2006 IEEE International Conference on
Multisensor Fusion and Integration for Intelligent Systems
September 3-6, 2006, Heidelberg, Germany

MoB01.4

Integration of Multiple Sensors using Binary Features in a Bernoulli Mixture Model

F. Ferreira and V. Santos
Department of Mechanical Engineering
University of Aveiro, 3810-193 Portugal
Email: cfferreira,vsantos@mec.ua.pt

J. Dias
Institute of Systems and Robotics
University of Coimbra, 3030-290 Portugal
Email: jorge@isr.uc.pt

Abstract— **This article reports on the use of a Bernoulli Mixture model to integrate features extracted independently from two or more distinct sensors. Local image features(SIFT) and multiple types of features from a 2D laser range scan are all converted into Binary form and integrated into a single binary Feature Incidence Matrix(FIM). The correlation between the different features is captured by modeling the resultant FIM in terms of a Bernoulli Mixture Model. The integration of binary features from different sensors allows for good place recognition. The use of binary features also promises a much simpler integration of features from dissimilar sensors.**

I. Introduction

We are interested in implementing an algorithm for place recognition for a mobile robot equipped with different sensors. While the use of features obtained from laser range scans and from images is common, we have attempted to simultaneously utilise features extracted from both types of sensors.

To perform integration of data from multiple sensors we have first represented the features from each sensor in binary form. Further processing of this binary data is required to capture the correlations that are present in the data; correlations that must be taken into account when we deal with noisy perception.

A. Feature-Based methods using Range Sensors and Vision

Feature extraction from a laser range scan attempts to detect distinct properties of a portion of the scan or of the whole scan. These must be [relatively] invariant to changes in the scan view point(the place from which the scan was taken), to the presence of view-obstructing objects and to [small]changes that occur in dynamic, real-world environments.

Dudek and MacKenzie in [1] provide a method by which different real-world objects are modeled as 2D *lines* and the scan data is segmented into lines matching these corresponding objects. The extraction of lines from the laser scan continues to be a popular approach in the segmentation of laser scan data, see [2] and [3] for recent reviews of line-extraction algorithms. Ribeiro and Gonçalves in [4] utilise vertical *edges* (corners in the 2D laser scan) to obtain mobile robot localisation estimates. In [5], Arsenio and Ribeiro mount a Laser Range Finder (LRF) atop a Pan and Tilt Unit (PTU) so as to obtain 3D landmarks consisting of vertical edges. In outdoor environments too, range scanners have been utilised to segment features. Manandhar and Shibasaki in [6] extract road, buildings, tunnels and other outdoor features by modeling 3D range data.

A different approach which eschews segmentation into simple primitive features favours the description of a section or sections of the 2D Laser scan in some reduced variable space. This is the approach used in [7] where each feature extracted from the laser range scan is given a symbol and each scan is described in the form of a string for example mMmMmMmMmDCm(local maxima (M), a discontinuity (D), a local minima (m), or a connection (c)). Other methods use 'sections' of the laser range scan so as to minimise the effect that changes in one part of the scan will have on other parts.

Feature extraction for vision-based robots vary from local-image descriptors to properties derived from the intensity distribution over an entire image.

The use of local-image descriptors descriptors is characterized by two steps 1)the selection of points of interest and 2) their characterisation. The selection must be repeatable (even with changes in the conditions in which the images are taken) and the characterisation must be done using properties that, again, must be tolerant to changing viewpoints, lighting and other conditions.

Local-image features based on local image gradients are an important class of vision features. Baker in [8] attempts to create a generalised descriptor for local image features. The introduction to his thesis provides a perspective on the development of gradient based methods. The stability and repeatability of points extracted at local Maxima (or Minima) in gradient images that have been repeatedly smoothed using operators, has been known for some time [9] [10], and finally culminated in the Scale-Space theory proposed by Lindeberg[11]. In work that combined the lessons of Scale-Space with the reliable characterisation of features, Lowe [12] describes the use of gradient histograms taken at various points close to the point of interest. Our work utilizes these SIFT images features and we have introduced a modification that allows us to create and use these features for a sequence of images.

Another approach is to calculate properties (e.g. moments of some value) of interest within a region of the image as in [13]. A commonly used approach is to attempt to capture the distribution of intensities in the colour space using a histogram [14]. Along with area-based methods there are other contour based methods which seek to code the properties of contour of a regions and use novel ways to

1-4244-0567-X/06/$20.00 ©2006 IEEE.

match the these properties [15], [16].

B. Integration Of Features and Sensor Fusion

Vision sensors can be coupled with range sensors to aid the segmentation of scan data. Arras and Tomatis, [17] attempt to add a CCD camera to a robot already having a localisation system based on a LRF. The stand-alone LRF-equipped system achieves good performance in rooms in which the environment is made up of distinct features. With the aid of a vertical-edge extraction from the images, situations in which the LRF is prone to provide unpredictable or highly ambiguous data are reduced.

Baltzakis et al. in [18] propose a method in which images of the environment are taken simultaneously with 2D laser scans. A [simple] 3D model of the environment is built out of each range scan and the displacement of the robot in this 3D model is estimated by verifying the transformation required to match two laser scans. This 3D model is then verified in those places where the transformations are consistent with those estimated for the images.

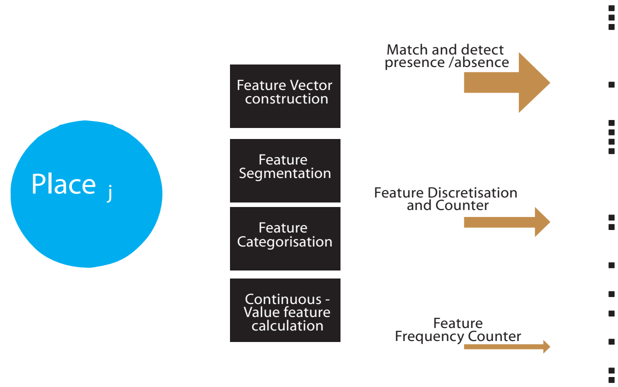
Other approaches have been proposed to combine sensors and, given the breadth of features in vision and range scans, the combinations are many (see the bibliography maintained by Keith Price at http://iris.usc.edu/Vision-Notes/bibliography/match-pl502.html).

Perception usually results in noisy data which is further compounded by a 'dynamic' environment. Gathering 'good' or unchanging features is one way of ensuring that there is some chance of localisation. It is even more important to build redundancy into the perception process. Place recognition, image retrieval and robot localisation methods (even with a single sensor) typically make use of a large number of features. Having to reduce the dimension of data from a sensor or from multiple sensors in order to make the procedures more tractable and robust to noisy data is a common problem. A solution to the problem is to perform data fusion according to some model that creates composite features (originating from the same or from different sensors and after registering the data of one sensor with another). This approach seeks to impose correlations that are suspected to exist in the data due to a particular environment.

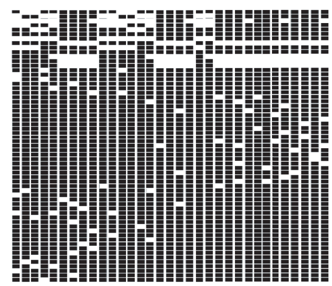
On the other hand a data-driven approach will attempt to extract the [strongest] correlations that are present in the data. Methods that reduce the dimension of features with continuous values are common in many perception fields including face recognition, speech recognition etc. Among these approaches, Mixture Models are a common solution to modeling data that is thought to a follow non-parametric distribution. Sajama and Orlitsky in [19] demonstrate the use Mixture models composed of Gaussian, Bernoulli and exponential distributions as a solution to the classification problem. Mclachlan and Peel's book [20] provides a good reference to the general topic of Finite Mixture Models.

C. Layout of this article

This section has named a few of the approaches that utilise feature extraction to address the problem of localisation.



(a) Creation of Vector of Binary Features.



(b) Feature Incidence Matrix where each column describes a place and each row a feature.

Fig. 1. Binary Features are created in different ways for different features.

The following sections will detail the features for vision, SIFT features in Section II, and an assortment of features from laser range scans in Section III. The actual method of modeling the features in terms of a Bernoulli Mixture Model to capture the correlations between features is then presented in Section IV. Finally, Section V lists ongoing work and the future scope of this approach.

As shown in Fig. 1(a) We have extracted features using different methods and converted them into binary form by one of the following:

- matching extracted features against a feature database to detect their presence(or absence)
- categorising features
- discretizing features with continuous values

We end up with a matrix of binary values Fig. 1(b), where each row denotes a particular feature that was extracted from one or more image or laser range scan. The presence of a 'one' in any column represents that the feature was observed in an image or laser scan taken at that place.

II. SIFT Features From Images

Since their introduction, SIFT features have been widely applied among others to object recognition [21] [22], in the panoramic assembly of images [23] and in image retrieval [24]. SIFT has also been used for robot localisation as in [25] [14] [26] [27].

Two factors affect the efficacy of the SIFT descriptors (for that matter, of any other descriptor). These are 1) the repeatability of the point extractor and 2) the robustness of the descriptor itself to changes in the viewpoint, orientation and changes in lighting, scale, etc.

Once the points are extracted from the images using a good corner extractor (see [28] for an evaluation of different extractors), the creation of the SIFT keys from the gradient histograms taken over a sufficiently wide area results in a long, robust descriptor of the local point.

While the strength of the SIFT features stems from the long descriptors, this same property presents new challenges since each image easily throws up hundreds of good features and calculating the distance between the vectors of each of these features with the vectors of the features obtained from any other image is computationally very demanding. For this reason Lowe[29], suggested the utilisation of a data structure, the KDTree, to perform matching. The principal advantage of the KDTree lies in its ability to quickly retrieve points even when these have descriptors of very large dimension. It is relatively fast to construct and when implemented efficiently (see [30] for an open-source implementation), turns the matching of SIFT features into a feasible task.

We have experimented with extracting between fifty and two hundred features per image from sequences containing up to three hundred and fifty VGA-size images taken as the robot moves along a path. One problem that we came up against was the fact that very similar images produced many features with similar SIFT keys. In certain situations this can prevent us from correctly building a KDTree. Since the KDTree must be built all at once (no accepted method exists for incrementally adding data to a KDTree) we adopted the following simple procedure to aid the construction of the KDTree. The procedure in Algorithm. 1 allows the trouble free construction of the tree and does not alter the way in which the SIFT features are retrieved and used. The added noise was small (less than 0.5%) and no significant degradation in performance of the tree at the time of retrieval of the points was verified.

Algorithm 1 Create KDTree

$N =$ total number of points to insert
$n =$ number of points already inserted
$F =$ number of points to insert at a time (per image possibly)
Require: $N \geq 2$
while $(N - n > 0)$ **do**
  Add tiny amount (less than 0.5%) of random noise to F points
  Query tree for F points to be inserted(without noise)
  Mark matched points in total of $n+F$ points for removal
  Add marked points as aliases of points still in tree.
  Destroy KDTree
  Create new KDTree only with unique points
end while
Destroy KDTree
Create new KDTree only with unique points and without noise

(a) Image 3. (b) Image 19. (c) Image 104.

Fig. 2. Three images from a sequence of 118 taken by Camera 1 along a hallway.

(a) Image 3. (b) Image 19. (c) Image 104.

Fig. 3. Three images from a sequence of 118 taken by Camera 2 along a hallway.

Figs 2(a) through 2(c) represent a sequence of 118 images taken along a hallway with the camera (Camera 1) looking into the hallway. The robustness of the SIFT features to changes in the scale and orientation of the images is seen in the way some features persist and appear over a large number of views as verified in Fig. 5.

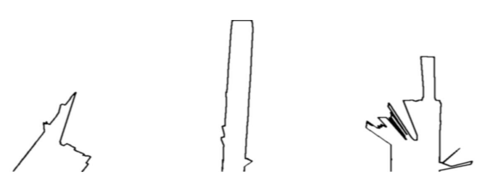
Figs 3(a) through 3(c) represent a sequence of 118 images taken along a hallway with the camera (Camera 2) looking onto one side of the hallway.

III. Features from Laser Range Scans

Unlike the features for vision, which are all of the same type(SIFT), we have used multiple types of features from the laser range scan, namely 1)wall-like, line features, 2) scan region properties and a 3) scan contour properties in the form of a vector that characterises 2D discontinuities in the plane of the scan using Hu moments[20].

A. Line Features

We extracted long lines from the points in the laser scan using a two-stage method. The extraction of short line segments (containing between 2-6 points) was performed using the incremental method, shown by [2] to be one of the effective methods for this type of application. These short lines are then fused into longer lines (segments of at least 2 meters) wherever possible. Binary features from the extracted line segments are created by classifying the number of extracted lines and their distance from the Laser range scan as shown in Table I.



(a) Laser Scan 3. (b) Laser Scan 19. (c) Laser Scan 104.

Fig. 4. Three laser range scans from a sequence of 118 laser scans.

Feature Number	Description
0	>= 1 lines at 4+ meters
1	>= 2 lines at 4+ meters
2	>= 3 lines at 4+ meters
3	>= 1 lines at 2+ & 4- meters
4	>= 2 lines at 2+ & 4- meters
5	>= 3 lines at 2+ & 4- meters
6	>= 1 lines at 2- meters
7	>= 2 lines at 2- meters
8	>= 3 lines at 2- meters

TABLE I

CATEGORISATION OF LONG LINES INTO BINARY FEATURES

Feature Number	Description
0	area >= 2 m ^{2}
1	area >= 4 m ^{2}
2	area >= 8 m ^{2}
3	$MaxDim$ >= 4meters
4	$MaxDim$ >= 8meters
5	$MaxDim$ >= 16meters
6	$MaxDim/MinDim$ >= 1
7	$MaxDim/MinDim$ >= 4
8	$MaxDim/MinDim$ >= 8

TABLE II

DISCRETISATION OF AREA AND PRINCIPAL DIMENSIONS INTO BINARY FEATURES.

B. Scan Boundary Features

We have employed features denoting regional properties including area enclosed within the scan boundaries and the ratio between the directions of maximum length and of minimum length. The values of the area and the lengths of the principal dimensions are then classified into the corresponding class as specified in Table II.

C. Hu Moment Features

We have also employed novel vector features that seek to describe the shape of the scan contour using Hu Moments[13]. This is a technique that is employed in the matching of images and is based on the observation that combinations of the centralised moments of an image are (quite) invariant to rotation, scale and reflection[20]. We have used the Hu image moments in a similar way to SIFT features -by collecting the 7-element features into a KDTree and matching the features extracted from new images.

IV. Place recognition using the Bernoulli Mixture-Model

Articles such as [31] and [32] go some way to demonstrate the usefulness of binary features. In [33] the context in which a word is used in a sentence is converted into multiple binary features. Similarly [34] and [35] seek to model training data as a sample of sets of binary features taken from a population of binary features, each distributed according to a mixture of Bernoulli distributions.

By gathering all the binary features into a single Feature Incidence Matrix(FIM), we obtain a representation of the features that were viewed at each place in the environment. As can be seen from a comparison of the Binary Feature Incidence Matrices for Camera 1, Camera 2 and the Laser Range Finder, in Fig. 5, the Features from laser and the camera 2 are fewer(per scan/image) but quite uncorrelated with the features from camera 1. This is an interesting observation whose usefulness will be confirmed in the results of the integration of the FIMs and application to the place recognition problem in the next section.

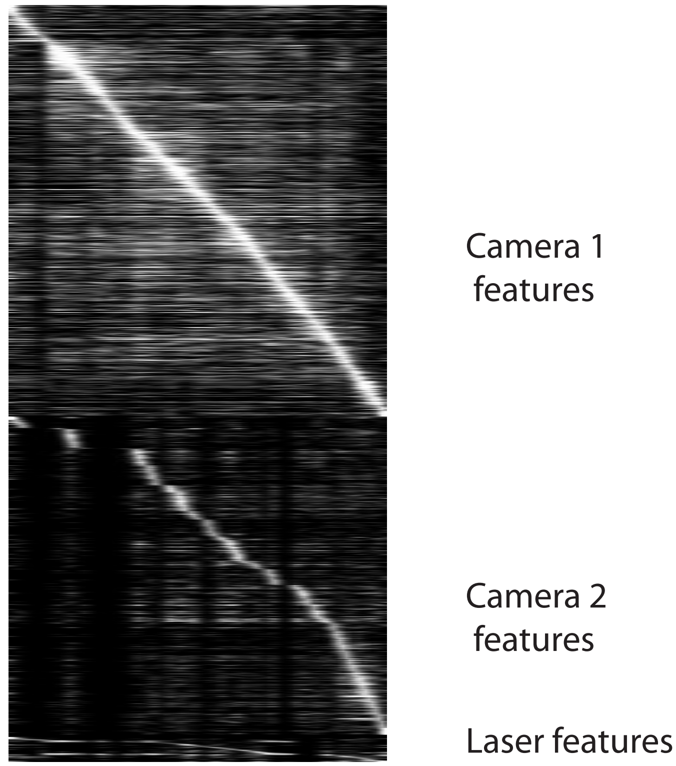


Fig. 5. The Feature Incidence Matrix For features from the Camera 1, Camera 2 and Laser Range Finder

A. Formulation of the Bernoulli Mixture-Model

Mixture models assume that there exists a finite number of distributions which, when mixed together in a particular proportion, result in a distribution that best describes the data we wish to characterise. The probability of an observed view V^{obs} can be written as in (1).

$$P(V^{obs}|\Theta) = \sum_{i=1}^{M} \alpha_i P_i(V^{obs}|\Theta_i) \tag{1}$$

The variables α_i, which represent prior probabilities of that component in the mixture model, are subject to the constraint $\sum_i \alpha_i = 1$. V^{obs} consists of vectors with binary values $\{0,1\}^N$ (there are N features in all) and Θ denotes the parameters of the distribution of the Mixture Model. Each component Θ_i is a multivariate vector of Bernoulli probabilities, of size N . If we have K views that were gathered during the training stage then ,

- Θ is an $N \times M$ matrix. Each column of Θ, Θ_i is a component of the mixture. Each component Θ_i is composed of N features.

- α is an $1 \times M$ vector. Each column of α, α_i is the proportion of mixture that is attributed to component Θ_i.

Most texts express the product in (1) in terms of a sum by employing the log-likelihood (2).

$$\mathcal{L}(\Theta|\mathcal{V}) = \sum_{k=1}^{K} log(\sum_{i=1}^{M} \alpha_i P(V^{obs}|\Theta_i)) \qquad (2)$$

The procedure of choice to obtain the parameters of the Mixture Model is the Expectation Maximisation(EM) method. The EM method [36] applied to the Mixture problem assumes that the data is only partially available. It becomes fully known through the use of a vector of coefficients denoted henceforth as the 'missing data', the 'hidden data' or the 'observed data', Z. We can now express the likelihood of the observations given the entire data as in (3). Z is a $M \times K$ matrix. Each column of Z, z_k is the vector that is applied to the components (besides the appropriate mixing coefficient) and allows us to obtain an observation in the training data set.

$$\mathcal{L}(\Theta|\mathcal{V}, Z) = \sum_{k=1}^{K} \sum_{i=1}^{M} z_{ki}(log(\alpha_i) + log(P(V^{obs}|\Theta_i))) \qquad (3)$$

B. Matching the Observed View, V^{obs}

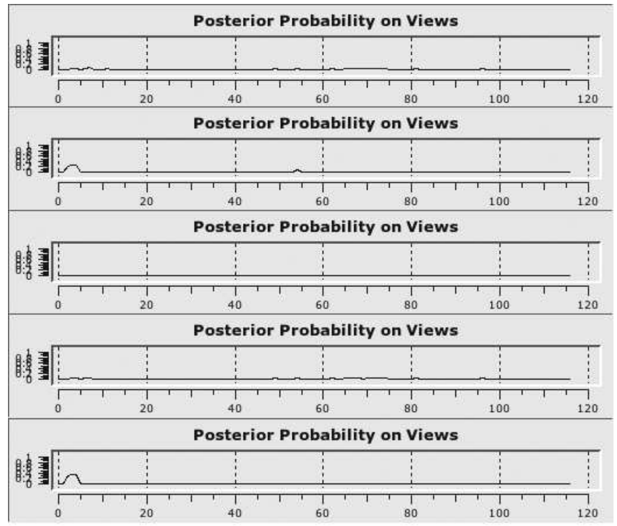
Our goal in performing this entire exercise is to match the observed view from the working sequence with one of the index views in the reference sequence. To achieve this we use the 'Classification interpretation' of the Mixture Model where both, the mixture parameters Θ and the hidden data Z, are used so that any observed view can be classified into one or more groups defined by the views that made up the original training data.

We choose to use the *Maximum Likelihood Estimation* method, a commonly used approach, to select the most probable place in the distribution given by (4). Wherever available, the prior probability of being at any state i can also be incorporated (or the states can be assumed to be equi-probable)

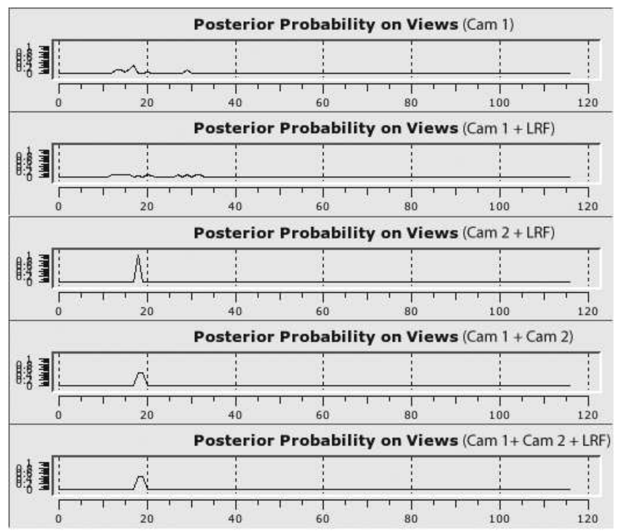
$$P(V=i|V^{obs}) = \frac{\sum_{j=1}^{M} P(V=i) z_{ij} \alpha_j P(V^{obs}|\Theta_j)}{\sum_{k=1}^{K} \sum_{j=1}^{M} P(V=k) z_{kj} \alpha_j P(V^{obs}|\Theta_j)} \qquad (4)$$

Fig. 6 shows the result of matching a second series of images(the working sequence) with the Bernoulli Mixture Model calculated from the Binary Feature Incidence Matrix in Fig. 5. The results of using different combinations of sensor data have been presented to demonstrate the effect of the integration of the different features.

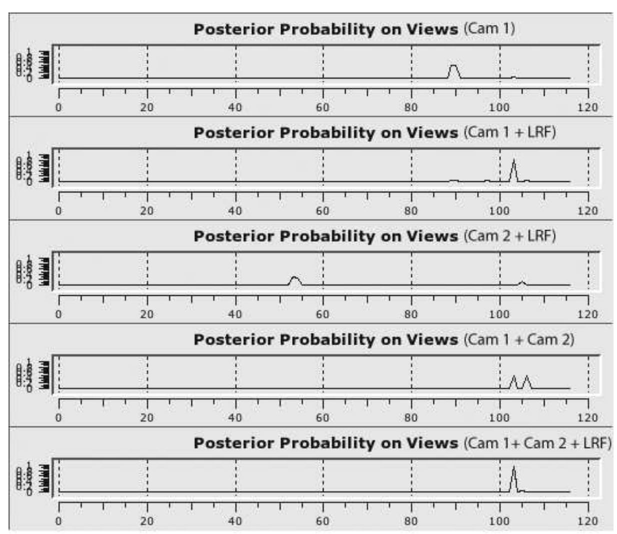
The results of matching are quite good given that no additional constraint was imposed on the matching and that the prior probability for matching any image was assumed to be uniform. As seen in the first situation, Fig. 6(a), while the features from Camera 1 and Camera 2 are not able to localize effectively, the features from the laser range sensor



(a) Posterior Probability for Image/Scan 2 of working sequence.



(b) Posterior Probability for Image/Scan 9 of working sequence.



(c) Posterior Probability for Image/Scan 52 of working sequence.

Fig. 6. Posterior Probabilities over the places in the Reference sequence for combinations of sensors.

help produce a good estimate when combined with the other sensors. In the case of Fig. 6(b), the presence of many unique features from Camera 2 help us to recognize the place correctly. Finally in Fig. 6(c), while no individual sensor was certain enough of a place(the combination of Camera 2 and the Laser Range Finder was way off) the combination of the three sensors resulted in the correct estimation of the place.

V. Conclusions and Scope of Future work

We have presented a method by which features from multiple sensors might be integrated through an intermediate step of conversion into binary features. We have utilized a Bernoulli Mixture Model to model the distribution of these binary features whose parameters are obtained using the well-known EM algorithm. Place recognition using binary features from vision was demonstrated to be quite good given unknown (uniform) prior probabilities.

Further development of the method is needed; for example a means to handle the probability of error in the binary data and the uncertainty that is inherent in individual sensor models must be developed. We intend to apply the method to a more varied set of features from additional sensors. Yet another issue involves the effect of having a dissimilar number of features from each sensor. In such a situation it is possible that a greater number of (less reliable) features might overwhelm a smaller number of features from another (more accurate) sensor.

References

[1] Gregory Dudek and Paul MacKenzie. Model based map construction for mobile robot localization'. In *Proceedings of Vision Interface*, pages 97–102, Toronto, Ontario, July 1993.

[2] Viet Nguyen, Agostino Martinelli, Nicola Tomatis, and Roland Siegwart. A comparison of line extraction algorithms using 2d laser rangefinder for indoor mobile robotics. In *International Conference on Intelligent Robots and Systems*, 2005.

[3] Daniel Sack and Wolfram Burgard. A comparison of methods for line extraction from range data. In *IAV 2004*, 2004.

[4] Maria Isabel Ribeiro and J Goncalves. Natural landmark based localisation of mobile robots using laser range data. In *IEEE EUROBOT 96*, 1996.

[5] Artur Arsenio and Maria Isabel Ribeiro. Active range sensing for mobile robot localization. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robotic Systems*, Victoria, Canada, October 1998.

[6] Dinesh Manandhar and Ryosuke Shibasaki. Feature extraction from range data. In *ACRS 2001 - 22nd Asian Conference on Remote Sensing*, volume 2, pages 1113–1118, Singapore, November 2001.

[7] James Fellers Sooyong Lee, Nancy M. Amato. Localization based on Visibility Sectors using Range Sensors. In *Procceedings of the IEEE Int. Conference Robot. Autom. (ICRA),*, pages 3505–3511, Jan 2000.

[8] Simon Baker. *Design and Evaluation of Feature Detectors*. PhD thesis, Columbia University, 1998.

[9] Jan J Koenderink. The Structure Of Images. *Biological Cybernetics*, 50(5):363–370, 1984.

[10] A.P.Witkin. Scale-space filtering. In *Proc. 8th Joint conference on Artificial Intelligence*, pages 1019–1023, Karlsruhe, W. Germany, 1983. 8th Joint conference on Artificial Intelligence.

[11] Tony Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Press, 1994.

[12] David G. Lowe. Object Recognition from Local Scale-Invariant Features. In *Proc. of the International Conference on Computer Vision, Corfu*, pages 1150–1157, 1999.

[13] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *IEEE Transactions on Information Theory*, 8(2):179– 187, February 1962.

[14] Henrik Andreasson and Tom Duckett. Topological Localization For Mobile Robots Using Omni-Directional Vision And Local Features. In *5th IFAC/EURON Symposium on Intelligent Autonomous Vehicles*, Instituto Superior Tenico, Lisboa, Portugal, July 2004. IFAC/EURON.

[15] Jukka Iivarinen, Markus Peura, Jaakko Särelä, and Ari Visay. Comparison of combined shape descriptors for irregular objects. In *British Machine Vision Conference*, 1997.

[16] Carmen de Trazegnies Otero. *3D Object Learning and Recognition System Based on Planar Views*. PhD thesis, Universidad de Málaga, Málaga, Spain, 2004. ISBN:84-689-5286-9.

[17] K. Arras and N. Tomatis. Improving robustness and precision in mobile robot localization by using laser range finding and monocular vision. In *3rd European Workshop on Advanced Mobile Robots*, 1999.

[18] Haris Baltzakis, Antonis Argyros, and Panos Trahanias. Fusion of range and visual data for the extraction of scene structure information. In *Proceedings of the International Conference On Pattern Recognition*, pages 40007–40012, Quebec, Canada, August 2002.

[19] Sajama and Alon Orlitsky. Supervised dimensionality reduction using mixture models. In *Proceedings of the 22 nd International Conference on Machine Learning*, Bonn, Germany, 2005.

[20] Geoffrey McLachlan and David Peel. *Finite Mixture Models*. John Wiley and Sons, 2000.

[21] Arthur Pope and David G. Lowe. Probabilistic Models of Appearance for 3-D Object Recognition. *IJCV*, 40(2):149 – 167, 2000.

[22] David G. Lowe. Local Feature View Clustering for 3D Object Recognition. In *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 682–688, Kauai, Hawaii, 2001.

[23] Matthew Brown and David G. Lowe. Recognising Panoramas. In *Tenth International Conference on Computer Vision (ICCV 2003)*, October 2003.

[24] Yan Ke, Rahul Sukthankar, and Larry Huston. An efficient parts-based near-duplicate and sub-image retrieval system. In *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*, pages 869–876, New York, NY, USA, 2004. ACM Press.

[25] Stephen Se, David Lowe, and Jim Little. Mobile Robot Localization and Mapping with Uncertainty using Scale-Invariant Visual Landmarks. *International Journal of Robotics Research*, 21(8):735–760, August 2002.

[26] David M. Bradley, Rashmi Patel, Nicolas Vandapel, and Scott M. Thayer. Real-Time Image-Based Topological Localization in Large Outdoor Environments. In *Proceedings of the 2005 IEEE Conf. on Intelligent Robotics and Systems*, August 2005.

[27] Nicolas Cuperlier, Mathias Quoy, and Philippe Gaussier. Navigation and Planning in an Unknown Environent Using Vision and a Cognitive Map. In *In Proceedings of The 2006 European Robotics Symposium*, 2006.

[28] Cordelia Schmid, Roger Mohr, and Christian Bauckhage. Comparing and Evaluating Interest Points. In *ICCV*, pages 230–235, Bombay India, 1998. Narosa Publishing House.

[29] David G. Lowe. Distinctive Image Features-From Scale-Invariant Keypoints. *IJCV*, 60, Issue 2(Issue 2):91–119, November 2004.

[30] Matthew B Kennel. KDTREE 2: Fortran 95 and C++ software to efficiently search for near neighbors in a multi-dimensional Euclidean space, 2004.

[31] Ata Kaban and Mark Girolami. Initialized and Guided EM-Clustering of Sparse Binary Data with Application to Text Based Documents. In *Proceedings of the 15th International Conference on Pattern Recognition (ICPR'00)*, volume 2, 2000.

[32] Xin Wang and Ata Kaban. Finding Uninformative Features in Binary Data. In *Sixth International Conference on Intelligent Data Engineering and Automated Learning (IDEAL '05)*, 2005.

[33] J. Gonzalez, A. Juan, P. Dupont, E. Vidal, and F. Casacuberta. A Bernoulli Mixture Model for Word Categorisation. In *Symposium Nacional de Reconocimiento de Formas y Analises de Imagenes*, 2001.

[34] Alfons Juan and Enrique Vidal. Bernoulli Mixture Models for Binary Images. In *International Conference on Pattern Recognition (ICPR'04)*, volume 3, pages 367–370, 2004.

[35] José García-Hernández, Vicent Alabau, Alfons Juan, and Enrique Vidal. Bernoulli mixture-based classification. In A. R. Figueiras-Vidal et al., editor, *Proc. of the LEARNING04, ISBN 84-688-8453-7*, Elche (Spain), October 2004.

[36] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, Series B,34:1–38, 1977.