The Third AFSS(1998). 252-257

Fuzzy Rule Based Multimedia Information Data Acquisition Method

Kabsuk OH[1], Kaoru HIROTA[1], Witold PEDRYCZ[2]

1. Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 226-8502, Japan.
Tel:+81-45-924-5686 E-mail:oks@hrt.dis.titech.ac.jp, hirota@hrt.dis.titech.ac.jp
2. Department of Electrical and Computer Engineering, The University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2

Abstract

A method of multimedia information data acquisition based on fuzzy rules is proposed, where the multimedia means the five senses of a human being. Observed information is characterized by *VAGOT* (visual, acoustic, gustatory, olfactory and tactile) time series data and the goal is to extract an appropriate subset of the *VAGOT* data based on a given instruction. Fuzzy rules based on visual and acoustic information are used to identify the appropriate time interval on the fireworks multimedia information.

Keywords: multimedia, fuzzy inference, information acquisition, real time processing

1. Introduction

Multimedia information fusion is increasingly recognized as central to the multimedia database and the multimedia computer system [1], [7]. It is very important to obtain appropriate real-time information for the multimedia information researches.

The multimedia, in this paper, means the five senses of a human being. Human being gets intelligent information data from outside world using the five senses. Similarly, computer is expected to obtain useful external data using various sensors such as CCD camera and microphone. Therefore, computer naturally deals with multimedia information based on the sensors, but taking current multimedia techniques into consideration, here multimedia information is considered centering around image (dynamic image) and sound.

When human being decide a temporal boundary of appropriate multimedia information data, the temporal boundary, which have uncertainty, is difficult to express by two valued logic. So, fuzzy rules are introduced to infer the uncertainty. It is also considered real time processing, because human being processes a great deal of information on the momentary from outside world.

In this paper, each media data is expressed by time series observation data. Observed data is represented by fuzzy membership functions with special feature of each medium. Fuzzy rules related to image are extracted by the change of characteristics of object in dynamic image. Similarly, there are fuzzy rules to judge the current situation from sound intensity. As one of the multimedia information fusion, a method of multimedia information data acquisition in information frame based on fuzzy rules is proposed, with joint fuzzy rules as well as real time processing. The final goal of this study is to obtain a useful outside world information data under an aim based on fuzzy rules.

In Section 2, the multimedia information acquisition method is proposed [6], [8]. The fuzzy inference engine is described in Section 3. Section 4 shows the experimental results to obtain fireworks multimedia information based on fuzzy rules.

2. Configuration of the multimedia information acquisition system

The study of each media information have been done independently as image processing, voice processing, and so on. If those processes are fused or unified then it can be anticipated more highly multimedia information processing. Based on these viewpoints, the proposed configuration of multimedia information data acquisition system is shown in Figure 1.

Multimedia Input Device regards data of each media as observed information characterized as a time series data. The Visual Information, Acoustic Information, Gustatory Information, Olfactory Information and Tactile Information are characterized by $V(t)$, $A(t)$, $G(t)$, $O(t)$ and $T(t)$ at observed time t, respectively. And the whole observed multimedia information can be characterized by $VAGOT$ as follows:

$$VAGOT_{\mathrm{T}} = \{V(t), A(t), G(t), O(t), T(t)\}_{t \in \mathrm{T}}, \quad (1)$$

where T shows the whole observed time.

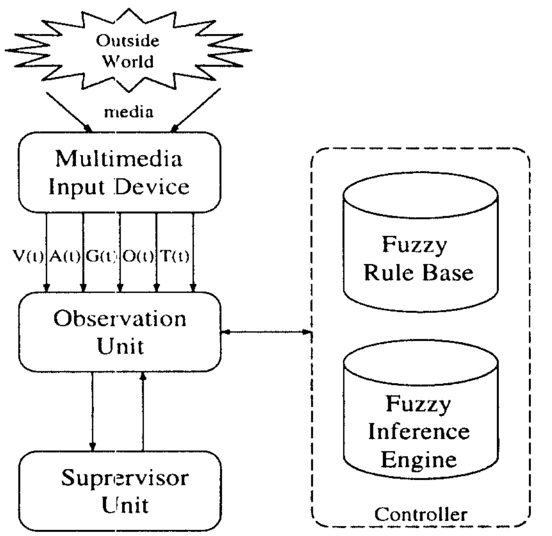

Fig. 1. Basic configuration of multimedia information data acquisition system.

When an instruction adjusted to some goal is indicated by the upper module (i.e., supervisor), the Observation Unit decides an appropriate time subset

$$[t_s, t_e] \, (\subset \mathrm{T}) \quad (2)$$

within T and obtains a multimedia information $VAGOT_{[t_s,t_e]}$ based on a given instruction as a type of information frame. But when instruction of the supervisor has temporal vagueness, the subset has to decide also if it needs ambiguous time interval. In this situation fuzzy rules are effectively introduced to infer the subset element with fuzziness. Which media to use and then how to combine the subset $[t_s, t_e]$ is determined by the supervisor.

But taking current multimedia technology and information input characteristics into account, here multimedia information data is considered centering around only the Visual Information and the Acoustic Information.

The difference method between image frames is a widely used preprocessing method in general image processing for the purpose of detecting moving objects or color changing [2]. So, this paper introduces an expression method of binary image which consists of the difference between background image and current image to analyze histogram of the binary image based on fuzzy membership function.

On the other hand, general sound processing such as frequency analysis requires considerable time for the complicated processing. Then here sound absolute intensity is represented by fuzzy membership function. Fuzzy rules with those membership functions decides the element of the goal subset.

3. Fuzzy inference engine for multimedia information data acquisition

Since the inception of fuzzy concept by Zadeh in 1965, the fuzzy logic has been used in a variety of situations. The capability of fuzzy logic to model the real-world situations has resulted in its wider application in diverse fields as well [3], [4], [5], [9].

A fuzzy set theoretic approach has the following advantages over traditional technologies: (1) it can easily deal with imprecise and vague properties, descriptions and rules, (2) it degrades more gracefully when the input information is incomplete, (3) a given task can be achieved with a more compact set of rules, (4) inference and uncertainty maintenance can both be carried out in one consistent framework.

The basic fuzzy rules are constructed by the Visual information, Acoustic Information and Visual and Acoustic Information.

Visual information has two membership functions, i.e., objects information on current image (V_{object}) and the difference between previous V_{object} and V_{object} ($V_{difference}$), e.g.,

IF V_{object} is LITTLE (MIDDLE, HUGE)
and $V_{difference}$ is SMALL (MEDIUM, LARGE)
THEN the t_s (or t_e) is M-th frame after,

where V_{object} is the difference value between previous image and current image.

Acoustic Information has also two membership functions, i.e., absolute intensity value of the current sound frame ($A_{intensity}$) and the temporal slope of the sound intensity (A_{slope}) which is obtained by linear regression method using previous five elements:

$$\{\bar{M}(t), \bar{M}(t-1), ..., \bar{M}(t-4)\},$$

$$\bar{M}(\tau) = \frac{1}{N}\sum_{i=0}^{N-1} M(\tau - i), N = 5. \quad (3)$$

Visual and Acoustic Information has two membership functions (V_{object} and $A_{intensity}$) or three membership functions (V_{object}, $A_{intensity}$ and $V_{difference}/A_{slope}$) or four membership functions (V_{object}, $V_{difference}$, $A_{intensity}$ and A_{slope}) as follows:

IF V_{object} is LITTLE (MIDDLE, HUGE)
and $A_{intensity}$ is WEAK (MIDDLE, STRONG)
and $V_{difference}$ is SMALL (MEDIUM, LARGE)
and A_{slope} is SMALL (MEDIUM, LARGE)
THEN the t_s (or t_e) is M-th frame after.

We call the unit that can infer based on those IF – THEN type fuzzy rules a fuzzy inference engine. Here, fuzzy inference is considered centering around MAX-MIN composition center of gravity method. Figure 2 shows the proposed algorithm of multimedia information acquisition based on fuzzy inference engine.

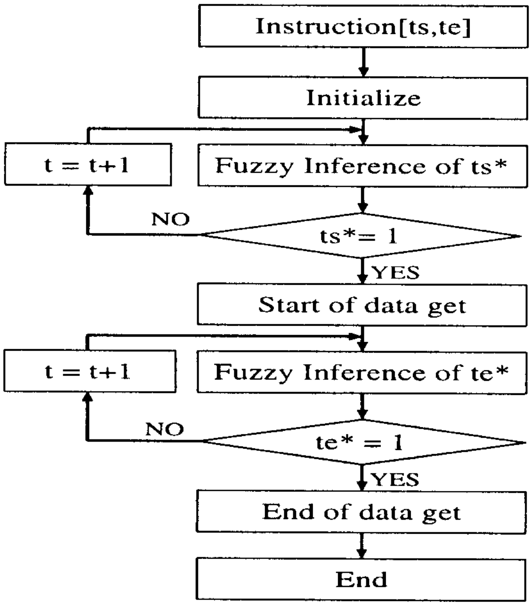

Fig. 2. flowchart of the algorithm to obtain multimedia information data.

4. Experiment of the multimedia information acquisition

Dynamic images and sound data for experiments are taken by digital video camera (Panasonic Co., NV-DJ1) with a microphone and converted to AVI file on video capture board (Ulead Systems Co., Media-Studio Version 3.0) of IBM PC. The image size is 320 × 240 pixels in RGB 256 color. The sampling between two image frames is 10 frame/sec. A sound frame is PCM type, 8 bit resolution and the sampling time is 11kHz.

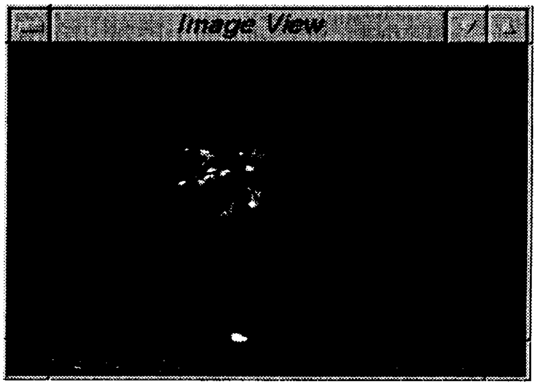

Fig. 3. An example of the fireworks (1 frame).

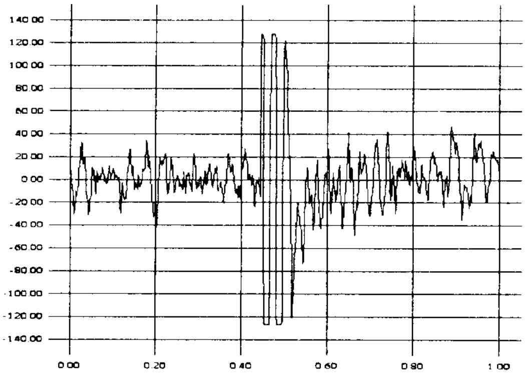

Fig. 4. The example wave form of launch sound(for 0.1 second).

4.1 *Information acquisition experiment of the fireworks*

Japanese fireworks are selected to confirm the validity of the above mentioned multimedia information acquisition system. Here, the Visual Information and Acoustic Information can be expressed as follows:

$$\{V(t), A(t)\}_{t \in T}. \quad (4)$$

The samples of the dynamic images and the sound waveforms are shown in Figures 3 and 4, respectively.

When the instruction of Supervisor Unit with the aim of the distance recognition between current camera position and the fireworks origin is "*Get the multimedia information VA* $_{[t_s,t_e]}$ *from the time* t_s *to the time* t_e*, where* t_s *is the time when the launch sound of the firecracker reaches, and* t_e *is the time when the firework finishes the fire opening process*",
then the goal is to extract the multimedia information $VA_{[t_s^*,t_e^*]}$ data, where t_s^* and t_e^* are the estimations of t_s and t_e, respectively.

In this example, the given instruction is crisp/fuzzy, i.e., the t_s is crisp and t_e is fuzzy.

Although the Visual Information cannot recognize the t_s completely, the fireworks environment can estimate it to some extent. The Acoustic Information can find the t_s clearly, it is crisp. So, in order to obtain t_s, the two sets of input membership functions are prepared for sound intensity of current frame ($A_{intensity}$) and the number of the firework color pixels (V_{object}).

On the other hand, the fuzzy t_e can be predicted by the dynamic image information to some extent. Therefore, the two sets of input membership functions are selected for the firework color pixels on the current image frame (V_{object}) and difference of the pixels between current and previous image frame ($V_{difference}$).

Experiments have been done on 3 fireworks and the results are shown in Table 1. Obtained multimedia data have various information. For example, we could compute the distance between the place of the launch of the firecracker and the camera position which was a goal of the supervisor as follows:

$$331.45 \times \sqrt{T/273} = 652.41526(m). \quad (5)$$

It is about 652m, because the time interval was 1.9 second and the temperature was about 20°C, i.e., $T = 273 + 20°K$.

The result of multimedia information acquisition, see the No.1 of Table 1, is shown in Fig.5 – 9, where Fig.5 is the time when the launch sound of the firecracker reaches, Fig.6 is the fire opening start time, Fig.7 is the most beautiful scene, Fig.8 is the time when the explosion sound of the Fig.6 scene reaches, Fig.9 is the time when the firework finishes the fire opening process.

TABLE 1
THE PERCENTAGE RATE OF POSSESSION OF FIREWORK COLOR PIXELS OF THE START TIME(t_s^*) AND THE END TIME(t_e^*). INSIDE [] INDICATES THE ABSOLUTE VALUE [0, 128] OF SOUND INTENSITY.

	at the time t_s^* % [intensity]	at the time t_e^* % [intensity]
No.1	1.5 [23.3]	46.6 [12.4]
No.2	0.5 [21.3]	42.8 [12.8]
No.3	0.7 [25.2]	45.5 [20.1]

5. Conclusion

A multimedia (dynamic image and sound) information data acquisition method based on fuzzy rules is proposed. When an instruction is given, observed multimedia information is characterized by *VAGOT* time series data and the proposed method with simple fuzzy inference is able to obtain the fireworks information data in real time. This method is applied to one experiment that is related with the temporal boundary subset of crisp/fuzzy. In this experiment, multimedia infor-

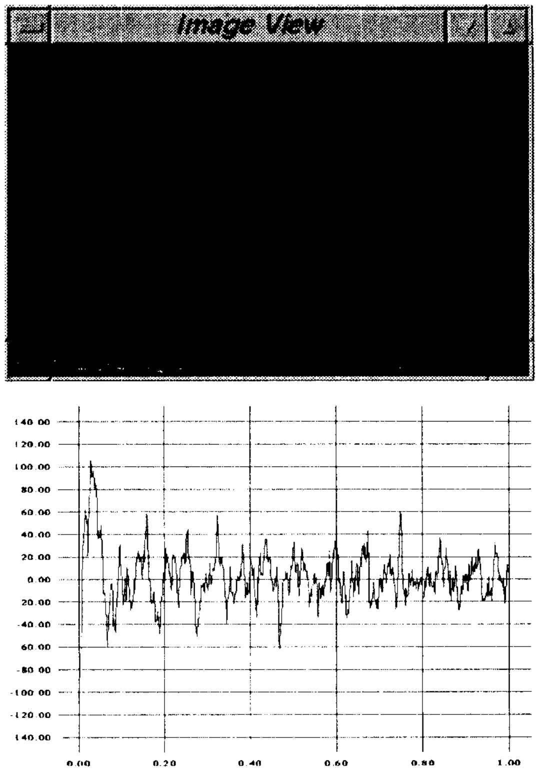

Fig. 5. The image and sound at Start frame of No.1

mation of the fireworks are completely obtained.

References

[1] T. Tomii, H. Arisawa "Video Image Modelling and Manipulation Language for Multimedia Database", The trans. of ICEIE D-II (in Japanese), Vol. J79-D-II, No. 4, pp. 520–530, 1996

[2] H. Choi, T. Agui, M. Nakajima, T. Yokomae "Extraction of Passing Car and Its Number Plate Using a Series of Images", The trans. of ICEIE D-II (in Japanese), Vol. J73-D-II, No. 3, pp. 392–398, 1990

[3] C. C. Lee "Fuzzy logic in control systems: Fuzzy logic controller – Part I", IEEE Trans. Syst. Man Cybern., Vol. 20, No. 2, pp. 404–418, 1990

[4] C. C. Lee "Fuzzy logic in control systems: Fuzzy logic controller – Part II", IEEE Trans. Syst. Man Cybern., Vol. 20, No. 2, pp. 419–435, 1990

[5] Y.ARAI, K.HIROTA "Fuzzy Hierarchical Pattern Recognition for Robotics Applications", Advanced Topics in Artificial Intelligence, Lecture Notes in A.I., 1342, pp. 274–281, 1997

[6] K. Oh, K.Hirota *Multimedia Information Acquisition for Robotics* , Soft Computing on Robotics, Springer-Verlag Press, to appear

[7] Y.Nakagawa, K.Hirota *A Propose on Fuzzy Communication System with An Intention Understanding for the Multimedia Computer System*, the second Joint Symposium related System Information (in Japanese), pp. 351–354, 1995

[8] K. Oh, K.Hirota *A Multimedia Information Acquisition Method Based on Fuzzy rules*, The World Automation Congress 1998 (WAC '98), accepted

[9] Y.Nakagawa, K.Hirota and W.Pedrycz "Dynamic Image Understanding on General Roads in Japan Using Fuzzy Frame Knowledge Base with Spool Base", Int. J. of Scientia Iranica, Vol. 2, No. 4, pp. 347–356, 1996

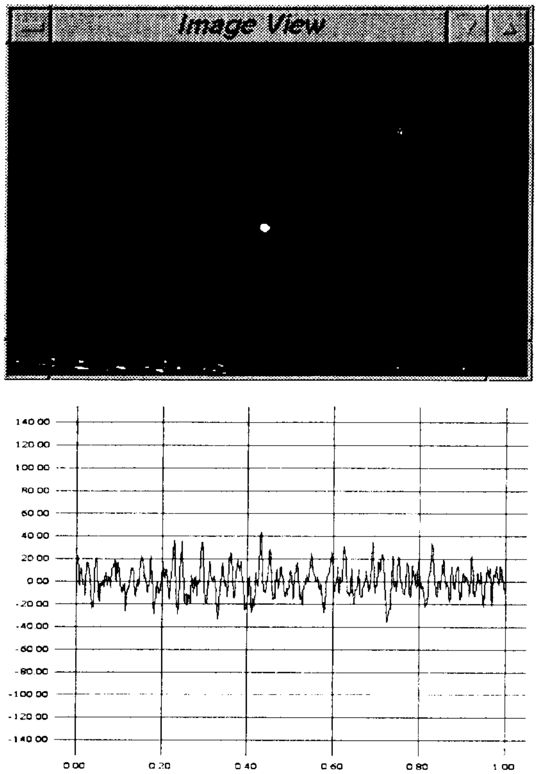

Fig. 6. The image and sound at 20-th frame of No.1

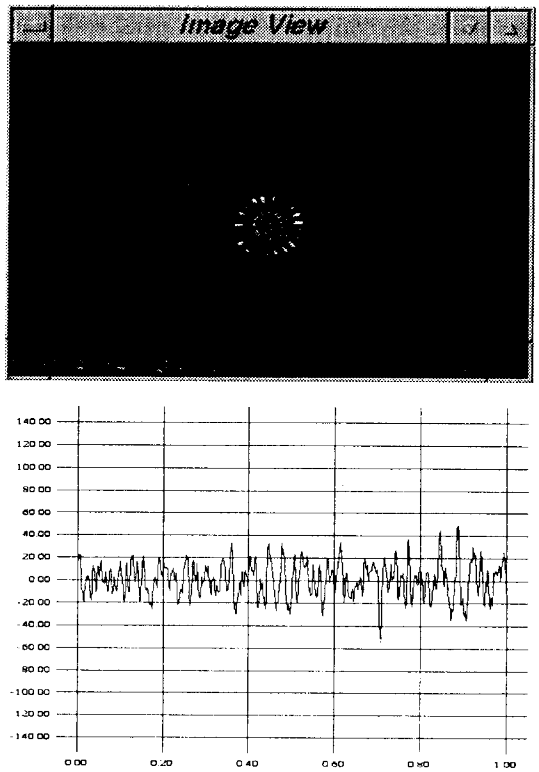

Fig. 7. The image and sound at 34-th frame of No.1

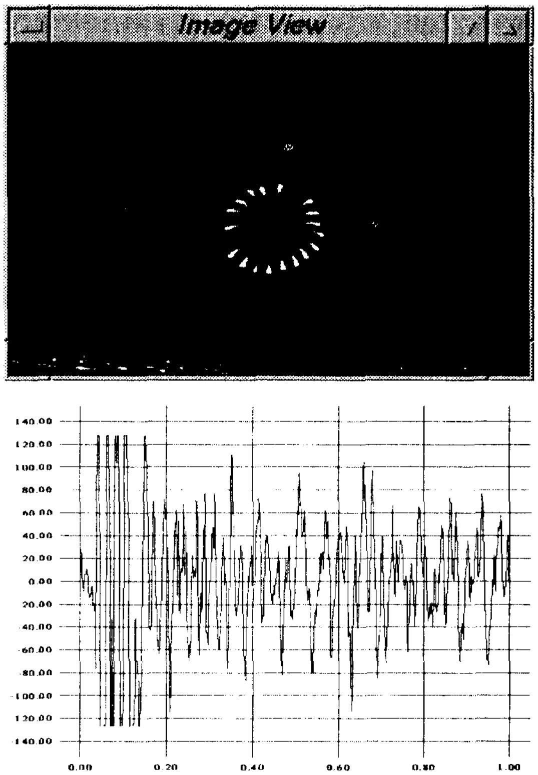

Fig. 8. The image and sound at 40-th frame of No.1

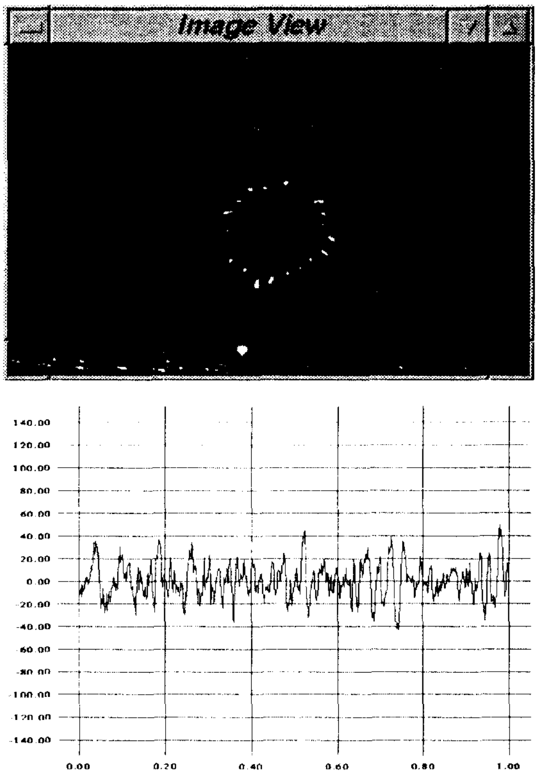

Fig. 9. The image and sound at End frame (47-th) of No.1