A Fuzzy Rule-Based System with Ontology for Summarization of Multi-camera Event Sequences

Han-Saem Park and Sung-Bae Cho

Dept. of Computer Science, Yonsei University
Shinchon-dong, Seodaemun-ku, Seoul 120-749, Korea
sammy@sclab.yonsei.ac.kr, sbcho@cs.yonsei.ac.kr

Abstract. Recently, research for the summarization of video data has been studied a lot due to the proliferation of user created contents. Besides, the use of multiple cameras for the collection of the video data has been increasing, but most of them have used the multi-camera system either to cover the wide area or to track moving objects. This paper focuses on getting diverse views for a single event using multi-camera system and deals with the problem of summarizing event sequences collected in the office environment based on this perspective. Summarization includes camera view selection and event sequence summarization. View selection makes a single event sequence from multiple event sequences as selecting optimal views in each time, for which domain ontology based on the elements in an office environment and rules from questionnaire surveys have been used. Summarization generates a summarized sequence from a whole sequence, and the fuzzy rule-based system is used to approximate human decision making. The degrees of interests input by users are used in both parts. Finally, we have confirmed that the proposed method yields acceptable results using experiments of summarization.

1 Introduction

Recently, the popularization of digital cameras and advancement of data compression and storage techniques made it possible for most people to access and use the video data easily [1]. Accordingly, people can obtain the video data in various ways from using the CCTV in the public space to using a personal portable device. The video data are more useful than text documents, voices and still images because they contain much more specific and realistic information. Therefore, it has become very important to extract information that people want to search, and the studies of analyzing and summarizing video data have been investigated from all over the world [1, 2].

There are various types of target video for summarization. Some videos are made by experts like movie, news, and sports, and the some other videos are collected by researchers in a preset indoor environment. The first type has clear scenes and shots divided by backgrounds. Summarization process of these videos includes content analysis, structure parsing and summarization [2]. Content analysis step maps low-level features to high-level semantic concepts. This can be

L. Rutkowski et al. (Eds.): ICAISC 2008, LNAI 5097, pp. 850–860, 2008.

conducted automatically using image processing and recognition techniques [3] or manually [4]. Structure parsing step divides the video into scenes and shots based on the result of content analysis, and summarization step provides significant parts of videos using analyzed information in previous steps. The second type videos are collected by researchers. They usually collect the data in an indoor environment and use a multi-camera system [5-6]. These videos do not have clear scenes because the backgrounds are static indoor environments, so they are divided by either the location of the users [5] or activities occurred in that domain [6]. The analyses of activities or events can provide valuable information for studies on human behavior, the design of indoor environment, etc. Besides, the summarization service can help people to remember what they did in the past [5].

This paper targets the video collected in an office environment with the multi-camera system. The proposed method subdivides the target videos into event sequences, selects the optimal camera views, and summarizes event sequence considering the degree of interests input by users. For view selection, we have made the domain ontology describing the elements in the office and the rules from questionnaire surveys. For summarization, we have used the fuzzy rule-based system to approximate a human decision making process in events evaluation [8]. The users should input the degrees of interests to each event, person, and object so that the system can provide personalized summary of target video.

2 Multi-camera System

Most conventional multi-camera systems used the multiple cameras to cover wide area and to track moving persons and objects. J. Black et al. developed the multi-camera system that tracked and extracted the moving object in an outdoor environment [9], and they attempted to apply the system to an indoor environment. Another main purpose of using the multi-camera system is in a security. F. Porikli set several cameras in a building to track the moving objects and summarize that information [10]. Recently, studies for summarization and retrieval or studies for human activity analyses have been investigated. Y. Sumi et al. analyzed the subjects using the sensors including multiple cameras and provided a simple summarization [6], G. C. Silva et al. exploited the multi-camera system to summarize and retrieve information of persons living in a home-like ubiquitous environment [5].

However, the multi-camera systems have another use where they can provide diverse views using the multiple cameras. Because a single view for a certain activity or event might not provide the exact information, we can obtain this correct and diverse information through multiple views. C. Zhang et al. extracted the event information using IR tags and four cameras, and provided the retrieval system that searched for event the user wanted with a simple query [7]. They focused on an advantage of diverse views by the multi-camera systems. We also

focus on this possibility, and also we used the users' degrees of interests to events, persons, and objects to provide the personalized summary of event sequences.

3 Multi-camera Office Environment

3.1 Setting Multi-camera System in Office Environment

To collect the office event sequence, we set eight cameras in the lab, shown in the left figure of Figure 1. We set all cameras to focus on the same area so that the system can provide diverse views for a single event. The right figure illustrates a captured example of different views.

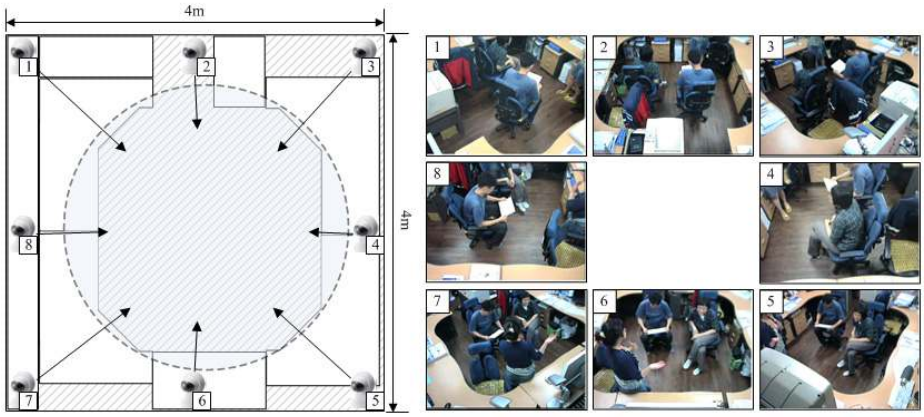

Fig. 1. Cameras set in an office environment and the target area (left) and an example of event captured by the multi-camera system (right)

3.2 Annotation of Events, Persons, and Object

This paper regards the collected video as an event sequence and provides the summarization service with that sequence. All basic information including events, persons, objects, and their positions have been annotated manually, and these works have been performed based on the event definition as follows.

- Entry (A),
  if stand (A, entrance-area) and face (A, in)
- Leaving (A),
  if stand (A, entrance-area) and face (A, out)
- Calling (A),
  if hold (A, phone) and speak (A)
- Vacuuming (A),
  if hold (A, vacuum cleaner) and stand (A, center-area)
- Eating (A),
  if hold (A, food)

- Resting (A),
  if rest (A, corner-area) or stretch (A, corner-area)
- Work (A),
  if sit (A, corner-area) and {use (A, computer) or hold (A, document)}
- Printing (A),
  if exist (A, printer-area) and hold (A, printout)
- Conversation (A, B),
  if {exist (A, x) and exist (B, y) and close (x, y)} and
  {speak (A) or speak (B)}
- Meeting (A, B, C),
  if {exist (A, x) and exist (B, y) and exist (C, z) and close (x, y, z)} and
  {speak (A) or speak (B) or speak (C)} and
  {hold (A, document) or hold (B, document) or hold (C, document)}
- Seminar (A, B),
  if {stand (A, screen-area) and sit (B, center-area)} and speak (A)

These eleven events are normal ones that could happen in the office environment. We have decided these events according to the related works dealing with human activities or events. Some of these works classified the events based on objects such as phone, table, chair, book, keyboard [11], and some other works classified the events based on persons and their activities [12]. Many of the previous video summarization studies have performed the annotation automatically or semi-automatically using image processing and pattern recognition techniques [1, 3]. The current version of the system relies on manual annotation, and it will be replaced by automatic annotation in the future work.

4 The Proposed Method

Figure 2 provides an overview of the proposed summarization system. As mentioned before, summarization process is divided into a view selection module and a summarization module. View selection, which selects an optimal view for a single point in time, is performed using rules based on domain ontology. Summarization is performed using fuzzy rule-based system. In summarization, fuzzy system evaluates each event in event sequence, and then events with high evaluation score are selected as important ones.

4.1 User Input

To summarize an office event sequence considering users' personal interests, we attempt to use user input of the degree of interest (DOI) to each event, person, and object. User interest is an important factor to design an interface or to interact with users. A user input has a form of integer between 0 (not interested) to 3 (interested a lot), and it can be used for view selection and the personalized summarization.

We have segmented the event sequence into shots. Shot, a basic unit for the video data, is defined as a set of consecutive frames with the same event, person,

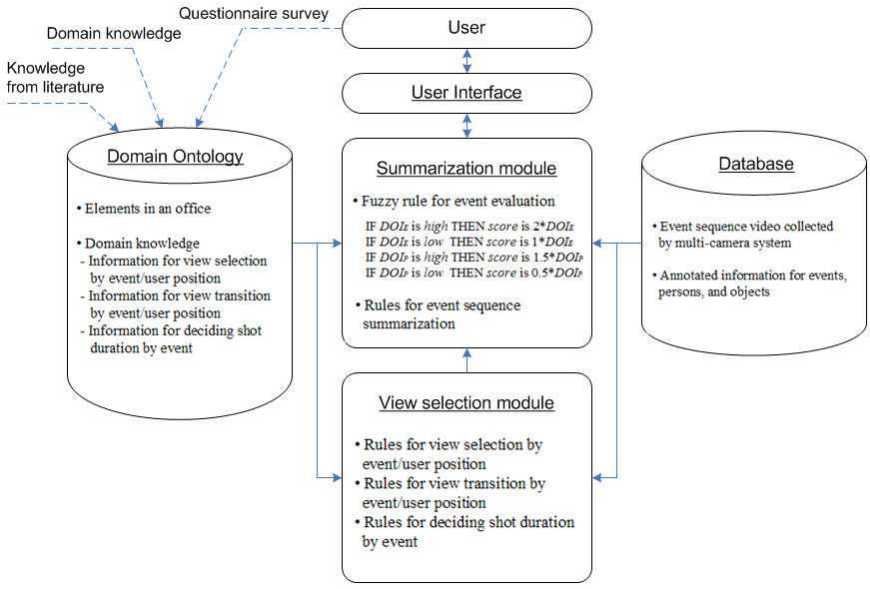

Fig. 2. An overview of the proposed summarization system

and object. DOI value has been calculated by shot, and DOI those one shot Si are defined as follows.

$$DOI^2_{E_j,S_i} = \sum_{frame f \in S_i} DOI_{E_j}(f) \tag{1}$$

$$DOI^2_{P_k,S_i} = \sum_{frame f \in S_i} DOI_{P_k}(f) \tag{2}$$

$$DOI^2_{O_l,S_i} = \sum_{frame f \in S_i} DOI_{O_l}(f) \tag{3}$$

Equations (1) to (3) show the DOI value for one shot S_i to event E_j, person P_k, and object O_l. Adding DOI values of all frames in a shot, its square root is used as a final DOI value so that the duration does not have a significant influence.

4.2 Domain Ontology

Domain ontology comprises elements in an office and domain knowledge for view selection. The former describes each element in office environment and their relationships, and the latter describes information required to design rules for view selection. Figure 3 shows the element description of "Place" and "Meeting". Domain knowledge includes information for camera view selection and view

```
<owlr:Rule>
        <owlr:antecedent>
                <owlr:individualPropertyAtom owlr:property="locate">
                        <owlr:variable owlr:name="A">
                        <owlr:variable owlr:name="2">
                </owlr: individualPropertyAtom>
                <owlr:individualPropertyAtom owlr:property="happen">
                        <owlr:variable owlr:name="Work">
                </owlr: individualPropertyAtom>
        </owlr:antecedent>
        <owlr:consequent>
                <owlr:individualPropertyAtom owlr:property="view">
                        <owlr:variable owlr:name="3">
                </owlr: individualPropertyAtom>
        </owlr:consequent>
</owlr:Rule>
```

Fig. 3. Domain knowledge description: An information description for view selection

transition by event and user position, and information for deciding shot duration by event. Most of this information is based on the questionnaire surveys and their analyzed results. Questionnaires were surveyed by ten graduate students. Figure 4 illustrates an example of information description for view selection, which means "If person A locates in area 2 and event Working happens, and then view #2 should be selected." Representation format has referred the syntax for ORL (Owl Rule Language) [13].

4.3 View Selection and Event Sequence Summarization

View selection generates a single event sequence from multi-camera event sequences. It includes a selection of an optimal event at the same point in time and a selection of an optimal view among views showing the same event. Previous works exploited simple rules to select one sensor among many of them. Y. Sumi et al. selected a sensor that had a high priority based on predefined priority [7]. To select an optimal camera view considering all variables including

```
Procedure ViewSelection
var       N: the number of events in given time point
          M: the number of views for a given event
function  SelectEvent(Ei): a function that select an event with the highest DOIE
          SelectView(Ei, Vj): a function that select a view for an event Ei based on rules
                              in domain ontology
begin
        for i=1 to N
                SelectEvent(Ei)
        for j=1 to M
                SelectView(Ei, Vj)
end
```

Fig. 4. A pseudo-code for a view selection

IF (DOI_E is high and DOI_P is high and DOI_O is high) THEN score = $2\times DOI_E + 2\times DOI_P + 1\times DOI_O$

IF (DOI_E is high and DOI_P is high and DOI_O is low) THEN score = $2\times DOI_E + 2\times DOI_P + 0.25\times DOI_O$

IF (DOI_E is high and DOI_P is low and DOI_O is high) THEN score = $2\times DOI_E + 0.5\times DOI_P + 1\times DOI_O$

IF (DOI_E is high and DOI_P is low and DOI_O is low) THEN score = $2\times DOI_E + 0.5\times DOI_P + 0.25\times DOI_O$

IF (DOI_E is low and DOI_P is high and DOI_O is high) THEN score = $0.5\times DOI_E + 2\times DOI_P + 1\times DOI_O$

IF (DOI_E is low and DOI_P is high and DOI_O is low) THEN score = $0.5\times DOI_E + 2\times DOI_P + 0.25\times DOI_O$

IF (DOI_E is low and DOI_P is low and DOI_O is high) THEN score = $0.5\times DOI_E + 0.5\times DOI_P + 1\times DOI_O$

IF (DOI_E is low and DOI_P is low and DOI_O is low) THEN score = $0.5\times DOI_E + 0.5\times DOI_P + 0.25\times DOI_O$

Fig. 5. TSK fuzzy rules for event evaluation

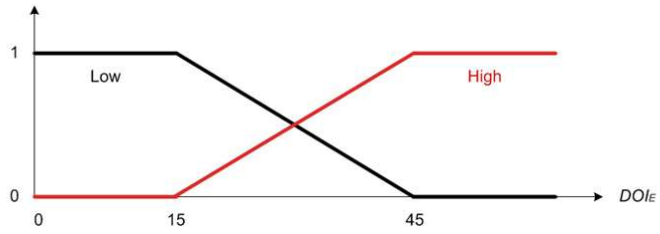

Fig. 6. A fuzzy membership function for DOI_E

user input, the proposed system constructed the domain ontology with the domain knowledge through questionnaire survey. Figure 4 shows a pseudo-code explaining a view selection process.

Next, the system evaluated each event in single event sequence generated by view selection to summarize it. TSK fuzzy system has been utilized in this paper. Fuzzy rules designed by domain knowledge in domain ontology are as follows. Getting users' DOI values as an input, the system calculates the final score for each event.

Before the fuzzy inference with these fuzzy rules, a fuzzy membership function shown in Figure 6 has been used to fuzzify the user inputs, DOI values, so that they could be used as input variables of the fuzzy system. It depicts a fuzzy membership function for DOI_E, and functions for DOI_P and DOI_O also have the trapezoidal function. This type of membership function is very simple, but widely used [14].

In summarization step, important events are selected by the rank based on the evaluated score. Here, shots of one long event, which splitted into several events due to an event happening in-between, cannot be selected more than twice, and events with low evaluated score were excluded from summary. Duration of each event is also based on the domain ontology, and central frames were selected for summary.

5 Experiments

5.1 Scenario and Data Collection

Experimental data were collected in the office environment presented in section 3. We designed a realistic scenario that could happen in an office assuming three persons in one day (from 9:00 a.m. to 6:00 p.m.). Figure 7 illustrates this scenario. Here, EN, PR, CONVERS, CA, and LE stand for entry, printing, conversation, calling and leaving, respectively.

For data collection, Sony network camera (SNC-P5) were used, and video were saved with the resolution of 320 240 and frame rate of 15 fps using MPEG video format.

5.2 Result of Summarization

Based on the scenario in Figure 7, we have made a single event sequence with view selection, and user inputs were assumed as Table 1 and 2. DOI values to objects were assumed as 0.

First, we have selected a single event sequence with view selection process. Most events in selected view were related to person C and event Vacuuming, Printing, Meeting, and Seminar, which had high DOI values. Subsequently, experiments for summarization have been conducted using this selected event sequence. This sequence contains 29 shots, and user inputs were assumed as in section 5.2. Figure 10 provides finally evaluated score by shots using fuzzy rules from DOI_E and DOI_P values.

In Figure 8, shots with the highest scores are shot #20 and shot #23 #26. The common characteristics of these shots are as follows. They are related to

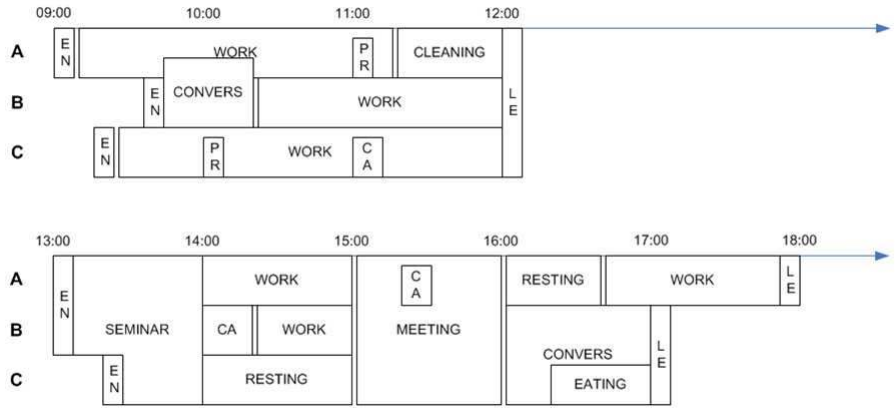

Fig. 7. A scenario of office events in one day

Table 1. User input (DOI value to event)

Event	DOI_E
Entry	2
Leaving	1
Calling	1
Vacuuming	3
Eating	0
Resting	0
Work	1
Printing	3
Conversation	2
Meeting	3
Seminar	3

Table 2. User input (DOI value to person)

Person	DOI_P
A	1
B	1
C	3

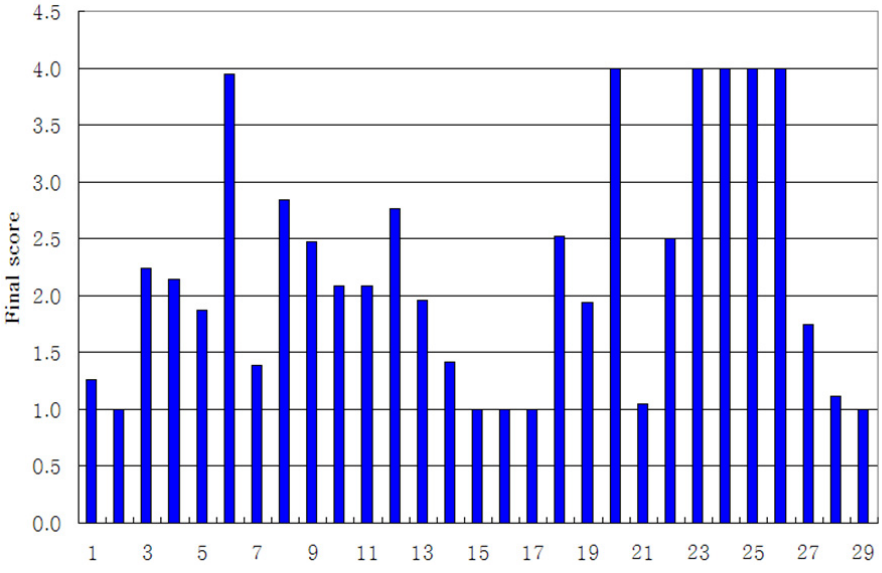

Fig. 8. Change of finally evaluated score by shot number

Table 3. A summarized event

S#	C#	F#(start)	F#(end)	Person	Event	Score
3	6→1	400	570	**C**	Entry	2.2
4	1	616	690	**C**	Work	2.1
6	1→6→1	950	1230	**C**	**Printing**	3.9
9	3→6→3	1810	2050	A	**Printing**	2.5
10	1	2065	2140	**C**	Calling	2.1
12	6→3→1→6	2395	2750	A	**Vacuuming**	2.8
13	1→6	2750	2900	**C**	Leaving	2
20	2→3	4216	4365	A, B, **C**	**Seminar**	4
22	1	5081	5155	**C**	Resting	2.5
25	2	6441	6515	A, B, **C**	**Meeting**	4
26	1	6731	6805	B, **C**	**Meeting**	4
27	6	7241	7335	**C**	Leaving	1.8

events with high DOI_E values. They are also related to person C, which has a high DOI_P value, and they have long duration. Shot #6 is exceptionally short, but it is about Printing event, which also has a high DOI_E value, and person C.

Table 3 shows a summarized event sequence. Here, S# means the shot number, C# means the camera view number, and F# means the frame number, and more than two views have been selected for C# in case of events with movement. Score represents a finally evaluated score by fuzzy rules. View selection in view transition has been performed using domain knowledge in domain ontology as described in section 4.2.

6 Conclusions

We collected the video with office events in a multi-camera environment and proposed the summarization system. The system generated a single event sequence from manually annotated event sequences and summarized the sequence using fuzzy rule-based system. For view selection, the domain ontology based on questionnaire surveys and literatures was used. Also, users' degrees of interest were used for personalized summarization. With experiments of view selection and summarization, we confirmed that the summarized event sequence was acceptable.

Future work will focus on the automatic annotation of office events. The design and implementation of user friendly interface to use the proposed method is required to help easy use and effective presentation. Also, subjective test to evaluate the experimental result will be performed.

Acknowledgement. This research was supported by MIC, Korea under ITRC IITA-2008-(C1090-0801-0011).

References

1. Zhu, X., et al.: Hierarchical video content description and summarization using unified semantic and visual similarity. Multimedia Systems 9(1), 31–53 (2003)
2. Li, Y., et al.: Techniques for movie content analysis and skimming. IEEE Signal Processing Magazine 23(2), 79–89 (2006)
3. Tseng, B.L., et al.: Using MPEG-7 and MPEG-21 for personalizing video. IEEE Multimedia 11(1), 42–53 (2004)
4. Petkovic, M., Jonker, W.: An overview of data models and query languages for content-based video retrieval. In: Int. Conf. on Advances in Infrastructure for E-Business, Science, and Education on the Internet. l'Aquila, Italy (2000)
5. Silva, G.C., et al.: Evaluation of video summarization for a large number of cameras in ubiquitous home. In: Proc. of the 13th ACM Int. Conf. on Multimedia, pp. 820–828 (2005)
6. Farwer, Sumi, Y., Mase, K., et al.: Collaborative capturing and interpretation of interactions. In: Pervasive 2004 Workshop on Memory and Sharing of Experiences, pp. 1–7 (2004)
7. Zhang, C.C., et al.: MyView: Personalized event retrieval and video compositing from multi-camera video images. In: Smith, M.J., Salvendy, G. (eds.) HCII 2007. LNCS, vol. 4557, pp. 549–558. Springer, Heidelberg (2007)
8. Dorado, A., et al.: A rule-based video annotation system. IEEE Trans. on Circuits and Systems for Video Technology 14(5), 622–633 (2004)
9. Black, J., Ellis, T.: Multi-camera image tracking. Image and Vision Computing 24, 1256–1267 (2006)
10. Porikli, F.M., Divakaran, A.: Multi-camera calibration, object tracking, and query generation. In: IEEE Int. Conf. on Multimedia and Expo, vol. 1, pp. 653–656 (2003)
11. Fogarty, J., et al.: Predicting human interruptibility with sensors. ACM Trans. on Computer-Human Interaction 12(1), 119–146 (2005)
12. Oliver, N., et al.: Layered representations for learning and inferring office activity from multiple sensory channels. Computer Vision and Image Understanding 96(2), 163–180 (2004)
13. Horrocks, I., Patel-Schneider, P.F.: A proposal for an OWL rules language. In: Proc. of the 13th Int. World Wide Web Conf., pp. 723–731 (2004)
14. Lertworasirikul, S., et al.: Fuzzy data envelopment analysis (DEA): A possibility approach. Fuzzy Sets & Systems 139(2), 3–29 (1998)