Learning and Vision Mobile Robotics Group Research Report 2002-2003

J. Andrade, F. Moreno, R. Alquézar, J. Aranda, J. Climent, A. Grau,
E. Muñoz, F. Serratosa, E. Staffetti, J. Vergés, T. Vidal, and A. Sanfeliu
Institut de Robòtica i Informàtica Industrial, UPC-CSIC
Llorens Artigas 4-6, 08028 Barcelona, Spain
lrobots@iri.upc.es

Abstract

This article presents the current trends on wheeled mobile robotics being pursued at IRI-UPC-CSIC. It includes an overview of recent results produced in our group in a wide range of areas, including robot localization, color invariance, segmentation, tracking, visual servoing, and object and face recognition.

Keywords: robot localization, color invariance, segmentation, tracking, recognition.

1 INTRODUCTION

The Learning and Vision Mobile Robotics Group, and 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 3 PhD students. The group, consolidated in 1996, has given rise to 3 PhD thesis and 7 final year projects. Within the last 6 years, the group has published 7 peer reviewed journal articles, 7 book chapters, 6 conference proceeding editorials, and presented articles in over 40 international conferences (35 indexed in SCI), and 15 national conferences. Furthermore, within the past two years, the mobile robotics platforms developed in our group have been portrayed numerous times on live and printed media [5, 7, 10, 19, 20, 21].

Financial support comes mainly from a continued set of projects from the Interministerial Council of Science and Technology (CICYT) TAP96-0629-C04-03, TAP98-0473, and DPI2001-2223.

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 contributed new insight in the classical simultaneous localization and map building problem, from a control systems theory point of view [1, 2, 6]. Furthermore, we have developed new feature validation techniques that improve the robustness of typical map building algorithms [3, 4].

Moreover, very good results have been achieved in the tracking of subjects under varying illumination conditions and in cluttered scenes. On the one hand, we have mastered now the use of histogram based techniques to color segmentation and illumination normalization [8, 22, 23]. On the other hand, we have tested different statistical estimation paradigms to track subject candidates using not only color information but shape as well [11, 12]. Most of our video demonstrations from last year show results in this topic.

With respect to 3d object recognition and subject identification, we have formalized and validated a compact representation for a set of attributed graphs. We call this new formulation a *function described graph*, and it borrows the capability of probablisitic modelling from random graphs. FDG's are the result of a long time effort in the search for a compact representation of multiple view model descriptors [13, 14, 15, 16, 17, 18].

In the following sections we summarize the key contributions from three selected publications [1, 11, 14]. Each of them tackles very different problems typically encountered in mobile robotics applications.

3 SIMULTANEOUS LOCALIZATION AND MAP BUILDING

To univocally identify landmarks from sensor data, we study several landmark representations,

XXIV Jornadas de Automática
León, September 10-12, 2003, ISBN 84-931846-7-5

and the mathematical foundation necessary to extract the features that build them from images and laser range data. The features extracted from just one sensor may not suffice in the invariant characterization of landmarks and objects, pushing for the combination of information from multiple sources.

Once landmarks are accurately extracted and identified, the second part of the problem is to use these observations for the localization of the robot, as well as the refinement of the landmark location estimates. We consider robot motion and sensor observations as stochastic processes, and treat the problem from an estimation theoretic point of view, dealing with noise by using probabilistic methods.

The main drawback we encounter is that current estimation techniques have been devised for static environments, and that they lack robustness in more realistic situations. To aid in those situations in which landmark observations might not be consistent in time, we propose a new set of temporal landmark quality functions, and show how by incorporating these functions in the data association tests, the overall estimation-theoretic approach to map building and localization is improved. The basic idea consists on using the history of data association mismatches for the computation of the likelihood of future data association, together with the spatial compatibility tests already available.

Special attention is paid in that the removal of spurious landmarks from the map does not violate the basic convergence properties of the localization and map building algorithms already described in the literature; namely, asymptotic convergence and full correlation.

We contribute also an in depth analysis of the fully correlated model to localization and map building from a control systems theory point of view. Considering the fact that the Kalman filter is nothing else but an optimal observer, we analyze the implications of having a state vector that is being revised by fully correlated noise measurements. We end up revealing theoretically and with experiments the strong limitations of using a fully correlated noise driven estimation theoretic approach to map building and localization in relation to the total number of landmarks used.

Partial observability hinders full reconstructibility of the state space, making the final map estimate dependant on the initial observations, and does not guarantee convergence to a positive definite covariance matrix. Partial controllability on the other hand, makes the filter beleive after a number of iterations, that it has accurate estimates of the landmark states, with their corresponding Kalman gains converging to zero. That is, after a few steps, innovations are useless. We show how to palliate the effects of full correlation and partial controllability.

Any map building and localization algorithm for mobile robotics that is to work in real time must be able to relate observations and model matches in an expeditious way. Some of the landmark compatibility tests are computationally expensive, and their application has to be carefully designed. We touch upon the time complexity issues of the various landmark compatibility tests used, and also on the desirable properties of our chosen map data structure. Furthermore, we propose a series of tasks that must be handled when dealing with landmark data association. From model compatibility tests, to search space reduction and hypothesis formation, to the actual association of observations and models.

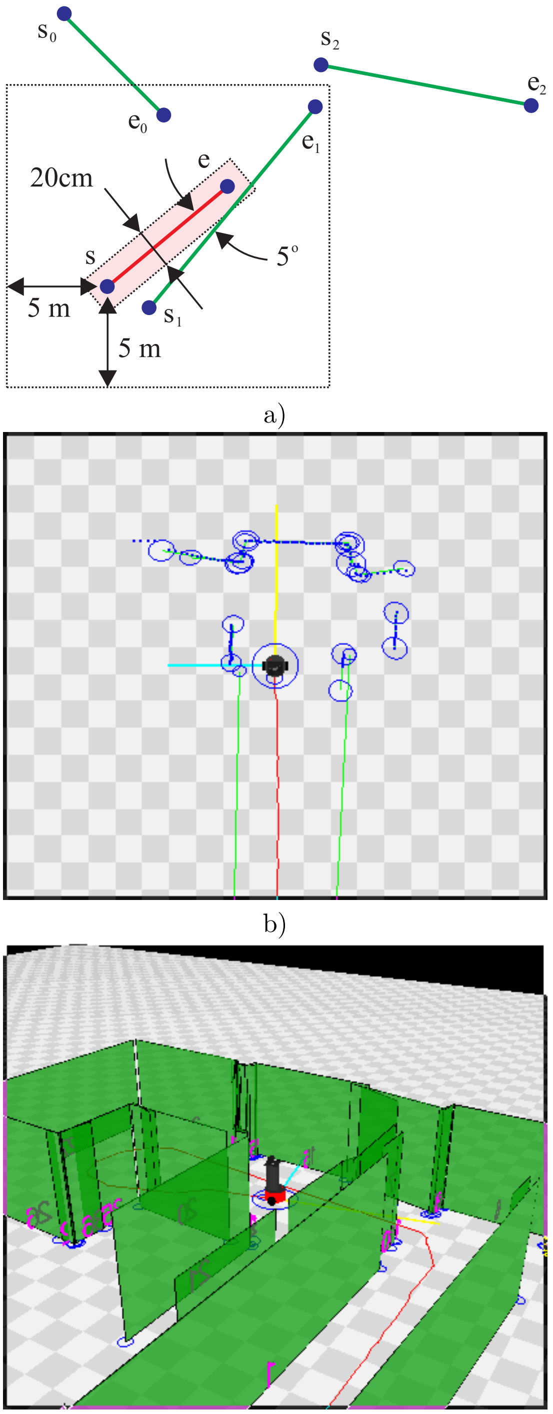
Figure 1a shows some of the model compatibility heuristics devised for the validation of straight lines extracted from laser range data into walls. Frame b shows data as extracted from a laser range finder, and 2σ covariance ellipses around hypothesized landmark estimates. The third frame shows a virtual reality model of the map constructed during a run of the algorithm.

4 FUSION OF COLOR AND SHAPE FOR OBJECT TRACKING UNDER VARYING ILLUMINATION

Color represents a visual feature commonly used for object detection and tracking systems, specially in the field of human-computer interaction. For such cases in which the environment is relatively simple, with controlled lighting conditions and an uncluttered background, color can be considered a robust cue. The problem appears when we are dealing with scenes with varying illumination conditions and confusing background.

Thus, an important challenge for any color tracking system to work in real unconstrained environments, is the ability to accommodate variations in the amount of source light reflected from the tracked surface.

The choice of different color spaces like HSL, normalized color rgb, or the color space $(B-G, G-R, R+G+B)$, can give some robustness against varying illumination, highlights, interreflections or changes in surface orientation for an analysis of different color spaces. But none of these transfor-



a)

b)

c)

Figure 1: Concurrent mobile robot localization and map building. a) Hypothesis search range for walls extracted from a laser range scan. b) The blue dots indicate sensor raw data coming from a laser range finder. The green lines represent walls inferred from consecutive readings. The red lines indicate the estimated robot trajectory. c) Graphical representation of the map built.

mations is general enough to cope with arbitrary changes in illumination.

Instead of searching for color constancy, other approaches try to adapt the color distribution over time. One such technique is to use Gaussian mixtures models to estimate densities of color, and under the assumption that lighting conditions change smoothly over time, the models are recursively adapted. Another option is to parameterize the color distribution as a random vector and to use second order Markov model to predict the evolution of the corresponding color histogram. These techniques perform much better than the mere change of color space representation, but have the drawback that they do not check for the goodness of the adaptation, which can still lead to failure.

The fusion of several visual modules using different criteria offers more reliability than methods that only use one feature. As an example, systems that track in real-time an individual might model the head of a person by an ellipse and use intensity gradients and color histograms to update the head position over time. In [12], color histograms are fused with stereovision information in order to dynamically adapt the size of the tracked head. These real time applications however are constrained only to the tracking of elliptical shapes.

A new methodology that addresses the problems present in the approaches described above, results in a robust tracking system able to cope with cluttered scenes and varying illumination conditions. The fusion is done using the CONDENSATION algorithm that formulates multiple hypothesis about the estimation of the object's color distribution and validates them taking into account the contour information of the object [9].

Four sets of sequence results are summarized in Figure 2 to illustrate the robustness of our system under different conditions. In the first experiment we show how our system is able to accommodate color by applying it over a synthetic sequence of circles moving around and changing randomly its color. In the upper left frame of the figure the path of the color distributions for the tracked circle is shown. The second experiment is to track a colored rectangle. It has to be pointed out that in the previous experiment we used the *RGB* color space, but in the present and subsequent experiments the color space used was the $(B-G, G-R, R+G+B)$ in order to provide robustness to specular higlights. The last two experiments, correspond to outdoor scenes, where although the change in illumination conditions is limited, they are useful to show that our method

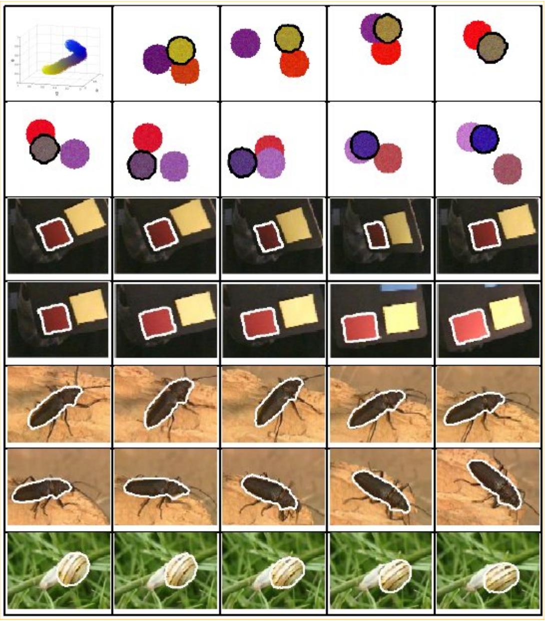


Figure 2: Results of the 4 experiments.

works with non-uniform shapes (third experiment of a beatle tracking), and in cluttered scenarios (fourth experiment of a snail tracking).

5 FUNCTION-DESCRIBED GRAPHS FOR MODELLING OBJECTS REPRESENTED BY SETS OF ATTRIBUTED GRAPHS

A *function-described graph* (FDG) is a model that contains probabilistic and structural descriptions of a set of attributed graphs (AGs) to maintain, to the most, the local features of the AGs that belong to the set and other AGs that are "near" them, as well as to allow the rejection of the AGs that do not belong to it or are "far" of them. Let us consider, as an example, the 3D-object modelling and recognition problem. The basic idea is that only a single FDG is synthesised from the graphs that represent several views of a 3D-object. Therefore, in the recognition process, only one comparison is needed between each model represented by an FDG and the unclassified object (view of a 3D-object) represented by a graph.

Random graphs are on the other hand, one of the earliest approaches used to represent a set of AGs. In this approach, AGs are extended to include probabilistic information. Wong et al. first defined the General Random Graphs (GRGs) for modelling classes of patterns described by AGs through a joint probability space of random variables ranging over pattern primitives (vertices) and relations (arcs). Due to the computational intractability of GRGs, caused by the difficulty in estimating and handling the high-order joint probability distribution, First-Order Random Graphs (FORGs) were proposed for real applications. Strong simplifications were made in FORGs to allow the use of random graphs in practical cases; more precisely, the following assumptions were made: a) the random vertices are mutually independent; b) given values for the random vertices, the random arcs are independent; c) the arcs are independent of the vertices except for the vertices that they connect.

FDGs can be seen as a type of simplification of the GRGs, different from FORGs, in which some structural constraints are recorded. A drawback of FORGs is that the strong assumptions about the statistical independence of nodes and arcs may lead to an excessive generalisation of the sample graphs when synthesising a FORG. To alleviate this weakness, a qualitative information of the joint probabilities of two nodes is incorporated into FDGs, thus improving the representational power of FORGs with a negligible increase of computational cost.

5.1 APPLICATION OF FDGS FOR MODELLING AND RECOGNITION OF OBJECTS

FDGs are applied here to 3D-object representation and recognition. The attribute of the vertices is the average hue of the region (cyclic range from 0 to 49) and the attribute of the edges is the difference between the colours of the two neighbouring regions.

We first present an experimental validation of FDGs using artificial 3D-objects in which the adjacency graphs have been extracted manually and afterwards we present a real application on an image database in which the graphs have been extracted automatically. The advantages of the experimental validation are that the results do not depend on the segmentation process and that we can use a supervised synthesis, since we know which vertices of the AGs represent the same planar face of the object. Thus, we can discern between the effects of computing a distance measure using different values of the costs on the 2nd-order relations. In the real application, we show the capacity of FDGs to keep the structural and semantic knowledge of an object despite the noise introduced by the segmentation process and an automatic synthesis.

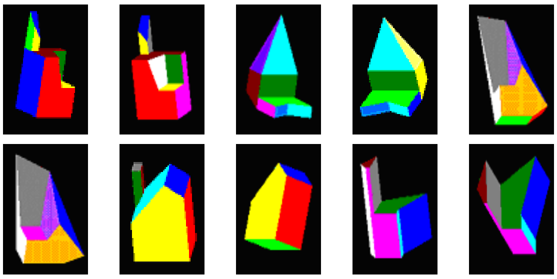


Figure 3: 10 views of 5 artificial objects designed with a CAD program.

5.1.1 SUPERVISED SYNTHESIS ON ARTIFICIAL OBJECTS

We designed five objects using a CAD program (see Figure 3). After that, we took five sets of views from these objects and from these views we extracted a total of 101 adjacency graphs. To obtain the AGs of the test set and of the reference set, we modified the attribute values of the vertices and arcs of the adjacency graphs by adding zero-mean Gausian noise with different variances. Moreover, some vertices and arcs where inserted and deleted randomly in some cases. The FDGs were synthesised using the AGs that belonged to the same 3D-object and using the synthesis given a common labelling from a set of AGs described in.

Table 1 shows the ratio of correctness for different levels of noise and applying several costs on the antagonisms and occurrences. We see that the bests results appear always when we use moderate 2nd-order costs. Furthermore, when noise increases, the recognition ratio decreases drastically when we use high costs but there is only a slight decrease when we use moderate costs. Moreover, in Table 2 we compare the FDGs (with 2nd-order costs) to other two methods. The FDG classifier always obtains better results than the 3-Nearest Neighbours and the Random Graph classifiers. The difference of the ratio between FDGs and the other two methods increases when the noise also increases. FORGs obtain better results than the 3-N.N. only when the noise is high.

5.1.2 UNSUPERVISED SYNTHESIS ON REAL LIFE OBJECTS

Images were extracted from the database COIL-100 from Columbia University. It is composed of 100 isolated objects and for each object there are 72 views (one view each 5 degrees). Adjacency graphs are obtained by color segmentation. Figure 4 shows 20 objects at angle 100 and their segmented images with the adjacency graphs. 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). FDGs were synthesised automatically using the AGs in the reference set that represent the same object. The method of incremental synthesis, in which the FDGs are updated while new AGs are sequentially presented, was applied. We made 6 different experiments in which the number of FDGs that represents each 3D-object varied. If the 3D-object was represented by only one FDG, the 36 AGs from the reference set that represent the 3D-object were used to synthesise the FDG. If it was represented by 2 FDGs, the 18 first and consecutive AGs from the reference set were used to synthesise one of the FDGs and the other 18 AGs were used to synthesise the other FDG. A similar method was used for the other experiments with 3, 4, 6 and 9 FDGs per 3D-object.

Similarly to the previous experimental results, the correctness is higher when 2nd-order relations are used with a moderate cost. The best result appears when each object is represented by 4 FDGs, that is, each FDG represents 90 degrees of the 3D-object. When objects are represented by 9 FDGs, each FDG represents 40 degrees of the 3D-object and 4 AGs per FDG, there is poor probabilistic knowledge and therefore the costs on the vertices and arcs are coarse. Moreover, when objects are represented by only 1 or 2 FDGs, there are too much spurious regions (produced in the segmentation process) to keep the structural and semantic knowledge of the object.

References

[1] J. Andrade-Cetto. *Environment Learning for Indoor Mobile Robots.* PhD thesis, UPC, Barcelona, Apr. 2003.

[2] J. Andrade-Cetto and A. Sanfeliu. Concurrent map building and localization on indoor dynamic environments. *Int. J. Pattern Recogn.*, 16(3):361–374, May 2002.

[3] J. Andrade-Cetto and A. Sanfeliu. Concurrent map building and localization with landmark validation. In *Proc. 16th IAPR Int. Conf. Pattern Recog.*, volume 2, pages 693–696, Quebec, Aug. 2002. IEEE Comp. Soc.

[4] J. Andrade-Cetto and A. Sanfeliu. Concurrent map building and localization with temporal landmark validation. Accepted for presentation in 2003 IEEE International Conference on Robotics and Automation, 2003.

Table 1: FDGs ratio of correctness.

Num. Vertices	**Ins & Del**	**0**	**0**	**0**	**0**	**0**	**1**	**2**	**1**
Standard	**Deviation**	**0**	**2**	**4**	**8**	**12**	**0**	**0**	**8**
Cost Antag	**Cost Occu**								
Moderate	Moderate	100	98	97	95	92	89	85	83
High	Null	100	92	89	87	84	61	54	57
Null	High	100	91	89	88	85	62	59	59
High	High	100	95	90	86	80	60	53	56
Moderate	Null	100	92	91	91	87	80	75	75
Null	Moderate	100	95	92	91	86	81	77	76
Null	Null	100	90	89	88	86	70	67	68

Table 2: FDGs (moderate 2nd order costs), FORGs and 3-NN ratio of correctness.

Num Vertices Ins. or Del.	**0**	**0**	**0**	**0**	**0**	**1**	**2**	**1**
Standard Deviation	**0**	**2**	**4**	**8**	**12**	**0**	**0**	**8**
FDGs (Moderate costs)	100	98	97	95	92	89	85	83
Random Graphs (FORGs)	100	90	89	88	86	70	67	68
3-N.N. (Edit Op. Distance)	100	98	82	62	52	90	58	58

[5] Canal Blau TV. *The IRI Robots in Canal Blau*, May 2002.

[6] A. Checa, J. Andrade, and A. Sanfeliu. Construcció de mapes per robots móbils equipats amb sensors làser de profunditat. Technical Report IRI-DT-03-01, IRI, UPC, Mar. 2003.

[7] El Periódico de Catalunya. *El Robot Doméstico Llama a la Puerta*, Apr. 2002.

[8] A. Grau, J. Climent, F. Serratosa, and A. Sanfeliu. Texprint: A new algorithm to discriminate textures structurally. In T. Caelli, A. Amin, R. P. W. Duin, M. Kamel, and D. de Ridder, editors, *Proc. IAPR Int. Workshop Syntactical Structural Pattern Recog.*, volume 2396 of *Lect. Notes Comp. Sci.*, pages 368–377, Windsor, Aug. 2002. Springer-Verlag.

[9] M. Isard and A. Blake. Condensation - conditional density propagation for visual tracking. *Int. J. Comput. Vision*, 29(1):5–28, Aug. 1998.

[10] La Politècnica, Num 4. UPC. *Criatures Cibernètiques*, Nov. 2001.

[11] F. Moreno, J. Andrade-Cetto, and A. Sanfeliu. Fusion of color and shape for object tracking under varying illumination. In F. J. Perales, A. J. C. Campilho, N. P. de la Blanca, and A. Sanfeliu, editors, *Proc. 1st Iberian Conf. Pattern Recog. Image Anal.*, volume 2652 of *Lect. Notes Comp. Sci.*, pages 580–588, Puerto de Andratx, Jun. 2003. Springer-Verlag.

[12] F. Moreno, A. Tarrida, J. Andrade-Cetto, and A. Sanfeliu. 3d real-time head tracking fusing color histograms and stereovision. In *Proc. 16th IAPR Int. Conf. Pattern Recog.*, volume 1, pages 368–371, Quebec, Aug. 2002. IEEE Comp. Soc.

[13] A. Sanfeliu, R. Alquézar, J. Andrade, J. Climent, F. Serratosa, and J. Vergés. Graph-based representations and techniques for image processing and image analysis. *Pattern Recogn.*, 35:639–650, Mar. 2002.

[14] F. Serratosa, R. Alquézar, and A. Sanfeliu. Function-described graphs for modelling objects represented by sets of attributed graphs. *Pattern Recognition*, volume 36, pages 781–798, 2003.

[15] F. Serratosa, R. Alquézar, and A. Sanfeliu. Estimating the joint probability distribution of random vertices and arcs by means of second-order random graphs. In T. Caelli, A. Amin, R. P. W. Duin, M. Kamel, and D. de Ridder, editors, *Proc. IAPR Int. Workshop Syntactical Structural Pattern Recog.*, volume 2396 of *Lect. Notes Comp. Sci.*, pages 252–260, Windsor, Aug. 2002. Springer-Verlag.

[16] F. Serratosa, R. Alquézar, and A. Sanfeliu. Synthesis of function-described graphs and clustering of attributed graphs. *Int. J. Pattern Recogn.*, 16(6):621–655, 2002.

[17] F. Serratosa, A. Sanfeliu, and R. Alquézar. Modelling and recognizing 3d-objects de-

Figure 4: The 20 selected objects and the segmented images with the AGs.

scribed by multiple views using function-described graphs. In *Proc. 16th IAPR Int. Conf. Pattern Recog.*, volume 2, pages 140–143, Quebec, Aug. 2002. IEEE Comp. Soc.

[18] E. Staffetti, A. Grau, F. Serratosa, and A. Sanfeliu. Oriented matroids for shape representation and indexing. In F. J. Perales, A. J. C. Campilho, N. P. de la Blanca, and A. Sanfeliu, editors, *Proc. 1st Iberian Conf. Pattern Recog. Image Anal.*, volume 2652 of *Lect. Notes Comp. Sci.*, pages 1012–1019, Puerto de Andratx, Jun. 2003. Springer-Verlag.

[19] Televisió de Catalunya. *Robot Casolà*, May 2003.

[20] Televisión Española. *Informe Semanal*, Jan. 2002.

[21] UPC Libraries. *More Than Just Machines*, Jan. 2003.

[22] J. Vergés-Llahí and A. Sanfeliu. Colour constancy algorithm based on the minimization of the distance between colour histograms. In F. J. Perales, A. J. C. Campilho, N. P. de la Blanca, and A. Sanfeliu, editors, *Proc. 1st Iberian Conf. Pattern Recog. Image Anal.*, volume 2652 of *Lect. Notes Comp. Sci.*, pages 1066–1073, Puerto de Andratx, Jun. 2003. Springer-Verlag.

[23] J. Vergés-Llahí, A. Tarrida, and A. Sanfeliu. New approaches to colour histogram adaptation in face tracking tasks. In *Proc. 16th IAPR Int. Conf. Pattern Recog.*, pages 381–384, Quebec, Aug. 2002. IEEE Comp. Soc.