Combining Motion and Contrast for Segmentation

WILLIAM B. THOMPSON, MEMBER, IEEE

Abstract—A method is presented for partitioning a scene into regions corresponding to surfaces with distinct velocities. Both motion and contrast information are incorporated into the segmentation process. Velocity estimates for each point in a scene are obtained using a local, nonmatching technique not dependent on any prior boundary determination. The actual segmentation is accomplished using a region merging procedure which combines regions based on similarities in both brightness and motion. The method is effective in determining object boundaries not easily found using analysis applied only to a single image frame.

Index Terms—Motion, region merger, scene analysis, segmentation, velocity estimation.

I. Introduction

MOTION is a crucial property of many visual environments. Knowledge of object velocities and trajectories is clearly important for scene interpretation. Motion is also useful as a cue for scene segmentation. Velocity information may be used to link adjacent but visually dissimilar surfaces or to divide surfaces not easily separable by static criteria alone. Often, ambiguous object boundaries in a single image frame are easily resolved when dynamic effects are evaluated based on a sequence of frames. This paper describes a technique for combining both motion and brightness information into a single segmentation procedure which determines the boundaries of moving objects.

The most straightforward approach to locating moving surfaces is to compare a sequence of accurately registered image frames searching for areas of change. If velocity estimation is not required, simple pixel-at-a-time subtractive techniques applied to two frames may be appropriate [1]. In order to determine speed and direction of moving objects, differences must be evaluated over longer sequences [2]. By using both difference measures and gray scale information, segmentation of some scenes is possible using much shorter image sequences [3].

A second approach involves tracking identifiable image structures from frame to frame [4]-[6]. These systems are capable of accurately estimating surface translation, but usually only for a relatively sparse sampling of points. Potter has described

Manuscript received April 20, 1979; revised March 14, 1980. This work was supported by the National Science Foundation under Grant MCS-78-20780.

The author is with the Department of Computer Science, University of Minnesota, Minneapolis, MN 55455.

0162-8828/80/1100-0543$00.75

a similar system designed to estimate velocity at every point in a scene and thus allow for segmentation based on motion [7]. At each point in one frame, a skeletal template is defined. This template is then searched for in a second frame. The size of the template is adaptive, depending on the nearest prominent gray scale discontinuity in each of four directions. This adaptability allows for effective velocity estimates even in the center of large, smooth surfaces. A potential limitation of the approach is that no use is made of the gray scale structure of an image except for the definition of a discontinuity threshold. Thus, difficulties may arise when templates cannot be accurately determined using a simple thresholding criteria. The technique has not yet been demonstrated on complex, realistic scenes.

Much of the work in dynamic scene analysis depends on the availability of segmented image frames. If accurate information on the location of visible boundary elements is available, sophisticated techniques may be used to track objects and resolve ambiguities due to occlusion [8], [9]. If a static interpretation is available, higher order cognitive analysis can actually yield English-like descriptions of the motion [10]. Because these techniques depend on extensive analysis of each frame, they cannot be used to assist in the original segmentation task. (They may, however, provide information useful for the processing of subsequent frames.)

The next sections describe a method for using information about translational surface motion to aid in the determination of moving object boundaries. Velocity estimates are made for each point in a scene using a local, nonsearch technique requiring no prior boundary detection. These estimates are then combined with information about contrast boundaries to produce a partitioning of the scene into regions with distinct velocities. By combining contrast and velocity effects, object boundaries may be found using only two frames of an image sequence. Results demonstrate that dynamic analysis may aid significantly in the segmentation of relatively complex scenes.

II. Local Area Velocity Estimation

To use motion as a basis for segmentation, velocities must be determined in a manner not dependent on extensive static analysis of connected boundaries. In this section, an efficient procedure is described for implementing the nonmatching velocity estimation technique proposed in [11] and based on the earlier work of Limb and Murphy [12] and Cafforio and Rocca [13]. The technique is shown to be effective for translational motion of rigid objects (no significant rotations, scale changes, or deformations). Both the dominant velocities in the scene and the velocity of the underlying surface at each point are computed. The method operates on local properties rather than by identifying a feature in one frame and then searching for the corresponding feature in another frame. Velocity information is developed by relating the time variation of image intensity at a point due to motion and the spatial variation of intensity over object surfaces. This relation does not uniquely determine velocity, but it does constrain the possible speeds and directions that a moving surface may have. If, however, a sufficient number of image points correspond to a surface with uniform velocity, the true velocity vector can be determined using clustering techniques.

Spatial variability is commonly represented by a gradient field G. An incremental change in intensity, di, due to a spatial shift, ds, of the underlying surfaces giving rise to the intensity distribution satisfies the relationship

$$di = -G \cdot ds.$$

(The minus sign is introduced because it is the surface and not the observation point that is moving.) By relating translational differences to velocity and time, at a single image point we have

$$\frac{di}{dt} = -G \cdot v = -(G_x v_x + G_y v_y)$$

where G_x and v_x are the x-axis components of the gradient and velocity vectors and G_y and v_y are the y-axis components. di/dt and the gradient values may be estimated directly from an image sequence. Thus, v_x and v_y are constrained by a linear relationship: at each point, knowing the values of di/dt and G allows the specification of a family of possible values of v. A clustering approach [11] may be used to determine the dominant velocities in the scene and to classify each image point based on estimated speed and direction.

Significant efficiencies are possible with this technique. Unlike search and matching approaches, complexity for a single moving surface increases linearly with maximum possible speed, rather than as the square of the maximum speed. Even when multiple moving objects require several passes through the image pairs, the process is more efficient than search and matching. The use of a linear constraint relationship for possible velocities allows implementation of the clustering with a small number of purely fixed-point arithmetic operations. In addition, the algorithm may be easily decomposed into pipelined and/or parallel steps. The primary limitation of the technique is that it cannot directly analyze gradually varying disparity due to rotation or motion along the optical axis.

III. Motion and Segmentation

Moving object segmentation procedures should, as much as possible, incorporate information about both static and dynamic features of a scene. Because the local area velocity estimation technique requires no prior image partitioning, the resulting velocity map may be used as a primary source of data for motion-based segmentation. Discontinuities in image brightness are obviously also important. Not only do they give static cues to the existence of surface boundaries, but they provide contextual information important to the interpretation of the velocity maps. Velocity maps obtained from local analysis will have large areas where no effective determination of velocity can be made. Neither matching nor local area estimation techniques can measure the motion of surfaces with near uniform intensity. Local velocity information is most readily available near surface edges where significant variations in intensity are present. Thus, a relationship exists between the location of

surface boundaries (determined by static analysis) and the reliability and availability of motion map information (a product of dynamic scene analysis).

We suggest a region merger approach to integrating these two sources of segmentation information [14]. Region analysis guarantees closed boundaries while at the same time minimizing the effects of local noise. In most region-based systems, an image is first partitioned into a large number of primitive segments. An iterative process is then applied which merges smaller segments into larger and larger regions. A pair of adjacent regions is merged if selected image properties along the boundaries of the two candidate regions satisfy certain similarity constraints. Eventually, region boundaries approximate the outlines of surfaces of near uniform intensity.

This process easily extends to dynamic scenes. Region boundaries correspond to portions of the velocity map for which values are likely to be available and accurate. We can classify a region based on the nature and distribution of local velocity values along its perimeter. This classification may suggest merging visually dissimilar regions if they are adjacent and have the same overall velocity. At the same time, merges of similar regions may be prevented if those regions correspond to surfaces with distinctly different velocities. (See [15] for another example of using region classification for purposes of segmentation.)

A moving surface should show a single, dominant velocity around its border, except perhaps where the border is parallel to the direction of motion. A region in which more than one dominant velocity is present, or where a velocity is associated with a portion of its boundary but is not present in other parts of the boundary with a similar orientation, cannot be confidently labeled with a single velocity. These properties may be used to classify regions based on motion. First, the velocity map values along the region perimeter are tabulated, counting only those points at which the direction of estimated motion differs significantly from the direction of the region boundary at the point. In most cases, the region is assigned the most frequently occurring velocity along its perimeter. Some regions, however, cannot be accurately classified. Such regions commonly are either too small to be effectively analyzed, stationary, near occlusion boundaries, or at the center of large uniform moving surfaces. These regions are assigned an "undefined" velocity labeling if at least one of three acceptance criteria is not met. 1) The total count of dominant velocity values should exceed a specified standard to ensure that there is a sufficient basis for classification. 2) Larger regions require more velocity points to correctly evaluate motion. Consequently, the count of dominant velocity points should represent a significant portion of the total boundary points not parallel to the dominant velocity. 3) If two or more different velocities all have relatively large counts for a particular region, no single velocity assignment is possible.

Region merging is carried out in a manner similar to the system described by Brice and Fennema [14]. Two consecutive frames of an image sequence are used to compute velocity map values while one frame of the sequence is chosen to provide static contrast information. The single frame is partitioned into elementary 4-connected regions having identical gray scale and velocity map values. Pairs of adjacent regions likely to correspond to the same surface are then combined. Adjacent regions with the same (nonzero) velocities are always merged. Visually similar adjacent regions are combined only if there is little likelihood of them having different velocities. Often, the motion of regions in the interior of larger surfaces cannot be accurately estimated. Such regions are consolidated until velocity labeling is possible.

Region merging takes place in two phases. The first phase is designed to deal with the smaller regions resulting from the initial partition while the second takes advantage of the increased accuracy of velocity labeling possible with larger regions. Merger decisions are based both on measures of local contrast along the boundary separating two regions and on velocity labelings dependent on velocity map values along the complete perimeter of each region. In the first phase, visual similarity between a pair of adjacent regions is measured in a manner comparable to the "phagocyte heuristic" used by Brice and Fennema: similarity is proportional to the ratio of the length of the low contrast portion of the common boundary to the length of the perimeter of the smaller of the two regions. If a region has a nonzero velocity labeling, then it is merged with its most similar neighbor having the same label. If the region is unlabeled or has no neighbors with the same label, it is merged with its most similar neighbor if the degree of similarity exceeds a specified threshold. Any regions resulting from a merger operation are classified and assigned the appropriate velocity label. Regions are considered as candidates for merger in order of increasing perimeter until no further mergers are possible.

In the second phase, regions are again considered in order of increasing perimeter and are merged based on a contrast criteria modified by velocity labeling. To determine the boundaries of only those objects which are moving, the termination criteria are designed to depend purely on velocity classification. (Identifying boundaries of stationary objects is considerably more difficult.) Regions may only be merged if they have the same velocity label, if they are both unlabeled, or if at least one of the regions is very small. If more than one neighbor of a merger candidate satisfies one of these conditions, then the neighbor having the lowest average contrast along the common boundary is chosen. The process continues until no further mergers are possible.

This process effectively combines gray scale and motion information into the region merger process. Using both intensity and velocity for the initial partition aids in identifying low contrast portions of boundaries not visible in the input image but strong enough to indicate a moving surface. In the first phase, adjacent regions with the same (nonzero) velocity label are always merged. Unlabeled regions and regions with different labels can be merged, but only if they satisfy the similarity condition. This is reasonable, particularly for small regions, as the initial velocity assignments may not always be accurate. By the second phase, velocity estimates are sufficiently well

determined that mergers between regions with differing labels are prohibited. A real strength of the approach comes from its ability to consolidate unlabeled regions until they include a sufficient number of velocity map values to allow for accurate classification.

The process is relatively insensitive to the settings of the various parameters with the exception of the minimum number of velocity map values needed for region labeling. This threshold must be balanced to the accuracy of the local area velocity estimate. If chosen too low, early merges will be based primarily on velocity; if too large, early merges will ignore velocity. A second problem arises from the use of 4-connected regions. If diagonal motion is possible, boundary orientation must be estimated using more than just the direction of the boundary element at the point in question. Finally, the process is obviously incapable of producing an effective segmentation if surface boundaries are of such low contrast that they are not detectable using either static or dynamic analysis.

IV. Results

The scene segmentation technique described above has been successfully applied to a variety of image sequences. Three examples are presented in this paper involving the motion of from one to three distinct objects. The examples are chosen to demonstrate the effectiveness of the procedure in commonly occurring scene types not easily partitioned with a purely static analysis. Both textured and nearly uniform brightness surfaces are present. Many object boundaries are relatively indistinct. Prominent, high contrast edges sometimes divide surfaces of a single object. Despite these potential difficulties, the method produces a reasonable estimate of moving object boundaries in all the examples.

Figs. 1, 2, and 3 show the initial image pairs, the individual pixel velocity estimates determined from local analysis, and the final region boundaries produced by the segmentation process. Because the local area velocity analysis depends on intensity differences between frames, photometric accuracy (or at least repeatability) is important. As a result, all original image pairs were normalized to have the same mean and variance in order to minimize photographic distortions of intensity. Electronic sensing of an image (vidicon, image dissector, etc.) would remove the need for this step (see [11]). The original scenes were digitized as 8-bit monochrome images of varying size, averaging about 100×100 pixels. Local velocity estimates were calculated for each point in the original pair. The resulting velocity map and the first frame of the original were both subsampled at every other point of every other line to provide a reduced resolution input for the region merger process. Note that using only the first frame for contrast information introduces a bias into the system. If the frame rate is sufficiently rapid, a better approach might be to employ a three frame sequence with the first and last used to compute local motion and the middle frame used for determining contrast. Finally, the subsampled original was requantized to 4 bits to limit the regions resulting from the initial partition to a manageable number.

Fig. 1(a) shows a real outdoor scene in which a single object is moving to the left with a displacement of 3 pixels between frames. Local area analysis correctly determined the velocity and produced the velocity map shown in Fig. 1(b). (Light areas indicate points with a velocity of 3 to the left. No other nonzero velocities were found.) Note that the velocity map itself is not sufficient for segmentation purposes as no velocity assignment is made at the center of uniform brightness, moving surfaces. Fig. 1(c) shows the results of the region merging process overlayed back onto the first frame of the original sequence. While not perfect, the outline is a reasonable approximation of the boundary of the truck. The protuberance at the front of the truck is a spare tire. Although indistinct in the original image, it is correctly identified as part of the moving object.

Fig. 2(a) shows two objects moving toward one another, each at about 3 pixels per frame. The example was chosen because it characterized a situation in which objects are not clearly separated from their backgrounds and the objects themselves are composed of a number of visually dissimilar surfaces. Object velocities were again correctly determined. Fig. 2(b) and (c) show the local area velocity estimation and the final segmentation. (In Figs. 2(b) and 3(b), each velocity is assigned a different gray scale value.) A number of errors occur, particularly in the left object. The boundaries which are missed, however, are those which are neither apparent as contrast edges in the original nor as motion discontinuities in the velocity map. A more accurate analysis would require either higher level knowledge, observations over more frames, or a presumption that nonadjacent regions with the same velocity should be linked.

Fig. 3 shows three objects moving in different directions towards the center of the scene. The upper left object moves 1.5 pixels between frames, the upper right object moves 3 pixels, and the lower object moves 2 pixels. Direction was correctly determined with speeds estimated at 2, 3, and 3 pixels per frame, respectively. In this example, the objects are somewhat better differentiated from their background. On the other hand, each of the objects is made up of several visually distinct surfaces. Fig. 3(b) and (c) show the local velocity estimates and segmentation. As can be seen, all moving objects are located with reasonable accuracy. In particular, the method is effective in linking adjacent regions with the same velocity but separated by very high contrast gray scale edges. To demonstrate the difficulty of segmenting this image pair using just differencing techniques, Fig. 4(a) shows those points in Fig. 3(a) which differ by at least the ΔI threshold used for local area velocity estimation. Fig. 4(b) shows a similar difference picture except that both input frames were first smoothed by the same blurring function as used for velocity mapping. While these difference pictures localize portions of the moving surfaces, they are not sufficient by themselves as the basis for segmentation.

V. Conclusions

The estimation of boundaries of moving objects in a scene sequence is an important part of the analysis of time-varying imagery. Often, these boundaries are difficult to locate in a single frame. Thus, boundary detection should not depend on extensive static analysis. The examples above demonstrate an

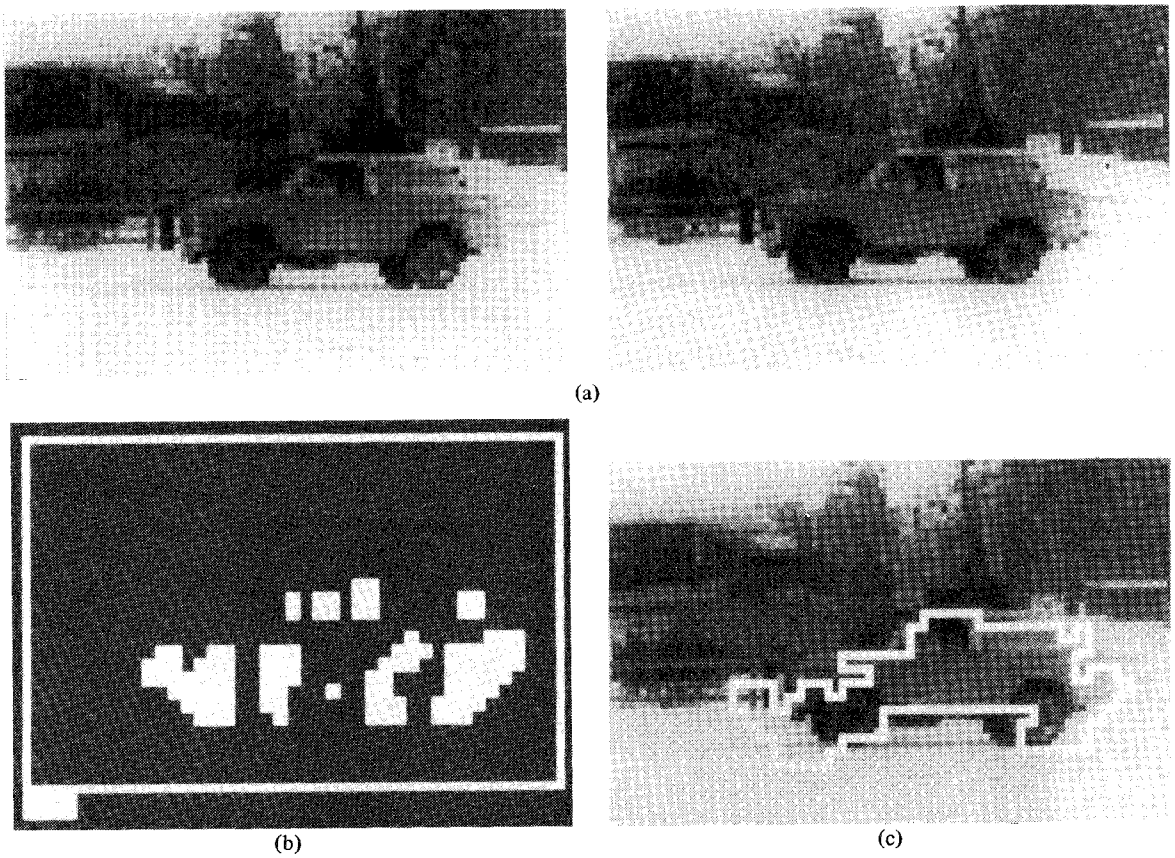


Fig. 1. (a) Single moving object. (b) Velocity map. (c) Segmentation.

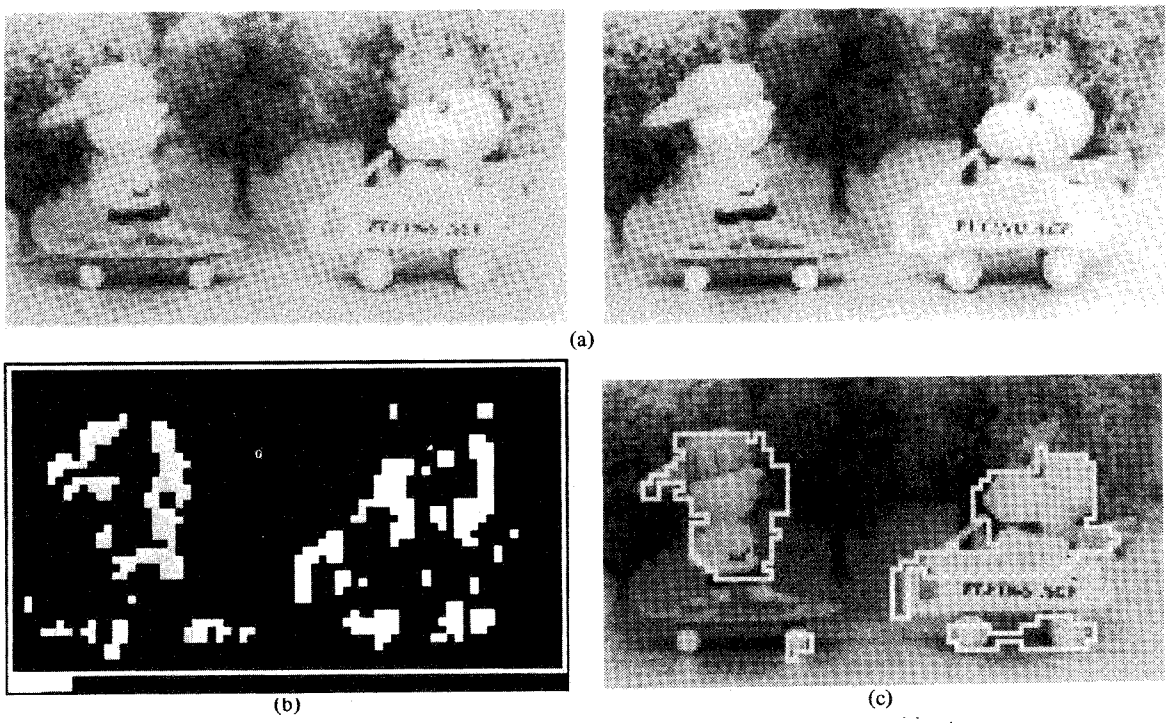
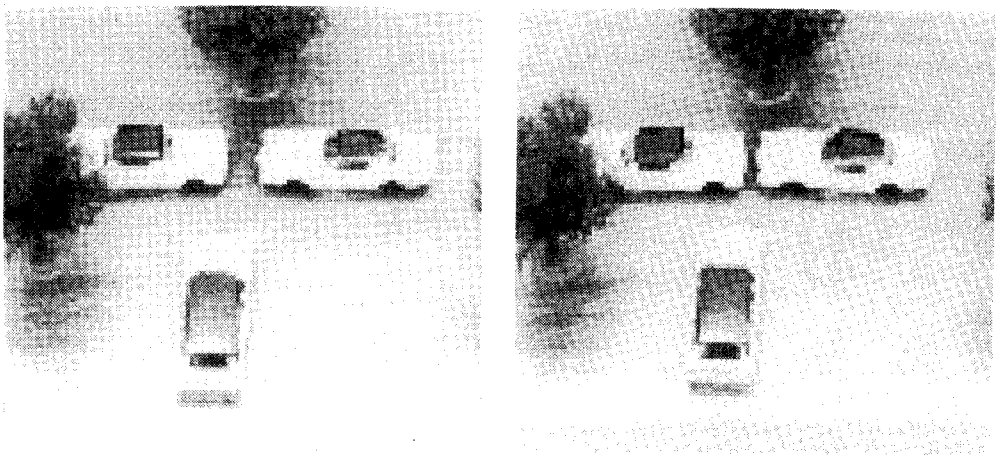


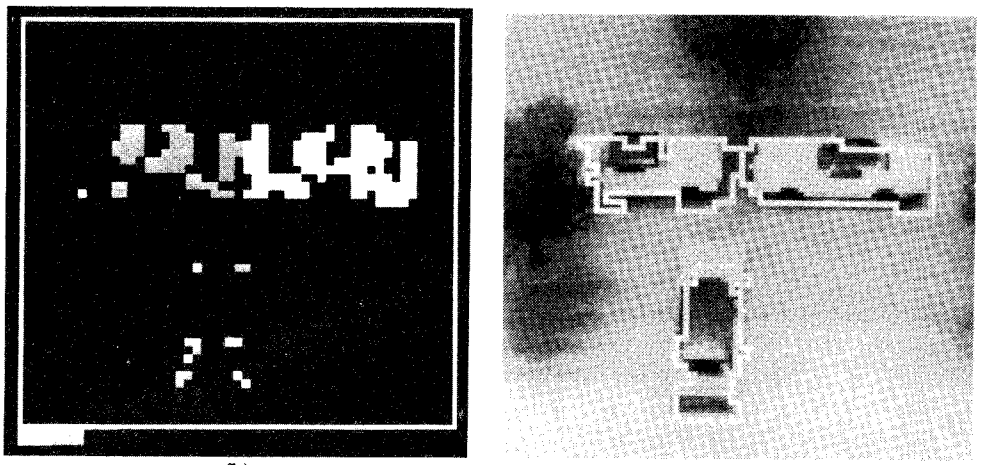
Fig. 2. (a) Two moving objects (toy problem). (b) Velocity map. (c) Segmentation.

effective technique for developing an initial segmentation of a dynamic scene by incorporating both static and dynamic information into the segmentation process.

Velocity information is extracted prior to any segmentation by a technique which relates spatial gradient to intensity change over time. The procedure is computationally efficient and provides an accurate estimation of the translational motion in a scene, even when several moving objects with different ve-

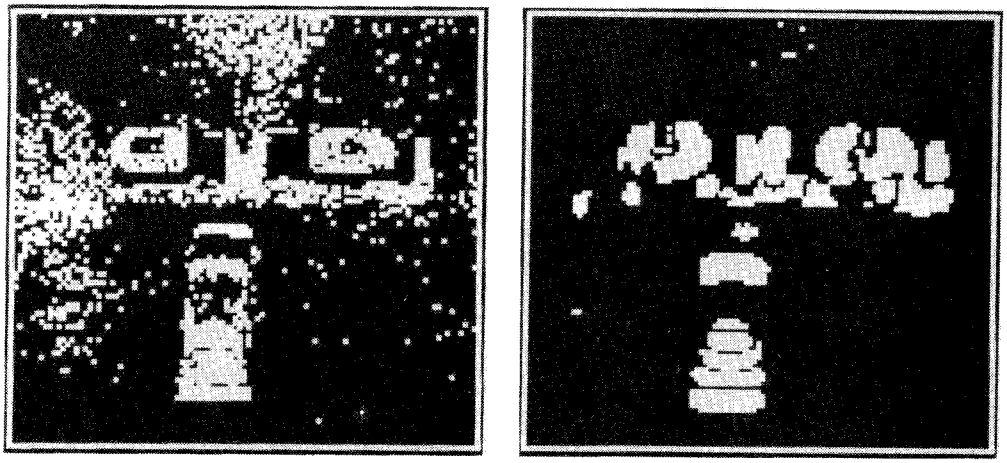


(a)



(b) (c)

Fig. 3. (a) Three moving objects. (b) Velocity map. (c) Segmentation.



(a) (b)

Fig. 4. (a) Above threshold differences in Fig. 3(a). (b) Differences in Fig. 3(a) with preblurring.

locities are present. In addition, a velocity estimate for each local area in the scene is produced. Because these estimates do not depend on any boundary analysis, they may be used as an independent source of information for scene segmentation.

Local area velocity estimation alone is not a sufficient basis for segmentation. Such analysis cannot effectively deal with the motion of uniform brightness surfaces or the translation of surface boundaries oriented parallel to the direction of motion. However, by combining velocity information with more traditional contrast based boundary cues, effective segmentation of

moving objects is possible. In the system described above, motion analysis is incorporated into a region merging procedure. A region-based approach is particularly appropriate because local velocity information is in general most reliably available at the perimeter of regions corresponding to surface boundaries. First, both local velocity estimates and brightness are used to establish an initial partition. The simultaneous use of velocity and gray scale decreases the possibilities of a low contrast boundary being missed. Next, regions are merged based on a combined brightness and motion criteria. The final result is a set of regions corresponding to connected areas of the scene with common velocities.

Limitations of the approach include the computational inefficiencies of region merging and the restriction to translational motion, although rotations may be handled by a more sophisticated (but likely less efficient) local area velocity estimator. Advantages include: straightforward incorporation of motion information into an existing, well understood segmentation procedure; need for only two consecutive frames of an image sequence; no need for prior segmentation; and ability to deal with images which are difficult to segment by traditional, static analysis. Most importantly, the technique described in this paper demonstrates that motion and contrast information may be effectively combined at the lowest levels of scene analysis.

Acknowledgment

R. Hummel contributed a number of important suggestions leading to the efficient implementation of the local area velocity estimator. C. Lemche and S. Barnard provided additional useful suggestions and criticism. W. Franta and the staff of the Special Interactive Computing Laboratory of the University of Minnesota supplied excellent facilities for undertaking this research.

References

[1] R. L. Lillestrand, "Techniques for change detection," *IEEE Trans. Comput.*, vol. C-21, pp. 654–659, July 1972.

[2] R. Jain and H.-H. Nagel, "On the analysis of accumulative difference pictures from image sequences of real world scenes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-1, pp. 206–214, Apr. 1979.

[3] R. Jain, W. N. Martin, and J. K. Aggarwal, "Segmentation through the detection of changes due to motion," *Comput. Graphics Image Processing*, vol. 11, pp. 13–34, Sept. 1979.

[4] J. A. Leese, C. S. Novak, and V. R. Taylor, "The detection of cloud pattern motions from geosynchronous satellite image data," *Pattern Recognition*, vol. 2, pp. 279–292, Dec. 1970.

[5] H. P. Moravec, "Towards automatic visual obstacle avoidance," in *Proc. 5th Int. Joint Conf. Artificial Intell.*, Aug. 1977, p. 584.

[6] S. T. Barnard and W. B. Thompson, "Disparity analysis of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, pp. 333–340, July 1980.

[7] J. L. Potter, "Scene segmentation using motion information," *Comput. Graphics Image Processing*, vol. 6, pp. 558–581, Dec. 1977.

[8] J. K. Aggarwal and R. O. Duda, "Computer analysis of moving polygonal images," *IEEE Trans. Comput.*, vol. C-24, pp. 966–976, Oct. 1975.

[9] W. K. Chow and J. K. Aggarwal, "Computer analysis of planar curvilinear moving images," *IEEE Trans. Comput.*, vol. C-26, pp. 179–185, Feb. 1977.

[10] N. Badler, "A concept model for the description of image sequences," in *Proc. Milwaukee Symp. Automat. Comput. Contr.*, Apr. 1976, pp. 377–381.

[11] C. L. Fennema and W. B. Thompson, "Velocity determination in scenes containing several moving objects," *Comput. Graphics Image Processing*, vol. 9, pp. 301–315, Apr. 1979.

[12] J. O. Limb and J. A. Murphy, "Estimating the velocity of moving images in television signals," *Comput. Graphics Image Processing*, vol. 4, pp. 311–327, Dec. 1975.

[13] C. Cafforio and F. Rocca, "Methods for measuring small displacements of television images," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 573–579, Sept. 1976.

[14] C. R. Brice and C. L. Fennema, "Scene analysis using regions," *Artificial Intell.*, vol. 1, pp. 205–226, 1970.

[15] J. M. Tennenbaum and H. G. Barrow, "Experiments in interpretation guided segmentation," *Artificial Intell.*, vol. 8, pp. 241–274, June 1977.

William B. Thompson (S'72–M'74), for a photograph and biography, see p. 340 of the July 1980 issue of this TRANSACTIONS.