ON CLASSIFYING TIME-VARYING EVENTS

John K. Tsotsos
Computer Systems Research Group
and
Division of Cardiology, Toronto General Hospital
University of Toronto
Toronto, Ontario, Canada

Abstract

This paper describes a methodology for the classification of time-varying phenomena and discusses several experiments that demonstrate aspects of its performance. The key idea is that the scheme is a knowledge-based "expert" system, using a knowledge base (KB) of time-varying concepts to drive the control structure. This control structure is based on feedback concepts [1] but expands on them so that the semantic organizational components of the knowledge base are intimately tied to the control scheme. The basic mechanism has been described in [2], and will only be briefly described below. We will concentrate on the characteristics of the scheme, its successes and failures, and examples of its operation in this paper.

Introduction

Recently, there has been a substantial amount of research into solutions for the correspondence problem for time-varying imagery [3], [4], [5]. This problem may be stated as follows: given an object A in one image, and a successive image, determine which of the objects in the successive image corresponds to object A so that the resultant motion characteristics are compatible with the motion history and capabilities of object A. In each of these efforts, the researchers identified the crucial open problem in the computer analysis of time-varying imagery is that of motion recognition or classification. We wish to propose a methodology for tackling this problem.

Our system, called ALVEN, is a knowledge-based expert system for human left ventricular performance assessment. Each knowledge unit in the KB (knowledge base) defines a particular motion concept ("upwards", "contract", "move", "extend", etc.) using a frame-based representational formalism designed to accomodate the semantic components necessary for motion description [6]. The frames are organized using the IS-A (class/sub-class) and PART-OF (part/sub-part) primitives with additional relationships between frames provided by temporal and similarity (characterisitics of mutual exclusiveness between frames) connections. The similarity connections are of prime importance here because they define the time-course of differences between two motion concepts, and therefore provide a mechanism for choosing alternate hypotheses depending on the matching failures of a given hypothesis. Left ventricular motion concepts are defined in terms of simpler general motion concepts, each one being represented as a slot in the motion frame, with constraints between slots providing additional spatial and temporal relationships. These motion definitions were initially derived from text books and from discussions with the cardiologists at our unit (Toronto General Hospital). Experimentation is providing more refined versions of these definitions.

The paradigms of competition and cooperation among hypotheses and hypothesize-and-test form the basis of our recognition control structure. The key feature of the control structure is that it is driven by the organization of the knowledge base, that is, by the primitive relations between knowledge units. Hypotheses are ranked on the basis of certainty factors. Each hypothesis, when activated, receives an initial certainty factor equal to that of the hypothesis that activates it. A modified relaxation process is then used to update the certainty factors. The relaxation process is based on "conceptual adjacency" that specifies which hypotheses are competitors and which ones are complementary and in what respect. The compatibilities between hypotheses that are necessary for the RLP (relaxation labelling process) are derived in a dynamic fashion, depending on which conceptual adjacencies are present between hypotheses that are active during the course of the input image sequence. The best hypotheses (highest ranked) are used to derive the expectations for the next image.

A scheme similar to, but not as sophisticated as [3], is used for the vision aspects of processing, that is, the use of expectations to guide the search for objects in the image and determine their correspondencies to the previous image. Changes are described in terms of location changes of points, length changes of axes and perimeters, and area changes These primitive kineses provide an intermediate representation for relating quantitative changes to qualitative ones.

This low level data drives the activation of hypotheses that attempt to describe the exhibited motions. The kineses for a particular object are matched against the hypothesized motions, and matching failures between expected and actual kineses are recorded. These failures are represented in terms of exception frames which contain any information necessary so that proper selection can be made of alternate hypotheses. This selection is made via the similarity links which are present in the hypotheses. The activation of more specialized hypotheses proceeds along the IS-A axes in the KB, while motions that are aggregates of simpler motions are activated via the PART-OF links.

An Example

The basic control structure has been implemented and the early stages of testing and performance analysis have proven quite successful. Currently, films taken at 30 or 60 images per second of tantalum markers, implanted into the myocardium of the left ventricle during open heart surgery, are being analyzed. The contraction and relaxation patterns of a typical left ventricle are shown in figures 1 and 2 respectively. In this patient, the anterior (right) wall is hypokinetic (exhibits less than normal motion, i.e., less extent of motion, but no exhibited anomalies in the direction of motion) and the remainder of the left ventricle is normal, according to the radiologists reports.

The knowledge base of motion concepts currently includes descriptive terms such as "inwards", "outwards", "contract", "expand", "extend", "lengthen", "shorten", "uniform contraction", "uniform expansion", "asynchrony", "dyskinesis", and many lower level terms such as "leftwards", "rightwards", "upwards", "downwards", "translate", "approach", "recede", "no motion", "area change", "no area change", etc. as well as other left ventricular specific concepts. The example motions in Figures 1 and 2 are interesting for the following reason. The standard definition of a uniform contraction (or relaxation) motion pattern used by cardiologists at our unit, and therefore the one currently represented in our KB, is that all markers in the left ventricular wall move inwards (or outwards for relaxation), and normal contraction (relaxation) implies a uniform contraction (relaxation) with each marker exhibiting approximately the same extent of motion, during the whole phase. So one would expect, if ALVEN functions properly, that the frame embodying the motion definition for uniformitity would be instantiated, while the one for normal motion would not be. On analysis of this film by ALVEN, it was revealed, and later verified by manual analysis, that no "uniform" motion of the type described above was present. In fact, the left ventricle exhibited some paradoxical motion (wrong direction with respect to the left ventricular phase) in addition to hypokinesis. (The actual output is too voluminous to include here: for each of the 11 markers and for the left ventricle as a whole, there are between 10 and 30 instances of motion descriptive terms.) This is an example of where the qualitative definition used by the cardiologists who simply view the film in order to analyze it, was demonstrated to be inadequate at a quantitative level. Clearly, since cardiologists are quite capable of diagnosing heart disease properly, there is a more abstract level of processing that they do for these films, but which they perhaps have difficulty in articulating. This has lead us to the inclusion of aggregations of marker motions, so that several markers are grouped into segments. At the segment level, the uniformity "seen" by the radiologist was also recognized by ALVEN, as well as the hypokinesis, and several instances of asynchrony (non-uniform onset of motion). Therefore, for this example, ALVEN's analysis was shown to be consistent, yet more complete than, the radiologist's report. We are working on several more such films in an attempt to further refine the motion definitions in our KB.

Characteristics of the Updating Scheme

An important component of KB expert systems is the "hypothesis scoring function" that is present in virtually all such systems. This function is usually tailored for the application, is based at least in part on Bayesian criteria and for the most part, is very difficult to analyze or to extend to other applications. We believe that an important next step in the construction of expert systems is the inclusion of a scoring scheme whose performance can be analyzed in "engineering" terms. By this we mean that a set of standards must be put forth so that one scoring scheme may be compared to another in a meaningful way. Only in this way, can the system's limitations, successes, and possible extensions be really understood. We will propose a set of attributes of updating schemes for time-varying event analysis and then discuss the performance of our scheme with respect to those attributes.

As described in [2], the updating scheme is based on relaxation labelling [7]. Each iteration of the RLP represents a single image pair of the image sequence, and compatibilities are dynamically determined depending on the semantic relationships present between their corresponding frames in the KB. These semantic relationships include temporal constraints, global constraints (IS-A), local constraints (PART-OF) and similarity (mutual exclusion).

With respect to this updating function, we have conducted a series of experiments in order to study its characteristics. The intent was to see what relationships are present among the following attributes of the updating process with respect to a particular domain:

a) the number of active, competing hypotheses;
b) the certainty thresholds for hypothesis instantiation and deletion;
c) the scene sampling rate;
d) the semantic relationships between hypotheses;
e) the image sequence noise level ("temporal" noise); and,
f) the duration of the shortest event defined in the KB, and thus, the shortest event that should be recognized by the system.

The experiments were carried out according to the following criteria:

a) The trials are run using a sequence of random numbers to represent matching successes and failures, and are adjusted for noise levels. For example, if there is a 10% noise level, then any random number below 0.1 is considered to be a matching failure for the correct hypothesis, while the same range is used for matching successes for the incorrect ones. The range of random numbers is 0.0 to 1.0.

b) Each data point is the average time value required for instantiation of the correct hypothesis obtained over thirty trials.

c) The decision threshold, that is, the certainty value that must be achieved before a particular hypothesis is instantiated is related to the number of hypotheses in the following manner. On examination of the relationship between certainty and time for two competing hypotheses, the resultant curves strongly resemble exponentials [2]. For this reason, the notion of a "time constant" from electronics was borrowed for our purposes and defined as the time at which the certainty of a hypothesis reaches

$$(1-\frac{1}{N})*(1-e^{-1})+\frac{1}{N}$$

which simplifies to:

$$0.632+\frac{0.368}{N}$$

where N is the number of competing hypotheses for a single object's motion. We shall see also, below, how changing this threshold affects the updating scheme's performance. The deletion threshold is defined by:

$$\frac{e^{-1}}{N}$$

We will present results from three such experiments. The first experiment, which also appeared in [2] and is included here for completeness and comparison purposes with the remaining ones, examined the relationship among the number of competing hypotheses, the number of inter-image descriptions needed before a decision can be made, and temporal noise.

Temporal noise is the analog of static noise in time. Static noise is the term used to denote the degree of uncertainty in the precise intensity value of a particular point in an image. If the image has a large amount of static noise, then there is a high uncertainty in the characteristics of a particular picture element, and therefore, there is a high uncertainty in the data produced by the image analysis component in a computer vision system in terms of relationships between intensity values (for example, gradients). Temporal noise describes the degree of uncertainty of a particular event occurring at a specific time instant, i.e., the characteristics of the time instant in terms of which events are present or absent. In our formulation, a large of amount of temporal noise will mean that there is a correspondingly large amount of uncertainty in the nature of the changes that the image analysis component "observes" at a particular time instant, i.e., the presence or absence of change or the amount of change.

In this first experiment, the only semantic relationships that play a part are "similarity" and "PART-OF", and only one hypothesis of the competing set is true. As can be seen in Figure 3, a family of curves is produced that satisfies our intuitive notions of what should be occurring. These notions include:

a) the more temporal noise there is, the greater the number of images that must be examined before a decision can be made,

b) at the 50% noise level there is no information in the signal,

c) the greater the number of competing hypotheses, the more images that must be examined in order to discriminate from among them.

An additional interesting fact is that the minimum number of inter-image descriptions needed for a decision in any case is 2. Under the conditions in the experiment, this implies that no motion could be recognized unless it appeared in at least three images. We shall see below that this is a characteristic of the decision threshold, and that in certain cases, this can be improved. However, this is a valuable property for our classification scheme to possess: we have the ability to control the response of the system. This provides us with a handle on the "focus" problem present in so many "expert" systems in the face of extraneous or erroneous data.

Response of a focus mechanism based on feedback is not instantaneous - there is an inherent delay. Past expert systems, regardless of their domain of application, have attempted to provide a completely updated view of the analysis of the problem domain at each instant during processing. Therefore, any errors in the input data would be completely integrated as if they were valid data, rather than waiting to see if the trend is indeed that which is suggested by the data. The focus should not shift abruptly in a temporal classification scheme: it should only change if the stimulus dictating the change is present over a significant period of time. Such time periods are clearly problem dependent. Our relaxation process accumulates the evidence of the matching history of a hypothesis, and thus, it is very difficult for an isolated error event to negate a large positive history of successful matching (or, for that matter, a history of matching failures).

Another interesting observation can be made from this graph. If there are, for example, 5 competing hypotheses, under 10% noise conditions, then the system must examine at least 10 inter-image descriptions (11 images) in order to discriminate from among them. This means that if the motion concepts represented by the hypotheses have expected minimum duration D, then the sampling rate of the film must be at least 11/D in images per second. This can be generalized to the following relationship:

$$SR_{min}=\max_{i}\frac{(\tau_i+1)}{dur_i}$$

This relationship ties together the minimum required scene sampling rate for proper classification, the number of images required for classification from the graph for each set of competing hypotheses, and, the shortest expected duration for that set of hypotheses. This relationship is important because it places a constraint on the data presented to our motion classifier: only if this is satisfied will the shortest event in the KB be recognized properly. Scene sampling rate is clearly an important consideration in the design of a classifier for time-varying events, just as it is in the reproduction of a time-varying signal. In the case of signal reproduction, using the standard definition of the Nyquist rate, one must sample the signal at least at a rate of twice the maximum frequency present, or in other words, 2 divided by the minumum waveform "period" or duration for periodic signals since frequency is the inverse of period. Our relationship also has this same form, replacing the constant 2 by a value that can be extracted from the figure, (greater than or equal to 2). We are interested in classification and not in simple reproduction: however, this similarity is quite satisfying conceptually.

In our second experiment, we investigated the effect of the IS-A relationship on the competing hypotheses of the first experiment. Here, all the competing hypotheses had the same IS-A parent (see Figure 4a), i.e., the same global constraints, and this constraint is satisfied in the image sequence. Thus, the IS-A constraint signifies "complementary" hypotheses as opposed to the similarity relationship that specifies "competing" hypotheses. The results are shown in Figure 5. Note the sharp difference in slope of the curves between this figure and Figure 3, except for the 50% noise case (as is proper). The effect of the IS-A relationship significantly speeds up the decision time for large numbers of hypotheses and when noise is present. This verifies our expectations. Knowledge can be structured by providing descriptions at varying levels of abstraction, from coarse to detailed. Feedback can then be present between levels, with the coarse descriptions placing strong, more global constraints on the detailed descriptions, thus assisting in the removal of isolated noise stimuli. So, the IS-A relationship not only provides a conceptual benefit from the point of view of knowledge organization, but also provides performance benefits. In addition, if none of the IS-A sons matches the data in the image sequence, so that no decision can be made from amongst those hypotheses, but the IS-A parent matches the data, a conceptual description could still be provided at a more abstract level. This would further enhance the "intelligent" behaviour of our system in that it does not just give up if no descriptive term in the KB at that level is appropriate, but instead, it provides a more general view of the motions.

Let us further investigate the IS-A relationship. Suppose two levels of IS-A ancestors are considered as in Figure 4b. In this case, and in agreement with our expectations, no significant change from the previous case was observed. This structure does not change the nature of the competition at the bottom level and thus no change should be seen. Thus, the addition of ancestors in the IS-A hierarchy is "free" - we gain in conceptualization, and do not pay a price in terms of performance

Suppose that the structure of the IS-A hierarchy is changed to that in Figure 4c. In this case, there are two levels of IS-A, and both levels exhibit competition amongst IS-A siblings. The results of this experiment are shown in Figure 6. The slopes of the curves are again less than those of Figure 3, but not by as much as in Figure 5. The situation here becomes rather complex for explanation at an intuitive level. In effect, we have entire "branches" of the hierarchy in competition with each other, with each level affecting the other. The greatly beneficial effects of the IS-A constraint, as in the first experiment with IS-A, apply only for the top level of this hierarchy and are partially negated by the straight competition present in the second level. The top level of IS-A assists in discrimination, as well as re-inforcing IS-A sibling hypotheses, while the lower level only re-inforces IS-A sibling hypotheses. These effects cause the additional time delay observed in comparing the cases in Figures 5 and 6.

Summarizing the conclusions of these experiments, we see that:

a) IS-A provides a practical conceptual tool for knowledge organizing that is to be used for recognition.

b) The number of images of data required for classification is decreased by the imposition of the global constraints provided by the IS-A relationship.

c) Knowledge should be structured along IS-A in such a way so that it reflects the fact that competition amongst IS-A "branches" does not lead to as good performance results as does competition amongst IS-A siblings.

In a previous paragraph, motivation was given for the selection of the decision threshold. Suppose that this is now changed. If we decrease the parameter k in the expression

$$\left(1-\frac{1}{N}\right)*\left(1-e^{-k}\right)+\frac{1}{N}$$

from its value of 1.0 (as used for the previous experiments), we should observe that, obviously, we can make decisions faster . We repeated the first experiment for varying values of k (0.9, 0.8, 0.7, 0.6, 0.5), but without noise, and observed that the lower limit of 2 inter-image descriptions remains at 2 until k becomes 0.6 when it falls to 1. This is clearly also the absolute limit for motion. We must be careful, however, in the use of a threshold such as this. In exchange for a faster response, we sacrifice our stable and non-erratic focus of attention. Only in cases where there is no extraneous or erroneous data should such a low certainty threshold be used. In addition, we would expect that for large amounts of temporal noise, lowering the decision threshold may degrade performance to such a degree so that the updating scheme is rendered useless. Determination of the limits on temporal noise with respect to decision threshold is a goal of a future experiment.

Conclusions

Experimentation is a powerful tool, used extensively throughout the scientific world. Artificial intelligence researchers introduce complex schemes for the solution of difficult tasks, and for the most part, a clear understanding of these complex schemes can only be obtained via experimentation and performance analysis, and not only by the demonstration of a few examples. In this paper, a classification scheme for time-varying events has been briefly described and several of its characteristics have been presented. We believe that our experimentation has demonstrated that IS-A is a powerful structuring tool for knowledge and that gains can be had both conceptually and computationally. Although our system, ALVEN, has successfully handled several image sequences, both of human left ventricles and of moving dots, we believe that the results of these experiments provide much stronger statements on the appropriateness and performance of our methodology than do examples. Several attributes of motion classification schemes were presented, however, analysis of our scheme is not yet complete with respect to all of them. Research is continuing particularly for the temporal constraints present in the knowledge base.

Acknowledgements

I would like to acknowledge the contributions made to this research by John Mylopoulos, Dominic Covvey, and Steve Zucker. Fincancial support was gratefully received from the Canadian Heart Foundation and from the Natural Sciences and Engineering Research Council of Canada.

References

1. Zucker,S.W., "Production Systems with Feedback", in ***Pattern-Directed Inference Systems***, ed. by Waterman and Hayes-Roth, Academic Press, 1978.

2. Tsotsos, J.K., Mylopoulos, J., Covvey, H.D., Zucker, S.W., "A Framework For Visual Motion Understanding" ***IEEE Trans. on Pattern Analysis and Machine Intelligence***, Nov. 1980.

3. O'Rourke, J., Badler, N., "Model-Based Image Analysis of Human Motion Using Constraint Propagation", ***IEEE Trans. on Pattern Analysis and Machine Intelligence***, Nov. 1980.

4. Ullman, S., ***The Interpretation of Structure from Motion***, MIT Press, Cambridge, Mass., 1979.

5. Rashid, R., "Towards a System for the Interpretation of Moving Light Displays", ***IEEE Trans. on Pattern Analysis and Machine Intelligence***, Nov. 1980.

6. Tsotsos, J.K., "A Framework for Visual Motion Understanding", CSRG TR-114, University of Toronto, 1980.

7. Zucker, S.W., Hummel, R.A., Rosenfeld, A., "An Application of Relaxation Labelling to Line and Curve Enhancement", ***IEEE Transactions on Computers***, Vol. C-26, April 1977.

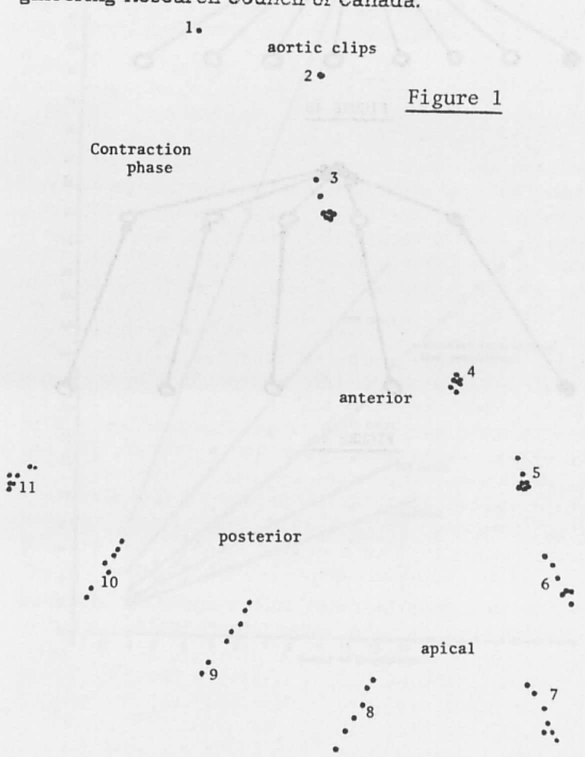

Figure 1

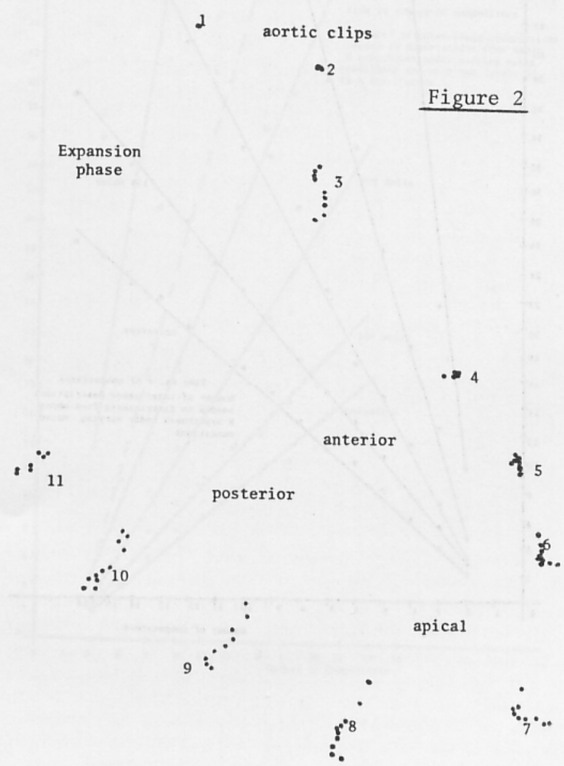

Figure 2