Sharing Landmark Information using
Mixture of Gaussian Terrain Spatiograms

Damian M. Lyons, *Member, IEEE*

Abstract— In this paper we evaluate the use of a novel spatial histogram, the terrain spatiogram, as a common representation for exchanging landmark information between robots working as a team to map an area. Individual robots use range sensors to provide the spatial dimension of the spatiogram and video for the image dimension. We have previously shown that terrain spatiograms can be shared between robots in a heterogeneous team to recognize landmarks and to fuse observations from multiple sensors or multiple platforms.

A terrain spatiogram using a mixture of Gaussians (MOG) model is introduced and a corresponding normalized spatiogram similarity measure defined. Two methods to generate a MOG terrain spatiogram are presented and compared experimentally using indoor and outdoor landmark information transferred between two different models of robot equipped with differently configured stereocameras.

I. INTRODUCTION

A team of robots working to collaboratively and quickly generate a map of a site showing hazards, obstacles, traversable routes, etc, will need to register their local maps. One effective way to do this is communicate landmark information to each other. In previous work [9], we proposed 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. Based on Birchfield and Rangarajan's image spatiogram [2], the terrain spatiogram represents image data and corresponding 3D terrain spatial information rather than image spatial information. We showed that this representation allows effective sharing of landmark information between differently equipped platforms. In this paper, we present a terrain spatiogram based on a mixture of Gaussians model. We introduce two methods to fuse multiple views in this model and present an experimental evaluation of each.

Previous work is reviewed in Section II of the paper. In Section III, we recap the terrain spatiogram notation from [9] and its extension to a mixture of Gaussians framework. Section IV presents two approaches to fusing information from multiple views in a mixture of Gaussians framework. In Section V the experimental evaluation of each is reported.

II. PRIOR WORK

One of the principle uses of a landmark in navigation and mapping is to allow the recognition of a place. Approaches to robot map representation include topological maps, maps based on places and their interconnection, and metric maps, maps based on accurate spatial measurements [16]. A *cognitive map* [12] is a biologically inspired, primarily topological map composed of *natural landmarks* [5] identifying places, the edges identifying routes between places and augmented with navigation and hazard information. In this framework, a robot needs to be able to select and recognize landmarks. Robots working as a team can function more efficiently if they can share landmark information, allowing them to fit their local maps together correctly and coordinate exploration and mapping activities. Other robots in the team can also serve as visual landmarks, leveraging a valuable, additional source of localization information.

Landmark selection refers to the process of determining which parts of the environment can function as effective landmarks and landmark recognition refers to the process of identifying previously selected landmarks. In this paper we restrict our attention to landmark recognition. The primary focus will be place recognition in a topological map. Micro-landmarks whose appearance is independent of scale and rotation, e.g., SIFT features [14], are commonly used in metric mapping. A collection of these micro-landmarks are matched to localize the robot accurately. The approach presented here, in contrast, uses a small number of macro-landmarks to recognize place within a topological map.

However, landmark recognition is also important in metric mapping for loop closure. Ramos et al. [13] shows that 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.

Another representation that combines depth and image information is Birchfield and Rangarajan [2]'s *spatial histogram* or *spatiogram*. The image spatiogram extends an image histogram with a Gaussian distribution per histogram bin that summarizes the image location for the image pixels that fall in that histogram bin. In [9], we note that 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 called the result a *terrain* spatiogram (as opposed to image spatiogram). We showed in [9] that a landmark could be represented by a stored terrain spatiogram in landmark-centered cylindrical coordinates and that this representation enabled effective landmark recognition on one robot of landmarks seen by another robot with a different sensor configuration.

However, the terrain spatiogram model proposed by [9] followed [2] in using a Gaussian distribution per bin, limiting how well it could represent outdoor landmarks, where colors could have multimodal distributions. In this paper, we present a reformulation of the terrain spatiogram model to include a mixture of Gaussian distribution per bin and we introduce present similarity measures for the new model.

III. THE TERRAIN SPATIOGRAM APPROACH

A. 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 $\{0,...,|P|\}$ such that $h_I(b)=n_b$ and

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

where δ_{ib} is equal to 1 iff the ith pixel is in the bth equivalence class and 0 otherwise, and η is a normalizing constant. 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:

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

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

Birchfield & Ragajaran define a histogram as a first order spatiogram, a formulation that also allows for second and higher order spatiograms. 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. [11] developed a normalized spatiogram comparison measure (one in which $\rho(h,h)=1$),

making it much more intuitive to use ρ to match two spatiograms.

B. 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 about the definition 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 function $d(p)$ is introduced that maps a pixel at position p to its three dimensional location in the viewed
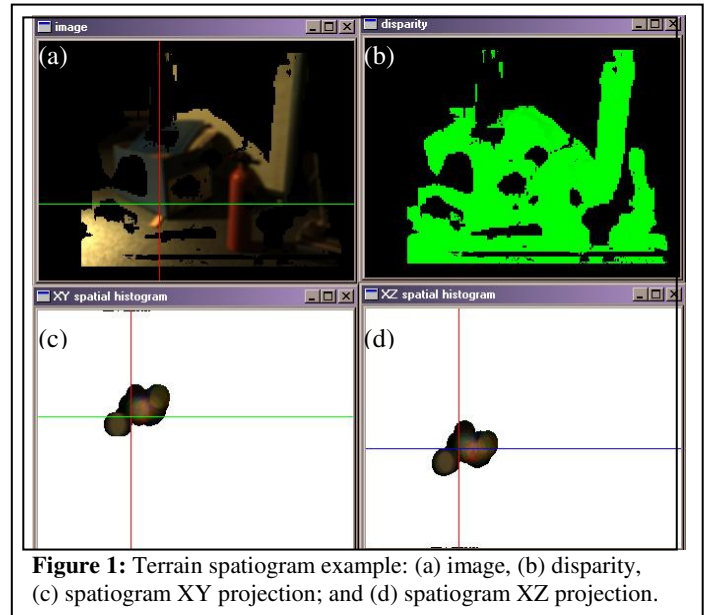

Figure 1: Terrain spatiogram example: (a) image, (b) disparity, (c) spatiogram XY projection; and (d) spatiogram XZ projection.

scene and the definition of the function δ_{ib} is modified so that $\delta_{ib} = 1$ iff the ith pixel is in the bth equivalence class and its stereo disparity is defined, 0 otherwise. The spatial moments for a *terrain spatiogram* then become:

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

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

Figure 1(a) shows an image taken from a Pioneer AT3 robot using the SRI SmallVision [7] software and Videre digital stereohead[1]. Fig. 1(b) is a monochrome disparity map. Figs 1(c,d) are an illustration of a terrain spatiogram calculated as follows: Terrain spatiograms $R(b)$, $G(b)$ and $B(b)$ with $|B|=32$ were taken for the red, green and blue color channels of this image. Fig. 1(c) is the projection of the three spatiograms onto Cartesian XY space; and Fig. 1(d) the projection onto Cartesian YZ space. The X axis is horizontal in all. The

[1] Model STH-MDCS3

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 Σ_b with that color.

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 spatial information must be landmark-centered rather than robot-centered [9]. We employ a variant on the normalized spatiogram measure introduced by [11] to compare two terrain spatiograms h and h':

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

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]

C. Mixture of Gaussian (MOG) Terrain Spatiograms

We argue that a unimodal bin distribution makes this representation less useful in representing the appearance of outdoor landmarks effectively for several reasons:

1. A multimodal color distribution can be a useful feature to distinguish a landmark in a complex outdoor scene (e.g., compare Fig. 2(c) and (d)).

2. When unimodal terrain spatiograms from multiple sources (sensors or platforms) are combined the resultant spatiogram may over-generalize and become less effective for landmark identification purposes.

A MOG terrain spatiogram is defined as:

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

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

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

The definition of normalized similarity ρ, needs to be modified to accommodate the m_b component. We define the normalized similarity of two mixture of Gaussian spatiograms h and h' as follows:

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

where we define ψ_b^{mm} and η_{bij} as

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

$$\eta_{bij} = 2(2\pi)^{0.5} |\Sigma_{bj} \Sigma'_{bi}|^{0.25}$$

---

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

Since a single Gaussian bin distribution is a special case of a mixture of Gaussians, we can also use this measure to compare the normalized similarity of a Gaussian spatiogram and a mixture of Gaussians spatiogram.

IV. CALCULATING MIXTURE OF GAUSSIAN SPATIOGRAMS

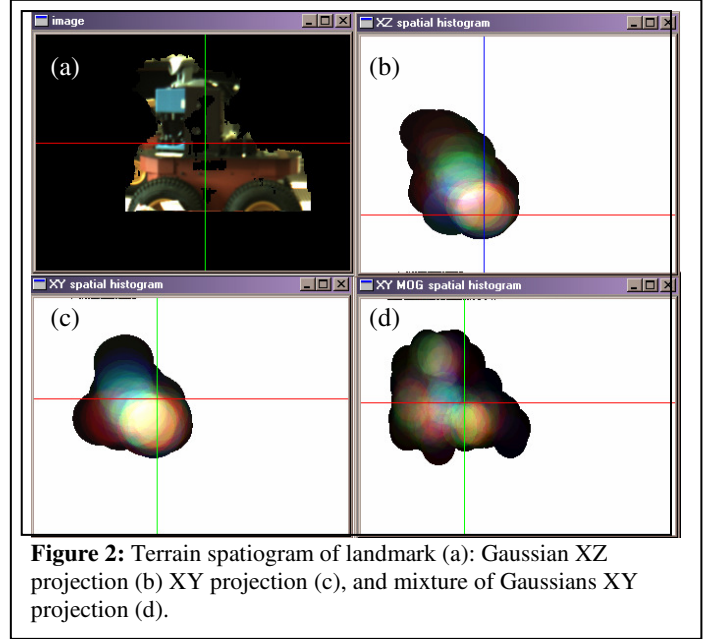Two approaches are proposed here to construct a mixture of Gaussians terrain spatiogram: clustering and fusion. The

Figure 2: Terrain spatiogram of landmark (a): Gaussian XZ projection (b) XY projection (c), and mixture of Gaussians XY projection (d).

next section presents an experimental evaluation of the proposals.

A. Clustering

To be useful in this application, a clustering approach needs to be fast. For example, a 3-channel 32 bin histogram needs to perform 96 clustering steps just to generate the spatiogram. For this reason, we propose a simple k-means based clustering.

1) **Cluster initiation.**
    1. Select a cluster center at random.
    2. Select furthest data point from this as next center.
    3. Repeat 2 until m cluster centers selected.

2) **K-means.**
    1. Assign each data point to its closest cluster.
    2. Recalculate clusters as centroids of assigned points.
    3. Repeat 1 and 2 until the average distance from a point to its cluster center does not change more than $\varepsilon=0.001$
    4. Calculate the variance of points in each converged cluster.
    5. Calculate the cluster weight as the number of points in the cluster divided by the total number of points.

Figure 2 shows an example of a mixture of Guassians terrain spatiogram calculated using this clustering method. Fig. 2(a) shows a Pioneer AT3 robot viewed by a second Pioneer DX2 robot using a stereocamera. Figs. 2(b) and (c) show the RGB terrain spatiogram in XY and YZ projections,

calculated as described for Fig. 1. Fig. 2(d) shows the mixture of Gaussian terrain spatiogram calculated using K-means in XY projection. Note that the yellow wheels show up as a single, centered point in Fig. 2(c) but as two distinct points in Fig. 2(d).
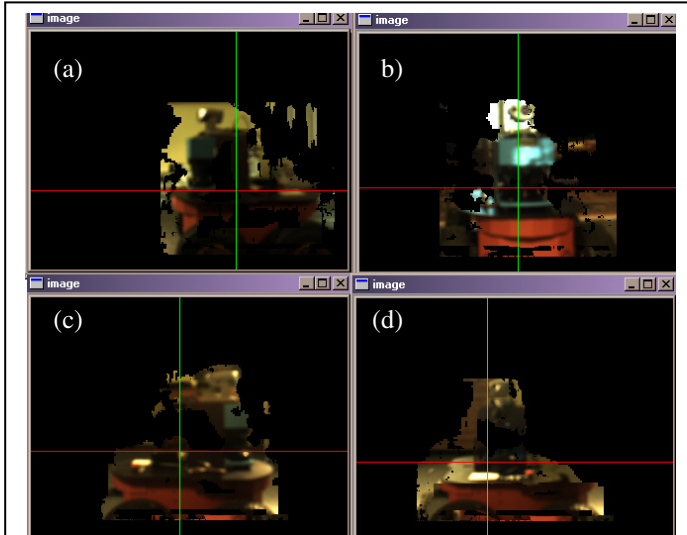

Figure 3: Four views of a Pioneer AT3 taken from Pioneer DX2 using stereocamera (black areas within image indicate no disparity).

Panoramic and omnidirectional cameras have been used in robot navigation for some time (e.g., [8] etc.) and there is evidence that panoramic processing is used in some kinds of insect navigation [4]. 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. However, to take advantage of this, we need to be able to combine multiple views into a terrain spatiogram.

On solution is simply to combine data from multiple views in the clustering process. Figure 3 shows four views of a Pioneer AT3 robot (left, front, right and back, 3(a-d) respectively) taken from a stereocamera. If the data from all
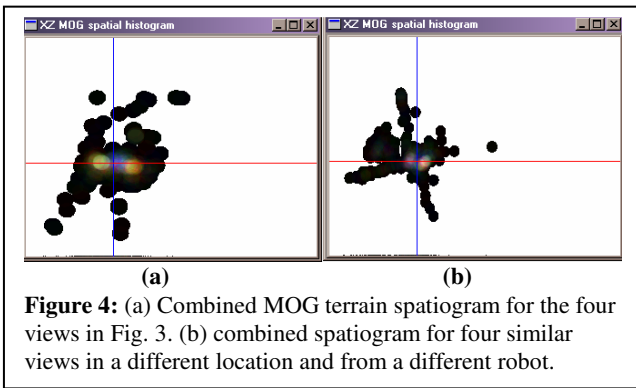

Figure 4: (a) Combined MOG terrain spatiogram for the four views in Fig. 3. (b) combined spatiogram for four similar views in a different location and from a different robot.

four views are clustered (with appropriate rotations of 0, $\pi/2$, π, $3\pi/2$ for each set of data), then the terrain spatiogram in Fig. 4(a) is the result. Fig. 4(b) shows a combination of four similar orthogonal views in a different location and from a

different robot. In this paper, we restrict our study to combining views taken by the same robot.

B. Fusion.

Aspect graphs represent a 3D object as a collection of views of the object [3]. Thus, another approach to build a mixture model for multiple views is to incorporate each view as a separate mixture member. The steps involved are as follows:

1) View collection.

    1.   Collect a single Gaussian terrain spatiogram per view, h_v.

    2.   Record the pose of the view in the landmark-based cylindrical coordinate frame, a_v.

2) View Fusion.

    1.   Translate the mean and variance of each bin in h_v by a_v.

    2.   Copy the modified h_v to the *vth* mixture

    3.   Repeat until all views/mixtures completed.

Figure 5(a) shows the result of fusing four Gaussian terrain spatiograms of the robot in Fig. 3. (The same views as used for Fig. 4(a)). Figure 5(b) shows the result of fusing four spatiograms of a similar robot in a different
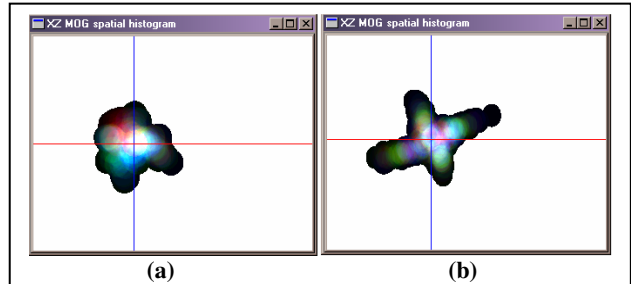

Figure 5: (a) Fused MOG terrain spatiogram for the four views in Fig. 3. (b) fused spatiogram for four similar views in a different location and from a different robot.

location and taken by a different robot. (Same views as Fig. 4(b)).

V. EXPERIMENTAL RESULTS

A. Experimental Procedure

The experiments were conducted using the same equipment as [9]: two Pioneer AT3 robots and one Pioneer DX2 robot as follows:

1. AT3-1: Pioneer AT3 equipped with a stereocamera (6mm lens) on a PTZ base;

2. AT3-2: Pioneer AT3 (passive target);

3. DX2-1: Pioneer DX2 equipped a 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. These landmarks are labelled OL1 to OL4. (See [9] for images and descriptions of these.)

2. 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. These landmarks are labelled R1 to R4.

3. In a separate location, the DX2-1 platform repeated these four measurements on the AT3-2 platform. These landmarks are labelled R5 to R8.

The landmark-centered terrain spatiograms for each landmark was constructed as follows:

1. The depth was sampled in an area of 20 pixels^2 around the image window center, and average depth established as the z origin of the landmark-centered frame.

2. The data was filtered by extracting only points within depth threshold z_{th} of the origin, $z < z_{th}$.

3. The RGB color values of these points were normalized to rgY values, since the lighting conditions under which the three experiments were run were markedly different.

4. Each landmark was used to produce three color channel spatiograms as described before.

B. Single-View Mixture of Gaussian by Clustering Results

A mixture of Gaussians terrain spatiogram was constructed for each landmark. The average time to calculate each color channel spatiogram for a 32 bucket, 4 mixture model was 0.051 seconds on a 1.4 GHz Pentium laptop. The four terrain spatiograms generated for the AT3-2 platform, R1 to R4, were compared to the outdoor landmarks OL1 to OL4 and to the landmarks taken from the DX2-1 platform, R5 to R8. The results are shown in Figure 6. R1 to R4 compare well to each other and to landmarks taken by the DX2-1 platform, R5 to R8 and are matched poorly to the outdoor landmarks. This again supports the thesis that terrain spatiograms are an effective way to share landmark information between different robot platforms. Note that the R1-R8 similarities in Fig. 6 are in fact lower than the single Gaussian similarities reported for the same landmarks in our previous work [9], where we reported that all robot landmarks match each other quite well (>0.9) and match the outdoor landmarks quite badly (<0.44). However, this is a reasonable result of the fact that the mixture of Gaussian

spatiograms represent the individual landmarks more accurately and hence allow for less generalization (and lower
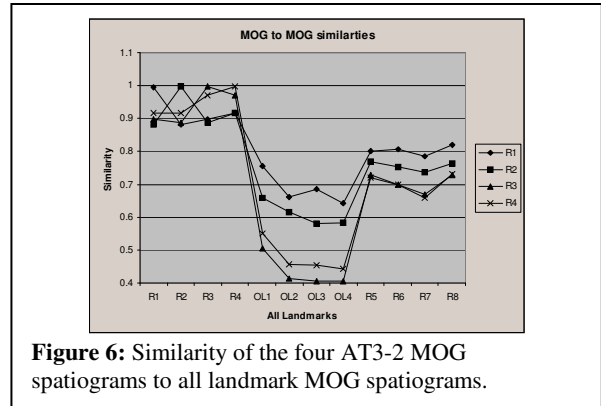

Figure 6: Similarity of the four AT3-2 MOG spatiograms to all landmark MOG spatiograms.

similarities) between robot landmarks.

C. Multiple-View MOG by Clustering Results

The R1 to R4 landmarks were combined into a single MOG spatiogram. The 12 landmark MOG spatiograms were compared to the combined MOG spatiogram. The results are shown in Fig. 7 (dashed line). While the robot landmarks are still distinguishable from the other landmarks, only R1 and R5 give good results. This is because all the other landmarks
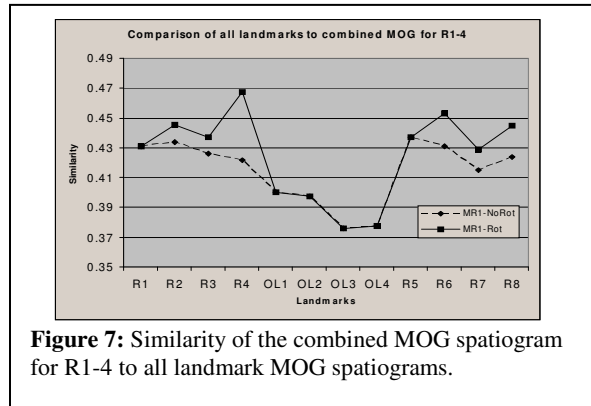

Figure 7: Similarity of the combined MOG spatiogram for R1-4 to all landmark MOG spatiograms.

were rotated when added to the combination, and hence are less similar. When the rotations are restored (Fig. 7. solid line) the matches are much better. The important implication
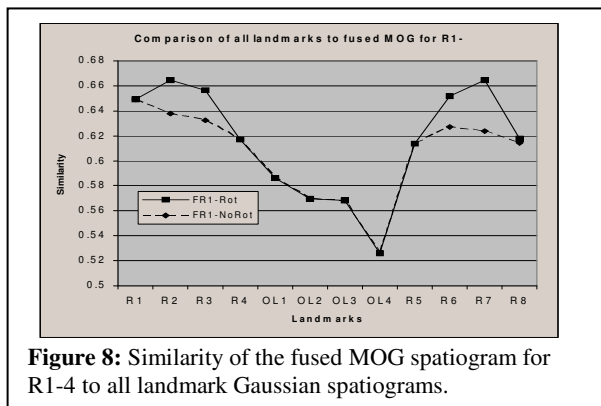

Figure 8: Similarity of the fused MOG spatiogram for R1-4 to all landmark Gaussian spatiograms.

is that the matching process may be able to yield landmark

orientation information. Problematically, however, the range of similarity values covering all landmarks is quite small.

D. Multiple-View MOG by Fusion of Aspects Results

The Gaussian spatiograms for R1 to R4 were rotated and added as mixture members to a fused MOG spatiogram. Each of the 12 landmark Gaussian spatiograms was compared to the MOG (treating a Gaussian spatiogram as a special case of a mixture with one member). The results are shown in Figure 8 (dashed line). There is good separation between the robot landmarks and other outdoor landmarks, but again R1 and R5 are the best matches. When the comparison is made with rotated landmark spatiograms, we get the solid line in Fig. 8.

VI.   DISCUSSION

Terrain spatiograms combine 3D spatial information from the environment with image information for landmark recognition of map places, loop closure in SLAM and for sharing information between mobile platforms working together to map a site. Previously we have shown that terrain spatiograms using on a single Gaussian per bin allowed effective communication of landmark information between two differently configured robots viewing the landmark under different conditions. In this paper we have introduced a terrain spatiogram with a mixture of Gaussians model per bin. This is arguably a more useful way to uniquely identify outdoor landmarks. We looked at two ways to populate this more complex model: a fast K-means based-clustering, and an aspect graph inspired approach. Our results show that

1. The MOG by clustering approach also allows effective sharing of landmark information between robots, but the increased specificity of the MOG representation means there is less generalization between views of a landmark that for the single Gaussian case.
2. The combination of many views into a single MOG spatiogram remained effective, and indicates that it may be possible to get not only similarity but also orientation from the comparison. However, the range of similarity values was small.
3. Surprisingly, the aspect graph inspired fusion approach retained all the same value as the clustering approach, but generated a wider range of similarity values.

There are a number of areas of future work suggested by these results. The ability to get both similarity and pose information begs a comparison with SIFT approaches using a much larger landmark set. A key part of such a study must be the robustness of both approaches to landmark occlusions; whereas occlusion might cause some SIFT micro-landmarks to drop out, reducing accuracy, will occlusion cause a macro-landmark terrain spatiogram to fail to be recognized at all? We propose to study how well the depth information available in a terrain spatiogram can be used to filter potential occlusions.

All the landmarks used in this study and in [9] were manually windowed and centered. Future work will add an automatic landmark extraction phase. Existing image-based approaches to selection include the *sum of absolute differences* [17] and *turn and look back* [1] methods, and methods that combine image and range data include [10].

Future work will also investigate the construction of terrain spatiograms from laser data as well as stereo data and the comparison of these with each other and with stereo.

REFERENCES

[1] Giovanni Bianco Alexander Zelinsky Miriam Lehrer, Visual landmark learning, *International Conference on Intelligent Robots and Systems* (IROS), 2000, Page(s):227 – 232.
[2] Stanley T. Birchfield, Sriram Rangarajan, Spatial Histograms for Region-Based Tracking, *ETRI Journal*, V29, N5, Oct. 2007.
[3] Cyr, C., Kimia, B., A Similarity-based Aspect-Graph Approach to 3D Object Recognition, *Int. Conf. on Computer Vision*, Vancouver, Brit. Columbia, Canada, July 7-14, 2001
[4] B. Cartwright and T. Collett. Landmark learning in bees. *Journal of Comparative Physiology*, 151(4):521 543, 1983.
[5] Dudek, G., Jenkin, M., Computational Principles of Mobile Robotics. Cambridge Press 2000.
[6] Hanspeter A. Mallot, Matthias O. Franz, Biological Approaches to Spatial Representation: A Survey, *Robotics and Autonomous Systems*, 65, 1999.
[7] Konolidge, K., Beymer, D., SRI SmallVision User's Manual v4.4d May 2007.
[8] 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.
[9] Lyons, D. M., "Sharing and Fusing Landmark Information in a team of Autonomous Robots" *SPIE Defense and Security Symposium: Multisensor, Multisource Information Fusion*, April 13-17, Orlando, FL 2009.
[10] Murrieta-Cid, R., Parra, C., Devy, M., Visual Navigation in Natural Environments: From Range and Color Data to a Landmark-based Model. *Autonomous Robots V13, N2, 2002.*
[11] O'Conaire, C., Smeaton, A.F., An Improved Spatiogram Similarity Measure for Robust Object Localization. *IEEE Int. Conf on Acoustics, Speech & Signal Proc.*, 15-20 Mar. 2007.
[12] 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
[13] 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.
[14] 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.
[15] 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.
[16] Thrun, S., Robot mapping: A Survey. Technical report CMU-CS-02-111, School of Computer Science, Carnegie-Mellon University, Pittsburgh PA Feb. 2002. In: *Exploring Artificial Intelligence in the New Millenium*, (Eds. Lakemeyer, G. and Nebel, B.) Morgan Kaufmann 2002.
[17] Craig Watman, David Austin, Nick Barnesy, Gary Overett and Simon Thompson "Fast Sum of Absolute   Differences Visual Landmark Detector," *IEEE Int. Conf.   on Robotics and Automation.* 26 April-1 May 2004 pp. 4827- 4832.