Automatic Summarization Method for First-person-view Video Based on Object Gaze Time

Keita Hamaoka[1,*], Yasuyuki Kono[1]

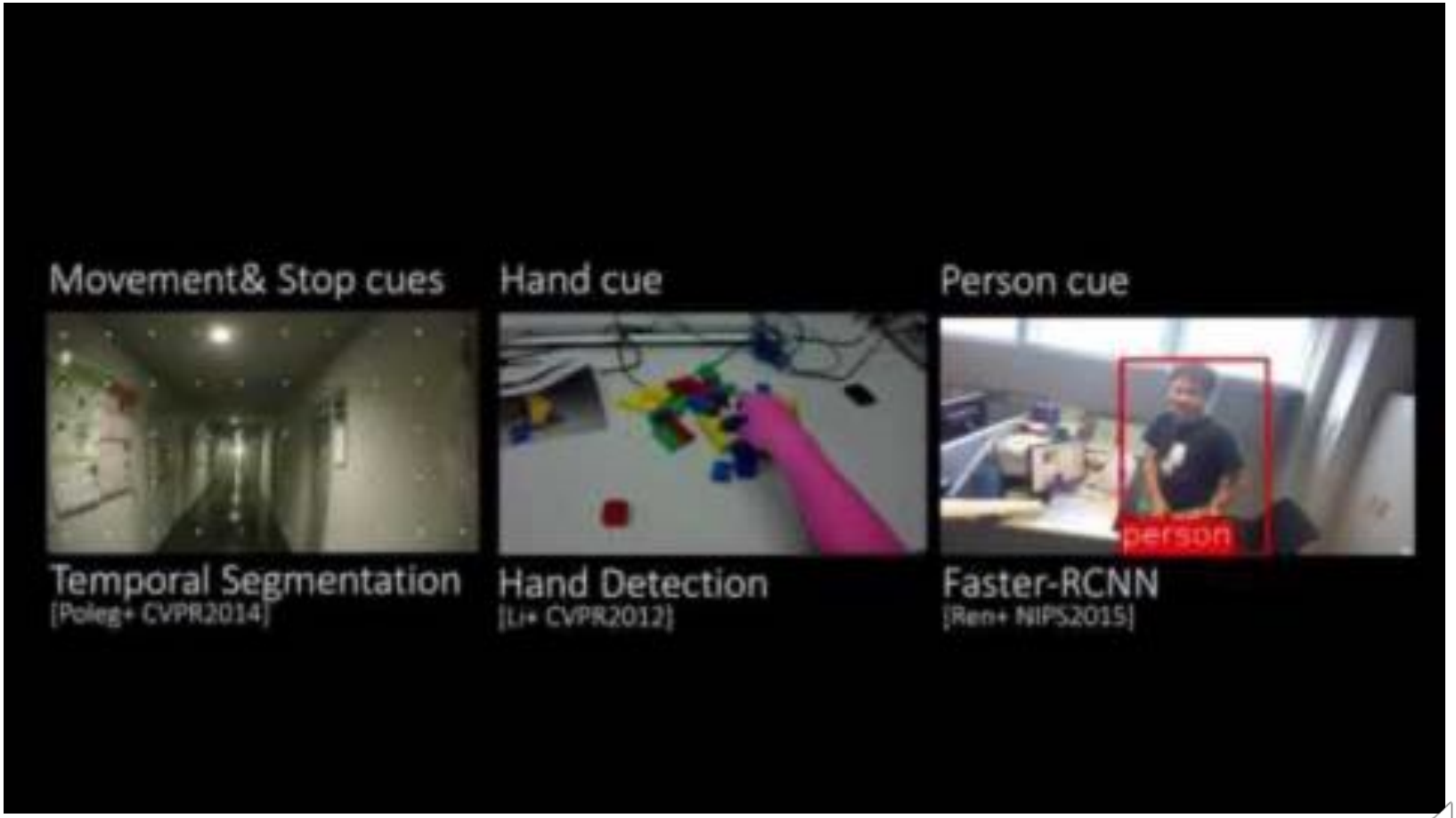
[1] Kwansei Gakuin University
{dwd11537, kono}@kwansei.jp

konolab
Human Computer Interaction

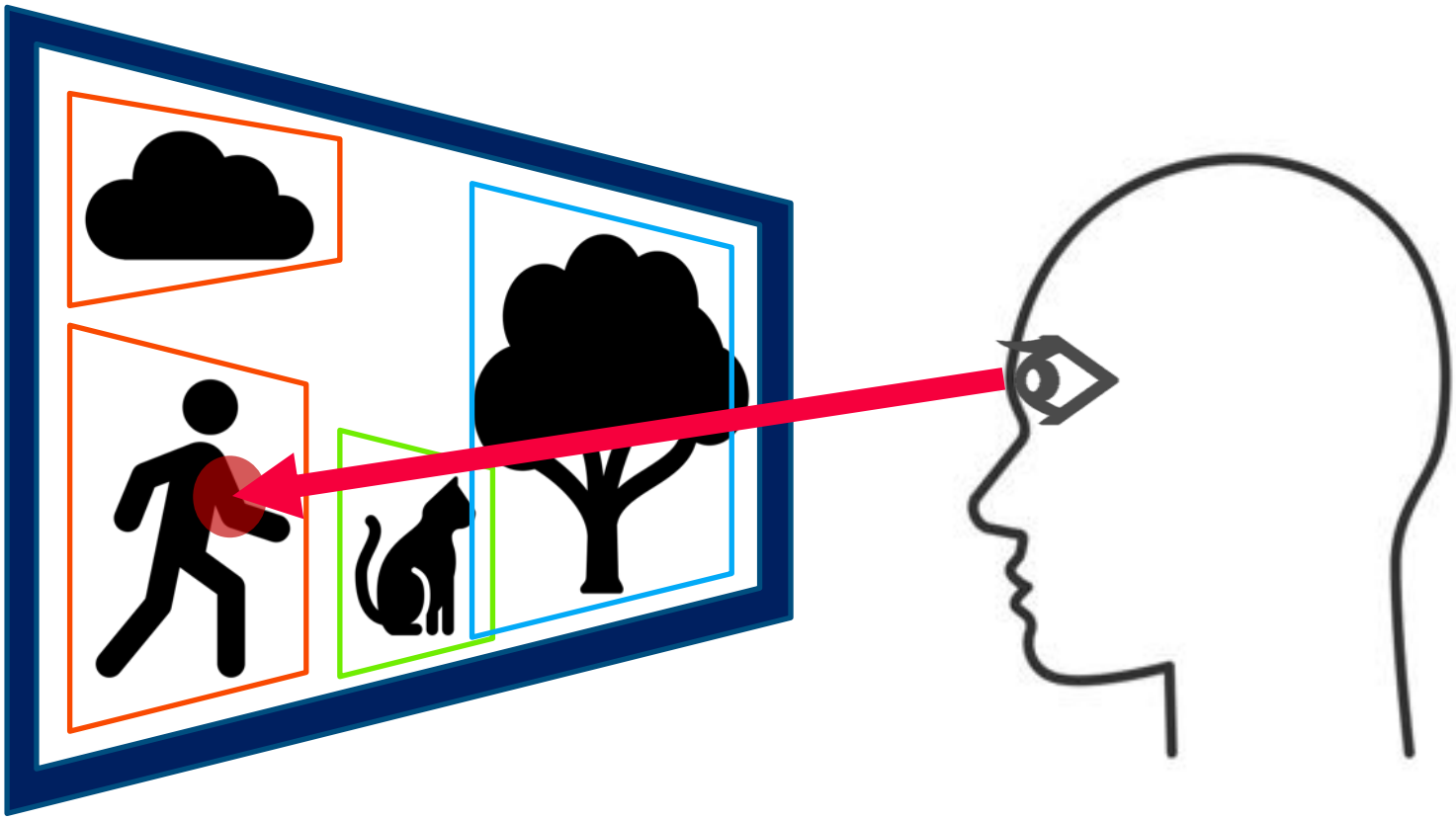
1 Introduction

konolab
Human Computer Interaction

1.1 Related Work



2 System Overview

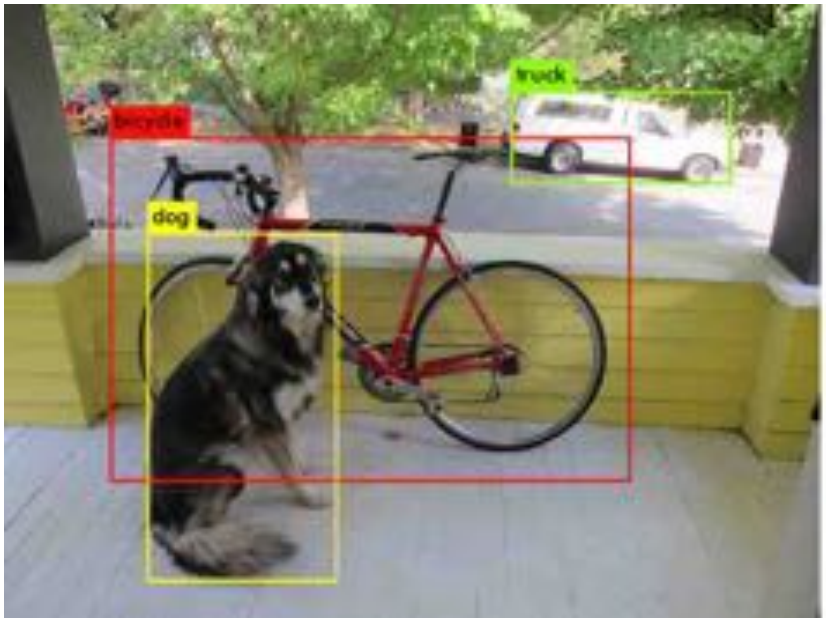


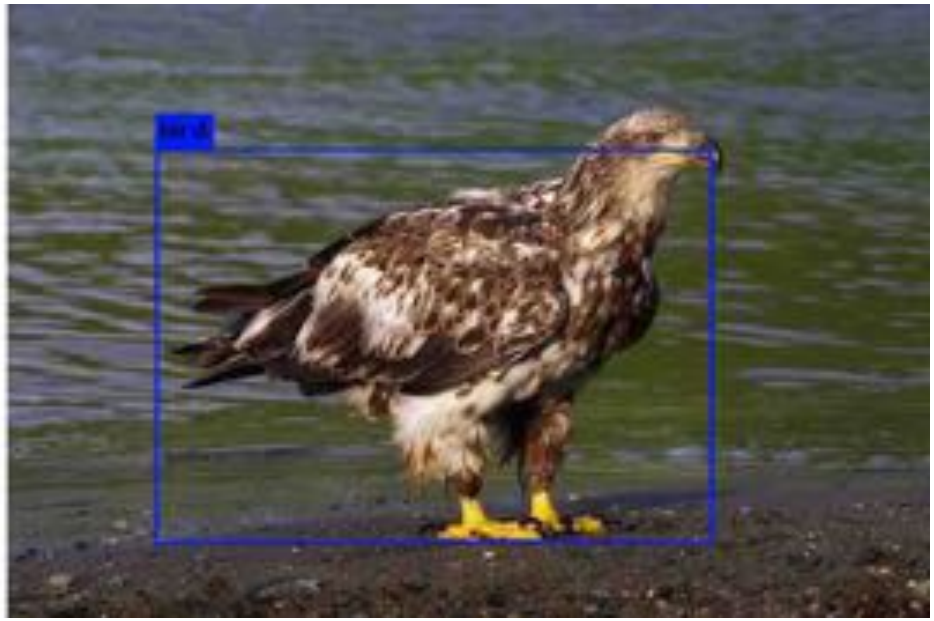
3.1 Gazing-point Extraction

- Since the acquired data is noisy, it is smoothed
- Smoothing using weighted average

$$P_{fixation} = \frac{(1P_0 + 2P_1 + \ldots + nP_{n-1})}{(1 + 2 + \ldots + n)}$$

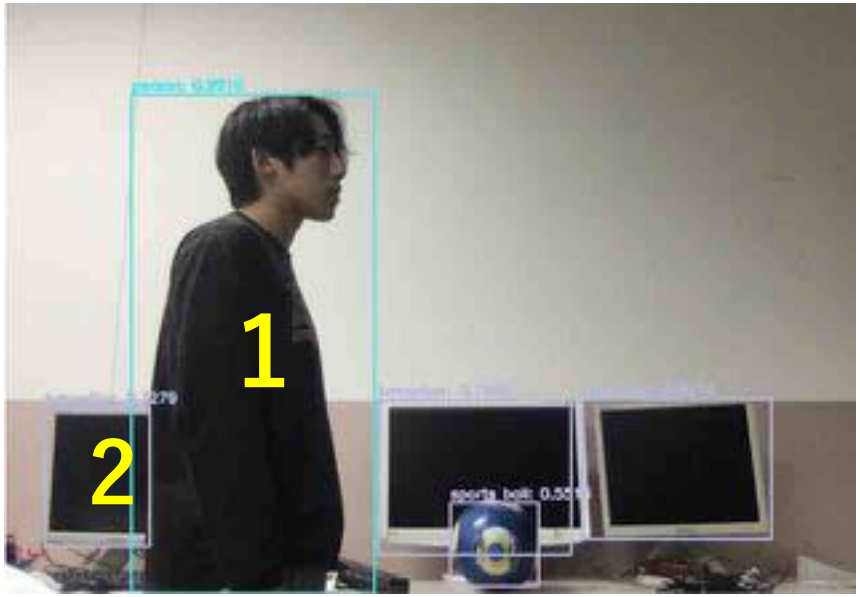
3.2 Object-area Extraction



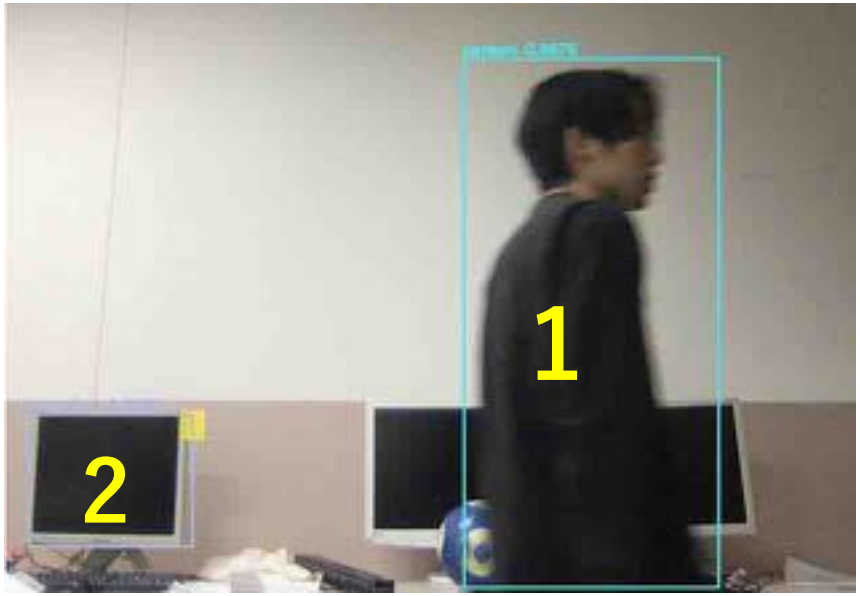


3.2 Object-area Extraction

- Compare Euclidean distance of each object area



Preceding frame

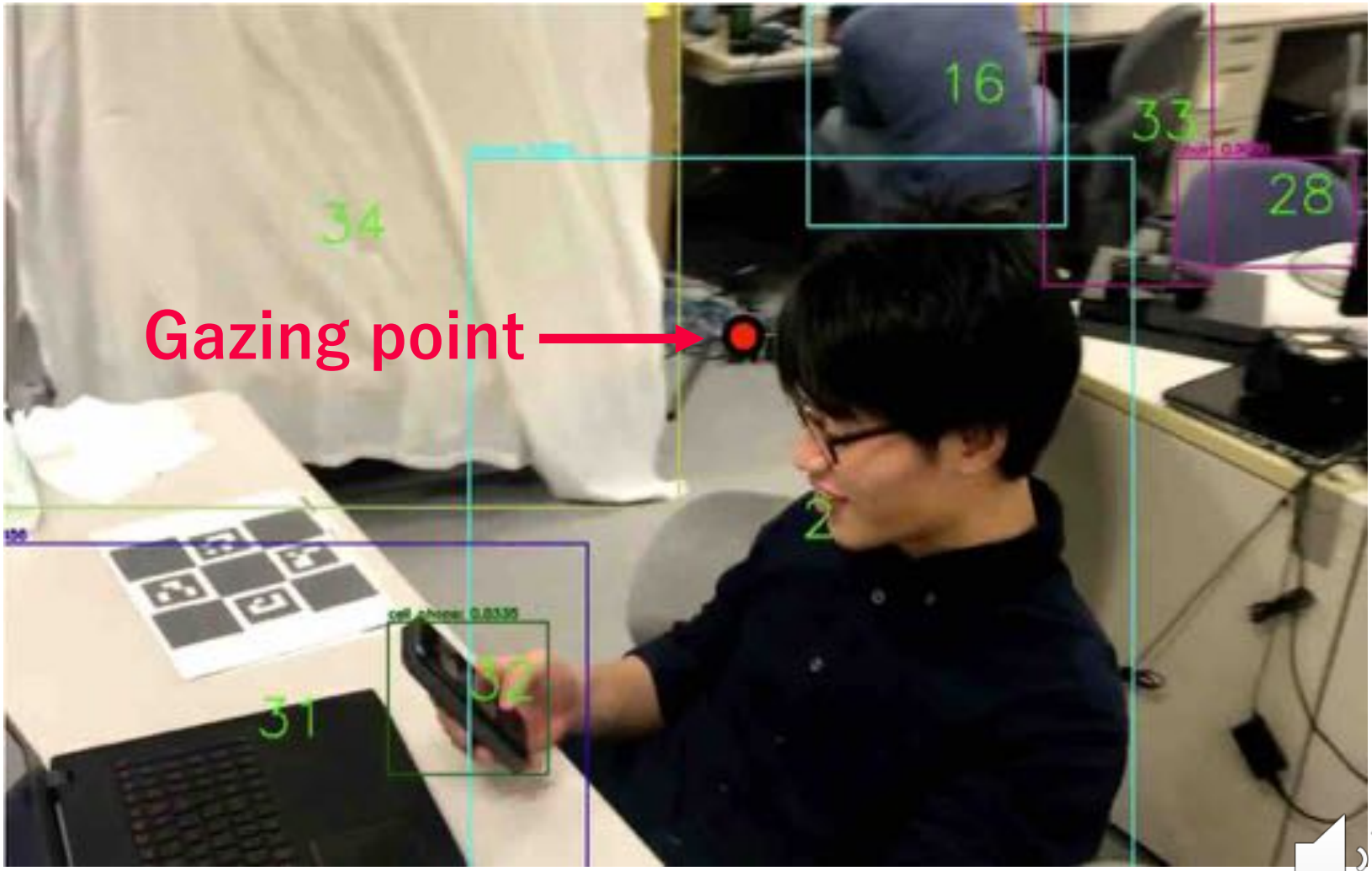


Current frame

3.2 Object-area Extraction

konolab
Human Computer Interaction

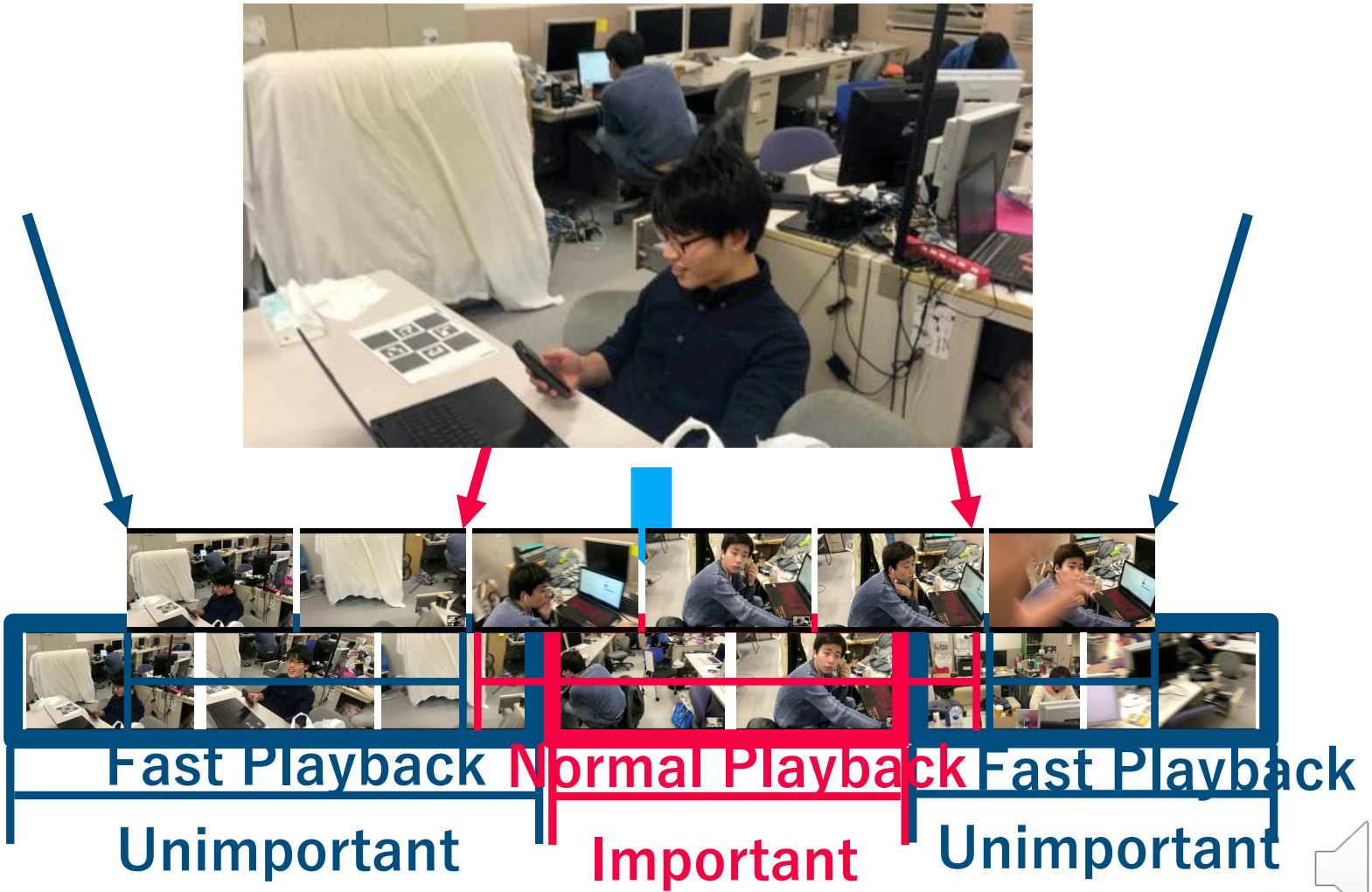
3.3 Object-gazing-time Calculation



4.1 Important-scene Extraction

- Set scene importance based on the gaze time for an object
- If the gaze time for an object exceeds the threshold, set it as an important scene
- Threshold can be changed

4.2 Generation of Summary Video Based on Importance of the Scenes



konolab
Human Computer Interaction

5.1 Experimental Method

- Investigate the usefulness of this system
- Compare with summary video that randomly extracted important scenes
- 9 subjects (6 males and 3 females)

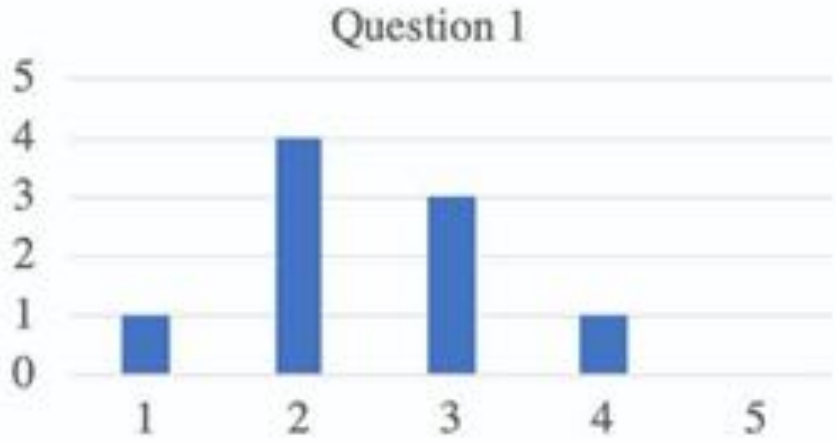
5.1 Experimental Method

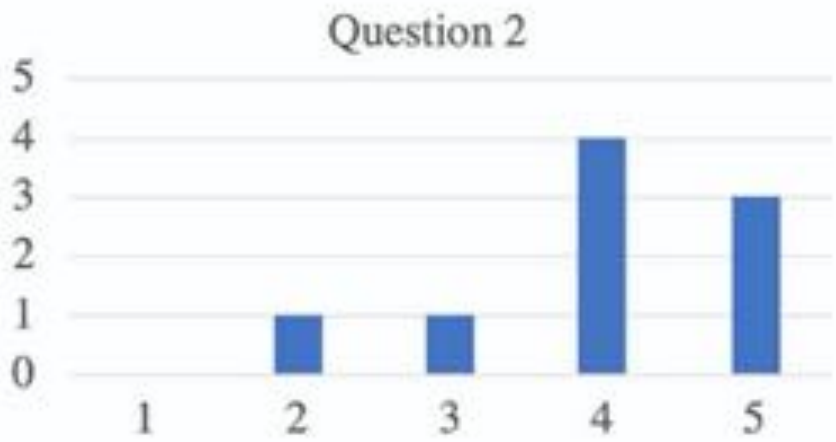
- Summarize questionnaires based on a 5-point Likert scale

	Questionnaire content
Question 1	Did you feel tired while watching the summary video?
Question 2	Is this system useful for high-speed viewing of videos?
Question 3	Is the interest reflected in the video after the summary? (The summary video generated by the proposed system)
Question 4	Is the interest reflected in the video after the summary (The summary video in which important scenes are randomly extracted)

5.2 Result

	Questionnaire content
Question 1	Did you feel tired while watching the summary video?
Question 2	Is this system useful for high-speed viewing of videos?

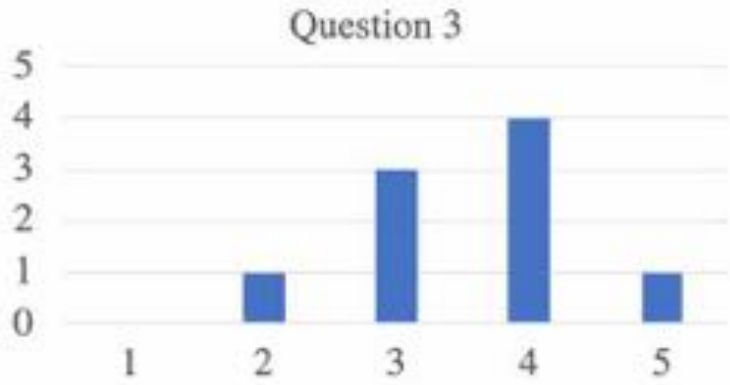


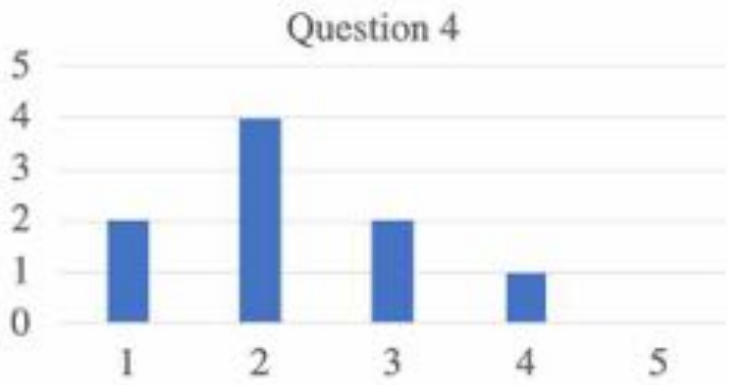


	Average	Standard deviation
Question 1	2.4	0.83
Question 2	4.0	0.94

5.2 Result

	Questionnaire contents
Question 3	Is the interest reflected in the video after the summary? (The summary video generated by proposed system)
Question 4	Is the interest reflected in the video after the summary (The summary video in which important scenes are randomly extracted)





	Average	Standard deviation
Question 3	3.7	0.81
Question 4	2.2	0.92

6. Concluding Remarks

- Purpose : The summarization of long-term first-person video
- Technique : An automatic summarization system based on object gaze time
- Result : The usefulness of this system was confirmed
- Future Work : Regarding a user's pupil diameter