

THREE DIMENSIONAL MOVEMENT ANALYSIS OF DYNAMIC LINE IMAGES

Saburo Tsuji, Michiharu Osada, and Masahiko Yachida
Department of Control Engineering
Osaka University
Toyonaka, Osaka 560, Japan

As a part of a cine-film understanding system, this paper presents a dynamic scene analyzer which separates moving objects from the background and analyzes their motion patterns in dynamic line images such as cartoon films. Finding correspondence of regions and their segments in a sequence of input frames by a new flexible template matching method, the analyzer tracks moving regions and their segments in the dynamic images. Next, a similarity test of segment movements detects the background movement, and classifies the regions into stationary and non-stationary ones. Each in the latter group is further labeled as partly occluded, false, or moving, by examining motion patterns of its segments. Finally, the analyzer merges segments of each moving object into groups having similar motion patterns in order to obtain a meaningful partition corresponding to its components such as hands or legs.

1. INTRODUCTION

In recent years, people have studied dynamic scene analysis, which analyzes a sequence of pictures taken at a constant time interval, such as a cine film or a video record, to interpret it into meaningful descriptions on constituents of the scene and their temporal changes, as well [1,2]. Many papers have been published on detection and tracking of moving objects with [3,4] or without [5] their models in temporal records of natural scenes, interpretation of a sequence of silhouetted images into a model of static and moving constituents [6], and analysis of dynamic line images by utilizing a camera model [7] or by utilizing procedural knowledge about an actor and static objects to describe the dynamic image in terms of meaningful actions [8].

Most systems utilize a priori knowledge and models of the constituents which are very useful to efficiently analyze a large number of consecutive frames, and to solve occlusion problems. They were, however, designed to accept only limited spatio-temporal patterns as inputs; the dynamic image must satisfy both or one of strong restrictions that (1) the constituents show two dimensional movements, and (2) their shapes are almost rigid.

This paper presents a more general motion analysis system that analyzes a sequence of line images, a cartoon film, in which a personified animal moves and rotates freely in a three dimensional world. The analysis is difficult because of the following reasons: (1) Since the line image of the object consists of curved components, it seems difficult to find strong local features which could be utilized as useful cues for motion analysis or segmentation of the moving object into meaningful parts such as its arms or legs. (2) Since the moving object and its parts move and rotate in the three dimensional world, and since

occlusion often occurs, the shape, size, and structure of each constituent change from frame to frame.

(3) The background or stationary objects may move in the image because of camera movements to track interesting objects.

An approach to overcome the above-mentioned difficulties is to develop a scene analyzer which utilizes effectively temporal patterns. The proposed system has been designed along this approach; knowledge about the shape and structure of each object in the scene, by which usual computer vision systems interpret static scenes, is not utilized, but similarities of both spatial and temporal patterns are iteratively examined with an aim of extracting useful information or cues for higher level analysis of the dynamic scene. In other words, the analyzer is a subsystem of a film-understanding system [8], and its task is to divide the scene into meaningful parts; for example, it separates moving objects from the background, and also subdivides a moving object into components showing similar movement patterns. One can arrange the results into a structural description of the detected objects on properties, spatial relations, and motion patterns.

2. SEGMENTATION OF SCENE BY MOTION ANALYSIS

2.1 INPUT IMAGE

Two animation films containing about 100 frames, in which a personified animal (Moomin) moves freely in a three dimensional world, are used to test the performance of the system. Fig.1 displays binary pictures of sampled frames in Film 1; the animal runs from the right side of the picture and then jumps over a stone. Another film, Film 2 shown in Fig.2, is one for testing the effect of the camera movement. We perceive it as if the animal moves

leftwards while it moves its hands and legs but its location is almost unchanged, because objects, trees, in the background translate rightwards at an almost constant velocity.

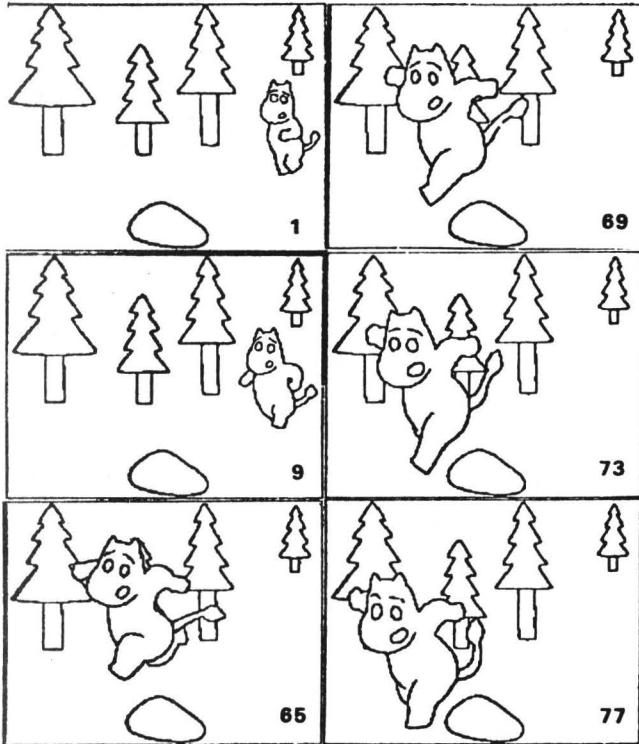


Fig.1 Sample Frames of Film 1.

An animal runs from the left side, jumps over a stone.

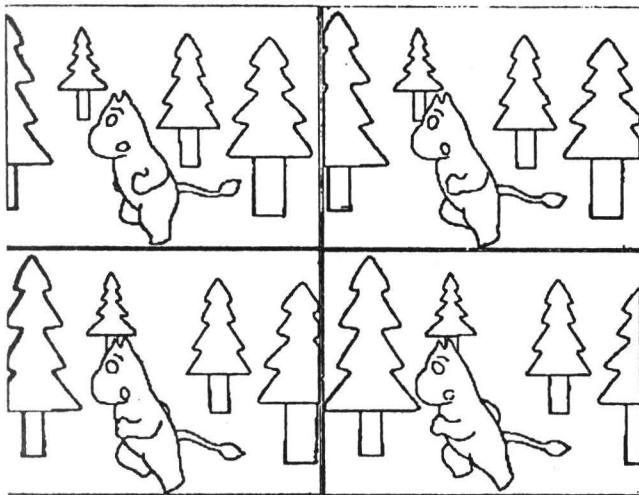


Fig.2 Sample Frames of Film 2.

Although the location of the animal changes little, it is perceived as it runs leftwards, because the objects in the background move to rightwards.

If a computer observes Film 1 from its first frame, where the animal is separated from the other objects, it can easily create a model of the scene in which the segmentation of each constituent is established. Then, the successive frames will be analyzed by following moving constituents with the aid of the scene model. However, we try to analyze a more complex case; a frame in which occlusion has already occurred, for example the 65th frame, is selected as the first input picture to the computer, and the successive frames are then analyzed without the already segmented model of the scene.

A TV camera is used as an input device of the system, and a digitized (200 by 200 4bit) image of each frame is converted into a binary picture by thresholding. Applying a simple thinning operator, we track every arc in the picture, and arrange the line image into lists of arcs; an arc is described in terms of end points, connected arcs, and chain coded pixels on it. Illustrated in Fig.3 is a graphical representation of the arc list of the 65th frame of Film 1.

2.2 SEGMENTATION OF SCENE BY MOTION ANALYSIS

Our system is a lower level motion analyzer which segments scenes into their constituents by examining their motion patterns, and interprets them into moving, stationary, or partly occluded objects. Movement patterns of short segments of a moving object are also analyzed in order to subdivide the object into meaningful moving parts such as hands or legs.

Most motion analyzers extract movements of the objects in consecutive frames, but they do not analyze how different parts of a moving object show different motion patterns. Our system is characterized by extraction of motion patterns of sufficiently short segments, and also by tests of similarities of the segment movements. The analysis is divided into the following phases.

(1) Interframe comparison of segments

At first, we detect regions in the scene and establish their correspondence to those of the next frame. Then, we divide all arcs into sufficiently short segments, and follow them in the line image sequence to measure how they are moving. A flexible template matching method has been developed to find correspondence between the segments in the consecutive frames. The movement patterns of the segments are examined in the next three phases.

(2) Separation of non-stationary regions from the background

The objects in the background may also move in the image because of the camera movements, therefore we need to know which objects are really moving. Detecting the background movement as that of the majority of the segments in the scene, we classify the regions in the scene into stationary and non-stationary ones.

(3) Labeling non-stationary regions

The movements of the segments in the non-stationary regions are examined to classify them into partly occluded objects (for example, Region ABCD in Fig.3), moving objects, or false objects that are areas surrounded by moving objects and static objects (Region EFGHIJE in Fig.3). The results are used to create and update a scene model. This labeling is not always correct, and the analysis is backtracked when the analysis of the later frames finds the wrong labeling in the scene model.

(4) Grouping of segments in a moving object

Finally, the similarities of the segment movements in each moving object are analyzed in order to divide the object into meaningful parts.

We have implemented the system on a mincomputer HP2108A with 256 kbyte buffer memory. The program is written in FLISP, a modified version of INTER LISP, which can control lower level programs written in FORTRAN.

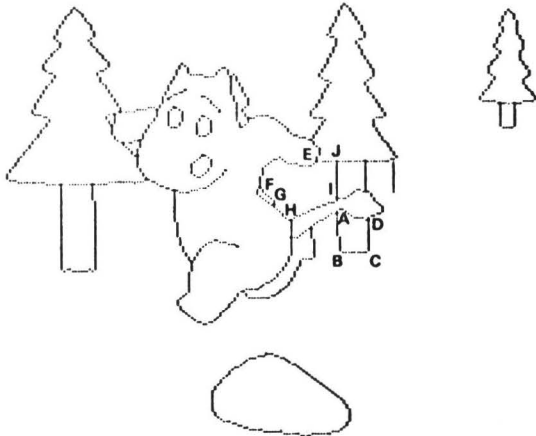


Fig.3 Graphical Representation of Arc Lists of the 65th Frame of Film 1.

3. INTERFRAME COMPARISON OF SEGMENTS

3.1 CORRESPONDENCE OF REGIONS

As mentioned in the previous section, our motion analyzer follows the movements of short segments in the sequence of line images with the aim of obtaining an interpretable goal partition of the dynamic scene. At first, the system examines correspondence of regions in the consecutive frames, and then it analyzes that of each short segment in them.

We can easily divide the input scene into regions, and obtain a list of regions for each frame, in which a region is described in terms of a location of its centroid, its size (perimeter), and a list of arcs (its boundary and internal edges). Now, we examine correspondence of regions in the sequence of frames by a simple template matching method, because changes of the line images between two consecutive frames are small.

Let R be a region in the current frame. At first, we search the region list of the next frame for a small number of regions, one of which possibly corresponds to R , such that both locations of their centroids and their perimeters are almost same as those of R . Then, the boundary of R is used as a template, and we find the best matched one of the selected regions by the following method.

At first, a distance map of the template, an array of numbers representing distance to the nearest point on the template, is created in order to easily compute a distance from the template to a region boundary by simply summing distance values in the map at every point on the boundary and then normalizing by the boundary length [9]. By a hill climbing method, we search for the best matched location of each selected region boundary to the template, and a region having the minimum distance value is selected. If the distance value is less than a threshold, we consider it as corresponding to R . Eliminating the detected pair from the region lists, we iteratively apply the method to find correspondence of all regions. The results are successful, unless changes in the structure of the line image such as appearance of a new region do not occur between two frames. The unmatched regions in this case are also recorded for higher level motion analysis.

3.2 TRACKING OF SEGMENTS IN CONSECUTIVE FRAMES

Let us consider how we can find correspondence of short segments in two regions whose correspondence has been established. Since the line image of the animal consists of curved segments, we cannot find useful singular points in it, such as true vertices in a polygonal world [6], by which we can easily establish their true correspondence between two frames. Although T joints give us useful information about segmentation between objects, two T joints in different frames not always correspond each other, because they belong to a category of *false vertices* in the polygon world.

Another difficulty for examining the correspondence of the segments is that the shape of the moving object changes from frame to frame. The body of the animal moves and rotates in the three dimensional world, and its moving parts such as its hands and legs independently rotate about different axes. Therefore, the template matching method, which assumes shape rigidity, is not adequate for examining correspondence of each portion of the object.

A flexible template matching method called as *rubber mask* technique [10] was developed for matching of a distorted pattern to its template. The method, however, is not applicable to our problem because it assumes a priori knowledge about the structures of patterns which are built in the matching program.

Therefore, we have developed a new flexible template matching method for finding correspondence of each

short segment in a line pattern to that in a partly distorted one. Consider a problem to match each part of Pattern 1 shown in Fig.4 (a) to Pattern 2 in Fig.4 (b). Because of the shape distortion, we cannot establish a complete or almost complete matching of the two patterns. However, we can establish a good match of a considerable long arc of Pattern 1 to a part of Pattern 2 as shown in Fig.4 (c). Also we can match the other arcs by shifting the template (Fig.4(d),(e),(f)). Thus, we can consider points A,B,C,D, which belongs to the matched portions in the successive partial matching process, as corresponding to A', B', C', and D', respectively.

Our method utilizes this matching process. Let R and R' be two regions which consist of sets of arcs $\{a_i\}$ and $\{a'_i\}$ respectively. At first, we search $\{a_i\}$ and $\{a'_i\}$ for the best matched pair of considerably long arcs a_k and a'_m by a similar method as that for matching regions. Next, a_k is divided into segments $l_1, l_2, \dots, l_n$ shorter than a predetermined threshold value.

If one cannot match a_k precisely to a'_m , we shift each segment by a small amount to obtain a better matching to a'_m and also its connected arcs as shown in Fig.5. As a result, a point which is a common terminal point of two segments in a_k , for example point 2 in Fig.5, is mapped on two different points in a'_m . Therefore, the correspondence of a segment is described with the transition of a representative point, a point located at near its center. After this type of matching process is accomplished for the selected arc, then its connected arcs are also precisely matched by the same procedure. Thus, we get a correspondence table of R to R' on the segments. We iterate the above-mentioned procedure to find correspondence of the successive frames, and representative points of segments in the first input frame are tracked by interpolation of those in the later frames. Results of tracking of segments in a moving region and a partly occluded objects from the 65th frame to the 70th frame in Film 1 are illustrated in Fig.6.

4. SEPARATION OF MOVING REGIONS FROM BACKGROUND

4.1 DETECTION OF BACKGROUND MOVEMENT

Stationary objects in the background may move in the image because of camera movements such as panning, tilting, and zooming, therefore, the system needs to detect how the background is moving in consecutive images. We consider that there are many more segments in the background than in moving objects, and they show almost a similar movement pattern; a horizontal or vertical movement by panning or tilting, and a radial movement by zooming.

Therefore, histograms of the segment movements will give us valuable information about the background movement. Fig.7 illustrates two histograms that frequencies of segment movement

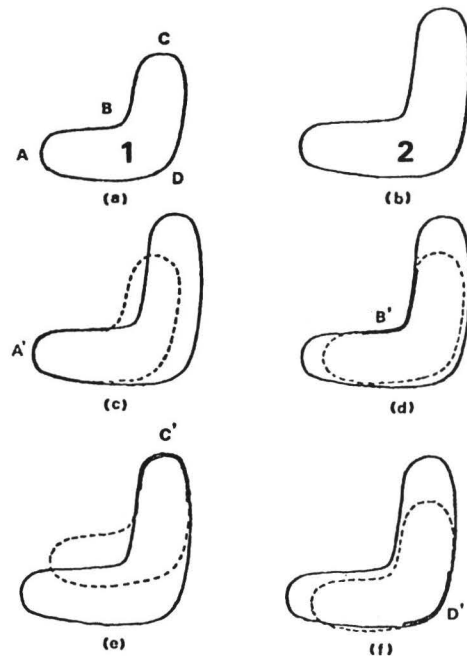


Fig.4 Flexible Template Matching.

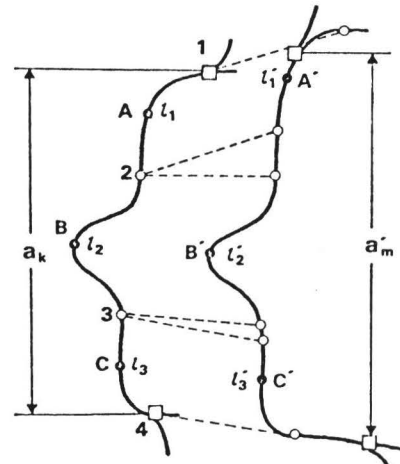


Fig.5 Correspondence of Segments.

Arc a_k is matched to Arc a'_m . a_k is divided into segments l_1, l_2 and l_3 , and the best matched location of each segment to a'_m and its connected arcs is detected by a hill climbing method. As the result, correspondence of A, B, C, to A', B', C' is established.

vectors in Film 1 and Film 2 are mapped on a X-Y plane. The sharp peak at (0,0) in Fig.7 (a) shows that the background of Film 1 is still, and the peaks of Fig.4 (b) suggest that Film 2 was taken while the camera was panning to track the moving object.

Subtracting the background movement from the movement vectors of all segments, we separate non-stationary regions, as those containing moving segments, from stationary ones.

4.2 LABELING OF NONSTATIONARY REGIONS

The non-stationary regions are not always parts of moving objects. Partly occluded regions of stationary objects by moving objects and regions surrounded by moving objects and stationary objects are also registered as non-stationary by the above-mentioned separation process.

Therefore, we create a scene model after analyzing several consecutive frames. The model is a collection of descriptions of the regions on their movements, and we label a non-stationary region as moving if most of their segments are moving, and as partly occluded if a considerable long sequence of segments is not moving (a region shown in Fig.6 (b)). Next, successive frames are sequentially analyzed, and labels of corresponding regions are examined. If the structure of the line image changes between two frames, for example a new region appears, the system marks the frame in the model for higher level analysis. The system, however, manages a simple case that a partly hidden region is merged into the empty background as Region EFGHIJE in Fig.3 which disappears at the 69th frame. It is considered as a false region, and the analyzer backtracks to correct the wrong label.

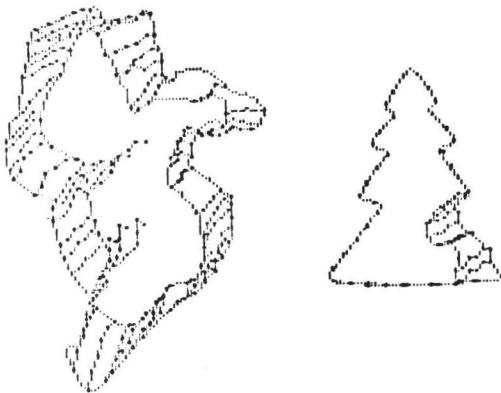


Fig.6 Tracking of Segments.

5. GROUPING OF SEGMENTS IN A MOVING OBJECT

The final process is to segment a moving region into meaningful parts showing similar movements. Since the movement vectors of all segments in each moving region have been analyzed, we can divide them into groups by a similarity test of their displacements. At first, we subtract a displacement vector of the centroid of the moving region from the movement vector of every segment in it. Fig.8 shows an example of the relative movement vectors displayed on the figure of the moving region. One can find a strong movement similarity for each group of segments such as its leg, body, and hand, therefore its segmentation is easy. We iterate comparing of each two adjacent vectors and merging them into a group

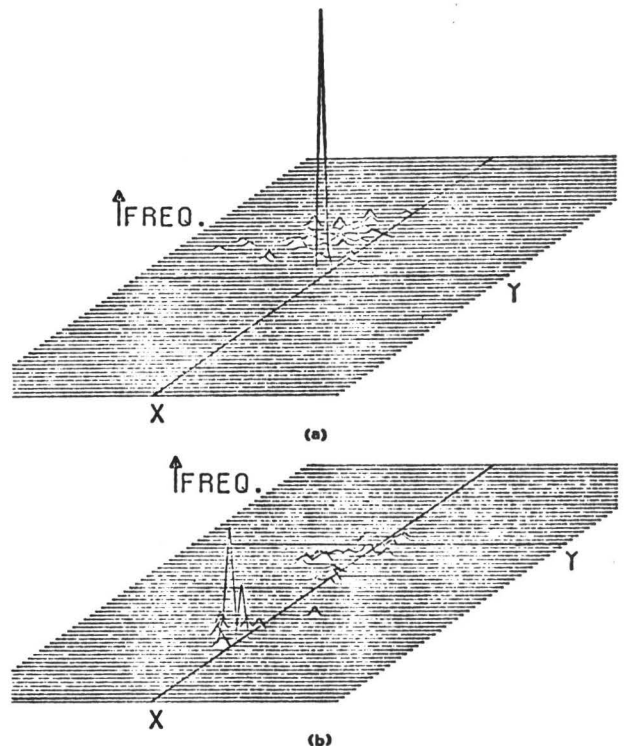


Fig.7 Detection of Background Movement.

A sharp peak in (a) shows the background is still, and peaks on X axis in (b) indicate that the camera was panning when the film was taken.

if the difference of two vectors is smaller than a threshold. Fig.8 (c) shows the result of segmentation. Next, each set of segments between two groups is divided into two parts by a test of the movement similarity. Thus, segmentation of the moving object into its hand, leg, body, and head is accomplished.

6. DISCUSSION

We have developed a motion analyzer which separates non-stationary regions from stationary ones and subdivides the moving region into meaningful components on the basis of segment movements. Usefulness of a flexible template matching method for establishing correspondence of segments in consecutive frames is shown. Similarity tests of segment movements are also useful to detect the background movement and to segment a moving region into its moving components.

However, the scene model used is very simple, and the acquired knowledge about the properties of the regions by analyzing the previous frames such that each region has been occluding (or occluded by) what regions or not, region is not used to analyze a new coming frame. We could easily implement this

function, an essential function for dynamic scene analysis, to our system, and we would be able to save much computing time.

Also many important problems such as (1) analysis of the structural change of the line image, (2) semantic interpretation of moving and stationary regions, and (3) understanding meaning of the movements, are left for future research.

REFERENCE

- 1] W. N. Martin and J. K. Aggarwal, "Dynamic Scene Analysis" *Computer Graphics and Image Processing*, Vol.7, pp. 356-374, 1978.
- 2] H. H. Nagel, "Analysis for Image Sequences" *Proc. of 4th Int. Joint Conf. Pattern Recognition*, pp. 186-211, 1978.
- 3] M. Yachida, M. Asada and S. Tsuji, "Automatic Motion Analysis System of Moving Objects from the Records of Natural Processes" *ibid.*, pp. 7 1978.
- 4] Y. Ariki, T. Kanade and Y. Sakai, "An Interactive Image Modeling and Tracing System for Moving Pictures" *ibid.*, pp. 681-685, 1978.

5] R. Jain, D. Miltzer and H. H. Nagel, "Separating Non-stationary from Stationary Scene Components in a Sequence of Real World TV-Images" *Proc. 5th Int. Joint Conf. Artificial Intelligence*, pp. 612-618, 1975.

6] J. K. Aggarwal and R. O. Duda, "Computer Analysis of Moving Polygonal Images" *IEEE Trans. Comput.*, Vol. C-24, pp. 966-976, 1975.

7] N. Badler, "Three-Dimensional Motion from Two-Dimensional Picture Sequences," *Proc. of 2nd Int. Joint Conf. Pattern Recognition*, pp. 157-161, 1974.

8] S. Tsuji, A. Morizono and S. Kuroda, "Understanding a Simple Cartoon Film by a Computer Vision System" *Proc. 5th Int. Joint Conf. Artificial Intelligence*, pp. 609-610, 1977.

9] H. G. Barrow, J. M. Tennenbaum, R. C. Bolles and H. C. Wolf, "Parametric Correspondence and Chamfer Matching: Two New Techniques for Image Matching," *Proc. 5th Int. Joint Conf. Artificial Intelligence*, pp. 659-663, 1977.

10] B. Widrow, "The 'Rubber Mask' Technique - I. Pattern Measurement and Analysis" *Pattern Recognition*, Vol.5, pp. 175-198, 1973.

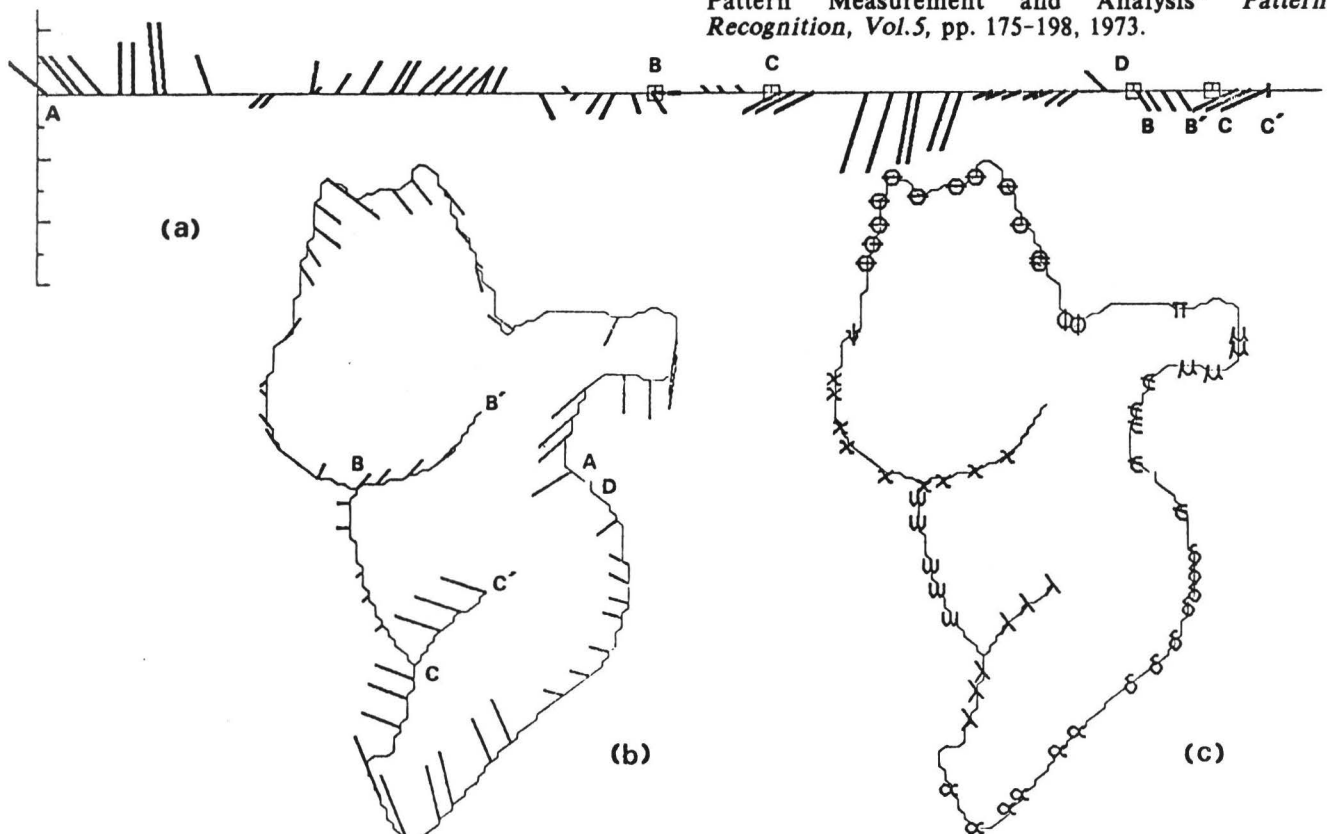


Fig.8 Grouping of Segments by Similarity of Segments Movements.

Segment movement vectors of a moving region from the 65th frame to the 70th frame relative to the centroid of the region are displayed (a) along the arcs and (b) on the image at the 65th frame. The segmentation result is shown in (c).