UNDERSTANDING A SIMPLE CARTOON FILM BY A COMPUTER VISION SYSTEM

Saburo Tsuji, Akira Morizono, and Shinichi Kuroda
Department of Control Engineering
Osaka University
Toyonaka, Osaka 560, Japan.

This paper describes a primitive intelligent system that analyzes and understands simple cartoon films of a dynamic mini-world; each film contains a dynamic line image in which an actor, a personified frog named Besi, and objects such as a tree or a rock exist. The first goal of our research is to give the machine vision system the capability of understanding what the frog is doing or meanings of its actions.

Input Film

In order to simplify the analysis of the films, we assume that they have the following properties. (1) There exists only one scene in a film. (2) The used cine camera stood still or moved horizontally at a constant velocity while it took the picture. (3) Besi and other movable objects show two-dimensional movements.

A film on a simple scenario, Scenario 1, is used to test the elementary functions of the system. Fig.l shows sample frames of the film.

Scenario 1: Besi hides behind a rook, appears from it, walks toward a tree, jumps up and takes a fruit from it.

System Implementation

In order for the system to attain the goal, it must be an integrated system that first identifies the actor and the objects and detects their movements, next recovers patterns of their motions, and finally deduces actions from a sequence of the motion patterns and relations between the actor and the objects by utilizing the knowledge about the mini-world. In other words, it must be provided with (1) a scene analyzer to identify the the actor and the objects, (2) a motion analyzer to recognize the motion patterns, (3) a deduction subsystem to understand meanings of the motions, (4) a model of the viewed film which is genenrated and updated by the system. The deduction subsystem also behaves as a question-answering system, and we can test how the system understands a film by giving it questions on the film and observing answers deduced from the film model and the knowledge about the mini-world. Fig.2 shows an overall diagram of the film-understanding system. We implement it on a system of two mini-computers, HP2108A (64kbytes memory, and 256kbytes bulk memory for LISP program) and PDP8/E (24kbytes memory, and 48kbytes bulk memory for storing pictures).

Scene Analysis

Examining local patterns in a 3x3 window at every point in each 128x128 digitized picture of the film, SCENE ANALYZER finds feature points, and then it follows arcs in them. Thus, we obtain a segment list which describes the input picture.

When the first frame is accepted, SCENE ANALYZER has no knowledge about the input scene. It first searches the segment list for loops and long straight lines. They are used as cues for successive suggestions and tests of the hypotheses that interprete the line image by the procedural knowledge SHAPE. For example, the knowledge about the shape of Besi is described as follows. (l)Find two almost elliptic loops such that one encloses the other; they corresponds to Besi'8 body. (2)Find a big loop over these loops; it corresponds to Besi's face. (3)Find a short path connecting them; Besi's neck. If the above procedure gives the result of a success, Besi is identified, and then the segments corresponding to the detailed part of Besi such as its legs, eyes, and hands are interpreted. The result of analyzing the frame of Fig.1(a) is shown in Fig.3(a); the tree is found by a combination of an inverted T joint and a big loop, and then it suggests a hypothesis that small loops near the tree would be fruits, and the procedural knowledge on the shape of the fruit tests these loops. The rock and Besi are interpreted as an

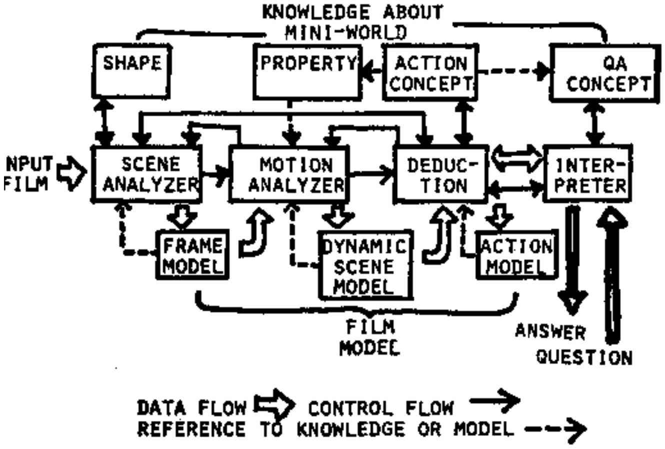

Fig.2 Organization of Film-Understanding System.

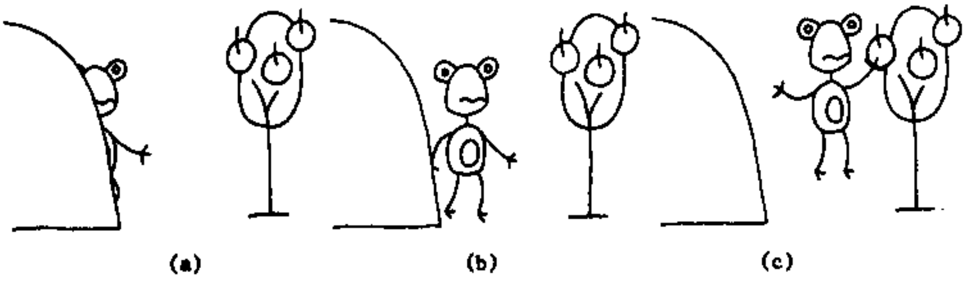

Fig.1 Sample Frames of Film on Scenario 1.
(a) Besi appears,
(b) walks toward a tree,
(c) jumps up and catches a fruit from it.

unknown object because SHAPE has not any knowledge on the rock. The interpretation of the frame is stored in FRAME MODEL.

Since the difference between two successive frames is very small, FRAME MODEL suggests the location and property of each segment of the objects in the coming frame. Thus, we can reduce the computing time for analyzing the new frame by 1/10. SCENE ANALYZER also examines the correspondence of each part of the unknown object with that of the preceding frame, and it discriminates an unknown moving object from a still one (the rock). When the frame of Fig.1(b) is analyzed, the cues for suggesting that the moving object would be Besi is recovered, and SHAPE identifies all parts of Besi (Fig.3(c)).

Motion Analysis

MOTION ANALYZER compares the locations of the actor and the objects in successive frames, and judges which one is really moving. Since the relative movement of a still object is caused by a camera movement, MOTION ANALYZER tests whether most objects in a picture move in a similar way. If so, the displacement is considered as the camera movement, and the origin of the coordinates is shifted so as to compensate the movement.

Next, MOTION ANALYZER parses the film into scene groups; each group has successive frames in which the actor moves in a similar way. The actor's movement is described in DYNAMIC SCENE MODEL (Fig.4). The film is segmented into A groups, and Scene Group 1 is described as (M BESI RIGHT), Besi moves rightwards. Changes of the relations between the actor and the objects are also examined; for example, MOTION ANALYZER finds a fruit, an apple, is moving in Scene Group 3, so that it examines its relation to Besi, and adds (TC BESI APPLE) Besi touches the apple to the model. Also it examines the validity of the relation in the past frames, and updates the model.

Deduction of Actions

Now let us consider how people infers concepts of actions from a descrption Besi moves rightwards. We probably deduce Besi walks rightwards after reasoning the following items. (l)Moving one is a human being. (2)He moves to a horizontal direction implies he walks or runs. (3)If his legs move fast then the motion is run, else walk. (4)Right is a horizontal direction. (5)Besi is equivalent to a human being. (6)The reason why one moves to a direction is to transfer itself.

A priori knowledge about the mini-world such as (4) and (5) is stored in a semantic network PROPERTY. ACTION CONCEPT stores the knowledge about each action:[a] Static conditions are predicates which must be true for assuring the specific action is performed;(1) for the above example, [b] Dynamic conditions are described in the procedural knowledge for logical reasoning from state transitions: (2) (3) for walk. [c]Deduced knowledge is one assured by the actual performance of the action; (6) for the above example.

The deduction subsystem utilizes these concepts to deduce actions from DYNAMIC SCENE MODEL. The subsystem finds an action matched to a motion pattern, and then it tests other conditions for the action. For example, the success in matching (M BESI RIGHT) to a description (M X Y) in an action move makes the subsystem test the other conditions, X is animate, X comes near to Y. Since the both conditions are satisfied, (MOVE BESI TREE) is deduced. The description also matches to a condition for walk, and (WALK BESI TREE) is deduced by examining other conditions for walk. After similar deductions, the film is described as (((WALK BESI TREE))((JUMP BESI APPLE)) ((JUMP BESI APPLE)(HOLD BESI APPLE))((HOLD BESI APPLE))). Finally, the system searches for actions unifying a series of actions, and it deduces (((CATCH BESI APPLE))). Fig.5, examples of dialogue of QA-system, shows how the system understands the film.

```
* HELLO
GIVE ME ANY QUESTION

* WHAT IS BESI ?
BESI IS FROG

* WHY JUMP BESI APPLE ?
BECAUSE TO NEAR APPLE

* WHAT DID BESI WHEN MOVE APPLE ?
TIME IS SD4
BESI HOLD APPLE

* WHAT DID BESI ?
BESI CATCH APPLE

* HOW DID BESI ?
WALK BESI TREE
JUMP BESI APPLE
CATCH BESI APPLE
HOLD BESI APPLE

* WHEN JUMP BESI ?
TIME IS SD2 OR SD3

* WHY CATCH BESI APPLE ?
BECAUSE TO HOLD APPLE

* THANK YOU
YOU ARE WELCOME
```

Fig.5 Examples of Dialogue

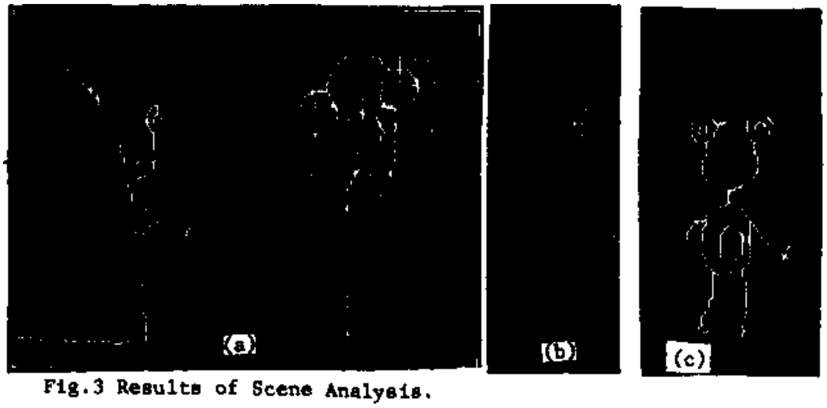

Fig.3 Results of Scene Analysis.

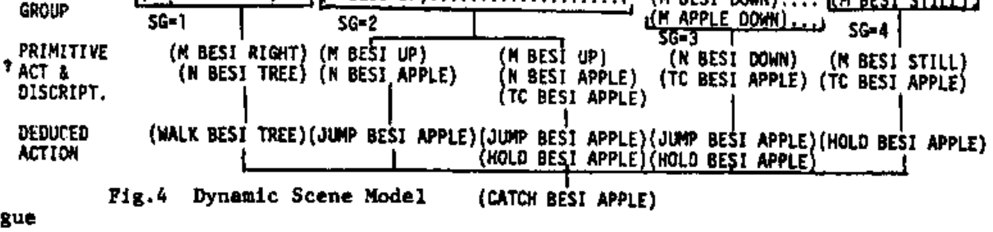

Fig.4 Dynamic Scene Model