

EXTRACTION OF MOVING OBJECT IMAGES THROUGH CHANGE DETECTION

Ramesh Jain
Department of Computer Science
Wayne State University
Detroit, MI. 48202

W. N. Martin, J. K. Aggarwal
Departments of Electrical Engineering
and Computer Sciences
The University of Texas at Austin
Austin, Tx. 78712

Abstract: This paper discusses a scheme for extracting the images of moving objects in dynamic scenes. Differencing operations are used to identify areas containing moving objects. The images of the moving objects can then be obtained by focusing the segmentation processes on these restricted areas. Thus motion is used as a cue to segmentation.

1. INTRODUCTION

Computer vision systems are usually composed of subsystems for segmentation and interpretation. Although motion has been suggested as a powerful cue for segmentation and has attracted increasing attention from researchers [6,7], little research has been concerned with the use of motion for segmenting images. Potter [8], and Fennema and Thompson [2], proposed schemes for segmentation using the velocity estimated at individual pixels. Jain et al. [3], and Jain and Nagel [4] showed that the image of a moving object may be extracted by using interframe differencing operations. Their scheme [4], however, has limited applicability because images can be extracted only when they have been displaced completely from their position in a specified reference frame, which is usually the first frame of the sequence.

We have developed an approach for recovering the images of moving objects, which exploits the image transformations due to motion. This approach has been implemented on a DEC-10 in Pascal. The input to the program is a sequence of registered frames of a dynamic scene. The output is in the form of binary images, i.e., two-dimensional masks, of the moving objects, or greylevel images obtained from the input frame using the mask. This paper describes our approach. The concepts and details of the method are given in [5].

2. OVERVIEW

On receiving each raw image, a condensed frame is prepared [4,5]. Starting with the second frame, a difference picture, denoted DP, is prepared for each frame by comparing it with the preceding frame. Regions are found in the difference picture, and then the previous frame edge picture and current frame edge pictures are prepared. As has been shown by Jain and Nagel [4], it is possible in many cases to determine whether a DP region is due to the covering of the background, due to uncovering of the background, or due to both the covering and uncovering of the background. Let us denote these three types of DP regions as type O, B, and X, respectively. For a given DP region, the type can be found by computing a ratio called CURPRE, which is defined as

$$\text{CURPRE} = \text{CC/CP}$$

where CC (CP) is the number of points which are both extreme points of a DP region and edge points in the current (previous) frame. The extreme points of a region are the leftmost and rightmost 1 for a row, and the topmost and bottommost 1 for a column. As has been shown in [4], this ratio is much less than 1 for type B regions, much higher than 1 for type O regions, and near 1 for type X regions.

After the classification of the DP regions, our approach comprises two main phases which are applied to each pair of consecutive frames in the input image sequence. In the first phase, depending on the class of the DP region, a process is applied to obtain a region in the previous (or current) frame which is an estimate of the mask of the moving object image. This phase is called the recovery process. The second phase modifies this estimate of the mask by testing whether or not the motion of an object having the image

This research was supported in part by the Air Force Office of Scientific Research under grant AFOSR 77-3190 and in part by the National Science Foundation under grant ENG 74-04986.

corresponding to the mask could result in the transformations indicated by the difference picture. This phase is called the refinement process.

3. RECOVERY PROCESSES

We use region growing for O and B type regions to obtain masks in previous and current frames. This method is applied to the DP regions which are either O or B type, since a region which is classified as one of these two types is considered to be a core mask. A region is grown by taking each nonregion pixel which has a horizontal neighbor within the region and comparing its greylevel to that of the adjacent region pixel. If the pixels are similar then the nonregion pixel is added to the region. The greylevels are taken from the previous frame if the DP region is type B and from the current frame if it is type O. A similar process is applied to nonregion pixels which have vertical neighbors within the region. Both processes are then iterated until no new pixels are added to the region. The region thus obtained represents the previous frame object mask when grown from a type B region and the current frame object mask when grown from a type O region.

4. REFINEMENT PROCESSES

In this section we consider various refinement processes, which are used to improve the masks obtained by the recovery process. The details of these processes may be found in [5].

SAME OBJECT. After recovering masks in the previous and current frames, for B and O type regions, respectively, the following method of refining masks is used. In this refinement, masks are modified by using the fact that a moving object has masks in the previous as well as current frames. Some parts of the masks may not be recovered due to noise or coincidences. The same object refinement modifies the masks to account for such lost parts.

Same object refinement is also important in processing DP regions formed by occluding objects. If region K is type O while regions I and J are type B and the masks of regions I and J in the previous frame have some overlap with the mask of region K in the current frame, then objects whose images do not occlude in the previous frame have occlusion in the current frame. Similarly, occlusion of objects in the previous frame may also be found.

Type X regions become important for occluding object analysis in the following case. If region K is type X while regions I and J are B type and if the masks of regions I and J in previous frames touch or overlap region K in

the difference picture, then objects whose images do not have occlusion in the previous frame have occlusion in the current frame. The region K is the result of occlusion of the objects. This region is grown in the current frame to obtain a mask of the composite object in the current frame. In such a situation, if the regions I and J are type O, then region K is grown in the previous frame.

TERMINATION REFINEMENT. Normally for a rigid object moving completely in the field of view, difference entries result at both its leading and trailing ends. This fact may be utilized for refining those regions which have been incorrectly grown by including non-object points (those points which do not belong to the object image) in the object mask during the region growing process.

GAP FILLING. For a variety of reasons the object masks resulting from the recovery processes may have holes or significant concavities. We use "gap filling" to check to see if such areas are representative of the object image or just side effects of the recovery processes when applied to the particular object shape and motion.

SAME FRAME REFINEMENT. When analyzing a sequence of frames, say frames $F_1, F_2, \dots, F_n$, a difference picture is prepared for each successive pair of frames (i.e., F_1 and F_2 , F_2 and F_3 , F_3 and F_4 , ..., F_{n-1} and F_n). The object masks are extracted in each frame. For each frame, F_2 through F_{n-1} , object masks are extracted twice. For a moving object, the two masks extracted at different times but from a given frame should be the same. Due to noise and inexact recovery processes, this may not be the case in most practical situations. However, the fact that the masks are of the same object in the same frame is exploited for refining the mask.

5. RESULTS

Due to space limitation, we present the results for only one input sequence. The sequence shows the motion of two blocks in a scene containing many other objects. In Figs. 1, 2 and 3 we show three frames from the sequence taken in our laboratory. The difference picture for the first two frames is shown in Fig. 4, while the region types and CURPRE values are given in Table 1. The object masks recovered by the program are shown in Figs. 5 and 6 for the previous and current frames, respectively. After processing the third frame of the sequence, the program can apply the same frame refinement to yield the masks of the moving objects for the second frame. The moving objects start occluding in

this frame; thus a single mask is formed for the composite object, as shown in Fig. 7. The image of the objects is obtained from this mask as presented in Fig. 8. Note that all other objects have been neglected as they are stationary.

6. CONCLUSION

The present version of our program was implemented to support our belief that the transformations in frames of a scene due to motion may help in the recovery of the images of moving objects. The results of the current evolutionary stage of the program indicate that it is possible to recover the images of moving objects using only low level knowledge, even from the frames of nontrivial scenes.

The processes implemented in the current version are based on the knowledge derived from simple idealized scenes as presented in [4,5]. It is clear that a deeper understanding of the image transformations due to motion will be useful in developing better recovery and refinement processes.

REFERENCES

- [1] J.K. Aggarwal and R.O. Duda, "Computer Analysis of Moving Polygonal Images," *IEEE Trans. on Computers*, vol. C-24, 966-976, 1975.
- [2] C.L. Fennema and W.B. Thompson, "Velocity Determination in Scenes Containing Several Moving Objects," A contribution of the Central Research Laboratory of 3M Co., St. Paul, MN., and the Dept. of Computer Science, University of Minnesota, Minneapolis, MN., 1978.
- [3] R. Jain, D. Militzer and H.H. Nagel, "Separating Non-Stationary Scene Components in a Sequence of Real World TV Images," *Proc. of the 5th Int. Joint Conf. on Artificial Intelligence*, 612-618, Aug. 1977.
- [4] R. Jain and H.H. Nagel, "On a Motion Analysis Process for Image Sequence From Real World Scenes," Tech. Report, University of Hamburg, April 1978 (to appear in *IEEE Trans. Pattern Analysis and Machine Intelligence*).
- [5] R. Jain, W.N. Martin, and J.K. Aggarwal, "Segmentation Through the Changes Due to Motion," *Computer Graphics and Image Processing* (in press).
- [6] W.N. Martin and J.K. Aggarwal, "Dynamic Scene Analysis," *Computer Graphics and Image Processing*, vol. 7, 356-374, 1978.
- [7] H.H. Nagel, "Analysis Techniques for Image Sequences," *IJCPR-78*, Kyoto, Japan, Nov. 1978.
- [8] J.L. Potter, "Scene Segmentation Using Motion Information," *Computer Graphics and Image Processing*, vol. 6, 558-581, 1977.

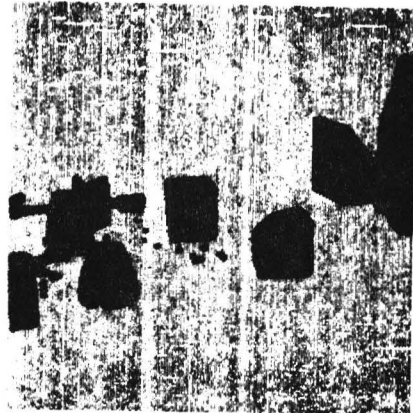


Figure 1. First frame of a laboratory sequence.

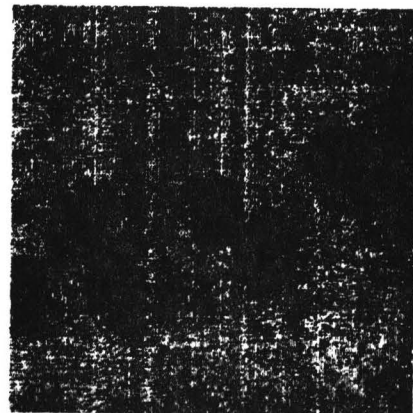


Figure 2. Second frame.

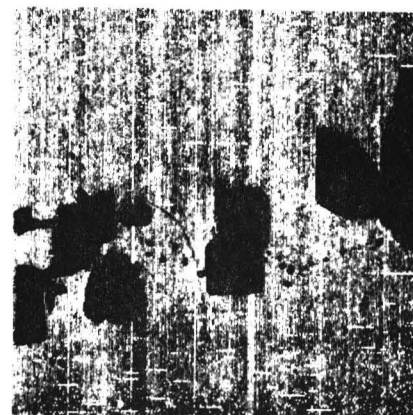


Figure 3. Final frame.

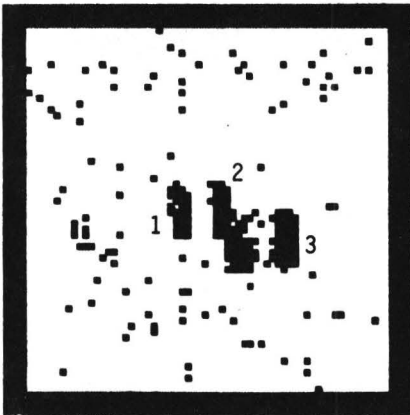


Figure 4. Difference picture for the frame pair of Figs. 1 and 2.

TABLE 1

Values for the difference picture of Fig. 4.

<u>REGION</u>	<u>TYPE</u>	<u>CURPRE</u>
1	B	0.416
2	X	1.545
3	B	0.545

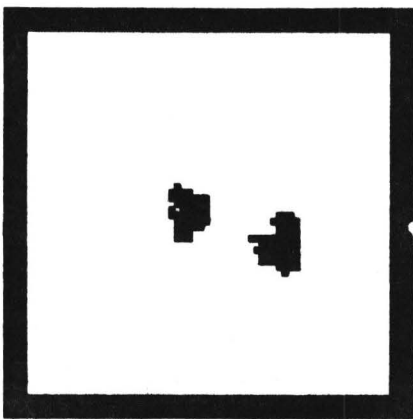


Figure 5. Previous frame object masks for the moving objects of Fig. 1.

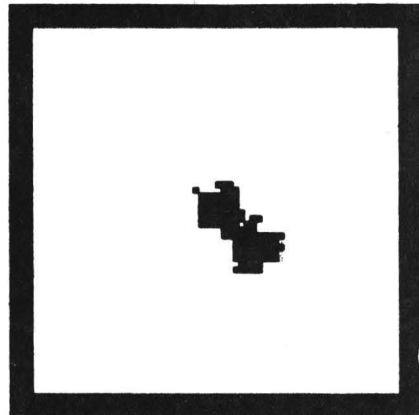


Figure 6. Current frame object mask for Fig. 2.

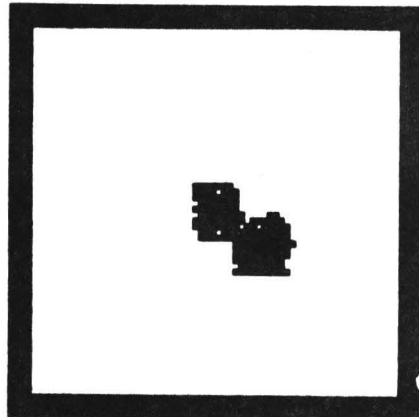


Figure 7. Refined object mask for Fig. 2.

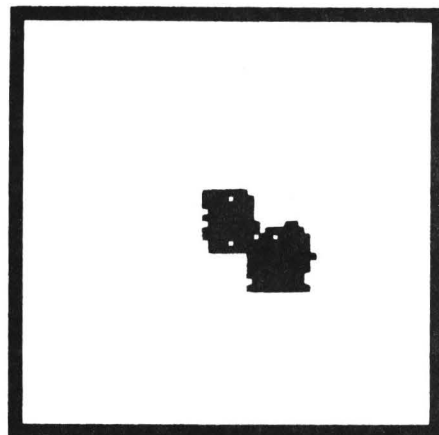


Figure 8. Moving object images extracted by the object mask of Fig. 7.