SOME NOTES ON MOTION UNDERSTANDING

John K. Tsotsos
Department of Computer Science
University of Toronto

The issues of recognition of and re-presentation for motion concepts, with all their semantic richness, have been largely ignored by the AI community. However, there are exceptions, such as Miller [1], Abrahamson [2], and in particular, Badler [3]. Badler considered image sequences attempting to provide English descriptions for motions recognized in the images. He provided precise definitions for direction-als and adverbials and also outlined ideas on representation of the verb semantic components from [1].

These works do not claim to solve the problem of motion understanding. Firstly, a coherent mechanism is lacking for structuring the defining components of verbs. In addition, temporal relationships between components and a way for determin-ing the duration of a verb from its compo-nents are needed. Secondly, the design and control structures of a system which accepts visual input and recognizes motion concepts has not been outlined in much detail.

This short note describes on-going work in these two areas and is detailed in [4].

Frames [5] are used for structuring verb definitions and have the following characteristics. There are slots with selectional restrictions which become bound to participating objects in the scene. Functional relations (active knowledge com-ponents) between concepts (particularly temporal concepts) are needed, such as mathematical relations and concepts which aid in determination of which of the recog-nized objects fill which slots. Prerequi-sites act as triggers for activation of the frame. These prerequisites correspond exactly to the semantic components of the verb [1]. Instantiation of a frame occurs only after the following has taken place: the frame is activated; each of its prere-quisites must be instantiated; all temporal relationships must be verified; and, the duration of the verb must be determined. The start-time of the verb is the maximum value of the start-times of its prerequi-sites, while the end-time is the minimum of their end-times. That is, all prere-quisites must exist during a common time interval. If this condition is violated, the frame is de-activated, and the condi-tion is checked on recognition of the start- and end-times for each prerequisite.

A frame interpreter has been implemen-ted which performs the above functions. However, it also performs one other function to be described below.

The necessary control structures are now outlined. (The overall structure is similar to that described by Freuder [6].) There is an independent low level module which can recognize objects in the scene, compute spatial relations between objects on command, and recognize when an object changes location (assuming a stationary observer). It independently instantiates objects in the scene and their locations. The spatial relations are computed only when it is queried about an object. The other major module is the frame interpreter as described above. Its other function is to provide guidance for the low level. A frame concept is activated by the instan-tiation of one of its prerequisites. How-ever, the fact that the object moves cannot be one of the activating prerequisites because this would activate all frames. For this fact, the interpreter queries the low level about basic spatial relations for the object. On activation, the interpreter then queries the low level about the re-maining prerequisites. The low level per-forms computations to determine when and if these are instantiated until it is instruc-ted to delete these from its query list (when the frame is de-activated).

A basic question here is whether or not the low level becomes swamped with queries during processing of the scene. In the simple scenes considered, an inter-esting fact was observed which is also reported in [6]. The query list grows un-til part way through the scene and then shrinks in size as concepts are instantia-ted.

The frame collection now includes 12 of Miller's primitive motions [1] with several verbs in each class. The concepts described above have been sufficient for the verbs attempted thus far.

References
[1] Miller, G.A., "English Verbs of Motion: A Case Study in Semantics and Lexical Memory", in Coding Processes in Human Memory, ed. by Melto § Martin, 1972.
[2] Abrahamson, A.A., "Experimental Analy-sis of the Semantics of Movement", in Explorations in Cognition, ed. by Norman § Rumelhart, 1975.
[3] Badler, N.I., "Temporal Scene Analysis: Conceptual Descriptions of Object Movements" TR-80, U. of Toronto, 1975.
[4] Tsotsos, J.K., "A Prototype Motion Understanding System", TR-93, U. of Toronto, 1976.
[5] Minsky, M., "A Framework for Represent-ing Knowledge", in Psychology of Computer Vision, ed. by P.H. Winston, 1975.
[6] Freuder, E.C., "A Computer System for Visual Recognition Using Active Knowledge", AI-TR-345, MIT AI LAB, 1976.