Supervised Object Knowledge Learning for Image Understanding

Anna Bosch, Xavier Muñoz, Jordi Freixenet, and Joan Martí

Computer Vision and Robotics Group.University of Girona
Campus de Montilivi 17071.Girona,Spain
{aboschr,xmunoz,jordif,joanm}@eia.udg.es

Abstract: An object learning system for image understanding is proposed in this paper. The knowledge acquisition system is designed as a supervised learning task. Therefore, the role of the user as teacher of the system is emphasized, which allows to obtain the object description as well as to select the best recognition strategy for each specific object. An object description is acquired by considering different model representations provided from several representative examples in training images. Moreover, different recognition strategies are built and applied to obtain initial results. Next, teacher evaluates these results and the system automatically selects the specific strategy which best recognise each object. We provide a feedback with the user until he/she is satisfied with the knowledge achieved by the system. Experimental results are shown and discussed.

Key words: image understanding, supervised learning, object recognition, srategy selection, image processing.

1. INTRODUCTION

The goal of image understanding systems is the identification of objects in visual imagery and to establish three-dimensional relationships between the objects themselves, or the objects and the viewer. Although humans perform this perception in an immediate and effortless manner, it is not a trivial task to adequately describe a scene significantly. It requires the integration of different image processing techniques, pattern recognition algorithms, and artificial intelligence tools, in order to transform images into visual descriptions of the world [4].

An image understanding system can also be considered as a knowledge-based vision system, because such system requires models that represent prototype objects and scenes. Hence, two important issues must be taken into account: (1) the way in which the model knowledge is organised and stored, and (2) how this knowledge is acquired. However, while knowledge representation has become a permanent focus of interest and a large number of proposals can be found in the literature (see [1], [4]), knowledge acquisition tools are still in their infancy [1]. A small number of systems are oriented to facilitate the entry of knowledge or carry out some form of automated learning. In contrast, most of the existing systems had to incorporate this new model knowledge by hand, and code-encapsulated data. Examples are the Schema system [2] and the region analyser of Ohta et al. [8], which are successful systems and works of reference.

Nevertheless, nowadays most vision researchers agree that the success of scene description systems lies on their ability to learn from experience and training.

Automated learning must consider the acquisition of object models, as a description of the object attributes, as well as the selection of the strategy used to find and recognise the object in an image. In fact, not all objects are defined in terms of the same attributes, and even these attributes may be used in various ways within the matching or interpretation process. Therefore, the learning system must take a flexible and multifaceted recognition strategy into account. A large number of object recognition strategies have been proposed to achieve a particular goal. However, there is not a perfect strategy for all objects and very little research in the field of computer vision has gone into the problem of determining the best recognition strategies [3].

In this paper we propose an object learning system for image understanding, which addresses the problem of automatic recognition strategy selection. Inspired on relevance feedback techniques used on image retrieval systems, the knowledge acquisition system is designed as a supervised learning task, which involves the user as teacher and part of the learning process. Therefore, this learning allows us to obtain the object description as well as to select the best recognition strategy for each specific object.

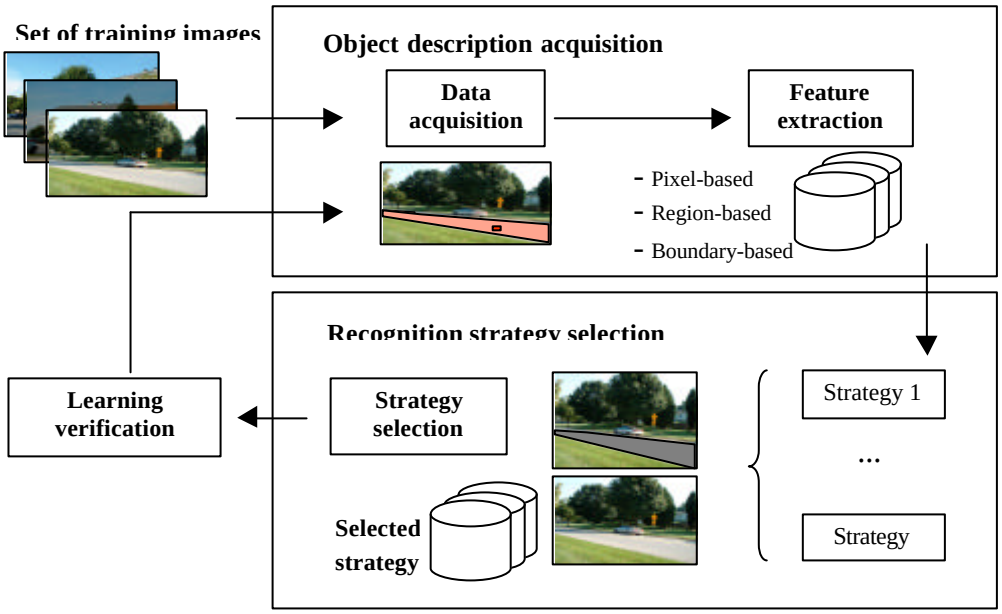

Figure 1 - Scheme of the proposed knowledge acquisition process, working as a supervised learning task.

The remainder of this paper is structured as follows: Section 2 describes the proposed supervised object knowledge learning, focusing on the object description, the recognition strategy design and the strategy selection. Experimental results proving the validity of our proposal appear in Section 3. Finally, conclusions are given in Section 4.

2. LEARNING PROPOSAL

Our object learning approach has been designed as a supervised task, which emphasises the role of the user as the responsible of teaching the system. Firstly, the teacher selects some representative examples of objects in training images. From these

examples, a description of each object is acquired by considering different model representations. Next, several strategies to recognise the object are built and applied in order to obtain initial recognition results. These results are evaluated by the teacher, and the system automatically selects the specific strategy which best recognise each object. A global scheme of the proposed knowledge acquisition process is shown in Figure 1.

2.1. Object Description Acquisition

The teacher first selects meaningful examples in the training images by clicking inside the object of interest. This simple selection allows to extract the whole object and to compute and register different model representations, which provide a complete description of the object. Specifically, the acquired information is composed by:

Pixel-based description: from the selected (clicked) point, a set of neighbouring pixels are extracted and considered as samples of the object pixels. Next, a large number of colour and texture features of these pixels are measured. We initially consider the whole set of features as candidates to characterise real objects. In particular, 28 colour features related to different colour spaces, and a set of 8 co-occurrence matrix based texture features are computed for every pixel.

Region-based description: a colour texture active region, which integrates region and boundary information [7], grows from the selected point in order to segment the region corresponding to the whole object. Colour and texture descriptors, as well as shape information based on Fourier descriptors [9], are extracted in order to describe and characterise the object of interest.

Boundary-based description: based on the active region segmentation, we build and store the geometric shape of objects. Boundary of the segmented region provides an initial approximation to the object contour. Moreover, due to the difficulty to obtain accurate boundaries, the teacher has the possibility to interfere in this process refining (or correcting) the object boundary by hand. As result, object edges are obtained as shape model.

2.2 Object Recognition Strategies

The object descriptions can be used in different ways in order to recognise the object. We initially included into our system three different object recognition strategies, which differ on: (1)the description model they use, and (2) the philosophy which the strategy relies on. Thus, recognition strategy A is a top-down strategy based on pixel description; recognition strategy B follows a pure hybrid strategy based on an initial segmentation; and recognition strategy C is an hybrid strategy able to refine initial segmentation results. Figure 2 shows the recognition of the *road* object by these strategies in order to illustrate the basic flow and main stages of them. More in depth,

Recognition Strategy A: The top-down strategy consists on the direct search of a specific object by exploiting information concerning the objects characteristics. A decision tree of hyperplane nodes that recursively divide the feature space into target samples (pixels belonging to the object) and non-target samples (pixels belonging to the remaining objects) is designed. Moreover, a feature selection process allows choosing the features that best characterise each object. The obtained tree will be later used on

recognition process to directly look for pixels belonging to the object. More details of this strategy can be found in our previous proposal [6].

Recognition Strategy B: A pure hybrid strategy for image understanding starts with a bottom-up stage of general purpose segmentation to part the image into meaningful regions. Next, a top-down strategy is performed over these results to specifically find objects of interest. We take as a basis of this implementation a previously developed unsupervised colour texture segmentation technique based on active regions, which integrates region and boundary information [7]. As in the previous strategy, the subset of region features which best characterise each object are chosen by a feature selection process. Selected features are then considered to look for the segmented regions on the image, which match with the object model.

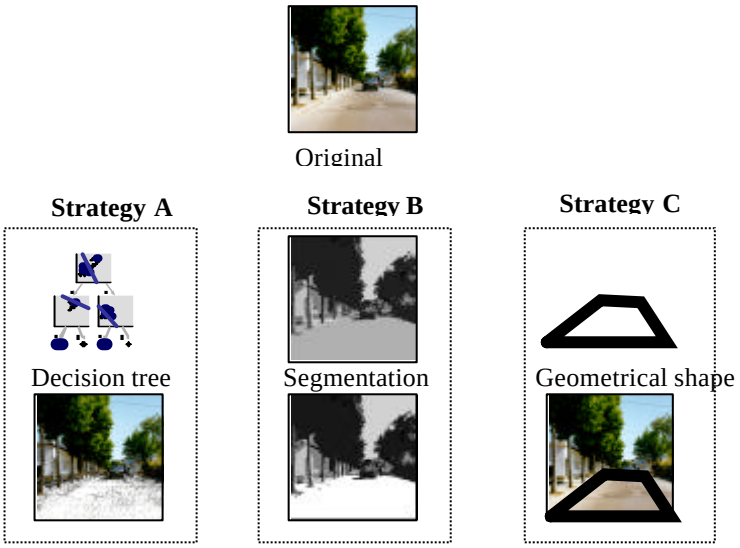

Figure 2 - Recognition strategies. Strategy A: Top-down strategy, pixel recognition using a decision tree. Strategy B: Hybrid Strategy, recognition using color, texture and shape of the regions. Strategy C: Hybrid Strategy based on the boundary of regions

Recognition Strategy C: Strategy C is also a hybrid strategy, which relies on a previous segmentation stage. However, initial segmented results are refined to obtain on accurate object recognition based on the known shape model and the use of snakes. Segmented regions with similar shape to the object of interest are considered to estimate the initial position of the snake. Next, the deformable template optimises an energy function by iteratively altering the shape of the contour so that the best match between the deformed template and the regions/edges in the image is obtained [5].

2.3. Recognition Strategy Selection

Once the process of recognition strategies design is complete, the best specific strategy to recognise each object must be determined. This is the key stage of our proposal; inspired on relevance feedback techniques used on content-based image retrieval systems, the role of the user is emphasised and he/she is involved as a vital part of the learning process. With the help of the teacher feedback, the system is able to evaluate the different recognition strategies and to learn which is the best strategy for each object.

Therefore, given a set of training images, recognition methods are launched together to find all the instances of the given object. Obviously, these strategies will provide different results. In some cases, recognition methods will coincide, while can differ in other ones. Basically, because a strategy misses an object apparition or, contrarily, it gives a false positive. These recognition results are visually retrieved to the teacher in order to evaluate their quality. In front of these results, the teacher will mark the found instances to indicate if they are well recognised or not. In other words, if the strategy (or strategies) which obtained this recognition was right or wrong. We provide the teacher with three levels of correctness: highly correct, correct and wrong. Although the use of more levels could probably provide more information, we consider it would be lesser friendly for the user to interact with the system. When the teacher has evaluated the results, this information allows the system to measure the score of each strategy and finally select the strategy which best recognise the object.

The learning process ends with a final verification step. A visual feedback is provided by means of recognising the object in the training images. Obviously, selected strategy will be used for each object. The visual feedback guides the teacher, giving the option to interfere in the learning process by introducing new training images.

3. EXPERIMENTAL RESULTS

The experimental results were obtained over a large set of images extracted from the image database of the University of Massachusetts (*http://vis-www.cs.umass.edu/vislib/Outdoor*), and the image database of Ohio State University (*http://sampl.eng.ohio-state.edu/~sampl/data/stills/color/index.htm*). These images include a large variety of natural objects such as *sky*, *ground*, *grass* and *leaves*, as well as artificial objects such as *house*, *car* and *traffic signal.* In general, the learning of all these objects was achieved with a high degree of success. We observed that colour and textural information were usually selected for describing natural objects, while some of the man-made objects were only described by its shape information. In what follows we report on some of the obtained results in order to illustrate the learning system performance.

Figure 3 - Results obtained on the learning of the objects sky, road, grass, leaves, car and trafic signal.

In Figure 3 we show the results obtained on the learning of five different objects (*objects sky, road, grass, leaves, car* and *trafic signal*). Results were obtained after training the system with different images for each object under different outdoor conditions such as time of the day, the season or the weather. For the *sky*, the *road* and the *grass*, the strategy B based on the region information was selected. Basically, the learning process chooses colour features in order to characterise these three objects. In contrast, the *leaves* object was described by the strategy A, which is based on pixel information. As well as colour, texture features were used to describe this object. Finally, we show two examples for which the strategy C is selected to perform the recognition. In these cases, the boundary information of the *car* and the *traffic signal*

was used to describe the objects. The system was trained using 15 images and the training process tooks approximately 3 hours, most of this time performing the feature extraction and feature selection.

4. CONCLUSIONS

An object knowledge acquisition method for image understanding was described. The process has been designed as a supervised learning task, which emphasises the role of the user as system teacher. From some examples provided by the teacher, the system extracts the information required to describe the object. Moreover, the best recognition strategy for each specific object is automatically selected.

Future extensions of this work are oriented to the improvement of the strategy selection in order to make possible the selection of a combination of different strategies, and not only one of them, as the best method to recognise an object. Furthermore, new recognition strategies will be included into the system.

3. REFERENCES

1. D. Crevier and R. Lepage. (1997) Knowledge-based image under-standing systems: A survey. *Computer Vision and Image Understanding* 67(2):160–185.

2. B. Draper, R. Collins, J. Brolio, A. Hanson, and E. Riseman. (1989). The schema system. *International Journal of Computer Visio*n, 2:209–250.

3. B. Draper, A. Hanson, and E. Riseman. (1996) Knowledge-directed vision: Control, learning, and integration. *Proceedings of the IEE*E, 84(11):1625–1637.

4. R. Haralick and R. Shapiro. (1993) *Computer and Robot Visio*n, vol-ume II. Adison-Wesley, Reading, Massachussets.

5. A. K. Jain, Y. Zhong, and S. Lakshmanan. (1996) Object matching using deformable templates. *IEEE Transactions on Pattern Analysis and Machine Intelligenc*e, 18(3):267–278.

6. J. Martí, J. Freixenet, J. Batlle, and A. Casals. (2001) A new approach to outdoor scene description based on learning and top-down segmentation. *Image and Vision Computin*g, 19:1041–1055.

7. X. Muñoz, J. Freixenet, X. Cufí, and J. Martí. (2003) Active regions for colour texture segmentation integrating region and boundary information. In *IEEE International Conference on Image Processin*g, Barcelona, Spain.

8. Y. Ohta. [1985] *Knowledge-based Interpretation of Outdoor Natural Color Scene*s. Pitman Publishing, Boston, Massachussets.

9. D. Zhang and G. Lu. (2002) *Generic fourier descriptor for shape-based image retrieval.* Multimedia and Expo. IEEE International Conference on , vol 1:425 - 428