Research at the Learning and Vision Mobile Robotics Group 2004–2005

J. Scandaliaris, R. Alquézar, J. Andrade, J. Aranda, J. Climent, A. Grau, J.-M. Mirats, F. Moreno, F. Serratosa, J. Vergés, T. Vidal, A. Sanfeliu

Institut de Robòtica i Informàtica Industrial, UPC-CSIC

Llorens Artigas 4-6, 08028 Barcelona, Spain

www-iri.upc.es/groups/lrobots

lrobots@iri.upc.edu

Abstract

This article presents the current trends on wheeled mobile robotics being pursued at the *Learning and Vision Mobile Robotics Group* (IRI). It includes an overview of recent results produced in our group in a wide range of areas, including robot localization, color invariance, segmentation, tracking, audio processing and object learning and recognition.

1 Introduction

The Learning and Vision Mobile Robotics Group, an interdisciplinary team of researchers, is the product of a joint effort between the *Institut de Robòtica i Informàtica Industrial* and the *Departament d'Enginyeria de Sistemas, Automàtica i Informàtica Industrial* at the *Universitat Politècnica de Catalunya*, and the *Departament d'Enginyeria Informàtica i Matemàtiques* at the *Universitat Rovira i Virgili*.

Headed by Prof. A. Sanfeliu, as of today, it embraces 5 professors, 2 posdoctoral associates, and 7 PhD students. The group, consolidated in 1996, has given rise to 3 PhD thesis and 7 final year projects.

Financial support comes mainly from a continued set of projects from the former Interministerial Council of Science and Technology (CICYT), DPI2001-2223 and DPI2004-05414.

2 Current Research Areas

During the last year, our efforts have been tailored at giving our mobile platforms the ability to navigate autonomusly in unknown structured settings. In this sense, we have continued our contributions to the classical simultaneous localization and map building problem, focusing on the observability and stability problems [1, 2, 3, 12, 13].

Moreover, very good results have been achieved in dealing with the problem caused by varying illumination. On the one hand, we have developed histogram based techniques for visual tracking and for the robust description of images taken from a mobile robot. [10, 11]. On the other hand, we have proposed a new figure-ground segmentation technique for image sequences with varying illumination conditions based on the integration of shape and color information [5, 6].

With respect to 3D object recognition and identification, we have formalized and validated a random graph (RG) representation for modelling sets of attributed graphs. We call this new formulation *second-order random graphs* (SORGs). A basic feature of SORGs is the inclusion of second-order relations between nodes and arcs. This new representation allows to describe better the features of 3D objects[7, 8, 9].

In the following sections we summarize the key contributions from five selected publications [3, 4, 6, 8, 10]. Each of them tackles very different problems typically encountered in mobile robotics applications.

3 Unscented transformation of vehicle states in SLAM

The most widely accepted tool for solving the SLAM problem is the Extended Kalman Filter (EKF). One drawback however with the use of the EKF, is in the linear propagation of means and covariances. Vehicle and sensor models are usually of a very high nonlinear nature, and the effects of linearization required in the EKF can lead to filter divergence.

This situation has prompted the use of particle filters for a non parametric approximation of vehicle and map probability density functions in SLAM. Particle filters approximate the state space by random sampling the posterior distribution, and may require many samples to accurately model the nonlinear effects in both vehicle and measurement models. A middle ground is to use a deterministic approach for the nonlinear propagation of means and covariances. One such solution is the use of the Unscented Kalman Filter (UKF). An unscented transformation is similar to a particle filter in that it samples the pdf, but instead of doing it randomly, a careful selection of deterministic sigma points is made so as to preserve the moments of the distribution.

Deterministically choosing the particles is a computationally efficient solution for the nonlinear propagation of means and covariances, but doing so for the full state vector in SLAM may not be appropriate. There is no need to use particles in the computation of the map prior, given its linear nature. Thus, by using the Unscented Transformation (UT) only for the vehicle states we are able to reduce the computational complexity (compared to a full UKF), and to produce, at the same time, tighter covariance estimates (See Figure 1). The detailed formulation on how to compute the UT only for the vehicle states within the EKF approach to SLAM is in [3].

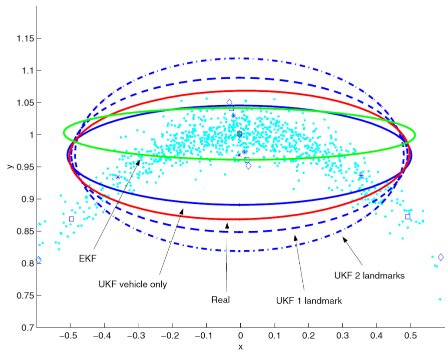
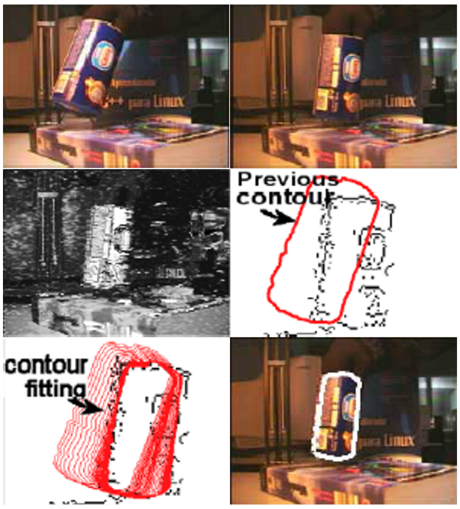


Figure 1: In red (x): actual mean and covariance; in green ($+$): linear transformation of mean and covariance; in blue ($*$): nonlinear transformation of mean and covariance for robot only; in blue dashed ($\Box$): nonlinear transformation for robot and one landmark; in blue dash dot ($\diamond$): nonlinear transformation for robot and 2 landmarks.

4 Figure-ground segmentation in non-stationary environments based on shape and a multi-hypothesis Fisher color model

Color is a visual cue that is commonly used in computer vision applications such as object detection and tracking tasks. In environments with controlled lighting conditions and uncluttered background, color can be considered a robust and invariant cue, but when dealing with real scenes with changing illumination and confusing backgrounds, the apparent color of the objects varies considerably over time. In the literature, the techniques that cope with change in color appearance can be divided in two groups. On the one side, there are approaches searching for color constancy; but in practice, these methods work on artificial and highly constrained environments. On the other hand, there are the techniques that generate a stochastic model of the color distribution, and adapt it over time.

The drawback in all these approaches is that they assume a smooth and slow color variation that can be predicted by a dynamic model based in only one hypothesis (usually by a weighting function). However, this as-



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

Figure 2: Results of an abrupt change of illuminant and object position (commented in text).

sumption does not suffice to cope with general scenes, where the dynamics of the color distribution might follow an unknown and unpredictable path. In [5], we have suggested the use of a multihypothesis framework to track color objects in such situations, where the color could be approximated by an unimodal distribution. In [6], we deal with multicolored objects.

In order to cope with unconstrained environments, we propose a system with the following main features, that represent contributions with respect to previous works:

- Fisher color model: Instead of using the classical RGB, normalized color rgb, HSV or XYZ color spaces, we propose the use of a color space efficient for the discrimination between foreground and background classes, based on the 2D projection of the R, G and B components on the plane obtained from a nonparametric LDA.

- Multihypothesis framework: The use of a particle filter formulation to predict the color distribution in following iterations, offers robustness to abrupt and unexpected changes in the color appearance.

- Integration of color and shape: The fusion of both vision modules is done in a final stage introduced to the classical CONDENSATION algorithm, and makes our method suitable to work in cluttered scenes.

The results of one experiment where there is and abrupt change of both illumination and object position are shown in Figure 2. Fig. 2a and 2b are the two consecutive frames. At least one of the multiple hypothesis of color distributions performs a good a posteriori probability map (Fig. 2c) that is used to fit the contour (Fig. 2d, 2e). For all the relevant details about this technique, see [6].

5 A color constancy algorithm for the robust description of images collected from a mobile robot

This work presents a framework for describing a set of images taken at nearly the same pose in a simple, generic and compact way. A requirement imposed on any such descriptor is that it should be representative of the actual content of the set at pose p and at the same time be unaffected by the presence of variations.

There are three main sources of variations. The first one is the pose p, that has to be relatively constant among the set of images. A second source of variations is the possibility of objects appearing, disappearing or just changing their relative positions in the scene. A third source is due to the objects' relative positions change with respect to that of light sources, or even due to temporal changes of the sources themselves.

To cope with the first two sources of variations we employ the α-trimmed average histogram to integrate the images into one descriptor by analogy to the problem of representing a group of frames in a video by their color histograms. The interest of this aggregation scheme is that of reducing the effects of outliers such as those belonging to objects that appear or disappear in the frames embodied in the set of images while easily encompassing either the mean or the median by tuning the α parameter. Additionally, the use of histograms minimizes the importance or relative positions of objects in the scene.

The third source of variability, illumination, needs special consideration. When taking images from a mobile robot, no control on the illumination can be done and the images belonging to the same set of images may even have been taken at different times, implying big light variations which affect the color of the objects in the scene (See Fig. 3). Our aim is to find the most likely color transformation T which maps the pixel colors of image I_a as close to those of image I_c as possible, hence reducing the color variation in the set of images. For determining this transformation the algorithm first calculates the set of all feasible mappings. This set consists of all the mappings taking each of the colors present in one image to each of the colors present on a reference image. The best mapping is then selected by maximizing a likelihood function. This likelihood function depends both on the probability of each mapping, which is estimated from the histogram of the set of mappings, and on a simmilarity measure between the histograms of the two images. Figure 4 shows the prefformace of the algorithm in one of the scenes used in the experiments. Details about the error measure and the technique just described can be found in [10].

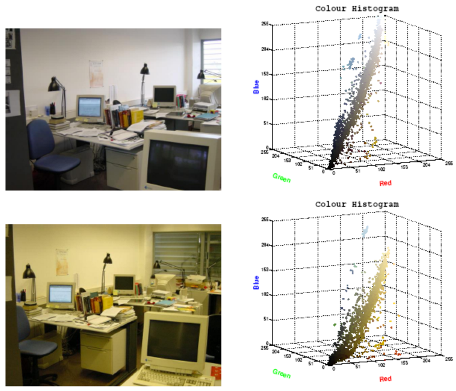


Figure 3: Images taken nearly at the same pose.

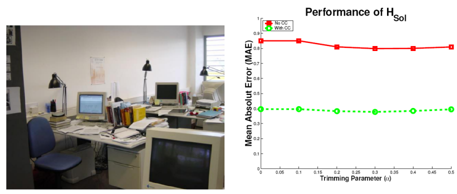


Figure 4: Scene and performance plot.

6 Second-order random graphs for modeling sets of attributed graphs and their application to object learning and recognition

Attributed Graphs (AGs) has been used to solve computer vision problems for decades and in many applications. Some examples include recognition of graphical symbols , character recognition, shape analysis, 3D-object recognition and video and image database indexing. In these applications, AGs represent both unclassified objects (unknown input patterns) and prototypes. Moreover, these AGs are typically used in the context of nearest-neighbour classification. That is, an unknown input pattern is compared with a number of prototypes stored in the database. The unknown input is then assigned to the same class as the most similar prototype.

The main drawback of representing the data and prototypes by AGs is the computational complexity of comparing two AGs. The time required by any of the optimal algorithms may in the worst case become exponential in the size of the AGs. For applications dealing with large databases, this may be prohibitive.

To alleviate these problems, some attempts have been made to try to reduce the computational time of matching the unknown input patterns to the whole set of models from the database. Assuming that the AGs that represent a cluster or class are not completely dissimilar in the database, only one structural model is defined from the AGs that represent the cluster, and thus, only one comparison is needed for each cluster. Random Graphs have been defined as a good representation of multiple AGs. A RG can be obtained by the union of the AGs, according to some synthesis process, together with its associated probability distribution. In this manner, a structural pattern can be explicitly represented in the form of an AG and an ensemble of such representations can be considered as a set of outcomes of the RG. Within these methods, three of most important are First-Order Random Graphs (FORGs), Function-Described Graphs (FDGs) and Second-order Random Graphs.

6.1 Matching views of 2D projections of 3D objects by random graphs

The idea is to represent 2D views of a 3D object by means of random graphs and then to obtain the model as the synthesis from the graphs that represent the 2D views of a 3D object. Once the model has been learned, the recognition process is based on applying a distance measure among the input graph (the graph that encodes the 2D view of a scene object) and the object models. The input graph is assigned to the model graph with the minimum distance measure value. We apply this method to learn and recognize objects from the COIL-100 from Columbia University. We did the study with 100 isolated objects, where each one is represented by 72 views (one view each 5 degrees). The test set was composed by 36 views per object (taken at the angles 0, 10, 20 and so on), whereas the reference set was composed by the 36 remaining views (taken at the angles 5, 15, 25 and so on).

The learning and recognition process was as follows: (1) perform colour segmentation of each individual object view image; (2) create an adjacency graph for each one of the segmented regions of each object view; and (3) transform the adjacency graph in an attributed graph (AG) using the hue feature as the attribute for each node graph. The learning process was based on 36 views of each object and for each object, we synthesise four random graphs, the first of one grouping the views from 0^o to 90^o, the second one grouping from 95^o to 180^o and so on. We used four different techniques for the representation of random graphs: AG (Attributed Graph), FORG (First Order Random Graph), FDG (Function Described Graph) and SORG (Second Order Random Graph). Fig. 5 shows 20 objects at angle 100^o and their segmented images with the adjacency graphs. The best result appears when the SORG and FDG representations were used, although the best is the SORG representation. Fig. 6 shows the ratio of recognition success of the 100 objects using different object representation and distance measures. This figure also shows the result of describing individually each object view by means of an AG and then comparing each input AG against the rest of the prototype AG. The reader is directed to [8] for all the relevant details.

Figure 5: Some objects at angle 100 and the segmented images with the AGs

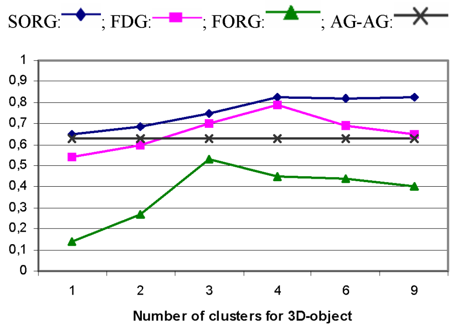


Figure 6: Ratio of recognition correctness of the objects using SORG, FDG, FORG and AG-AG.

7 Non-speech sound feature extraction

Sound offers advantages for information systems in delivery of alerts, duration information, encoding of rapidly incoming information, representing position in 3-D space around the person and her localization. Hearing is one of human beings most important senses. After vision, it is the sense most used to gather information about our environment. Despite this, little research has been done into the use of sound by a computer to study its environment. The research that has been done focuses mainly on speech recognition, while research into other types of sound recognition has being neglected. In robotics, non-speech audio has been ignored in front of artificial vision, laser beams and mechanical wave sensors beyond the audible spectrum. But the study and modeling of non-speech audio can help greatly to robot navigation and localization in the space domain. The existing research in non-speech sound is incipient and focuses on signal processing techniques for feature extraction with the use of neural networks as a classification technique. In this work a new technique based on pattern recognition techniques in order to locate a robot in the space domain by non-speech audio signals is proposed. The feature space will be built with the coefficients of model identification of audio signals. Due to their non-stationary property wavelet decomposition is needed as a preprocessing step. We also propose a technique (transform function) to convert the samples in the feature space into the space domain, based in the sound derivative partial equation.

We propose a new localization in space domain approach from non-speech audio signals that will be applied on a robot in an industrial environment, the approach follows the next steps: 1) measurement and data preprocessing. 2) MAX models signals identification by the wavelet transform; 3) feature selection, feature extraction and its correspondence with the space domain. Non-speech audio signal generated by any audio source (industrial machinery, appliances, etc.) is continuous by its nature. Preliminary, non-speech signal preprocessing includes sampling the analog audio signal with a specific frequency and to convert it into a discrete set of samples. Sampling interval should be chosen in such a way that essential information be preserved.

Some experiment have been carried out with MARCO robot in a real environment with a CNC milling machine as a non-speech audio source. The rate of spatial recognition for unknown positions are close to the 95% in average respect their actual position.

With the methodology used in this work we have achieved some interesting results that encourage the authors to keep on walking in this research field. The introduction of more that one audio source is also a new challenge. The experimental results show a narrow correspondence with the sound physical model and this demonstrates a high reliability of the proposed methodology. This work has been presented in [4].

References

[1] J. Andrade-Cetto and A. Sanfeliu. The effects of partial observability in SLAM. In *Proc. IEEE Int. Conf. Robot. Automat.*, pages 397–402, New Orleans, Apr. 2004.

[2] J. Andrade-Cetto, T. Vidal-Calleja, and A. Sanfeliu. Multi-robot C-SLAM: Simultaneous localization, control and mapping. In *Proc. IEEE ICRA05 Workshop Network Robot Syst.*, Barcelona, Apr. 2005.

[3] J. Andrade-Cetto, T. Vidal-Calleja, and A. Sanfeliu. Unscented transformation of vehicle states in SLAM. In *Proc. IEEE Int. Conf. Robot. Automat.*, pages 324–329, Barcelona, Apr. 2005.

[4] Y. Bolea, A. Grau, and A. Sanfeliu. Non-speech sound feature extraction based on model identification for robot navigation. In A. Sanfeliu and J. Ruiz-Schulcloper, editors, *Progress in Pattern Recognition, Speech and Image Analysis*, volume 2905 of *Lect. Notes Comput. Sci.*, pages 723–731, Havana, Nov. 2003.

[5] F. Moreno-Noguer and A. Sanfeliu. Adaptive color model for figure-ground segmentation in dynamic environments. In A. Sanfeliu and J. Ruiz-Schulcloper, editors, *Progress in Pattern Recognition, Speech and Image Analysis*, volume 2905 of *Lect. Notes Comput. Sci.*, pages 37–44, Havana, Nov. 2003.

[6] F. Moreno-Noguer and A. Sanfeliu. Integration of shape and a multihypotheses fisher color model for figure-ground segmentation in non-stationary environments. In *Proc. 17th IAPR Int. Conf. Pattern Recog.*, volume Vol.4, pages 771–4, Cambridge, UK, Aug. 2004. IEEE Comp. Soc.

[7] A. Sanfeliu. The use of graph techniques for identifying objects and scenes in indoor building environments for mobile robots. In *Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR Int. Workshops, SSPR 2004 and SPR 2004*, volume 3138 of *Lect. Notes Comput. Sci.*, pages 30–41, Lisbon, Portugal, Aug. 2004. Springer-Verlag.

[8] A. Sanfeliu, F. Serratosa, and R. Alquezar. Second-order random graphs for modeling sets of attributed graphs and their application to object learning and recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(3):375–96, 2004.

[9] F. Serratosa and A. Sanfeliu. Distance measures between attributed graphs and second-order random graphs. In *Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR Int. Workshops, SSPR 2004 and SPR 2004*, volume 3138 of *Lect. Notes Comput. Sci.*, pages 1135–44, Lisbon, Portugal, Aug. 2004. Springer-Verlag.

[10] J. Verges-Llahi and A. Sanfeliu. A color constancy algorithm for the robust description of images collected from a mobile robot. In *Progress in Pattern Recognition, Image Analysis and Applications*, volume 3287 of *Lect. Notes Comput. Sci.*, pages 232–40, Puebla, Mexico, Oct. 2004.

[11] J. Vergés-Llahi, J. Aranda, and A. Sanfeliu. Visual tracking system for a mobile robot using colour histograms. In *Proc. 5th IFAC/EURON Sym. Intell. Auton. Vehicles*, Lisbon, Jul. 2004.

[12] T. Vidal-Calleja, J. Andrade-Cetto, and A. Sanfeliu. Conditions for suboptimal filter stability in SLAM. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, volume 1, pages 27–32, Sendei, Sep. 2004.

[13] T. Vidal-Calleja, J. Andrade-Cetto, and A. Sanfeliu. Estimator stability analysis in SLAM. In *Proc. 5th IFAC/EURON Sym. Intell. Auton. Vehicles*, Lisbon, Jul. 2004.