

ALVEN: A Study on Motion Understanding by Computer*

John K. Tsotsos and John Mylopoulos

Department of Computer Science, University of Toronto

H. Dominic Covey

Cardiovascular Unit, Toronto General Hospital, &
Computer Systems Research Group, University of Toronto

Steven W. Zucker

Department of Electrical Engineering, McGill University

Abstract

This paper presents an overview of a framework for the analysis of motion from sequences of images by computer. The framework includes:

- (a) A representation of knowledge for motion concepts that is based on Minsky's frame theory and semantic networks;
- (b) Associated algorithms for recognizing these motion concepts. These algorithms implement a form of feedback, by allowing competition and co-operation among local processes. They also allow a change of attention mechanism that is based on similarity links between knowledge units.

The framework is being realized with a system called ALVEN, whose structure and operation are described. The purpose behind this system is to provide an evolving research prototype for experimenting with the analysis of certain classes of biomedical imagery, and for refining and quantifying the body of relevant medical knowledge.

1.0 INTRODUCTION

The main goal of this research, from an AI viewpoint is the development of a framework for Motion Understanding by Computer. This framework is expected to provide:

- (a) A representation of knowledge for motion concepts, based on frames [Minsky75] and semantic networks [Brachman79].
- (b) The description of basic motion concepts such as verbs, shape transformations and temporal information in terms of the knowledge representation proposed, based on past work in motion representation [Badler75], [Tsotsos76];
- (c) The development of algorithms for the analysis of pictorial data in terms of a motion knowledge base. These algorithms will be expected to generate a description of the objects and motions found in the pictorial data at different levels of abstraction, from detailed LV wall kinematics to summary statements using English-like verbs. The recognition control structure is based on the concept of competition and co-operation among local hypotheses [Zucker78] both at the low (image) level [Rosefeld,Zucker, Humel76] and at higher (conceptual) levels.

ALVEN (A Left Ventricular Wall Motion Analysis Computer Consultant), is a system for the analysis of cinecardioangiograms (films of the human left ventricle). Given such a film, the system's job is to generate a conceptual des-

cription of the shapes and motions exhibited in the film by the left ventricular wall, noting abnormalities and unusual occurrences.

From a medical point of view, ALVEN's importance is based on its ability to provide an assessment of the human left ventricle (herein-after LV), as a functioning muscle. The aim of the research is twofold:

- (a) To produce a system that can aid in the analysis of cinecardioangiograms, and which produces consistent and objective descriptions of left ventricular wall dynamics,
- (b) To provide a framework for the representation of medical knowledge about the dynamics of the human left ventricle.

ALVEN is a first step towards achieving these objectives. Past descriptions of ALVEN may be found in [Tsotsos78] and [Tsotsos79].

2.0 REPRESENTATION OF KNOWLEDGE

2.1 What Needs to be Represented?

The main concern of this representational formalism is how motion concepts are to be represented and organized into a general motion concept and problem-specific concept knowledge base. There are several types of motion concepts:

- a) speed of an object
- b) trajectory of motion
- c) descriptive motion verbs
- d) qualitative trajectory descriptions - directionals
- e) changes of an object's properties
- f) temporal concepts: duration, start and end time, relative timing of events.

In addition, the representation must be flexible enough to handle "fuzzy" data, different levels of abstraction including linguistic levels, and to represent enough information for approximate regeneration of analyzed motions.

2.2 Representational Building Blocks

The basic entities of the representation are called frames and are used to model abstract concepts such as that of HEART or BEATING. The instantiations of the abstract concepts of the real world are modelled in the representation by frame tokens (or simply tokens). The INSTANCE-OF relationship relates tokens to the frames of which they are instances. Thus, "john's LV", an instance which represents a particular left ventricle, is related to "LV" through the INSTANCE-OF relationship.

* This paper is a summary of [Tsotsos79].

The internal structure of a frame is made up of two distinct components: information on how to instantiate the frame (fill the slots which make up the semantic components of the concept which the frame defines), and information on what concepts must be present in the description derived thus far in the analysis in order for the defined concept to apply. These two components are termed dependant and the prerequisite respectively. Both parts are made up of slots. The difference lies in the default values of the slots. The slot defaults for the dependant slots are made up of references to slots of the prerequisite part and function calls which use these values to calculate the instantiated value of the slot. Defaults in the prerequisite slots give expected values for observed data. Slots have associated with them a name, a frame type, constraints that must be satisfied by any candidate value and exceptions that are to be raised if those constraints are not satisfied.

2.3 Frame Organization

The frames which are part of a knowledge base are organized into an IS-A hierarchy in terms of the (partial order) relationship IS-A [Brachman79].

The prerequisite slots of a frame must conform to specific conditions as well. If frame A defines a motion concept for object C, the prerequisite slots must describe:

- a) the motion sequences for C, including any parallel descriptions which must be added together to form the overall motion of C;
- b) the motion sequences of all objects C' such that C' is a PART-OF C, thus conforming to the PART-OF hierarchy of C.

In addition to these two static organizational concepts, the representation allows the use of similarity links as an aid in transferring control between contexts and classifications during recognition.

Similarity links are associated with the frame concept defined and are based on exceptions which occur in the prerequisite components of that frame. There are three major components to the data stored along with a similarity link:

- a) compatibility expression - This Boolean expression contains the necessary conditions which the two linked frames must share in order for the similarity link to be activated (i.e., the "similarities").
- b) exception expression - This is a list of Boolean conditions in which the predicates are exceptions which may be raised during matching of the frame. Each of the list's elements has associated with it a time when the exception should occur. In effect, what such an expression defines is the time course of matching failures in one frame which indicate that the other frame is more relevant (i.e., the time sequence of "differences").

- c) activate-support list - This is a list of generic frame names which are to be activated or supported by the similarity link.

2.4 Left Ventricular Motion Concepts

The left ventricle is standardly described by splitting up the outline into 3 segments (when viewed in the Right Anterior Oblique projection) - posterior, apical and anterior.

While the form of LV beats is not known precisely, there are some subjective data that are relevant for an initial version of the knowledge base (KB). For example, there is a separate motion frame for each of 7 phases of an LV cycle. Some phases define contraction of the LV (maximum ejection, reduced ejection), others expansion (rapid inflow, diastasis, filling by atrial contraction) and two define a state of rest (isovolumic contraction, isovolumic relaxation - isovolumic is the key here). Therefore the phase frames are IS-A linked to the frames for "contract" and "expand" as is appropriate. In each frame, all of the segments of the LV are associated with the event or sequence of events that define the segment's motions during that phase. Correspondingly, the frame for "normal LV beat" would be made up of the events which represent each phase of the LV, linked in a cycle with appropriate timing information so that together they define the cyclic beats.

Abnormalities are represented using the same structure. The segments which have abnormal motions have separate frames which describe abnormalities expected for that type of motion. Thus, for a particular abnormality, the frame would be made up of the appropriate normal and abnormal segments and phases.

3.0 A KNOWLEDGE-BASED FEEDBACK MODEL FOR ANALYSIS OF TIME-VARYING IMAGERY

There are three main components to the feedback model: (i) image analysis using system expectations; (ii) comparison of the results to the system's expectations; and (iii) adjustment of the system's state and expectations.

3.1 Image Analysis

The starting point for the system is an initial state (say, "NORMAL" LV motion is expected). Thus, the normal LV motion frame is activated along with all of its parts and generalizations (along the IS-A hierarchy). This allows generation of the first set of expectations for the next image. In our case, the first LV outline has been traced in the image. The first expectation is this same outline with a window of acceptable "normal" outlines built upon it. This is reasonable because at end-diastole (i.e., the time when the LV has completed its expansion) there is very little motion in the time spanned by the first 2 or 3 images.

The LV outline is found using an expectation guided relaxation process. Predictions enter the

relaxation process in two ways: the search region and the expected edge orientation. Relaxation is only applied to the points for which a gradient operator has produced responses, i.e., in the search region. The expected orientation is included in the computation by associating with each sub-region in the region an orientation. This orientation is then used to bias the updating function to favour edge labels of the expected orientation (and those immediately compatible with that orientation). This information is coded into a "strength" function which becomes part of the relaxation updating rule.

We call the resultant outline the "essential trace" of the LV. Motion is detected for the object if the locations of any of its parts differ between the two consecutive images. Correlation of moving shape features between the two LV outlines is done in a heuristic manner. Using the correlation information, values are determined for transformation predicates. The essential kineses for an object are the changes of its physical properties and the changes in its location.

3.2 Change of Attention Mechanism

The system maintains a set of active hypotheses throughout the recognition process. This set is initially filled with the motion hypothesis the user believes is exhibited by a particular object (or objects).

Active hypotheses are organized by their "conceptual adjacency", defined in terms of the primitive relationships used to organize the knowledge base (IS-A, PART-OF and time relationships).

The system adds hypotheses to its list of active ones via activated similarity links and via the "next" temporal constraint. Similarity links are activated when one of the exceptions specified in the exception-expression is satisfied in addition to the hypothesis compatibility expression being satisfied. An active similarity link is also propagated along the motion PART-OF hierarchy in order to ensure the proper new context for the newly activated hypothesis is activated as well.

Hypotheses are deleted from the active list in two ways. If the hypothesis' expected maximum duration is exceeded, it is removed. If the values of its definitional slots can be found, it is instantiated. If the slots cannot be filled it is abandoned completely. The second way of removing hypotheses is by simple thresholding based on the leading hypothesis' certainty. Deletion of a hypothesis is propagated to all of its parts, all parent subjects, and all IS-A sons (recursively).

3.3 Focus of Attention Mechanism

The set of active hypotheses requires further organization. The prediction mechanism needs to know which are the best hypotheses at any instant in the processing, so that it can base its expectations on them. The hypotheses will be ordered by means of the certainty which the system has in them. On activation, all hypotheses are given a 50-50 chances - a certainty of 0.5. For each image

afterwards, this certainty is updated using the paradigm of competition and cooperation among conceptually adjacent hypotheses.

The updating is done by a second modified relaxation labelling rule for which compatibility are defined between hypotheses depending on the type of conceptual adjacency they exhibit. For example, mutually exclusive hypotheses are those with a common direct IS-A ancestor. Hypotheses which follow or precede each other in time have large negative compatibilities to aid in their time-wise segmentation. Hypotheses which are the prerequisite slots of another hypothesis have large positive compatibilities with their parent hypothesis. A hypothesis has a compatibility dynamically set with itself for its following time interval depending on its matching progress. Finally, hypotheses linked by similarity links also have dynamically set compatibilities depending on whether one supports the other during that time interval.

4.0 PROJECT STATUS

Currently, we are in the final stage of a prototype implementation of ALVEN. The low level modules (essential trace and essential kineses) have been implemented and are being tested. The data collection module has been implemented and films are being processed. A syntax for the representation of knowledge has been defined and a compiler for entering and maintaining the KB has been implemented.

References

- Badler, N., "Temporal Scene Analysis: Conceptual Descriptions of Object Movements", TR-80, Dept. of Comp. Sci., U. of Toronto, 1975.
- Brachman, R., "On the Epistemological Status of Semantic Networks", in *Associative Networks*, N. Findler, (ed.), Academic Press, 1979.
- Minsky, M., "A Framework for Representing Knowledge", in, *The Psychology of Computer Vision*, Winston, (ed.), McGraw-Hill, 1975.
- Tsotsos, J., "A Prototype Motion Understanding System", TR-93, Dept. of Comp. Sci., U. of Toronto, 1976.
- Tsotsos, J. et al., "Gross and Segmental Wall Motion Analysis in Dynamic Cardiac Imagery", *Proc. Computer Applications in Medical Care*, Washington, D.C., 1978.
- Tsotsos, J. et al., "A Framework for Visual Motion Understanding", AI Memo 79-2, Dept. of Comp. Sci., U. of Toronto, 1979.
- Zucker, S. et al., "An Application of Relaxation Labelling to Line and Curve Enhancement", *IEEE Trans. on Computers*, vol.C-26, 1977.
- Zucker, S., "Production Systems with Feedback", in *Pattern-Directed Inference Systems*, Waterman and Hayes-Roth, (ed.), Academic Press, 1978.