



Title	UNDERSTANDING OF THREE-DIMENSIONAL MOTIONS IN TRIHEDRAL WORLD
Author(s)	浅田, 稔
Citation	大阪大学, 1982, 博士論文
Version Type	VoR
URL	https://hdl.handle.net/11094/2813
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

January 1982

UNDERSTANDING OF THREE-DIMENSIONAL
MOTIONS IN TRIHEDRAL WORLD

Minoru Asada

Department of Control Engineering
Osaka University

This work is supported by a Grand-in-Aid for Scientific Research from the Ministry of Education, Culture and Science, Japanese Government and submitted to the Department of Control Engineering, Osaka University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy, January 1982.

UNDERSTANDING OF 3D MOTIONS

ACKNOWLEDGEMENT

I offer my thanks to the following people who contributed to this thesis.

to Prof. Saburo Tsuji who supervised, advised and encouraged me in this work,

to Dr. Masahiko Yachida who contributed many ideas and helpful suggestions to me,

to Prof. Takeshi Kasai and Dr. Norihiro Abe who served as readers and gave me helpful and valuable comments,

to Prof. Yoshifumi Sakurai, Prof. Kimisuke Shirae, Prof. Kokichi Tanaka and Prof. Ryoji Suzuki for their constructive readings of this thesis,

to members of Prof. Tsuji's Laboratory for their supports in various kinds of things.

Finally, I'd like to thank to my wife and two my sons for their patiences and supports.

ABSTRACT

This paper describes how a computer can understand three-dimensional motions of blocks. Input is a sequence of line drawings of rigid trihedral blocks projected orthogonally on an image plane. The block can show complex coincidental motions such that a block rotates around a joint attached on one side of another moving block.

At first, a labeling scheme and an object-to-object matching method are applied to the image sequence to segment the images into individual blocks and find correspondences of their vertices between frames. A transition table of junction labels and contextual information are used to analyze structural changes of the line drawings.

Secondly, the shape rigidity property of three vertices on a block is used to evaluate geometrical parameters, such as orientations and edge lengths, then their interframe rotational movements are determined.

In order to analyze coincidental motions, the computer finds a main body in a jointly moving group such that motions of the other objects viewed from it can be simply represented, and constructs hierarchical scene representations.

Finally, consistent properties of the translational movements through multiple frames are found by a hypothesis-and-test method, and the system segments the image sequence into several parts which are described with different consistent properties.

Thus, the dynamic scene is segmented into the simple and natural representation similar to what a human would perceive from the given input.

TABLE OF CONTENTS

	page
ACKNOWLEDGEMENTS	i
ABSTRACT	ii
TABLE OF CONTENTS	iii
1. INTRODUCTION	1
1.1 Research of Computer Vision	1
1.2 Analysis of Time-Varying Images	2
1.3 Goals of our system	3
2. ANALYSIS OF IMAGE SEQUENCE IN TRIHEDRAL WORLD	5
2.1 Input Images	5
2.2 Motion Understanding	5
2.3 Analysis and Representation of Dynamic Line Images	9
3. STRUCTURAL ANALYSIS OF LINE IMAGES IN TRIHEDRAL WORLD	12
3.1 Junction Labelling in Dynamic Scene	12
3.2 Matching Objects between Frames	13
3.3 Label Changes by Rotation of Single Object	13
3.4 Changes in Structure of Line Images Containing Plural Objects	21
3.5 Analysis and Interpretation of Non-trihedral Vertices	24
3.6 Examples of Segmentation and Matching	27
4. MEASUREMENT OF INTERFRAME CHANGE	31
4.1 Reconstruction of 3-D Geometry	31
4.2 Measurement of Interframe Rotational Components	34
5. UNDERSTANDING OF 3-D MOTIONS	39
5.1 Hierarchical Representation for Articulated Motions	39
5.2 Motion Descriptor and Consistent Property	42
5.2.1 Representation of 2-D motion	42
5.2.2 Reconstruction of 3-D motion	46
5.2.3 Motion representation with consistent property	51

5.3 Experimental Results	53
5.3.1 Motion representations of main body and subpars	53
5.3.2 Segmentation of image sequence by motion	57
6. DISCUSSION	62
6.1 Problems in Each Phase	62
6.2 Future Research	62
7. CONCLUSIONS	66
APPENDIX Precise Estimation of Z-co of Vertices	68
BIBLIOGRAPHY	70

1. INTRODUCTION

1.1 RESEARCH OF COMPUTER VISION

Computer vision which reconstructs three-dimensional scene descriptions from two-dimensional pictures has been extensively studied since the classical paper by Roberts [Roberts-68] was published. Real world scenes, however, can be extremely complex, then many people have focused their efforts on interpretations of line drawings of the blocks world, a mini-world composed of opaque polyhedra, to find general principles of the scene understanding; for example, heuristics for decomposing scenes into solids [Guzman-68], assigning meanings to lines and junctions [Huffman-71 and Clowes-71], utilization of constraints to understand complex scenes with shadows [Waltz-75], and analysis in dual spaces [Mackworth-71 and Huffman-77].

Now, to cope with real scenes, many researchers have been studying how physical constraints can help the computer to determine 3-D properties of scene from 2-D imagery. For example, "Shape from Shading" [Horn-75, Ikeuchi-81 etc] which obtains a surface gradient of each point by using relations among light intensities on an image plane, light directions and the reflectance on objects,

"Texture Gradient" [Witkin-81 etc] which also obtains the surface gradient of each small area by examining how each pattern on the object can be deformed by observing from a certain view point when the object has homogeneous patterns (texture) on its surface,

"Surface Contour" [Barrow-81 etc] which tries to interpret line drawings as three-dimensional surfaces, based on constraints on local surface orientation along extremal and discontinuity boundaries,

"Structure from motion" [Ullman-79] which interprets dynamic images as motions of rigid objects and reconstructs their three-dimensional structure from multiple images, and so on.

Applying each method, we can obtain the 3-D properties of a local area from 2-D imagery, such as the 3-D coordinates of a

point, the edge direction, surface gradient and so on. The results of these analysis are called "*needle map*" by Horn, "*intrinsic images*" by Tenenbaum and "*2-1/2 sketch*" by Marr [Marr-78]. The results, however, have not been integrated to a 3-D model of object, which is often called "*Object-centered model*" by Marr and, for example, it can be instantiated by "*Generalized cylinders*" [Agin-76]. This problem is too difficult and inherently concerns on the matter how we can recognize an object in a 3-D world from a retinal image. Particularly, in the case of "*Structure from Motion*" which deals with time-varying images, the computer needs not only to reconstruct the 3-D model of objects but also to integrate the spatial properties, such as positions and orientations of objects in every frame, into their motion representation. Few systems, however, have dealt with the latter problem.

In this paper we propose a new method by which a computer understands 3-D motions from time-varying images.

1.2 ANALYSIS OF TIME-VARYING IMAGES

We, human beings, can easily recognize the 3-D world surrounding us through our eyes. It seems that we process a large amount of information from our retinal images which are dynamically changed by movements of objects and/or our eyes, rather than from a static image like a picture. Then, in recent years, many people have directed their attention to a new field of the computer vision, the analysis of time-varying images [Martin-78, Nagel-78]. Most of them, however, involve with problems of interpreting changes in the image sequences as two-dimensional movements; for example, movement detection and segmentation [Jain-79, Thompson-80 etc], tracking of particle-like objects [Greaves-75, Yachida-78 etc] or polygons [Aggarwal-75], and analysis of dynamic behavior of heart walls [Kaneko-73, Yachida-80 etc].

An important theoretical basis for computer understanding of three-dimensional motions was given by Ullman [Ullman-79] who has studied the problem from a rather different direction, understanding of psychological evidences. Applying his "*structure from motion theorem*" or similar mathematical theories [Roach-80, Meiri-80], we can construct computer programs which reconstruct three-dimensional geometry of each object in the

scene with [Nagel-81, Webb-81] or without a constraint of motion [Ullman-79, Asada-80] from the image sequence under an assumption that the corresponding points between frames are given. However, it does not mean that the computer understands the three-dimensional motions. Although this scheme gives us information on the interframe changes of the location and orientation of each object, the computer cannot interpret the changes into a motion description similar to what a human would perceive given the image sequence to one eye.

Another important but unsolved problem is to establish the correspondences of points between frames. As objects move and rotate in the three-dimensional world, their surfaces, edges, and vertices appear or disappear from frame to frame. Finding general solutions to determine the correspondences in the time-varying real world scenes is too difficult. We, therefore, and because of the knowledge accumulated in the research on the static scene analysis, choose scenes of the time-varying blocks world to study both this correspondence problem and the above-mentioned motion representation problem.

Roach and Aggarwal [Roach-79] developed a system which finds, tracks, and represents linear motions of convex polyhedra through a series of images. Their emphasis was laid on tracking of each block in raw image data by a hierarchical matching process, and the system does not accept the rotational movements which give us more information than the translational ones.

1.3 GOALS OF OUR SYSTEM

We have studied, from a rather theoretical point of view, how a computer, given the sequence of two-dimensional images of scenes as input, can understand general motions composed of both translations and rotations of the blocks. In the mini-world, the blocks show complex coincidental motions such that a block can rotate around a joint attached on one side of another moving block.

To achieve the goal, we use two assumptions as cues for the analysis. One is the rigidity of each object (or part) by which the computer can discover corresponding parts between dynamic line images and can reconstruct the 3-D geometry of each object, and the other is the consistency of motion and the relation between moving objects by which the system can understand

complex coincidental motions. It is reasonable for a computer to use these assumptions in order to analyze the time-varying images since we seem to utilize them very often to understand the 3-D world.

The following problems have been studied by using these assumptions.

- (1) Analysis of changes of the line drawing between frames and finding the correspondence of each vertex.
- (2) A method for evaluating 3-D geometry of each block and its interframe rotational movements.
- (3) Hierarchical representation of the coincidental motions.
- (4) Finding and representation of consistent properties of the movements through multiple frames.

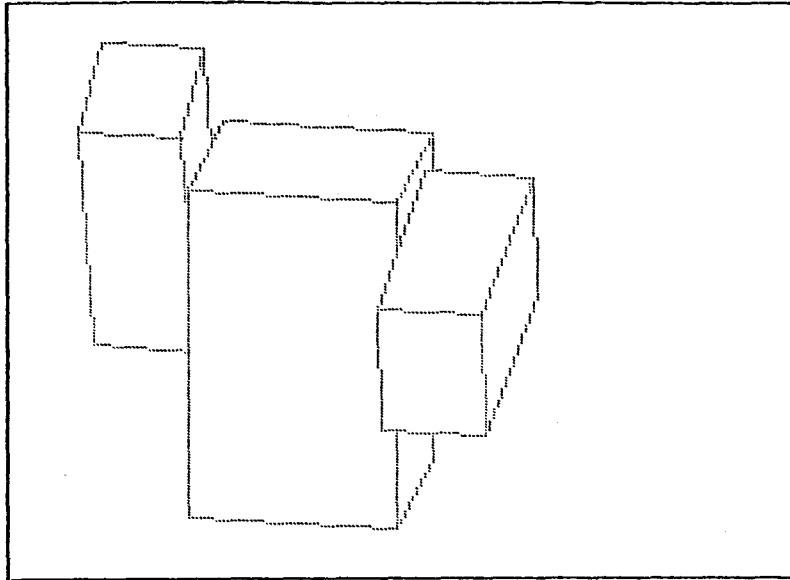
2. ANALYSIS OF IMAGE SEQUENCE OF TRIHEDRAL WORLD

2.1 INPUT IMAGES

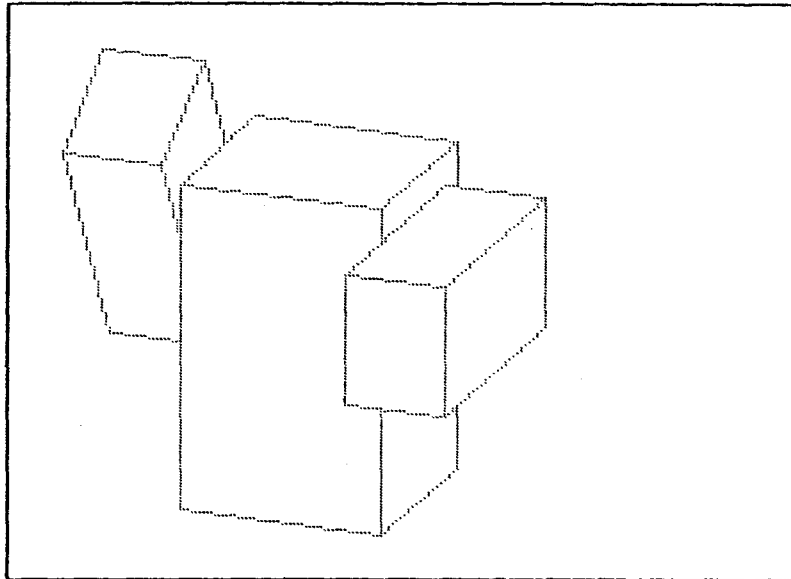
In order to focus on the fundamental problem of the motion understanding, we study scenes satisfying the following restrictions. Input is a sequence of noise free line drawings of rigid trihedral blocks projected orthogonally on an image plane. The line drawings are generated by computer, and each vertex is represented as a point in a 256 by 256 image array. It is assumed that time interval between frames is short enough that only small changes can occur. The block's motions are general; they consist of both translations and rotations. Some blocks might show complex coincidental motions. Fig.2.1 displays an example in which two small blocks rotate in different directions around joints attached on both sides of another moving block.

2.2 MOTION UNDERSTANDING

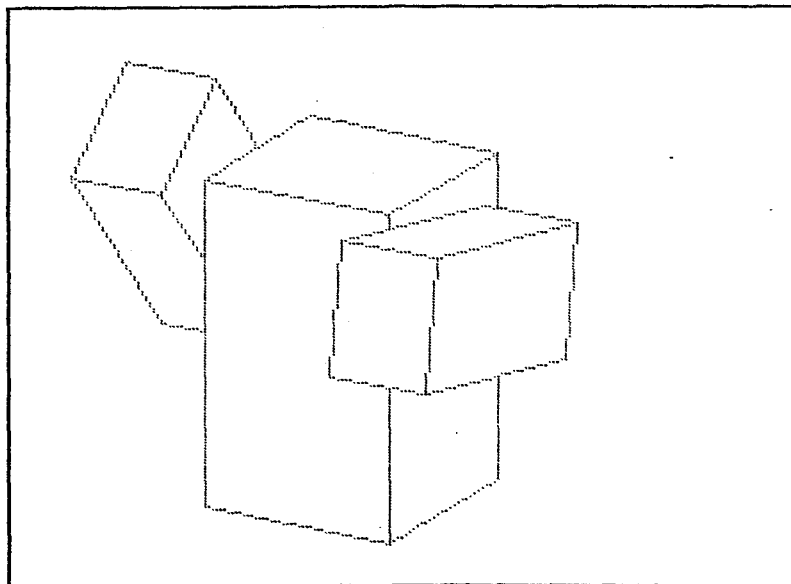
Now let us consider how we can analyze and represent such complex motions. If the correspondences of vertices between frames are determined, we can reconstruct 3-D geometry of each block and evaluate its interframe movements. Therefore, we can describe the motion with a sequence of these interframe movements. This type of descriptions, however, are not useful for the complex motions. Suppose that a computer vision system sees a man walking and describes the motion of his hands in this manner. Since the motion is measured and described in the camera's coordinate system, the representation is complex and it gives us little information. We, human beings, describe the motion in a compact yet informative form such that a walking man swings his hands. In this case, neither the global nor viewer's coordinate systems are used, and a coordinate system fixed to a moving object is adopted instead. We will apply this type of representation to our problem. Note that the representation is meaningful when we obtain also a good representation of the motion of the coordinate system. In other words, we need not only the finding of a suitable coordinate system for a simple representation of the object's motion but also the good representation of the motion of another object to which the



(a)

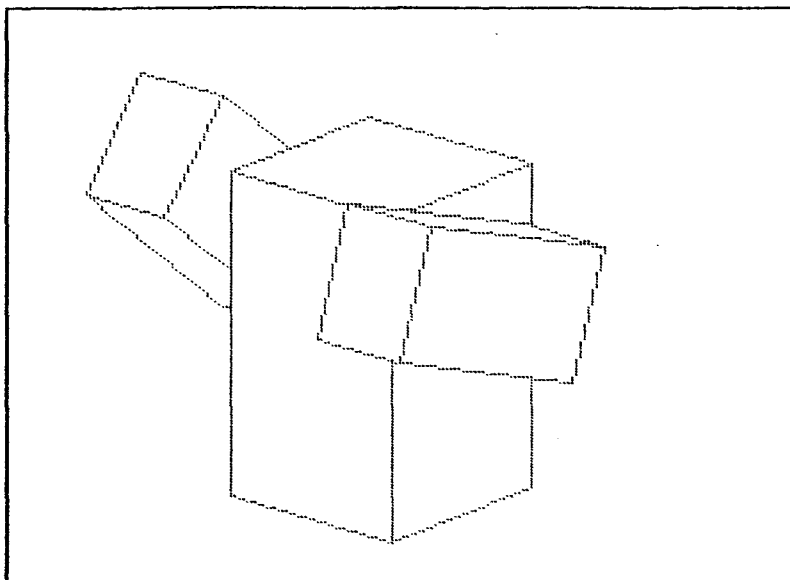


(b)

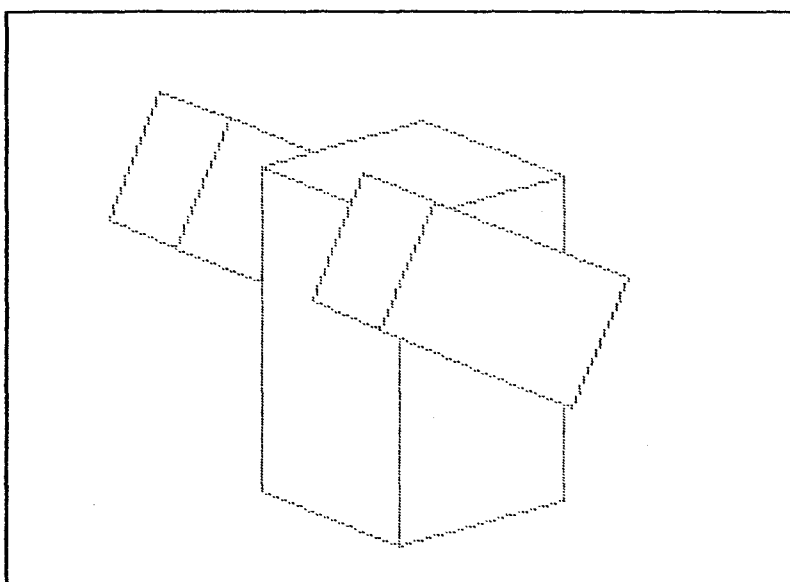


(c)

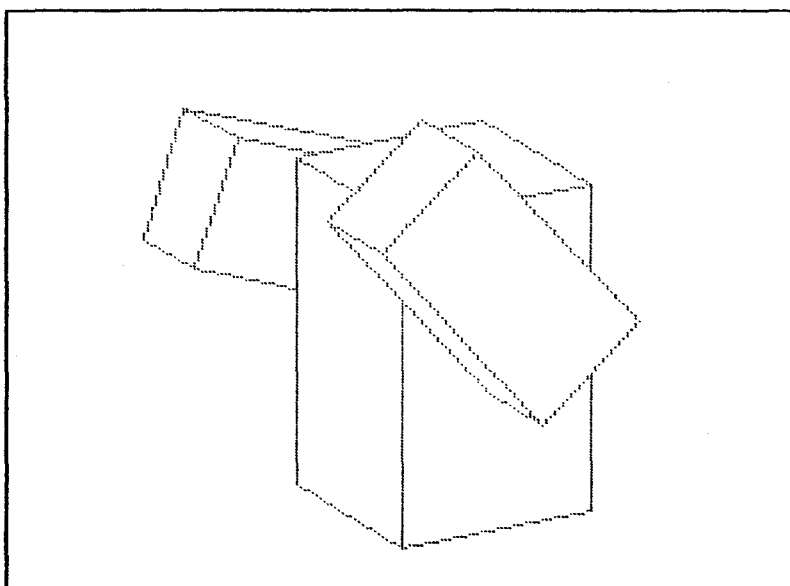
Fig. 2.1 (continued)



(d)

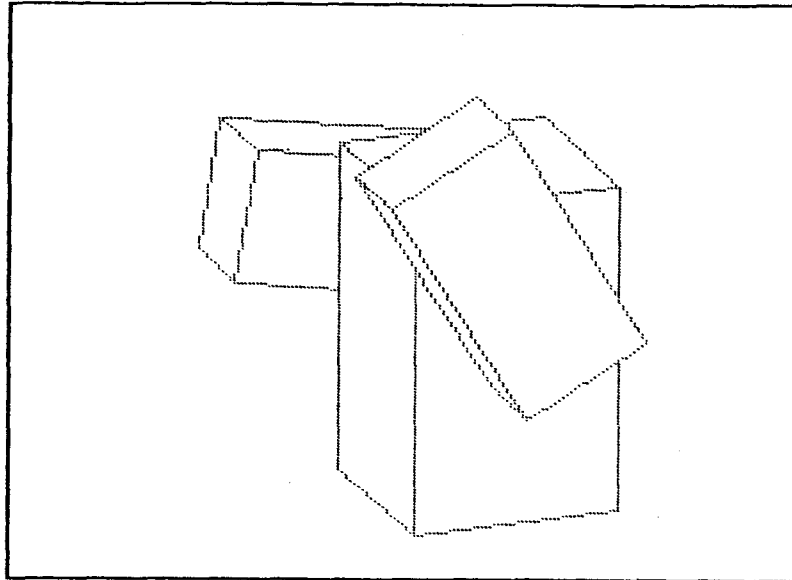


(e)

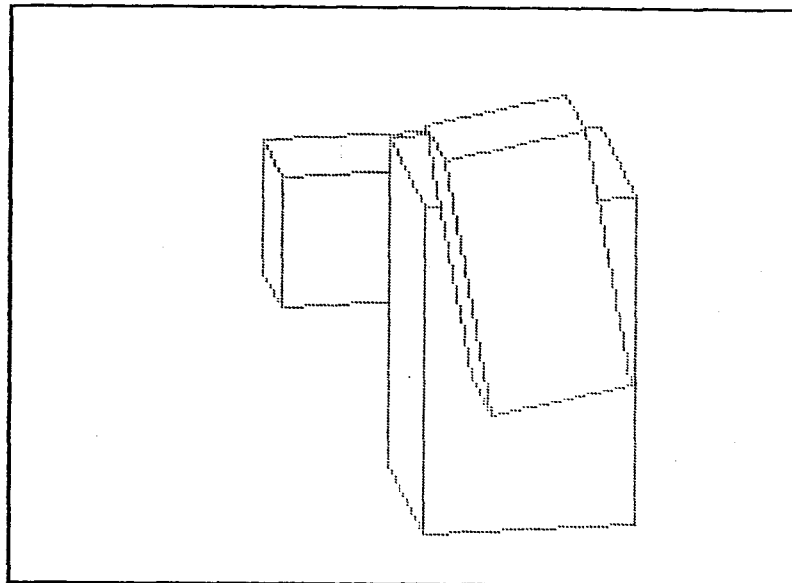


(f)

Fig.2.1 (continued)



(g)



(h)

Fig.2.1 Input image sequence.

OBJECT 1 (the largest block) is moving rightward and turning about a vertical axis through its top and bottom faces, and OBJECT 2 and OBJECT 3 revolve up and down around horizontal axes attached on the both sides of OBJECT 1.

coordinate system is fixed.

In this paper, we consider the simplicity of the motion representation as follows. If the properties of the motion are constant in a considerably long period, it can be described in a small number of terms. Therefore, the constancy of the two components of the motion, a translation and a rotation, is highly desirable for the simple representation. We select two measures for evaluating the goodness of the representation; invariance of the translational vector and the orientation of the rotation vector. We will show the invariance of the orientation is very useful for finding the suitable coordinate system. The constancy of the translation is used to find a good representation of movements determined by three views.

If a group of objects show a coincidental joint motion such that a block is rotating around an axis fixed to another translating and rotating block, the motion of the former is more complex than that of the latter. In such a case, the latter, detectable as of a simpler motion whose rotation vector does not change in its orientation, is considered as the main body of the group. We can find a good motion representation of the main body such that its translation is also simple. The other objects are likely to be the subparts of the group, if their motions described in the coordinate system fixed to the main body are simpler than those in the viewer's coordinate system. Thus, the coincidental motion is described in a hierarchical form, which is similar as the motion description by human beings.

2.3 ANALYSIS AND REPRESENTATION OF DYNAMIC BLOCKS WORLD

The first phase of our dynamic scene analysis is to determine a structural model, a set of models of individual blocks in each frame, from the image sequence; the analyzer segments every input frame into objects and establishes their correspondences. The junction labeling scheme such as Huffman's [Huffman-71] is very effective for static scene analysis. For applying it to the dynamic scene, we need to overcome the difficulties as follows: (1) appearance and disappearance of the junctions, (2) transition of the label of each vertex, and (3) non-trihedral junctions resulting from the accidental alignments. Investigating how the labels change as the block rotates, we introduce a transition table of the junction labels

for interpreting the dynamic blocks world. The changes in the structure of the line drawings are analyzed by utilizing the transition table and the contextual information obtained from the models of individual blocks in the consecutive frames. We use an object-to-object matching method to establish the correspondence of each junction in the image sequence, and assign meanings to the changes.

In the second phase, we evaluate the geometrical parameters of the models, such as orientations and edge lengths of the blocks, from the image sequence by utilizing the shape rigidity, an important cue for movement analysis. *Structure from motion theorem* by Ullman [Ullman-79] uses four points in three frames to determine the displacements of these points. His method, however, has the following drawbacks: (1) the verification of the rigidity of four points is necessary, (2) it is difficult to determine which of two solutions is true and which reflective, and (3) the displacements evaluated by this method would have large errors.

The model of each object obtained by the first phase is very useful to eliminate the drawbacks. The shape rigidity of any three points on a block is obvious. Referring to the label (+ or -) of an edge, the system can select the real solution from the two candidates. Since the images are orthogonal projections of the scene, any information on the distance from the image plane is not available. We can, however, determine the distance from a vertex to another vertex on a same object, by utilizing the shape rigidity property. The distance obtained from three vertices in three views by our method is more accurate than Ullman's because the vertices are much more aparted. Moreover, distance data obtained from many triplets of frames containing four vertices are integrated to make the solution more reliable; unreliable data are discarded and reliable ones are averaged. From these 3-D data, we can determine the rotational components of the interframe movements of each object.

The final stage is to find a simple and natural representation of the motions of the blocks through the image sequence. Since some objects may move in a group showing coincidental motion, we search the scene for objects of which motions are simple, that is, their rotational vectors do not change in the orientations; these objects may be the main body

of the group. Next, a hypothesis that another block showing a complex rotational movement is a subpart of the group moving with the main body is proposed, then it is verified if its relative motion (rotation) to the main body is simple.

Finally, the computer tries to describe the translational movements of the objects; at first the description of the main body and then that of other parts viewed from the main body. In the description of the interframe movement, the orientation of the rotational axis and the rotational angle are uniquely determined. On the contrary, the translation cannot be uniquely determined but depends on the position of the rotational axis. That is, we can set the rotational axis at a certain place so that the translation can have the consistent property through multiple frames, such as a constant velocity, a constant acceleration and so on. We prepare computer programs to examine these consistent properties and apply them to the motion of the object. If one of them is satisfied in many frames, then its motion is described by that property. Otherwise, those frames are interpreted that the movement of the object abruptly changed then.

Thus, in order to understand the 3-D motions, the computer represents them with the simple and natural descriptions similar to what a human would perceive.

3. STRUCTURAL ANALYSIS OF LINE IMAGES OF DYNAMIC BLOCKS WORLD

3.1 JUNCTION LABELING OF DYNAMIC SCENE

In this chapter, we discuss a method for constructing models of moving individual objects by examining line semantics and utilizing contextual information between consecutive frames. The well-known line labeling methods, proposed and developed by Huffman [Huffman-71], Clwes [Clwes-71] and Waltz [Waltz-75], are successfully applied to the static blocks world. Following Huffman's scheme, we design an analyzer of the line images of the time-varying blocks world, because his algorithm is simple and easy to be programmed yet it is capable of assigning meanings to lines and junctions in most static images.

For the analysis of the dynamic blocks world, we need, however, to augment the analyzer's ability because of the following difficulties. (1) Huffman's method assumes that the viewer is in a general position, therefore it does not accept images of non-trihedral vertices resulting from the accidental alignments. These types of junctions, however, frequently appear in our images as objects move and rotate in the three-dimensional world. (2) Our scene model should represent the interframe movements of individual objects, therefore we need to find the correspondence of each vertex between the frames. It is very difficult to establish the correspondence, because the change in the image structure often occurs; labels of junctions change, and some junctions appear or disappear from frame to frame.

Labels of lines used in this paper are same as those used by Huffman; symbols of +, - and an arrow indicate that the attached line is a convex, concave and occluding edge, respectively. Huffman's 18 types of junctions (six Ls, five FORKS, three ARROWS, and four Ts) and additional four types of junctions (four special ARROWS described in later) are defined as legal. Junctions resulting from the accidental alignments are defined as illegal and are named N (Non-trihedral) type.

Each line drawing in the image sequence is independently labeled and, if possible, it is segmented into objects. The process of the labeling and segmentation of the line drawing is as follows. At first, the exterior boundary of the line image

is labeled with a clock wise sequence of arrows. Next, the constraints of the labeling are propagated to the adjacent junctions. Each branch of the propagation is suspended when it arrives at an N junction with more than three legs or when the assignment of any legal label to a line is not possible because of illegal ARROWS, FORKS, or Ts explained in later. Areas surrounded by a clockwise sequence of occluding edges are extracted from the images as objects, if all junctions in them are interpreted as legal.

3.2 MATCHING OBJECTS BETWEEN FRAMES

We use an object-to-object matching method to establish the correspondences of the vertices between frames. Starting with a frame in which the image is segmented and their models are constructed, the matching of the object propagates to both directions in the image sequence. We utilize a search tree method, a similar method as the labeling which utilizes the external constraints for pruning, for matching each model to the adjacent images.

The basic assumption that only small changes can occur between two consecutive frames gives useful constraints. Each vertex in the model is allowed to move in a small neighborhood in the matched image or to disappear. As a result, the matched image has only a small number of candidates for each of the model's vertices. A more strict constraint is obtained when the labeling scheme was able to segment the matched image into objects, because it is easy to find which object is matched to an object in the model from the distance and/or size information.

Although these neighboring constraints are useful, they cannot deal with the changes in structures of the line images. The following sections in this chapter discuss how these structural changes occur as objects move in the scene, and shows the external constraints or rules obtained by the analysis are useful for pruning of the search tree and finding of the true correspondences.

3.3 LABEL CHANGES BY ROTATION OF SINGLE OBJECT

At first, we investigate how junction labels can vary as an isolated object moves, and arrange the rules on the label

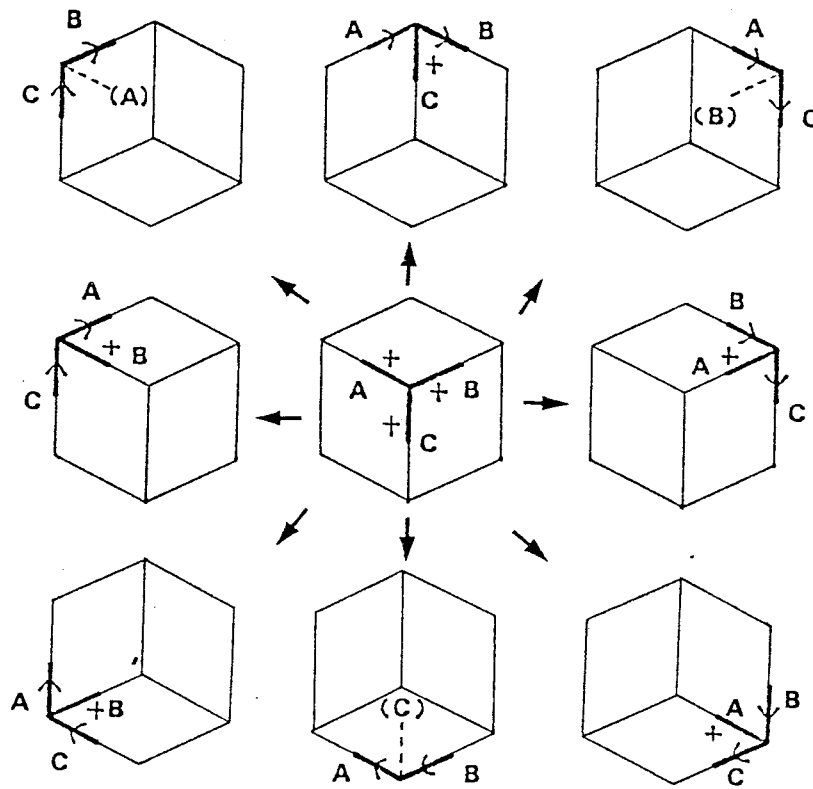


Fig.3.1 Transitions of a type 1 vertex by rotation.

transitions into a table. Referring to this transition table, the analyzer can find correspondences of the junctions between the consecutive frames. Since the transitions of junction labels of the isolated object can be caused only by its rotational movements, we investigate whether one junction label can be changed to the other and what type of rotation causes it. Fig.3.1 demonstrates the transitions of a FORK with each line labeled with a + symbol.

Fig.3.2 summarizes the transitions of junctions. Junctions appearing as projections of vertices of a same type is shown in each row (the type number tells how many octants of space around the vertex are filled with solid material). The junctions can transit each other in a same row, however the rotation of the isolated object does not cause the transition to any member of a different type. The original junctions, projected images of vertices with three visible surfaces, are shown in the left side. As the object rotates, the original junction transits to another member with one or two occluding edges, and the direct transition from the member to an obscured state is possible; it disappears if the object rotates further to a certain direction. The right column shows junctions which appear in the transition processes. Their shapes are same as those of Ts, but they corresponds to real vertices and, hence, we discriminate these junctions as special ARROWS from the Ts when we match a model object to the consecutive frame.

Now let us study in more detail the transition rules which will be used in the matching process. By analyzing the label transitions, we can find constraints of a change in a line label of a junction to changes of the other lines. Fig.3.3 (a) illustrates an example of rules: if the symbol of a line of a type 1 ARROW transits from + to an inward arrow, then another line with an inward arrow must disappear and the other line does not change. The transition of a + symbol of a type 3 ARROW to an inward arrow allows three types of changes shown in Fig.3.3 (b). We examine such rules for all possible transitions and arrange them into a transition table in order for a computer to utilize them. Fig.3.4 shows the transition rules for the all types vertices.

Next, we show how these rules are applied to find the correspondences of vertices between frames by a simple example;

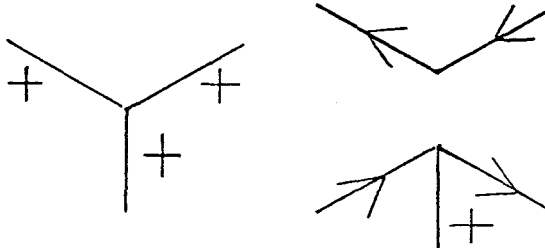
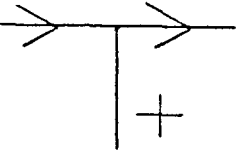
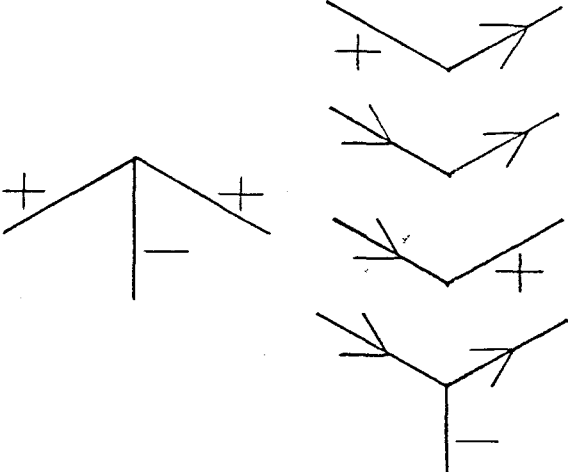
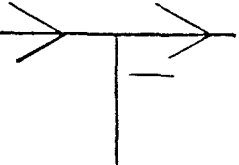
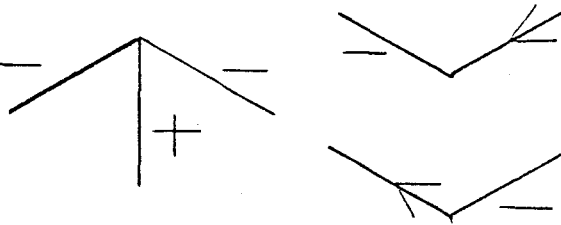
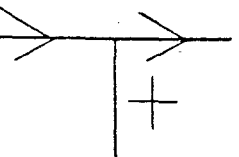
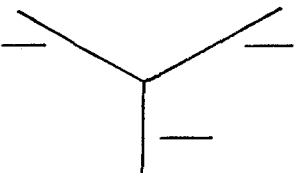
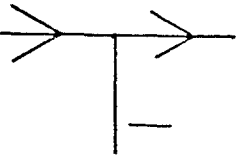
Junction Type	Family of Junction Labels	Special ARROW
I		
III		
V		
VII		

Fig.3.2 Junction transitions.

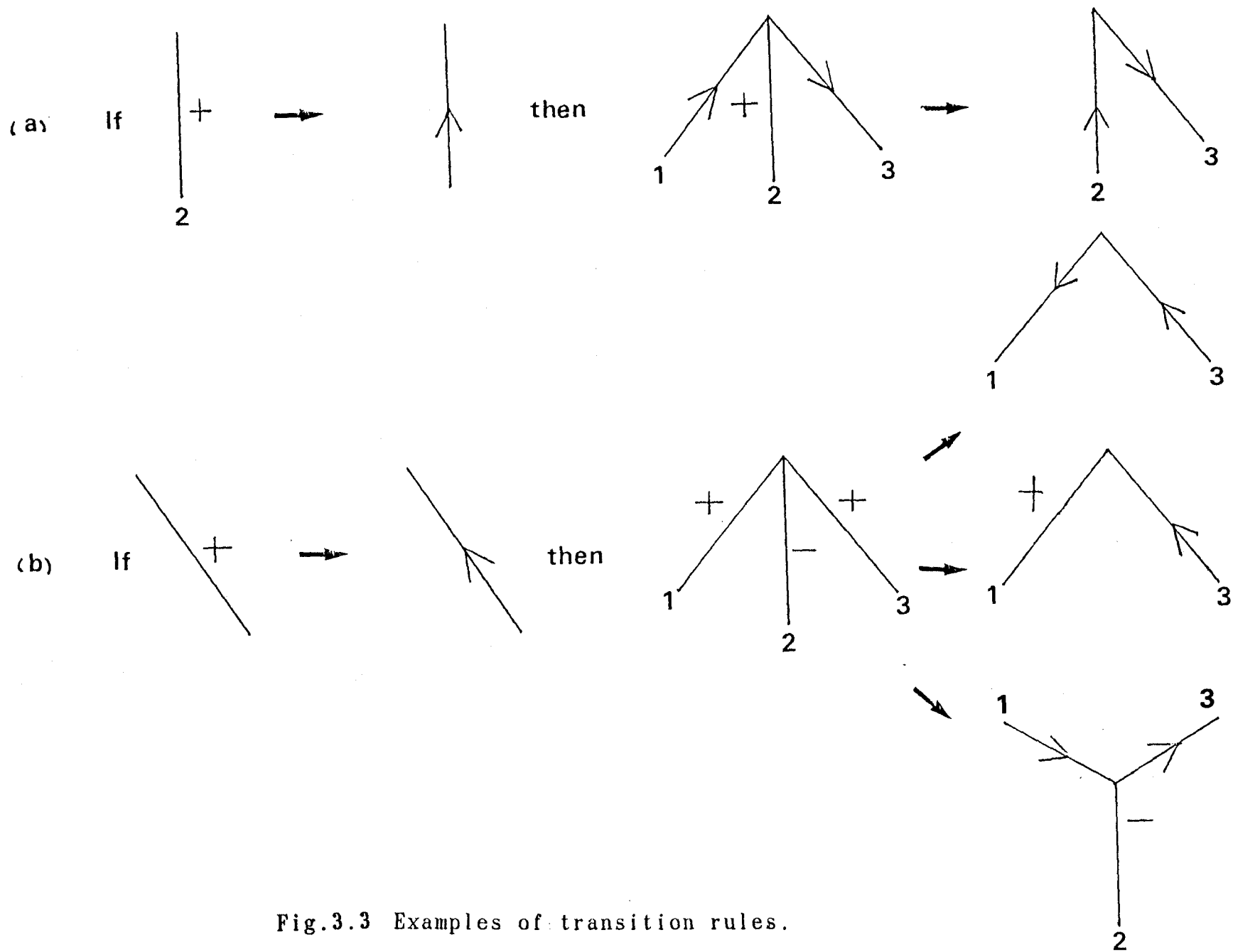


Fig.3.3 Examples of transition rules.

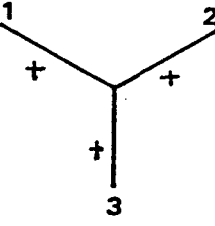


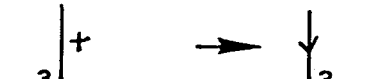


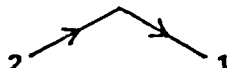
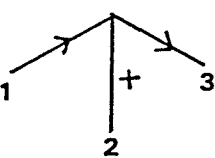


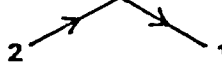
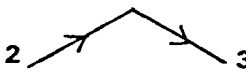
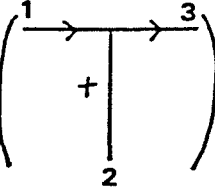
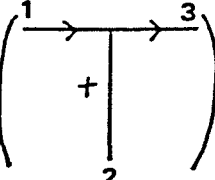
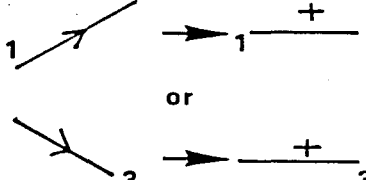
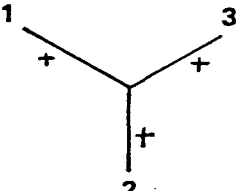
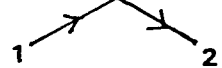
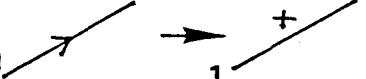
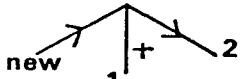
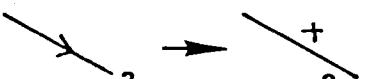
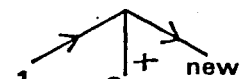
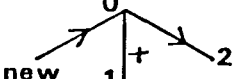
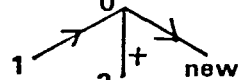
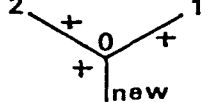

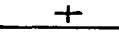
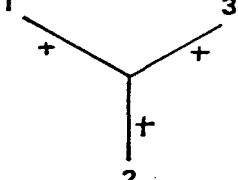
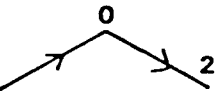

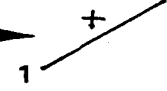


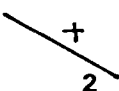
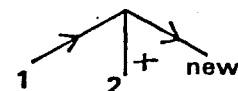

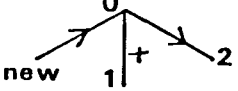
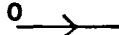
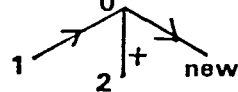

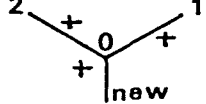
Type	Junctions	Allowable Change of a Line Label	Results
I		 → 	 or 
		 → 	 or 
		 → none	
	 	 or   → none	
		 or   → none	
		 → none	disappear
		 →  or  → 	
		any line → none	disappear
		 → 	
		 → 	
		none → 	
		none → 	
		none → 	

Fig.3.4 (continued)

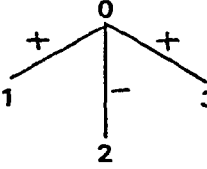
















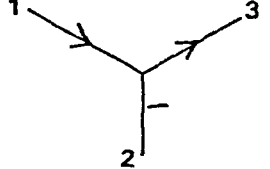







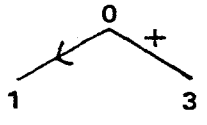




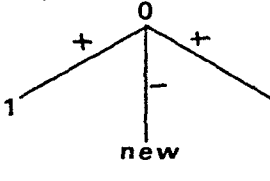

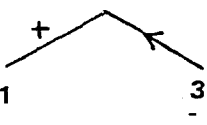




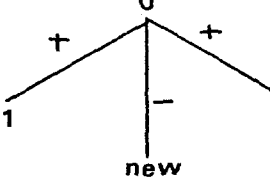

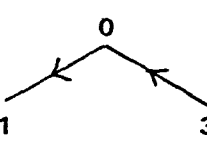













Type	Junctions	Allowable Change of a Line Label	Results
III		 → 	  or  
		 → 	  or  
		 →  or  → 	
		 → none	  or   or  
		 →  or  → 	
		 → none	disappear
		 →  or  → 	
		 → none	disappear
		 → 	 
		 → 	 
		 →  or  → 	
		any line → none	disappear

Fig.3.4 (continued)

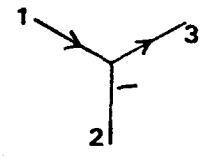
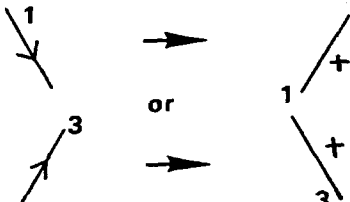
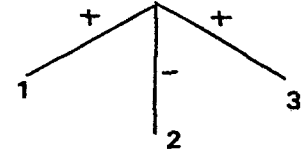
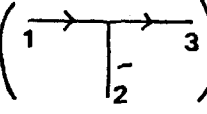
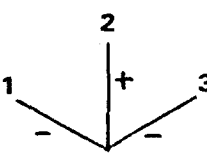
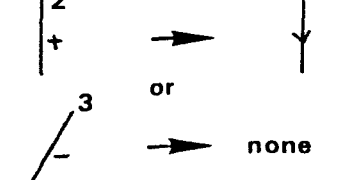
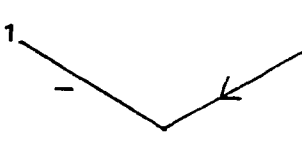
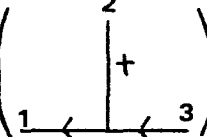
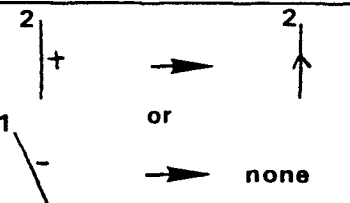
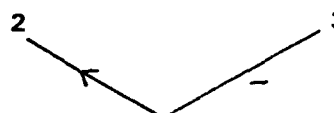
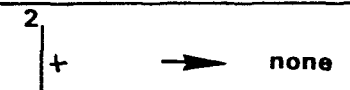
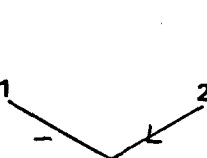
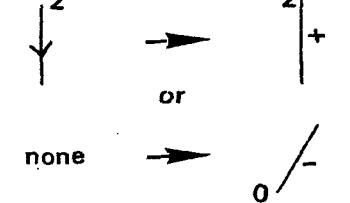
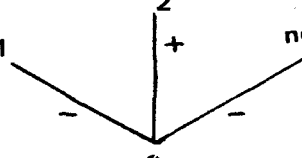
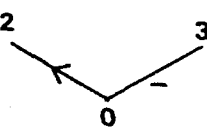
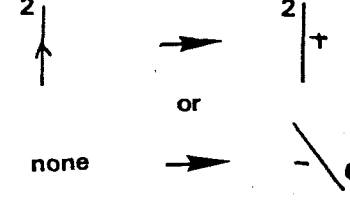
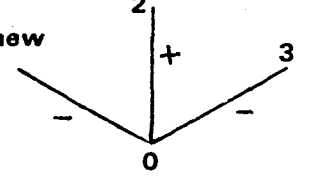
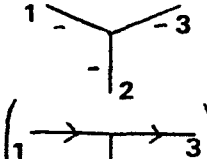
$T_{y_{pe}}$	Junctions	Allowable Change of a Line Label	Results
III			
		any line \rightarrow none	disappear
V			
			
			disappear
			
		any line \rightarrow none	disappear
			
		any line \rightarrow none	disappear
		any line \rightarrow none	disappear

Fig.3.4 Transition table of junctions.

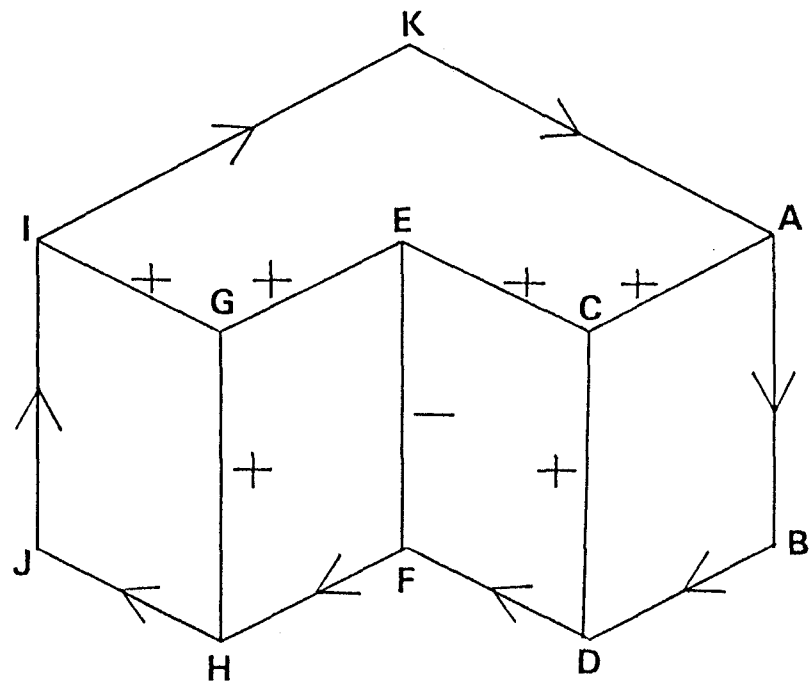
matching of a model (Fig.3.5 (a)) to a line image (Fig.3.5(b)). At first, C in the model is matched to the junction 1, a junction with same features as C and, then, the matching operation propagates in the line image. Judging the matching of E as a fail because the junction to be matched (junction 3) is of a different type, we backtrack and try to match C to junction 3. As a result, we find that there must be changes in the labels of two edges, CD and CE, of the model if the correspondence of C with 3 is correct. From these changes, we can predict changes of the adjacent junctions in the model by referring to the transition tables. If each label of the matched image coincide with one of those predicted from the model, then the matching is considered as successful. In this case, CD is transitted to an upward arrow, then D is an L junction and DF must disappear. Since CE is with an arrow, E must be one of three types shown in Fig.3.3 (b). Without any difficulty, we can match these junctions of models to the image, and the matching operation propagates until all junctions in the models are either matched or interpreted as disappearing vertices.

This process of prediction and matching is useful to interpret the unmatched parts as appearing or disappearing. For example, when the analyzer tries to match A to 1, it searches the transition table and finds that AB and AK are allowed to be with +, and the two junctions are considered as matched. From this information, we predict K and B must have new occulding edges and, if an unmatched part of the image (for example, junction 10) connect with these edges, assign a meaning to it that it appeared between the two frames.

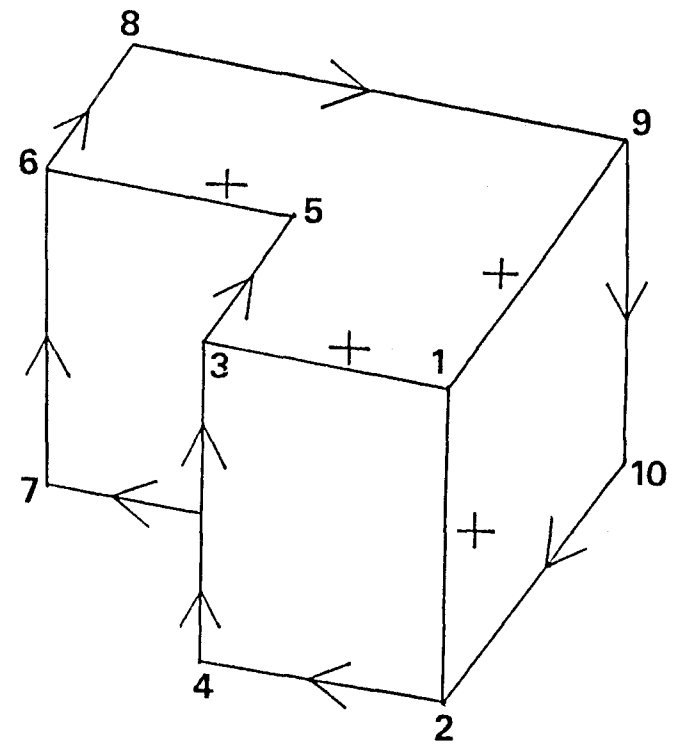
3.4 CHANGES IN STRUCTURES OF LINE IMAGES CONTAINING PLURAL OBJECTS

Changes of the line images of the dynamic scene containing plural objects are more complex. We first study the line drawing without the N type junctions and, then, examine the difficulties resulting from the accidental alignments in the next section.

When an object occludes a portion of another object, T junctions appear on the occluding edges of the object. The T junctions give, hence, very important clues for finding and analyzing the occlusion. The matching process of an object in



(a)



(b)

Fig.3.5 Matching of model (a) to image (b).

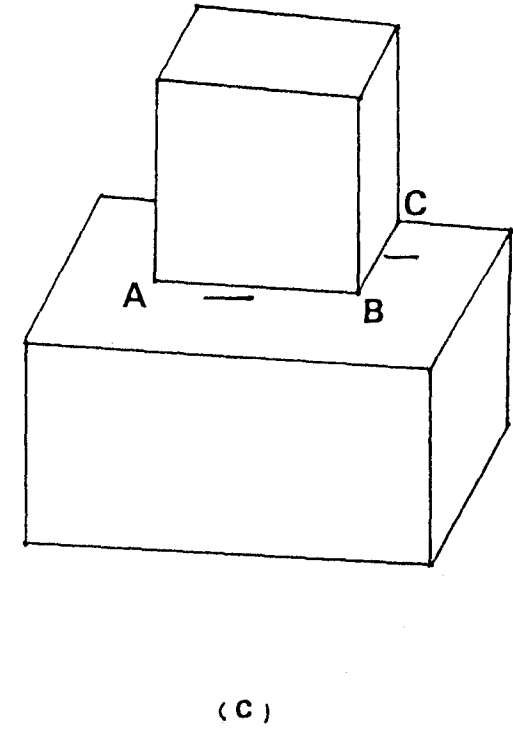
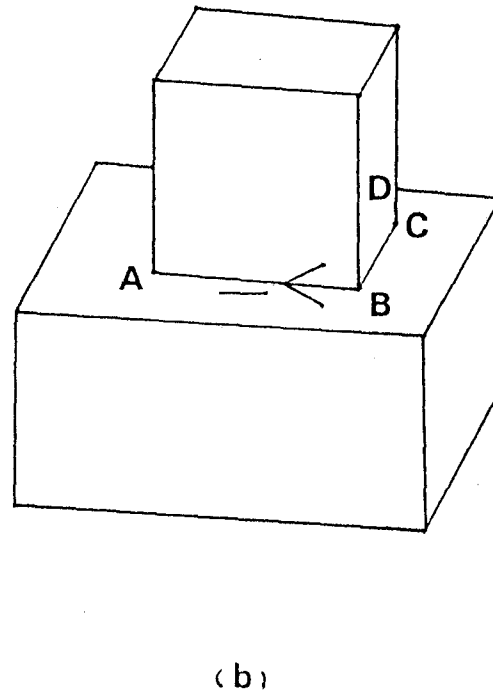
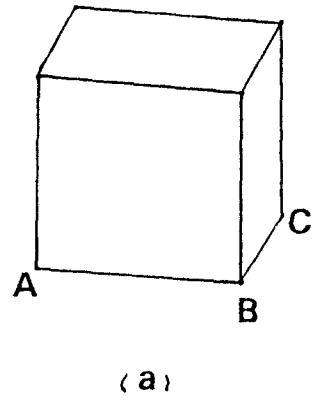


Fig.3.6 Transitions to a different type of vertices.

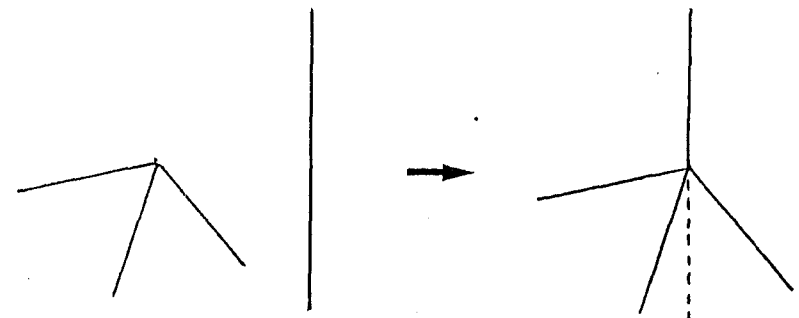
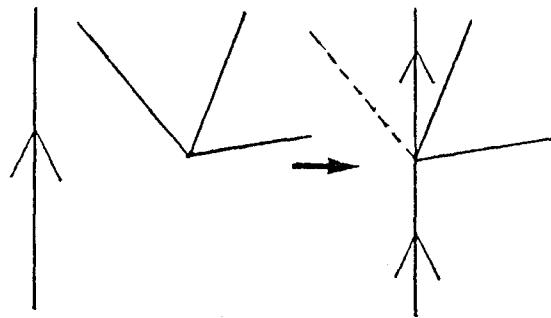
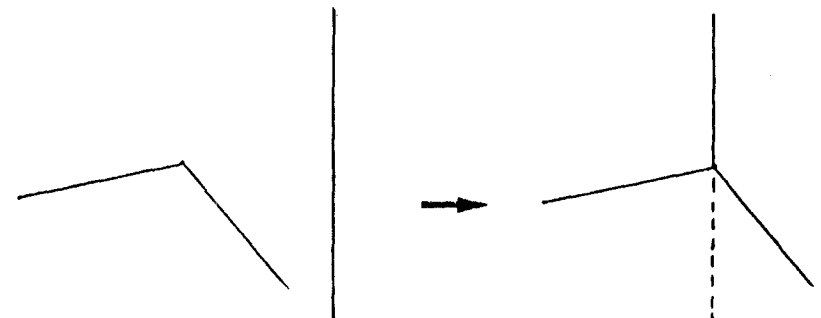
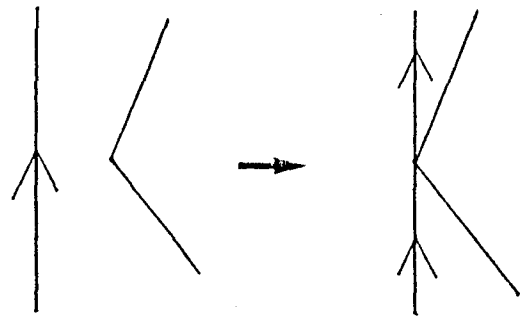
the model to another image interprets some T-like junctions as special ARROWS, otherwise the edges corresponding to the stems of Ts are considered as occluded at the joints and, therefore, as connecting to unmatched portions, giving a similar interpretation as that of disappearing edges discussed before.

Another difficulty in analyzing the changes of the line drawing is that the types of vertices change or appear to change when objects are seen above or in front of other objects. It is well known that a simple scene as shown in Fig.3.6 (b) has two interpretations; a cube floating above a block or an integrated objects with them (we consider the case that the cube rests on the block is a special one of the former that the distance between them is zero). Comparing Fig.3.6 (a) with (b), one may consider that some vertices change their types; for example, a vertex B transits from type 1 to type 5. Fortunately, the contextural information between (a) and (b) gives true labels; we can easily select one of two interpretations by the object to object matching. However, there exist exceptions (see Fig.3.6 (c)) that the labeling is unique but AB and BC are with a different label from Fig.3.6 (a). Note that the junction C in Fig.3.6 (c) is an N type resulting from the accidental alignment to which a matching method discussed in later is applicable.

3.5 ANALYSIS AND INTERPRETATION OF NON-TRIHEDRAL VERTICES

Now let us consider the problem of analyzing line drawings which contain illegal junctions resulting from the accidental alignments. Since so many types of N junctions are possible to appear, we do not intend to construct a catalog of them for finding and interpreting the illegal junctions. Instead, we utilize a common strategy for analyzing the dynamic scenes. The strategy is that if a decision is not made with the current scene, the uncertainty is carried forward in the hope that it can be resolved in the later scene [Roach-79]. In other words, we expect there exist frames in the image sequence for each object such that it is viewed from a general position and, thus, the object's image is interpretable. Once the object is found, the N junctions in the adjacent frame are interpreted by the object-to-object matching method.

Befor applying the above idea, we need, however, to know what type of accidental alignments can occur and how they



(a)

(b)

Fig.3.7 Examples of illegal junctions resulted from accidental alignments.

interfere the line labeling. Waltz [Waltz-75] classified the accidental alignments into three types, one of which is excluded from our case because we do not deal with scenes with shadows. One type of the alignments, occurring when an edge between the eye and a vertex appears to be part of the vertex, are detectable, because the N junction of this type are with more than three legs as illustrated in Fig.3.7 (a). Since the edge is visible as a line passing through the junctions, this type of alignments and junctions are called E (Edge) type. Another feature of an E type junction is that the other legs lie in one side of the line. The other type of alignments occur when a vertex is closer to the eye than an edge which appears to be but is not part of the vertex. Since the vertex of the object is visible in the image, this type is called V (Vertex) type. As shown in Fig.3.7 (b), V type alignments generate not only the above-mentioned extra line type junctions but also ARROWs, FORKs and Ts. When the edge is aligned to an L junction, many types of illegal ARROWs, FORKs and Ts are possible to appear. We can assign the legal labeling to some of the line drawing containing it as Fig.3.6 (c), but many line drawings with the V junctions are interpreted as impossible scenes. Thus, we have two indications of illegal junctions; the number of lines radiating from each junction and the interpretation as the impossible scene.

Finally, we consider a procedure for matching the model to an N junction in the image. We first test whether all N junctions in the image have two features of E type; a straight line through the junction and legs lying in one side of the line. Those having the features are regarded as E-like, and the straight lines through the junctions are matched to occluding edges of the object models. Note that the other two legs may correspond to an L or a partly obscured ARROW or FORK. When a model's junction, predicted from the constraints by the method described before, is matched to an E-like junction, the matching is considered as successful (1) if the model's junction is an L which we can match to the two legs, or (2) if the model's junction is an ARROW or a FORK of which two legs are matched to the two legs of the E-like junction.

When a junction is judged as V type, that is, an N with three legs or one with extra legs but judged as not E type, we

match an L predicted from the model to a part of the V junction. If the matching is in success, the other leg is considered as a stem of a T junction.

Thus, we can interpret the junctions resulted from the accidental alignments unless two vertices are aligned. We argue that such alignments seldom occur; therefore we could skip the frame and try to find the correspondences to the next frame.

3.6 EXAMPLE OF SEGMENTATION AND MATCHING

This section shows an example of the junctions labeling and object matching in the dynamic scene. The results of applying Huffman's scheme to images in Fig.2.1 (b), (d) and (e) are shown in Fig.3.8 (a), (b) and (c), respectively. Note that only (a) is perfectly labeled but (b) and (c) have junctions with "?" of which any unique and legal interpretation is not possible. For example, N' and P' are N type junctions with extra lines and legal labeling of WI' is not possible.

Utilizing the model of OBJECT 2 (the block at the right side) in Fig.2.1 (c), the analyzer tries to match the original junction O. By using the neighboring constraints, we find two candidates O' and W, and O' is selected first because it is also a FORK. (If W is selected first, the matching of W fails at finding a corresponding point to J.) When the analyzer matches P in the model, P' is selected as only one candidate. Since P' is V type, one of the four arms of P' must be a stem of T junction. Then, the analyzer generates four candidates in each of which one arm corresponds to a stem of T, and deletes those cases which are not reasonable to the labels that are correctly interpreted in labeling. Applying the transition rules to the reasonable cases, the analyzer selects a correct one which does not contradict to the adjacent junctions.

Fig.3.9 (a) shows the model of vertex P (Fig.2.1 (c)) and the four candidates of the corresponding non-trihedral vertex p' (Fig.2.1 (d) or Fig.3.8 (b)) are shown in Fig.3.9 (b),(c),(d) and (e). (d) and (e) indicate the vertices (P'-O'K'F' and P'-K'F'M') which must not exist physically, and a stem of T in (c) cannot exist there although the vertex P'-O'M'F' is able to be as a FORK of Type 1. Only (b) indicates a reasonable one and therefore non-trihedral vertex P' is decomposed into a line P'F' (a stem of T) and an ARROW (P'-M'O'K') to which the matching of

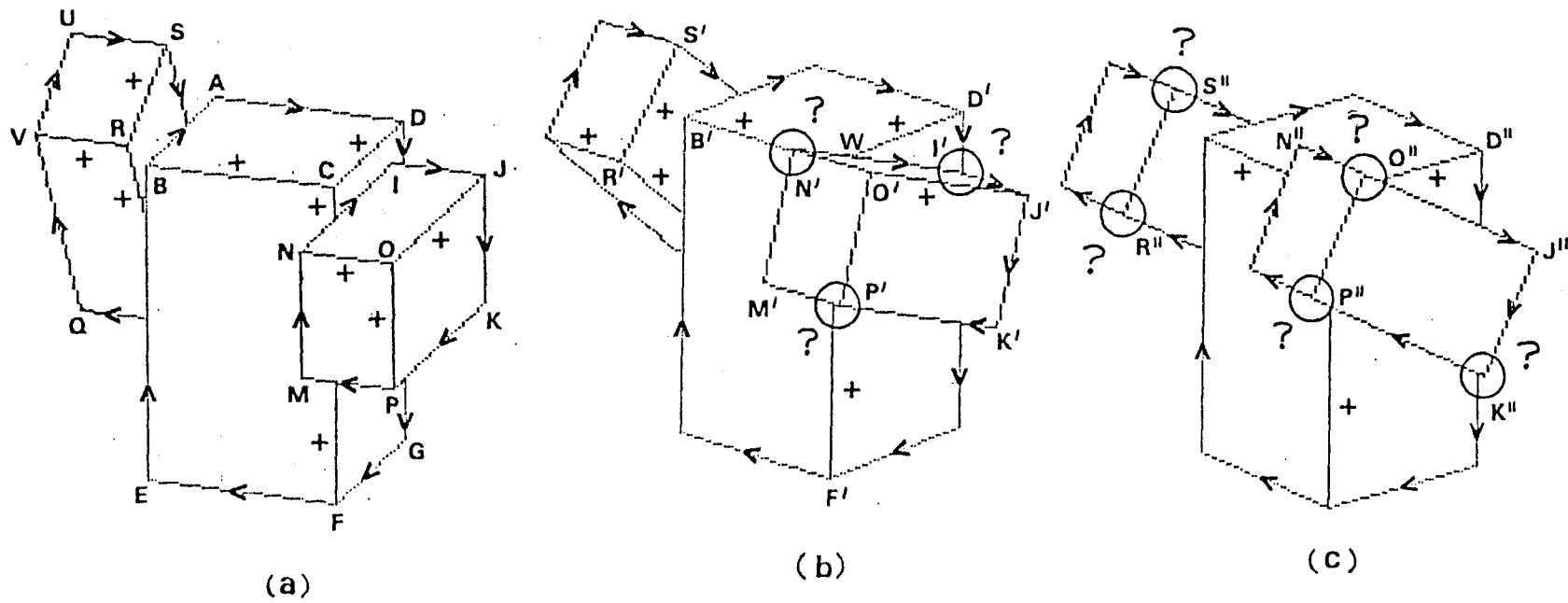


Fig.3.8 Results of junction labeling.

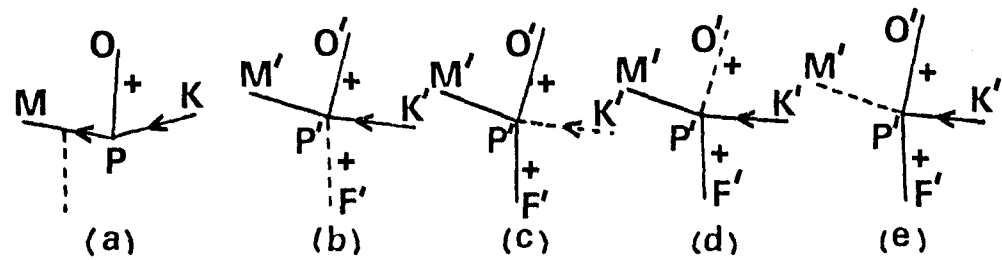


Fig.3.9 Analysis of non-trihedral vertex.

P is successful. In the matching process of N, a junction at I' appears as a FORK but it is a V type and is decomposed into a line I'D' and an L junction (I'-N'J'). Thus we can accomplish the matching of all vertices on OBJECT 2.

Fig.3.8 (c) is analyzed as follows. The labeling is not completed, because there exist two illegal junctions at K" and O", and a line S"R" was not given a unique label. The V type junction at O" is decomposed into a line O"D" and a special ARROW junction (O"-J"P"N") when a junction (O-JPN) in the model is matched. When the analyzer matches an ARROW junction at S in the model to S", it regards S" as a special ARROW with two occluding edges and a convex edge.

The matching is successful for the image sequence shown in Fig.2.1, and the movement and label change of each vertex on the three detected objects are tracked. Table 1.1 shows how the state of each junction has changed.

OBJECT 1.		(A)	(B)	(C)	(D)	(E)	(F)	(G)
Frame\Junction		(A)	(B)	(C)	(D)	(E)	(F)	(G)
(a)		L	A	F	A	L	A	L
(b)		*	*	*	*	*	*	*
(c)		*	*	*	*	*	*	*
(d)		*	*	o	*	*	*	*
(e)		*	*	*	*	*	*	*
(f)		*	*	*	*	*	*	*
(g)		*	*	*	*	*	*	*
(h)		*	*	F	*	*	*	*

OBJECT 2.		(H)	(I)	(J)	(K)	(M)	(N)	(O)	(P)
Frame\Junction		(H)	(I)	(J)	(K)	(M)	(N)	(O)	(P)
(a)		o	L	A	L	L	A	F	A
(b)		*	*	*	*	*	*	*	*
(c)		*	*	*	*	*	*	*	*
(d)		*	*	*	*	*	*	*	*
(e)		*	o	L	*	*	L	sA	sA
(f)		L	*	*	A	A	*	A	F
(g)		*	*	*	*	*	*	*	*
(h)		*	*	*	*	*	*	*	*

OBJECT 3.		(Q)	(R)	(S)	(U)	(V)
Frame\Junction		(Q)	(R)	(S)	(U)	(V)
(a)		L	F	A	L	A
(b)		*	*	*	*	*
(c)		*	*	*	*	*
(d)		o	*	*	*	*
(e)		o	sA	sA	*	L
(f)		*	A	F	A	*
(g)		*	*	*	*	*
(h)		*	*	*	*	*

* is same as the above junction label

A : ARROW junction

sA : special ARROW junction

F : FORK junction

L : L junction

o : occluded

Table 3.1 States of each junction at every frame.

4. MEASUREMENT OF INTERFRAME MOVEMENTS

In the previous chapter, we have analyzed the sequence of line images and constructed the qualitative models of the objects. Now we evaluate their 3-D parameters such as edge lengths and orientations from the image sequence, and determine the 3-D movements in each time interval between frames. Since the projection type of scene to image is orthogonal, the x and y coordinates of the vertices are self-evident. This chapter discusses how we can evaluate the z -coordinates of vertices on an object, and determine the rotational components of the interframe movements. The representation problem of the translational components is discussed in the next chapter.

4.1 RECONSTRUCTION OF 3-D GEOMETRY

Ullman [Ullman-79] has shown a method for reconstructing a 3-D model of each object, a collection of 3-D coordinates of points on it, at each frame from an image sequence of a scene containing tachistoscopic dots attached on moving objects. The input images are the orthogonal projections of the scene, therefore any information is not available on the object's movements in depth. Utilizing his *structure from motion theory*, we can obtain the exact model (to within a reflection) from four points in three views. If additional information such that the points are on a rigid body is given, the numbers of the points and/or views needed to determine the 3-D parameters decrease. Roach and Aggarwal [Roach-80] and Meiri [Meiri-80] have studied this problem for the sequence of images being the central projections of the time-varying scene.

This section presents another method for evaluating the 3-D parameters of the blocks from the image sequence of which the projection type is orthogonal. The advantages of our method over Ullman's result from the utilization of knowledge about the scene, the model of each block, which has been obtained by the analysis described in the previous chapter. The evaluation of the 3-D parameters is possible from three points in three views, because we know which vertices are on a same object. Referring to the label (+ or -) of an edge, we can decide which of two solutions is real, which reflective.

Let us consider the transition from non-collinear three points O,A,B in the first frame to those in the second frame as shown in Fig.4.1. For simplicity, the second image was shifted in such a way that O does not move between the two frames. We select O as the temporal origin of the coordinate system and allign the z-axis with the optical axis of the camera. Let $Z_j(i)$ be the z-coordinate of the point j in the i-th frame. Next three equations are derived for four unknown variables ($Z_A(1), Z_B(1), Z_A(2)$ and $Z_B(2)$)

$$X_A(1)^2 + Y_A(1)^2 + Z_A(1)^2 = X_A(2)^2 + Y_A(2)^2 + Z_A(2)^2 \quad (1)$$

$$X_B(1)^2 + Y_B(1)^2 + Z_B(1)^2 = X_B(2)^2 + Y_B(2)^2 + Z_B(2)^2 \quad (2)$$

$$X_A(1)X_B(1) + Y_A(1)Y_B(1) + Z_A(1)Z_B(1) = X_A(2)X_B(2) + Y_A(2)Y_B(2) + Z_A(2)Z_B(2) \quad (3)$$

By deleting $Z_A(2)$ and $Z_B(2)$ and rearranging (3), we have the following equations for $Z_A(1)$ and $Z_B(1)$.

$$PZ_A(1)^2 - 2QZ_A(1)Z_B(1) + RZ_B(1)^2 = RP - Q^2 \quad (4)$$

where

$$P = X_B(2)^2 + Y_B(2)^2 - X_B(1)^2 - Y_B(1)^2 \quad (5)$$

$$Q = X_A(2)X_B(2) + Y_A(2)Y_B(2) - X_A(1)X_B(1) - Y_A(1)Y_B(1) \quad (6)$$

$$R = X_A(2)^2 + Y_A(2)^2 - X_A(1)^2 - Y_A(1)^2 \quad (7)$$

The equation (4) represents a curve of second order with its center at the origin on the $Z_A(1)$ - $Z_B(1)$ plane. This gives only a relation between $Z_A(1)$ and $Z_B(1)$. Next, we examine another relation between the first and third frames and obtain the following equations.

$$SZ_A(1)^2 - 2TZ_A(1)Z_B(1) + UZ_B(1)^2 = US - T^2 \quad (8)$$

where

$$S = X_B(3)^2 + Y_B(3)^2 - X_B(1)^2 - Y_B(1)^2 \quad (9)$$

$$T = X_A(3)X_B(3) + Y_A(3)Y_B(3) - X_A(1)X_B(1) - Y_A(1)Y_B(1) \quad (10)$$

$$U = X_A(3)^2 + Y_A(3)^2 - X_A(1)^2 - Y_A(1)^2 \quad (11)$$

Then, $(Z_A(1), Z_B(1))$ is given as the cross points of the curves (4) and (8). Thus, two sets of the solutions are obtained, and we can determine which of them specifies the z-coordinates of the vertices by utilizing the labels of edges. The procedure is given in Appendix I.

The z-coordinates obtained by the above method sometime have large errors, because the errors in finding the locations of the junctions in the images are magnified in the reconstruction process. Especially when the object moves in such a way that the shape of the triangle OAB changes little,

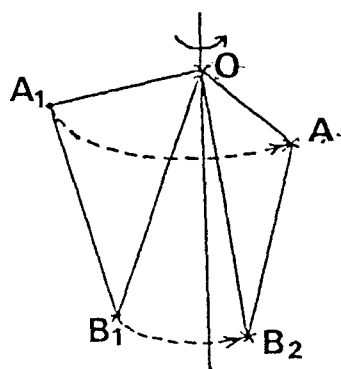
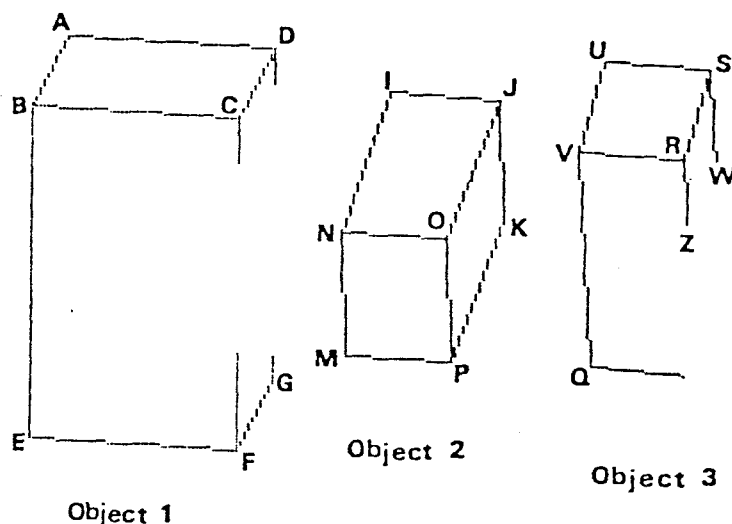


Fig.4.1 A triangle OA_1B_1 rotates about an axis to OA_2B_2 .



OBJECT 1

3D-co\Junction	A	B	C	D	E	F	G
x-co	0	-13	58	71	-13	58	71
y-co	0	-24	-28	-4	-137	-141	-117
z-co	0	66-1	78-2	12-2	26+1	37-0	-29+1

OBJECT 2

3D-co\Junction	I	J	K	M	N	O	P
x-co	-19	16	18	-34	-34	0	3
y-co	49	47	6	-39	3	0	-41
z-co	-87+1	-82-2	-105-0	-30+1	-6+2	0	-24+1

OBJECT 3

3D-co\Junction	Q	R	S	U	V
x-co	-30	0	8	-28	-36
y-co	-71	0	30	32	0
z-co	-67+1	0	-36+1	-42+0	-6+2

Fig.4.2 Three-dimensional parameters of vertices on the three objects at the first frame of the image sequence shown in Fig.1. The parameters are shown as the relative positions to the temporal origin on each object. The z-coordinates are shown with the errors evaluated from their true values.

the solutions are unreliable and, moreover, they may not have real values. In order to obtain more reliable solutions, we analyze the transitions of non-coplanar four points on the object instead of the three points, and examine the reliabilities of the solutions. We utilize all the possible combinations of frames in which these four points are visible, and discard the solutions of which reliability indices are less than a threshold. The other solutions multiplied with the weights proportional to the reliability indices are averaged to obtain reliable z-coordinates of these points. The detailed procedure is shown in Appendix I. The z-coordinates of the vertices in the later frames are easily obtained by (2) and (3).

The method is applied to the image sequence of Fig.2.1, and reasonable solutions are obtained for most vertices. However we cannot obtain solutions of some junctions, because the reliabilities of the solutions are too low or there exist less than two frames in which the points are visible. We try to evaluate the z-coordinates of these points by utilizing the 3-D parameters of other vertices on the same object. For example, the z-coordinate of Q in Fig.4.2 is not obtained at first since the number of frames available is too small (3 frames, (a),(b),(c) in Fig.2.1) and the reliabilities of the solutions are low. However, the z-coordinates of U, V, R, and S are reliably obtained. Therefore, we can determine the parameters of the plane UVRs. By using the constraints that the length of QV and the angle between QV and the plane are constant in each frame, we can calculate the z-coordinate of Q. Similarly, the orientation of lines SW and RZ are determined.

In this way, we utilize the spatio-temporal information in the image sequence to precisely evaluate the 3-D parameters of the models. Fig.4.2 shows the 3-D coordinates (relative to the temporal origin on each object) of junctions in the first frame in Fig.2.1 by this method.

4.2 MEASUREMENT OF INTERFRAME ROTATIONAL COMPONENTS

This section describes how the rotational components of the interframe movements are decided from the 3-D models in the successive frames. Fig.4.3 displays an example of block's movement. We consider that the object rotates around an axis through a temporal origin O (we can set it at any position on

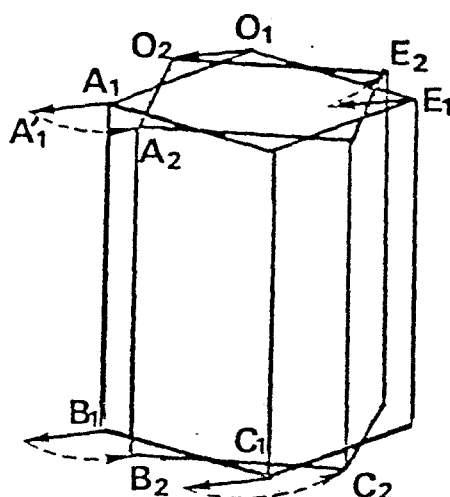


Fig.4.3 The object are first translated by a vector shown in solid arrows, and then rotated as shown in arrows in broken lines. O_1 and O_2 are the temporal origin before and after the motion.

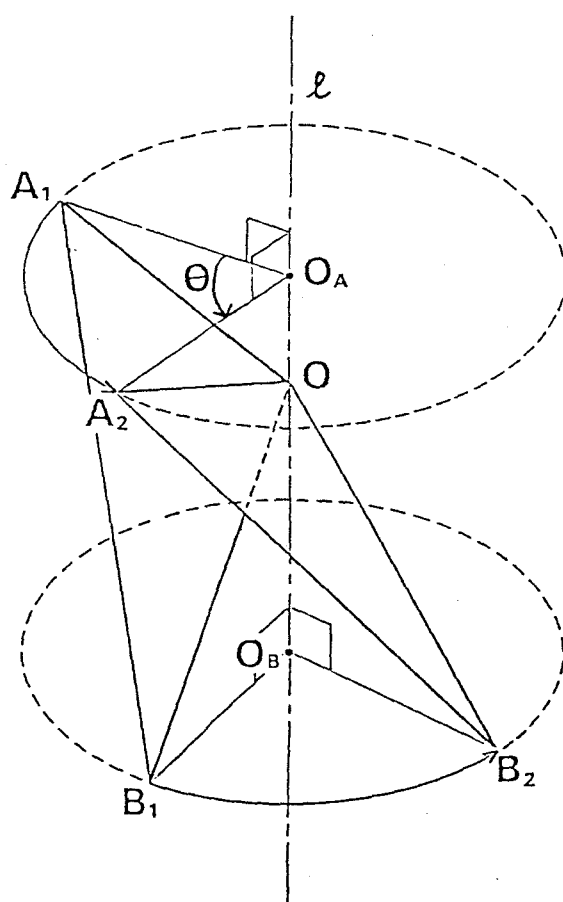


Fig.4.4 Method for determining the orientation of the rotational axis.

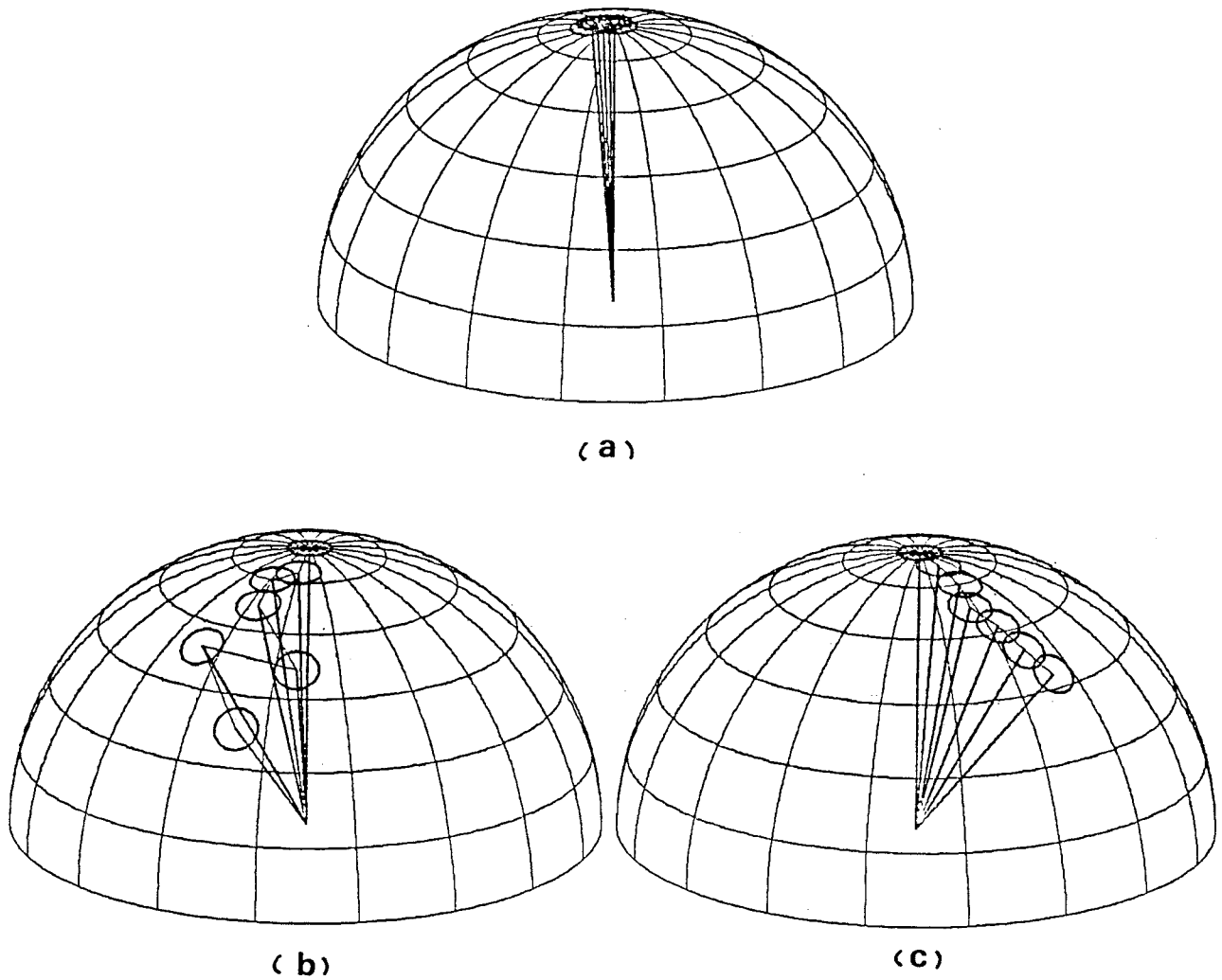


Fig.4.5 Orientations of the rotational axes.

The orientations of the rotational axes of OBJECT 1, 2, and 3 are shown in (a), (b), and (c), respectively. The locations of the small circles show the orientations in the time intervals between frames, and their sizes are proportional to the rotational angles. For convenience, the rotational axes between the first and second frames are set at the zeniths of the hemispheres.

the object) after the traslation from O_1 in one frame to O_2 in another frame as shown in Fig.4.3. Solid lines and broken lines represent the translational motion vector and the rotational one, respectively. Fig.4.4 helps to understand the method for evaluating the orientation of the rotation axis. The plane OA_1B_1 rotates around an axis ℓ through O and comes to OA_2B_2 . The points A_1, A_2 (B_1, B_2) are on a circle with its center at O_A (O_B) on the axis ℓ . Note that A_1A_2 and B_1B_2 are perpendicular to ℓ . Since the 3-D models specify the locations of these points, the orientation of ℓ is easily determined. If there are more visible vertices on the object, we use the least mean squares method to evaluate the orientation. The rotation angle $A_1O_AA_2$ is also computed by using the equation of ℓ .

Table 4.1 gives the rotational movements, the orientations of the rotational axes and the rotational angles, of each object between the consecutive frames. Fig.4.5 shows how the orientation of the rotational axis of each object changes through the image sequence. The orientation of each interval of frames is shown as a small circles (the size of the circle is proportional to the rotational angle) on a hemisphere. Note that the rotation axis of OBJECT 1 orients to an almost same direction through the whole sequence while those of OBJECT 2 and OBJECT 3 considerably change in the orientations.

Interframe	Orientation of Rotational axis	Rotational Angle (deg.)
OBJECT 1		
1-->2	(88, 158, 112)	9.5
2-->3	(90, 160, 110)	10.1
3-->4	(91, 161, 109)	10.4
4-->5	(90, 159, 111)	9.3
5-->6	(89, 161, 109)	10.3
6-->7	(90, 160, 110)	10.4
7-->8	(92, 158, 112)	9.7
OBJECT 2		
1-->2	(47, 52, 67)	-14.3
2-->3	(50, 58, 57)	-14.1
3-->4	(55, 57, 53)	-14.0
4-->5	(60, 62, 43)	-14.1
5-->6	(59, 81, 33)	-16.8
6-->7	(72, 63, 33)	-14.2
7-->8	(76, 83, 15)	-14.1
OBJECT 3		
1-->2	(46, 136, 95)	14.1
2-->3	(51, 141, 88)	14.5
3-->4	(56, 145, 82)	13.8
4-->5	(60, 146, 76)	13.7
5-->6	(66, 149, 72)	14.1
6-->7	(73, 152, 69)	14.2
7-->8	(80, 154, 66)	14.1

Table 4.1 Parameters of rotational components.

5. UNDERSTANDING OF 3-D MOTIONS

The final goal of our system is to understand the object's 3-D motions from the 2-D image sequence. We consider that the understanding of motion is not to describe a record of the interframe changes but to interpret them into a simple and natural representation through multiple frames. We first introduce a hierarchical representation to describe a complex coincidental motion using the descriptions of interframe movements. Next, we consider a reasonable representation of motions on a plane and then extend it to the 3-D cases. Finally, consistent property of motion are used to represent the motions of each object simply and naturally.

5.1 HIERARCHICAL REPRESENTATION FOR ARTICULATED MOTIONS

Since we can reconstruct 3-D geometry of each object and evaluate its interframe movements, we can describe the motion with a sequence of these interframe movements. This type of descriptions, however, are not useful for the complex motions. Suppose that a computer vision system sees a man walking and describes the motion of his hands in this manner. Since the motion is measured and described in the camera's coordinate system, the representation is complex and it gives us little information. We, human beings, describe the motion in a compact yet informative form such that a walking man swings his hands. In this case, neither the global nor viewer's coordinate systems are used, and a coordinate system fixed to a moving object is adopted instead. We will apply this type of representation to our problem. Note that the representation is meaningful when we obtain also a good representation of the motion of the coordinate system. In other words, we need not only the finding of a suitable coordinate system for a simple representation of the object's motion but also the good representation of the motion of another object to which the coordinate system is fixed.

Here, we consider the simplicity of the motion representation as follows. If the properties of the motion are constant in a considerably long period, it can be described in a small number of terms. Therefore, the constancy of the two components of the motion, a translation and a rotation, is highly desirable for

the simple representation. We select two measures for evaluating the goodness of the representation; invariance of the orientation of the rotation vector and the translation vector. We will show the invariance of the orientation is very useful for finding the suitable coordinate system. The constancy of the translation is discussed in the next chapter.

If a group of objects show a coincidental joint motion such that a block is rotating around an axis fixed to another translating and rotating block as shown in Fig.2.1, the motion of the former is more complex than that of the latter. We could obtain a better motion representation, a hierarchical representation, than individual motion descriptions in the viewer's coordinate system.

In order to analyze the motion of the group, we could use the descriptions of spatial relations between objects; for example, supported-by or touched-with. Instead, we utilize the hypothesize-and-test procedure, a common strategy used in AI researches, based on the information of the movements. We find objects of which axis of rotation do not change in the orientations as candidates for main bodies of the coincidental joint motions. Next, the hypothesis that another object is a subpart of the jointly moving groups is tested; if the relative motion of the object to a candidate is simple, in the sense that the orientation of the rotational axis is constant, then these two objects are considered as moving together.

From Fig.4.5, we can easily select OBJECT 1 as the candidate of the main body because the orientation of its rotational axis is almost constant. Since OBJECT2 and OBJECT 3 may be the subparts, the system describes its motion relative to the main body and examines the constancy of its motion in the image sequence. In order to detect the relative motion of OBJECT2 and OBJECT 3 to the main body, the system rearranges its orientations. Here, the system transform the coordinate system of scenes in Fig.2.1 (b),(c),...,(h) in such a way that the main body did not rotate from the position in Fig.2.1(a). Fig.5.1 (a) and (b) show the orientations of the rotational axis of OBJECT2 and OBJECT 3 viewed from OBJECT 1. These figures imply that their motions relative to the main body are simple rotations around the invariant axis. Thus, the coincidental motion is described in a hierarchical form as shown in Fig.5.2, which is

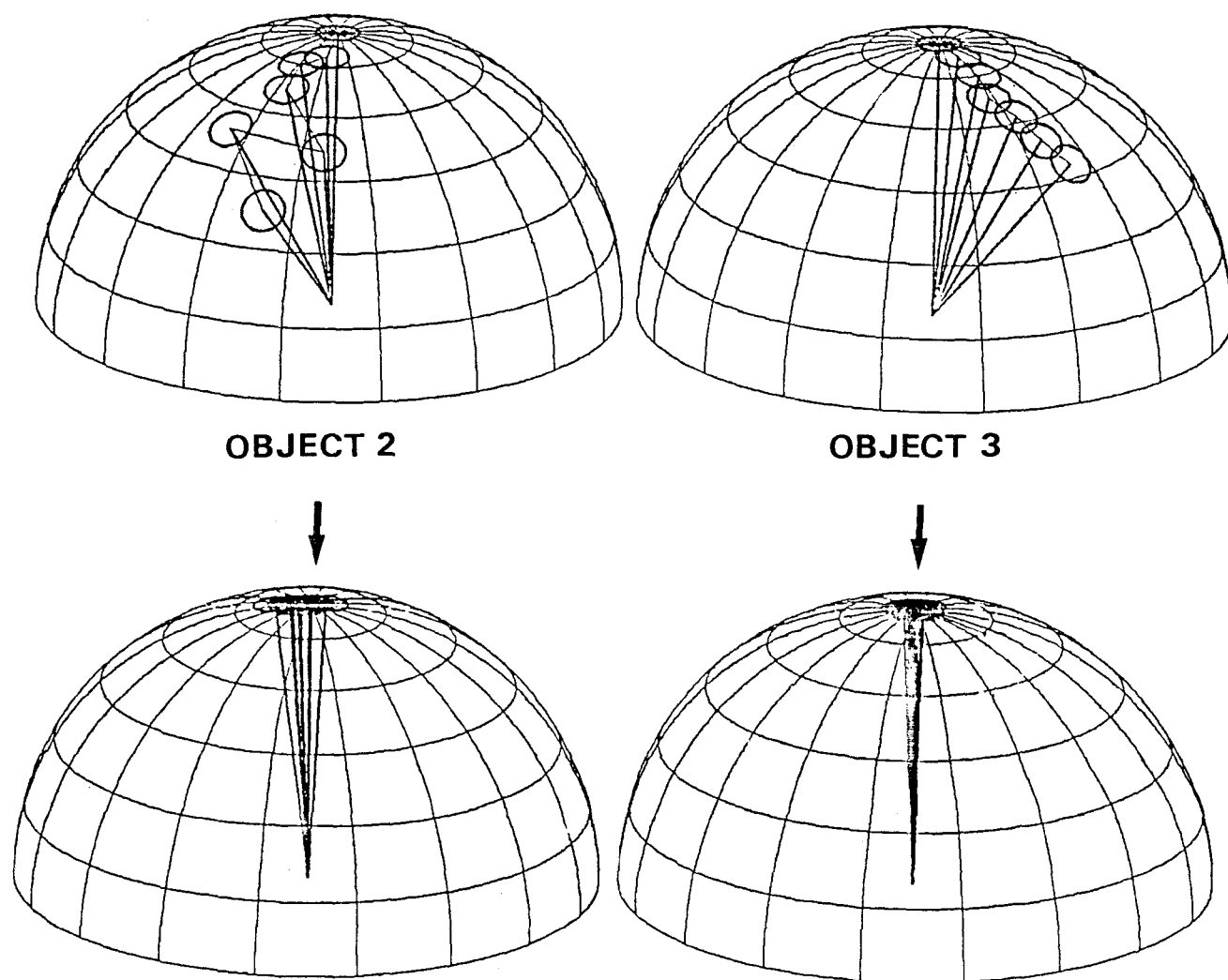


Fig.5.1 Orientations of rotational axes of OBJECT2 and OBJECT3 before and after transformation to a coordinate system fixed to OBJECT1.

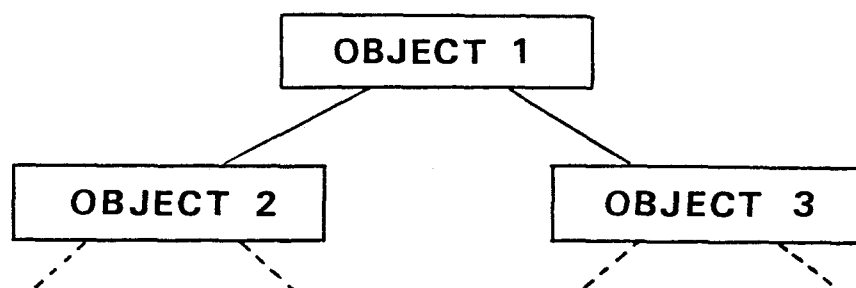


Fig.5.2 Hierarchical representation of jointly moving objects.

similar as the motion description by human beings.

5.2 MOTION DESCRIPTER AND CONSISTENT PROPERTY

In this section, we find the simple and natural representations of the motions, particularly about the translation using consistent properties of the motions.

5.2.1 REPRESENTATION OF 2-D MOTIONS

At first we consider how 2-D movements between frames are represented. Fig.5.3 shows a simple example of the 2-D motions; a line segment AB moves from A_1B_1 to A_2B_2 between the first and second frames. If the vector A_1B_1 equals to A_2B_2 , there exists a pure translation as shown in Fig.5.3 (a). Otherwise, we can decompose the movement into a rotation and a translation. For the sake of convenience, we consider that the segment AB translates after the rotation as shown in Fig.5.3 (b).

The rotation angle is uniquely determined as the angle between A_1B_1 and A_2B_2 . The translation, however, is not unique but depends on the location of the center of rotation which we can set at any location as shown in Fig.5.3 (b) and (c). It seems reasonable to select the simplest description such that the magnitude of the translation is minimum. We can make the translation zero (the proof is easy), and the center of rotation is uniquely determined as shown in Fig.5.3 (d). Thus, the 2-D movement between two frames is represented by either a pure rotation or a pure translation.

Next, we examine whether the above representation is applicable to motions through multiple frames. Fig.5.4 shows a motion of a line AB on a plane through three frames. A_1B_1 , A_2B_2 and A_3B_3 represent the locations of AB in the first, second and third frames, respectively. If we apply the interframe representation first to the movement between the first and the second frames, and then to the second and the third frames, the center of rotation suddenly jumps from C_{12} to C_{23} at the second frame, and this representation is unnatural. It seems more natural that the translation of the center of rotation has a consistent property through three views. The simplest consistent property is that the velocity is constant, and we can select the location of the center at the first frame such that it moves at a constant velocity from the first to the third frames by the

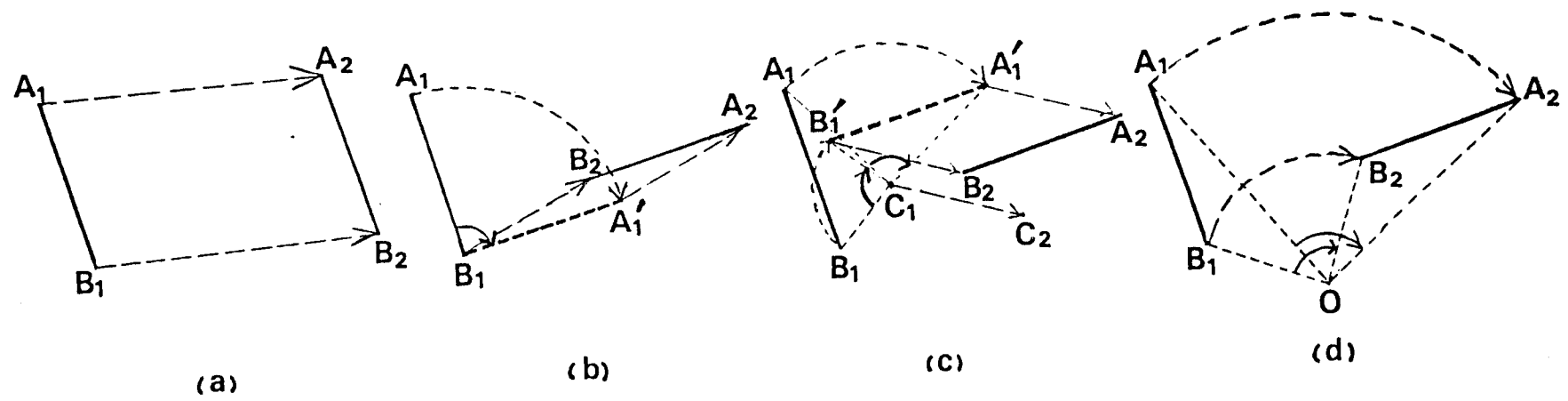


Fig.5.3 Representations of 2-D movements between two frames.

method described in later. O_1 , O_2 and O_3 in Fig.5.4 show the locations of the center of rotation determined by this method. Thus, we have a natural representation of the movements through three views that the line segment rotates around a point while the point moves at a constant linear velocity.

Let us study how we can select a reasonable representation of movements through n views. Let $A(i)$ and $O(i)$ be 2-D vectors which represent the locations of point A and the center of rotation O in the i th frame. Let $L(i) = \text{col}(L_x(i), L_y(i))$ be the traslation vector from the i th to $(i+1)$ th frames.

We have the following equations of mapping A and O in the i th frame to those in the $(i+1)$ th frame.

$$A(i+1) = T(i)(A(i) - O(i)) + O(i) + L(i) \quad (12)$$

$$O(i+1) = O(i) + L(i) \quad (13)$$

where,

$$T(i) = \begin{pmatrix} \cos\theta(i) & -\sin\theta(i) \\ \sin\theta(i) & \cos\theta(i) \end{pmatrix}$$

From (12) and (13), we have

$$A(i+1) = T(i)A(i) + (I - T(i))\left(\sum_{K=1}^{i-1} L(K) + O(1)\right) + L(i) \quad (14).$$

Thus, there are $2(n-1)$ equations and $2n$ unknown variables (two variables of $O(1)$ and $2(n-1)$ variables of $L(i)$). We can therefore select additional two relations between the unknown variables such that the motion description has a consistent property.

When the number of frames is three, there exist four mapping equations against six unknowns; the translational components $L_x(1), L_y(1), L_x(2), L_y(2)$ and x, y -coordinates of the center of rotation. Then we adopt next two additional relations,

$$L_x(1) = L_x(2) \quad (15)$$

$$L_y(1) = L_y(2) \quad (16)$$

which represent the constancy of the velocity, and the solution of (14), (15) and (16) gives the parameters of the representation.

Now we consider another example, the motion reconstruction from four views. In this case, there are eight unknown variables; $O_x(1)$, $O_y(1)$, $L_x(1)$, $L_y(1)$, $L_x(2)$, $L_y(2)$, $L_x(3)$ and $L_y(3)$. There are two alternatives to make the representation reasonable as shown in Fig.5.5. If we use the following two

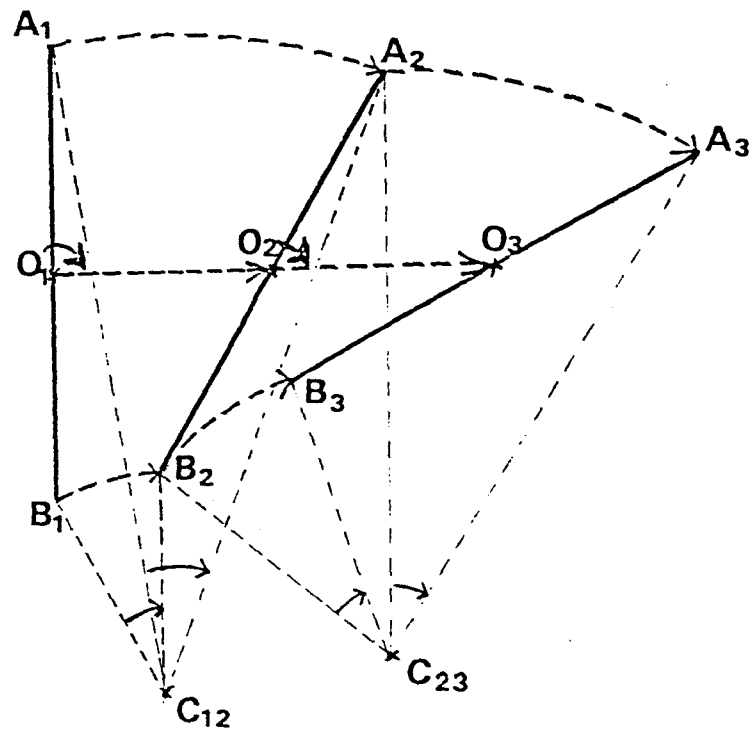


Fig.5.4 Representations of 2-D motion in consecutive three frames.

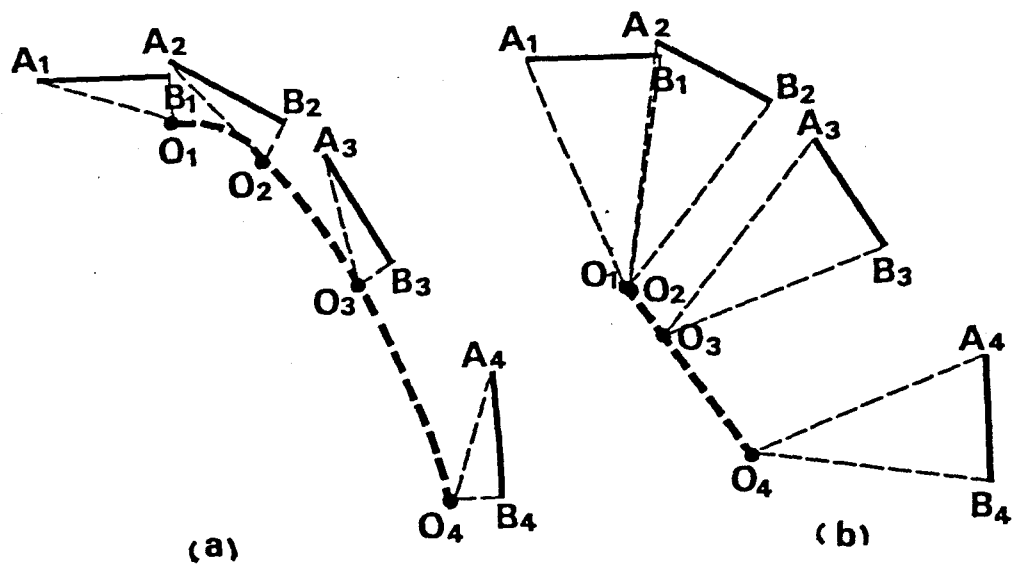


Fig.5.5 Two interpretations of a 2-D motion in consecutive four frames.

relations between the unknown variables

$$L_x(3) - L_x(2) = L_x(2) - L_x(1) \quad (17)$$

$$L_y(3) - L_y(2) = L_y(2) - L_y(1) \quad (18)$$

then the acceleration of the translation is constant. (Fig.5.5 (a))

Instead,

$$L_y(2) = (L_y(1)/L_x(1)) * L_x(2) \quad (19)$$

$$L_y(3) = (L_y(1)/L_x(1)) * L_x(3) \quad (20)$$

are used, then the direction of the translation is constant. (Fig.5.5 (b)) We need to use other knowledge on the object's motion to judge which representation is more adequate.

5.2.2 RECONSTRUCTION OF 3-D MOTIONS

Now, let us consider a method for reconstructing 3-D motions from a 2-D sequence of a limited number of frames. We assume that the orientation of the rotational axis and the rotational angle are given by the method described in the chapter 4. The representation of 3-D motion is very similar as that of 2-D motion; the translation is not unique but depends on the location of the rotational axis. We, therefore, find reasonable representations based on the method described in the previous section.

Since we reconstruct 3-D motions from a 2-D image sequence, the reconstructed motions depend not only on the number of views but also on the projection type used. The 3-D models reconstructed from the orthogonal projections do not give any information about the translation in depth, while the locations and orientations of the object to the viewer are available from the sequence of the central projections if the scale factor is known.

Another problem arises, in the case of more than three views, that the orientation of the rotational axis in some cases changes. In the previous section, the computer constructs the hierarchical representation of the moving objects, and such motions are reduced to the combination of the simple rotations. The following analysis assumes the invariance of the orientation.

MOTION RECONSTRUCTION FROM CENTRAL PROJECTIONS

This section discusses the problem of the 3-D motion

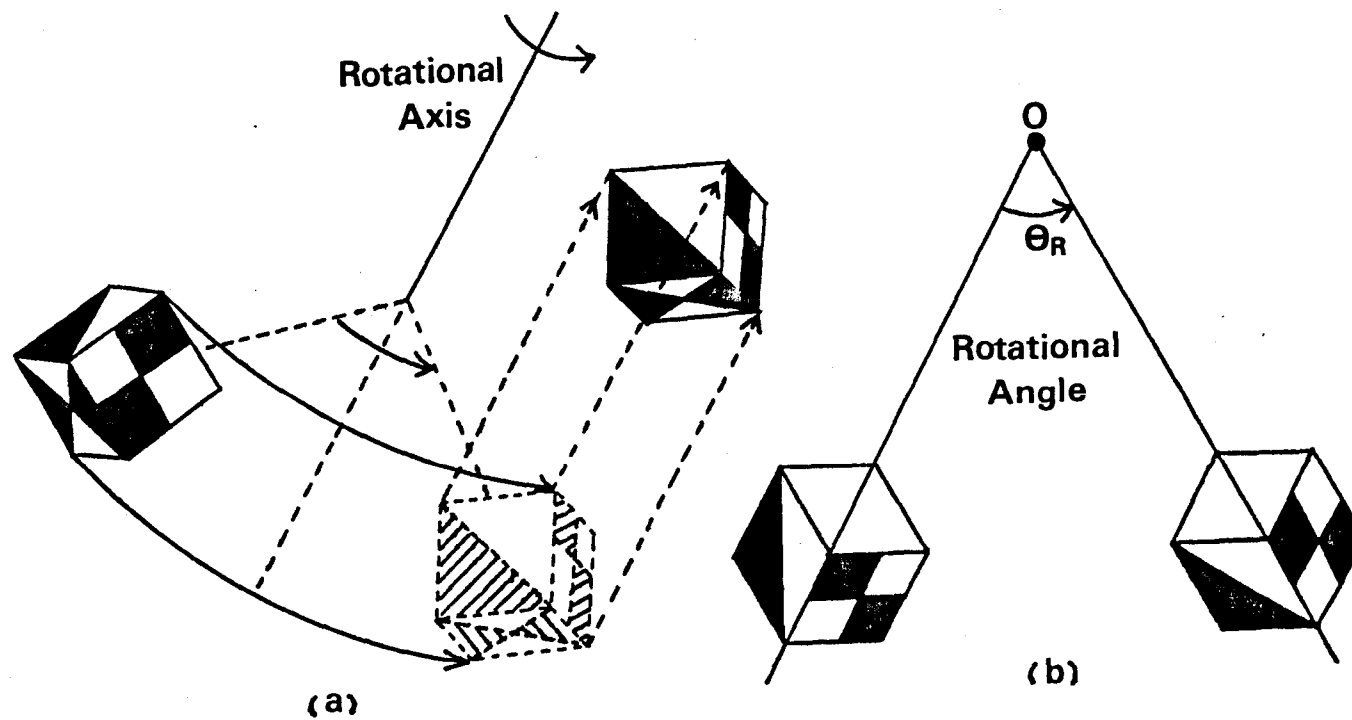


Fig.5.6 Representation of 3-D movement between two frames.

reconstruction from a sequence of the central projections of the scene. We advance the discussion by assuming that the scale factor is known; the positions and orientations of the object to the viewer in all frames are known.

[I] FROM TWO VIEWS : A movement between two frames is decomposed into one rotation and one translation, and the translation is not unique but depends on the position of the rotation axis. We can uniquely determine the translation if other constraints are introduced.

Now we analyze the 3-D movements in the scene by extending the method developed for the 2-D motion representation. Since the object rotates around an axis of which orientation is already known, we transform the xyz-coordinate system to the XYZ-coordinate system whose Z-axis is parallel to the rotation axis. By this transformation, the rotation around the axis in the xyz-space (Fig.5.6 (a)) is reduced to the rotation around a point O on the X-Y plane as shown in Fig.5.6(b), and the movement along the Z-axis consists of a pure translation.

For the given two frames, there are five unknowns; translational components ($L_x(1), L_y(1), L_z(1)$) and ($O_x(1), O_y(1)$), the position of the rotational axis. $L_z(1)$ is uniquely determined from the 3-D models of the first and second frames, because the location of the rotational axis does not change it. Thus, four unknowns are left like the case of 2-D motion. We can uniquely determine ($O_x(1), O_y(1)$) by adding two relations that $L_x(1)$ and $L_y(1)$ are zero. This representation is meaningful because it is very simple and natural. Thus, the 3-D movements between two frames are represented by a rotation around a certain axis and a translation parallel to this axis.

[II] FROM MULTIPLE VIEWS : Let us study how we can select a reasonable representation of 3-D motions when the number of views increases. Generally speaking, we have $3(n-1)$ linear equations which map points in the first frame onto the second, third, ... and nth frames if n frames are given. There are $3(n-1)+2$ unknowns; two unknowns specifying the location of the rotational axis at the first frame and $3(n-1)$ unknowns of translations between consecutive two frames. We consider the relations between them in the above-mentioned O-XYZ coordinate system. We can eliminate $(n-1)$ unknowns, because $L_z(i)$ ($i=1, \dots, n-1$) are uniquely determined by the transformation.

Therefore, $(n-1)$ linear equations about the translation along Z-axis are not needed. The result is similar to the 2-D motion: there exist $2n$ unknowns and $2(n-1)$ linear equations, and we can use additional two relations between the unknowns to make the motion description have a consistent property.

Now we consider an example, the motion reconstruction from three views. In this case, there are six unknowns; $(L_x(1), L_y(1))$, $(L_x(2), L_y(2))$, the translational components on the X-Y plane, and $(O_x(1), O_y(1))$, the location of the rotational axis in the first frame. If we use the following two relations between the unknown variables

$$L_x(2) = (L_z(1)/L_z(2)) * L_x(1) \quad (21)$$

$$L_y(2) = (L_z(1)/L_z(2)) * L_y(1) \quad (22)$$

then the direction of the translation is constant. If the translation along the Z-axis is constant through the three frames, then

$$L_x(2) = L_x(1) \quad (23)$$

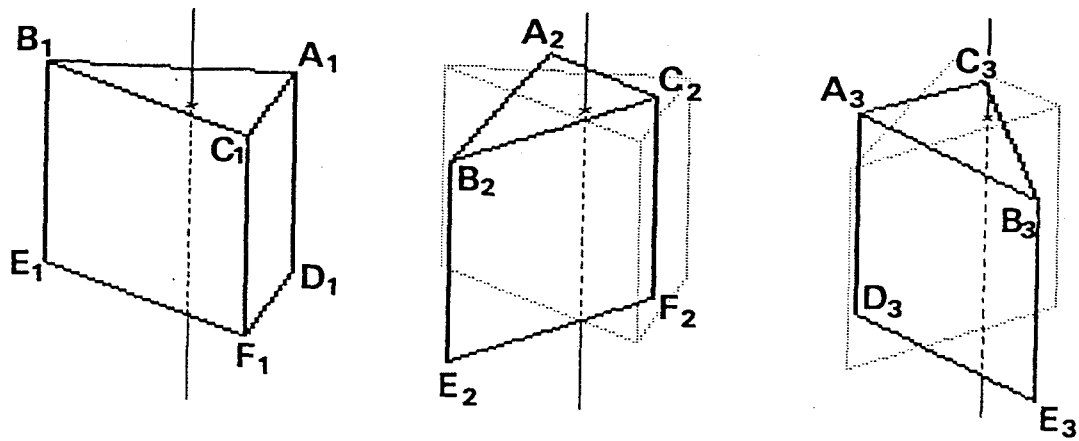
$$L_y(2) = L_y(1) \quad (24)$$

and the translation is constant through the sequence.

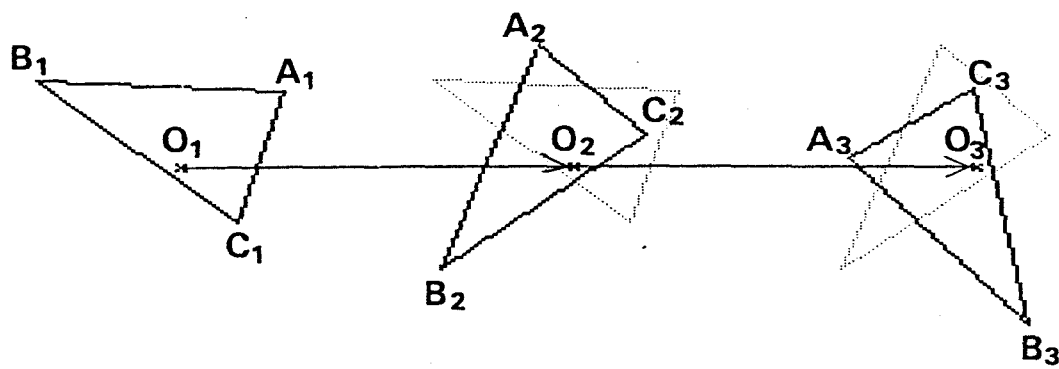
MOTION RECONSTRUCTION FROM ORTHOGONAL PROJECTIONS

Now, let us consider how the computer can represent the 3-D motions given a certain number of images of the orthogonal projections. Since the projection type is orthogonal, the z-components of the object movements are not detectable. Let us regard these components as being zero and proceed the analysis. We can, thus, determine the X,Y coordinates of points on the object and obtain a representation of the interframe movements. The introduced assumption, no movement in depth, results in some strangeness in the description of motion, but it resembles unnatural perception we know by experience. If we imagine observing a person running to the direction in depth at a long distance through a telescope, the image is an almost orthogonal projection of the scene and our perception would be that he does not make any approach to us although he actually does.

Let $(x(i), y(i))$ be the location of the projection of a point on the object to the i th image. The X,Y,Z coordinates of the point are determined as



(a)



(b)

Fig.5.7 Representation of 3-D motion between three frames.

$$\begin{pmatrix} X(i) \\ Y(i) \\ Z(i) \end{pmatrix} = P \begin{pmatrix} x(i) \\ y(i) \\ z(i) \end{pmatrix} \quad (25)$$

where P is the transformation matrix from the 0-xyz coordinate system to the 0-XYZ coordinate system. Let $O(I)$ be a vector representing the location of the rotational axis in the X-Y plane at the i th frame, and $L(I)$ be a displacement vector of the temporal origin on the object. We have the following very similar equations to those for 2-D motions.

$$\begin{pmatrix} X(i+1) \\ Y(i+1) \end{pmatrix} = T(I) \begin{pmatrix} X(i) \\ Y(i) \end{pmatrix} + (I-T(I)) \left(\sum_{K=1}^{I-1} L(K) + O(I) \right) + L(I) \quad (26)$$

If n frames are given, we have $2(n-1)$ equations and $2(n-1)+2$ unknowns; translational components $(L_x(i), L_y(i))$ and the X,Y coordinates of the rotational axis at the first frame. In a similar way to the 2-D case, we use additional two constraints of the unknowns. The equations (15) and (16) which constrain the velocity to be constant in three views are used again. Thus we have a representation that object rotates around the axis which moves at a constant velocity as shown in Fig.5.7.

When more than three views are given, more general property is considered.

5.2.3 MOTION REPRESENTATION WITH CONSISTENT PROPERTY

From above discussions, we can see that the more frames are given, the reconstructed motion is more general and not simple. Therefore, we consider that the constant velocity in three frames is the most fundamental property for the simple representation of motion. That is, when more than three views are given, the movements in all consecutive three frames are described by the constant linear velocity. If the translational components are almost same, the motion through the whole views is a rotation and a translation of the constant velocity. If there exists a considerable change in a linear velocity, we proceed to analyze them and represent in a more general motion description such as a constant acceleration. Thus, we utilize the hypothesis-and test procedure and segment the image sequence into several parts which have different consistent properties of motion each other.

Let us show an example applied to the scene as shown in Fig.5.8. In this figure, a rotating object moves from the

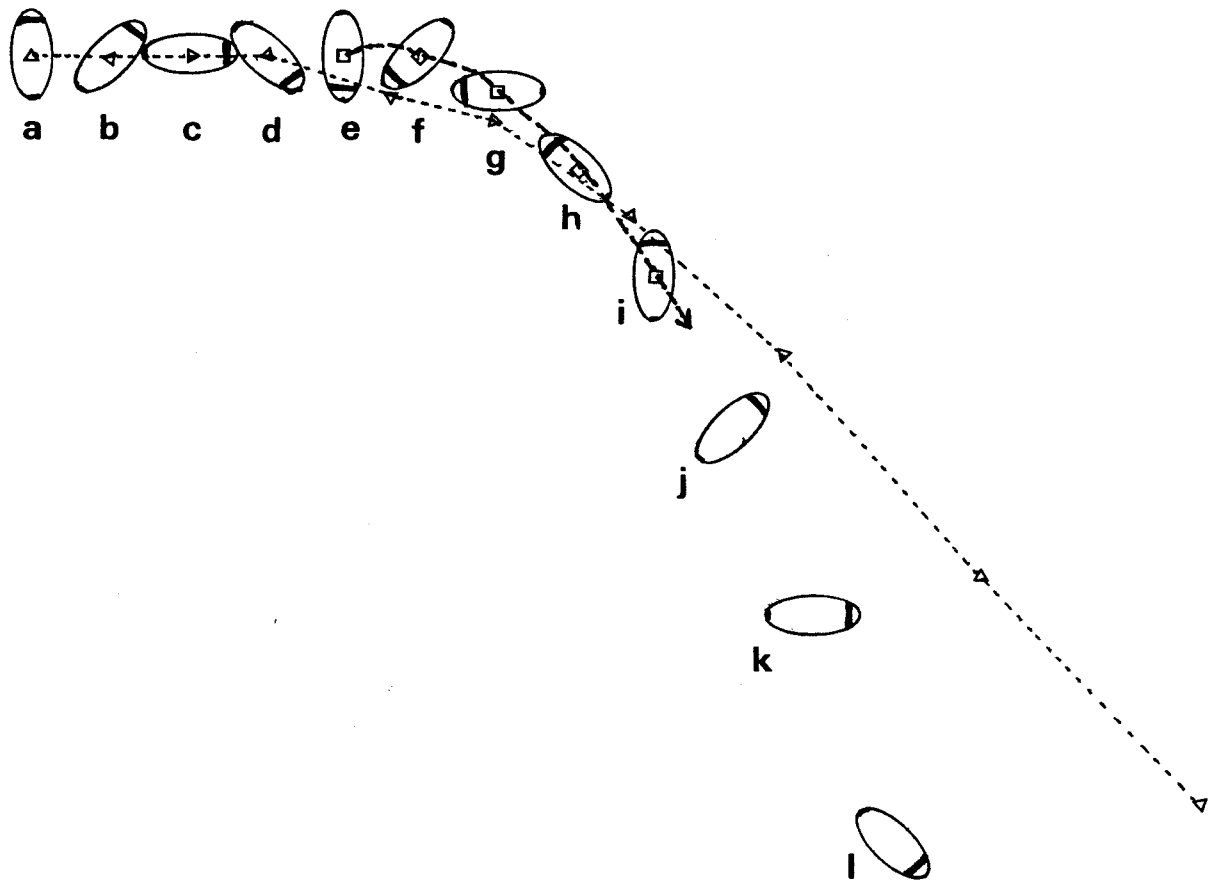


Fig.5.8 Input image sequence of 2-D motion.

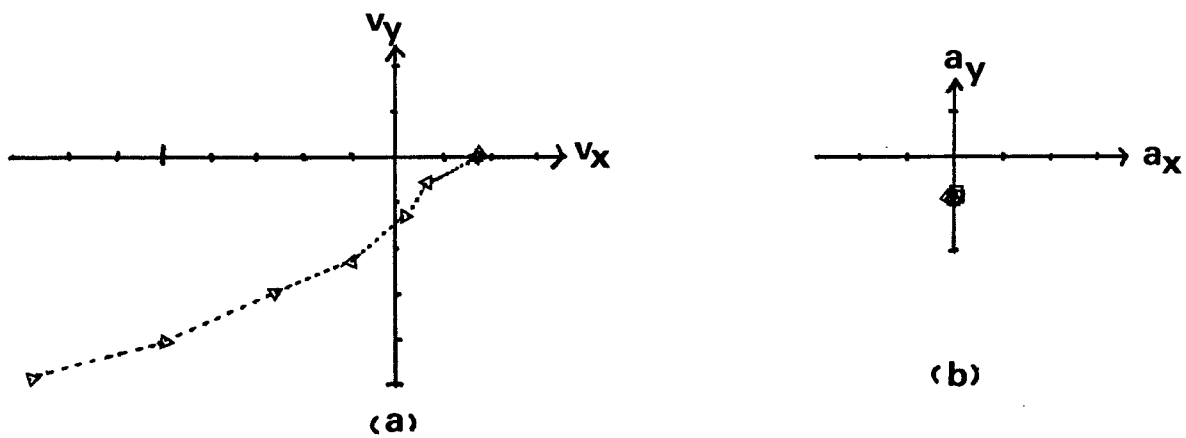


Fig.5.9 x-y components of constant translation (a) and constant acceleration.

upper-left to the right direction and it suddenly changes the direction of its motion to the lower-right at the center of picture as if it falls from a cliff. At first the movement in all consecutive three frames ((a,b,c), (b,c,d), ..., (j,k,l) in Fig.5.8) are described by a constant linear velocity. Small triangles indicate the position of the center of rotation at the first frame in each set of consecutive three frame, and the x-y components of the translation are displayed in Fig.5.9 (a). Both figures tell us that the components of the translation is increasing after the frame (e) and we can predict that the class of its motion changed from the constant velocity to a constant acceleration.

Then, using the equations (14), (17) and (18), the computer describe the movement in those consecutive four frames after the frame (e) in Fig.5.8, in the form of a constant acceleration. In this figure, small rectangles indicate the center of rotation at the first frame in each set of consecutive four frames and Fig.5.9 (b) shows the x-y components of the acceleration estimated. From these, we can see that the property of constant acceleration is satisfied in those frames after the frame (e) and the motion of the object in Fig.5.8 is segmented into two parts the constant linear velocity (a) to (f) and the constant acceleration (g) to (l), respectively.

5.3 EXPERIMENTAL RESULTS

In this section, we show the results which are obtained by applying the methods described in the previous section to our scenes.

5.3.1 MOTION REPRESENTATIONS OF MAIN BODY AND SUBPARTS

In the section 5.1, the system segmented the coincidental moving objects into the main body (OBJECT 1) and two subparts (OBJECT 2 and OBJECT 3) using the constancy of the orientation of the rotational axes. Here, we show the results applied the above-mentioned method to each object.

We, at first, represent the motion of the main body. The movement in all consecutive three frames in Fig.2.1 are described by the constant linear velocity. Table 5.1 shows the parameters of this description. The location of the rotational axis (OX,OY) is represented in the O-XYZ coordinate system

interframe	(L_x , L_y)	(O_X , O_Y)
(a)→(c)	(8.5, 0.4)	(48.4, 28.5)
(b)→(d)	(8.3, 0.3)	(45.2, 23.2)
(c)→(e)	(7.6, -0.5)	(43.5, 22.7)
(d)→(f)	(8.9, -1.2)	(49.3, 24.5)
(e)→(g)	(9.8, -2.1)	(43.9, 21.5)
(f)→(h)	(7.1, 1.8)	(42.6, 22.9)

Table 5.1 Parameters of translational components and the position of rotational axis of OBJECT 1.

relative to the main body at the first frame and the x-y components of the translation are represented on the image plane. From this table, we can see that OBJECT 1 (the main body) rotates to right around an axis through its top surface and translates to rightward.

Next, let us consider to represent the motion of subparts. In the case of subparts, the computer has to consider the consistent property of their motions relative to the main body. To simplify the parameters of the motion of a subpart, we transform the 0-xyz coordinate system (its x-y plane is same as the image plane) to the 0-XYZ coordinate system whose Z-axis is parallel to the rotational axis of a subpart at the first frame. The following parameters are considered to represent the motion of the subpart as shown in Fig.5.10.

Z_1 : the distance in the direction of Z-axis between the main body and the subpart.

OX, OY : the coordinates of the rotational axis of the subpart.

$L_x(i), L_y(i), L_z(i)$: the translational components of the subpart relative to the main body from the i-th frame to the (i+1)-th frame. Although the number of translational components of main body is two, that of subparts is three because each component of the translation affects the position of the subpart on the image plane by the rotation of the main body.

Thus, there are $2(n-1)$ mapping equations against $3n$ unknowns (Z_1, OX, OY and $3(n-1)$ translational components $L_x(i), L_y(i), L_z(i)$) if n frames are given. We can, therefore, select additional $(n+2)$ relations between the unknowns.

[case A] When the number of frames is two, there exist two mapping equations against six unknowns. Then we adopt next four relations,

$$Z_1 = C_1 \quad (C_1 : \text{constant}) \quad (27)$$

$$L_x(1) = L_y(1) = L_z(1) = 0 \quad (28), (29), (30)$$

which represent the motion of subpart by only a rotation at a certain fixed position to the main body.

[case B] If three frames are given, there exist four mapping equations against nine unknowns. Then, we adopt next five relations among the unknowns,

$$Z_1 = C_2 \quad (C_2 : \text{constant}) \quad (31)$$

$$L_x(1) = L_x(2) \quad (32)$$

$$L_y(1) = L_y(2) \quad (33)$$

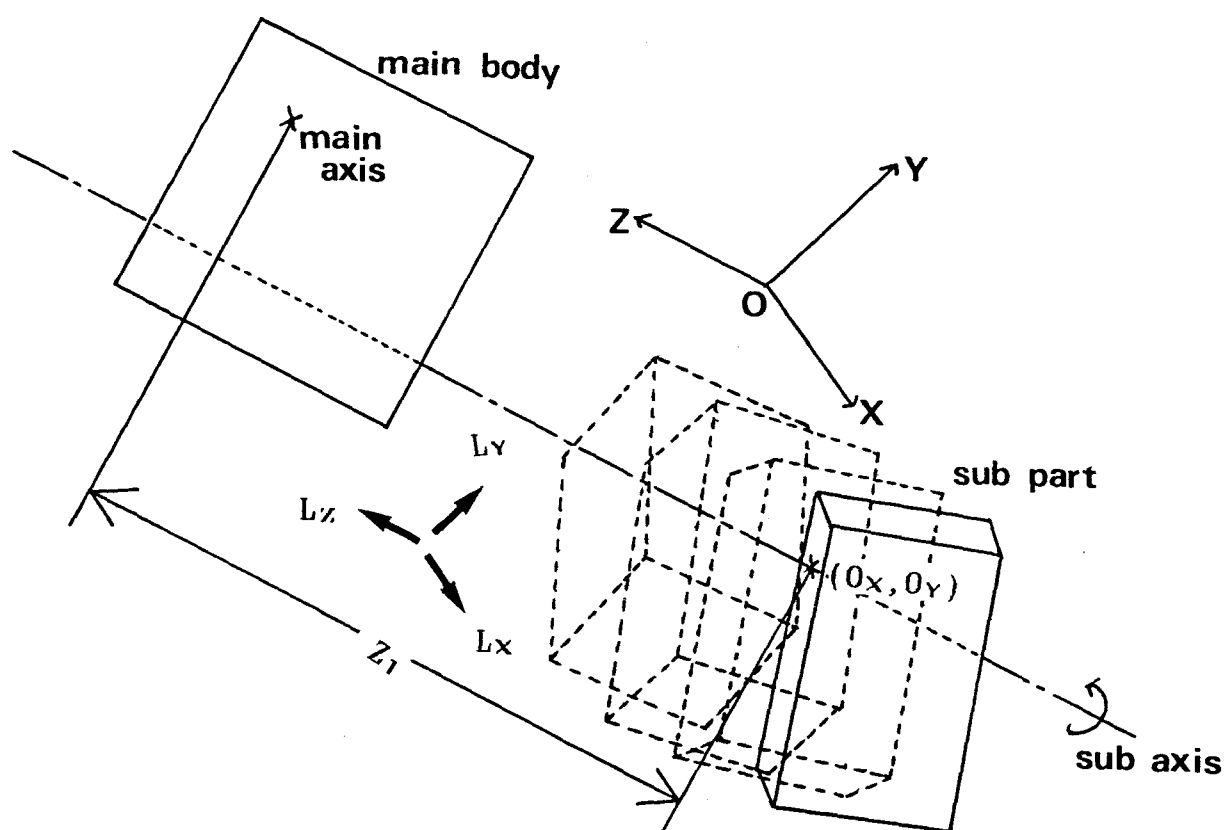


Fig.5.10 Motion parameters of a subpart to the main body.

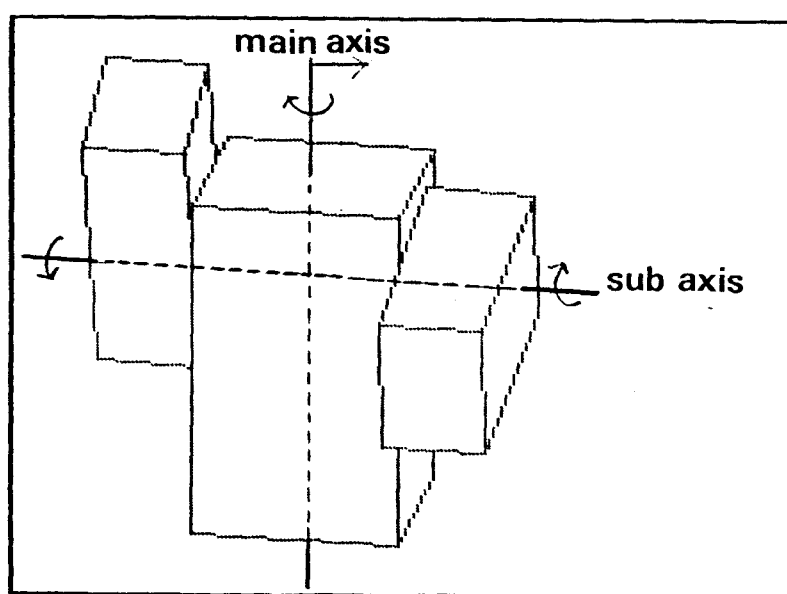


Fig.5.11 Model of joint motion obtained by analysis.

$$L_z(1) = L_z(2) = 0 \quad (34)$$

which represent the translation of the subpart by the constant linear velocity on the plane perpendicular to the rotational axis of the subpart and no translation in the direction of this axis. We can often see such a motion in some arms of a certain industrial robot.

[case C] When the number of the frames is four, there exist six mapping equations against twelve unknowns. Then, we use following six relations,

$$L_x(1) = L_x(2) = L_x(3) \quad (35)$$

$$L_y(1) = L_y(2) = L_y(3) \quad (36)$$

$$L_z(1) = L_z(2) = L_z(3) \quad (37)$$

which represent the constancy of the velocity.

Though, in the case of articulated motions, subparts which is often called "arms" are considered to move such a way that explained in the case B, we use the constancy of the translation to the main body (case C) since the system use it in the case of the main body. Then, the computer calculates each parameter against all consecutive four frames ((a,b,c,d), (b,c,d,e), ..., (e,f,g,h) in Fig.2.1). From these calculations, we can see that the all components of the translation are almost zeros and that each subpart is nearly touching with the main body and the rotational axes of the two subparts (OBJECT2 and OBJECT3) are almost same.

As a final result, the system obtains enough information to interpret the dynamic scene shown in Fig.2.1 into the description that OBJECT 1 turns right with OBJECT 2 and OBJECT 3, and OBJECT 2 revolves up at one side of OBJECT 1, OBJECT 3 revolves down at the opposite side as shown in Fig.5.11.

5.3.2 SEGMENTATION OF IMAGE SEQUENCE BY MOTION

In the case of the scene as shown in Fig.2.1, the constancy of the translation is satisfied through the whole frames. Generally speaking, however, moving objects do not always keep their consistent motion,

but often change the properties of their motions. Here, we show an example of segmenting image sequence by the consistency of the motion. Fig.5.12 shows the input image sequence in which the three moving objects are moving jointly as if a man walks and turns the corner with his arms swinging. In this case, the 3-D

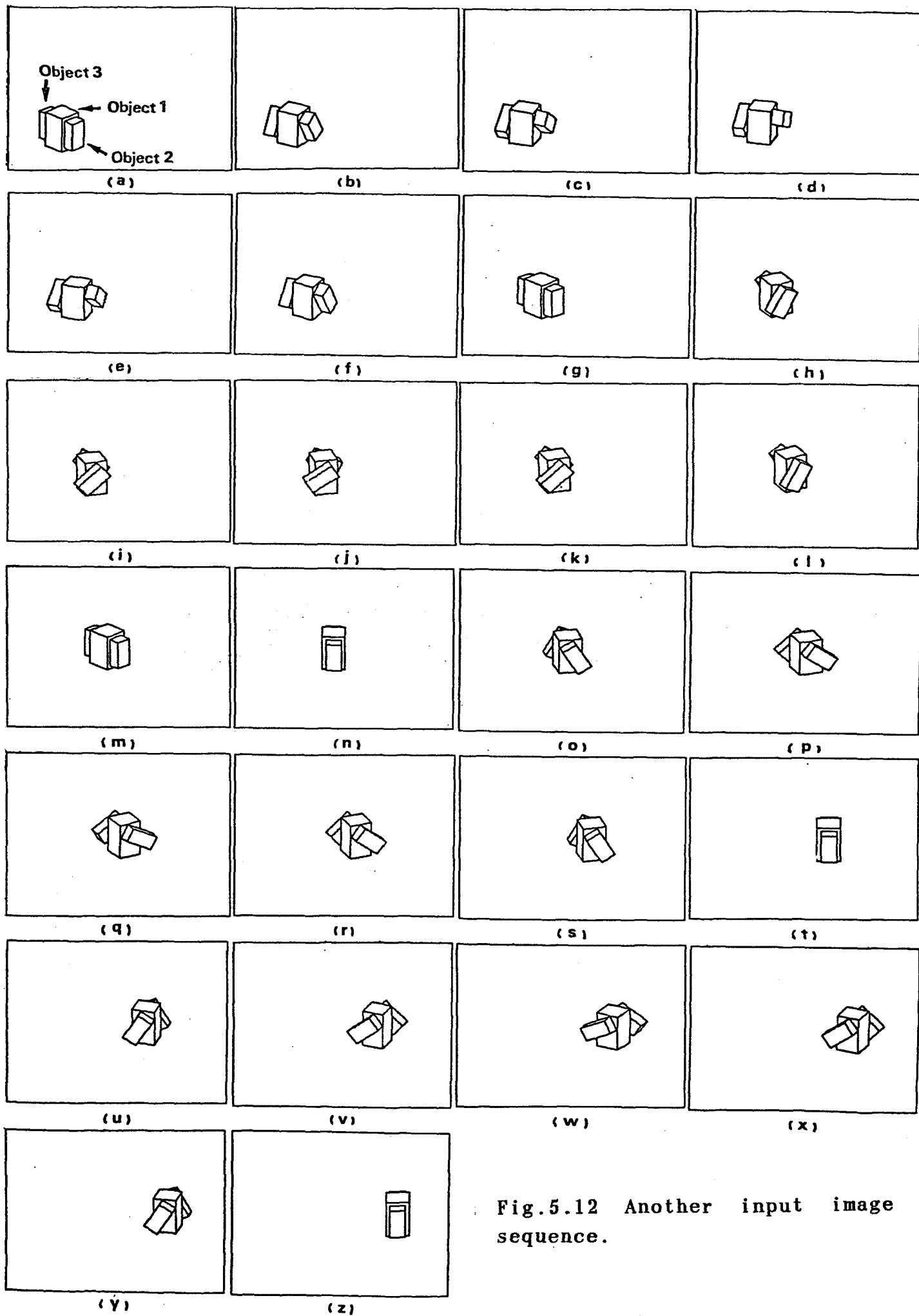


Fig.5.12 Another input image sequence.

geometries of OBJECT 1 and OBJECT 2 are obtained by the method described in the section 4.1, but that of OBJECT 3 cannot be obtained because the visible part of OBJECT 3 is too small. Therefore, only OBJECT 2 is a subpart of the main body (OBJECT 1).

Fig.5.13 and 14 show the results of the application of the representation with constant velocity of the main body to every three consecutive frames. In Fig.5.13, the coordinates of the rotational axis in the first frame of the three consecutive frames are shown in the transformed coordinate system whose Z-axis is parallel to the rotational axis. Also, x and y components of the evaluated constant linear velocity in every three frames are represented in Fig.5.14 on the coordinate system of the image plane. Both figures tell that those consecutive three frames including (m) and (n) in Fig.5.12 have different values from other combinations of consecutive three frames.

Since those frames before frame (m) and after frame (n) can be simply represented by the constant linear velocity, the computer does not apply more general property but interprets it as a considerable change from (m) to (n) and segments the image sequence into two parts; from (a) to (m) and from (n) to (z).

As a result, the computer understands and represents the dynamic scene into simple and natural descriptions of the motion similar to what a human would perceive from the given input.

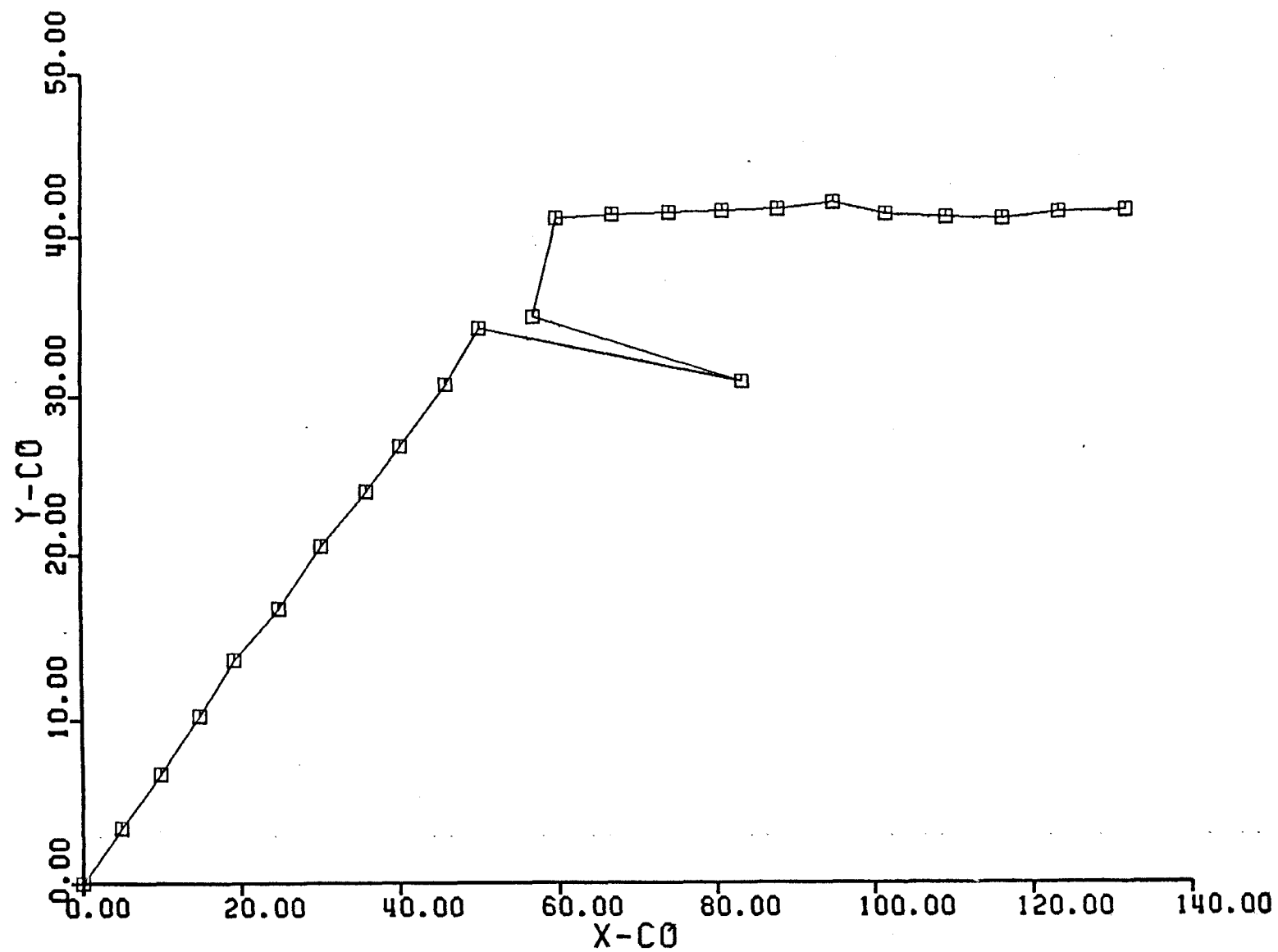


Fig.5.13 The coordinates of the rotational axis evaluated by assuming the constant linear velocity.

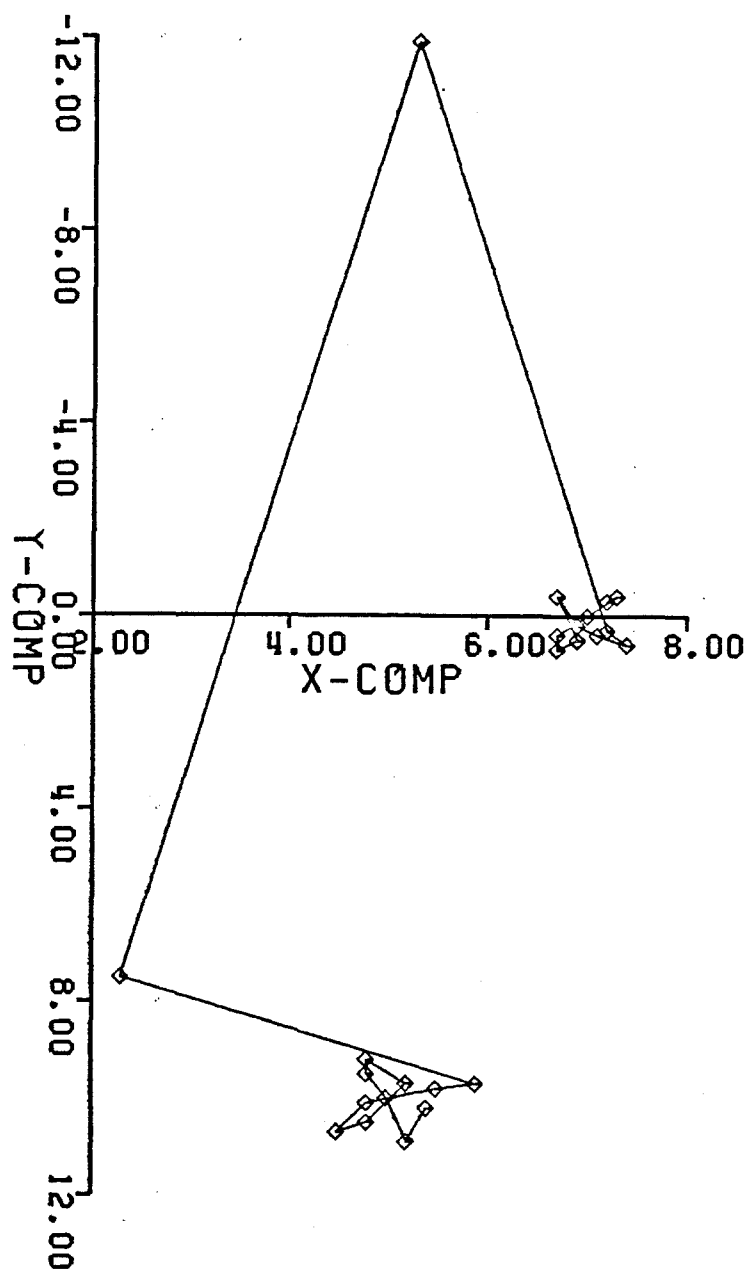


Fig.5.14 x and y components of the evaluated constant linear velocity.

6. DISCUSSION

6.1 PROBLEMS IN EACH PHASE

The described scheme of understanding the 3-D motions in the image sequence consists of three phases; the analysis of changes in structure of the line images, the reconstruction of 3-D geometry of each object, and the motion representation. We do not provide the scheme with capability of communicating each other. In order to analyze raw image data instead of the noise free line images we used, facilities like the blackboard are useful, because we need to integrate spatio-temporal information obtained from a number of subsystems. Even for the noise free line images, the communication between the first and second phases would be useful, because the information of the 3-D parameters could considerably help the object matching process.

The object-to-object matching method uses heuristics which cannot guarantee the matching is always correct, and further studies are needed.

The experiments of reconstruction of 3-D geometry assume that the errors in finding the locations of the junctions in the input images are very small (less than a half pixel size). Although the experimental results are satisfactory, we need further theoretical and experimental studies because the locations of the junctions in the line drawings extracted from the raw image data have much larger errors. Particularly, in describing the motion of subparts, parameters of their motions cannot be correctly obtained unless the motion of their main body can be described exactly.

In this paper, we utilize the consistency of the translation vectors and the constancy of the direction of the rotation vectors as measures for the goodness of representation. Therefore, we cannot deal with other classes of motions such as periodic motions. Methods for representing a more wide class of motions should be developed.

6.2 FUTURE RESEARCH

There are two methods which project a 3-D scene onto the 2-D image plane. One is the central projection and the other is the orthogonal projection. The former is a real projection through a

camera lens and the reconstruction process of 3-D geometry from 2-D images by this projection is more complex than the latter because the computer has to solve non-linear equations. The latter is an approximate projection of a real scene at a long distance through a telescope and it cannot detect the translation in the depth direction. We can, however, easily understand a 3-D world from images by the orthogonal projection and the reconstruction process of 3-D geometry is simple, that is, the computer only solve the quadratical equations.

We, human beings, seem to use these two projections well ; Our brains have 3-D models for individual objects by the orthogonal projections and we perceive the spatial locations of objects by changing a scale factor on the retinal image. In order for a computer to understand 3-D motions of object including translations in the depth direction, a new method can be considered that analyzes 2-D images by assuming the orthogonal projections locally and the central projections globally, instead of using only the central projections.

To develop such a method, we first have to investigate the relation between the errors in the solutions of 3-D geometries for the objects obtained by assuming the orthogonal projections and their motions (particularly their translation in the depth direction). This problem seems us interesting and relevant to human perception of both the spatial locations and the motions of objects in a 3-D world.

We have dealt with line drawings because of the following two reasons. One is that we want to avoid such a low-level processing as a feature extraction from gray level images, which is very time-consuming and does not guarantee desirable results. The other is that we think that only line drawings can represent physical properties of objects such as surface orientations, space curve, relations between objects and so on. For example, let us watch the line drawings as shown in Fig.6.1 ; a little girl in the popular cartoon "Dr. SLUMP ARALE-chan". Each part of this drawing represents the shape of each part of her style, well. For example, shapes of helmet, her face and fingers; folds of her mantle and so on.

Then, we consider to extend our method in this paper to more general dynamic line drawings such as animated cartoons as shown Fig.6.1. To do this, we have to study the following matters ;



Fig.6.1 An example of line drawing from a popular animated cartoon.

how a computer can assign physical meanings to each part of line drawings and take a correspondence of them between dynamic line images, how the computer can reconstruct the 3-D geometries of objects from multiple images, and what kinds of constraints can carry out these matters through the time-varying images. Those problems seem relevant deeply to our understanding of general line drawings and implicit skills of drawers of animated cartoons and painters.

7. CONCLUSIONS

We describe an image understanding scheme which can find a good representation of the dynamic blocks world given a two-dimensional line image sequence.

In the first phase, a labeling scheme and an object-to-object matching method are applied to the dynamic line images in order to segment them into individual blocks and to find correspondences of their vertices between frames. The feature of this scheme is that unlike the usual methods which use two-dimensional pattern matching techniques such as the correlation of intensity distribution, the system interprets the structural changes of line drawings as object motions in a trihedral world by using a junction transition table and contextual information between consecutive frames. They are very useful to analyze a complex scenes containing accidental alignments of a vertex to an edge which are inhibited in the static image analysis.

In the second phase, the shape rigidity property of three vertices on a block is used to evaluate geometrical parameters such as orientations and edge lengths. The advantage of our method over usual ones results from utilization of knowledge about the scene, the model of each block obtained in the previous phase, and from statistical evaluation of solutions from all possible combinations of three vertices; the solution whose reliability indices are less than a threshold are discarded and the remains are averaged to obtain reliable solutions. As a result, relative errors of estimated edge length are less than 5 percents and interframe rotational components are precisely obtained.

In the third phase, consistent properties of motion are hypothesized and verified in order to analyze the coincidental motions and to find a simple and natural representation of motion. In the analysis of coincidental motions, an object whose rotational axis does not change in the orientation is considered as the main body of a jointly moving group and the movements of other objects are tested whether they can be interpreted as the subparts of the group. Thus, the system can construct the hierarchical representation of jointly moving objects, and

interpret the visual complex motions of subparts as simple movements to the main body.

In the description of the interframe movements, the translation cannot be uniquely determined but depends on the position of rotational axis while the orientation of the rotational axis and the rotational angle are uniquely determined. That is, we can set the rotational axis at a certain place so that the translation has the consistent property through multiple frames, such as a constant velocity, a constant acceleration and so on. Then, the system use hypothesis-and-test procedure for finding consistent properties through multiple frames; if a property cannot satisfied in many frames, other properties are applied. And the image sequence is segmented into several parts which have different consistent properties each other.

Thus, the system segmented the dynamic scene into simple and natural representations similar to what a human would perceive from the given input.

Although the trihedral world seems us a very restricted domain, we can use most of the proposed ideas for analyzing real dynamic scenes. The methods used in the second and third phases do not require the objects are blocks and, therefore, are applicable to all rigid motions if the correspondences of points on surfaces of the objects are found.

APPENDIX *PRECISE ESTIMATION OF Z-COORDINATES OF VERTICES*

Suppose that a line drawing is labeled as shown in Fig.A, and we evaluate the z-coordinate of each point from four points O,A,B,C in $m>3$ views. At first, the first three frames are used. We independently analyze the transitions of two triangles OAB and OBC, and evaluate the z-coordinate of each point by using (4)-(11). Thus, the following solution pairs are obtained (they are arranged in such a way that both Z_{B-OAB} and Z_{B-OBC} are positive).

$$(Z_{B-OAB}, Z_A), (-Z_{B-OAB}, -Z_A), (Z_{B-OBC}, Z_C), (-Z_{B-OBC}, -Z_C)$$

where Z_{B-OAB} and Z_{B-OBC} are the z-coordinates of the point B obtained from OAB and OBC, respectively.

Next, the reliabilities of the solutions are examined. We consider that if the solutions are unreliable, the probability that Z_{B-OAB} and Z_{B-OBC} are almost same is very small. Therefore, we use the following reliability index.

$$R(Z_B) = 1 - |Z_{B-OAB} - Z_{B-OBC}|/C,$$

where C is a positive constant.

If $R(Z_B)$ is less than a predetermined threshold C_T , then we discard the solution and evaluate the z-coordinates by using a different triplet of the views.

If $R(Z_B)$ is larger than C_T , Z_B is obtained as the average of two solutions, or

$$Z_B = (Z_{B-OAB} + Z_{B-OBC})/2.$$

We rearrange the solution pairs and obtain the following triplets :

$$(Z_A, Z_B, Z_C), (-Z_A, -Z_B, -Z_C)$$

as the two sets of solutions.

The plane OAB is represented by

$$ax + by + z = 0 \quad \text{for } (Z_A, Z_B, Z_C)$$

$$-ax - by + z = 0 \quad \text{for } (-Z_A, -Z_B, -Z_C),$$

where $a = (Z_B Y_A - Z_A Y_B) / (X_A Y_B - X_B Y_A)$

$$b = (Z_A X_B - Z_B X_A) / (X_A Y_B - X_B Y_A).$$

Since the edge OB is labeled with a +, we have the following inequalities

$$aX_C + bY_C + Z_C < 0 \quad \text{for } (Z_A, Z_B, Z_C)$$

$$-aX_C - bY_C + Z_C < 0 \quad \text{for } (-Z_A, -Z_B, -Z_C).$$

One of the triplets satisfying the above relation is selected as

the true solution. Thus, we obtain the reliable parameters in the first frame from the three views.

Since there exist m views, we have $m-1C_2$ triplets of views to evaluate the parameters, and their solutions are obtained by applying a similar but simpler procedure. (The true solution is easily selected from the solution pairs as those having the same sign as Z_B obtained already.) The solutions judged as reliable are multiplied with a weight w which is a linear function of the reliability index, and then they are averaged to obtain the final evaluation of Z_A , Z_B , and Z_C . We use $C_s=50$, $C_T=0.9$ and $w=R(z)-0.88$ for the experiments described in this paper.

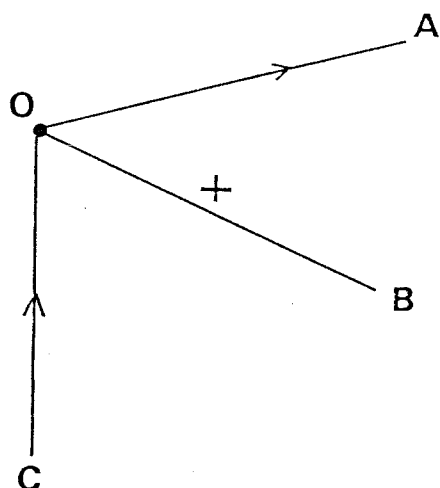


Fig.A A junction used for reconstruction process.

BIBLIOGRAPHY

[Aggarwal-75] J. K. Aggarwal and R. O. Duda, "Computer analysis of moving polygonal images, *IEEE Trans. Comput.*, vol. C-24, pp. 966-976, Oct., 1975.

[Agin-76] G. J. Agin and T. O. Binford, "Computer description of curved objects," *IEEE Trans. Computer*, vol.C-25, pp.439-449, April, 1976.

[Asada-80] M. Asada, M. Yachida and S. Tsuji, "Three dimensional motion interpretation for the sequence of line drawings," in *Proc. 5th Int. Joint Conf. Pattern Recognition*, Miami, USA, Dec., 1980, pp.1266-1275.

[Asada-81] M. Asada, M. Yachida and S. Tsuji, "Reconstruction of three dimensional motions from image sequence," in *IEEE Proc. Pattern Anal. and Image Processing*, 1981, Dallas, Aug., 1981, pp.88-90.

[Barrow-81] H. G. Barrow and J. M. Tenenbaum, "Interpreting line drawings as three-dimensional surfaces," *Artificial Intell.*, vol.17, no.1-3, pp.75-116, 1981.

[Clowes-71] M. B. Clowes, "On seeing things," *Artificial Intell.*, vol.2, no.1, pp.79-116, 1971.

[Greaves-75] J. O. Greaves, "The software strucure for reduction of quantized video data of moving organisms," *Proc. of IEEE*, vol.63, pp. 1415-1425,

[Guzman-71] A. Guzman, "Decomposition of a visual scene into three-dimensional bodies," in *AFIPS Proc. of Fall Joint Comput. Conf.*, Vol.33, 1968, pp.291-304.

[Horn-75] B.K.P Horn, "Obtaining shape from shading information" in *The Psychology of Computer Vision*, P. H. Winston, Ed. New York: McGrew-Hill, 1975, pp.115-155.

[Huffman-71] D. Huffman, "Impossible objects as nonsense

sentences," in *Machine Intelligence 6*, B. Meltzer and D. Michie, Eds. Edinburgh, Scotland: Edinburgh University Press, 1971, pp.295-323.

[Huffman-77] D. A. Huffman, "A duality concept for the analysis of polyhedral scenes," in *Machine Intelligence 8*, E. Block and D. Michie, Eds. Edinburgh, Scotland: Edinburgh University Press, 1977, pp.475-482.

[Ikeuchi-81] K. Ikeuchi and B. K. P. Horn, "Numerical shape from shading and occluding boundaries," *Artificial Intell.*, vol.17, no.1-3, pp.141-184, 1981.

[Jain-79] 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, 1979.

[Kaneko-73] T. Kaneko and P. Mancini, "Straight-line approximation for the boundary of the left ventricular chamber from a cardiac cine-angiogram," *IEEE Trans. Biomed. Eng.*, vol. BME-20, pp. 413-416, 1973.

[Mackworth-71] A. K. Mackworth, "Interpreting pictures of polyhedral scenes," *Artificial Intell.*, vol.4, no.2, pp.121-137. 1971.

[Marr-78] D. Marr, "Representing visual information," in *Computer Vision Systems*, Hanson and Riseman, Ed. New York: Academic Press, 1978.

[Martin-78] W. N. Martin and J. K. Aggarwal, "Dynamic scene analysis: A survey," *Comput. Graphics and Image Processing*, vol.7, no.3, pp. 356-374, 1978.

[Meiri-80] A. Z. Meiri, "On monocular perception of 3-D moving objects," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, no.6, pp. 582-583, 1980.

[Nagel-78] H. H. Nagel, "Analysis technique for image sequences,"

in *Proc. 4th Int. Joint Conf. on Pattern Recog.*, Kyoto, Japan: Nov. 1978, pp.186-211.

[Nagel-81] H. H. Nagel, "Representation of moving rigid objects based on visual observation," *Computer*, vol.14, no.8, August 1981, pp.29-39.

[Roach-79] J. Roach and J. K. Aggarwal, "Computer tracking of objects moving in space, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-1, no.2, pp.127-135, 1979.

[Roach-80] J. W. Roach and J. K. Aggarwal, "Determining the movements of objects from a sequence of images, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, no.6, pp. 554-562, 1980.

[Roberts-63] L. G. Roberts, "Machine perception of three-dimensional solids," in *Optical and Electro-Optical Information Processing*, J. T. Tippet et al., Eds. Cambridge, MA: MIT Press, pp.159-197, 1963.

[Thompson-80] W. B. Thompson, "Combining motions and contrast for segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, no.6, pp.543-550, 1980.

[Ullman-79] S. Ullman, *The Interpretation of Visual Motion*, Cambridge, MA: MIT Press, 1979.

[Waltz-75] D. Waltz, "Understanding line drawings of scenes with shadows," in *The Psychology of Computer Vision*, P. H. Winston, Ed. New York: McGraw-Hill, 1975, pp.19-95.

[Webb-81] J. A. Webb and J. K. Aggarwal, "Visually interpreting the motions of objects in space," *Computer*, vol.14, no.8, August, 1981, pp.40-46.

[Witkin-81] A. P. Witkin, "Recovering surface shape and orientation from texture," *Artificial Intell.*, vol.17, no.1-3, pp.17-45, 1981.

[Yachida-78] M. Yachida, M. Asada and S. Tsuji, "Automatic motion analysis system of moving objects from the records of natural processes," in *Proc. 4th Int. Joint Conf. Pattern Recognition*, Kyoto, Japan: Nov. 1978, pp. 726-730.

[Yachida-80] M. Yachida, M. Ikeda and S. Tsuji, "A plan-guided analysis of cineangiograms for measurement of dynamic behavior of heart wall, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, no.6, pp. 537-543, 1980.