The Use of Graph Techniques for Identifying Objects and Scenes in Indoor Building Environments for Mobile Robots

Alberto Sanfeliu

Institut de Robòtica i Informàtica Industrial (UPC-CSIC)
Universitat Politècnica de Catalunya (UPC)
sanfeliu@iri.upc.es
http://www-iri.upc.es/groups/lrobots

Abstract. In this paper we present some of the common issues that appear when we try to recognize objects in indoor scenes of a building, and we describe some strategies for recognizing them by using graph techniques. These scene images are captured by the colour cameras of a mobile robot, which in the learning phase, learn the objects by taken a set of 2D images of the projective object views. Then afterwards, the robot must identify the objects once its moves through the area that has been used to learn the objects. We describe two strategies to use graph techniques for object and scene recognition, some algorithms and preliminary results.

1 Introduction

Computer vision in autonomous mobile robotics is a very well known topic that is being treated by many research groups [7]. However, the use of perception techniques to automatically learn and recognize the environment and the objects located on it is probably not so well known, although there are also a number of research work on the area of robot vision [2,5,13,14,15,19]. We will describe in this paper some of the research that we are doing in the area robot vision for mobile robots and more specifically, the one related to the graph techniques. One part of our research has been concentrated in the development of techniques to capture and process the information that surrounds a robot, taking into account that this information can be captured by diverse perception sensors (colour video cameras, stereo vision, laser telemeter, ultrasonic sensors, etc.) and the sensors related to robot movement (odometers).

We have focused our research in the development of "robust" techniques that must be as much as possible, "invariant" to illumination, colour, surface reflectance, sensor uncertainty, dead reckoning and dynamic environments. However, this wish is not always possible. We also orient our research to develop techniques to learn the perceptive world, in order to create a data base that can be used later on, by robots.

In this paper we present some results of the use of graph techniques [16] for the process of identification of objects and scenes in indoor environments for mobile robots.

In the first section we describe some of the common issues in images acquired by a robot in indoor building environments, in the second section we show two strategies

A. Fred et al. (Eds.): SSPR&SPR 2004, LNCS 3138, pp. 30–41, 2004.

used for identifying objects, in the third section we summarize two graph methods based on the segmentation-recognition strategy and in the fourth section we present several ideas to use detection-recognition methods based on graph techniques. Finally, we present some results.

2 Common Issues in Robotic Scene Images of Indoor Buildings and Their Implication in Object Recognition

In order to recognize objects in a scene, we first have to capture them and create a data base. If the objects are isolated and environment conditions do not change, there are not object appearance variations between the learning and the identification phase. However, when a robot moves around an environment, the appearance of the objects may change between the both phases due to diverse issues. Let us describe some common issues that there exist in indoor building scene images and their implication in the process of learning and recognition.

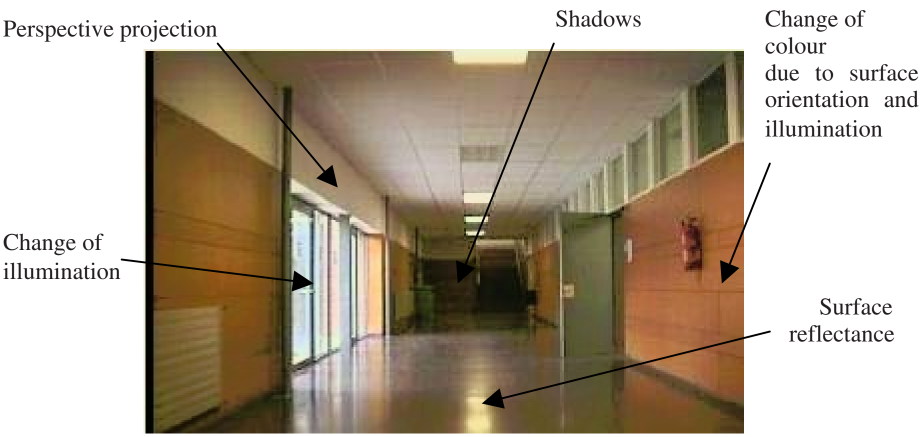

Fig. 1. Some issues on image scenes of indoor buildings

Some of the issues that produce discrepancies are the following ones (Fig. 1 and 2):

- Perspective projection due to the camera model.
- Partial occlusion due to camera point of view or due to the intersection of an obstacle between the object and the camera.
- Colour modification due to the surface orientation, surface reflectance, multiple illumination sources or sensibility of camera sensor.
- Surface reflectance
- Surface texture
- Shadows produced by other objects or by the own concavities of the object.
- Confusing background.

Some of these issues have a direct impact in the scene as far as the object recognition process is concerned. Specifically we can enumerate the following ones:

- *The separation of the object from the background*: If the features that differentiate the object from the background are sensible to the aforementioned problems, then the extraction of the object is not easy. For example, if a segmentation process is used, then the segmentation features must be invariant to colour, surface reflectance, etc. If there are shadows or the background of the object is confusing then the separation is even worse.
- *The detection of the object surface features*: The object surface colour, surface reflectance and surface texture are usually not invariant, although in some cases, these problems can be partially overcome.
- *The extraction of geometric object features*: Due that the scene is captured by means of a camera, then the perspective projection must be taken into account. This issue produce sensitive variations on the extraction of geometric features (area, centre of geometry of a surface, angles between contour lines, etc.).
- *The image view of a 2D projection of a 3D object*: A 3D object has usually multiple 2D views which depend on the orientation of the object with respect to the camera. The number of views usually depends on the number of potential object rotations and the number of concavities of the object.
- *The partial occlusion of an object*: This issue produce an important reduction of the visibility of the object which has a direct impact on the identification of the object.

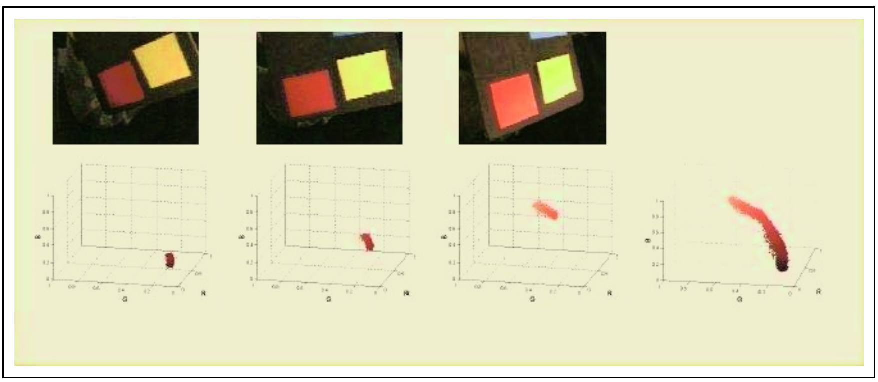

Fig. 2. Typical reflectance problems of a colour (red) planar surface: (a) a sequence of a red planar surface; (b) RGB map of the colour distribution of the sequence of the planar surface

In order to overcome some of these problems, for example, we can apply invariant techniques (colour constancy methods [9,24], projective invariants, etc.) or to fuse information of diverse sensors (colour-disparity for segmentation [1], colour-disparity-edges-motion-SSD for visual servoing [13], textons-contours-regions fors segmentation, etc.). However, these techniques are usually not enough for object recognition due to the stochastic variability of the perceptive features and to the last two commented issues: the orientation of the object and the partial occlusion. For these reasons, graph matching can be a good candidate for object learning and recognition. Moreover, if we consider attributed graphs, then the stochastic variability can be included in the nodes and arcs. If additional, it is used a distance measure to compare a graph view against object graph views of a data base, then we can also cope with the variability of the graph topology.

However, the use of graph techniques for object recognition has at least a big drawback than still have not been overcome: the time complexity of the matching process. This issue is under study.

3 Strategies for Object Recognition

There are two strategies for recognition of objects in scenes: segmentation-recognition and detection-recognition. The first one is a general approach, where it is not essential to have a priori knowledge, to extract the objects for the recognition process. It can be applied to any type of scene and the objects to identify can be partially occluded. This technique has some drawbacks in indoor images, for example, they are time consuming and very dependent on the segmentation process. The second one, detection-recognition, requires having a good knowledge of the objects to detect and moreover, the objects can not be partially occluded. The advantages are that the time complexity can be reduced and that the algorithms can be adjusted to diminish the feature extraction dependency. We will present both strategies from the point of view of the application of graph techniques.

- *Segmentation-Recognition*: Often called bottom-up strategy, it is a good approach for applying graph techniques for object recognition. In this case, the whole image can be seen as a graph and the goal is to find a sub-graph in the image graph that match one of the graphs of the object data base. We have been working in this area and we have developed several techniques. We will describe one technique based on random graphs for matching a 2D view of a scene object against a data base of 3D objects. We also will explain another one, which use oriented matroids to index 2D views of image objects.

- *Detection-Recognition*: This strategy is something similar to a top-down strategy but with some special features. In this case, the objective is to detect potential zones where there can be objects or zones of interest and then, apply a method to find the object in that zone. If the objects to identify have distinguish features then these features can be used to detect the object. This strategy has been applied successfully in object detection using non graph techniques, for example in human face detection [25]. We can also think about other techniques that are in between these two strategies, which do not require a pure segmentation process neither to have too much knowledge of the objects for recognition.

4 Segmentation-Recognition Graph Techniques for Object Recognition

We have previously described some common problems that we can find in a scene image of an indoor building, and the consequences that they produce to the objects that we have learned and we want to identify later on. Since these issues can produce big variations between the image captured in the learning phase and the image captured in the identification phase, then we need robust methods to cope with these variations.

The methodology is to segment an image, extract the graph features and then apply a graph method to identify an object against a data base of reference objects. The main issue of this methodology is to segment well the image, which is often not the case. Since there is not a good segmentation, the graph matching technique must overcome the potential variability of the extracted graph with respect to the "ideal" graph. Usually graph or sub-graph isomorphism techniques are not the most appropriate ways to identify an object due to aforementioned problems, besides a potential partial occlusion of the object to identify. It is usually required to apply distance measure methods which allow coping with the variability between the object graph and the reference one. There have been developed several well known graph techniques than can be applied to object recognition, for example [6,11,12,22,26,27, 28,29], although we only will summarize two techniques that our group have developed based on this strategy.

4.1 Matching Views of 2D Projections of 3D Objects by Using Oriented Matroids

The idea is to represent 2D views of a 3D object, by means of topological properties of the regions of the segmented image and then, to create a table with each one of the topological representations. Then the identification process is based on matching the input representation of one scene view, to the table of the topological representations of the 2D object views. In this case the graph representation of a segmented image is reduced to a list of ordered chains of symbols (denominated co-circuits), where each co-circuit is the spatial combination of regions based on two reference regions.

A topological representation is created by using the oriented matroid theory by means of encoding incidence relations and relative position of the elements of the segmented image, and by giving local and global topological information about their spatial distribution. The result is a set of co-circuits [3] of sign combinations that relates segmented regions with respect to the convex hull of two selected regions of the scene. The details of this process are explained in [22]. The set of co-circuits obtained is projective invariant, which is an important feature for the representation of the model objects. Fig. 3 shows the segmentation and process indexing of one object and Table 1 shows the resulting indexes of the object.

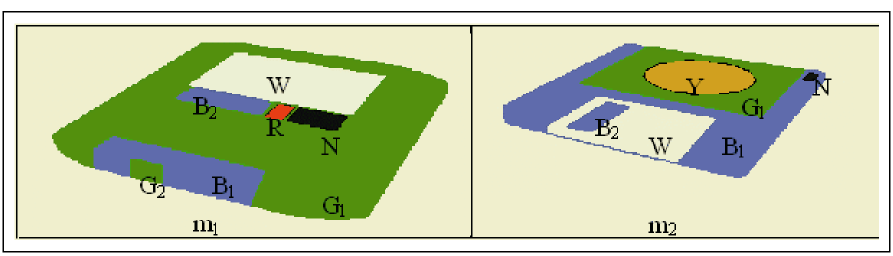

Fig. 3. Segmentation and process indexing of two objects

The result of the indexing process looks as follows:

Table 1. Index result of the process indexing of the images of Fig. 3. The first column is the baseline area from where the segmented regions are related. 0 means the region is inside the baseline area; - the region is one the left side; + the region is on the right side; and * means the region does not exist in the segmented image

	W	R	Y	G_1	G_2	B_1	B_2	N	Object
WR	0	0	*	0	0	0	-	+	m_1
WY	0	*	0	0	*	0	0	-	m_2
WG_1	0	*	*	0	*	*	*	*	m_1
WG_1	0	*	0	0	*	0	0	0	m_2
WG_2	0	0	*	0	0	+	0	0	m_1
WB_1	0	0	*	0	0	0	0	0	m_1
WB_1	0	0	*	+	+	+	0	+	m_2
WB_2	0	0	*	+	+	+	0	+	m_1
WN	0	0	*	-	-	-	-	0	m_1
WN	0	*	+	+	*	0	0	0	m_2
RG_1	*	0	*	0	*	*	*	*	m_1
…	…	…	…	…	…	…	…	…	
B_2N	+	0	*	-	-	-	0	0	m_1
B_2N	-	*	+	+	*	+	0	0	m_2

The matching process is done by comparing the set of co-circuits of the 2D projection view of the scene, to the set of co-circuits of the data base. The time complexity of the matching process is polynomial with respect to the number of segmented zones of the scene image. The reason of the reduction of the time complexity is due to two reasons: the elimination of labelling process; the comparison against a set of co-circuits which number is polynomial with respect to the number of segmented zones in the worst case.

4.2 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. Fig. 4 shows the process of learning (synthesis of the object graph views) and recognition.

Object views are often represented by graphs, and one robust representation is based on attributed graphs (AG). However, in order to synthesize AG we need a more general model representation, which is called Random Graph (RG). The generalization of these graphs is denominated General Random Graphs (GRG) which has theoretically, great representation power, but they need a lot of space to keep up with the associated data. We have defined several simplifications to the GRG to reduce the space and also to diminish the time matching complexity. Wong and You [27] proposed the First-Order Random Graphs (FORGS) with strong simplifications of the GRG, specifically they introduce three assumptions about the probabilistic independence between vertices and arcs which restrict too much the applicability of these graphs to object recognition. Later, our group introduced a new class of graphs called Function-Described Graphs (FDG) [20] to overcome some of the problems of the FORG. The FDG also considers some independence assumptions, but some useful 2° order functions are included to constrain the generalisation of the structure. Specifically a FDG includes the antagonism, occurrence and existence relations which apply to pairs of vertices and arcs. Finally, we have expanded this representation, [17,18] by means of Second-Order Random Graphs (SORG), which keep more structural and semantic information than FORGs and FDGs. These last types of representation have led to the development of synthesis techniques for model object generation (by means of 2D projections of a 3D object) and graph matching techniques for graph identification.

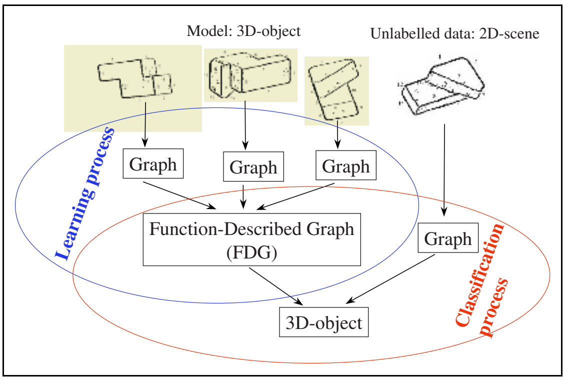

Fig. 4. Learning and classification processes in the classifiers that use only one structural representation per model

The time complexity of this method is in the worst case, exponential with respect to the number of nodes of the graph. This time complexity can be reduced pruning the number of combinations by using some ad-hoc information of the objects and the images to be applied.

5 Detection-Recognition Graph Techniques for Object Recognition

As we have commented, there are other ways to recognize objects in scenes, where graph techniques can be used. The strategy is to detect zones of potential objects and then apply classification techniques to recognize an object in that zone. The graph techniques can be used in the detection of the zones, in the classification process or at the same time, in the detection-classification. In this last case, the technique can be used, for example, as an indexing method. We will describe in this section only detection techniques, since once a zone has been detected, the methods described in the previous section can be used.

Three general detection approaches can be applied:

- *Global Search Detection*: The idea is to generate a global graph of the full image and then look for a specific sub-graph that has the potential to be a zone object. A typical technique to represent the complete image is the Voronoi diagram representation [10]. If not attributes are used, then we can apply general sub-graph matching techniques. When we have attributes, for example the colour or the area of the regions, then we can prune the potential matches using the node and arc attributes. More sophisticated techniques can also be applied, for example to grow zones using a potential field, where the function can be related to the fan in and fan out of the graph or the node and arc attributes.

- *Raster Search Detection*: The idea is to pass a window through the full image which has a basic graph structure and attributes of the zone to be located. When the match distance between the reference graph and the graph that are extract under the window limits, is higher than a threshold, then the zone is identified. One potential method to apply is a PCA approach for fast retrieval of structural patterns in attributed graphs [26]. In this case, in order to detect the nodes we can use a pre-segmentation process which outcome is a planar graph, for example using one step of the technique [25] or the technique [8] which nodes are spanning trees. Since we look for potential graph zones instead a full matching process, this algorithm can do the process in polynomial time.

- *Probabilistic Search Detection*: The idea is to probabilistically take some initial starting points where to grow a graph. This methodology has been applied successfully in diverse fields, for example in segmentation, path planning or salience detection. From the starting point of the image, we can grow the graph without restrictions, that is, looking for the neighbour nodes by using general rules, or to grow the graph imposing a graph reference model. In the last case the idea is similar to the raster search detection methodology, but in a probabilistic way.

6 Some Results

We show in this article two examples of identifying objects and images by means of graph techniques. The first one is applied to learn and recognize 3D objects by means of their 2D projection views and the second one, it is the learning and recognition of image scenes by means of oriented matroids. In the first example, the images come from a standard data base, and in the second, the images have been acquired by the colour camera of the robot. In both examples we used the segmentation-recognition strategy, where the segmentation was based on the colour of the image pixels using the method described in [8]. Moreover, in Fig. 8, we present the set of images that we are using and the segmentation results.

For the first example we used a set of objects extracted from the database 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 0º to 90º, the second one grouping from 95º to 180º 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). The learning techniques (synthesis of graphs) are described in [18]. For the recognition process we used the distance measures explained in [18,20].

Fig. 4 shows 20 objects at angle 100º and their segmented images with the adjacency graphs. FORGs, FDGs and SORGs 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 random graphs, FORGs, FDGs and SORGs, that represents each 3D-object varied. The best result appears when the SORG and FDG representations were used, although the best is the SORG representation. Fig. 5 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.

For the second example, we used two set of examples: (1) 10 different reference images of an indoor building, and from each one, three images were taking at different position and orientation by the colour camera of a mobile robot; and (2) a sequence of several hundred of images acquired by the robot. Figure 6.b shows an image taken from three different views, their segmented images and the learning process. Fig. 8 shows two images of the image sequence.

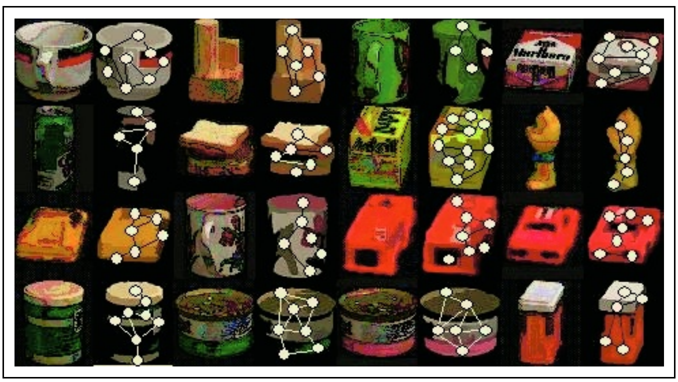

Fig. 5. Some objects at angle 100 and the segmented images with the AGs

The learning and recognition process was the following one: (1) perform colour segmentation of each image scene; (2) extract the co-circuits of each image; and (3) construct a data base joining the co-circuits. We applied a distance measure between co-circuits to identify the image. See [21] for details. For the images of Fig.7, 74% of the images where well recognized and for a sequence of images of Fig.8, 100% of the images were well classified.

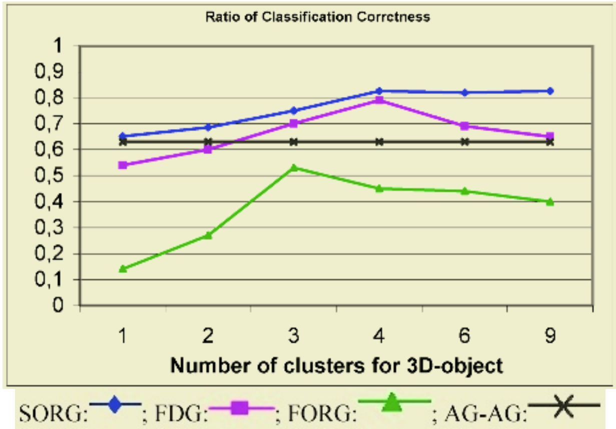

Fig. 6. Ratio of recognition correctness of the objects using SORG, FDG, FORG and AG-AG

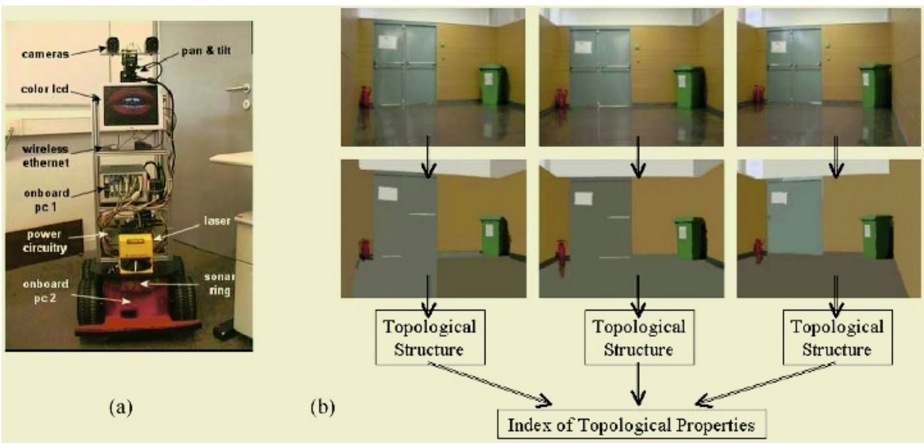

Fig. 7. (a) ANNA mobile robot; (b) learning process using three different views

Fig. 8. Some images taken from the mobile robot called Marco

Fig. 9. Segmentation results of indoor buildings

7 Conclusions

In this paper we present some common issues that we find in robotics when a robot must use computer vision techniques for identifying objects and scenes in indoor buildings. We also explain several strategies used to locate and identify objects and some graph techniques applied in the identification process. We are at present testing these techniques in several sequences of indoor building images in order to see the robustness of them.

References

1. J. Andrade and A. Sanfeliu: Integration of perceptual grouping and depth. Proceedings of the 15th International Conference on Pattern Recognition, Barcelona, September 3-7, IEEE Computer Society, Vol. 1, (2000), pp. 295-298.
2. S. Belongie and J. Malik: Matching with shape contexts. Proceedings IEEE Workshop on Image and Video Libraries, June 12, (2000), pp. 20-26.
3. A. Björner, M.L. Vergnas, B. Sturmfels, N. White, G.M.: Oriented matroids. Volume 43 of Encyclopedia of Mathematics and its Applications. Cambridge University Press, (1993).
4. H. Bunke: Recent developments in graph matching. Proceedings of the 15th International Conference on Pattern Recognition, Barcelona, September 3-7, IEEE Computer Society, Vol. 1, (2000).
5. C. Chen and A. Kak: Robot vision system for recognizing objects in low-order polynomial time. IEEE Trans. on System, Man and Cybernetics, 18(6), Nov. (1989), pp. 1535-1536.
6. W.J. Christmas, J. Kittler and M. Petrou: Structural matching in computer vision using probabilistic relaxation. IEEE Transactions on PAMI, vol. 17, (1995), pp. 749-764.
7. G.N. DeSouza and A.C. Kak: Vision for mobile robot navigation, a survey. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, No. 2, Feb. (2002).
8. P.F. Felzenswalb and D.P. Huttenlocher: Image Segmentation Using Local Variation. Proc. of the IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition, (1998), pp. 98-104.

9. G. Finlayson and S. Hordley: Colour by correlation: a simple, unifying framework for colour constancy. IEEE Trans. on PAMI, 23(11): 1209-1221, Nov. (2001).
10. I. G. Gowda, D. G. Kirkpatrick, D. T. Lee and A. Naamad: Dynamic Voronoi diagrams, IEEE Trans. on Information Theory, Vol. IT-29, No.5, Sept. (1983).
11. B.Huet & E.R.Hancock: Relational object recognition from large structural libraries. Pattern Recognition, Vol. 35, (2002), pp: 1895-1915.
12. X.Jiang, A.Münger and H. Bunke: On median graphs: Properties, algorithms and applications. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 23, No. 10, (2001), pp: 1144-1151.
13. D. Kragié and H. I. Christensen: Cue integration for visual servoing. IEEE Trans. on Robotics and Automation, Vol. 17, No. 1, Feb. (2001).
14. J. Malik, S. Belongie, J. Shi and T. Leung: Textons, contours and regions: cue integration in image segmentation. Proceedings of the Seventh IEEE International Conference on Computer Vision, Vol.2, Sept. (1999), pp.918-925.
15. S. Nayar, S. Nene and H. Murase: Subspace methods for robot vision. IEEE trans. on Robotics and Automation Vol. 12 No. 5, (1996), pp. 750-758.
16. A. Sanfeliu, R. Alquezar, J. Andrade, J. Climent, F. Serratosa and J. Verges: Graph-based representations and techniques for image processing and image analysis. Pattern Recognition, Vol. 35, (2002), pp. 639-650.
17. A. Sanfeliu and F. Serratosa: Learning and recognising 3D models represented by múltiple views by means of methods based on random graphs. IEEE Int. Conference on Image Processing, Barcelona 2003 (ICIP 2003), Sept. (2003).
18. A. Sanfeliu, F. Serratosa and R. Alquezar: Second-order graphs for modelling sets of attributed graphs and their application to object learning and recognition. Int. J. Pattern Recognition and Artificial Intelligence, Vol. 18, No.3, (in press), (2004).
19. K.Sengupta & K.L.Boyer: Organizing large structural model bases. IEEE Trans. on PAMI, Vol. 17, No. 4, (1995), pp: 321-331.
20. F. Serratosa, R. Alquezar and A. Sanfeliu: Function-described graphs for modelling objects represented by sets of attributed graphs. Pattern Recognition, Vol. 36, (2003), pp. 781-798.
21. F. Serratosa, T. Grau and A. Sanfeliu: Distance between 2D-scenes base on oriented matroid theory. Proc. of the International Conference on Pattern Recognition, ICPR'04, Cambridge, August 23-26, (2004).
22. E. Staffetti, A. Grau, F. Serratosa and A. Sanfeliu: Shape representation and indexing based on region connection calculus and oriented matroid theory. In Discrete Geometry for Computer Imagery, Lecture Notes in Computer Science, Springer-Verlag, Napoles, (2003).
23. S. Umeyama: An eigendecomposition approach to weighet graph matching problems. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 10, No.5, Sept. (1988).
24. J. Vergés-LLahí, A. Sanfeliu: Colour constancy algorithm based on object function minimization. Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA), Puerto Andrax, Mallorca, 4-6 June. Lecture Notes in Computer Science, (2003).
25. P. Viola and M. Jones: Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, 8-14 dec. (2001), pp. 511-518.
26. L. Xu and I. King: A PCA approach for fast retrieval of structural patterns in attributed graphs. IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics, Vol. 31, No. 5, Oct. (2001).
27. A.K.C. Wong and M. You: Entropy and distance of random graphs with application to structural pattern recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 7, (1985), pp. 599-609.
28. E.K. Wong: Model matching in robot vision by sub-graph isomorphism. Pattern Recognition, Vol. 25, (1994), pp: 287-304.
29. H. Zhang and H. Yan: Graphic matching based on costrained Voronoi diagrams, Proceedings of the Fifth International Symposium on Sygnal Processing and its Applications, Brisbane, 22-25 August, (1999).