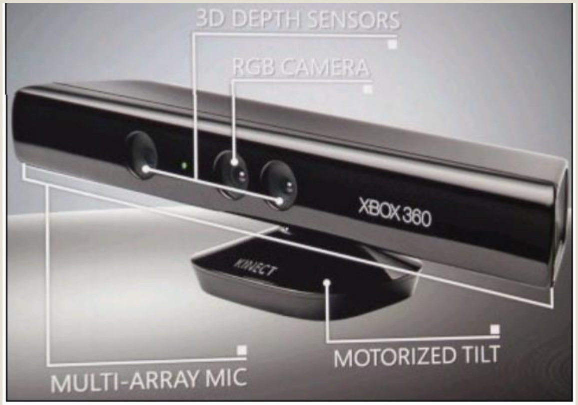
(b)
Pseudocolor disparity image



(c)
Point Cloud Image

A point cloud is a set of (x,y,z) points
that may include color information (r,g,b)

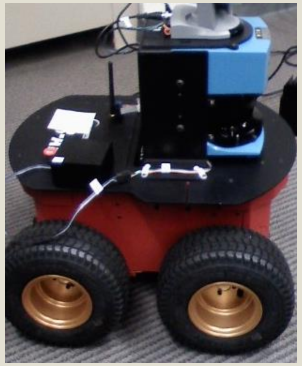
Sensors: Kinect RGB-D sensing



Other RGB-D sensing methods?

- Many similar sensors:
  - Swiss Ranger SR4000,
  - Asus Xtion PRO,
  - PMD CamCube,
  - Softkinectic Depthsense

- Could use Camera + Distance sensor combinations,
  - e.g. Camera + Laser Ranger

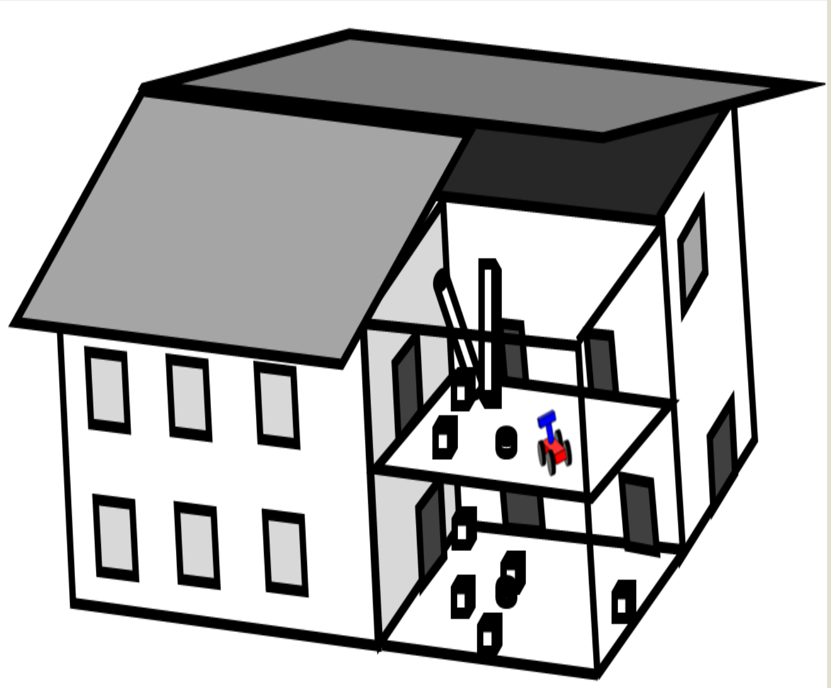


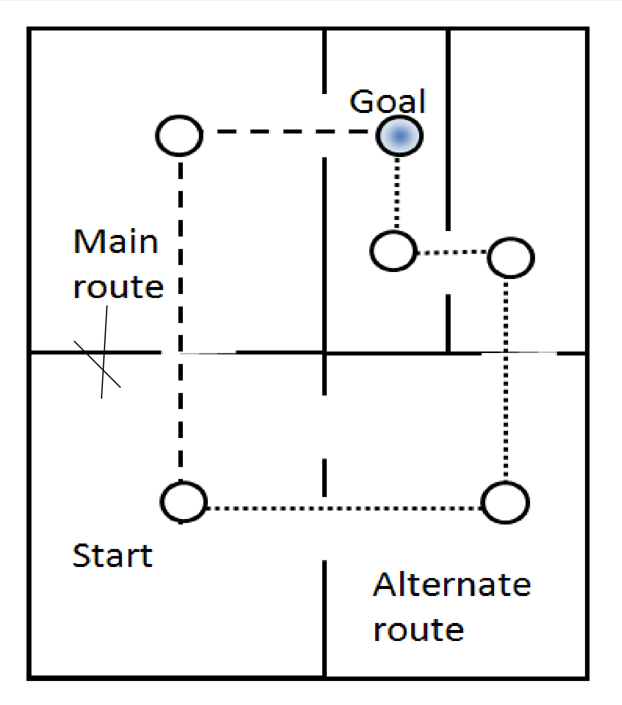
Overview

- Robots and Sensors, especially RGB-D Sensors
- The wayfinding problem for autonomous robots, and the role of landmarks
- Approaches to representing landmarks
- Image Histograms and Terrain Spatiograms
- Handling occluded landmarks
- Automatically selecting landmarks
- Comparing the performance of terrain spatiograms with some other approaches.
- Conclusions

Wayfinding;
Example: for Search and Rescue

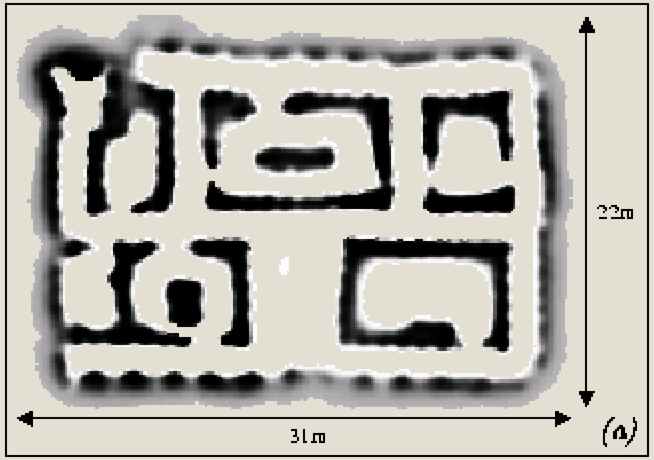
- Where am I in this building?
- Can I construct an ad-hoc map as I go?
- Can I recognize when I return to the same location I have been in?





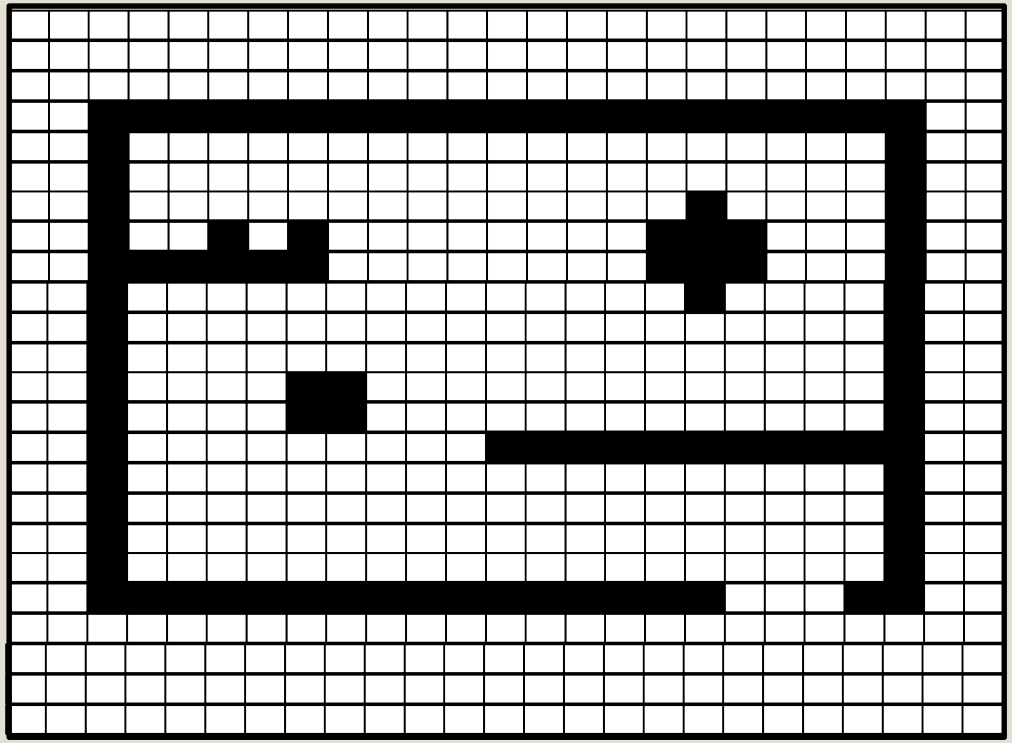
Navigation and Motion Planning

- **Construct Maps** from Sensor Data
- **Localization** of robot on Map
- **Planning motions** of the robot

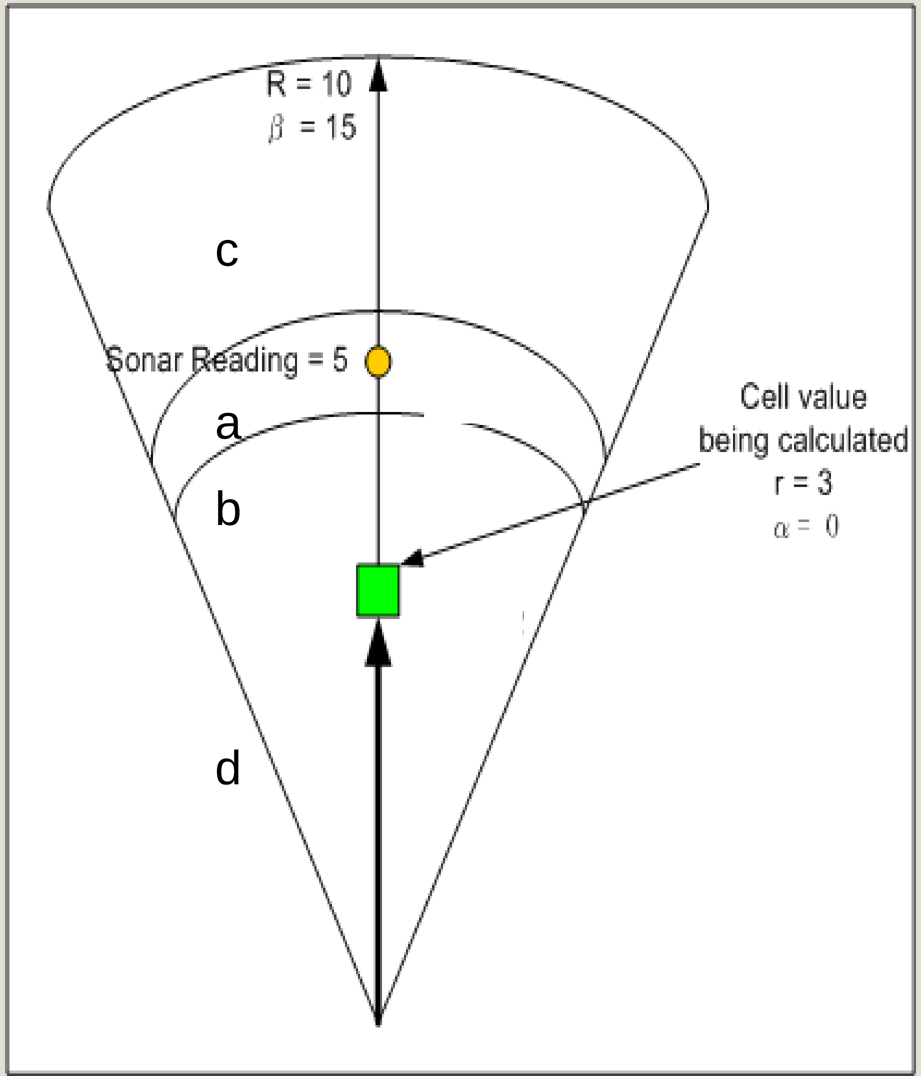


Spatial Occupancy Maps

- Two dimensional grid Morevec & Elfes 1985, Elfes 1987
- Each cell describes the occupancy of a corresponding area
- Probabilistic occupancy map: each cell contains the probability of that area being occupied

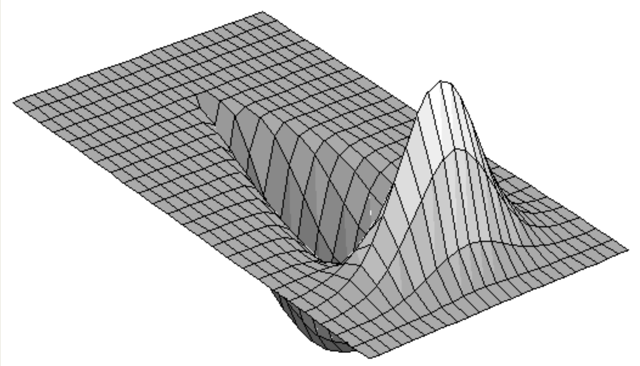


Sonar Cone on Occupancy Grid



Four regions:

a. Is probably occupied
b. Is probably empty
c. Status is unknown
d. Outside the beam



Combined region a and b
Sonar probability model
(Moravec & Elfes 1984)

Example Mapping
(Fox, Baumgard & Thrun 1999)

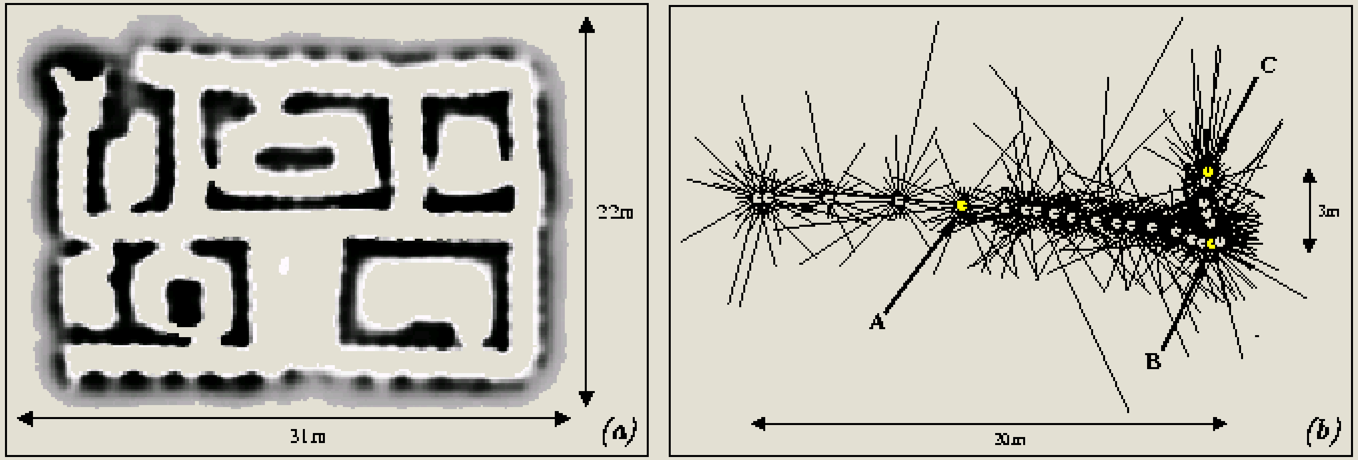
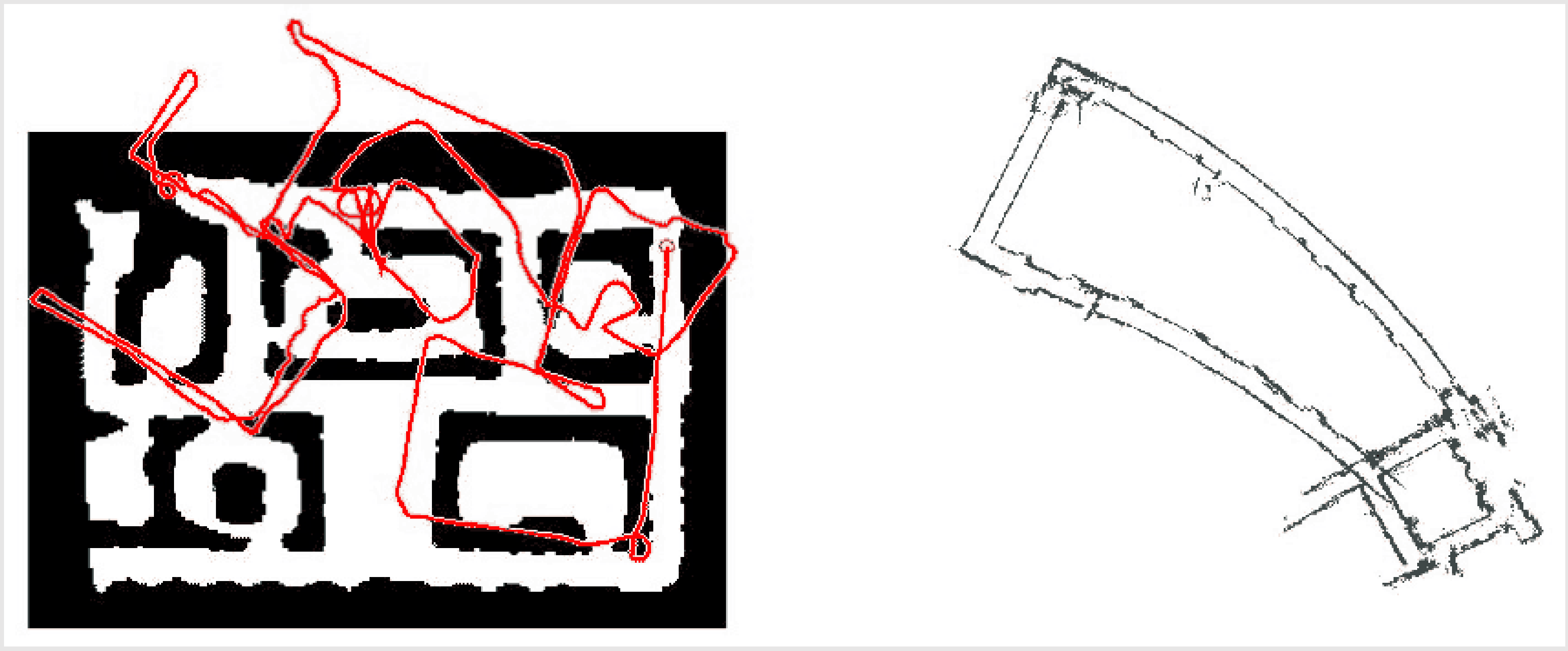


Fig. 16. (a) Occupancy grid map of the 1994 AAAI mobile robot competition arena. (b) Trajectory of the robot and ultrasound measurements used to globally localize the robot in this map.

Localization

- Where exactly is the robot with respect to the map?
- Why? – because odometry is error-prone even when corrected by gyroscopes, inclinometers and other sensors
- Probabilistic approach: What is the probability that the robot is at position x given the motions and sensing so far

Example from FNT’99



SLAM

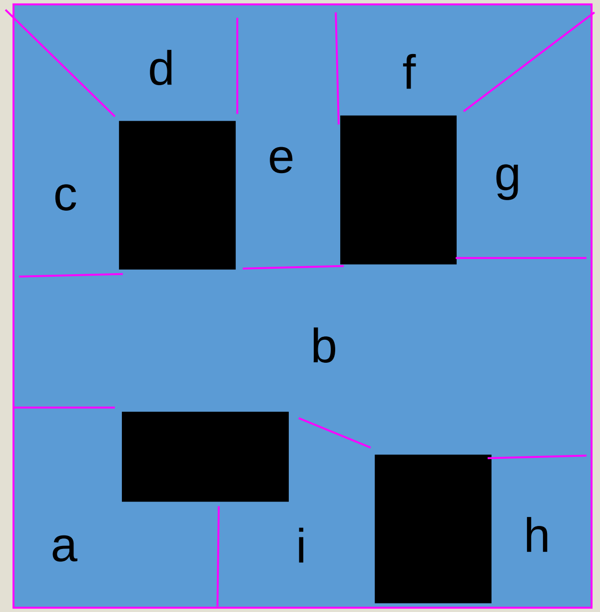
- SLAM
  - Simultaneous Localization And Mapping
  - Figure out where we are and what our world looks like at the same time
- Localization
  - Where are we?
  - Position error accumulates with movement
- Mapping
  - What does the environment look like?
  - Sensor error (not independent of position error)

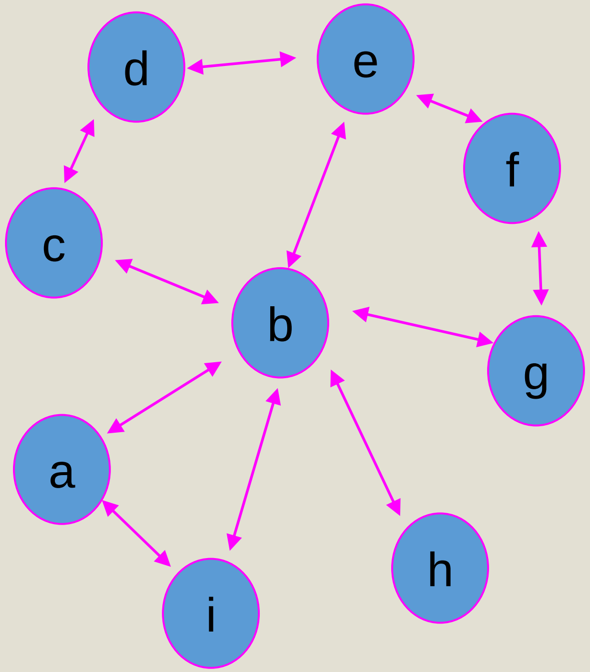
LOOP CLOSURE:
Knowing you are at a previously visited spot allows reduction of error

Expectation Maximization (*EM*)

- Find most likely map (and poses)
- Expectation step (E-step)
  - Calculate probabilities of robot poses for current guess of map
- Maximization step (M-step)
  - Calculate single most likely map for distribution of robot poses
- Iterate

Topological representation





G=(V,E); V={a,..i}; E={{a,b},{a,i},…}

- Represents space as a set of vertices and a set of edges
- Represents the connectivity between ‘places’
- May or not represent geometric details
- May or not contain metric information

Pros and Cons

- Can represent arbitrary size spaces
- Can contain metric information
  - on edges (distances between places)
  - at nodes (local metric map)
- Can be searched with standard graph search algorithm
- Worst case for metric information: Each place only identified by local sensor signature (i.e. visual signature).

The Role of Landmarks

- How to determine when you have ‘closed a loop’ that is, returned to a spot visited earlier, in a metric map
- How to determine when you have arrived at a place in a topological map

Note: ‘landmark’ is often used to denote any sensory feature stored, including edges, lines, regions, etc.

In our usage here, a landmark will be a macro, three-dimensional combination of geometry and texture used for navigation.

This is arguably similar to informal sense of the word landmark.

Why this approach to Landmarks

- Easy to use RGB-D data
- Allow easy (frequent and fast) collaboration between (heterogeneous) robot team members to support local map alignment
- Support human-readable annotation of a map

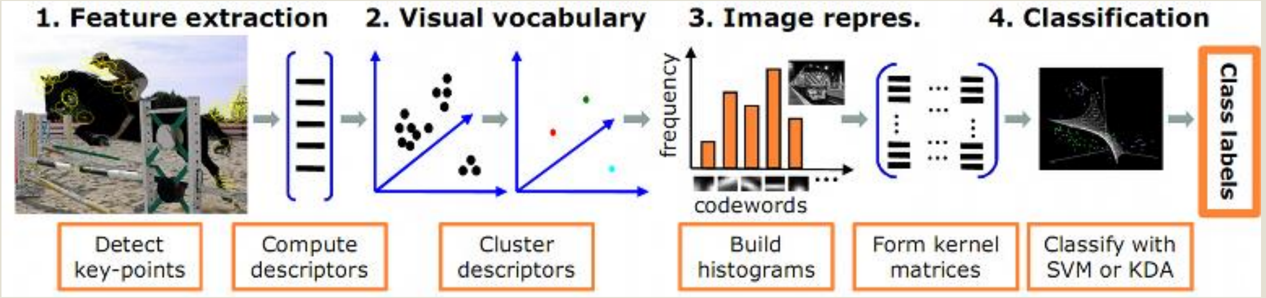
Our approach: Represent a landmark by an abstracted chunk of scene geometry and appearance information.

Overview

- Robots and Sensors, especially RGB-D Sensors
- The wayfinding problem for autonomous robots, and the role of landmarks
- Approaches to representing landmarks
- Image Histograms and Terrain Spatiograms
- Handling occluded landmarks
- Automatically selecting landmarks
- Comparing the performance of terrain spatiograms with some other approaches.
- Conclusions

Prior Work: Representing Natural Landmarks

- Visual templates (Belkenius 1998)
- 360^{o} scenes (Pinette 1994, Franz et al 1998, Fiala 2002)
- Select landmarks whose appearance is independent of scale and rotation – SIFT features (Se et al 2001)
- Planar quadrangles matched by homography (Hayet et al 2002)
- Structural relations of line segments (Frommberger 2006)
- Isomap low-dimensional location and image descriptions for landmarks (Ramos et al 2007)
- Bag of words representations (e.g., CLARET)



Overview

- Robots and Sensors, especially RGB-D Sensors
- The wayfinding problem for autonomous robots, and the role of landmarks
- Approaches to representing landmarks
- Image Histograms and Terrain Spatiograms
- Handling occluded landmarks
- Automatically selecting landmarks
- Comparing the performance of terrain spatiograms with some other approaches.
- Conclusions

Histograms

Histogram:

Let $I : P \rightarrow V$, value $v \in V$ of a pixel at location $p \in P$;

a histogram of I, written h_I maps equivalence classes B on V to the set $\{0,\ldots,|P|\}$ such that

$$h_I(b) = n_b = \eta \sum_{i=1}^{|P|} \delta_{ib}$$

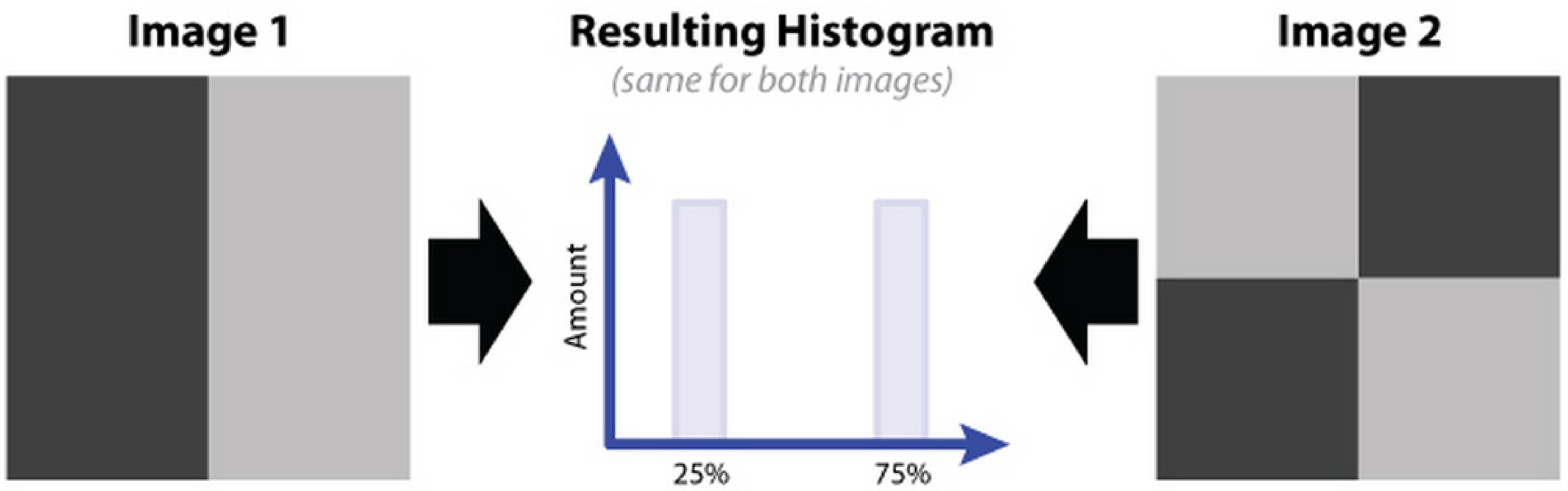


Figure 1: Two different black and white images that have the same histogram.

Histograms & Spatiograms

Histogram:

Let $I : P \rightarrow V$, value $v \in V$ of a pixel at location $p \in P$;
a histogram of I, written h_I maps equivalence classes B on V to the set $\{0,\ldots,|P|\}$ such that

$$h_I(b) = n_b = \eta \sum_{i=1}^{|P|} \delta_{ib}$$

Spatiogram:

Adds information about <u>where</u> values occur in the image: $h_I(b) = \langle n_b , \mu_b , \Sigma_b \rangle$

Spatiograms

Different visual objects

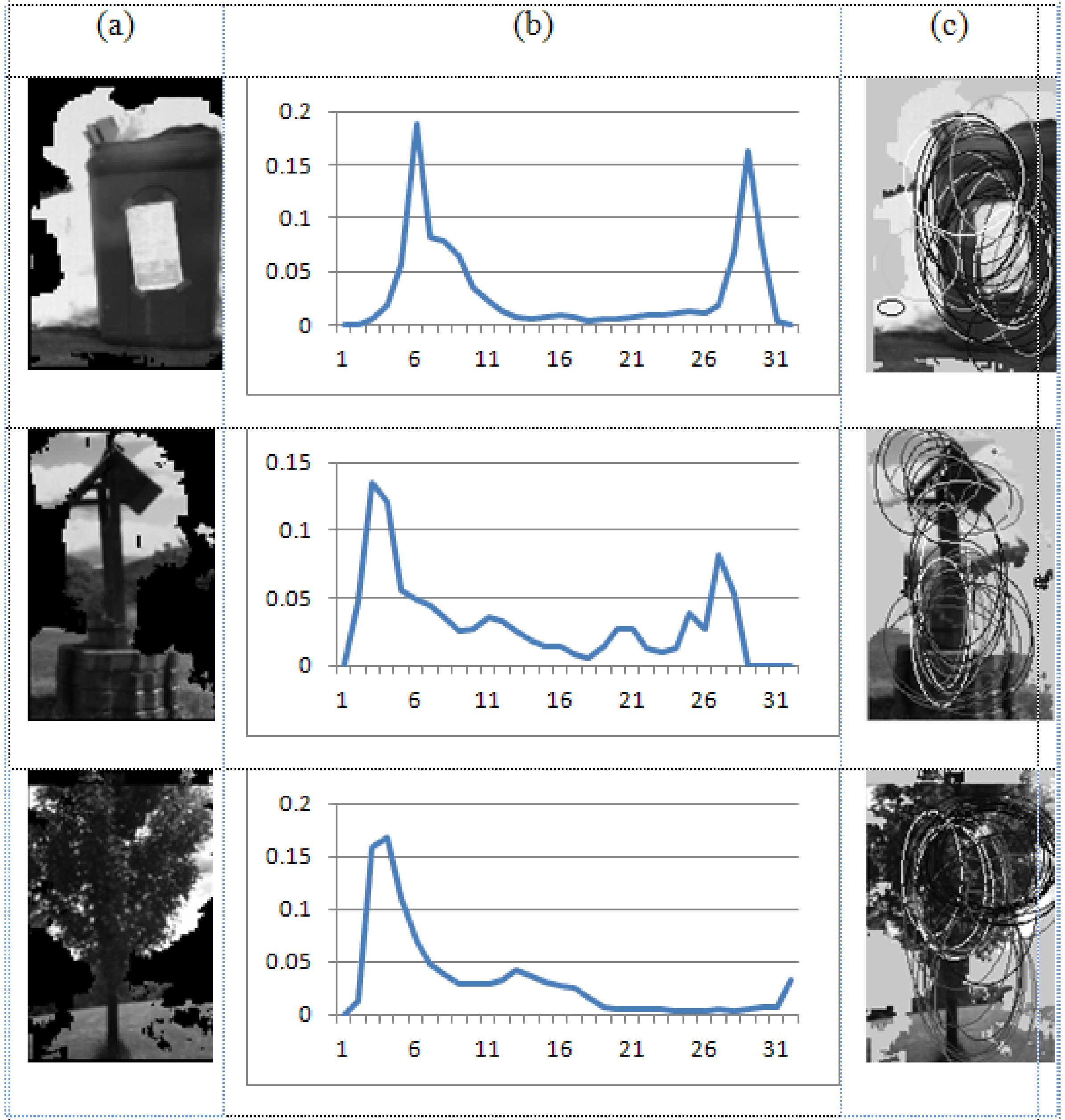


Figure 8-2: Landmark images (a), histograms (b), spatial means (c).

Spatiograms

The same visual object

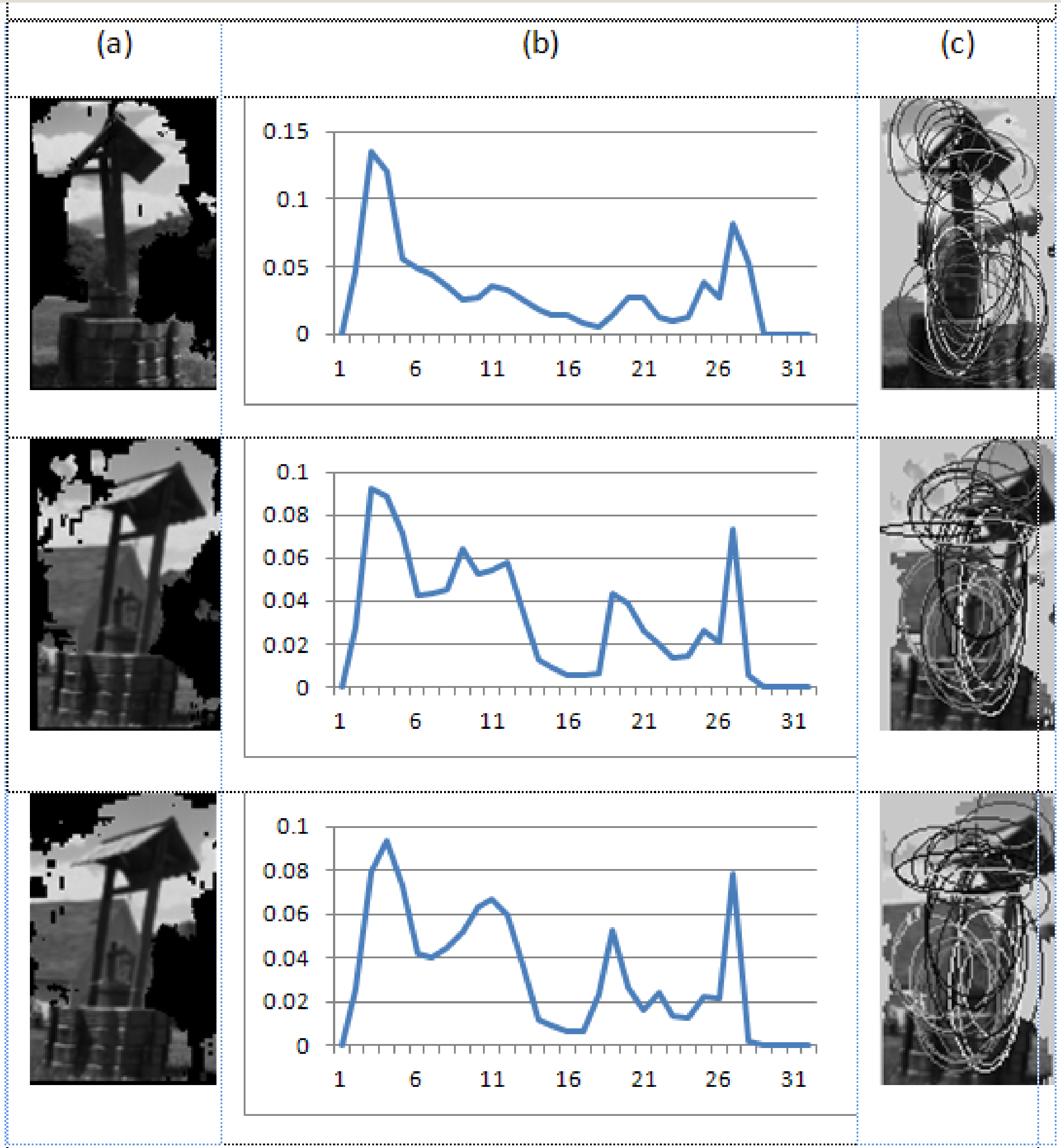


Figure 8-3: Landmark image (a), histogram (b), spatial means (c).

Terrain Spatiogram (TSG)

- A delta function δ_{ib} = *1 iff* the i^{th} pixel is in the b^{th} equivalence class, **and** its 3D location information is available, 0 otherwise.
- A function *d(p)* that maps a pixel at position *p* to its corresponding 3D location so that spatial statistics can refer to 3D (geometric) locations.
- Object-centered, cylindrical or rectangular coordinates.

Terrain Spatiogram

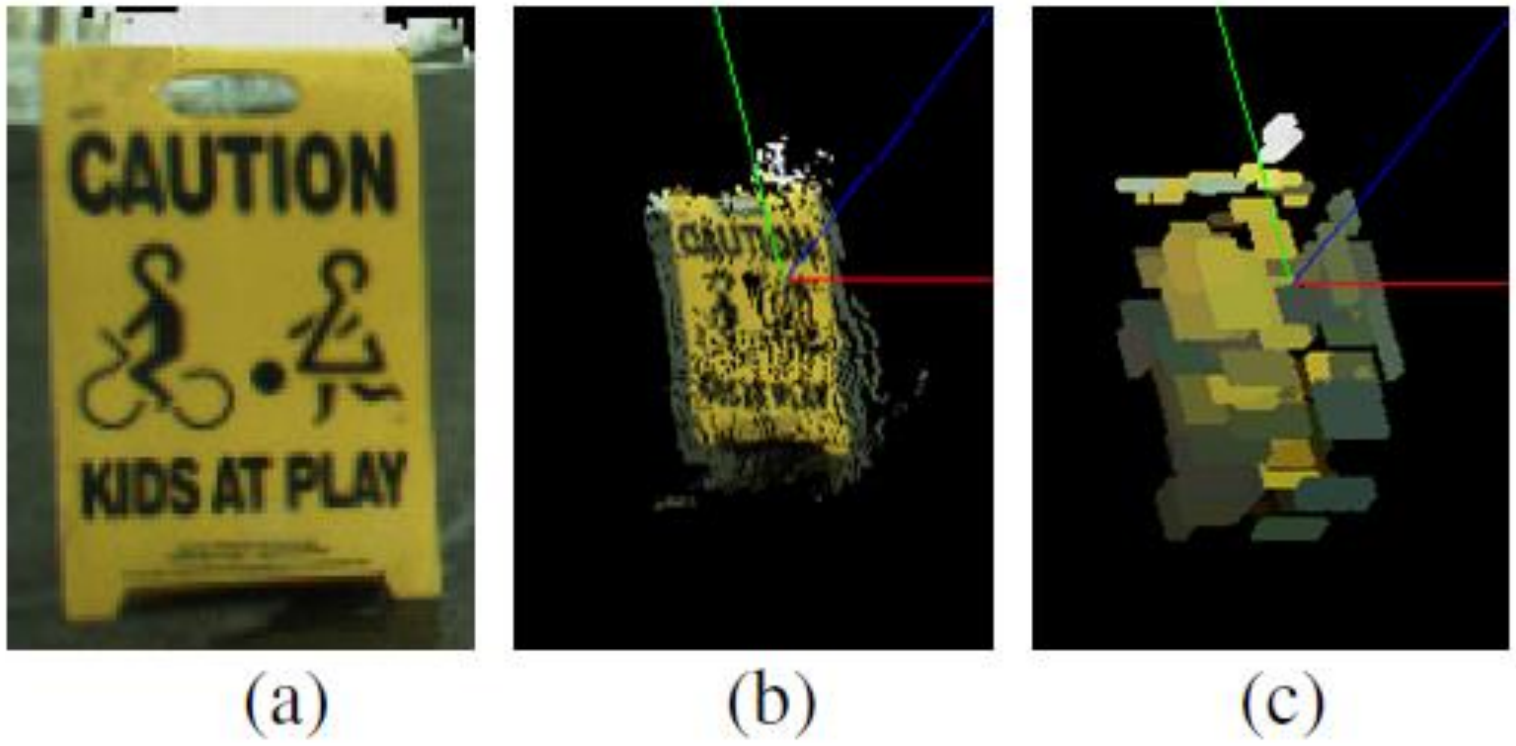
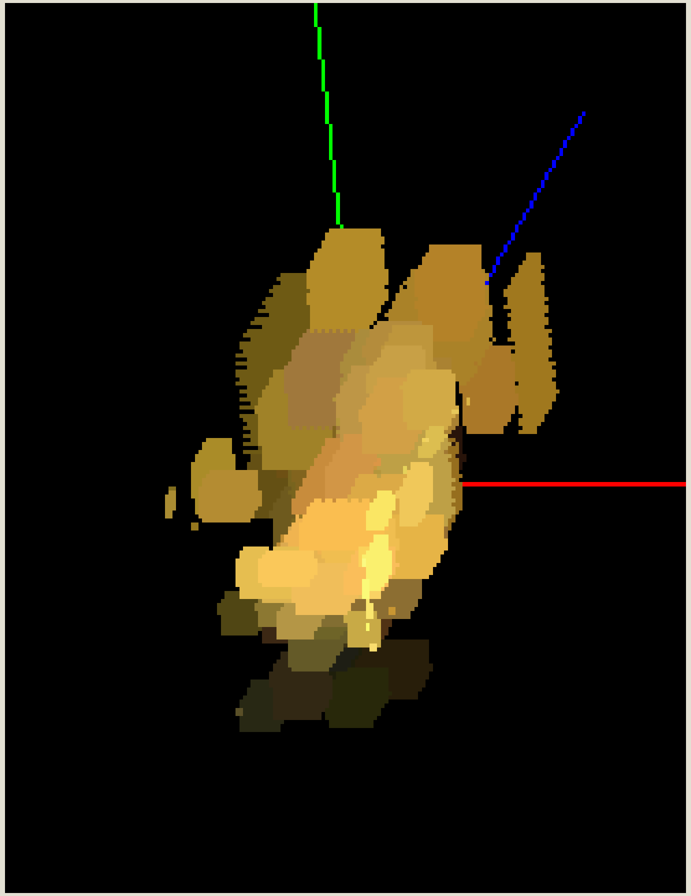


Figure 4: Terrain Spatiogram (TSG) Example
(a) Original image; (b) pixels mapped to depth; (c) TSG

Terrain Spatiogram

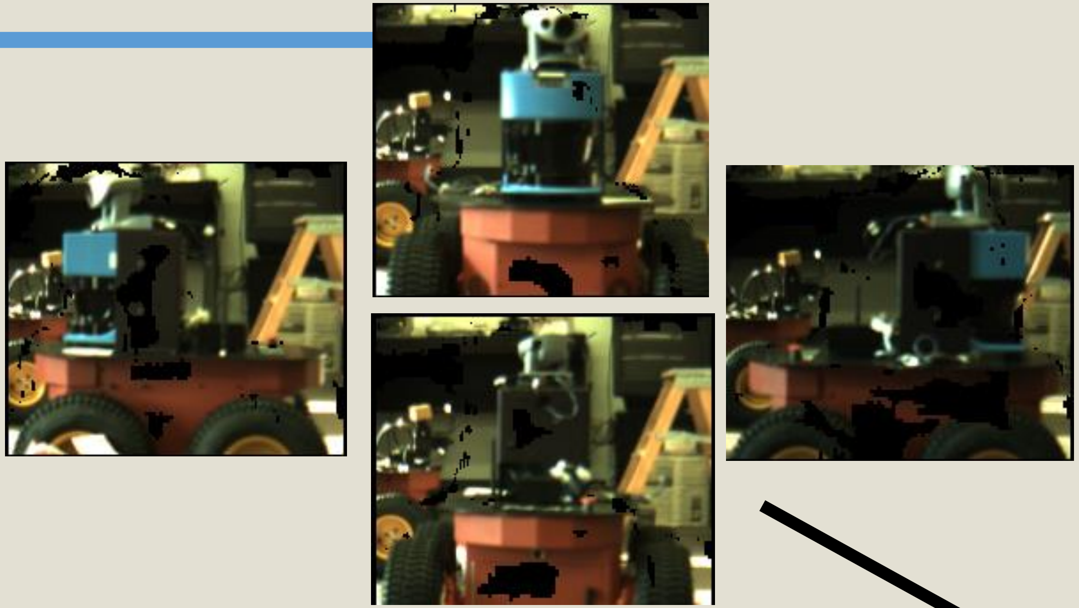
Video texture,
Pixels with valid
disparity

Monochrome
Disparity map

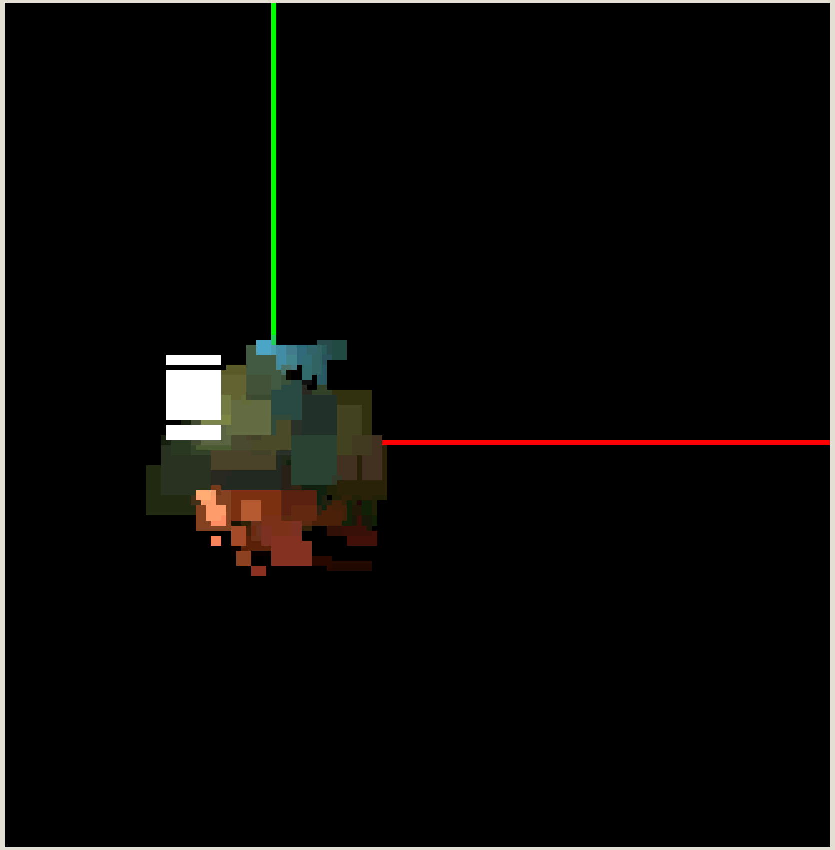


TSG calculated from stereo data

Another Motivation: Fusing Multiple Views



A single TSG that contains data from multiple views



TSG = Spatiogram with 3D Spatial Moments

- Collect set of Landmark Images
- Subregion of image with Landmark
- Update histogram using color information: $bin(p) = r + g\, s_b + b\, s_b^2$
- Update mean using depth information
- Update covariance using depth and mean

=> Set of h, one for each landmark

$$h(b) = \langle n_b, \mu_b, \Sigma_b \rangle, \; N(\mu_b, \Sigma_b)$$

$$n_b = \sum_{i=1}^{|P|} \delta_{ib}$$

$$\mu_b = \frac{1}{\sum_{j=1}^{|P|} \delta_{jb}} \sum_{i=1}^{|P|} d(p_i)\delta_{ib}$$

$$\Sigma_b = \frac{1}{\sum_{j=1}^{|P|} \delta_{jb}} \sum_{i=1}^{|P|} (d(p_i) - \mu_b)(d(p_i) - \mu_b)^T \delta_{ib}$$

Recognizing a landmark: Comparing TSGs

Normalized similarity (O'Conaire et al 2007)

$$\rho(h,h') = \sum_{b=1}^{|B|} \psi_b \sqrt{n_b n'_b}$$

Where $\psi_b = 2(2\pi)^{0.5} |\Sigma_b \Sigma'_b|^{0.25} N(\mu_b ; \mu'_b , 2(\Sigma_b + \Sigma'_b))$

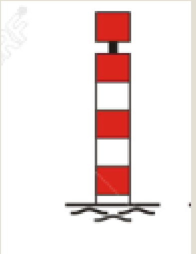
Step 1: Collect TSG h from current image subregion

Step 2: Identify landmark λ *from list L of landmark TSG using :*

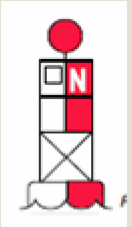
$$\arg\max_{\lambda \in L} \rho(h, h_\lambda)$$

Mixture of Gaussian (MoG or GMM) TSGs

$$h(b) = \langle n_b,\ m_b = ((\alpha_{b1}, \mu_{b1}, \Sigma_{b1}), \ldots, ((\alpha_{bm}, \mu_{bm}, \Sigma_{bm}))\ \rangle$$



$$p(x \mid m_b) = \sum_{i=1}^{m} \alpha_{bi} N(x; \mu_{bi}, \Sigma_{bi})$$



$$\psi_b^{mm} = \sum_{i=1}^{m} \alpha_{bi} \sum_{j=1}^{m} \alpha'_{bj} \eta_{bij} N(\mu_{bj};\ \mu'_{bi}, 2(\Sigma'_{bi} + \Sigma_{bj}))$$

Need to do clustering
to find mixture members!

Overview

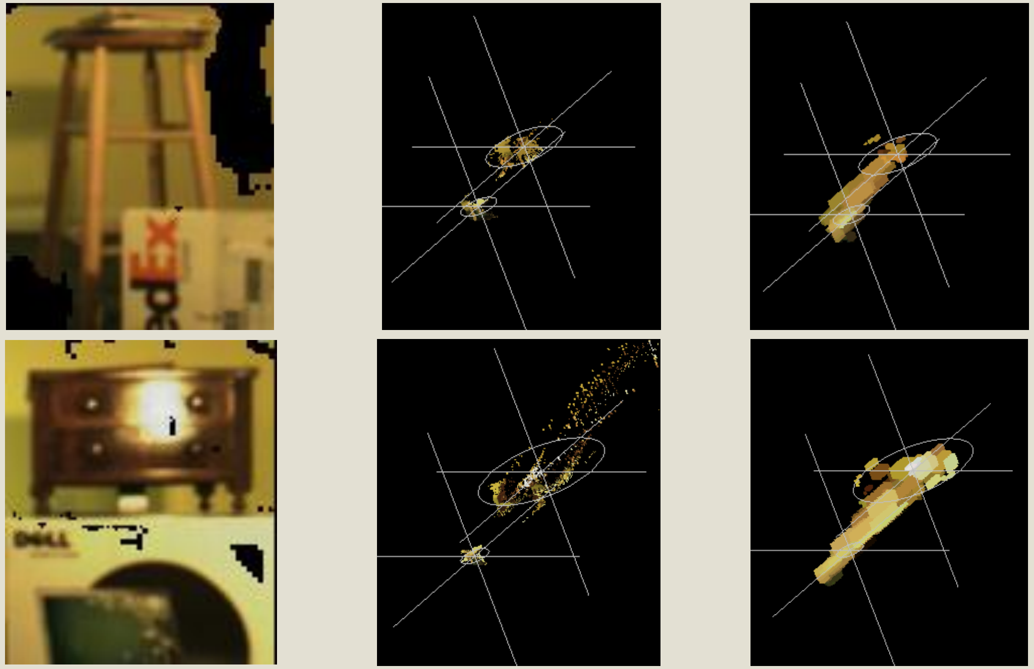
- Robots and Sensors, especially RGB-D Sensors
- The wayfinding problem for autonomous robots, and the role of landmarks
- Approaches to representing landmarks
- Image Histograms and Terrain Spatiograms
- Handling occluded landmarks
- Automatically selecting landmarks
- Comparing the performance of terrain spatiograms with some other approaches.
- Conclusions

Occlusion

- Landmark Occlusion is a depth related phenomenon
- A landmark is occluded when an occluding object
  - hides a portion of the landmark
  - as a consequence of being between the sensor and the landmark

Identifying Occlusion

(a) (b) (c)



(d) (e) (f)

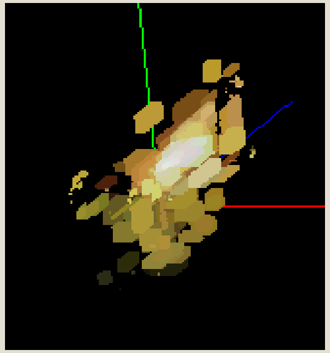
XZ is ground plane
Y is height

An occlusion will always have a separate cluster center in lower Z than the landmark!

- Occluded Landmark left image of stereo pair (a, d);
- Perspective view of image pixels mapped to absolute depth (b, e);
- Perspective view of terrain spatiogram with XZ cluster center and 1SD circle (c, f) from K-Means clustering

Steps in Occlusion Filtering

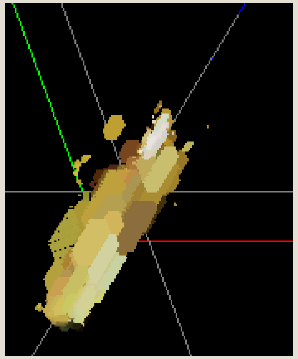
Unoccluded landmark



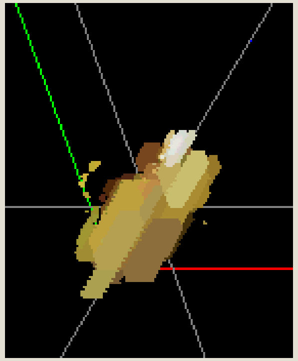
TSG before trimming

TSG after trimming outliers

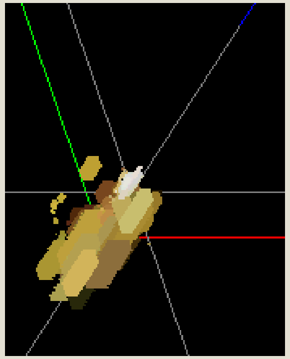
Candidate (occluded)
Landmark



candidate cluster
center



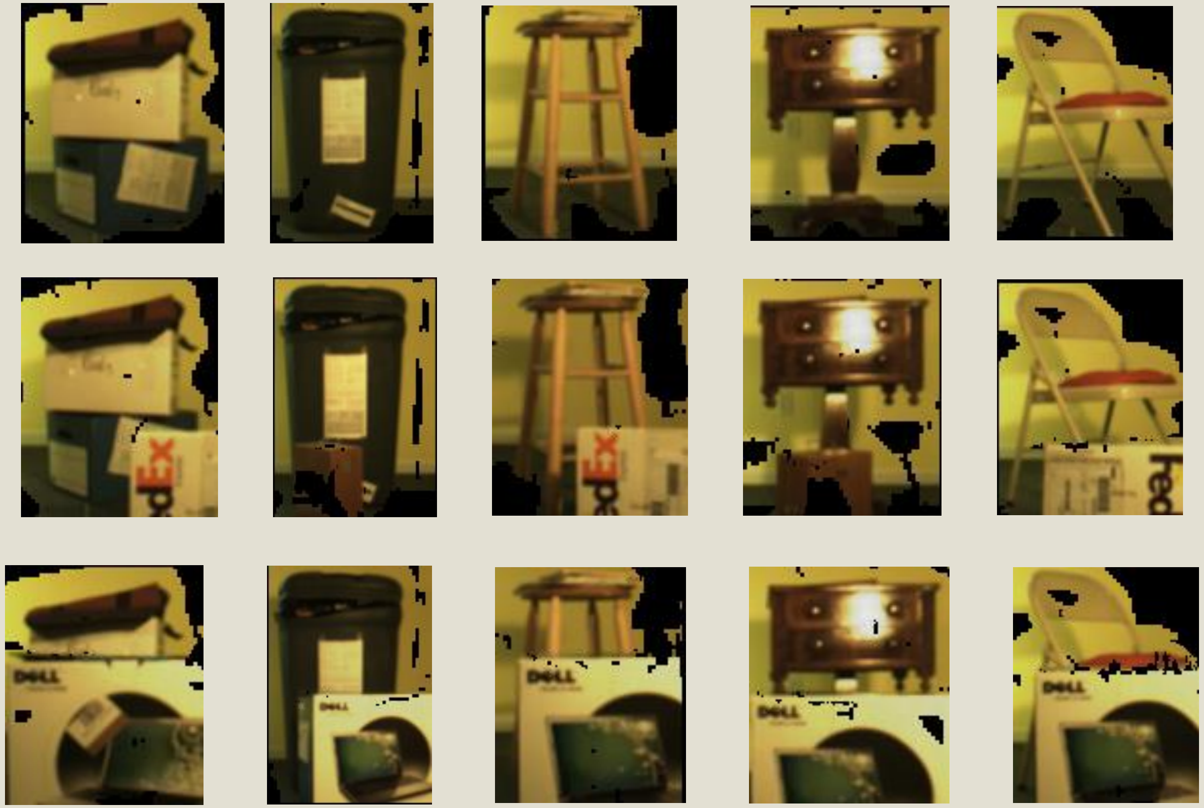
TSG trimmed
to landmark moments



translated
to Z origin

Unoccluded and Occluded Landmarks

(a) (b) (c) (d) (e)



(1)

(2)

(3)

Results

Table 1: Confusion Matrix for Landmarks.

a	b	c	d	e	
1	0.434	0.463	0.385	0.416	**a**
0.483	1	0.417	0.459	0.335	**b**
0.486	0.351	1	0.545	0.61	**c**
0.41	0.4	0.533	1	0.449	**d**
0.485	0.258	0.61	0.486	1	**e**

Table 2: Direct Normalized Comparisons

	ρ11	ρ22	ρ33	ρ12	ρ13
a	1	1	1	0.815	0.485
b	1	1	1	0.828	0.697
c	1	1	1	0.571	0.405
d	1	1	1	0.868	0.632
e	1	1	1	0.835	0.483

Occlusion-Filtered Landmarks

Table 3: Occlusion Filtered Normalized Comparisons.

	ρ1'2'	ρ1'3'	ρ1'2' %change	ρ1'3' %change
a	0.905	0.694	11.113	42.86
b	0.893	0.885	7.871	26.92
c	0.632	0.549	10.721	35.628
d	0.917	0.812	5.687	28.574
e	0.914	0.611	9.536	26.455

Draped Landmarks

(4)

Table 4: Normalized Comparisons with draped landmarks.

	ρ14	ρ1'4'	ρ1'4' %change
a	0.727	0.694	-4.53
b	0.83	0.864	4.095
c	0.867	0.92	6.034
d	0.748	0.799	6.738
e	0.623	0.581	-6.701

Overview

- Robots and Sensors, especially RGB-D Sensors
- The wayfinding problem for autonomous robots, and the role of landmarks
- Approaches to representing landmarks
- Image Histograms and Terrain Spatiograms
- Handling occluded landmarks
- Automatically selecting landmarks
- Comparing the performance of terrain spatiograms with some other approaches.
- Conclusions

Landmark saliency architecture

Objective:
Automatically extract TSG landmarks from RGB-D data based on visual saliency and which are similarly salient to humans.

- Saliency consists of three components (Raubell & Winter 2002)
  - Visual attraction
  - Structural attraction
  - Semantic attraction

Visual attraction

- Input is RGB-D images

$I_c = \{ c_{ij} = (v_1, v_2, v_3) \mid i \in 1..n, j \in 1..m \}$

$I_d = \{ d_{ij} = (x_1, x_2, x_3) \mid i \in 1..n, j \in 1..m \}$

- Retinal cone responses: Red-green/Blue-yellow (Schoss & Palmer 2009)
- => CIELab color opposition space
- Visual Attraction Module applied to Depth and Color images in parallel

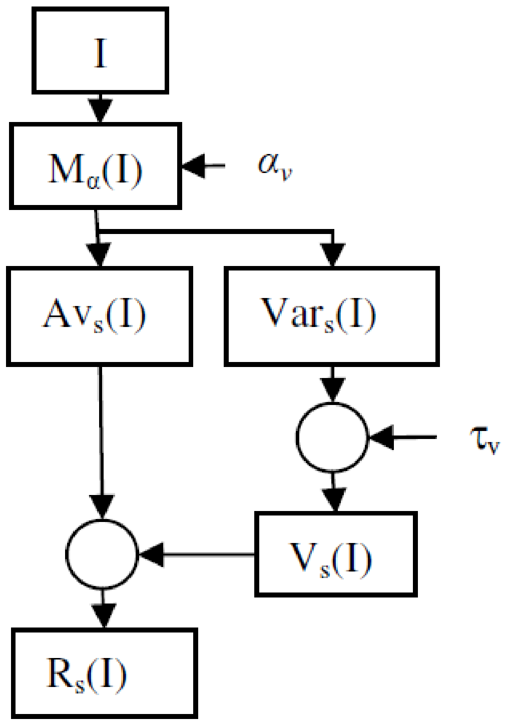


Figure 1: Visual Attraction Module

Structural Attraction

- Input is Rs(Ic) and Rs(Id)
- Three structural attractiveness properties:
  - Region area
  - Aspect ratio
  - Fused attractiveness

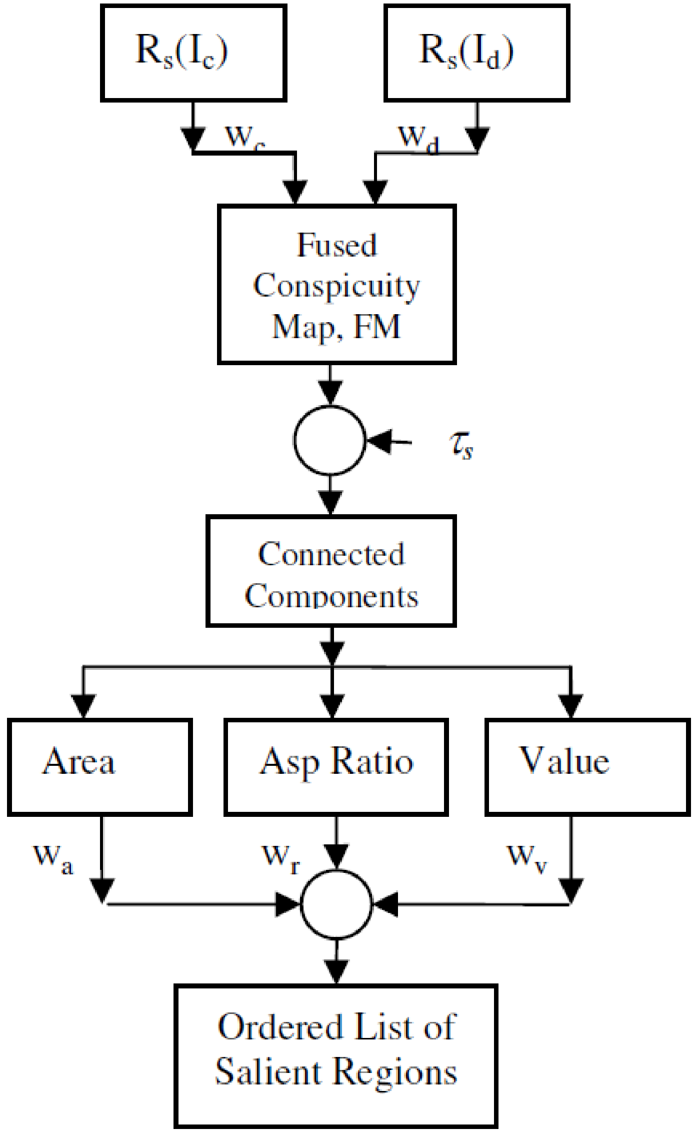


Figure 3: Structural Attractiveness Module

Example

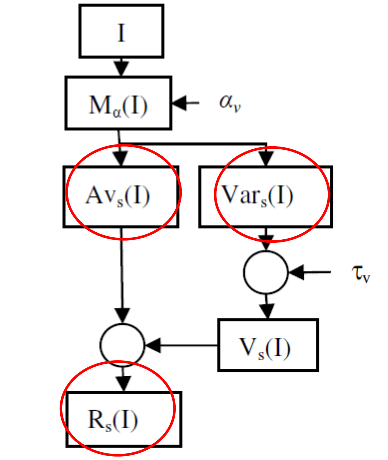


Figure 1: Visual Attraction Module

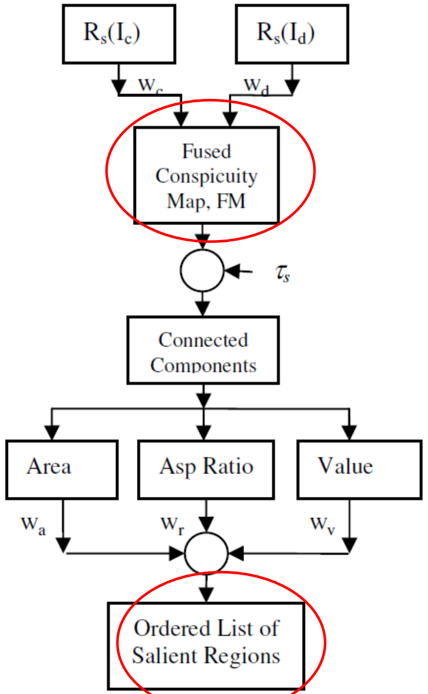


Figure 3: Structural Attractiveness Module

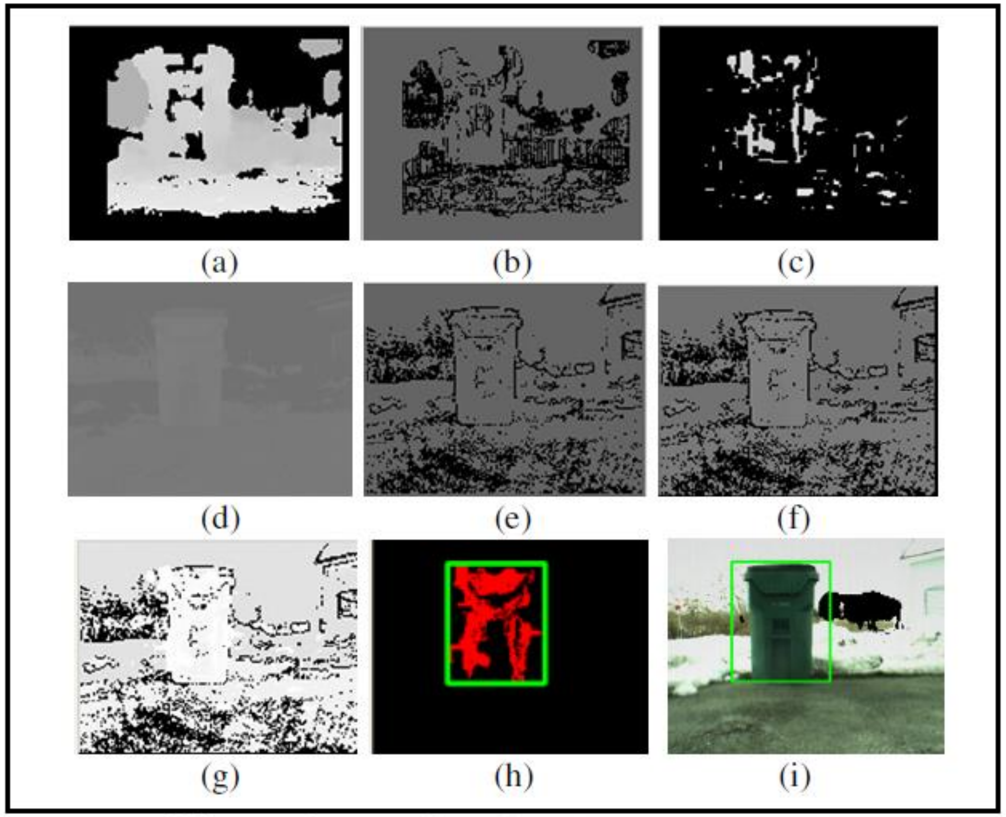


Figure 2: Landmark Saliency Example
(a-c): $Av_s(I_d)$, $Var_s(I_d)$, and $R_s(I_d)$;
(d-f): $Av_s(I_c)$, $Var_s(I_c)$, and $R_s(I_c)$;
(g-i): Fused Conspicuity map, Top saliency region, original image showing top region. Brighter is more salient in a-g.

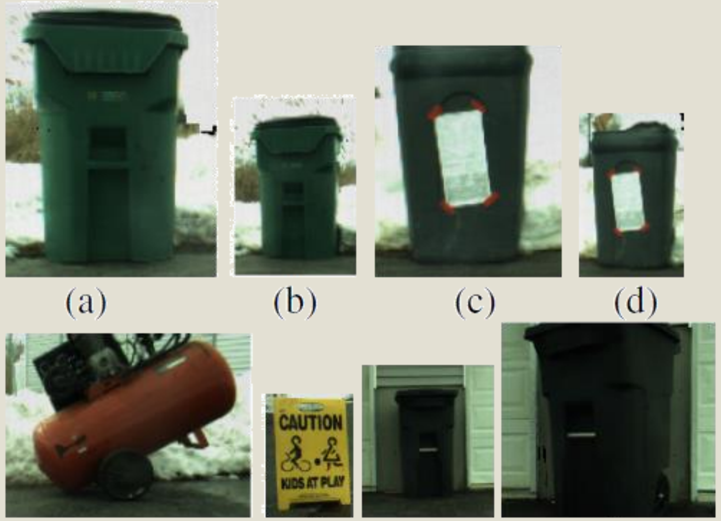
Semantic attractiveness

- All the seven settings for masks, thresholds and weights in the visual and structural modules (α_v , τ_v, w_c ,w_d , τ_s, w_a ,w_r , w_v)

- α_v : This parameter allows the salience of the input components to be reversed or masked
- τ_v: This controls how smooth surfaces need to be to show up as salient.
- w_c ,w_d : These two mutually dependent parameters indicate how important spatial information is relative to color information.
- τ_s: This controls how salient a fused region needs to be to appear in the list of regions.
- w_a ,w_r ,w_v : These three mutually dependent parameters control the relative attractiveness of large regions versus small regions, vertical regions (tall) versus horizontal (squat) regions and high versus low fused visual attractiveness.

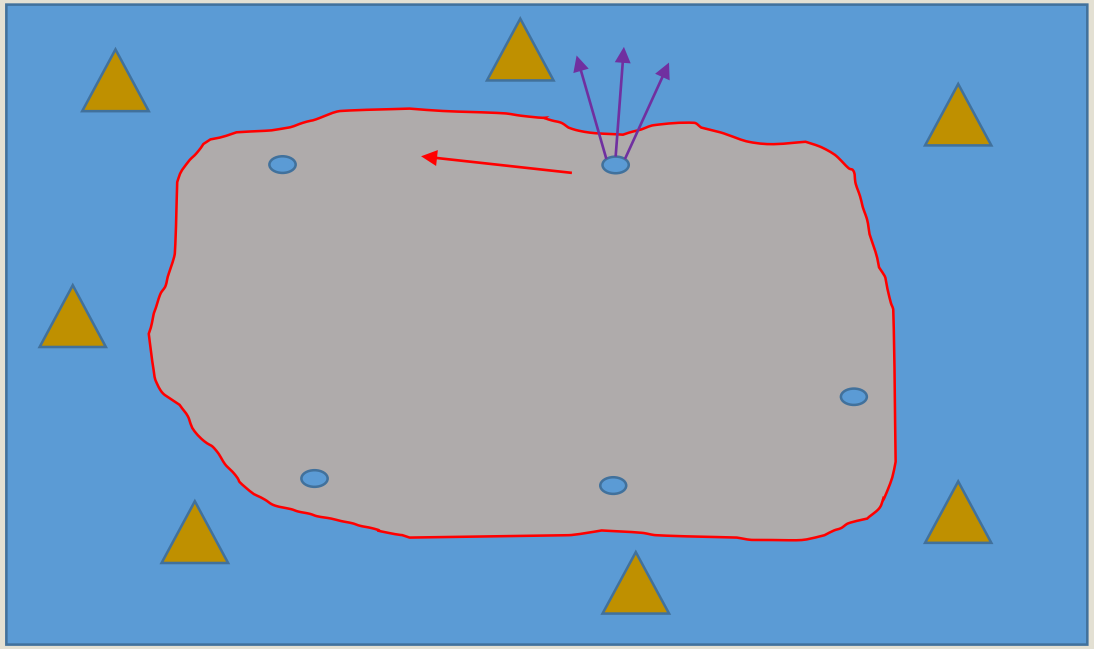
Experiments

- Pioneer 3-AT, Videre stereocam (f=6mm), Biclops PT base.
- Robot followed loop around 7mx10m blacktop area.
- Stopped at regular distances and collected images at (80,90,100) looping away from blacktop.



(a) (b) (c) (d)

(e) (f) (g) (h)



Recognition results

- Univariate TSG for each LSA candidate (46 in total)
- Filtered to top 3 matches per candidate (ρ>0.6); leaving 7 landmarks with 3 poses (21 TSGs).
- 21x21 Confusion matrix

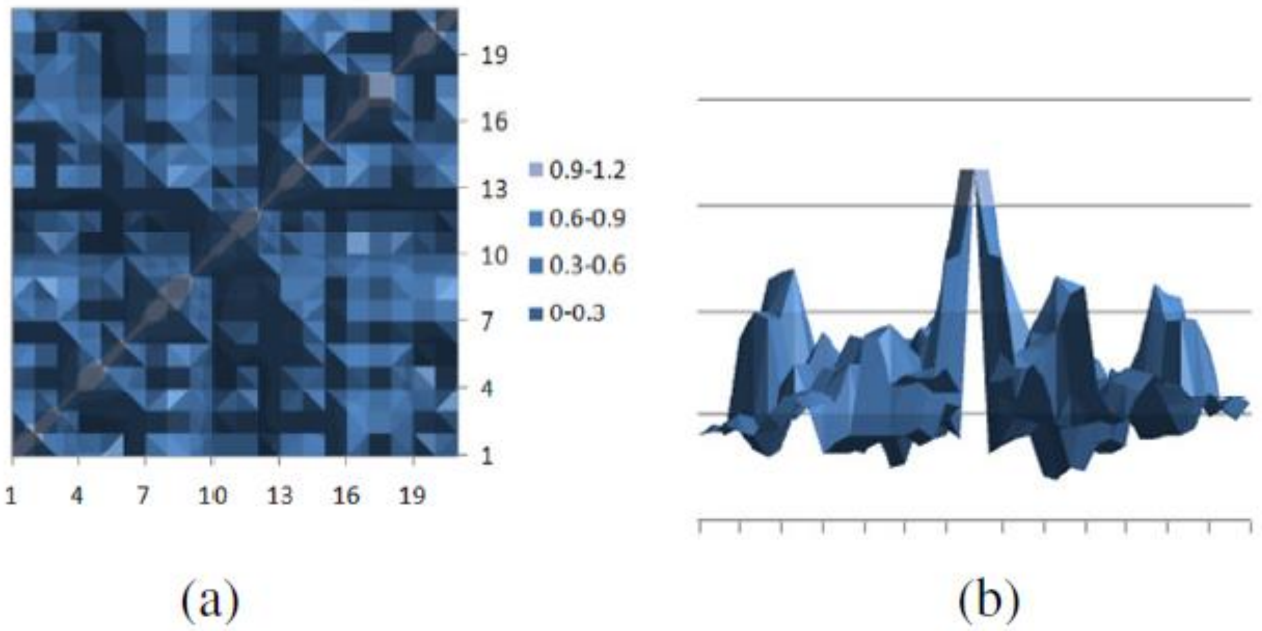


Figure 7: Confusion Matrix for TSG Comparisons

Overview

- Robots and Sensors, especially RGB-D Sensors
- The wayfinding problem for autonomous robots, and the role of landmarks
- Approaches to representing landmarks
- Image Histograms and Terrain Spatiograms
- Handling occluded landmarks
- Automatically selecting landmarks
- Comparing the performance of terrain spatiograms with some other approaches.
- Conclusions

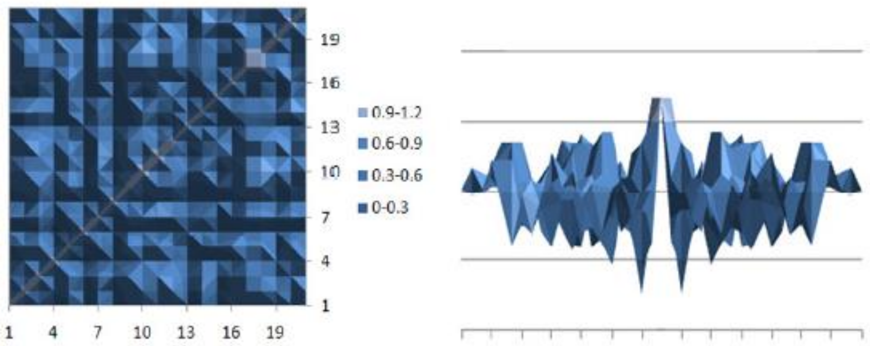
Comparison with two other approaches

Table 1: Comparison of Confusion Matrix Means

Method	Diagonal	Off-Diagonal
TSG	0.79	0.37
SQDIFF	0.59	0.59
HISTO	0.84	0.68

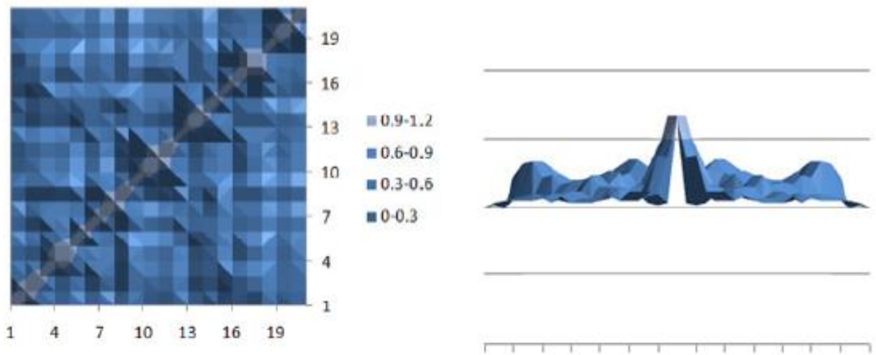
Table 2: Variance Ratios for each method

Method	Var Ratio
TSG	0.77
SQDIFF	0.41
HISTO	0.68



(a) (b)

Figure 8: Confusion Matrix for SQDIFF Comparisons



(a) (b)

Figure 9: Confusion Matrix for Histogram Comparisons

Conclusions

- Terrain Spatiogram Landmark Representation

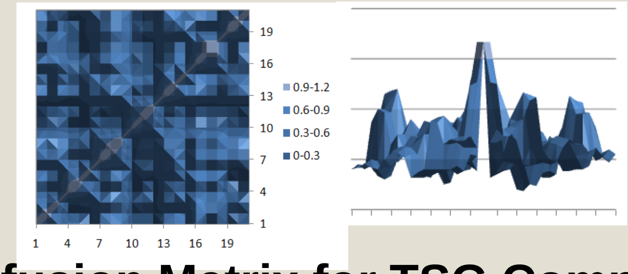
Represents landmark as abstract chunk of scene texture and geometry

- Discussed:
  - Simplifies recognition of occluded landmarks
  - Can be automatically selected by robot as it travels
  - Has good recognition characteristics
- Did not discuss:
  - How to share TSG among robot team members and with people
  - How to construct TSG from multiple orientations of same object

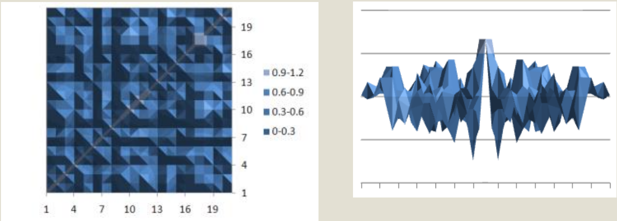
Selection and Recognition of Landmarks using Terrain Spatiograms

Damian M. Lyons
Fordham University

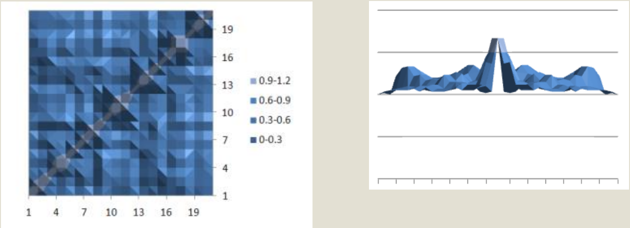
- A team of robots share information about observed landmarks
- Terrain spatiograms (tsg) combine spatial and image landmark data
- Saliency architecture autoselects landmarks
- tsg reliably recognizes autoselected landmarks
- Improves on SQDIFF & histogram approaches



Confusion Matrix for TSG Comparisons



Confusion Matrix for SQDIFF Comparisons



Confusion Matrix for Histo Comparisons