

Sharing and fusing landmark information in a team of autonomous robots

Damian M. Lyons

Robotics & Computer Vision Laboratory
Department of Computer & Information Science
Fordham University, Bronx, NY 10458, USA
lyons@cis.fordham.edu

ABSTRACT

A team of robots working to explore and map a space may need to share information about landmarks so as register local maps and to plan effective exploration strategies. In this paper we investigate the use of spatial histograms (spatiograms) as a common representation for exchanging landmark information between robots. Each robot can use sonar, stereo, laser and image information to identify potential landmarks. The sonar, laser and stereo information provide the spatial dimension of the spatiogram in a landmark-centered coordinate frame while video provides the image information. We call the result a terrain spatiogram. This representation can be shared between robots in a team to recognize landmarks and to fuse observations from multiple sensors or multiple platforms. We report experimental results sharing indoor and outdoor landmark information between a two different models of robot equipped with differently configured stereocameras and show that the terrain spatiogram (1) allows the robots to recognize landmarks seen only by the other with high confidence and, (2) allows multiple views of a landmark to be fused in a useful fashion.

Keywords: Mobile Robots, Navigation, Mapping, Landmarks, Sensor Fusion

1. INTRODUCTION

A team of mobile robot can play an important role in an urban search and rescue framework by exploring, mapping, guiding and assisting the incident support team members (IST) at the scene of the disaster and, via the site information collected, also at the IST base of operations control center. The activities of the robot team comprise only a portion of the activities that IST members must perform, and in this work we focus on the challenges of just one of these robot activities: Early deployment at the site to collaboratively and quickly generate a map of the site showing hazards, obstacles, traversable routes, and potential victims; as time proceeds the map can be refined by the team. However, a fast, accurate and human communicable map will allow rescuers to deploy more quickly, effectively and safely.

In this paper, we propose a novel landmark representation, the *terrain spatiogram*, which is designed to allow the easy fusion of data from multiple sensors and multiple platforms and to facilitate the sharing of landmark information between mobile platforms in the team. The concept is based on Birchfield and Rangarajan's spatiogram [1][2] extended to represent 3D terrain information rather than image information. Previous work is reviewed in Section 2 of the paper. In Section 3, we introduce the terrain spatiogram notation, as well as its relation to the standard spatiogram, the spatiogram difference operation and its extension to a Mixture of Gaussians framework. Section 4 presents our experimental results demonstrating that the approach is effective in sharing landmark information between mobile platforms and in fusing multiple landmark views.

2. PRIOR WORK

Deploying robot technology for US&R operations is very challenging, including at the least challenging problems in terrain mobility, navigation, situational awareness and human-machine interaction [11]. Odometry and other dead-reckoning approaches are disadvantaged by the slippery and unstable terrain [3]. We will adopt the biomimetic approach of a *cognitive map* [13], and consider our map to be composed of in the main of *passive, natural landmarks* [6] identifying places, the edges identifying routes between places and augmented with navigation and hazard information. The map will include some metric information by necessity (e.g., [16]) since that may be necessary for navigation, communication or rescue-planning information for IST members, but we will view its construction as primarily topological. The landmarks will be passive and natural, with one exception: *other robots in the team can also perform as visual landmarks*, leveraging a valuable, additional source of localization information.

Birchfield and Rangarajan [1][2] proposed an extension to the concept of a histogram for appearance based video tracking. Their *spatial histogram* or *spatiogram* augments an image histogram with a Gaussian distribution per histogram bin that summarizes the image location for that histogram bin. If a robot is equipped with range sensing equipment in addition to a visual sensor, then it is possible to relate the image positions of the spatiogram to Cartesian coordinates relative to the robot. If this spatial information is used rather than image spatial information, we refer to the result as a *terrain spatiogram* (as opposed to image spatiogram). A landmark in our cognitive map will be represented by a stored terrain spatiogram in landmark-centered cylindrical coordinates. Ramos et al. [14] shows that such a combination of depth and image information can be a powerful tool for landmark recognition. They employ Tenenbaum's *Isomap* to learn low-dimensional location and image descriptions for landmarks to implement loop-closure for outdoor SLAM.

One of the principle uses of a landmark is to allow the recognition of a place. Once the landmark is matched, then the bearing of the robot from the place can be calculated. If sufficient landmarks are seen, then the robot can be localized with respect to the landmarks. Many approaches to mapping and navigation rely on being able to identify landmarks independent of their scale and rotation and using a lot of these matches to localize the robot. In that case, it preferable to select landmarks whose appearance is independent of scale and rotation – SIFT features [15]. Experimental indicate however that insects can extract landmark orientation as well as bearing information [10] by matching the surrounding panoramic view as seen from the landmark. The advantage of extracting the additional orientation information is that it leads to a more accurate estimate of the robot position with fewer landmarks (since every landmark now provides two separate pieces of information). Our biological inspired approach matches the panorama not of the surrounding scene (which in an urban disaster scene is probably chaotic and may be dynamic) but the more static panorama of the landmark viewed (in the best case) from all sides.

Panoramic and omnidirectional cameras have been used in robot navigation for some time [4][7][9] and there is evidence that panoramic processing is used in some kinds of insect navigation [5]. We can consider the terrain spatiogram in cylindrical coordinates to be analogous to an omnidirectional camera image but with camera normal facing in – towards the object – rather than out – towards the environment.

The approach proposed here has a number of key features:

1. The landmark information can be easily transformed to a global coordinate frame and shared between robots.
2. Since localization information for the robots in a team can contain uncertainty, an important feature of the terrain spatiogram is that it allows for both robust matching, and for extracting landmark orientation information from the match. This provides robustness and additional information to refine localization.
3. The same landmark seen from a different orientation or by a different robot provides information that can be fused into a global terrain histogram for the landmark.

3. THE TERRAIN SPATIOGRAM APPROACH

3.1 Spatiograms

Let $I : P \rightarrow V$ be a function that returns the value $v \in V$ of a pixel at a location $p \in P$ in the image. The histogram of I captures the number of times each pixel value occurs in the range of the function I . Consider a set, B , of equivalence classes on V , a histogram of I , written h_I maps B to the set $\{0, \dots, |P|\}$ such that $h_I(b) = n_b$ and

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

where δ_{ib} is a delta function equal to 1 iff the i^{th} pixel is in the b^{th} equivalence class and 0 otherwise, and η is a normalizing constant so that together all n_b sum to 1. A spatiogram or spatial histogram adds information about where values occur in the image:

$$h_I(b) = \langle n_b, \mu_b, \Sigma_b \rangle$$

where μ_b , Σ_b are the spatial mean and covariance of the values in the class b defined as:

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

Birchfield & Ragajaran define a histogram as a first order spatiogram, a formulation that also allows for second and higher order spatiograms. We will restrict our attention to second order spatiograms in this work.

They also introduce an approach to comparing two spatiograms as the spatially weighted sum of similarities

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

where ψ_b evaluates the spatial means of bins in h in the spatial distributions of h' and where ρ_n compares the bin values. O'Conaire et al. [12] developed a normalized spatiogram comparison measure (one in which $\rho(h, h) = 1$). This makes it much more intuitive to use ρ to match two spatiograms.

3.2 Terrain Spatiograms

The spatial dimensions used by Birchfield & Ragajaran and others are the spatial dimensions of the image and a primary use of spatiograms has been for color-based tracking in video images. Note that there is nothing which constrains the spatial dimensions to be in the image. If, for example, the image information comes from a stereo camera, then the spatial information can be three-dimensional depth information. The spatiogram definitions can be modified as follows to allow for three-dimensional spatial data:

- (1) The delta function $\delta_{ib} = 1$ iff the i^{th} pixel is in the b^{th} equivalence class, 0 otherwise, and its stereo disparity is defined, and
- (2) the function $d(p)$ is introduced that maps a pixel at position p to its equivalent three dimensional value. The spatial moments then become:

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

To distinguish this formulation from spatiograms with image spatial dimensions, we will refer to this as a *terrain spatiogram*.

Figure 1 shows an image taken from one of the robots using the SRI SmallVision [8] software and Videre digital stereohead¹: Fig. 1(a) is the left video image, Fig. 1(b) is the disparity map, where brighter colors indicate greater disparity, and Fig. 1(c) shows video pixels mapped to a three-dimensional space with origin in the center of the left camera. Terrain spatiograms $R(b)$, $G(b)$ and $B(b)$ with $|B|=32$ were taken for the red, green and blue color channels of this scene and are shown in Figure 2: Fig. 2(a) shows a close-up of the scene but only showing video pixels for those points with a calculated stereo disparity; Fig. 2(b) a monochrome stereo disparity image of the scene; Fig. 2(c) the projection of the three spatiograms onto Cartesian XY space; and Fig. 2(d) the projection onto Cartesian YZ space. The X axis is horizontal in all. The spatiogram image is a composite of all three color channel spatiograms and is constructed by traversing the buckets of all three spatiograms and for each bucket value (r_b , g_b , b_b) drawing an ellipse of one standard deviation with that color.

3.3 Comparing Spatiograms

For a robot to recognize a landmark, it computes a terrain spatiogram of the landmark and then compares that spatiogram with the terrain spatiograms of a list of stored landmarks. The terrain spatiogram must be stored in a landmark-centred rather than robot-centred coordinate system. We employ a variant on the normalized spatiogram measure introduced by [12] to compare two terrain spatiograms h and h' :

¹ Model STH-MDCS3

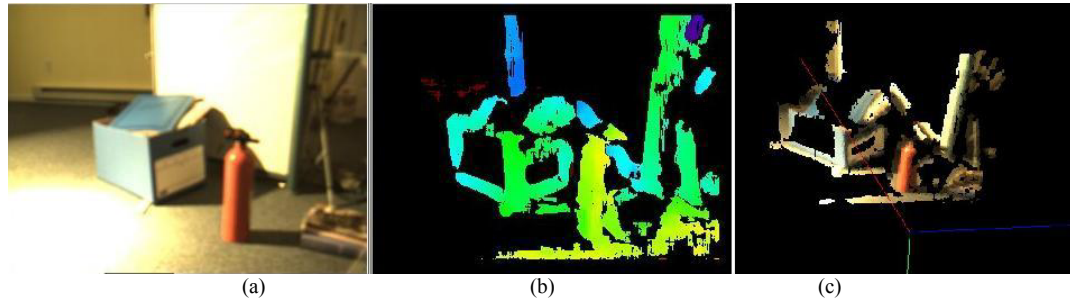


Figure 1: Stereo camera image (a), pseudocolor disparity image (b), depth image (c)

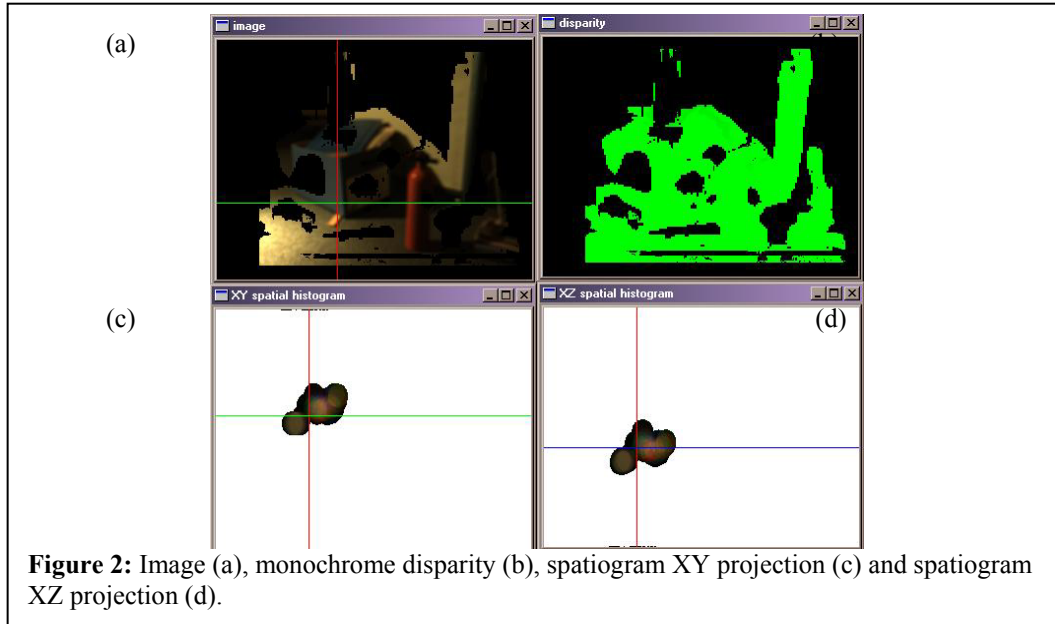


Figure 2: Image (a), monochrome disparity (b), spatiogram XY projection (c) and spatiogram XZ projection (d).

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

where

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

is the normalized probabilistic spatial weighting term.² If the set of stored terrain spatiograms is L then

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

is the stored landmark that best matches the observed landmark.

3.4 Mixture of Gaussian (MOG) Terrain Spatiograms

The usual spatiogram definition is unimodal. This can be a useful approximation; however, it is too simplistic to represent the appearance of urban search and rescue landmarks effectively. Furthermore, when unimodal terrain spatiograms from multiple sources (sensors or platforms) are combined the resultant may over-generalize and become less effective for landmark identification purposes. For both of these reasons we introduce a Mixture of Gaussian model.

A MOG terrain spatiogram is defined as:

² It can be easily verified that $\rho(h, h) = 1$.

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

where μ_{bi} , Σ_{bi} are the i th mixture parameters and α_{bi} is the weight or mode probability of the i th mixture. The probability for bin b of the spatial location x is given as usual as

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

An expectation-maximization (EM) algorithm can be used to select the mixture parameters when the terrain spatiogram is calculated from stereo information for example.

The definition of normalized similarity ρ , needs to be modified to accommodate the m_b component. We define the normalized similarity of a Gaussian spatiogram h , and a mixture of Gaussians spatiogram h^m as follows:

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

where

$$\psi_b^m = \sum_{i=1}^m \alpha_{bi} \eta_{bi}^m N(\mu_b; \mu_{bi}^m, 2(\Sigma_{bi}^m + \Sigma_b))$$

$$\eta_{bi}^m = 2(2\pi)^{0.5} |\Sigma_b \Sigma_{bi}^m|^{0.25}$$

The combination $\alpha_{bi} \eta_{bi}$ normalizes the difference over all mixture components. The mixture model also allows us a novel way to combine multiple views of a landmark from the same or different platforms without the issue of over-generalization: *each view contributes a single mixture to the MOG terrain spatiogram.*

4. EXPERIMENTAL RESULTS

4.1 Experimental Procedure

The experiments were conducted using two Pioneer AT3 robots and one Pioneer DX2 robot as follows:

1. AT3-1: Pioneer AT3 equipped with sonar and stereocamera (6mm lens) on a PTZ base;
2. AT3-2: Pioneer AT3 equipped with SICK laser ranger and PTZ camera;
3. DX2-1: Pioneer DX2 equipped with sonar and stereocamera (12 mm lenses) on fixed base.

The AT3-2 platform played a passive role in the experiments, acting as a robot landmark, and the other two robots were used to collect stereo depth information from which landmark terrain spatiograms were computed.

Landmark data collection was carried out as follows:

1. The AT3-1 platform was used to collect data on four outdoor landmarks composed of stacked construction debris. The robot was manually guided to the vicinity of the landmarks and each was manually windowed.
2. Since we expect that the members of the robot team will come into visual contact on occasion as they map the disaster site, the information about the relative position of the two platforms can be used to improved localization and mapping. The AT3-1 platform was driven in a one meter circle around the AT3-2 platform which functioned as a robot landmark. Stereo data was collected at four points on the circumference; the front left, back and right of the AT3-2 platform. The robot landmark was manually windowed.
3. In a separate location, the DX2-1 platform repeated these four measurements on the AT3-2 platform.

The terrain spatiograms for each landmark was constructed by first converting the RGB color values to normalized rgY values. Since the lighting conditions under which the three experiments were run were markedly different, the conversion to normalized rgY was necessary. Each landmark was used to produce three color channel spatiograms and the spatiogram images in the next section are constructed from the three color channel spatiograms as described in Section 3.2

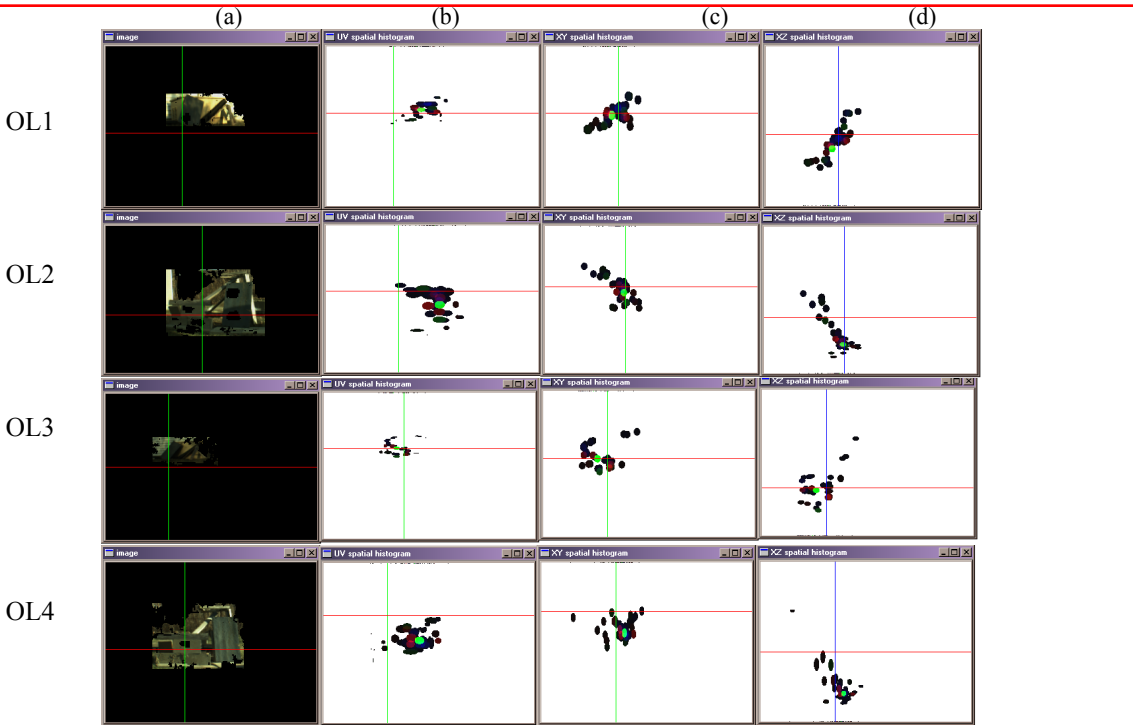


Figure 3: Outdoor landmarks OL1 through OL4. Column (a) is the video image, (b) is the image spatiogram; (c) is the XY projection and (d) is the XZ projection of the terrain spatiogram.

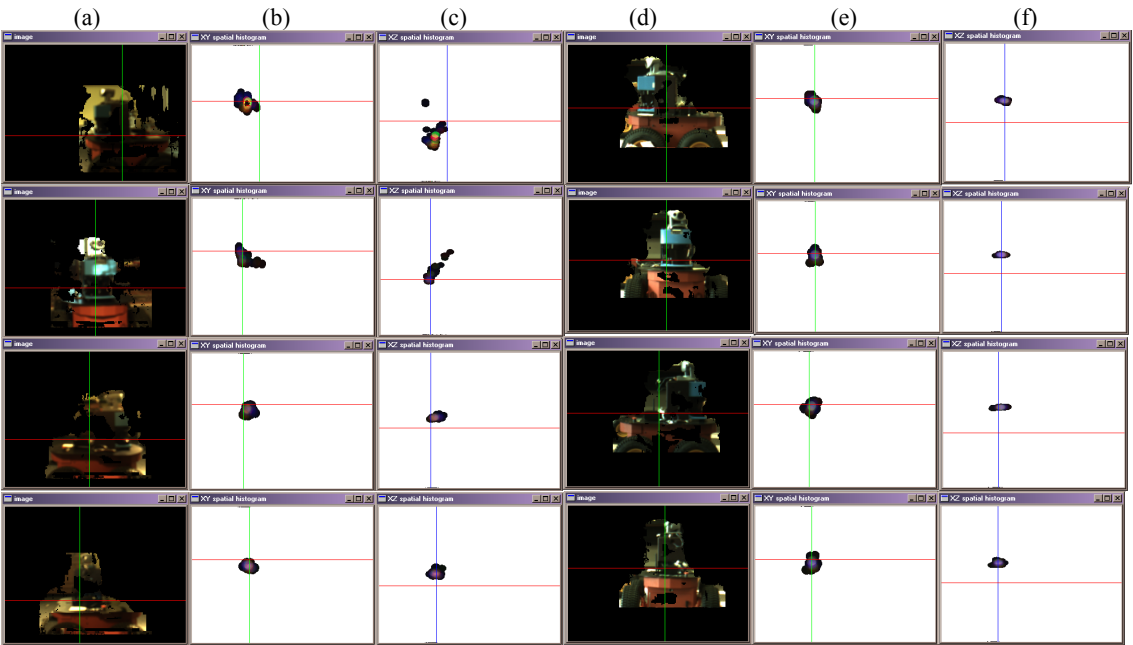


Figure 4: DX2-1 landmarks R1 through R4 (columns (a-c)) and AT3-2 landmarks R5 through R8 (columns (d-f)). Columns (a, d) are the video image; (b,e) are the XY projection and (c,f) are the XZ display of the terrain spatiogram.

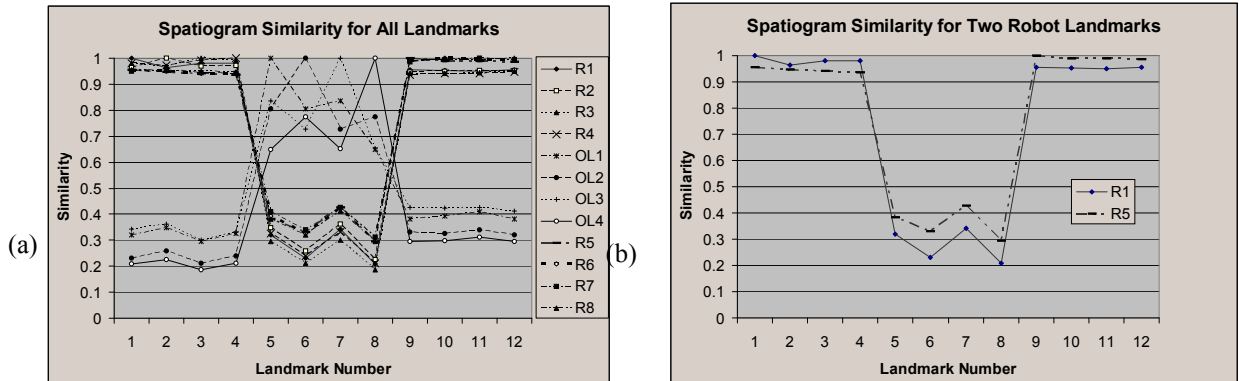


Figure 5: (a) Similarity values for all Landmarks (Landmarks 1-4 are R1-R4; 5-8 are OL1-OL4; 9-12 are R5-R8); (b) Similarity values for first DX2-1 and AT3-1 landmarks.

4.2 Landmark Comparison Results

The outdoor landmark terrain stereograms are shown in Figure 3. The video image is shown in the first column, where the black regions are places where no disparity data could be collected. The image spatiogram is shown in the second column. The XY and XZ projections of the landmark-centered terrain spatiogram are shown in the final two columns. Figure 4 (a-c) shows the four AT3-2 landmark terrain spatiograms collected by the DX2-1 platform and Figure 4 (d-f) shows the four AT3-2 landmark terrain spatiograms collected by the AT3-1 platform.

The 12 terrain stereograms (R1-4, OL1-4, R5-8) were compared to each other using the spatiogram similarity measure introduced in Section 3.3. The result is shown in Figure 5(a). The horizontal axis of the graph is the 12 landmarks in order (R1-4, OL1-4, R5-8). There is one line on the graph per landmark, and that line shows the similarity value of that landmark compared to each other landmark. There is a clear pattern in the graph:

1. All robot landmarks match each other quite well (>0.9) and match the outdoor landmarks quite badly (<0.44). There is much greater variability in the mutual outdoor landmark matches than there is in the matches between the robot landmarks.
2. The terrain spatiograms from the DX2-1 platform match the AT3-1 terrain spatiograms despite being taken by a different platform with different stereo cameras (and lenses).
3. The close up view (Figure 5(b)) shows that the first DX2-1 landmark matches itself best and next best matches the first AT3-1 landmark (i.e., the same robot pose in each). The same pattern is shown here for the first AT3-1 landmark.

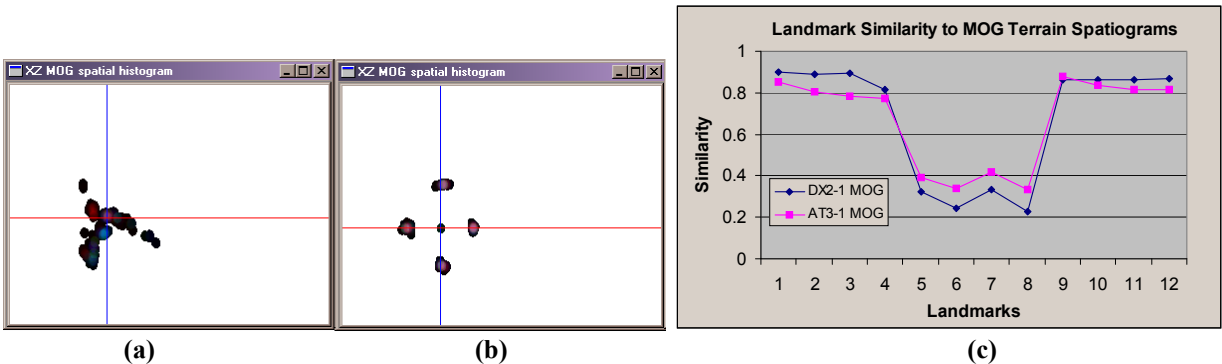


Figure 6: MOG Terrain Spatiogram for the four DX2-1 Landmarks (a); for the four AT3-1 Landmarks (b); normalized MOG similarity for all landmarks compared to these two MOG terrain spatiogram landmarks.

4.3 Landmark Fusion Results

The four landmarks from the DX2-1 were fused into a single MOG terrain spatiogram. Each landmark contributed one member of the mixture and the mixture coefficients were set equally to $\alpha_{bi}=0.25$. The second, third and fourth landmark terrain spatiograms were rotated by $\pi/2$, π , $3\pi/2$ respectively to model the robot pose when the data was collected. The four landmarks for the AT3-2 platform were similarly composed into a single MOG terrain spatiogram. The XZ projections of the two MOG terrain spatiograms are shown in Figure 6(a) and 6(b).

Each of the 12 landmarks was compared to the two MOG terrain spatiograms using the modified similarity expression in Section 3.4. The result is shown in Figure 6(c). The pattern is similar to Figure 5(a) in that the robot landmarks match each other better, no matter which platform they come from, than they match the outdoor landmarks. Note that the R1 and R5 landmarks match the MOG terrain spatiograms best - this is because all the other landmarks were rotated to put them in the right orientation in the MOG, and hence match the unrotated MOG less well. The DX2 and AT3 landmarks match both MOGS well, again supporting the benefit of this approach in sharing landmark information between platforms.

5. SUMMARY

In this paper we introduce an approach called terrain spatiograms to sharing landmark information between robot platforms. The approach combines 3D spatial information in a landmark-centered frame with image information. Experimental results are presented to show that landmark information can be effectively shared between two different robot platforms. The approach is extended with a mixture of Gaussian model and results are shown to demonstrate that this is an effective way to fuse multiple views of a landmark.

Future work will investigate the construction of terrain spatiograms from sonar and laser data as well as stereo data and the comparison of these with each other and with stereo data. All the single view landmarks in this paper were unimodal spatiograms. In future work we will evaluate the calculation of single view MOG spatiograms using an EM algorithm.

References

- [1] Stanley T. Birchfield, Sriram Rangarajan, Spatiograms Versus Histograms for Region-Based Tracking, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Diego, California, June 2005.
- [2] Stanley T. Birchfield, Sriram Rangarajan, Spatial Histograms for Region-Based Tracking, *ETRI Journal*, V29, N5, Oct. 2007.
- [3] Andreas Birk and Stefano Carpin Rescue Robotics - a crucial milestone on the road to autonomous systems, *Advanced Robotics Journal*, 20 (5), VSP International Science Publishers, 2006.
- [4] Briggs, A., Detweiler, C., Mullen, P., and Scharstein, D., Matching Scale-Space Features in 1D Omnidirectional Images. *Computer Vision and Image Understanding*, 103(3):184-195, September 2006.
- [5] B. Cartwright and T. Collett. Landmark learning in bees. *Journal of Comparative Physiology*, 151(4):521-543, 1983.
- [6] Dudek, G., Jenkin, M., Computational Principles of Mobile Robotics. Cambridge Press 2000.
- [7] Mark Fiala and Anup Basu, Robot Navigation Using Panoramic Landmark Tracking. *15th Int. Conf. on Vision Interface*, Calgary, Canada, May 27-29, 2002.
- [8] Konolidge, K., Beymer, D., SRI SmallVision User's Manual v4.4d May 2007.
- [9] Wen-Nung Lie, Tom C.-I. Lin, Ting-Chih Lin, Keng-Shen Hung, A robust dynamic programming algorithm to extract skyline in images for navigation, *Pattern Recognition Letters*, Volume 26, Issue 2, January, 2005.
- [10] Hanspeter A. Mallot, Matthias O. Franz, Biological Approaches to Spatial Representation: A Survey, *Robotics and Autonomous Systems*, 65, 1999.
- [11] Murphy, Robin R.; Burke, Jennifer L., Up from the Rubble: Lessons Learned about HRI from Search and Rescue, *Human Factors and Ergonomics Society Annual Meeting Proceedings, Cognitive Engineering and Decision Making*, pp. 437-441(5).
- [12] O'Conaire, C., Smeaton, A.F., An Improved Spatiogram Similarity Measure for Robust Object Localization. *IEEE Int. Conf on Acoustics, Speech and Signal Processing*, 15-20 March 2007.
- [13] J. O'Keefe and J. Dostrovsky. The Hippocampus as a spatial map: Preliminary evidence from unit activity in the freely moving rat. *Brain Res.*, 34:171-175, 1971
- [14] Ramos, F.T.; Nieto, J.; Durrant-Whyte, H.F., Recognising and Modelling Landmarks to Close Loops in Outdoor SLAM. *IEEE Int. Conf. on Robotics and Automation*, 2007.
- [15] Stephen Se, David Lowe, Jim Little, Local and Global Localization for Mobile Robots using Visual Landmarks, *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*, pp.414-420, 200.
- [16] Thrun, S., Burgard, W., and Fox, D., A Probabilistic Approach to Concurrent Mapping and Localization for Mobile Robots. *Machine Learning and Aut. Robots* 31/5 1998, pp. 1-25.