In the proceedings of the Intn'l conference of the North American Fuzzy Information Processing Society, NAFIPS 1999, New York, pp 710-714

Transitioning from Recognition to Understanding in Vision using Additive Cartesian Granule Feature Models

James G. Shanahan[+], James F. Baldwin[#], Barry T. Thomas[#],
Trevor P. Martin[#], Neill W. Campbell[#], Majid Mirmehdi[#]

[+] Xerox Research Centre Europe (XRCE)
Grenoble Centre, 6 chemin de Maupertuis
38240 Meylan, FRANCE
Email: Jimi.Shanahan@XRCE.Xerox.com

[#] Advanced Computing Research Centre
University of Bristol, Bristol,
BS8 1TR, ENGLAND

Abstract

Here we propose an approach to object recognition that facilitates the transition from recognition to understanding. The proposed approach begins by segmenting the images into regions using standard image processing approaches, which are subsequently classified using a discovered fuzzy Cartesian granule feature classifier. Understanding is made possible through the transparent and succinct nature of the discovered models. The recognition of roads in images is taken as an illustrative problem in the vision domain. The discovered fuzzy models while providing high levels of accuracy (97%), also provide understanding of the problem domain through the transparency of the learnt models. The learning step in the proposed approach is compared with other techniques such as decision trees, naïve Bayes and neural networks.

1. Introduction

Current learning approaches that have been successful within computer vision such as [1, 2] have mainly focussed on low-level image processing and object recognition, while tending to ignore high-level processing such as understanding. Previous work which has addressed this issue have tended to operate in abstract worlds [3] or on relatively simple problem domains [4]. Here we propose an approach to object recognition that facilitates the transition from recognition to understanding while also providing high levels of accuracy in rather difficult problem domains. The proposed approach begins by segmenting the images into regions using standard image processing approaches, which are subsequently classified using a discovered fuzzy Cartesian granule feature classifier (Section 2). Understanding is made possible through the transparent and succinct nature of the discovered classifiers. The approach is illustrated on a road classification problem (Section 3) and the classifier component of the approach is compared with other standard approaches in Section 4. In Section 5 we discuss the various classifier approaches from understandability, effectiveness and efficiency perspectives.

2. Overview of the approach

We address the problem of recognising object regions within the context of digital images of outdoor scenes. The problem is partitioned into two natural parts: region segmentation and region classification. Segmentation is achieved using standard image processing approaches (see Section 3 for details), whereas region classification is carried out by a classifier. Each segmented region is described using a variety of features such as colour, location, texture, shape etc.. The main goal of this research is to construct a classifier, automatically from examples, that provides high performance accuracy while, facilitating user understanding. In order to meet both criteria we propose to represent the classifier as an additive Cartesian granule feature models i.e. if-then-rules with weighted antecedents whose values are fuzzy sets defined over Cartesian granule features. Cartesian granule features [5, 6] are a new type of multidimensional feature defined over the Cartesian product of words drawn from the linguistic partitions of the constituent feature universes. Cartesian granules (characterised by fuzzy sets) provide an abstraction of the multidimensional universe by carving it into regions that are drawn together as result of indistinguishability, similarity, proximity or functionality. Figure 1 gives an illustrative example of how to extract a Cartesian granule fuzzy set corresponding to car positions in images from example car positions where the top left table corresponds to examples of car positions, corresponding linguistic descriptions and least prejudiced distributions (LPDs or probability distributions). Mass assignment theory provides a formal mapping between linguistic descriptions (fuzzy sets) and probability distributions [7].

The top middle graph corresponds to the initial Cartesian granule frequency distribution, where the granule characterisations (i.e. the fuzzy sets) are also shown. The top right graph depicts the Cartesian granule frequency distribution after updating with the LPD corresponding to the value of *40*. The right middle graph shows the Cartesian granule frequency distribution after updating with the LPD corresponding to the value of *60*. The right bottom graph displays the Cartesian granule frequency distribution after counting all the LPDs corresponding to the example car positions. Finally the left bottom graph depicts the corresponding Cartesian granule fuzzy set for car positions in images i.e. a linguistic summary of car positions in images in terms of the words *left, middle* and *right.*

[6] has shown that systems can be quite naturally described in terms of Cartesian granule features incorporated into rule-based models. Here a region-based fuzzy classifier is constructed automatically from example data (region feature values) using the G_DACG constructive induction algorithm [6, 8]. The G_DACG algorithm discovers good Cartesian granule features (i.e. the feature subsets and the feature universe abstractions). G_DACG is a population-based search algorithm (based on genetic programming), where each node in the search space is a Cartesian granule feature. G_DACG iteratively hones in on good Cartesian granule features based on the evolutionary operations of crossover, mutation and reproduction. Good Cartesian granule features are subsequently incorporated into rule based models (see for example Figure 2). Inference is carried out using evidential reasoning and semantic unification/match of class fuzzy sets and data fuzzy set [7]. Classification corresponds to taking the class corresponding to the maximum of the inferred results.

3. Vision Dataset

The Bristol Image Database [9, 10] consists of over 350 colour images of a wide range of urban and rural scenes. Figure 4 (lower left quadrant) depicts a typical urban scene in this database. Eighty images of typical outdoor rural scenes were selected from the Bristol image database. Subsequently these images (characterised by intensity) were segmented into *road* and *non-road* regions using the *k*-means segmentation algorithm, where *k* was set to 4. Previous results have shown the *k*-means algorithm to be effective [9] (see also the upper left quadrant of Figure 4). This resulted in 13,628 regions being generated. Feature values were subsequently generated for each region feature. Non-overlapping training, validation and test sets of regions were subsequently generated in a class-wise manner as follows: 70% of data allocated to training, 15% to validation and 15% to testing. Table 1 gives a sample-count breakdown for each class. For the road classification problem each segmented image region was described using a set of over sixty features, comprising of colour, location, orientation, size, shape and texture features. In order to reduce the complexity of the learning process a neural network-based "filter" feature selection algorithm was applied to this feature set [11]. This resulted in ten features been selected as representative features for task of road classification.

Table 2 describes the features that were selected for the subsequent induction step. The first three features correspond to the average luminance and colour differences in a region. The location of the region is expressed as the X and Y co-ordinates of the region centroid. Orientation is expressed as the sine and cosine of the angle of the principal axis. The next feature corresponds to the principle mode of the PCA (principle component analysis [12]) transformed region boundary description. The last two features arise from the use of a psychophysically plausible model of texture, based upon Gabor filters. In this case the features correspond to two high frequency (128 and 256) isotropic Gabor filters. A full description of all features is presented in [11].

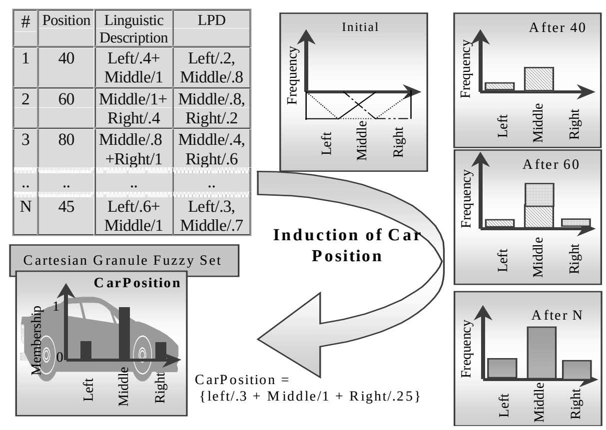


Figure 1: Induction of the Cartesian granule fuzzy set for to car positions in images.

4. Results

For the purposes of this work we reduced the number of object classes that can make up an image from twelve (considered in [6, 13]) to two generic classes: *road* and *not-road*. The G_DACG constructive induction algorithm was used to discover classification systems that classified image regions into *road* and *not-road*, in terms of additive Cartesian granule feature models, from example region data. The proposed approach was then evaluated against various other induction algorithms such as decision trees, neural networks and naïve Bayes, using a variety of performance criteria such as accuracy, understandability and efficiency.

Table 1: Object classifications for each region and corresponding sample counts.

Class No.	Class	# Train examples	# Validation examples	# Test examples
1	NotRoad	8381	1796	1797
2	Road	1157	248	249
TOTAL	13628	9538	2044	2046

Table 2: Selected features for each region that are considered for learning.

10 Selected FEATURES No.	Features
0	Luminance
1	Red-Green
2	Yellow-Blue
3	Centroid (X, Y)
5	Orientation 1
6	Orientation 2
7	Shape 1 (principle mode)
8	Texture G_{128} – high frequency, isotropic
9	Texture G_{256}– high frequency, isotropic

We applied the G_DACG constructive induction algorithm to the road classification problem. The reduced feature set of ten base features were considered and Cartesian granule features of dimensionality up to three with granularity ranges of [2, 12] were considered (while parsimony was promoted) thus yielding a search space of over 500,000 nodes. The G_DACG algorithm iterated for fifty generations and at the end of each generation five of the best Cartesian granule features were selected from the current population. The number of individuals in a generation was set to 120. The discovered features were then used to form additive Cartesian granule feature rule-based models. Backward elimination was also employed, eliminating extraneous lowly contributing features. The models were evaluated using the test dataset. Table 3 tabulates the results of some the more interesting additive Cartesian granule feature models that were discovered using G_DACG. In the case of the models presented in Table 3, the models were trained using equal numbers of examples for the *Road* and *Not-Road* classes. By equalising the example count across classes a minor improvement (of less than 1%) in test case accuracy was achieved over learning from the original skewed training set. The results presented correspond to models where the weights have been estimated using semantic discrimination analysis and also where the weights have been tuned using Powell-based optimisation [11].

For example when an additive model consisting of three one-dimensional Cartesian granule features, was formed respectively over the features *Luminance*, *Y-B* (Colour difference) and *Y-Position*, a classification accuracy of 95.5% (after tuning the weights) on unseen image regions was achieved. The feature universes in this model were linguistically partitioned using five words, which are characterised by uniformly placed trapezoidal fuzzy sets with 50% overlap. The corresponding additive rule base is presented in Figure 2. The linguistic descriptions, characterised by a Cartesian granule fuzzy sets, corresponding to the *luminance* for *Road* and *Not-Road* classes is presented in close-up detail in Figure 3. Notice that in the additive rule model in Figure 2 that the *Luminance* feature receives a lower weight than the other features involved in the decision-making process. This is due mainly because the *Luminance* linguistic summaries do not provide as a good a separation of concepts as the other features. Figure 4 presents a Java applet screendump that illustrates the results of applying this ACGF model to a k-means segmented image. The results are qualitatively very good from a classification perspective, however the low-level k-means and region growing segmentation process has under segmented parts of the image, thus leading to some areas of the image being misclassified.

Table 3: ACGF models discovered using the G_DACG algorithm

Dimension	Train	Valid %	Test%	Optimised Weights	Cartesian Granule Features
1D	92	95	95.5	No	((0 5)) ((2 5)) ((4 5))
2D	94	93.3	96.6	No	((0 5)(2 5)) ((2 5) (4 5))
1D	92	95	95.5	Yes	((0 5)) ((2 5)) ((4 5))
2D	93.9	93.5	96.7	Yes	((0 5)(2 5)) ((2 5) (4 5))

An additive Cartesian granule model composed of two two-dimensional features give a marginal improvement over the one-dimensional model (see Table 3 for details). The test confusion matrix for this model is presented in Table 4. The cells in the diagonal of the table correspond to the correctly classified regions for each class; for example 210 of the 249 (84.3%) *Road* regions were correctly labelled.

4.1 A Comparison with other Learners

The results obtained when additive Cartesian granule feature modelling was applied to the region classification problem were compared with those achieved using other standard induction approaches such as neural nets, naïve Bayes, and various decision tree approaches. For comparison purposes, all features in the reduced base feature set were made available to all approaches. The datasets used were the same as those used in the additive Cartesian granule feature modelling case. Table 5 summarises the results of various modelling approaches that were used on the road classification problem.

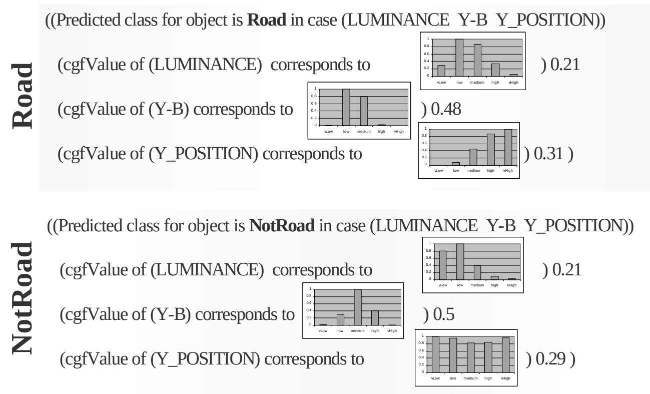
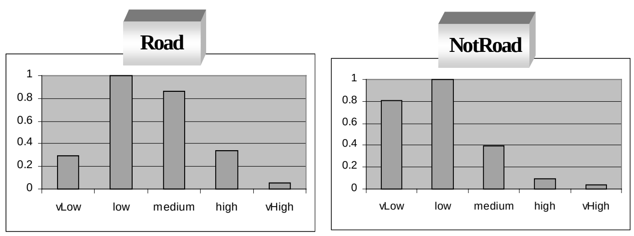


Figure 2: Additive Cartesian granule feature model for road classification.



{vLow:0.29, low:1, medium:0.86, high:0.33, vHigh:0.056}

{vLow:0.8, low:1, medium:0.39, high:0.09, vHigh:0.036}

Figure 3: Linguistic summary, in the form of Cartesian granule fuzzy sets, of *luminance* for Road and NotRoad Classes.

Table 4: Confusion Matrix generated by the discovered 2D model (optimised) in Table 3.

Actual\ Predicted	NotRoad	Road	Total	Class %Accuracy
NotRoad	1767	30	1797	98.3
Road	39	210	249	84.3

5. Discussion

One of the primary concerns of intelligent systems is that they should be able to interact naturally with their environment. One of the integral parts of many domains is the human, and consequently the intelligent system (agent) needs to interact with the human. This can be achieved by a variety of means and at many different levels such as a graphic display of trend data. However, one of the most natural forms of communication (and sometimes most effective) is through words. The proposed approach has generated a road classification system that enlightens the user about what a road is, in terms of *luminance* and other feature value descriptions. These descriptions are in terms of words such as *low* and *very low* – generic words in this case, but these could be assigned from a user-defined dictionary and supplemented with hedges such as *very*, *not so much*, etc. and with connectives such as conjunctions and disjunctions. Furthermore the weights associated with each feature inform the user of how important a particular feature is in the inference process.

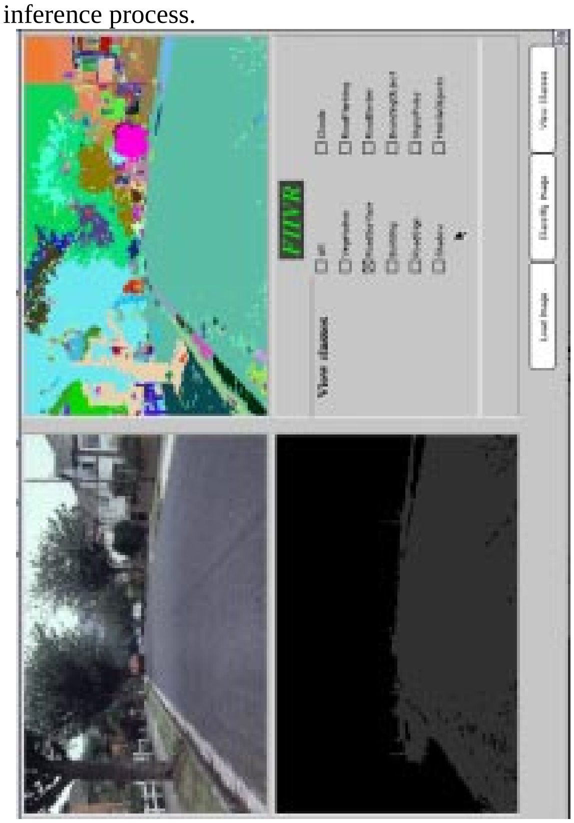


Figure 4: Screendump of a Java applet that displays the original image (top left quadrant), k-means segmented image (top right quadrant) and the results of region classification using a rule-based ACGF model (bottom left quadrant). The regions classified as road are highlighted in grey and the non-road regions are display in black.

The induced Cartesian granule feature model facilitates a transition from a low-level object recognition task to a high level understanding task which should greatly simply human computer interaction. This simplification comes from the expression of the knowledge in a form that is almost directly interpretable by the human user. The proposed approach, while facilitating machine learning, may also facilitate human learning and understanding through the generated anthropomorphic models.

With regard to the other approaches examined here, such as the ID3 and C4.5 algorithms, the induced models while being readable tend to be large and consequently makes understanding very difficult. In the case of neural networks and oblique decision trees [14], the induced knowledge is encoded in vectors of weights (and biases) which may prove difficult for a user to interpret and understand.

Table 5: Comparison of results obtained using a variety of machine learning techniques on the road classification problem.

Approach.	# of Features used	Accuracy(%)
Additive Cartesian granule feature Model	3	96.7
Naïve Bayes	10	96.2
Oblique Decision Trees	10	94.7
Neural Net	10	97
C4.5	10	92.75
ID3	10	92.5

A further consequence of readability and understandability is that it will generally increase user's confidence in the system and it can also enhance reliability. For example, the user may enhance the systems reliability by identifying a data deficiency or a variable deficiency.

The results of all modelling approaches seem to indicate a high regularity and good separablilty in the classes of the road dataset. All approaches do well but the induced additive Cartesian granule feature model outperforms most other approaches examined with respect to novel example classification.

From a classification task perspective, the induced models have similar computational requirements. From a learning perspective the induction of Cartesian granule feature models is more intensive. This arises mainly from not just identifying a model that provides high performance accuracy (the goal of most other induction algorithms), but also from identifying a model that glassbox in nature. This issue of identifying glassbox models, while having extra computational requirements, is compensated by the identification of models that facilitate understandability.

6. Conclusions

A new approach to object recognition, based upon a Cartesian granule feature classifier, has been proposed that facilitates the transition from recognition to understanding. The results for the proposed approach of additive Cartesian granule feature modelling indicate that such a system can classify image regions with a very high accuracy, using a very transparent model, which takes a little longer than other approaches to discover. The extra discovery time needed is mainly due to the search for a transparent model. Envisioned applications include content based image retrieval systems (CBIR). CBIR is an area which relies heavily on human-computer interaction, where interaction requires understanding, and thus would greatly benefit from the glassbox approach proposed here.

Acknowledgements

This work was supported by DERA Grant 92W69. James Shanahan carried out this work while at the University of Bristol.

Bibliography

1. Campbell, N.W., *et al.*, *Interpreting Image Databases by Region Classification.* Pattern Recognition, 1997. **30**(4): p. 555-563.
2. Pentland, A., R. Picard, and S. Sclaroff. *Photobook: Tools for Content-based manipulation of image databases*. in *SPIE Conf. on Storage and retrieval of image and video databases 2*. 1994.
3. Winston, P.H., ed. *The Psychology of Computer Vision*. . 1975, McGraw-Hill: USA.
4. Michalski, R.S., *et al.*, *Learning patterns in images*, in *Machine Learning and Data Mining*, R.S. Michalski, I. Bratko, and M. Kubat, Editors. 1998, Wiley: New York. p. 241-268.
5. Baldwin, J.F., T.P. Martin, and J.G. Shanahan. *Modelling with words using Cartesian granule features*. in *FUZZ-IEEE*. 1997. Barcelona, Spain: pp 1295-1300.
6. Shanahan, J.G., *Cartesian Granule Features: Knowledge Discovery of Additive Models for Classification and Prediction*, . 1998, PhD Thesis, Dept. of Engineering Maths, University of Bristol, Bristol, UK.
7. Baldwin, J.F., T.P. Martin, and B.W. Pilsworth, *FRIL - Fuzzy and Evidential Reasoning in A.I.* 1995: Research Studies Press(Wiley Inc.), ISBN 086380159 5.
8. Baldwin, J.F., T.P. Martin, and J.G. Shanahan, *System Identification of Fuzzy Cartesian Granule Feature Models using Genetic Programming*, in *IJCAI Workshop on Fuzzy Logic in Artificial Intelligence, Lecture notes in Artificial Intelligence (LNAI 1566) - Fuzzy Logic in Artificial Intelligence*, A.L. Ralescu and J.G. Shanahan, Editors. 1998, Springer. p. 26.
9. Campbell, N.W., B.T. Thomas, and T. Troscianko, *Automatic segmentation and classification of outdoor images using neural networks.* International Journal of Neural Systems, 1997. **8**(1): p. 137-144.
10. Mackeown, W.P.J., *et al. Contextual Image Labelling with a Neural Network*. in *IEE Vision, Speech and Signal Processing*. 1994.
11. Shanahan, J.G., *et al.*, *Transitioning from recognition to understanding in vision using additive Cartesian granule feature discovery models.* Journal paper in preparation, 1999: p. N/A.
12. Jolliffe, I.T., *Principal Component Analysis*. 1986, New York: Springer.
13. Baldwin, J.F., T.P. Martin, and J.G. Shanahan. *Automatic fuzzy Cartesian granule feature discovery using genetic programming in image understanding*. in *FUZZ-IEEE*. 1998. Anchorage, pp 960-965, USA: UK.
14. Murphy, S.K., S. Kasif, and S. Salzburg, *A system for induction of oblique decision trees.* Journal of Artificial Intelligence Research, 1994. **2**: p. 1-33.