Extracting a Domain Theory from Natural Language to Construct a Knowledge Base for Visual Recognition

Lawrence Chachere and Thierry Pun
Computer Science Center, University of Geneva,
24 rue General Dufour, CH-1211 Geneva 4, Switzerland
e-mail: chachere@cui.unige.ch

Introduction

As computer vision research shifts from smaller applications to more general object recognition problems which involve very large sets of object classes, there is an increasing amount of interest in feature indexing schemes. Feature indexing raises the important issue of how to construct a knowledge base of associations between key discriminative features and their respective object classes. It is clear that some amount of automation is necessary in order to build such knowledge bases for very large recognition systems.

Visual recognition is a very difficult domain for machine learning. Concept formation has two components: 1. finding appropriate attributes which give a well-behaved class membership function, and 2. given attribute vectors, finding the class membership function. Although most learning has been concerned with the latter problem, the first is at least as important for many domains, and probably more difficult [Rendell, 1990]. While the idea of using inductive learning to generate object class descriptions for a computer vision system is not new, e.g., [Shepherd 1983; Lehrer et. al, 1987; Cromwell and Kak 1991], attribute selection has tended to be either ad hoc or application oriented, since there is no universally established set of appropriate visual attributes.

An additional difficulty when considering vision problems is that many of the existing similarity-based inductive learning systems are not directly applicable if one takes into account the large numbers of training examples they typically rely on. There exists biological evidence that it is possible to formulate good prototypes with very few training examples, e.g., [Cerella, 1979]. A topic that is currently receiving a lot of attention in machine learning is the integration of domain theory within existing empirical techniques [Lebowitz, 1990]. Domain knowledge reduces the burden imposed on statistics for ensuring correctness of output.

This paper describes a proposed approach for learning characteristic features of objects which is being developed for use by the Geneva Vision System [Pun 1990]. The following section introduces the use of natural language as a heuristic knowledge source of visual information [Chachere, 1992]. The next two sections divide the concept formation problem in terms of defining appropriate attributes (extraction of natural language defined features) and inductive training. The final section shows some results that have thus far been obtained on an application domain of leaf images.

Natural Language as a Knowledge Acquisition Heuristic

Our system considers a feature significant if it is encoded as an adjective in natural language. Thus an attribute is deemed appropriate if it can take the value of one or more natural language features. There are several advantages of this feature selection criterion. First, natural language is an extremely rich and reliable source of perceptual information. Secondly, concepts defined by natural language are easy to understand by humans, e.g., it is much easier for a person to determine the correctness of output from the hue attribute than an ad hoc attribute such as an 'excess red' ratio, which is commonly derived from the red/green/blue representation of color images. Understandability of parameters is helpful for most complex AI problems which are difficult to theoretically verify. A third advantage is that verbal feature values can serve as heuristics during the induction process. For example, *yellow* can be considered a significant numerical interval within the hue attribute.

Perhaps the main inconvenience with a natural language criterion is that low level vision methods for deriving many of these features are not readily available. Vision methods have generally been customized for easily derivable numeric features, e.g., compactness, elongatedness, and various other aspect ratios, without a deep investigation of the utility of the various features for higher level object classification tasks. Natural language features tend to be symbolic; thus, the inevitable necessity of translating signals to abstract higher level representations is immediately put to question.

Extraction of Natural Language Features

A library of symbolic features was constructed through a systematic search for adjectives, by means

of standard references (computerized Webster's English dictionary and Roget's thesauruses). A vocabulary of 102 common adjectives describing shape and color has been obtained, for which feature extraction techniques are being developed. The shape adjectives are either global (e.g., *triangular, oval, cordate*), contour-descriptive (e.g., *smooth, dentate, jagged*), or localized (e.g., *pointed, rounded,* or *notched*). Some shape features can be further characterized by more specific adjectives, e.g., *dull* or *sharp*-pointed. Color adjectives consist of a basic set and subcolors of members of the basic set. The basic set is defined by a recent study of 98 different languages [Berlin and Kay, 1991] which suggested a linguistic universality and evolution towards 11 basic color terms *(white, black, red, green, yellow, blue, brown, gray, orange, purple,* and *pink)*. Texture and spatial adjectives are also important, but are more difficult and will be considered in future work.

The image signal to symbol transition is a knowledge engineering task which varies from feature to feature. While a few of these features have absolute definitions, the majority do not. The approach we use for extracting shape features is based on contour analysis. Contour chains are created using a Sobel edge detector. Global shape descriptors are then derived by template matching. To determine whether an object is triangular, an ideal geometric triangle is superimposed onto the region enclosed by the contour chain outlining the object. If the ratio of overlap between the two areas is extremely high, triangular is assigned as the global shape. A threshold for ratio of overlap will always be somewhat subjective, but is systematically determined by measurements taken on sample triangular-shaped images. Regarding localized shape descriptions, it is difficult to determine the significant locations on an object where local measurements should be obtained without the use of heuristics. One possibility is to use simple verbal heuristics. For example, our implementation will routinely examine the *base* and *top* of an object, or to the *tip* of a protrusion of an object.

For the purpose of color characterization, a color name knowledge base was created by processing images of labelled color samples and computing their respective hue, lightness, and saturation values from the RGB image signals. The examples included a prototype sample of each basic color, borderline samples (e.g., borderline reds: *orange red* and *violet red*), and proper sub-colors (e.g., red shades: *maroon, scarlet,* etc.). It is important to consider that lightness greatly influences the definitions of basic colors which are generally thought of as hues, both linguistically [Berlin and Kay, 1991] and biologically (the Bezold-Brucke Effect) [Wyszecki and Stiles, 1982]. For example, the concept of *yellow* implies a high lightness value; a 'yellow' hue with a low lightness value will be perceived either as brown or green. Since it is probably not possible to define perfect regular functions, in particular within the hue-lightness plane for chromatic colors in basic set (Figure 1), assignment of a basic color to an unlabeled region is accomplished using a nearest neighbor scheme in the 3-dimensional HLS space.

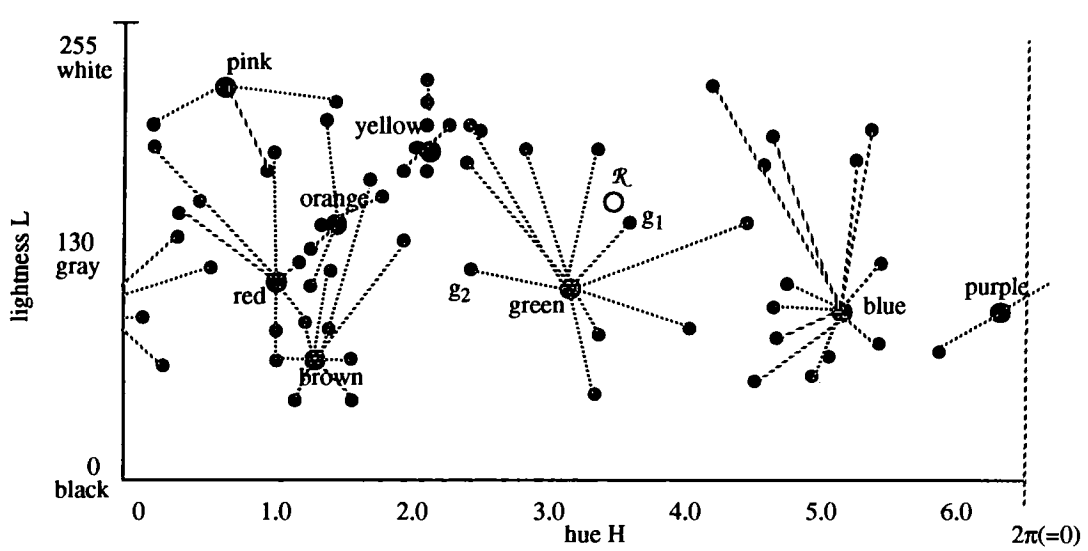

Figure 1: distribution of color samples on hue-lightness plane;g_1, g_2 are two sub-colors of green, respectevely Paris green and olive green) g_2 is an extreme member of the green class.

Given HLS coordinates of a fairly homogeneously colored region of an image, a color description is attached as follows. A nearest neighbor within the knowledge base of colors is determined using a weighted Euclidean distance measurement. The region $\mathcal{R}$ is assigned the basic color of its nearest neighbor (Fig. 1, $\mathcal{R} \rightarrow g_1 \rightarrow$ green), unless the nearest neighbor is an extreme data point (e.g. g_2) within its basic color class and the distance of the region's HLS to the class prototype is larger than the distance between that nearest neighbor and the class prototype. If such a case arises, the region is characterized as 'undefined color', indicating a gap in the knowledge base. Since color terms are defined subjectively rather than functionally, it seems inappropriate to have the system statistically guess color labels in unchartered areas of the HLS space.

In addition to the assignment of a basic color, a region may also be modified with more detailed features.If the HLS of a region is extremely close to the HLS coordinates of its nearest neighbor, it will also acquire the subcolor label of the neighbor. Various adjectives which modify the basic color description (e.g., *pale, dark, light, greenish*) are attached if applicable. For example, *dark* is appended to a blue region which has a relatively low lightness value within the blue class of blue colors. Standard definitions of subcolors help to define thresholds for these modifiers. For example, *navy* and *indigo* are dark shades of blue by definition, and thus are the lightness values corresponding to these samples within the knowledge base.

Training

Training in our system is accomplished using images of labelled objects. The label will include a basic class name, and optional hierarchical links to broader classes. It is assumed that all examples are labelled with basic level categories [Rosch, 1978]. It is possible that two random class members of a very broad category are likely to have more visual differences than similarities. Nevertheless, if higher level taxonomic categories are supplied, the system tries to formulate concepts in the same manner as for the basic categories. This use of taxonomic knowledge can greatly help to economize the learning procedure, since it allows evidence to be shared between sibling classes via inheritance from a parent category.

Both numeric and symbolic data are supplied from the feature extraction routines discussed in the previous section. Symbolic values always correspond to a natural language adjective. Numeric information cannot be ignored, since it is possible that certain numeric ranges within an attribute may not map to any natural language symbol. When a new example is supplied for a given class, all data is gathered from previous examples. Generalized features are created by applying basic generalization operators are to each attribute; interval closing for numeric values, and tree climbing and disjunction for symbols [Michalski, 1983].

The next step after generalization is to evalutate each generalized feature. This is accomplished by simple measurements of intraclass similarity and interclass differences:

Similarity =

$$\sum_{\text{symbols}} \frac{\text{fraction of class samples equal to symbol}}{\text{rank of symbol in class}}$$

Differences = (c1 * number of complete classes discriminated by feature) + (c2 / overall frequency of feature)

It is emphasized that if the set of attributes defined in the previous section is adequate, the statistical task need not be complicated. The similarity function decreases significantly with each symbol added to a disjunction, but does allow for exceptions. For example, an observation of 50 diamonds, 47 white, 1 yellow, 1 pink, and 1 blue, would yield ==> .94/1 + .02/2 + .02/3 + .02/4 = .962, indicating that white is generally characteristic of diamonds despite the variety of different exceptions. Interclass discrimination is not accomplished by an exhaustive proof that a given combination of features will uniquely distinguish a concept, but rather a by measuring the discriminative power and uniqueness of individual features. c1 and c2 in the above equation are weighted coefficientss, respectively corresponding to discrimination and uniqueness.

Additional evaluation factors include inheritance (a feature is not characteristic of a basic class if it is characteristic of a parent class) and word commonality biases derived from word corpuses (higher weight is given to features which are more commonly used in natural language). For simplicity of discussion, the important topic of spatial relationships, such as above(Part-A,Part-B), has been avoided. However, we are currently investigating natural language heuristics to help reduce some of the complexity of this problem. Regular or common patterns can be represented as unary predicates, rather than binary or higher order predicates, in terms of adjectives such as *palmate* or *4-legged*.

A set of characteristic features for an object class is finally produced by ranking all the evaluated features. The highest ranking features are chosen as characteristics, while features with low evaluations are discarded. Since the purpose of the output is for indexing

rather than a complete recognition, it is only important that the selected features are good keys for recognition, rather than an exhaustive concept description of the object class. It is important to note that the independence of these output characteristic features makes them suitable to a variety of recognition procedures (e.g., decision trees, parallel discrimination procedures).

An Example

For the purpose of testing the learning system, a database was created consisting of 275 images of non-compound leaves from European broadleaf trees. The data set consists of 55 differenct classes, 5 examples of each class. The size and distribution of this data set was established for two purposes: 1. To demonstrate the ability to formulate good concept descriptions from small sets of training examples; and 2. To demonstrate the ability to deal with discrimination problems involving large numbers of different classes. Each leaf image in this database is attached with four labels, a species name (the basic category), a genus name (e.g., maples, oaks), a family name (e.g., beech family, which includes oaks, beeches, chestnuts) and the inclusive label "leaf".

Table 1 displays initial results by based upon two of the feature functions, for the purpose of demonstrate discrimination of three selected leaf classes from the database (Figure 2). The apex of each leaf was located by finding the axis of symmetry and assuming that the leaves were oriented with the apex at the top of the image.The apex angle feature is computed by measuring the angle of the two edges extending from the apex point. Leaves have two significant parts, a blade and a petiole and therefore a color was determined for both subregions.

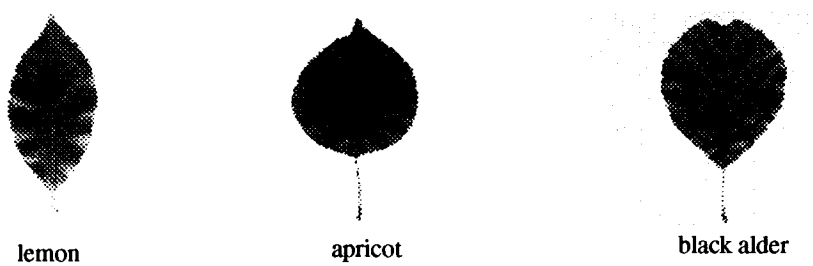

Figure 2: example images selected from 3 of the 55 leaf classes.

Leaf	Apex angle	Apex shape description	HLS values of blade and petiole	Color descriptions
lemon 1	1.2383	pointed	(2.34,126,63), (2.13,206,51)	green / yellow
lemon 2	1.25	pointed	(2.31,119,63), (2.11,199,50)	green/ yellow
lemon 3	1.0264	pointed	(2.38,155,69), (2.12,174,51)	yellowish green / yellow
apricot 1	1.5611	pointed	(1.75,88,45), (1.12,175,51)	dk brwnish green/dark pink
apricot 2	1.4745	pointed	(1.88,92,46), (0.98,100,59)	dark brownish green / red
apricot 3	1.419	pointed	(1.72,86,46), (0.99,123,58)	dark brownish green / red
black alder 1	3.8281	notched	(2.32,1.05,60), (2.25,165,54)	green / yellowish green
black alder 2	3.4735	notched	(2.09,97,55), (2.25,166,50)	green / yellowish green
black alder 3	3.5457	notched	(2.32,110,59), (2.32,163,48)	green / yellowish green

Table 1: extracted features. Apex angle ranges from 0 to 2π, values are "pointed" if $< \pi$, "notched" if $> \pi$, "straight" if very close to π. Hue H ranges 0 to 2π where pure red=$\pi/3$, pure green=π, pure blue=$5\pi/3$. Lightness L ranges from 0 to 255 where 0=black and 255=white. Saturation S is a percentage.

Based on the data in Table 1, the training procedure makes the following characteristic features: apricot → dark brownish green blade, pointed apex, red or dark pink petiole; black alder → notched apex, yellowish green petiole; lemon → yellow petiole, pointed apex; leaf → green blade. Since the three examples are from different families, no descriptions are formulated at that level. With more examples of different leaf classes, petiole color=yellowish green will become a default leaf characteristics rather than a black alder characteristic. The system would also eventually learn that red/dark pink is a very rare petiole color and thus it should emerge with a higher ranking within the set of apricot leaf characteristics.

The results for these examples match the classification descriptions are close to what one would find in a typical field guide to trees. The blade color of the apricot leaves is an exception.The HLS coordinates of those examples actually map into an undefined region within the color knowledge base slightly closer to a green data point than the next nearest neighbor, a brown point. Furthermore it has been determined that the numerical values are not quite correct. The discrepancy was traced to a loss of accuracy in the image acquisition process, using a PerfectScan^{TM} Mikrotek 600Z scanner. Higher quality scanners and cameras for image acquisition are likely to produce better results.

Conclusion

The ultimate goal in this research is to formulate good visual concepts about object classes in an efficient manner, without relying upon large numbers of training examples. There are still many natural language features which remain to be developed. The task of extracting such features may seem arduous. However, we believe that there is no shortcut for difficult AI problems; perhaps the key element is to have lots of knowledge [Lenat and Guha, 1989]. The uncertain rate of progress towards a general visual recognition system indicates that such an effort may well prove worthwhile.

Acknowledgments

This research is supported by a Swiss National Research Fund grant, 20-26475.89. The authors would like to express their deep gratitude to Catherine de Garrini for her major role in the implementation of feature extraction of shape adjectives.

References

Berlin, B. and Kay, P. (1991). Basic Color Terms: Their Universality and Evolution. University of California Press, Berkeley, CA.

Cerella, J. (1979). Visual Classes and Natural Categories in the Pigeon. Journal of Experimental Psychology: Human Perception and Performance, 5.

Chachere, L.F. (1992). Reasearch Summary. In workshop notes, Constraining Learning with Prior Knowledge, AAAI-92, 78-80.

Cromwell, R.L. and Kak, A.C. (1991). Automatic generation of object class descriptions using symbolic learning techniques In Proc. of the 9th National Conf. on Artificial Intelligence, 710-717.

Lebowitz, M. (1990). The Utility of Similarity-Based Learning in a World Needing Explanation. In Michalski, R.S., and Kodratoff, Y. (eds.), Machine Learning: An Artificial Intelligence Approach, volume 3, chapter 15, Morgan Kaufmann.

Lehrer, N.B., Reynolds, G. and Griffith, J. (1987). Initial Hypothesis Formation in Image Understanding Using an Automatically Generated Knowledge Base. DARPA, Image Understanding Workshop, 520-537.

Lenat, D.B. and Guha, R.V. (1989). Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project. Addison-Wesley.

Michalski, R.S. (1983). A Theory and Methodology of Inductive Learning. In Michalski, R.S., Carbonell, J.G., and Mitchell, T.M. (eds.), Machine Learning: An Artificial Intelligence Approach, chapter 4, Morgan Kaufmann.

Pun, T. (1990). The Geneva Vision System: Modules, integration and primal access. Technical Report no. 90.06, Computing Science Center, University of Geneva.

Rendell, L. (1990). Commentary to Lebowitz, M., The Utility of Similarity-Based Learning in a World Needing Explanation. In Michalski, R.S. and Kodratoff, Y. (eds.), op. cit.

Rosch, E. (1978). Principles of Categorization. In Rosch, E. and Lloyd, B.B. (eds.) Cognition and Categorization, 28-39, Lawrence Erlbaum Associates.

Shepherd, B.A. (1983). An appraisal of a decision tree approach to image classification. In Proceedings of IJCAI, 473-475.

Wyszecki, G. and Stiles, W. (1982). Color Science: Concepts and Methods, Quantitative Data and Formulae. Wiley.