Object Recognition based on Visual Grammars and Bayesian Networks

Elías Ruiz, L. Enrique Sucar
National Institute of Astrophisics, Optics and Electronics
Computer Science Department
{elias_ruiz, esucar}@inaoep.mx

Abstract

A novel proposal for object recognition based on relational grammars and Bayesian Networks is presented. Based on a Symbol-Relation grammar an object is represented as a hierarchy of features and spatial relations. This representation is transformed to a Bayesian network structure which parameters are learned from examples. Thus, recognition is based on probabilistic inference in the Bayesian network representation. Preliminary results in modeling natural objects are presented.

I Introduction

Most current object recognition systems are centered in recognizing certain type of objects, and do not consider their structure. This implies several limitations: (i) the systems are difficult to generalize to any type of object, (ii) they are not robust to noise and occlusions, (iii) the model is difficult to interpret. This paper proposes a model that achieves a hierarchical representation of a visual object in order to perform object recognition tasks, based on a visual grammar [Ferrucci *et al.*, 1996] and Bayesian Networks (BN's). Thus, we propose the incorporation of a visual grammar in order to develop an understandable hierarchical model so that from basic elements (obtained by any image segmentation algorithm) it will construct more complex forms by certain rules of composition defined in the grammar, in order to achieve object recognition in certain context (e.g. images of objects indoors). The importance of using a hierarchical approach is that it can build a more robust model to noise and occlusions, also the BN model can work with incomplete evidence. In addition, the model expresses the grammar in an understandable way to a human, in order to interpret the model and even modify the structure.

There are several works using a hierarchical approach [Chang *et al.*, 2011; Felzenszwalb, 2011; Melendez *et al.*, 2010; Zhu and Mumford, 2006]. We propose Symbol-Relation grammars (SR-Grammars) because they can represent relationships between elements using predicate logic; the transformation into a BN can deal with uncertainty and we can do inference in order to perform object recognition tasks.

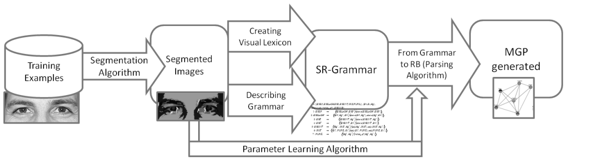


Figure 1: Training of the model. Starting from training images and a description of the object in terms of the lexicon and the visual grammar, the model generates a BN structure with its parameters.

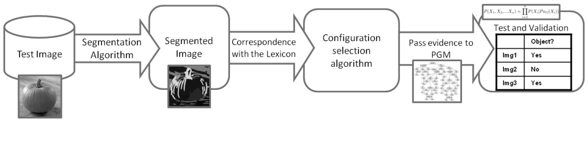

Test Image
Segmentation Algorithm
Segmented Image
Correspondence with the Lexicon
Configuration selection algorithm
Pass evidence to PGM
$P(X_1, X_2, \ldots X_n) = \prod_{i=1}^{n} P(X_i | Pa_G(X_i))$
Test and Validation
Object?
Img1 Yes
Img2 No
Img3 Yes

Figure 2: Testing stage. The test image is segmented with the visual dictionary and correspondences between regions and the visual lexicon are obtained. After that, the algorithm evaluates subsets of those regions with their spatial relationships that are candidates to be evaluated in the previously trained PGM in order to do inference. At the end, we obtain a result given by the PGM if there is an object in the image.

II Methods

The proposed method compromises two phases: (i) model construction and transformation to a BN (Fig. 1); and (ii) image pre-processing and object recognition using probabilistic inference (Fig. 2). Next we briefly describe the main steps of our method.

A Segmentation and Lexicon

The segmentation is performed with simple RGB quantization (32 levels) and edge extraction using Gabor Filters. Small regions are fused with other regions. The idea is to use a simple segmentation algorithm. These regions define a visual dictionary. Every region is described using several shape and color features. Similar regions by their features are considered as candidates to terminal elements in our grammar. All the terminal elements are described in a "Lexicon".

B SR-Grammars and Spatial Relationships

The object representation is based on SR-grammars [Ferrucci *et al.*, 1996] which can incorporate spatial relationships between elements. In our work, these relationships determine

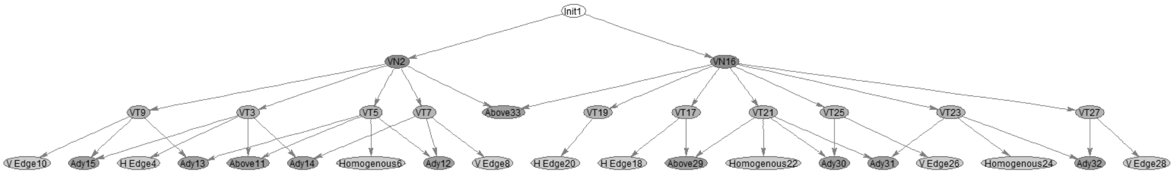


Figure 3: BN generated by the pumpkin. Evidence is passed only on the leaf nodes. Nodes with two parents represent relationship nodes. Leaf nodes with only one parent represent terminal elements.

the position of one object element with respect to another object element, at different levels in the hierarchy. Although there are different types of spatial relationships, in our model we use topological and order relationships, such as $inside\_of$ or $above$.

C Transformation of the grammar

We transform the grammar into a BN: for every production rule $Y^0 \rightarrow \langle \mathbf{M}, \mathbf{R} \rangle$, we produce the node Y^0 in the network and connect this node with all $x \in \mathbf{M}$. For every relationship $r(a, b) \in \mathbf{R}$ we produce the node r connected with its parents $a, b \in \mathbf{M}$.

D Parameter learning and object recognition

Once the BN is obtained, its parameters are learned from examples using EM, as the intermediate elements (nodes) in the BN are hidden nodes. Then, for recognition, the low level features (color segments and edges) are detected in the image in order to instantiate the low level nodes in the BN; this information is propagated to determine the probability of observing certain object.

III Results

We have applied this method for face detection [Ruiz *et al.*, 2011]. Here we present a simple example of how an object is represented using a grammar, and its transformation to a BN.

We describe the object pumpkin, by the following grammar:

GPUMPKIN=(VN,VT,VR,S,P,$\emptyset$);

VN={PUMPKIN,Stem,Fruit}; VT={Bh,Bv,Hg1,Hg2,Hg3}; VR={above,ady}; S=PUMPKIN; P:

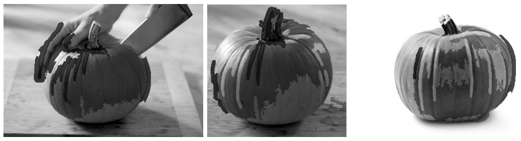
1: PUMPKIN^0 → <{Stem^2, Fruit^2}, {above(Stem, Fruit)}>

2: Stem^0 → <{Bh^2,Hg1^2,Bv^2,Bv^3}, {above(Bh^2,Hg1^2), ady(Bv^2,Hg1^2), ady(Hg1^2,Bv^3), ady(Bv^2,Bh^2), ady(Bh^2,Bv^3)}>

3: Fruit^0 → <{Bh^3,Bh^4,Hg2^2,Hg3^2,Bv^4,Bv^5}, {above(Bh^3,Hg2^2), ady(Bv^4,Hg2^2), ady(Hg2^2,Hg3^2), above(Hg3^2,Bh^4), ady(Hg3^2,Bv^5)}>

where the Lexicon is given by: i) Hg as homogeneous region, ii) Bh as horizontal edge and iii) Bv as vertical edge. Numbers associated to these elements represent minimal variations in its features (color). The BN generated from this grammar is depicted in the Fig. 3.

In this example the probabilistic inference detects the regions which probably represent a pumpkin. Results of detected objects are illustrated in Fig. 4. The configurations presented have maximum likelihood given by inference in the BN.[1]



(a) pumpkin1 (b) pumpkin2 (c) pumpkin3

Figure 4: Regions detected for a pumpkin example. In spite of the simple segmentation algorithm, the grammar helps to detect parts of the object.

IV Conclusion

A first stage in the design of a visual grammar to detect objects was described. This approach combines Symbol Relation grammars and Bayesian networks to describe an object in an image. This model was tested for natural objects (like pumpkins) with low level features as terminal elements with promising results. The next stages in our research are to learn the SR grammars from examples and apply this formalism to other classes of objects, such as those in service robotics scenarios.

References

[Chang *et al.*, 2011] L. Chang, Y. Jin, W. Zhang, E. Borenstein, and S. Geman. Context, computation, and optimal roc performance in hierarchical models. *IJCV*, 93(2):117–140, 2011.

[Felzenszwalb, 2011] P. F. Felzenszwalb. Object detection grammars. In *ICCV Workshops*, page 691. IEEE, 2011.

[Ferrucci *et al.*, 1996] F. Ferrucci, G. Pacini, G. Satta, M. I. Sessa, G. Tortora, M. Tucci, and G. Vitiello. Symbol-relation grammars: a formalism for graphical languages. *Inf. Comput.*, 131(1):1–46, 1996.

[Melendez *et al.*, 2010] A. Melendez, L. E. Sucar, and E. Morales. A visual grammar for face detection. In Angel Kuri-Morales and Guillermo Simari, editors, *AAI - IBERAMIA 2010*, volume 6433 of *LNCS*, pages 493–502. Springer, 2010.

[Ruiz *et al.*, 2011] E. Ruiz, A. Meléndez, and L. E. Sucar. Towards a general vision system based on symbol-relation grammars and bayesian networks. In Jürgen Schmidhuber, Kristinn R. Thórisson, and Moshe Looks, editors, *AGI*, volume 6830 of *LNCS*, pages 291–296. Springer, 2011.

[Zhu and Mumford, 2006] S. C. Zhu and D. Mumford. A stochastic grammar of images. *Foundations and Trends in Computer Graphics and Vision*, 2(4):259–362, 2006.

[1] Additional information of our model can be found on the website: http://srmodel.erzh.org