RESEARCH NOTE

Dynamic Image Understanding on General Roads in Japan Using Fuzzy Frame Knowledge Base with Spool Base

Y. Nakagawa[1], K. Hirota[1] and W. Pedrycz[2]

In this paper, fuzzy frame knowledge base with spool base for dynamic image understanding system on general roads in Japan is presented with experimental results. The fuzzy knowledge base with spool base consists of fuzzy inference engine, fuzzy frame and spool frame. The spool frame is divided into floating spool frame (the information of the ambiguously recognized results about an observed object) and anchored spool frame (the information about the uniquely determined object). It is a tool for adapting the dynamically changing ambiguous information. The validity and the effectiveness of the notion of spool base are confirmed through the experiments of the dynamic image understanding e.g. existence of a moving object coming forward from a far distance is confirmed first, then its color, velocity and occlusion with other objects are gradually recognized and finally its type (moving objects in general) is identified. Moreover, the system enables us to recognize objects under various lighting and weather conditions by using fuzzy knowledge base.

INTRODUCTION

Many studies have been reported on dynamic image understanding that deal with extracting a moving object in the observed images [1,2]. Many knowledge representation scheme for image understanding have also been studied [3].

The ultimate goal of this study is to realize a real-time human like robotics vision. A dynamic image understanding system using fuzzy frame knowledge base with spool base has been presented here [4,5]. The presented system enables us to: 1) extract objects under various lighting conditions (i.e. sunlight changing through time and weather conditions: sunny, cloudy and rainy), 2) represent recognized objects with membership value, 3) make the extracted object gradually clear, such as a human and 4) response to user questions with useful answers by phonetically presented Japanese. The system consists of an image preprocessing part, a fuzzy knowledge processing part and human interface part. It should be noted that the fuzzy knowledge processing part becomes the most important part in the dynamic image understanding system.

In the image preprocessing part, moving objects on general roads under various lighting conditions are extracted. By using rule-based fuzzy inference and difference images,

1. Department of Systems Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, 4259 Nagatsuta-Cho, Midori-Ku, Yokohama 226, Japan.
2. Department of Electrical and Computer Engineering, The University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2.

Scientia Iranica, Vol. 2, No. 4, © Sharif University of Technology, January 1996.

the system extracts the objects under various lighting conditions with shadow and reflection or in monotone-like color against a dark background.

The fuzzy knowledge processing part consists of fuzzy frame knowledge base (including fuzzy frame [6], anchored spool frame and floating spool frame) and fuzzy inference engine. The fuzzy frame includes fundamental knowledge on general roads such as human, cars, motorcycles, road and objects' movement. The fuzzy inference engine generates/updates an anchored spool frame or a floating spool frame depending on whether the recognition result is uniquely determined or not, respectively. One of the advantages of the floating spool frame is that it makes the extracted object gradually clear as the object gets closer. The system enables us to recognize objects' single or mutual movement using a sequence of X-Y coordinates.

The human interface part using natural language (phonetically presented Japanese) consists of a keyword extraction process and answer sentence generation process. In the keyword extraction process, keywords for checking fuzzy frame knowledge base are extracted from the user question sentence. The answer sentence generation process provides a useful answer for a user question using prototype sentences.

Finally, detailed experimental results are shown along with a numbers of selected frames to confirm the usefulness of the proposed system.

DYNAMIC IMAGE UNDERSTANDING SYSTEM CONFIGURATION

A dynamic image understanding system on general roads in Japan is being presented here. It consists of an image preprocessing part, a fuzzy knowledge processing part and a human interface part. The system configuration is shown in Figure 1.

The image preprocessing part observes a series of image frames by a CCD camera to extract the moving objects. It extracts sky region and road regions under the various lighting conditions based on rule-based fuzzy inference. Moving objects are included in the region between the sky and the road. The moving objects are extracted by the difference images method (e.g. [1]). The result of the image preprocessing part includes the extracted object's image, width, height, identification number and color which are transferred to the fuzzy knowledge processing part.

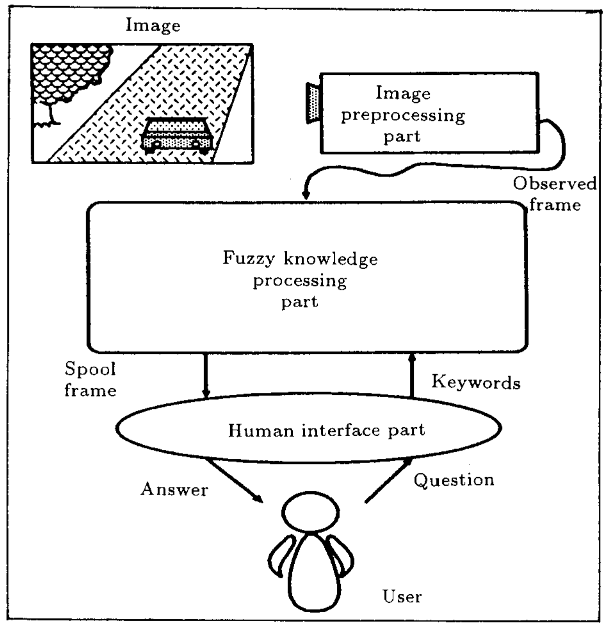


Figure 1. A dynamic image understanding system configuration.

The fuzzy knowledge processing part consists of a fuzzy inference engine and fuzzy frame knowledge base with a spool base. The fuzzy inference engine matches the extracted object image and the standard image fills in the fuzzy frame (that exists in the fuzzy frame knowledge base). Then, it compares the extracted object's width, height, color and speed with the slot values of fuzzy frame including the standard image file. Moreover, the fuzzy inference engine enables us to recognize objects' movement using a sequence of X-Y coordinates of the extracted objects. Finally it generates/updates an anchored spool frame or a floating spool frame for the extracted object depending on the result of matching.

The human interface part receives a question sentence (in phonetically presented Japanese) about the image world by the user. It extracts keywords from such a query and generates a useful answer for the question sentence using a set of prototype sentences. The details of these three parts are discussed in the following.

Image Preprocessing Part

The objective of the image preprocessing part are 1) extracting moving objects 2) getting information such as width, height, identification, number, color and X-Y coordinates in the image frame about the object and 3) identifying each object in a sequence of images.

Dynamic input images are observed by a CCD camera in RGB-color mode (512 × 432 pixels for each image frame). The sampling time between two consecutive image frames is 0.1 second considering the maximum legal speed (60km/h) of cars on general roads in Japan and computational power limitation. The RGB-color model is transformed into HLS-color model to utilize human subjective fuzzy evaluation on hue, lighting and saturation.

The object extraction is done as a result of identifying sky and road regions and difference images of the two image frames excluding sky and road regions.

An observed image is processed to extract the road region using two input three output fuzzy if ... then rules in Table 1 and the sky region using one input one output fuzzy if ... then rules. The membership functions of road region are shown in Figure 2. These rules are tuned up using the best experimental results for each lighting condition.

Table 1. Fuzzy if ... then rules for road region.

if	fine	& morning	then	S_H4, L_L8, L_H12
if	fine	& noon	then	S_H8, L_L9, L_H18
if	fine	& evening	then	S_H4, L_L8, L_H11
if	cloudy	& morning	then	S_H4, L_L9, L_H12
if	cloudy	& noon	then	S_H4, L_L12, L_H14
if	cloudy	& evening	then	S_H4, L_L11, L_H14
if	rainy	& morning	then	S_H3, L_L9, L_H17
if	rainy	& noon	then	S_H4, L_L10, L_H21
if	rainy	& evening	then	S_H4, L_L12, L_H18

S_H*, L_L*, L_H* stand for fuzzy labels of threshold of saturation, bottom of lighting and top of lighting, respectively.

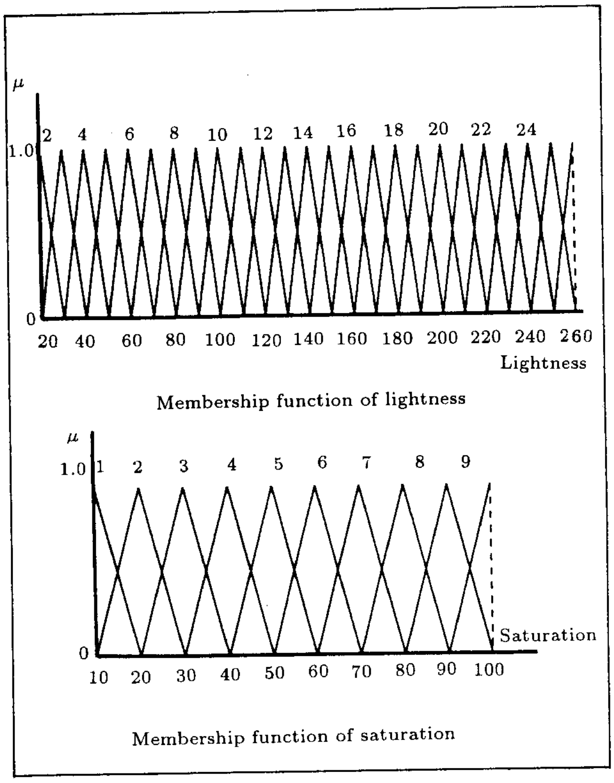


Figure 2. Membership functions of road.

Moving objects are extracted by applying an inter-frame difference operation for a sequence of the processed images mentioned above. The system could identify moving objects in order of appearance by labeling the extracted objects.

Each moving object is memorized in the form of a frame representation including the extracted object's image, width, height, identification number, color and X-Y coordinates in the image frame. These results are included in a frame form and stored until the next observation time. The result of image preprocessing part, called the observed frame, is transferred into fuzzy knowledge processing part.

Fuzzy Knowledge Processing Part

The fuzzy knowledge processing part consists of a fuzzy frame knowledge base and a fuzzy inference engine, as shown in Figure 3. The roles of this part are: 1) recognizing the extracted objects and their movements, 2) clarifying ambiguous objects and 3) updating the information about the extracted objects to answer the user's questions.

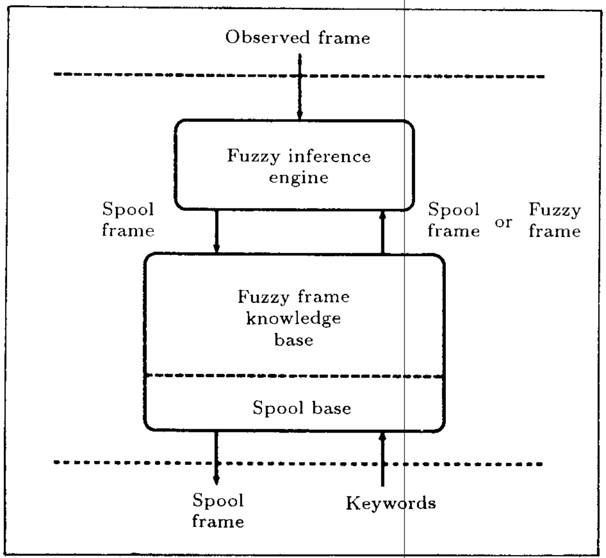


Figure 3. The configuration of fuzzy knowledge processing part.

Fuzzy Frame Knowledge Base

The fuzzy frame knowledge base consists of three kinds of frames, i.e. fuzzy frame, anchored spool frame and floating spool frame.

The fuzzy frame includes fundamental knowledge on general roads such as human, cars, motorcyles, road and objects' movement. These are given in advance considering the difference between movable objects and immovable (static) objects. A fuzzy frame is either an object category frame or a movement category frame. An example of the object category frame with spool frames is given in Figure 4.

The object category frame is connected with object subframes. Each object subframe has an object name, four slot items with membership grades and an adoptive threshold. The object name is a registered word to represent a recognized object such as pedestrian, bicycle, car and bus. Four slot items include one or more standard image files, one or several color names, a membership function of speed and membership functions of aspect ratio (of whole contour and several pieces of parts depending

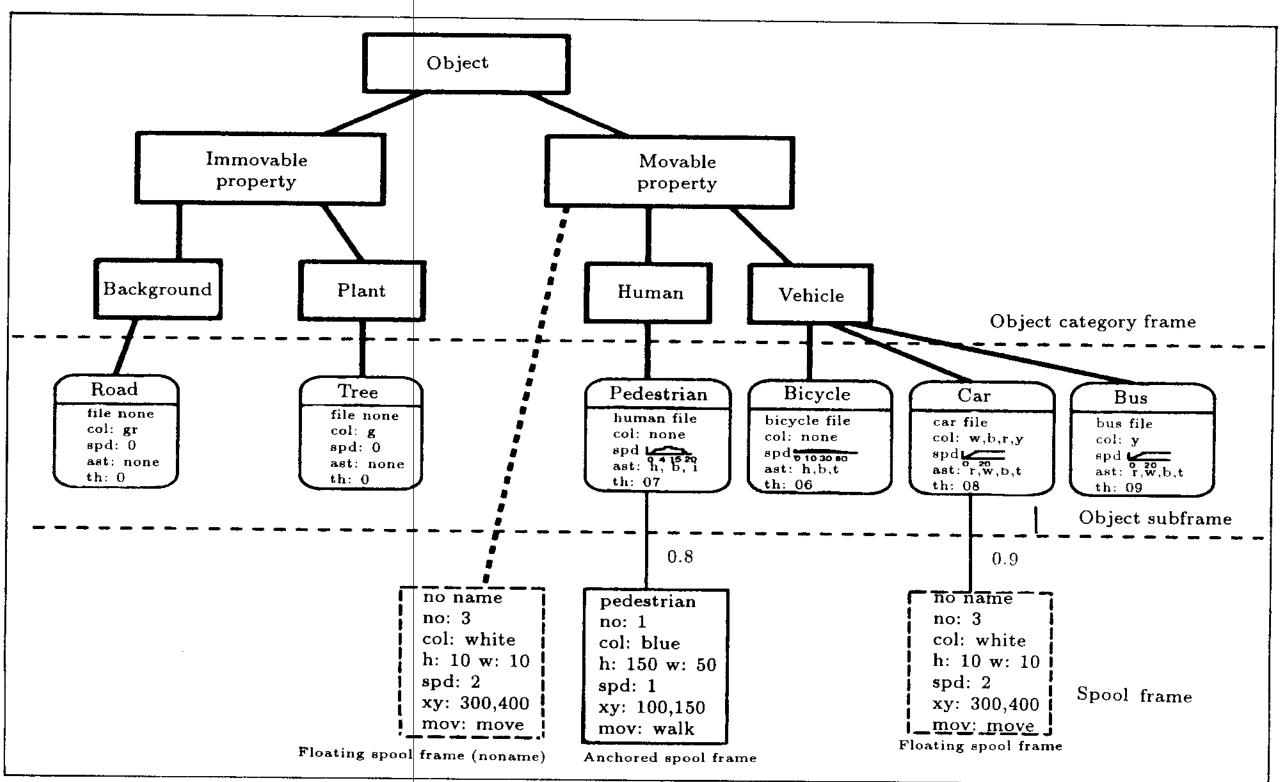


Figure 4. An example of the object category frame tree with spool frames.

on objects). The membership function defined on four items indicates the degree of importance in recognizing a labeled moving object as the object name. The adoptive threshold is referred to by the fuzzy inference engine when the fuzzy inference engine assigns the object name to the labeled moving object.

The movement category frames are connected with movement subframes. An example is given in Figure 5. Each movement subframe includes two slot items, i.e. movement masters and fuzzy if ... then rules with two fuzzy input (speed and direction) and one output (movement word).

The fuzzy inference engine generates/updates either the anchored spool frame or the floating spool frame for each labeled moving object, depending on the number of object subframes obtained. An anchored spool frame is generated/updated by the fuzzy inference engine when the inference engine obtains a unique object subframe for the observed frame. A floating spool frame is generated/updated when the fuzzy inference engine obtains plural (at most three for simplicity) object subframes for the labeled moving object. If the object has a sequence of X-Y coordinates stored in anchored spool frame or floating spool frame, the fuzzy inference engine recognizes the object movements. The anchored spool frame contains the name, identification number, color, height, width, speed, X-Y coordinates in the sequence of observed image frames and the object's movements. The anchored spool frame is fuzzily linked to the object subframe. The floating spool frame memorizes the names with membership grades suggested for the object, identification number, color, height, width, speed, X-Y coordinates in the sequence of observed image frames and the object's movements.

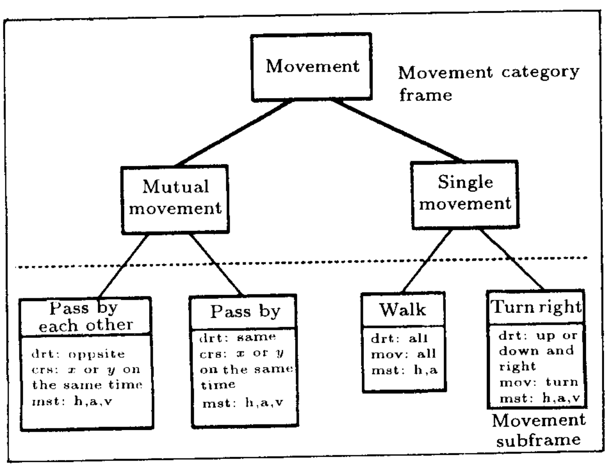


Figure 5. An example of movement category frames.

The spool frame makes it possible to express situations such as when 1) information changes dynamically, e.g. a small moving object at a far distance gradually becomes clear as it gets closer and 2) the same object is ambiguously recognized, e.g. an extracted moving object is referred to plural possible object names with membership grades.

Fuzzy Inference Engine

A fuzzy inference engine executes object name inference and movement understanding. It receives the observed frame from the image processing part and accesses the fuzzy frame knowledge base.

In the first step of the object name inference, the extracted object is compared with each of the selected images (by substituting the aspect ratio of the extracted object for the stored membership function of aspect ratio) in standard image files stored in the fuzzy frames of the fuzzy frame knowledge base. The extracted object's color values in HLS-color model are translated into a color name by using the membership functions for color as shown in Figure 6.

The aspect ratios of several sections of the whole, e.g. roof, window, body and tires in the case of cars, are then calculated and substituted into the corresponding membership functions stored in the object subframe, including the standard image files. These membership values are minimized. The speed of the extracted object is then substituted into the membership function in the three fuzzy frames. Finally, the color name of the extracted object is checked against the color names included in each of the three selected fuzzy frames (if included,

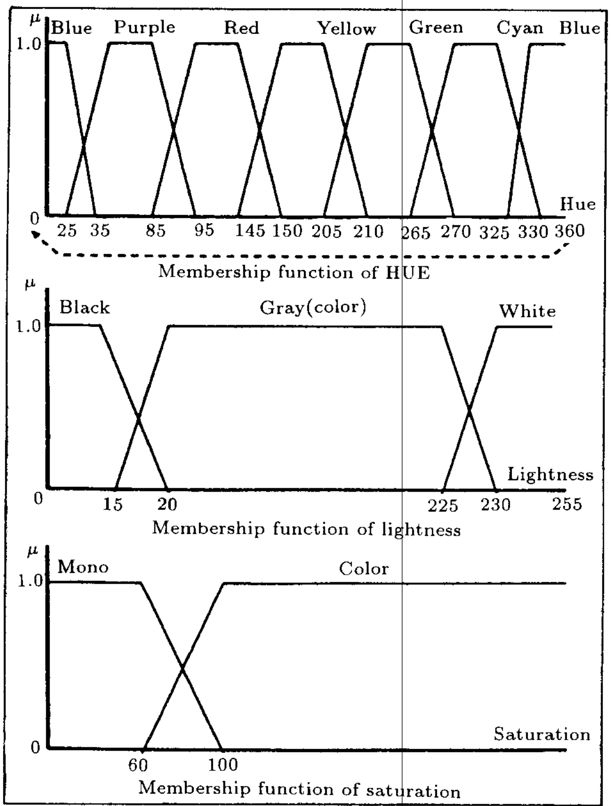


Figure 6. Membership functions of HLS-color model.

as in this case, the value 1 is assigned, otherwise 0).

The max-min composition is operated for the above mentioned three evaluation values and degree of importance in the selected object subframe. If this composed value is greater than, or equal to, the adoptive thresholds stored in the selected object subframes then the object names in the object subframes attached with the composed value are suggested for the object. If the composed value is less than all adoptive thresholds then the fuzzy inference engine generates/updates a floating spool frame with no object name for the extracted object. The floating spool frame with no object name is connected to the object category frame (see the floating frame connected with a dotted line in Figure 4). Moreover, if the object in the floating spool frame is specified, then the fuzzy inference engine exchanges from the floating spool frame to the anchored spool frame.

The movement understanding process selects preregistered words of movements from the sequence of X-Y coordinates and the objects' names. The sequence of X-Y coordinates of the object are stored in the spool frame. The object recognized the first time has no X-Y coordinates in the spool frame and its movement will not be analyzed. If the image frame includes plural objects already recognized then the object movements are analyzed in the order of single movement and mutual movement. Movement words of the extracted object are selected from a set of preregistered words stored in the movement subframes by applying the stored X-Y coordinates and X-Y coordinates in the observed frame to the if ... then rules that are stored in the movement subframe of the fuzzy frame knowledge base. If the extracted object is stored in an anchored spool frame then one movement word is chosen from the selected movement words in consideration of the extracted object name. In the case of floating spool frame, one movement word is determined from the selected movement words by using the name of the category frame connected with the floating spool frame.

The fuzzy inference engine proposed here enables us to: 1) make the extracted object clear step by step using spool frame as a human distinguishes gradually an object that is coming from a far distance and 2) understand the extracted objects' movements, not only single movements but also mutual movements with occlusion, e.g. identifying the movement of a fast car suddenly reappearing after a non-visible state when it is passing by a big slow car in a situation where they are both coming closer from far away.

Human Interface Part

The human interface part consists of a question sentence analysis process and an answer sentence generation process. It should be noted that the whole system presented employs a phonetically presented Japanese language to communicate with users.

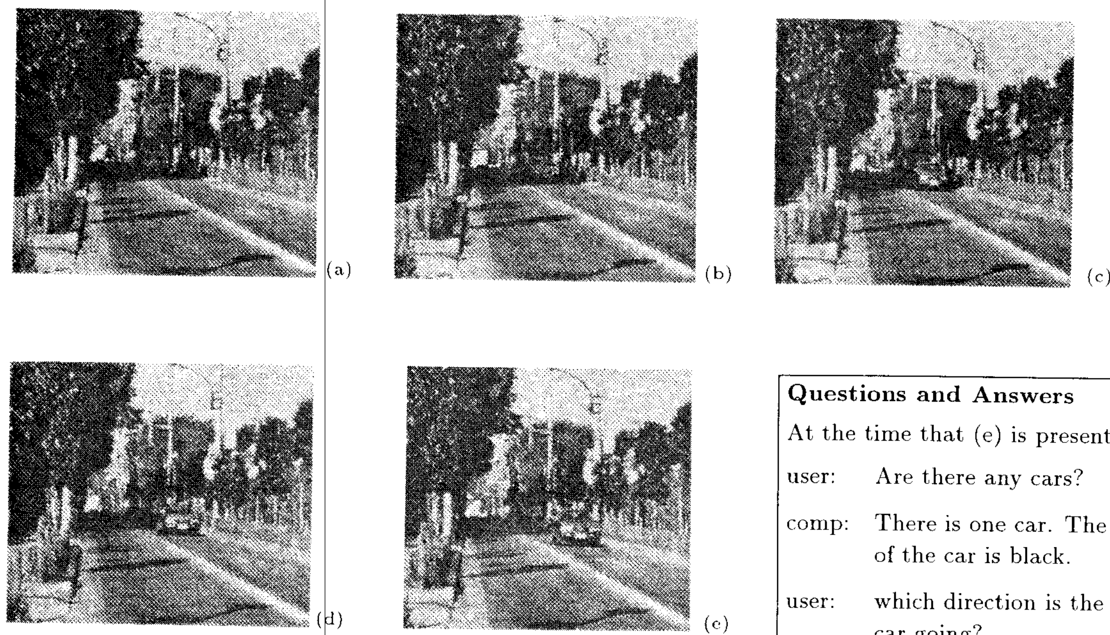
The question sentence analysis process is done in the order of lexical analysis, syntax analysis, semantic analysis, keywords extraction and the spool base check in the fuzzy frame knowledge base with spool base. Keywords are extracted from a collection of predefined words in the fuzzy frame knowledge base for each question given by users. As a result the keywords and checked spool frames are delivered to the answer sentence generation process. In the first phase of the answer sentence generation process the most desirable prototype sentence is selected from a set of already specified prototype sentences, where each prototype sentence becomes a complete sentence if a few appropriate words are filled in. The selected prototype sentence is modified into a complete answering sentence by using the results of the question sentence analysis process. If necessary, in order to add useful information to the answering sentence, the number of moving objects that are concerned with the question is counted.

Users can ask the system about the observed image world in natural language (phonetically presented Japanese) in the style of not only a yes/no type of question but also a what/which type of question. According to the situation, the human interface part enables us to create additional information intelligently to some extent (e.g., user: "Are there any cars?"; the system: "There is one car. The color of the car is white.").

EXPERIMENT OF THE DYNAMIC IMAGE UNDERSTANDING SYSTEM ON GENERAL ROADS IN JAPAN

The dynamic image understanding system using fuzzy frame knowledge base with spool base has been realized. The system consists of a CCD camera, an image processor Dr. IMAGE (Kawasaki Steel Co. LTD) and a work station SPARC Station10 (SUN Microsystems). The C language has been used for this system. The system was applied to the dynamic images observed on general roads in Japan from morning to early evening in various weather conditions (fine, cloudy and rainy). Figures 7, 8 and 9 are taken at intervals of 0.1 seconds for experiments, but here they are shown for every 0.5 seconds. Figure 7 shows a black car through tree shadow, since it was taken at noon in fine weather. The system could extract the object under the shadow at the time of Figure 7 (a) and stores the object information in the floating spool frame. The object information moves from the floating spool frame to the anchored spool frame at the time between Figure 7 (c) and (d) because the object was specified at this time. The first question simply asks at the time of Figure 7 (e) about the existence of a car, i.e. this is a typical example of yes/no type questions. The system identifies not only the existence of a contain object but also the number of cars in the first answer sentence and the color of the car in the following answer sentence by retrieving the information stored in the anchored spool frame. They are examples of additional useful information for users. The second query is an instance of what/which type of question. The system answers in which direction the car is running, showing that it could understand the movement of the extracted object.

Figure 8 shows two cars with occlusion. These pictures were taken during a rainy afternoon. It should be noted that the system extracted correctly these rather vague objects under the gray monotone circumstances in the rain by applying the result transferred from the image preprocessing part. At the time of Figure 8 (a), the information about the left car has already been stored in anchored spool frame and that of the right car is just stored in floating spool frame. The information about the right car moves from floating spool frame to anchored spool frame at the time of Figure 8 (c). Two cars pass by each other at the time of Figure 8 (d) and the information about the action "pass by" is stored in both anchored spool frames. The first question asks about the existence of cars at the time of Figure 8 (e). The systems responds to the question about the number of cars. This is a different answer from those of Figure 7 for the same question to avoid complication, i.e. in this case the system did not supply the colors of both cars since various object information is not indicated if the direct answer includes plural objects, i.e. it



Questions and Answers

At the time that (e) is presented.

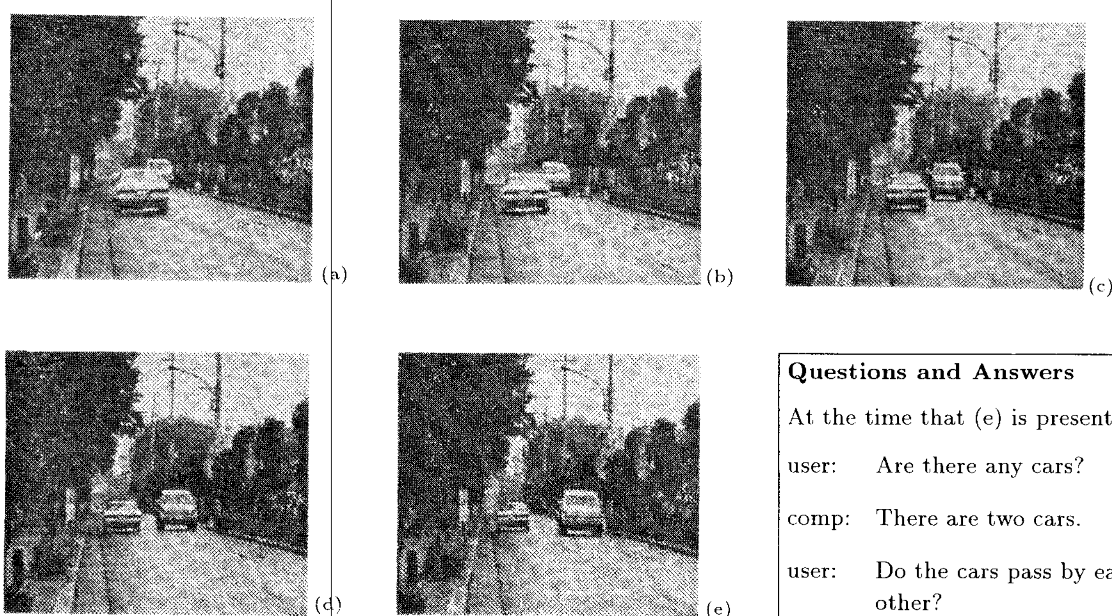
user: Are there any cars?

comp: There is one car. The color of the car is black.

user: which direction is the black car going?

comp: The black car is getting closer.

Figure 7. The observed images during a fine midday.



Questions and Answers

At the time that (e) is presented.

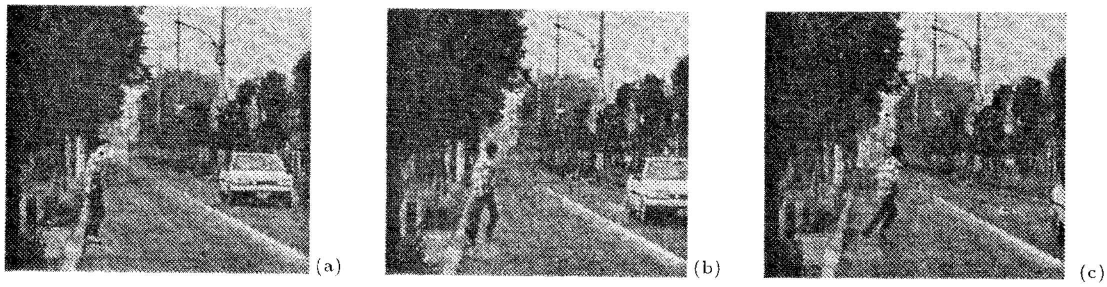
user: Are there any cars?

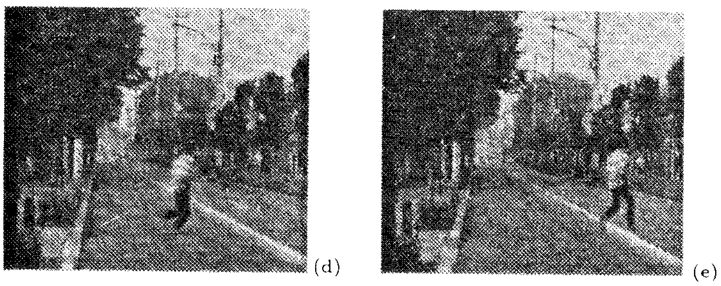
comp: There are two cars.

user: Do the cars pass by each other?

comp: Yes, they do.

Figure 8. The observed images during a rainy afternoon.





Questions and Answers

At the time that (e) is presented.

user:	Are there any cars?
comp:	Not at the present time.
user:	Has there been any cars?
comp:	There has been one car. The color of the car was white.
user:	What is there now?
comp:	There is a pedestrian.

Figure 9. The observed images during a cloudy evening.

is regarded as over useful information in this system. The system understands the mutual movement of the two cars. The second question concerns the cars passing by each other. The system identifies their mutual movement with yes/no. It shows that the system understands the mutual movement of the extracted objects in the image and updates the information of the extracted objects with spool frame especially the occluded car on the right side.

Figure 9 shows a car and a human, taken on a cloudy evening. The first question asks about the existence of a car at the time of Figure 9 (e). The system gives no affirmative however, implicates the existence of a car in the past, i.e. from the time of Figure 9 (a) to that of (c). The second question refers to the existence of a car in the past. The system gives the answer. It shows that the system updated the information about the car that ran out from the image frame at the time of Figure 9 (c). The last question asks a what/which type question. The system answers correctly what occurs in the present image frame. This example indicates that the system can trace the past situation correctly. Moreover, the possibility of a car coming in the left lane can be predicted from the back of the CCD camera because the information concerning the velocity of the human (who suddenly crossed the road) stored in the human's anchored spool frame varies slow (a)(b) at the left side of the road, fast (b)(c)(d) in the middle of left lane and slow (d)(e) in the center of the road (It should be noted that cars must keep left in Japan).

These experiments show that the system extracts the object under poor lighting conditions in Figures 7 and 8. By applying the spool frame, the system updates the information of the extracted object even if the extracted object is occluded. The system gives useful answers intelligently using the preregistered prototype sentences and the information in the fuzzy frame knowledge base with spool base.

CONCLUSION

A dynamic image understanding system on

general roads in Japan has been realized based on fuzzy inference and fuzzy knowledge base. It consists of image preprocessing part, fuzzy knowledge processing part and human interface part.

The object extraction under various lighting conditions even in bad weather and daylight changes is done by rule based fuzzy inference and difference image method in the image preprocessing part. The spool frame in the fuzzy frame knowledge base is proposed to provide a tool for expressing a gradually changing cognitive environment. The fuzzy inference engine enables us to recognize the extracted object and the objects' movements. The human interface provides a useful answer, that is not asked directly, for the user question. The system can answer both yes/no and what/which type questions. Finally, a few experimental results on general roads in Japan are demonstrated to confirm higher recognition capabilities such as correct moving objects extraction under difficult lighting conditions, mutual movements recognition with a large speed difference under occlusion and gradual understanding of objects coming closer.

Future study will cover the realization of a real time vision system for an autonomous mobile robot.

ACKNOWLEDGMENT

The authors would like to thank the referees for detailed and helpful comments.

REFERENCES

1. Choi, H., Agui, T., Nakajima, M. and Yokomae, T. "Extraction of passing car and its number plate using a series of images", *The Trans. of ICEIE D-II* (written in Japanese) **J73-DII** (3), pp 392–398 (1990).

2. Kawabata, A., Tanifuji, S. and Morooka, Y. "An image extract method for moving object", *J. of Information Processing Society of Japan* (written in Japanese), pp 395–442 (1987).

3. Rao, A.R. and Jain, R. "Knowledge representation and control in computer vision systems", *IEEE Expert Spring*, pp 64–79 (1988).

4. Hirota, K., Nakagawa, Y., Sakakibara, K., Tanaka, K., Kusanagi, K. and Yoshida, Y. "A real-time fuzzy dynamic image understanding system on general roads", *Proc. of 3rd IEEE International Conference on Fuzzy Systems* , Orlando, Florida, USA, **1**, pp 106–111 (1994).

5. Nakagawa, Y. and Hirota, K. "Fundamentals of fuzzy knowledge base for image understanding", *Proc. of FUZZ-IEEE/IFES'95*, Yokohama, Japan, pp 106–111 (1995).

6. Kakui, M., Umano, M., and Miyahara, H. "Fuzzy frame system", *The 6th Fuzzy System Symposium*, (written in Japanese) pp 247–250 (1990).