An Evaluation of Input Data Quality of Lifelog Analysis Application with a Framework based on Quantitative Index

Akika Yamashita
Ochanomizu University
2-1-1, Otsuka, Bunkyo-ku
Tokyo 112-8610, JAPAN
akika@ogl.is.ocha.ac.jp

Saeko Iwaki
Ochanomizu University
2-1-1, Otsuka, Bunkyo-ku
Tokyo 112-8610, JAPAN
saeko@ogl.is.ocha.ac.jp

Masato Oguchi
Ochanomizu University
2-1-1, Otsuka, Bunkyo-ku
Tokyo 112-8610, JAPAN
oguchi@computer.org

ABSTRACT

In recent years, by the improvement of the data acquisition technology and the development of storage, it has become greatly easier than before to collect lifelog that is to record the person's behavior as digital data. As a result, various lifelog analysis applications have been developed that offer the user profitable information such as person's action histories with an analysis of collected data by sensor terminals, video cameras, and so on.

However, in these lifelog analysis applications, the quality of the data that was collected from the sensor terminals and inputted to the application was not discussed in detail. Therefore, in this paper, we have focused on the quality of video image data and the acceleration data of objects. As a representative lifelog analysis application, we have chosen an application which verbalizes person's behavior from the data, and shown the influence of the quality of input data on the execution result of the application by a quantitative index.

An evaluation framework is proposed for the discussion of a correlation between input data and execution results of the application. As data processing methods, Bayesian Classifier and HMM are employed in this paper. With various conditions, it has been clarified how the quality of input data affects the result of the lifelog analysis application.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: User/Machine Systems—*Human information processing*

General Terms

Measurement

Keywords

lifelog analysis, data quality, quantitative index, Bayesian Classifier, Hidden Markov Model

ICUIMC'12 February 20-22, 2012 Kuala Lumpur, Malaysia

1. INTRODUCTION

In recent years, with a rapid improvement of devices having various sensors that becomes smaller and increased performance, including network cameras and smart phones which can record video image, motion-sensor data, and GPS data, it has become technically easier to collect real-world data. In terms of the accumulation of data, with increasing capacity of storage devices as well as the storage services on the Internet, we can store, share, and search large amount of data for free or very low cost.

A large amount of the recorded data, such as the record of human activities and health conditions, would be useful information for users when the data is applied to various kinds of data analysis processes. One of a representative system that performs such an analysis process is called "lifelog analysis application". Because of the development of data collection technologies and storage in recent years, a variety of lifelog analysis applications have been developed. For example, as shown in the related research works described later in this paper, there have been various applications that evaluate lifelog data in terms of the amount, enhance viewing of the data, and so on. However, in those applications, the quality of lifelog data has not been discussed in detail.

There are a lot of types of lifelog data, and its quality also varies widely depending on the conditions. Although it has become easier to accumulate lifelog data, we have no idea what extent of quality and what volumes of data should be collected for these applications. Thus, it is important to clarify this problem for the use of lifelog.

The purpose of this study is to evaluate the influence caused by the quality of input data to lifelog analysis applications with quantitative indicators. In other words, as shown in Figure 1, we suppose a typical lifelog analysis application of which the input data is video image and acceleration data of objects collected from a sensor space, and it provides users the result of analysis by applying some theoretical data processing to the input data. The data quality evaluation experiments have been executed to know how the difference of input data quality results in the output of the application. For example, while only a small number of frame drops or little noise will hardly influence the result of the lifelog analysis application, if there are a lot of dropped frames or noise, the application would not be able to output the correct results. Therefore, it is important to clarify the

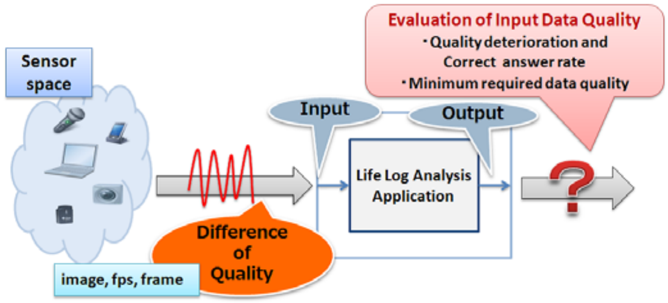

Figure 1: Input data quality evaluation on lifelog analysis application

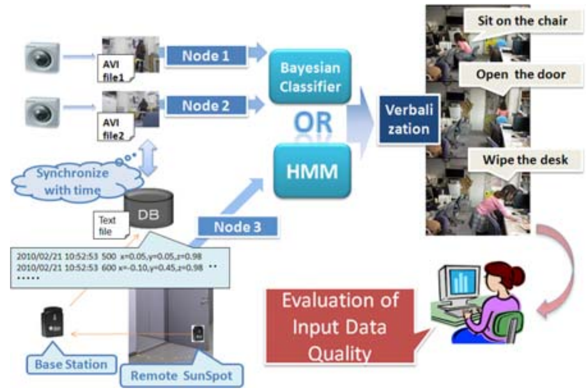

Figure 2: Execution environment of verbalization application

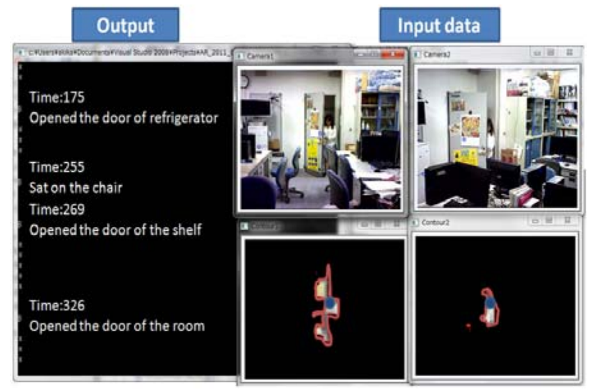

Figure 3: Behavior of verbalization application

quantitative indicators to show which level of data quality is required for the application to work correctly.

In this paper, two kinds of results from experiments that evaluate input data quality are shown. One of them is the relationship between the deterioration of input data quality and the correct answer rate of the application, and the other is the minimum quality of input data needed for the application to output the correct result. As the data processing method, two methods have been implemented and applied in these experiments, that is Bayesian Classifier and Hidden Markov Model (HMM) to compare their differences.

The remainder of this paper is organized as follows. The lifelog analysis application for evaluation experiments is described in Section 2, and an evaluation method is proposed in Section 3. The quality of input data for the application is described in Section 4, and the results of evaluation experiments are shown and discussed in Section 5. The related works are introduced and the position of this study is clarified in Section 6, and this paper is concluded in Section 7.

2. LIFELOG ANALYSIS APPLICATION THAT VERBALIZES HUMAN ACTIVITY

2.1 Behavior of verbalization application

In the evaluation experiments of this study, a lifelog analysis application that verbalizes human activity[1] has been used. This application is called 'verbalization application' in the rest of this paper. We have improved the application introduced in [1], and it works as follows:

This verbalization application describes human activity performed in the recorded data in natural language. As shown in the Figure 2, the input data of the verbalization application is video image data taken by two cameras set up in two different angles in the room, as well as acceleration data taken by the SunSPOT[2] attached to moving objects like door or chair in the room. When the verbalization application receives the input data, video image data from two cameras and the acceleration data, they are synchronized based on time and called as Node1 (video data from camera1), Node2 (video data from camera2), and Node3 (acceleration data), respectively. These nodes are processed with either Bayesian Classifier or HMM, and the verbalization application outputs verbalized expression and provides the information for users only when the predefined conditions are met.

The output of the verbalization application is the verbalized expression of human activity. For example, as shown in the Figure 3, if video images and acceleration data of human activity 'open the door of the room' are put into the application, a verbalized expression 'Opened the door of the room' is kept to be outputted as a result of the analysis while the human is opening the door in the recorded video.

We have chosen Bayesian Classifier and HMM to process the input data because these two methods are typical probabilistic models and their behavior is different with each other. Therefore, we can compare the results of Bayesian Classifier and HMM so as to prove the results are reliable if the both results are similar.

2.2 Development environment

In our experiments, verbalization application has been developed with Microsoft Visual C++ 2008 Express Edition, and image processing of video frame has been executed with OpenCV library[3]. Network camera for taking video is Panasonic BB-HCM715(130 megapixels, wired or wireless LAN)[4], acceleration data has been collected with SunSPOT[2].

3. PROPOSAL OF EVALUATION METHOD AND OVERVIEW OF ITS FRAMEWORK

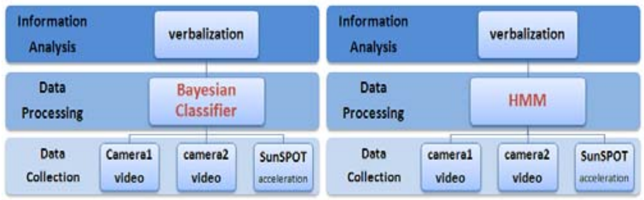

Figure 4: Framework for data quality evaluation (left:Bayesian Classifier model, right:HMM)

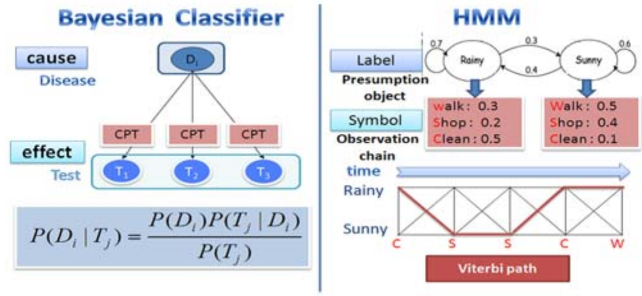

Figure 5: Examples of data processing layer in the data quality evaluation framework (left:Bayesian Classifier model, right:HMM)

3.1 Division of the process of verbalization application

The process of verbalization application is divided into three layers, which consists of 'data collection layer', 'data processing layer' and 'infomation analysis layer' as shown in Figure 4. We call this 'data quality evaluation framework' and have experimented data quality evaluation with this framework.

Data collection layer is the input part of sensor data used for analysis, data processing layer is the part of theoretical analysis process with data per nodes given from data collection layer, and information analysis layer is the output part of the analyzed result given from data processing layer.

In this paper, two kinds of different methods, that is Bayesian Classifier model and HMM, have been modeled in the data processing layer. We have compared and evaluated the influence of input data quality deterioration to the behavior of the application when verbalization has been performed through two different logical processes.

3.2 Data collection layer

The process of each layer of the data quality evaluation framework on our experiment will be described in section3.2, section3.3, section3.4. To begin with, data collection layer is the input part of video image data recorded by the network cameras and acceleration data collected by SunSPOT.

In the video data processing, the contours of the difference between the current image frame and the previous image frame for each frame are extracted (the red line shown in Figure 3), and a center of gravity of the portion surrounded by the contours (the blue point shown in Figure 3) is sought. The contours are supposed to be moving object that is a human, and the center of gravity is center of the human. The number of overlapping of the center of the human and an object (door, chair, and so on) is counted, and if the count exceeds a pre-defined threshold, Bit1 and Bit2 are marked which are bits for two network cameras. For the acceleration data processing, Bit3 is marked when the variation of x, y, z-axis acceleration collected each time exceeds a pre-defined threshold.

These three bits information is given to data processing layer.

3.3 Data processing layer

For data processing layer, two different logical processing methods of Bayesian Classifier model and HMM have been applied (Figure 5), and compared the effects of input data quality deterioration for the verbalization application when the input data was processed through two different processing methods.

First, Bayesian Classifier is a probabilistic inference model which describes causal relationship with conditional probability table (CPT) and predicts the cause from results. In the example shown in the left of Figure 5, CPT of a chance of testing positive (T_j) for the disease (D_i) is known and given. When we have the testing result (T_j) and want to know which disease (D_i) caused the result, we can predict the most plausible cause, that is, disease by getting D_i which maximizes the formula. At modeling of verbalization application in this research, 'human activity' means the cause and 'reaction of network cameras and SunSPOT' means the result. Details will be discussed later.

Next, HMM is a probabilistic model to estimate the system's internal state transitions based on Markov process from probability distribution of symbols according to each state. In the example of the right of Figure 5, only the transition probability of weather (rainy, sunny) is known and the occurrence probability of raining or sunny is unknown (that is Hidden). In addition, the probability distribution of three kinds of activities (walk, shop, clean) of a human during each weather is also given. In this case, we can predict the transition path of change of the weather from observed activity of people. This optimal path is called 'Viterbi path' and we have used the pattern of viterbi path to judge in the verbalization application. Details will be discussed later.

In the case of Bayesian Classifier model, it processes data per each frame, while in HMM, on the other hand, time series is important.

3.4 Modeling of verbalization application in the information analysis layer

3.4.1 *Information analysis based on Bayesian Classifier*

First, the modeling of verbalization application in the information analysis layer with Bayesian Classifier is described, as shown in Figure 6.

Three bits information of two network cameras and SunSPOT given from data collection layer are R_1, R_2, R_3, respectively. In the modeling of verbalization application, R_1, R_2, R_3 are the result nodes and they are reaction of two network cameras and SunSPOT, respectively. A_i is the cause node and it means activities of people and more than one activities can be defined.

For example, when the following three actions are dealt with;

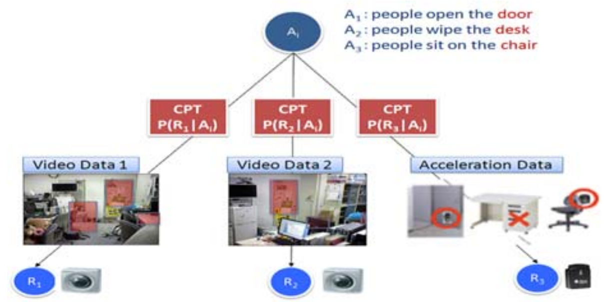

Figure 6: Modeling of verbalization application with Bayesian Classifier

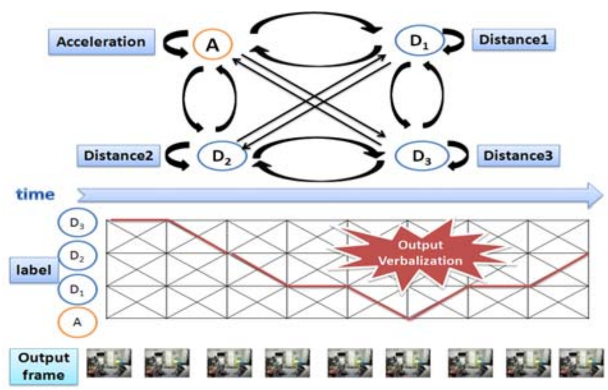

Figure 7: Modeling of verbalization application with HMM

- A_1 : people open the door
- A_2 : people wipe the desk
- A_3 : people sit on the chair

the most plausible cause is determined by finding A_i to maximize the formula below.

$$P(A_i|R_1,R_2,R_3) = \frac{P(A_i)P(R_1,R_2,R_3|A_i)}{P(R_1,R_2,R_3)}$$

3.4.2 Information analysis based on HMM

Next, the modeling of verbalization application and information analysis layer with HMM is described as shown in Figure 7. The following two kinds, total four states are defined and each states is labeled, and only transition probability between labels are known.

- A : Acceleration terminal SunSPOT reacts
- D_i : Distance between the center of a human and areas of pre-defined objects ($D_1 < D_2 < D_3$)

The output symbol for each state is the image frame of video data. When a pattern of the optimal viterbi path is obtained from each output image frame in chronological order, as the red line shown in Figure 7, if the distance between the human and the defined object becomes closer ($D_3 => D_2 => D_1$), stays there for a definite period of time($D_1 => D_1$), and Acceleration terminal SunSPOT reacts($D_1 => A$), then the system notices the human has performed some activities and outputs its verbalization. Note that patterns of the optimal viterbi path have been obtained by training more than one video data.

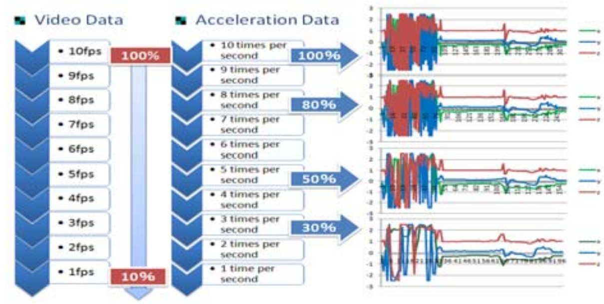

Figure 8: Obtained number of frames (the number of data per time) of video data and acceleration data

4. INPUT DATA QUALITY

The system of the verbalization application have been described, and the data quality evaluation experiments will be described in this section. In the case of input data quality evaluation experiments in this paper, the following two kinds of data qualities are evaluated:

- 'Number of frames (the number of data per time)' of video data and acceleration data
  (described in section 4.1)
- 'Image quality' of video data
  (described in section 4.2)

4.1 Quality of obtained number of frames

As the quality of 'obtained number of frames (the number of data per time)', the quality of video data and acceleration data, which is two kinds of input data of verbalization application, is divided into 10 steps respectively, as shown in Figure 8.

For the video data, the quality is divided into 10 steps from 10fps(frame per second), the maximum quality(100%), to 1fps, the minimum quality(10%). For the acceleration data which is collected more than 100 times per second by SunSPOT, the quality is also divided into 10 steps from 10 times per second, the maximum quality (100%), as the maximum quality of video data is 10fps, to 1 time per second, the minimum quality (10%). As shown in the right graph of Figure 8, the lower the quality of obtained number of frames of acceleration becomes, the rougher the graph becomes.

We have evaluated the correlation between input data quality difference and the correct answer rate of the verbalization application. The input to verbalization application of evaluation experiments is following three kinds.

- experimentA-1 : Only changing the quality of video data (acceleration data quality is fixed 100%)
- experimentA-2 : Only changing the quality of acceleration data (video data quality is fixed 100%)
- experimentA-3 : Changing the quality of both video data and acceleration data together

4.2 Image quality

Next, the 'image quality' focused on each frame of video data is described. In the evaluation experiments in this paper, the following four kinds of filtering for each image frame of video data artificially are executed and the quality is divided into 10 steps, as shown in Figure 9.

- experimentB-1 : filtering process intend for blurry image (smoothing).

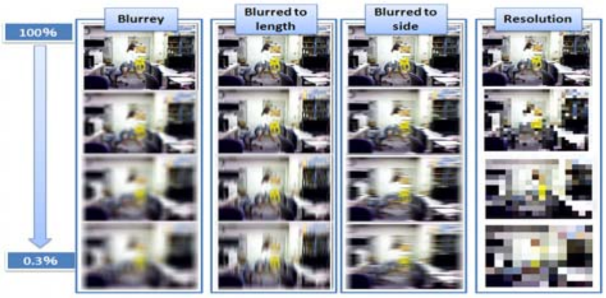

Figure 9: Image quality of each frame of video data

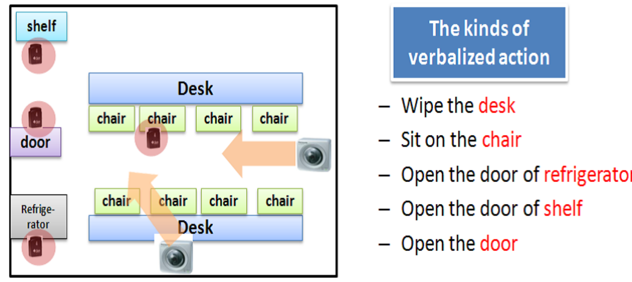

Figure 10: Experimental environment and actions to be verbalized

- experimentB-2 : filtering process intend for images blurred into length.
- experimentB-3 : filtering process intend for images blurred into side.
- experimentB-4 : filtering process intend for images the resolution deteriorated.

The wider the space (a square) of filter for each frame is, the lower the quality becomes. The image quality has been calculated with the following formula.

$$Q = \frac{1}{p^2} \times 100 \quad (\%)$$

Q:quality of imageĄCp:range of the filtered spaceĄithe length of one side of the squareĄj

5. EVALUATION

5.1 Experimental environment

The experimental environment is shown in Figure 10. In the sensor space where two network cameras and four SunSPOTs[1] exist as shown in the left of Figure 10, five kinds of human activities are verbalized as shown in the right of Figure 10. In the evaluation experiments, a correlation between the correct answer rate of verbalization application and the quality of input data (video data and acceleration data) has been evaluated.

In the experiments, both "experimental data" and "real data" are used. In the case of experimental data, the result of the evaluation is the average of more than one video data in which the same specific activities are intentionally performed. On the other hand, real data has been recorded at the same space for two days, in which people have performed their own activities naturally.

5.2 Evaluation method

How to calculate the correct answer rate is as follows: When the quality of both video data and acceleration data is 100%, the highest quality, verbalization is kept to be outputted during the occurrence of human behavior. The correct answer rate of this condition is regarded as 100%, and V_{100} is defined as the number of output in this case. Possible wrong verbalizations caused by the quality deterioration of input data are following three patterns.

- Though verbalization is outputted while human behavior is occuring, extra wrong output is also issued in addition to V_{100}. (V_{extra} is defined as the number of output in this case.)
- Wrong verbalization that is different from the human behavior is outputted. (V_{error} is defined as the number of output in this case.)
- Verbalization is not outputted even though human behavior is occurring.

[1]SunSPOT is attached to moving objects including chair, refrigerator, shelf, and door

The number of verbalization output is V_q when the input data quality is $q(0 < q <= 100)$. we have calculated the correct answer rate C with the following formula.

$$C = \frac{V_q - V_{extra} - V_{error}}{V_{100} + V_{extra} + V_{error}} \times 100 \quad (\%)$$

In addition, the minimum data quality required for the verbalization application to output correct answer is called the minimum required quality. This means that verbalization is outputted for all performed activities at least once. This is a low limit quality, which is before the condition that 'Verbalization is not outputted even though human behavior is occurring'.

5.3 Result of evaluation experiment

Figure 11-14 show the correlation between the change of the quality (the number of obtained frames/data, image quality) and the correct answer rate, and minimum required quality. The horizontal axis is the rate of the number of obtained frames/data per second and vertical axis is the correct answer rate. In the graph of minimum required data quality, the horizontal axis is the data quality that is changed in each case and vertical axis is the minimum required quality.

The three graphs (Video data, Acceleration data, Video data & Acceleration data) shown in Figure 11 and 13 are the results of experiment A-1 – A-3 described in section 4.1, respectively. They are the correlation between the rate of the number of obtained frames/data and correct answer rate. The Four graphs (Blurry, Blurred to length, Blurred to side, Resolution) shown in Figure 12 and 14 are the results of experiment B-1 – B-4 described in section 4.2, respectively. They are the correlation between image quality and correct answer rate.

The results of experiments with experimental data are shown in Figure 11 and Figure 12, and the results of experiments with real data are shown in Figure 13 and Figure 14. The difference between experimental data and real data is described in section 5.1.

The results of both Bayesian Classifier and HMM are shown in all the Figure 11-14. In the graph of minimum required quality, when the bar graph is higher, it means

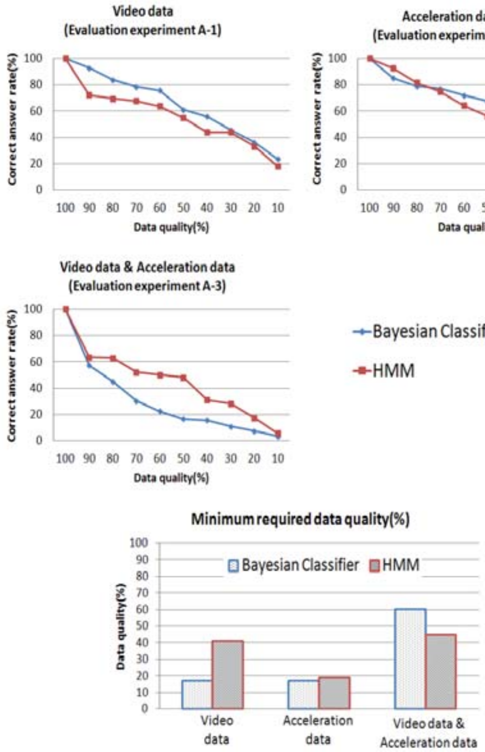

Figure 11: Correlation between the rate of the number of obtained frames/data per second and the correct answer rate (experimental data)

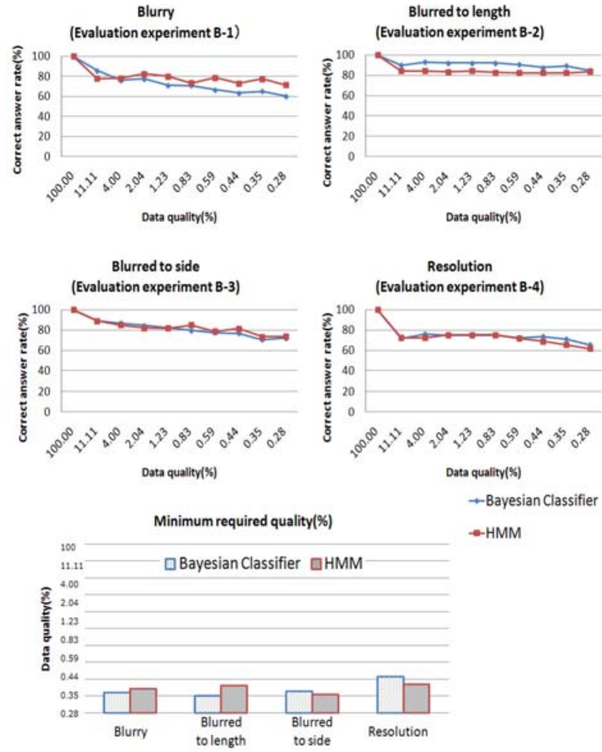

Figure 12: Correlation between the change of the image quality and the correct answer rate (experimental data)

that higher quality input data is required for verbalization application to output correct answer. On the other hand, when the bar graph is lower, it means that verbalization application can output correct answer even if lower quality data is inputted.

5.4 Discussions

5.4.1 The rate of the number of obtained frames/data and the correct answer rate

As shown in Figure 11 and Figure 13, when the input data quality becomes lower, the correct answer rate decreases. This is an appropriate result. As shown in Figure 11, the deterioration in either only video data or only acceleration data causes correct answer rate to drop to about 20%, while the deterioration in both video data and acceleration data decreases the correct answer rate to nearly 0%. In the case of real data shown in Figure 13 also, though the numerical value differs a little, the feature of the graph is almost the same.

According to this result, the quality deterioration of multiple input data has a lot of influence on the lifelog analysis application. However, if the quality deterioration occurs only in a part of input data, the percentage of correct answers can be kept up to 60% even if the input data quality drops around half.

In terms of comparison between Bayesian Classifier and HMM in the graph of video data, the line of HMM is drawn in lower place at all input data qualities and the bar graph of HMM is higher in the graph of minimum required quality.

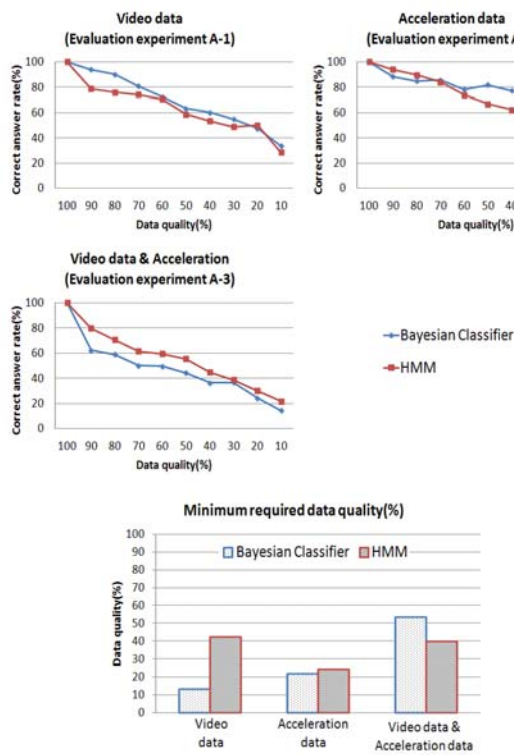

Figure 13: Correlation between the rate of the number of obtained frames/data per second and the correct answer rate (real data)

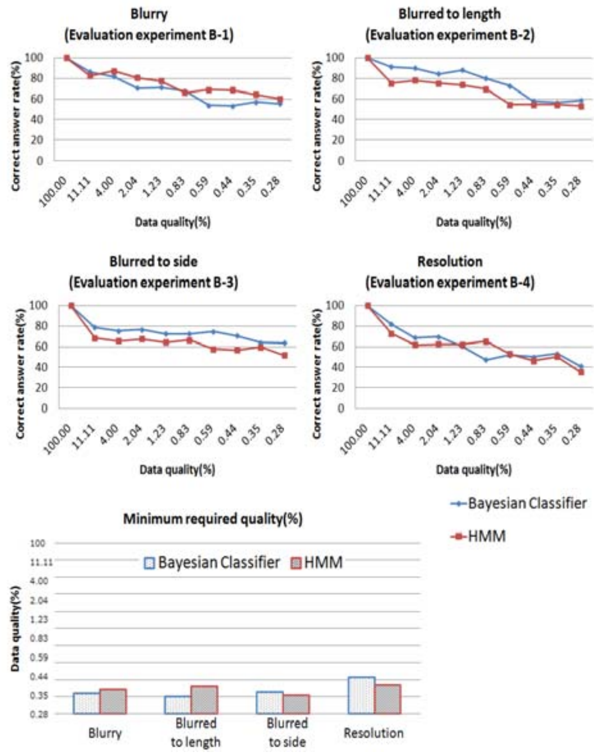

Figure 14: Correlation between the change of the image quality and the correct answer rate (real data)

Thus, to conclude, the process of HMM depends on image frames of video data and is more sensitive to the quality change of video data than that of acceleration data.

In the graph of acceleration data, judging from the height of the bar graph, verbalization application can output correct answer even if the quality of both Bayesian Classifier and HMM drop to about 20%. In the case of real data, about 40 − 50 % of correct answer is kept all the time. This means that while the features of both cases are almost the same, the acceleration data is less important for the real data. However we should increase the quantity of evaluation data and see whether the graph will be changed or not.

On the other hand, in the graph of video data & acceleration data, the deterioration of input data quality decrease correct answer rate earlier in both Bayesian Classifier and HMM. If the quality becomes lower in both cases, the case of Bayesian Classifier becomes lower than that of HMM. This result is also observed in both cases of using experimental data and real data.

5.4.2 The image quality and the correct answer rate

The graph of all kinds of image quality that are Blurry and Blurred to length and Blurred to side and Resolution, the correct answer rate only decreases to 60 - 80% as shown in Figure 12, and decreases to about 60% as shown in Figure 14. Therefore, to conclude, the deterioration of the rate of the number of obtained frames/data have much influence on the correct answer rate of the verbalization application than deterioration of image quality. As shown above, since almost the same feature graph is acquired in both cases of experimental data and real data, the result should have credibility.

5.4.3 Conclusion of discussion

Because the results of experimental data and real data are almost the same, these results are reliable even though it is necessary to increase the quantity of evaluation data in the future.

As the graphs of correlation between quality change of input data and the correct answer rate are shown with quantitative measure, we can confirm the input data quality required for the verbalization application clearly. As shown in the bar graph of minimum required data quality, it is not necessary to use the highest 100% quality input data to output correct answer. In some data like acceleration data, even the very low quality, about 20%, can output correct answer.

6. RELATED WORKS

A lot of studies to recognize human activity with analysis of lifelog data collected by sensor terminals have been performed and some of them are introduced[5][6]. In the reference [7][8], a large amount of acceleration data collected with sensor terminals attached to people was studied and used to recognize the human activity. As a result, increasing the volume of collected data would improve the recognition rate.

In the reference [9][10], activity information sharing system with which people can see others' activity and compare with himself has been provided by collecting acceleration and video data of human activities.

The collected lifelog has been intensively used for location-based systems such as [11], in which a method for predicting the future location of the human based on the lifelog was discussed. In [12], a method was discussed to build context modeling for recognition of a human behavior. They proposed a hierarchical spatio-temporal context modeling.

As shown above, various research works to make use of lifelog data acquired by the sensor nodes have been performed. However, our study is different from them because we have focused on the quality of input data and discussed the correlation between the quality and the results of the verbalization application in detail.

7. CONCLUSIONS

In this paper, we have proposed 'data quality evaluation framework' as a method to evaluate the influence of the verbalization application caused by the deterioration of input data quality. Data quality evaluation experiments are executed with two different kinds of data processing layer on the verbalization application. As a result, the correlation between data quality deterioration and the correct answer rate and the minimum required quality for the verbalization application has been shown quantitatively. About the data processing method, we have compared the results of Bayesian Classifier and HMM.

As the future work, we are going to accumulate more real data on a real environment of human life[13] and evaluate influences of deterioration in the input data quality to the lifelog analysis application. In addition, we would like to deal with sound data as input.

8. ACKNOWLEDGMENTS

The authors are deeply grateful to Prof. Ichiro Kobayashi and Ms. Erika Ochiai as they gave us a lot of useful advices

for our study.

9. REFERENCES

[1] E. Ochiai and I. Kobayashi. Efforts to verbalize human activity in given space based on the predict model. JSAI2010, 2G1-OS3-2, June 2010.

[2] Wireless sensor network devices: http://www.jp.sun.com/products/software/sunspot, SunSPOT memo wiki, http://www.klab.ai.kyutech.ac.jp/yatti/sunspot/index.php?Sun\%20SPOT\% 20memo\% 20wik

[3] Open Computer Vision Library: http://sourceforge.net/projects/opencvlibrary/

[4] PLANEX CS-WMV04NĄF http://www.planex.co.jp/product/camera/cs-wmv04n

[5] P. C. Ribeiro and J. Santos-Victor. Human Activity Recognition from Video: modeling, feature selection and classification architecture. HAREM2005, ACTIVITY RECOGNITION (I), September 2005.

[6] T. T. Truyen, H. H. Bui, and S. Venkatesh. Human Activity Learning and Segmentation using Partially Hidden Discriminative Models. HAREM2005, ACTIVITY RECOGNITION (II), September 2005.

[7] N. Kawaguchi, N. Ogawa, Y. Iwasaki, K. Kaji, T. Terada, K. Murao, S. Inoue, K. Kawahara, Y. Sumi, and N. Nishio. HASC Challenge2010:Construction of Wearable Accelerometer Sensor Corpus for Activity Recognization. DICOMO2011ĄC1E-1ĄCJuly 2011.

[8] N. Ogawa, K. Kaji, and N. Kawaguchi. Effects of Number of Subjects on Activity Recognization - HASC2010corpus -. DICOMO2011, 1E-2, July 2011.

[9] Y. Hattori, S. Inoue, and G. Hirakawa. Activity Recognization and Visualization for Activity Information Sharing System. DICOMO2011, 1E-4, July 2011.

[10] M. Nakamura, Y. Hattori, S. Inoue, and G. Hirakawa. Evaluatoin of Large Scale Activity Information Sharing System using Acceleration Data and Video Data. DICOMO2011, 2E-1, July 2011.

[11] M. Nishino, Y. Nakamura, T. Yagi, S. Muto, and M. Abe. A location predictor based on dependencies between multiple lifelog data. ACM LBSN2010, November 2010.

[12] J. Sun, X Wu, S. Yan, L. F. Cheong, T. S. Chua, and J. Li. Hierarchical spatio-temporal context modeling for action recognition. IEEE CVPR2009, June 2009.

[13] OchaHouse PukiWiki, Ubiquitous Computing Experiment House of Ochanomizu University: http://ochahouse.com/