ELSEVIER

Image and Vision Computing 18 (2000) 515–530

image AND vision COMPUTING

www.elsevier.com/locate/imavis

A review on strategies for recognizing natural objects in colour images of outdoor scenes

J. Batlle[a,*], A. Casals[b], J. Freixenet[a], J. Martí[a]

[a]*Computer Vision and Robotics Group—IIiA, University of Girona, 17071 Girona, Catalonia, Spain*
[b]*Department of Automatic Control and Computer Engineering, Polytechnical University of Catalonia, 08028 Barcelona, Catalonia, Spain*

Received 20 April 1999; received in revised form 23 August 1999; accepted 18 September 1999

Abstract

This paper surveys some significant vision systems dealing with the recognition of natural objects in outdoor environments. The main goal of the paper is to discuss the way in which the segmentation and recognition processes are performed: the classical bottom–up, top–down and hybrid approaches are discussed by reviewing the strategies of some key outdoor scene understanding systems. Advantages and drawbacks of the three strategies are presented. Considering that outdoor scenes are especially complex to treat in terms of lighting conditions, emphasis is placed on the way systems use colour for segmentation and characterization proposals. After this study of state-of-the-art strategies, the lack of a consolidated colour space is noted, as well as the suitability of the hybrid approach for handling particular problems of outdoor scene understanding.

Keywords: Outdoor scene description; Image segmentation; Object recognition; Control strategies

1. Introduction

Scene understanding constitutes a relevant research area involving the fields of Computer Vision and Artificial Intelligence. Most of the effort in this discipline has focused on improving image segmentation techniques and developing efficient knowledge representations that have enabled relevant results to be obtained. In particular, there is increasing interest in developing applications for outdoor environments such as vision-based mobile robot navigation systems [1–3]. However, outdoor scenes are particularly hard to treat in terms of lighting conditions due to weather phenomena, time of day and seasons.

The objective of a Scene Understanding System consists of recognizing and localizing the significant imaged objects in the scene and identifying the relevant object relationships. Consequently, a system must perform segmentation, region characterization and labelling processes. As Haralick and Shapiro suggested [4], the way to carry out these three tasks depends on the strategy used: bottom–up, top–down, or hybrid strategies, such as Fig. 1 outlines.

- The current bottom–up scheme was clearly defined by Fischler in 1978 [5]. Firstly, it is necessary to partition a scene into regions by using general-purpose segmentation techniques. These regions are then characterized by a fixed set of attributes, and the scene itself is characterized by linking the objects to each other. The labelling process requires an inference engine to match each region to the best object-model (a comparison of labelling methods can be found in Ref. [6]).
- Top–down approach starts with the hypothesis that the image contains a particular object or can be categorized as a particular type of scene [4]. The system will then perform further tasks in order to verify the existence of a hypothesized object. Systems that follow a top–down strategy are known as Special Purpose Vision Systems (SPVS). These systems carry out specific tasks: defining, structuring and applying knowledge relevant to their task domain. A typical SPVS applies specialized segmentation and recognition methods to each object to be recognized. Hanson et al. [7] refer to SPVS as "successful systems that cover a more humble goal."
- The hybrid approach is a mixture of the previous two. In fact, there are several schemes that can be considered as hybrid forms or variants of the top–down or bottom–up methods (see e.g. Refs. [8,9]). The most accepted hybrid approach is characterized by segmenting the input image using non-purposive techniques, i.e. non-semantic segmentation is used. Once the image is partitioned into regions, the objects are then identified by specialized

* Corresponding author. Tel.: +34-72418956; fax: +34-72418399.
E-mail address: jbattle@ei.udg.es (J. Batlle).

0262-8856/00/$ - see front matter
PII: S0262-8856(99)00040-2

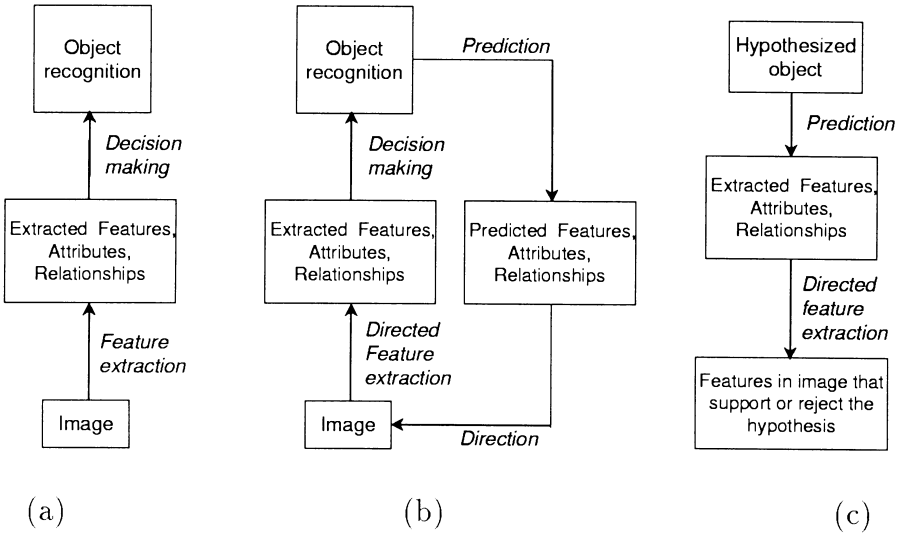


Fig. 1. Strategies for object recognition: (a) bottom–up; (b) hybrid; and (c) top–down.

procedures that search for a set of selected features. There are also hybrid systems that combine general-purpose and specific segmentation methods. For example, it is interesting to first segment the image into large regions (gross segmentation) and then apply the top–down scheme to look for particular objects inside these regions. In Image Understanding Systems of complex scenes (such as outdoor scenes), a top–down goal-directed approach is often required to correct errors from an initial bottom–up strategy [4,10].

This paper discusses the most relevant systems developed in recent years to the recognition of natural objects in outdoor environments. Special emphasis is placed on the way the systems perform the segmentation of outdoor scenes and the recognition of natural objects. In general, all the surveyed systems can also be considered as knowledge-based vision systems that make use of a large variety of knowledge representations, ranging from simple vectors to complex structures. Notice that there is no attempt to survey the different knowledge representation techniques; that has been carried out in other works [11–14]. The rest of this paper is structured as follows: assumptions, a scenario classification, and the related work, which concludes the introduction. Section 2 discusses strategies, and defines and classifies the surveyed systems. A detailed summary highlighting the principal characteristics of the analysed approaches is given in a table at the end of the section. In Section 3, the role of colour in outdoor vision systems is reviewed. Finally, the paper ends with some conclusions and suggests promising directions.

1.1. Assumptions and classifications

All the analysed systems in this survey use as input static colour images representing outdoor scenes. The approaches that use other sensor systems such as sonar, lasers or scanners have been excluded. The reason is, although it is widely known that these sensors are very useful in obtaining three-dimensional (3D) information, they do not provide colour information that is useful in identifying the nature of objects. All the reviewed work deals with characteristic ground-level images, like the ones in Fig. 2. In this sense, in Refs. [14,15] we proposed to classify the images to be described into four kinds of outdoor environments.

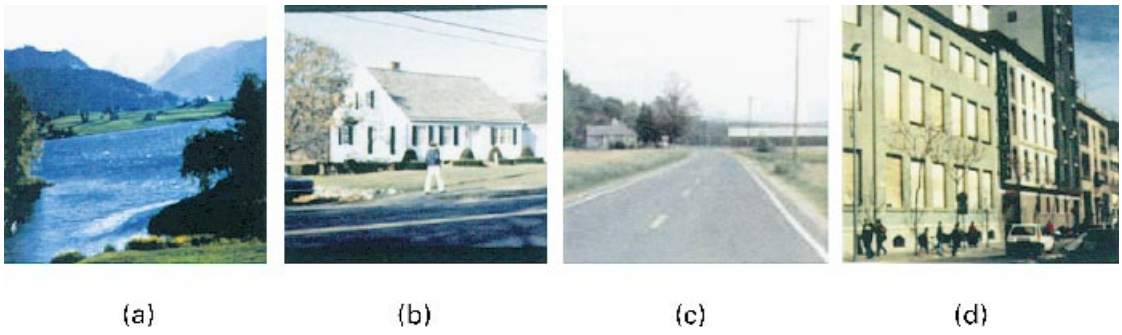


Fig. 2. An outdoor scenario classification, with images representing: (a) a natural scene; (b) a house scene; (c) a road scene; and (d) an urban scene.

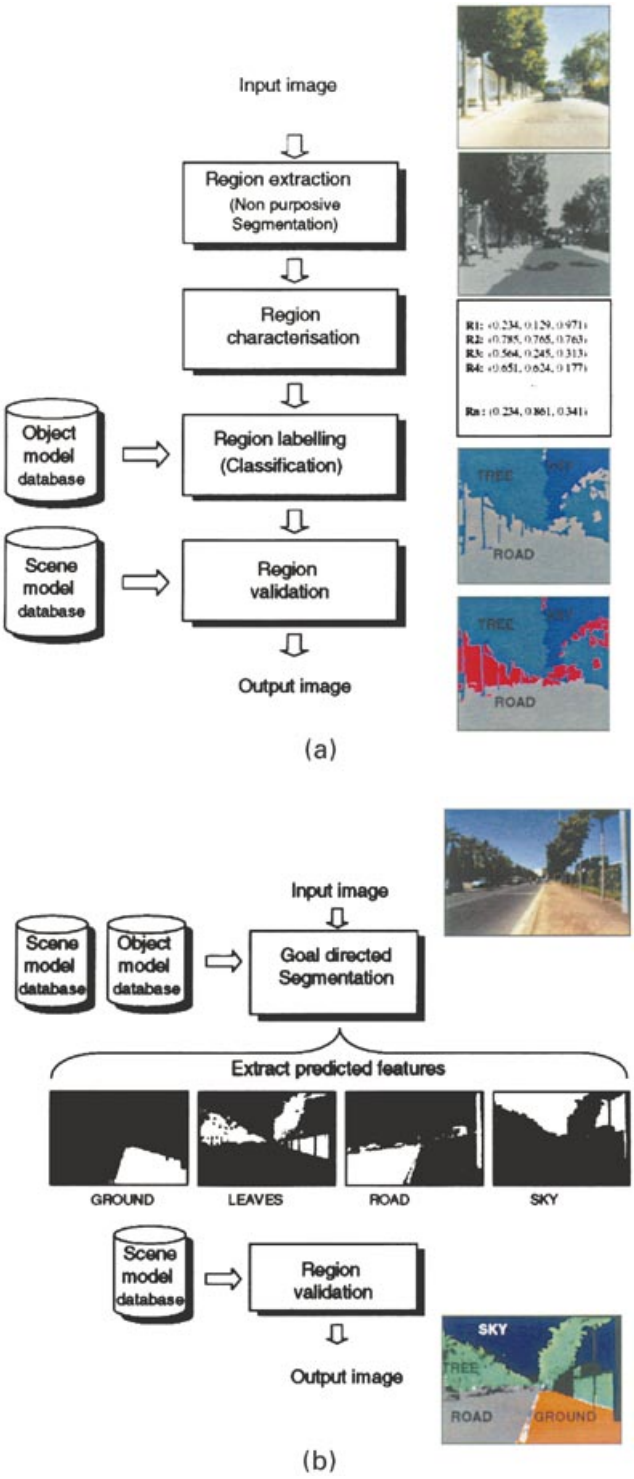


Fig. 3. A graphic example depicting the typical procedures of (a) bottom–up and (b) top–down approaches. In addition to the use of non-purposive or goal directed segmentation, both systems are also differentiated by the moment they incorporate the scene and object knowledge.

Natural scenes, which represent forests or fields, where in principle, only natural objects appear such as grass, ground, bushes, trees, sky and so on.
House scenes, which contain a typical house with a garden, trees and a road just in front of the main entrance.
Road scenes, which are represented by images taken on a road. The road is the essential element, and the remaining objects are often spatially referenced to it.
Urban scenes, which contain a typical urban setting such as pavement, sidewalk, buildings, cars, etc.

Fig. 2 shows four prototype images, with a different structuring level according to the type of scene. Therefore, urban and road scenes are more structured than house scenes, and obviously, more than natural scenes. Similarly, the presence of artificial (man-made) or natural elements vary in function of the scene and image. In other words, two images representing urban scenes do not necessarily contain the same objects, although it can be expected that there will be many similarities. Artificial objects such as cars, buildings, traffic signs and so on, are normally characterized by their straight shapes and homogeneous surface colour. Natural objects, such as trees, bushes, grass, etc, are normally highly variable in features, and are usually characterized by textured surfaces and dispersed colours. The information concerning the specific characteristics of objects is contained in the object model databases. On the other hand, the list of objects that are expected to be in an image and their relationships are contained in the scene model databases. The way the systems organize and store scene knowledge can vary from one to another, but the most used structures are rules and graph-like structures (graph-like includes nets, semantic nets, associative nets, tree structures,…). The use of scene models is required to perform several tasks such as to validate some initial results, to ensure a consistent description of the image, or even to guide the process of recognition (i.e. the invocation network of VISIONS/Schema System [16]). Therefore, scene knowledge plays a significant role in the selected strategy.

1.2. Related work

Some other surveys of outdoor scene understanding can be found in literature. However, these emphasize the concepts involved rather than the application itself (outdoor environments). Following this philosophy, Haralick and Shapiro [4] emphasized knowledge representations, control strategies and information integration. They discussed several key systems in order to illustrate various techniques and applications. In 1996, the VISIONS Group critically reviewed the core issues of knowledge-based vision systems [17]. They argued that knowledge-directed vision systems were typically limited for two reasons. First, low- and mid-level vision procedures used to perform the basic tasks of vision, were not mature enough at the time to support the ambitious interpretation goals of these systems. Second, the knowledge engineering paradigm used to collect knowledge (often a manual labour), was inadequate for gathering the large amounts of knowledge needed for more general systems. Draper et al. also suggested the changes they would have made if, in 1996, the VISIONS/Schema System had been designed again. Recently, Buxton [18] reviewed

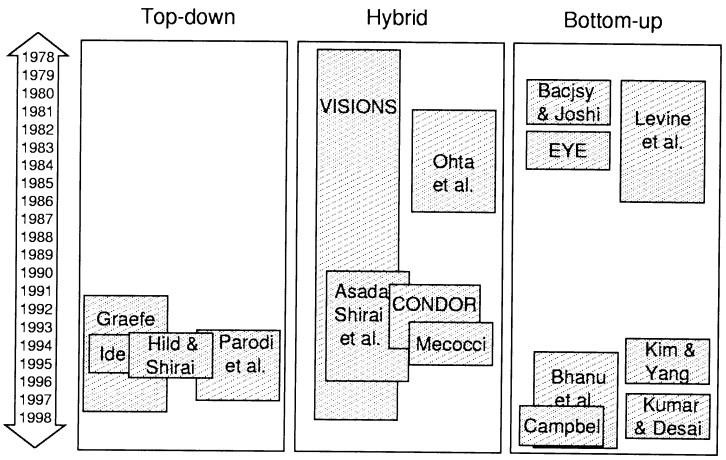


Fig. 4. A classification of a set of representative outdoor vision systems according to their strategy.

several vision systems with the view to give promising directions for future research. She emphasized the role of context, control and learning. She also argued that reasoning is the main focus of the work in visual interpretation and distinguishes four major approaches: constraint-based vision, model-based vision, formal logic and probabilistic frameworks. Several key systems illustrate these four approaches. Crevier and Lepage [13] reviewed the role of knowledge in knowledge-based vision systems. They classified the kinds of knowledge required for image understanding and examined how these kinds of knowledge have been represented. Crevier and Lepage pointed out promising trends, such as a standardized image understanding environment, agent-based representation and automatic learning techniques.

2. Bottom–up versus top–down approaches for object recognition

A key difference between bottom–up and top–down is the use they make of their model databases and, more specifically, in which part of the description process incorporate world knowledge (model databases) as depicted in Fig. 3 which shows the general procedures for both approaches. After obtaining a set of distinct regions by means of non-purposive segmentation algorithms, a bottom–up system extracts a feature vector for each obtained region. For example, the feature vector of the region named R1 in Fig. 3(a) is represented by a set of values corresponding to the measured parameters (colour, texture, shape) of that region. At this point, object model databases are needed in order to assign a label to each region. Finally, the overall consistency of the results are validated according to a scene model database. On the other hand, in the top–down approach the use of object and scene knowledge start in earlier stages. This knowledge rules the segmentation process as well as the later stage of region validation.

In spite of the restricted use of knowledge that bottom–up systems manage, they constitute a valid approach for handling unexpected (unmodelled) objects. An ideal bottom–up system would be able to give a complete description of a scene providing not only labelled regions but also the region descriptions for the unlabelled ones as feature vectors. This is an effective quality when the systems are applied to environments with a very poor knowledge. In contrast to this capability, bottom–up depends a great deal on its segmentation algorithm (non-purposive) and, as of this date, there are no perfect algorithms to solve this problem in outdoor scenes. To deal with this, the top–down approach does enable objects to be found through specific methods with the ability to handle intra-class variations, exceptions and particular cases. This is especially interesting in outdoor scenes because a scene can change its chromatic characteristics in a few seconds due to weather phenomena, time of day, seasons and shadows. These difficulties are hard to treat by using a pure bottom–up strategy. Therefore, top–down strategy will be more efficient since it is goal directed. As pointed out by Thorpe [3], general-purpose sensing is very difficult, but individual specialized modules dedicated to a particular task are continually gaining power. However, it does not imply that general-purpose perception is obsolete.

In the following, we will give a description of several key top–down, hybrid and bottom–up systems. Special emphasis is made on how these systems perform segmentation, characterisation and labelling processes. Segmentation is a key issue in object recognition, scene understanding and image understanding. The better the segmentation process, the easier the process of understanding and recognition. However, a perfect segmentation does not imply a perfect interpretation, as this only constitutes the first stage of in the whole description process. The available knowledge concerning objects and scenes, the features which characterise the regions, and the labelling techniques also play a central role in the description process. Segmentation in the bottom–up approach is carried out without using semantics related to the expected objects and/or scenes whereas in the top–down approach, the segmentation is specialized and

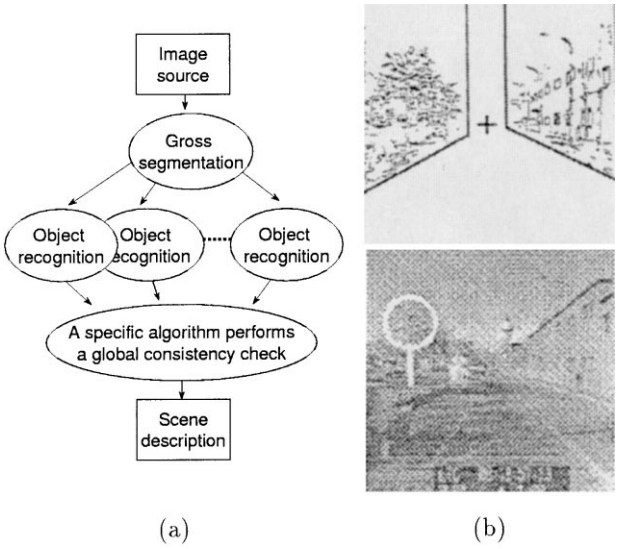


Fig. 5. Scheme of the top–down strategy system proposed by Campani et al. and Parodi and Piccioli for a traffic scene description (a), and an example of its performance on recognition of trees along the road borders (b): the polygonal approximation of the edges which belong to the lateral regions (top), and a recovered tree (bottom).

uses all the available knowledge related to the expected object and/or scene. Concerning characterization, bottom–up approaches extract features from regions, all of which are characterized by the same attributes, while in top–down approaches the characterization is specific because the regions themselves are extracted in a purposive way. Consequently, specific feature vectors are selected for each object model. Finally, the labelling methods in bottom–up systems treat all the objects equally, while top–down systems apply specific methods for each object to be recognized. The bottom–up approaches are based on regions and their objective consists of labelling each region. These approaches may cause some regions to be left without a label or have more than one label. The top–down approaches are based on the objects and search for regions that have the characteristics specified by the model. This focus can cause one pixel to have more than one label. The systems classified as hybrid

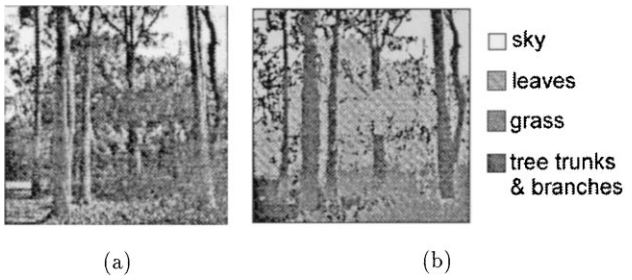


Fig. 6. An example of the results achieved by the system of Hild and Shirai: (a) a hard-edged natural scene to be described and (b) the recognized regions. This image is difficult to describe because all features are not well separated for the considered object models, and because of the high variability of the features across the object.

approaches are generally characterized by the use of general-purpose segmentation methods and specific procedures for labelling purposes, taking most of the advantages of top–down and bottom–up approaches.

Fig. 4 gives an overview of the surveyed systems, classified according to their strategy and arranged chronologically. It is interesting to note that systems based on special purpose segmentation did not arise until the 1990s. Until then, most of the efforts addressed bottom–up strategies as the best solution to build general purpose vision systems. Although the system proposed by Ohta can be considered as hybrid approach, it is ruled mainly by a bottom–up strategy, whereas only the VISIONS system had clearly begun with hybrid approach, which has remained a valid option until the present. Finally, as a result of the relevant advances AI techniques achieved during the last decade, there has been a resurgence of new bottom–up systems.

2.1. Top–down approaches

A Special Purpose Vision System acts to validate or reject an initial hypothesis about the contents of the image. With this aim, a purposive segmentation algorithm selects pixels or regions that will be candidates for the hypothesized object. Then specific algorithms are used to validate or reject the candidates. These systems are characterized by the use of constrained object or scene knowledge in all their processing tasks.

Relevant examples of two scene description systems and two natural object recognition systems have been selected. Campani et al. [19] and Parodi and Piccioli [20] proposed a method for the interpretation of traffic scenes, which combined gross segmentation and specific methods for recognising specific objects, as outlined in Fig. 5(a). Their strategy is based on treating each class of objects independently and exploiting the characterizing features of each class. Firstly, the general structure of the scene is recovered, exploiting a priori scene information and edge analysis (purposive gross segmentation). As a result, the image is partitioned into four regions: roadbed, sky and two lateral regions. Secondly, the objects are identified by specialized procedures that search for a set of selected features in determinate areas. Basically, objects are characterized by their supposed spatial disposition in the scene, their colour and edge parameters. The final step consists of gathering the information provided by the algorithms related to the different classes of objects in the scene, in order to obtain a globally consistent description of the scene in the form of a simple line drawing. Fig. 5(b) shows an example of performance by the localization of the two lateral regions (top), and the recognition of a tree plus the later reposition of the line drawn model of the tree (bottom).

Another purposeful (pure top–down) approach, proposed in 1993 by Hild and Shirai [21], consists of a two-stage interpretation system applicable to hard-edged natural

scenes. The first stage carries out specific object classification on the basis of known default models for local features such as hue, saturation and texture evaluated at pixel level. The classification of pixels is carried out by probabilistic methods using the Bayes decision rule. Then the system searches purposively for regions that satisfy qualitative constraints in the classified object images and selects object region candidates using multiple features such as shape, size, orientation, location, or spatial relationships with other objects (evaluated at region level). The second stage extends and refines the result of the first one by exploiting knowledge about the object. Finally, a symbolic description of the scene in terms of the main regions is achieved, as shown in Fig. 6.

A key top–down system for object recognition was developed by Efenberger and Graefe [22] and Regensburger and Graefe [23]. They proposed an object-oriented approach for detecting and classifying objects that could be obstacles for a mobile robot operating in an outdoor environment. Their approach is based on the principles of monocular dynamic vision for sensing the environment [24]. The goal of the system is to reach tangible results in a reasonable time, as required by mobile robots operating on roads. They decompose the global task into several independent sub-tasks, or modules, including the initial detection of possible candidates and the classification of candidates into false alarms and real physical objects. The key to their approach, in addition to the use of image sequences rather than a single one, is to apply specialized segmentation and recognition methods in order to find specific objects. The scheme used to recognize objects is roughly based on the following: first, limiting the search area; second, detecting objects in one image that are considered candidates; and third, tracking the candidates in order to validate them. Although neither colour nor texture features are used in their approach, they achieved reliable results.

As in the work of Graefe, a lot of object recognition systems use a top–down scheme. For instance, Ide et al. [25] attempted to automate the recognition of trunks or poles in urban scenes. Their method follows the pure top–down paradigm: it uses knowledge in order to limit the search area, to obtain candidates (pairs of vertical edge sequences) and finally, to validate them by using specific characteristics of poles, such as diameter and height. This system is goal directed in all its processing tasks.

In spite of the great number of top–down object recognition systems that can be found in literature (lane detection, vehicle and traffic sign recognition, military targets), there are only a very limited number of systems that try to understand the whole scene in a top–down way.

2.2. Hybrid approaches

Hybrid approaches are characterized by the use of non-purposive techniques in order to segment the input image into regions (bottom–up strategy), and a set of specialized procedures that search for selected features in that region (top–down strategy). It is important to emphasize that each object is characterized by its own features. A common final step consists of gathering the information provided by the algorithms, which is related to the different classes of objects in the scene, with the aim of obtaining a globally consistent description of the scene. We will call these systems “pure hybrid approaches” since their initial stage is general purpose, while their recognition stage is specific-object based.

From the mid-1970s the VISIONS system [7,16,26,27] has been evolving into a General Purpose Vision System for understanding images representing road and house scenes. The VISIONS Schema System provides a framework for building a general interpretation system as a distributed network of many small special-purpose interpretation systems. The VISIONS Schema System introduced the notion of a schema as an active process that encapsulates the knowledge about an object class. Each schema embeds its own memory and procedural control strategies, acting as an “expert” at recognizing one type of object. The system’s initial expectations about the world are represented by one or more “seed” schema instances that are active at the beginning of an interpretation. As these instances predict the existence of other objects, they invoke the associated schemas that in turn may invoke more schemas. In VISIONS, schema instances run as independent concurrent processes, communicating asynchronously through a global blackboard. The goal of such a schema is to collect the evidence necessary to validate or reject a hypothesis. In this sense, the VISIONS Schema System took a particular, simple view of evidence representation by using five confidence values representing different degrees of belief. Concerning object recognition, the VISIONS Group emphasizes flexible matching on a variety of characteristics, with the ability to handle exceptions and deviations. They pointed out that not all object classes are defined in terms of the same attributes, which may be used in various ways within the matching or interpretation process. The constant evolution suffered by VISIONS, with the continuous flow of new ideas and the update of the used methodologies and techniques, has become in itself a compulsory reference work for the study and the comprehension of outdoor scenes analysis systems.

Strat and Fischler [28] proposed a system, called CONDOR, to recognize objects in natural scenes. The central idea of its architecture is a special-purpose recognizing method designed for each object-class. Despite the fact that its architecture looks for specific objects, its goal consists of understanding the whole scene using non-purposive segmentation methods. In CONDOR, the knowledge is embedded in rules as condition action pairs, named “context sets”. These are employed in three types of rules: candidate generation, candidate evaluation and consistency determination. The rules enable the building of sets of mutually consistent candidate hypotheses (named cliques) which

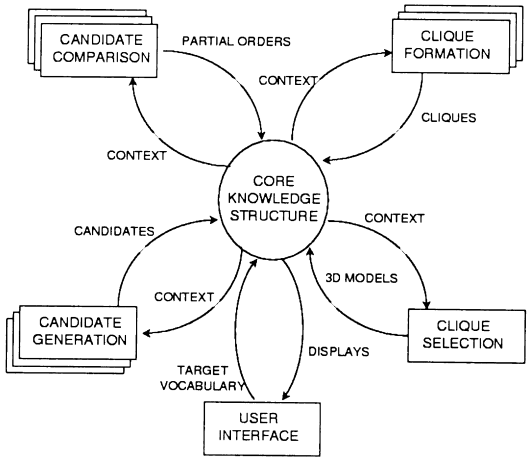


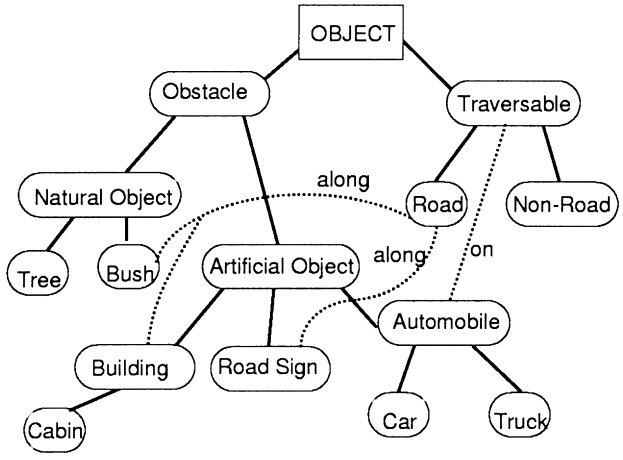
Fig. 7. General structure of the CONDOR system, the architecture of which is much like that of a blackboard system.

confer to the system its ability to deal with contexts (scenes) rather than independent objects. The addition of a candidate to a clique may provide knowledge that could trigger a previously unsatisfied context set. For example, once one bush has been recognized, it is a good idea to look specifically for similar bushes in the image. Fig. 7 depicts the general structure of the CONDOR system, where the processes act like demons watching over the knowledge database, invoking themselves when their contextual requirements are satisfied.

Four different processes: candidate generation, candidate comparison, clique formation and clique selection, interact through a shared data structure (the core) by exchanging knowledge (the context) and specific data structures (candidates, partial orders, cliques and 3D models). CONDOR represents a successful attempt to deal with the variability of outdoor images by associating a collection of simple procedures to each object model. Each procedure is competent only in some restricted contexts, but collectively these procedures offer the potential of recognizing a feature in a wide range of contexts.

From the late 1980s, Asada and Shirai [29], Hirata et al. [30] and Taniguchi et al. [31] developed another General Purpose Vision System intended to interpret colour images representing urban scenes by using 3D information. Their philosophy is more classical, since they proposed a sequential scheme that acts in two stages. First, using a general-purpose algorithm, the image is divided into regions having uniform brightness and colour. Second, the interpretation process is based on the search for specific objects in a predetermined order. For instance, the interpretation system starts from extracting road, sky and trees independently and then tries to recognize other related objects. The scene knowledge is organized as a network (graph) of object models (named frame structures) where each object model describes semantic constraints and the relationship to other objects. Fig. 8 shows the proposed road scene model and two examples of natural object models described as frames. It is interesting to note that the road scene model differentiates between natural and artificial objects, suggesting that they would require a different treatment. For instance, they consider that determining geometric properties for natural objects is not as significant as for artificial ones. An example of application of the proposed system is shown in Fig. 9, which illustrates some results achieved on a road scene. As output, the system provides the recognized objects as well as the non-interpreted regions.

Unlike the three previous hybrid systems, Ohta et al. [32] proposed a non-pure hybrid system. Although their proposal



(a)

```
(Road Frame
(IS-A traversable-object)
(CONSIST-OF a-pair-of-road-boundaries
center-part)
(PROPERTY constant-width smooth-curve)
(RELATION-TO-OTHER-OBJECTS
   (ON-THE-ROAD automobile)
   (ALONG-THE-ROAD bush building
automobile other-objects)))

(Bush Frame
   (IS-A natural-object)
   (PROPERTY
       (LOWER-THAN tree
   (RELATION-TO-OTHER-OBJECTS
   (ALONG road-boundary)))
```

(b)

Fig. 8. A graph structure (network of frame structure) depicting (a) the relationships between the expected objects of the scene and (b) two examples of object models represented as frame structures.

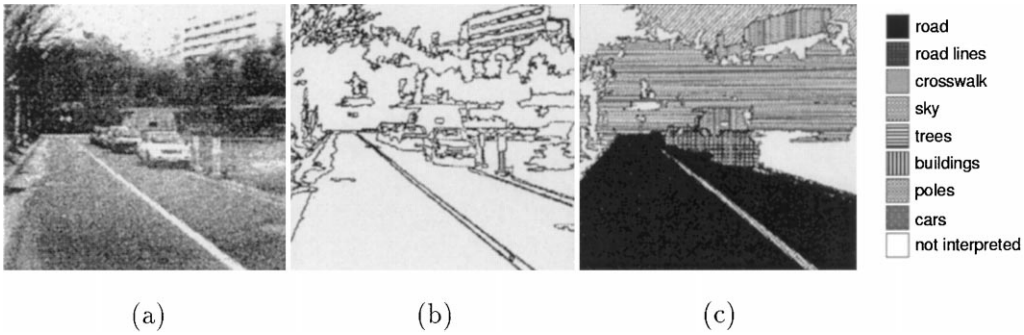


Fig. 9. (a) An urban image to be described, (b) the segmented results and (c) the final described image.

incorporates top–down control, their system is mainly governed by a bottom–up strategy as shown in Fig. 10(a). In fact, the system is considered to be a region analyser for colour images of urban scenes. It uses the top–down strategy in order to produce more detailed descriptions of some regions. For instance, once a region is labelled as a "building", it is a good idea to look for "windows" in a purposive way in that region. The knowledge of the task world is represented by rules, while the knowledge associated with the models, which describe properties and relations among objects, is organized as a semantic network (a graph structure). The bottom–up process can be summarized as follows.

- The image is first over-segmented by a region-splitting algorithm using 1D multi-histograms. Afterwards, a merging process organizes the data into a structured data network composed by patches (coherent regions).
- The largest patches are selected in order to generate a set of object labels and their respective degree of correctness. This set of tuples (region, label, degree of correctness) are called "plan", analogous to the clique structure in the CONDOR system.
- In order to ensure a consistent description of the image, the plan is evaluated by fuzzy rules contained in the production system. Each rule has a fuzzy predicate that describes a property of a certain object or a relationship among objects.

Another non-pure hybrid approach was proposed in 1995 by Gamba et al. [33] and Mecocci et al. [34], who presented an almost bottom–up system capable of giving a simple outdoor and indoor scene description in order to provide some help for blind people to navigate. The system works with some a priori knowledge about the scene structure (corridor and road scenes are analysed), and carries out a

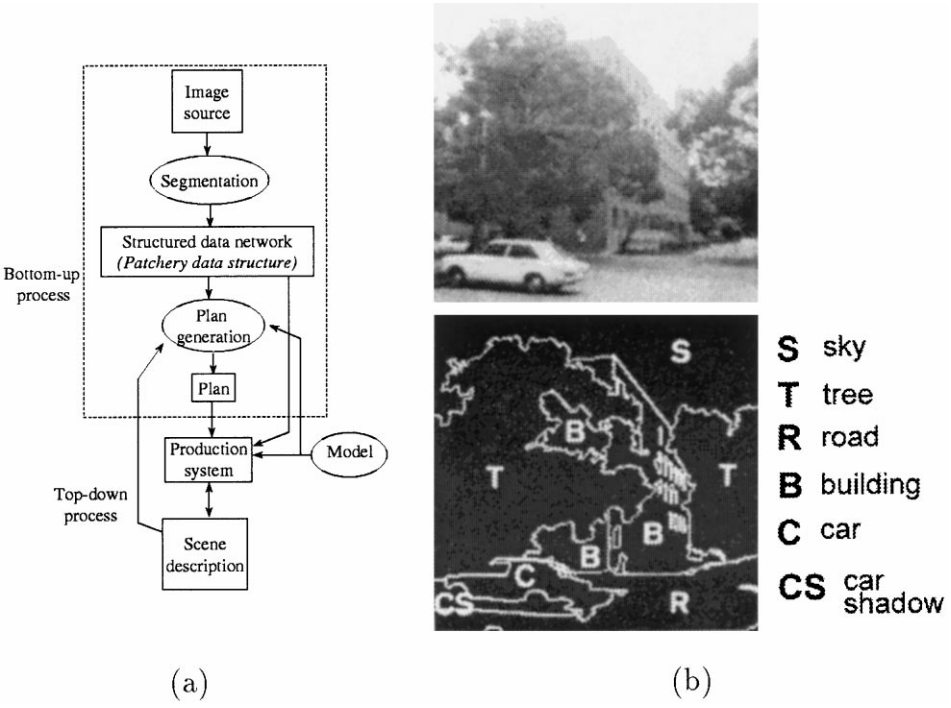


Fig. 10. (a) Outline of the Ohta region analyser, where bottom–up and top–down strategies cooperate, and (b) the results obtained on a urban scene. The gross regions are labelled by the bottom–up process (S,T,B,C), while top–down processes allow the achievement of more detailed results (CS).

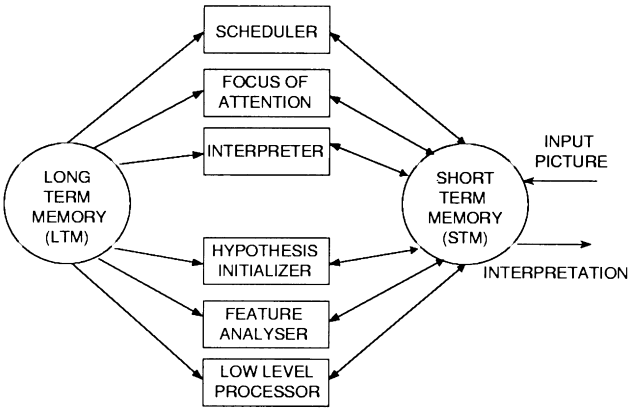


Fig. 11. Outline of the scheme proposed by Levine et al.

preliminary classification of the scene components into two classes: the man-made object class, characterized by planar surfaces and straight lines, and the fractal class, which represents natural objects, characterized by the lack of simple planar surfaces. The system recognizes fractal regions, lateral vertical surfaces, frontal vertical surfaces, horizontal surfaces, and unrecognized regions, and takes into account previously segmented regions and segments. The interpretation ends with constraint region rules (named criteria) and with rules for propagating adjacent regions.

2.3. Bottom–up approaches

In general, bottom–up systems model natural objects using the same attributes as other objects. Therefore, the high variability of these objects is not specially treated. As a result, the performance of the systems is highly dependent on the outdoor conditions.

In 1978, Bajcsy and Joshi [35] presented a world model for natural outdoor scenes in the framework of production rules. The system is based on finding relationships in the real world that impose a partial order on a set of objects. In order to accomplish these relationships, the system is structured in a database, a set of rules and an interpreter. The database is also structured in rules (named facts) that contain specific object knowledge, spatial relations between objects and subclasses (objects that are considered as a part of other objects). The interpreter carries out the matching between the database facts, as well as executing actions which arise in consequence of the applied rules. Although this is an early work on the outdoor scene description domain, it must be pointed out that the current validity of the ideas contained on this proposal, such as the use of an interactive query method to interface with the system, and the recognition is performed by checking only partial features. The later also allows, to some degree, dealing with the variability of natural objects.

Levine [36] and Levine and Saheen [37] proposed a scheme that consisted of a collection of analysis processors, each specialized in a particular task. The processors compete and cooperate in an attempt to determine the most appropriate label to assign to each region previously obtained by a simple one-pass region-growing algorithm. Conscious of the importance of segmentation, they proposed a more sophisticated algorithm that simultaneously deals with both edges and regions [38]. The system is based on the design of a rule-based expert system in the form of a modular set of processes and two associative memories, as depicted in Fig. 11.

The input image, the segmentation data, and the output are stored in the short-term memory, while long-term memory embodies the scene knowledge structured as rules. The segmentation algorithm is contained in the low-level processor while the features associated with each region are computed by the feature analyser processor. A set of initial hypotheses of each segmented region is then generated by the hypothesis initializer processor which attempts to match each region to all the object models stored in the long-term memory. These hypothesis are then verified by constraint relations given by rules which describe the world in terms of the conditions under which the objects may realistically coexist. These rules are contained in the interpreter processor and formally specify spatial and/or colour constraints involving two objects. The monitoring of the current state of knowledge of the scene is given by the focus of attention processor while the scheduler processor calls upon the appropriate processor to intervene. The system proposed by Levine et al. is highlighted by its extensibility, modularity and separability, which allowed a successful implementation. The first one enabled the addition of model data as well as control information while modularity focused basically on the model knowledge and the control mechanisms. Finally, the complete separability of the object and scene information from the program modules led to a suitable design. The experiments carried out on a typical house scene demonstrated the reliability of this scheme.

The EYE system, proposed in 1981 by Douglass [39], emphasized the integration of depth with semantic information to form a 3D model of a house scene. The images were first pre-processed and segmented into a set of regions of approximately uniform colour and texture using a segmentation algorithm that combined the advantages of edge detection and region growing. The final scene model is obtained by combining an iterative relaxation process with some hypothesized 3D information. In order to label the regions Douglass uses multiple criteria such as occlusion analysis, size evaluation, texture gradients, shadow and highlight analysis and linear perspective. The EYE system provides an interesting interpolation of object and scene model into a single representation structure which includes surface description of an object (colour, texture, size, boundary shape, curvature,…), logical relationship between objects (part–whole and classmembership), and spatial information on the 3D relationships between objects and their parts. Fig. 12 shows a detail of the proposed

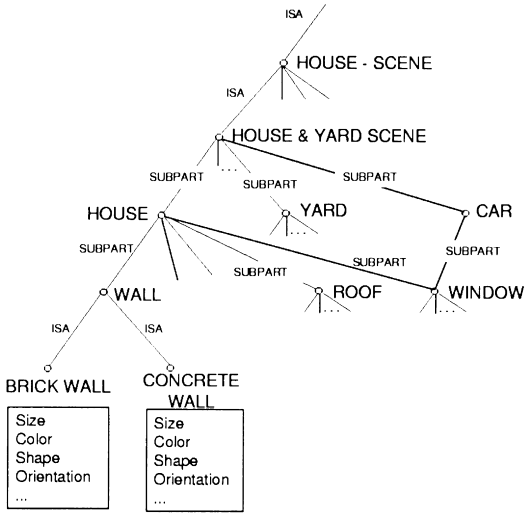


Fig. 12. A portion of the associative net used by the EYE system ISA links are used to denote class membership while SUBPART links not only define which parts comprise the whole, they also indicate the relative orientation between the parts in terms of pan and tilt angles, relative separation, and points and angles of intersection.

graph-like structure (named associative net) for a house scene where nodes are objects and the links are logical and spatial relationships among objects. The EYE system goes beyond most other existing image understanding systems that only intend to label regions. Indeed, EYE is the first system that attempts to interpret regions as 3D surfaces and to join the surfaces into a consistent model of the environment.

Since 1993 Kim and Yang [40,41] have been working on segmentation and labelling based on the Markov Random Field models, as an alternative to the relaxation labelling. Fig. 13 shows the global overview of their proposed strategy, which starts with an algorithm that segments the image into a set of disjoint regions. A region adjacency graph is then constructed from the resulting segmented regions

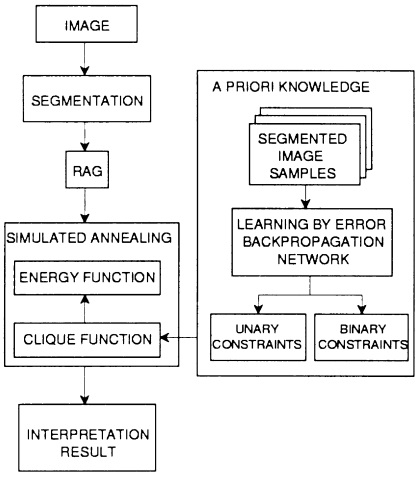


Fig. 13. Global overview of the Kim and Yang scene description system.

based on the spatial adjacencies. The problem is then formulated by defining region labels that are modelled as a Markov Random Field on the corresponding Region Adjacency Graph. The knowledge concerning the scene is incorporated into an energy function composed by appropriate clique functions which constrain the possible labels for the regions. Here a clique is a list of regions which border on each other. If the interpretation of the regions in a clique tends to be consistent with the feature measurements and the scene knowledge, the clique function decreases resulting in a decrease in the energy function. Optimal labelling results are obtained by finding a labelling configuration which minimizes the energy function by using a simulated annealing procedure. In designing clique functions, they consider only two types of clique encoded as unary and binary constraints. Unary constraints are used to recognize a region based on the feature of the region itself while binary constraints denote spatial adjacency compatibility between all the couples of object models. In order to handle the feature variability of outdoor scenes, they propose finding appropriate parameter values of the clique functions by error backpropagation networks. Formulating the image labelling problem based on the Markov Random Field model provides a systematic way of representing domain knowledge by means of clique functions, and facilitates finding optimal labelling through a general optimization algorithm such as simulated annealing. The proposal of Kim and Yang has evolved into an integrated scheme in which segmentation and interpretation co-operate in a simultaneous optimization process. Kumar and Desai [42] presented a similar work in 1996 in which they proposed a scheme for joint segmentation and interpretation in a multi-resolution framework by using the wavelet transform of the input image.

Recently, Campbell et al. [43] developed a system that is capable of labelling objects in road and urban scenes, thereby enabling image databases to be queried on scene content. The method is based by first segmenting the image using the k-means algorithm, which is demonstrated by the optimal method out of several segmentation algorithms. Each region is then described using a set of 28 features to represent the visual properties of the region (colour, texture, shape, size and position). Using a large database of ground-truth labelled images, a neural network (a multilayer perceptron) has been trained to act as a pattern classifier. The optimal network architecture, with 28 inputs and 11 label outputs, was found to have 24 nodes in the hidden layer. The system has been tested on a large number of previously unseen scenes and, on average, correctly classifies over 90% of the image area.

Neither the Campbell system nor any of the previous bottom–up systems consider the problem of outdoor conditions, such as metereological phenomena, time of day or time of year. In order to handle these problems, Bhanu et al. [44] presented an adaptive image segmentation system that incorporates a genetic algorithm to adapt the segmentation process to changes in image characteristics caused by

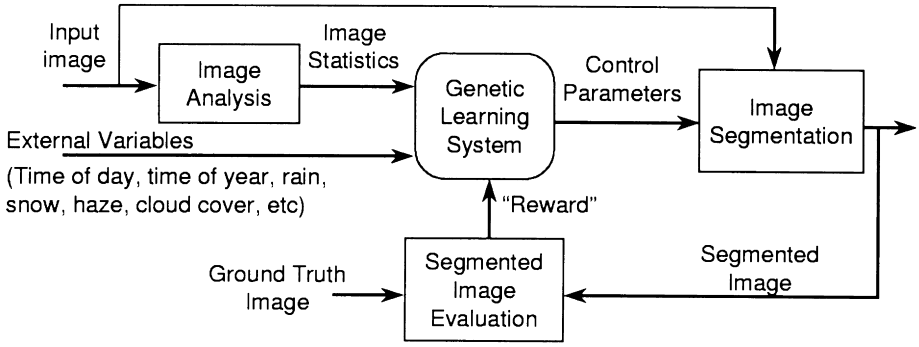


Fig. 14. Block diagram of the adaptive image segmentation process proposed by Bhanu et al.

variable environmental conditions. They consider an unstructured road scene belonging to a natural environment. The basis of their system is the Phoenix segmentation algorithm developed at Carnegie-Mellon University. This consists of recursive region-splitting that contains 17 different control parameters. Based on experimentation, they selected the two most critical parameters that affect the overall results of the segmentation process. The aim of their approach is to infer these parameters automatically. Once the image statistics and external variables (time of day and weather) have been obtained, a genetic learning component selects an initial set of segmentation algorithm parameters. The system needs a module to carry out the task of evaluating the segmentation results. Bhanu et al. proposed five different quality measures to determine the overall fitness for a particular parameter set. Fig. 14 shows a block diagram of the proposed algorithm which was tested and provided high quality segmentation results. A further improvement of that work was presented in 1998 [45], where it was attempted to automatically obtain the segmentation parameters by using a neural net. They introduced reinforcement learning as a method of improving the final performance of the system.

2.4. Summary

As a result of the detailed analysis of the systems, it can be stated that the main advantage of top–down approaches is that segmentation, characterization and labelling are specific and will permit handling the complexity of outdoor conditions. It is demonstrated that successful systems must take into account such variations, otherwise they are clearly restricted to the description of only some predetermined images in specific outdoor conditions (i.e. sunny images without shadows, only green trees). The disadvantages of top–down approaches are basically related to their inability to handle unmodelled objects and scenes. In general, the lack of knowledge will result in the impossibility of setting forth an initial hypothesis which is the basis of the top–down strategy. As far as the bottom–up approaches are concerned, their main advantage is the ability to handle unmodelled objects or scenes, being capable of giving descriptions of unrecognized regions. This can become a useful feature for learning tasks, not properly exploited until now by the existing bottom–up systems. In some way, using only the general-purpose methods became the main disadvantage of such approaches, because they are not able to deal with the great amount of specific cases which outdoor scenes can exhibit. A common drawback for both approaches consists of the emergence of labelling conflicts; at a pixel level on the purposive segmentation processes in top–down systems and at a region level on region labelling processes in bottom–up systems. The hybrid approaches take advantage of top–down and bottom–up strategy, since they are able to combine non-purposive segmentation with hypothesize-and-test methodology. Nevertheless, they also suffer from the labelling conflicts already mentioned.

For the purpose of providing an overview of the presented systems, Table 1 summarizes some of their most relevant features. The first column identifies the different systems by giving authors names with referred papers including the name of the system, if it exists. The next three columns categorize the strategy used, the type of scene and whether segmentation is purposive or not. The next column refers to how systems handle colour. In the next section we refer to this subject in depth. The next two columns give the list of recognisable objects for each analysed system and the features selected for object characterization. Despite the number of considered objects which cannot be understood as a criterion to evaluate the complexity of the system, it often gives an idea of what the problem is that the systems try to solve. Finally, the last three columns summarize the segmentation process and which algorithms are used, how object and scene knowledge is structured, and how the labelling process is performed.

3. Use of colour

Concerning segmentation and object characterization, outdoor scenes are especially complex to treat in terms of the lighting conditions. It is well known that chromatic characteristics of natural elements are not stable. As an example, Fig. 15 demonstrates how seasons affect an

Table 1
A summary of the outdoor vision systems analysed (strategy: T–D (top–down); B–U (bottom–up); H (hybrid); segmentation: P (purposive); NP (non-purposive); scene: U (urban); N (natural); H (house); R (road)

System identification	Strategy	Scene	Segmentation	Colour space	List and number of objects to recognise	Object charaterisation	Segmentation techniques	Object and scene knowledge representation	Labelling engine
Campani et al. [19] and Parodiand Piccioli [20]	T–D	U	P	Colour	Road boundaries, road signs, crosswalks, vehicles, buildings, trees. (6)	Spatial disposition in the scene, colour, and segments	A specific gross segmentation based on edges, colour, vanishing point detection	Encode the specific features of the objects inside the algorithm	A specific procedure that finds specific features
Hild and Shirai [21]	T–D	N	P	Hue, brightness	Tree trunks, branches, grass, leaves, sky. (5)	Shape, orientation, position, hue and texel dir.	Likelihood pixel classification	Feature vectors	Select the best candidate by shape processing
Efenbergerand Graefe [22] Regensburger and graefe [23]	T–D	R	P	Gray levels	Road, tree trunks, tree, rock, barrel, car. (6)	Edges and gray-levels	Edges and gray-levels	Sub-sampled images and prominent edge elements	2D correlation functions and special purposive methods
Ide et al. [25]	T–D	U	P	Gray levels	Poles. (1)	Diameter, height and layout	Vertical straight line detection	Encode the specific features of the object to recognize inside the algorithm	Recognition based on specific characteristics
VISIONS Draper et al. [16,26] and Hanson and Riseman [7,27]	H	R,H	NP	R, G, B, $(R + G + B)/3$, $(2R - G - B)$, $(2G - B - R)$, $(2B - R - G)$, H, S, V,Y, I, Q	ROAD SCENES: Sky, foliage, shoulder, trunk, sign-post, wire, warning-sign, phonepole, road, roof, building, roadline, grass, unknown. (14) HOUSE SCENES:Sky, tree, grass, bush, shutter, wire, house-wall, roof, roadline, road, film-border. (11)	Colour, texture, shape, size and spatial relations among objects	Combined histograming and region merging method	An schema and two graphs (part-of, invocation) for scenes, an schemas for the objects	Schemas based on solving specific confidence functions in order to verify hypothesis
CONDOR Strat [28]	H	N	NP	R, G, B	Geometric horizon, complete sky, complete ground, skyline, sky, ground, raised object, follage, bush, tree trunk, tree crown, tree, trall, grass. (14)	Colour, texture, geometric shapes, and spatial relations among objects	A set of specific context sets (rules) which include texture operators, edge operators, histogramming techniques, …	Semantic networks and context sets (rules as pairs of conditions actions)	a) Candidate comparison by likelihood methods b) Grouping mutually consistent hypothesis c) Select the best description
Asada and Shirai [29] Hirata et al. [30,31] Taniguchi et al. [31]	H	R	NP	(T, q, S), brightness, hue and saturation	Road, road lines, crosswalk, sky, trees, buildings, poles, cars, truck, not interpreted. (10)	Colour, heights, range information, and spatial relations among objects	Colour (they design an specific split and merge algorithm)	Each object is represented as a frame. A scene is represented as a network of frame structures	Rules and specific methods
Ohta et al. [47]	H	U	NP	$I1 = (R + G + B)/3$, $I2' = R - B$, $I3' = (2GR - B)/2$	Sky, tree, building, road, unknown, car, car shadow, building window. (8)	Colour, texture, position and shape	Region splitting using multihistograms	Semantic network	Instantiate production rules (which the condition part is a fuzzy predicate)
Gamba et al. [33]Mecocci et al. [34]	H	U	NP	Gray levels	Natural objects, lateral vertical surfaces, frontal vertical surfaces, horizontal surfaces, vanishing point, unrecognised. (6)	Segments and region localization respect to vanishing point	A specific region growing algorithm, edge analysis and vanishing point detection	Encode the specific features of the objects inside the algorithm	Criterias (rules)
Bajcsy et al. [49]	B–U	N	NP	Colour	Ground, sky, horizon skillines, tree. (4)	Colour, sizes and spatial relations among objects	Colour separation	Rules (named facts)	Partial match operations on rules and facts
Levine [36] Levine and Shaheen [37] Nazif and Levine [38]	B–U	H	NP	R, G, B	Bushes, car, fence, grass, road, roof, shadow, window. (12)	Colour and spatial relations among objects	Regions, edges and area	Rules	Constraint relations rules in order to verify the hypothesis
Douglass [39]	B–U	H	NP	H, S, I	Trees, house, grass, sky, car, street, window, brick wall, concrete wall, roof, ground. (11)	Colour, texture, boundary shape, size, curvature, and orientation	Edges, colour and texture (the algorithm combines edge detection and region growing)	Associative net (semantic net), where nodes are objects and links are logical and spatial relationships between objects	Probabilistic methods
Kim and Yang [40,41]	B–U	R, U	NP	R, g, b, r–b, intensity and saturation	Sky, foliage, road, grass, wall, roadline, window, footway, tree. (9)	Spatial disposition in the scene, colour, texture, and geometric features	Region growing	Feature vectors and graphs	Labelling the nodes of a Region Adjacency Graph by using a Simulated Annealing algorithm
Kumar and Desai [42]	B–U	R	NP	Grey level	Sky, tree, sidewalk, road. (4)	Spatial disposition in the scene, grey level, texture, and geometric features	K-means clustering	Feature vectors and graphs	Labelling the nodes of a Region Adjacency Graph by using a Simulated Annealing algorithm
Bhanu et al. [44] and Peng and Bhanu[45]	B–U	R	NP	R, G, B			A genetic algorithm selects parameters automatically of a region splitting algorithm		
Campbell et al. [43]	B–U	R	NP	$(3R + 6G + B)/10$, $(R - G + 1)/2$ and $(R + G - 2B + 2)/4$	Sky, vegetation, road marking, road, pavement, building, fence/ wall, road sign, signs/poles, shadow, mobile objects. (11)	28 features including colour, texture (Isotropic Gabor), shape and contextual information	K-means clustering	Feature vectors	Neural net

Fig. 15. Colour images of a tree through the seasons.

outdoor scene, where the colour, density and texture of objects can vary considerably. Moreover, the figure also shows how the spectral colour range of trees is affected by the continuous progression of a season, smoothly shifting from green to red. Buluswar and Draper [46] provide a survey detailing analysis and causes of colour variation due to illumination effects on outdoor images. The apparent colour of an object depends on illuminant colour, the reflectance of the object, illumination geometry (orientation of the surface normal with respect to the illuminant), viewing geometry (orientation of the surface normal with respect to the sensor) and sensor parameters. In outdoor images, at different times of the day, under different weather conditions, and at various positions and orientations of the object and camera, the apparent colour of an object can be different. Human beings have an adaptative mechanism called colour constancy that compensates for this colour shift. Unfortunately, no corresponding adaptative mechanism exists in machine vision systems, and the notion of colour associated with an object is precise only within the context of scene contions. Ideal object characterization will require flexible and dynamic models in order to adapt to the different phenomena.

Colour, being undoubtedly one of the most interesting characteristics of the natural world, can be computationally treated in many different ways. In many cases, the basic RGB components may provide very valuable information about the environment. However, the perceptual models, such as CIE (L,a,b) or HSI, are more intuitive and therefore enable the extraction of characteristics according to the model of human perception. The complexity of outdoor images emphasizes the need of the system to select a "good" colour space, which is of extreme importance to the segmentation tasks. Therefore, it is necessary to formulate the following question: *what is the best colour space to be applied in order to segment an image representing an outdoor scene?* This question has neither a single, nor a perfect solution. The colour space suitable for one segmentation algorithm is not suitable for others. Unfortunately, no better solution has been found. In the work reviewed, a wide range of proposals has been presented. Some authors, like Ohta, have proposed their own colour space. Ohta et al. [47] proposed a set of colour features, $I_1 = (R + G + B)/3$, $I_2' = (R - B)$ and $I_3' = (2G - R - B)/2$. The effectiveness of their colour feature set was discussed by a comparative study with other sets of colour spaces. The comparison was performed in terms of both the quality of segmentation results and the calculation involved in transforming data of *R*, *G* and *B* to other forms. Celenk [48] proposed a colour-clustering algorithm for segmenting colour images of natural scenes. He has performed a colour analysis method and proposed operating with the CIE ($L*, a*, b*$) uniform colour coordinate system L*, H^o and $C*$ (Luminance, Hue and Chroma). He argued that for colour clustering, it is

desirable that the selected colour features define a space having uniform characteristics. The proposed colour space approximately satisfies this property. Other authors have presented concrete solutions to concrete problems: Bajcsy et al. [49] proposed an approach to colour segmentation with the detection and separation of highlights by using hue and saturation. Luo et al. [50] proposed a fractal feature for textures of natural images that is stable under changes in lighting conditions. She and Huang [51] proposed working with the CIE-*uv* space and a texture parameter characterized by the fractal behaviour of the colour bands *R*, *G* and *B*. The fractal dimension measure is relatively invariant to changes in the image scale and contrast, while the chromaticity in the CIE-*uv* space reduces the effects of shadows and illumination gradients. Buluswar and Draper [52] present a technique that uses training images of an object under daylight with the view to learn about the way the colour of an object shifts along the RGB space. Similarly, Mori et al. [53] proposed the use of the *r*–*b* model (where *r* and *b* denote normalized red and blue components) in order to solve the problems of hue shift, due to outdoor conditions and shadows.

4. Conclusions and further work

In this paper we have reviewed some key outdoor scene understanding systems in order to point out the strengths and weaknesses of the top–down and bottom–up strategies. Special emphasis has been given to the modelling of objects and scenes, and how the high variability of outdoor scenes is treated by the systems. Top–down permits objects to be found through specific methods with the ability to handle intra-class variations, exceptions and particular cases. This is especially interesting in outdoor scenes because there is a change in their chromatic characteristics within a few seconds due to weather phenomena, time of day, seasons or shadows. These complexities have led to insurmountable difficulties with most of the bottom–up systems. Only the general purpose segmentation algorithm by Bhanu et al. takes into account the above-mentioned outdoor conditions. However, it would be desirable to include outdoor conditions on the processes of region characterization, labelling and modelling. Nevertheless, the capability of the bottom–up strategy in handling unmodelled situations is a valued feature for building reliable systems. In this sense, bottom–up systems are more suitable for the description of unknown environments because they are not constrained by prior knowledge. Furthermore, the quality of bottom–up in giving descriptions of unrecognised regions would be a useful input for further supervised learning tasks. These conclusions lead to the hybrid strategy as the best way to deal with outdoor conditions and unmodelled situations. From this study of state-of-the-art strategies, the lack of a consolidated colour space is noted. Neither colour space nor segmentation methods have proved to be the most suitable for treating images representing outdoor scenes. A wide range of proposals has been suggested ranging from general spaces for general outdoor images, to specific colour spaces for treating concrete problems.

In general, the results obtained to date by outdoor vision systems must be improved. It is widely assumed that the interpretation of scenes, which constitutes the final objective of computer vision, is a complex problem that has only been solved for very simplified scenes [54]. However, new and promising directions have arisen in relation to the concepts of growing and learning, as proposed by authors [55–57]. Those involve methods which add or update object classes to the systems in order to better describe the scene without having to recompile. Such an idea is not only based on automating the process of object characterization but on stating the procedures that are the most adequate for recognizing objects. Nevertheless, while the object-oriented paradigm has been proven to be useful in helping the development of IU, it may be unsatisfying in future due to its inability to deal with distributed computing. Distributed computing enables better response times to be obtained through parallel execution which is useful for image processing and artificial intelligence algorithms. In this sense, emerging technologies like the CORBA [58] specification may help. Following this specification, a set of objects-software can be programmed in heterogeneous programming languages, executed on heterogeneous workstations with heterogeneous operating systems and inter-operating in order to solve a concrete problem. An easy evolution of software systems could be especially useful for the IUE [59–61], since it is continuously growing with the contributions of many researchers.

As a final assessment, it can be stated that the development of a vision system with the aim of understanding the whole scene, where all the levels of vision are specific and purpose-oriented, still constitutes an important challenge. Although the surveyed systems report interesting results, there remains a great deal of work to be carried out in order to build a vision system with performances similar to that of the human eye and its visual perception system.

References

[1] A. Broggi, Vision-based driving assistance in vehicles of the future, IEEE Intelligent Systems November/December (1998) 22–23.
[2] E.D. Dickmanns, Vehicles capable of dynamic vision, in: Proceedings of the 15th International Conference on Artificial Intelligence, Nagoya, Japan, 1997.
[3] C.E. Thorpe, M. Hebert, Mobile robotics: perspectives and realities, in: Proceedings of the International Conference on Advanced Robotics, Sant Feliu de Guixols, Spain, 1995, pp. 497–506.
[4] R.M. Haralick, L.G. Shapiro, Knowledge-based vision, Computer and Robot Vision, II, Addison-Wesley, Reading, MA, 1993, pp. 493–533 chap. 19.
[5] M.A. Fischler, On the representation of natural scenes, in: A. Hanson, E. Riseman (Eds.), Computer Vision Systems, Academic Press, New York, 1978, pp. 47–52.

[6] A.M. Wallace, A comparison of approaches to high-level image interpretation, Pattern Recognition 21 (3) (1988) 241–259.
[7] A.R. Hanson, E.M. Riseman, in: Segmentation of natural scenes. A.R. Hanson, E.M. Riseman (Eds.), Computer Vision Systems, Academic Press, New York, 1978, pp. 129–163.
[8] G.J. Klinker, S.A. Shafer, T. Kanade, Color image analysis with an intrinsic reflection model, in: Proceedings of the Second IEEE International Conference on Computer Vision, December 1988, pp. 292–296.
[9] M. Mirmehdi, P.L. Palmer, J. Kittler, H. Dabis, Complex feedback strategies for hypothesis generation and verification, in: Proceedings of the Seventh British Machine Vision Conference, September 1996, pp. 123–132.
[10] T. Matsuyama, Expert systems for image processing: knowledge-based composition of image analysis processes, Computer Vision, Graphics and Image Processing 48 (1989) 22–49.
[11] D.P. Argialas, C.A. Harlow, Computational image interpretation models: an overview and a perspective, Photogrammetric Engineering and Remote Sensing 56 (6) (1990) 871–886.
[12] D.H. Ballard, C.M. Brown, Computer Vision, Prentice Hall, Englewood Cliffs, NJ, 1982.
[13] D. Crevier, R. Lepage, Knowledge-based image understanding systems: a survey, Computer Vision and Image Understanding 67 (2) (1997) 161–185.
[14] J. Martí, Aportation to the urban scenes description by means of approximate models (in Catalan), PhD thesis, UPC, Barcelona, Catalonia, 1998.
[15] J. Martí, J. Batlle, A. Casals, Model-based objects recognition in industrial environments for autonomous vehicles control, in: Proceedings of the IEEE International Conference on Robotics and Automation, Albuquerque, NM, April 1997, pp. 1632–1637.
[16] B.A. Draper, R.T. Collins, J. Brolio, A.R. Hanson, E. Riseman, The schema system, International Journal of Computer Vision 2 (1989) 209–250.
[17] B.A. Draper, A.R. Hanson, E.M. Riseman, Knowledge-directed vision: control, learning, and integration, Proceedings of the IEEE 84 (11) (1996) 1625–1637.
[18] H. Buxton, Visual interpretation and understanding, Technical Report CSRP 452, School of Cognitive and Computing Sciences, 1997.
[19] M. Campani, M. Capello, G. Piccioli, E. Reggi et al., Visual routines for outdoor navigation, in: Proceedings of the Intelligent Vehicles Symposium, Tokyo, Japan, July 1993, pp. 107–112.
[20] P. Parodi, G. Piccioli, A feature-based recognition scheme for traffic scenes, in: Proceedings of the Intelligent Vehicles Symposium, Detroit, USA, September 1995, pp. 229–234.
[21] M. Hild, Y. Shirai, Interpretation of natural scenes using multi-parameter default models and qualitative constraints, in: Proceedings of the Fourth International Conference on Computer Vision, Berlin, Germany, May 1993, pp. 497–501.
[22] W. Efenberger, V. Graefe, Distance-invariant object recognition in natural scenes, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Osaka, Japan, November 1996, pp. 1433–1439.
[23] U. Regensburger, V. Graefe, Visual recognition of obstacles on roads, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Munich, Germany, September 1994, pp. 980–987.
[24] V. Graefe, Dynamic vision systems for autonomous mobile robots, in: Proceedings of the IEEE/RSJ International Conference on Robots and Systems, Tsukuba, Japan, September 1989, pp. 12–23.
[25] A. Ide, M. Tateda, H. Naruse, A. Nobiki, T. Yabuta, Automatic recognition and stereo correspondence of target objects in natural scenes, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Raleigh, NC, July 1992, pp. 1597–1602.
[26] B.A. Draper, J. Brolio, R.T. Collins, A.R. Hanson, E.M. Riseman, Image interpretation by distributed cooperative processes, in: Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition, Ann Arbor, MI, June 1988, pp. 129–135.
[27] A.R. Hanson, E.M. Riseman, in: Visions: a computer system for interpreting scenes, A.R. Hanson, E.M. Riseman (Eds.), Computer Vision Systems, Academic Press, New York, 1978, pp. 303–333.
[28] T.M. Strat, Natural Object Recognition, Springer, Berlin, 1992.
[29] M. Asada, Y. Shirai, Building a world model for a mobile robot using dynamic semantic constraints, in: Proceedings of the 11th International Joint Conference on Artificial Intelligence, vol. II, Detroit, MI, August 1989, pp. 1629–1634.
[30] S. Hirata, Y. Shirai, M. Asada, Scene interpretation using 3-d information extracted from monocular color images, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Raleigh, NC, July 1992, pp. 1603–1610.
[31] Y. Taniguchi, Y. Shirai, M. Asada, Scene interpretation by fusing intermediate results of multiple visual sensory information processing, in: Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligence, Las Vegas, NV, October 1994, pp. 699–706.
[32] Y. Ohta, Knowledge-based Interpretation of Outdoor Natural Color Scenes, Pitman, London, 1985.
[33] P. Gamba, R. Lodola, A. Mecocci, Scene interpretation by fusion of segment and region information, Image and Vision Computing 15 (1997) 499–509.
[34] A. Mecocci, R. Lodola, U. Salvatore, Outdoor scenes interpretation suitable for blind people navigation, in: Fifth International Conference on Image Processing and its Applications, Edinburgh, UK, 1995, pp. 256–260.
[35] R. Bajcsy, A.K. Joshi, A partially ordered world model and natural outdoor scenes, in: A. Hanson, E. Riseman (Eds.), Computer Vision Systems, Academic Press, New York, 1978, pp. 263–270.
[36] M.D. Levine, A knowledge-based computer vision system, in: A.R. Hanson, E.M. Riseman (Eds.), Computer Vision Systems, Academic Press, New York, 1978, pp. 335–352.
[37] M.D. Levine, S.I. Shaheen, A modular computer vision system for picture segmentation and interpretation, IEEE Transactions on Pattern Analysis and Machine Intelligence 3 (5) (1981) 540–556.
[38] A.M. Nazif, M.D. Levine, Low level image segmentation: an expert system, IEEE Transactions on Pattern Analysis and Machine Intelligence 6 (5) (1984) 555–577.
[39] R.J. Douglass, Interpreting three-dimensional scenes: A model building approach, Computer Graphics and Image Processing 17 (1981) 91–113.
[40] I.Y. Kim, H.S. Yang, Efficient image labeling based on Markov Random Field and error backpropagation network, Pattern Recognition 26 (2) (1993) 1695–1707.
[41] I.Y. Kim, H.S. Yang, An integrated approach for scene understanding based on Markov Random Field model, Pattern Recognition 28 (12) (1995) 1887–1897.
[42] K.S. Kumar, U.B. Desai, Joint segmentation and image interpretation, Proceedings of Third IEEE International Conference on Image Processing 1 (1996) 853–856.
[43] N.W. Campbell, W. Mackeown, B.T. Thomas, T. Troscianko, Interpretation image databases by region classification, Pattern Recognition 30 (4) (1997) 555–563.
[44] B. Bhanu, S. Lee, J. Ming, Adaptative image segmentation using a genetic algorithm, IEEE Transactions on Systems, Man and Cybernetics 25 (12) (1995) 1543–1567.
[45] J. Peng, B. Bhanu, Closed-loop object recognition using reinforcement learning, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (2) (1998) 139–154.
[46] S.D. Buluswar, B.A. Draper, Color machine vision for autonomous vehicles, International Journal of Engineering Applications of Artificial Intelligence 2 (1998) 245–256.
[47] Y. Ohta, T. Kanade, T. Sakai, Color information for region segmentation, Computer Graphics and Image Processing 13 (1980) 222–241.

[48] M. Celenk, A color clustering technique for image segmentation, Computer Vision, Graphics and Image Processing 52 (1990) 145–170.
[49] R. Bajcsy, S.W. Lee, A. Leonardis, Color image segmentation with detection of highlights and local illumination induced by inter-reflections, in: Proceedings of the International Conference on Pattern Recognition, 1990, pp. 785–790.
[50] R.C. Luo, H. Potlapalli, D.W. Hislop, Natural scene segmentation using fractal based autocorrelation, in: Proceedings of the 1992 International Conference on Industrial Electronics, Control, Instrumentation and Automation, San Diego, CA, November 1992, pp. 700–705.
[51] A.C. She, T.S. Huang, Segmentation of road scenes using color and fractal-based texture classification, in: Proceedings of the IEEE International Conference on Image Processing, vol. III, Austin, Texas, November 1994, pp. 1026–1030.
[52] S.D. Buluswar, B.A. Draper, Non-parametric classification of pixels under varying outdoor illumination, in: Proceedings of the ARPA Image Understanding Workshop, Monterey, CA, November 1994, pp. 1619–1626.
[53] H. Mori, K. Kobayashi, N. Ohtuki, S. Kotani, Color impression factor: an image understanding method for outdoor mobile robots, in: Proceedings of the IEEE/RSJ Internationl Conference of Intelligent Robots and Systems, Grenoble, France, 1997, pp. 380–387.
[54] J. Amat, State of the art and future trends of computer vision (in Catalan), in: Proceedings of the First Workshop on Automation, Robotics and Perception, Barcelona, Catalonia, February 1996, pp. 15–25.
[55] B.A. Draper, Learning control strategies for object recognition, in: K. Ikeuchi, M. Veloso (Eds.), Symbolic Visual Learning, Oxford University Press, New York, 1997, pp. 49–76.
[56] J.H. Piater, E.M. Riseman, P.E. Utgoff, Interactively Training Pixel Classifiers, Proceedings of FLAIRS, AAAI Press, New York, 1998.
[57] M. Szummer, R.W. Picard, Indoor-outdoor image classification, in: Proceedings of the IEEE Internationl Workshop on Content-Based Acces of Image and Video Database, Bombay, India, January 1998, pp. 42–51.
[58] S. Baker, CORBA Distributed Objects. Uning Orbix, ACM Press/Addison-Wesley, New York/Reading, MA, 1997.
[59] D. Cooper, Image understanding environment. Technical report, EPSRC Summer School on Computer Vision, Manchester University, UK, 1 June 1998.
[60] J. Dolan et al., Solving diverse image understanding problems using the image understanding environment, in: Proceedings of the Image Understanding Workshop (IUW), 1996, pp. 1481–1504.
[61] J. Mundy et al. The image understanding environment program, in: Proceedings of the DARPA Image Understanding Workshop, San Diego, CA, 1992, pp. 185–214.