OCCLUSION IN DYNAMIC SCENE ANALYSIS$^{1,2}$

W.N. Martin$^3$ and J.K. Aggarwal$^{3,4}$

TR-80-1  
September 1980

---

$^1$This research was supported in part by the Air Force Office of Scientific Research under grant AFOSR 77-3190.

$^2$This report is reprinted from the paper that appears with the same title in the book, Digital Image Processing and Analysis, edited by J.C. Simon and R. Haralick. This book is published in conjunction with the NATO Advanced Study Institute 80/16 held in Bonas, France, June 23 - July 4, 1980.

$^3$Computer Sciences Department

$^4$Electrical Engineering Department
University of Texas at Austin
Austin, Texas 78712
OCCLUSION IN DYNAMIC SCENE ANALYSIS

W.N. Martin and J.K. Aggarwal

Departments of Electrical Engineering and Computer Sciences
The University of Texas at Austin
Austin, Texas 78712

Abstract: This paper presents several fundamental concepts necessary for the successful analysis of dynamic scenes containing occluding objects by discussing various systems which have been developed to perform this analysis. The dynamic scenes are represented by time ordered sequences of images. Data must be extracted from each of these images and then integrated into coherent information about the sequence as a whole.

1. INTRODUCTION

The term "dynamic scene analysis" has been used to refer to the process of analyzing time ordered sequences of images, and in an earlier survey [1] we have observed that for a computer vision system to recognize and understand the motion apparent in the image sequences several levels of analysis are required. Peripheral and attentive processes constitute two of the more important levels. The peripheral processes detect motion and direct the attentive processes to it, while the attentive processes track the movement and attend to the details of the objects in motion. Cognitive processes form a level in which the specific information acquired through the other levels is integrated with the goals and expectations of the vision system. In this paper we will discuss two systems [2,3] incorporating attentive level processes that have been developed at the University of Texas at Austin. Several other systems have been developed here and reported in [4,5,6,7,8], while additional surveys of the area can be found in [9,10].
2. DYNAMIC SCENE ANALYSIS WITHIN A DOMAIN OF POLYGONAL FIGURES

Our initial system [2,11] analyzed dynamic scenes containing arbitrarily complex rigid polygonal figures, possibly having holes. These figures were allowed to move with various rotational and translational velocities in planes parallel to the image plane. For this domain the polygons were required to be rigid, but were not restricted in their movement. In this way the polygons could occlude one another, combining their projections in the image plane to form apparent objects which changed shape as the occluding polygons changed their relative positions. Thus the main task of the system was to decompose the apparent polygons into constituents that corresponded to the visible portions of the actual polygons in the scene. This task was accomplished by tracking the rigid parts (called "real" vertices) of the apparent polygons and using the acquired information to interpret the changes in the remaining parts.

Here we should make the important point that dynamic scene analysis systems must have access to image features which remain constant through the image sequence or at least change so slowly as to be constant through subsequences of images. These image feature may be descriptors associated with various "tokens" in the image. Tokens are groupings of descriptors considered to be indicative of salient components of the scene. In [2] the tokens were the vertices of the apparent polygons and had descriptors for the spatial location of the vertex point, the length of the two polygonal sides incident on the vertex, and the interior angular measure of the vertex. Clearly, for moving polygons that can occlude one another only the latter descriptor would remain constant and thus it was used as the primary tracking feature.

The tracking, however, was mostly concerned with forming a correspondence between given scene components and the tokens in each image of the sequence. The angular measure at a vertex was not a completely sufficient tracking feature for several reasons. First, the angular measure was not a uniquely identifying property. Second, the analysis of the apparent polygons required the identification of the vertices formed by occlusion (called "false" vertices) even though their angular measure was changing. The tracking was accomplished with the aid of object models, created and continually updated by the motion analysis system. These models provided the additional constraints for identification by grouping the vertex tokens into ordered sets, i.e., objects, indicating the system's interpretation of the apparent polygons at any given time. The interaction between the tracking process and the interpretation structure, i.e., the object models, is another important feature of systems which attempt to analyze the motions of complex objects.
The example scene displayed in Figure 1 contained one large polygon rotating in a clockwise direction and initially obscuring two smaller polygons. A vertex of one of the obscured polygons appeared in Scene 2 while the remaining polygon first appeared in Scene 4. The interpretation of this scene derived by the system after five frames is shown in Model 5. Three separate objects were represented with two objects having only partial descriptions and the large polygon having its occluded parts restored by information from previous frames. These models provided the additional capability of forming complete object descriptions for polygons that were only partially visible in any given image.

3. DYNAMIC SCENE ANALYSIS WITHIN A DOMAIN OF CURVILINEAR FIGURES

The two subsequent studies [12,3] removed the constraint on polygonal objects. The first [12] made extensive use of a predictive model. The primary tracking was performed on a token created for each apparent object. The descriptors used were the spatial location of the centroid of the figure, the area, the direction of the major axis, and the size of the enclosing rectangle oriented to the major axis. For tracking of non-occluding figures the match between tokens in consecutive images was established on the basis of the latter three descriptors. If the figures occluded, however, the predictive model incorporated information from previous images to form expectations for the descriptor values. These expectations were then matched against the current token values. A disadvantage of the predictive scheme was that it required occluding objects to maintain constant motions. The reason for this restriction was that the descriptors retained in the tokens did not allow the updating of the motion estimates of the figures while they were occluding.

The second study [3] removed this restriction by modeling the figures with ordered sets of tokens. An edge detecting preprocessor [13] extracted the object boundaries from graylevel images of curvilinear planar figures which exhibited the following properties:

1. the edges in the images were the boundaries of figures;

2. when two figures overlapped the boundary between them was not discernible, thus occluding figures appeared as a single figure; and

3. the figures moved in planes parallel to the image plane, so that the scale of the images did not change.
A coordinate list representation of the closed boundaries was obtained from the images and inserted into a data base which contained all relevant information derived from the sequence. For this system the moving objects were to be tracked on the basis of shape features so an additional representation for each boundary was also incorporated into the data base. This shape representation was derived from a chain encoding of the subtended angle versus arc length, $\psi - s$, function as measured from an arbitrary starting point. A piecewise straight line approximation of the pictorial graph of the $\psi - s$ function constituted the representation which had several advantages, as listed in the following:

1. reduced the amount of storage required;
2. easily smoothed to attenuate noise;
3. could be processed to remove all effects of arbitrarily choosing the starting point of the $\psi - s$ function;
4. for matching shapes, differences in the functions corresponded to differences in shape; and
5. most importantly, effectively decomposed the shape into an ordered set of arcs, the tokens for this system.

The last of the above mentioned items arose from the fact that straight lines in the graph of the $\psi - s$ function corresponded to circular arcs in the image with the slope of a given line proportional to the curvature of the corresponding arc. The $\psi - s$ function for an example object is shown in Figure 2. Figure 3 displays the object boundary, as decomposed into the arcs associated with each straight line in the graph.

The tokens contained descriptors for the length and curvature of the arcs, while the ordered sets maintained the adjacency of the arcs along the boundary. In this way the correspondence necessary for tracking was established by matching shapes across image pairs. That is, subsets of the tokens associated with a boundary in one image were mapped into subsets of tokens for a boundary in the next image by matching the shape of the arcs described through those tokens. Thus contiguous arcs in one image which match, in the same order, to contiguous arcs in the second image were grouped into edge segments. This matching was performed by first choosing two arcs, one from each image of a consecutive pair, whose $\psi - s$ function lines had similar slopes and lengths. From these "seed" arcs an edge segment was "grown" by adding contiguous arcs to either end of the already matched segments until a dissimilarity in the curves was found. The dissimilarity of two curves was measured by the area between
the normalized graphs of their $\psi$-s function graphs. Two arcs were declared dissimilar when the measured value exceeded a preset threshold.

Edge segments grown in this way represented the portions of the object boundaries which had retained their shape through the sequence. Thus an edge segment related two views of some part of an actual object boundary. Motion measurements calculated from these views were then used to group the edge segments into object models under the assumption that edge segments which exhibit a common motion belong to the same object. Again tokens with appropriately constant descriptor elements and dynamic modeling were important factors in the success of the system.

A part of an example shown in Figure 4 was taken from a scene which contained three actual objects: one central stationary object; one object on the left side moving from top to bottom; and one object on the right side moving toward the upper left corner. The first two images, however, contained only one apparent object. When comparing the shapes of the first image to those of the second, the system formed four edge segments. These edge segments were then grouped into three object models based on motion measurements. The object models formed in this way were inserted into the data base with arbitrarily chosen names. In Figure 4 each edge segment was labeled with the name of the appropriate object model. The observant reader will have noticed that no edge segments were formed from the center section of the first two images. This was due to the extensive shape changes occurring in that part of the scene. It should also be noted that in the last image the upper object no longer overlapped the other object, causing the number of apparent objects to change. The shape matching procedures were able to handle this case and made the proper correspondence between the images of the last pair.

The salient features of the above system are as follows:

1. the figures were described by edges which were represented by both spatial coordinate lists and $\psi$-s functions;

2. figures were matched between consecutive images of the sequence on the basis of the shape information contained in the $\psi$-s functions;

3. the matched figures were used to calculate motion measurements; and

4. motion was used to form object models under the rigid body assumption.
The system used these features to analyze successfully several sequences, such as the example above, which were generated by our image dissector camera.

4. CONCLUSION

In this paper we have attempted to elucidate several important aspects of the problem of analyzing occlusion in dynamic scenes. These included the following: system considerations such as multilevel design with the peripheral, attentive, and cognitive functions implemented as separate yet cooperating processes; the use of constant or slowly changing object attributes in forming the correspondence between components, i.e., tokens, in consecutive images of the sequence; the necessity of having decomposable object representations for the identification of partially occluded objects; object model creation and maintenance for interpretation of the scene; and the fundamental importance of the interaction between component tracking and scene interpretation.

Acknowledgement: We wish to express our gratitude to Professor Simon and Professor Haralick for organizing and hosting the NATO ASI. The research presented here was supported by the Air Force Office of Scientific Research under Grant AFOSR 77-3190.

REFERENCES


Figure 1. Input sequence and generated model interpretation for a scene containing polygonal objects.
Figure 1. (Continued)
Figure 2. Graph of $\psi-s$ function for the curvilinear object shown in Figure 3.
Figure 3. Object boundary as segmented into arcs by the $\psi$-s function shown in Figure 2.
Figure 4. Input sequence and generated model interpretation for a scene containing curvilinear objects.
Figure 4. (Continued).