

Assisting Memories in Elderly Care by Recognizing and Documenting Daily Activities of Images

Raiki Saito*, Sinan Chen[†], Sachio Saiki[‡] and Masahide Nakamura[†]

*Graduate School of Engineering, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe, 657-8501, Japan

Email: raiki@es4.eedept.kobe-u.ac.jp

[†]Center of Mathematical and Data Sciences, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe, 657-8501, Japan

Email: chensinan@gold.kobe-u.ac.jp, masa-n@cmds.kobe-u.ac.jp

[‡]School of Data and Innovation, Kochi University of Technology, 185 Miyanokuchi, Tosayamada, Kami, Japan

Email: saiki.sachio@kochi-tech.ac.jp

Abstract—Memory impairment is a central manifestation of dementia that poses a formidable challenge for current medical interventions. While existing medications struggle to offer a cure, the potential of memory aid to compensate for these impairments and provide crucial support in daily life is evident. However, the present landscape of memory aid places a substantial burden on both caregivers and individuals grappling with dementia. This necessitates the exploration of novel memory aid approaches that alleviate this burden. The purpose of this paper is to propose an innovative memory aid methodology aimed at reducing the strain on caregivers. The primary goal is to investigate the feasibility of summarizing daily life videos into text through the application of Image Recognition AI. The methodology encompasses four key steps: recording daily activities using a Wearable Camera, extracting image data, generating *Tags* and *Caption* via a Image Recognition AI, and summarizing and formatting the resultant *Tags* and *Caption*. In this proposed method, daily life is captured using a Wearable Camera and input into the Image Recognition AI, specifically the recognize-anything-model. This model identifies *Tags* and *Caption* that characterize the objects within the still images. Subsequently, these outputs are analyzed chronologically and distilled using Markov chains, a technique for summarizing documentation. The focus of the case study lies in summarizing texts generated by the Image Recognition AI, with a specific exploration of their potential utility as memory aid for elderly individuals grappling with dementia.

Index Terms—image recognition AI; dementia; memory disorders; Web service

I. INTRODUCTION

Japan has a super-aging society, and the increase in the number of elderly people with dementia is an urgent issue that needs to be addressed as a society. According to the Ministry of Health, Labor and Welfare, the population of the elderly aged 65 and over is projected to reach 35 million, or 35.3 % of the total population, by 2040. It is estimated that more than 7 million of them are elderly people with dementia. This is a calculation that one in five elderly people is a dementia patient. Against this background, dementia prevention methods for the elderly and support for the elderly with dementia are needed.

Memory impairment is one of the core symptoms of dementia and is a symptom that is always seen in elderly people with dementia. Memory impairment is the inability to acquire necessary information and use it in the real world when necessary. Currently, memory aid such as caregivers providing

necessary information to elderly people with dementia and using tools such as notebooks are widely used for prevention and rehabilitation. However, as dementia progresses, motivation and understanding decline, and it is possible that instructions cannot be followed or the meaning of instructions cannot be understood.

With the above-mentioned memory aid, there are problems such as increasing the burden on caregivers if they do not follow the instructions, or neglecting the use of tools. A new memory aid method is needed to reduce the burden on caregivers and enable elderly people with dementia to continue. To reduce the burden on caregivers and enable continuous support, it is desirable to have a mechanism that can aid memory through natural behavior in daily life.

In this study, we aim that memory aid can be realized by recording daily life behavior as a document. The key idea of this study is to extract still images from daily life videos taken using a Wearable Camera, generate *Tags* and *Caption* that characterize the images using Image Recognition AI, and summarize them using document summarization technology. Finally, we verify whether it is possible to create a document from a daily life videos and discuss it as a memory aid for elderly people with dementia.

II. PRELIMINARIES

A. Needs for Digitization in Elderly Memory

Japan is now a super-aging society. According to the Ministry of Health, Labor and Welfare, the number of elderly people aged 65 and over in Japan is projected to reach 35 million, or 35.3 % of the total population, by 2040. It is estimated that more than 7 million of them are elderly people with dementia. This is a calculation that one in five elderly people is a dementia patient [1].

Against this background, dementia prevention methods for the elderly and support for the elderly with dementia are needed. Memory impairment is one of the core symptoms of dementia and is a symptom that is always seen in elderly people with dementia. Memory impairment is the inability to acquire necessary information and use it in the real world when necessary.

Currently, memory aid such as caregivers providing necessary information to elderly people with dementia and using tools such as notebooks are widely used for prevention and rehabilitation. However, as dementia progresses, motivation and understanding decline, and it is possible that instructions cannot be followed or the meaning of instructions cannot be understood.

Therefore, with the above-mentioned memory aid, there are problems such as increasing the burden on caregivers if they do not follow the instructions, or neglecting the use of tools. A new memory aid method is needed to reduce the burden on caregivers and enable elderly people with dementia to continue. Memory aid is the act of reminding or providing information to a person with dementia when they need it. Memory aid can compensate for memory impairment and support daily life. Currently, caregivers provide care for people with dementia, and memory aid such as notebooks are used for rehabilitation.

However, the former increases the burden on caregivers, and the latter is difficult for people with dementia to continue. In addition, excessive or appropriate care and rehabilitation may be a factor in behavioral and psychological symptoms (BPSD) when a person cannot adapt well to life due to core symptoms, and the person's personality, environment, and physical condition are added. In this way, a new memory aid method that does not interfere with the daily life of caregivers and enables elderly people with dementia to continue is needed.

B. Previous Studies

We have developed a virtual agent user interface (VAUI) as a previous study using agent technology. In implementing the agent, we used the MMDAgent [2] a toolkit for voice dialogue systems. In addition, we developed HNS-VAUI [3], which applies VAUI to the user interface of the home network system (HNS) [4], and have been conducting experiments and evaluations. The user information used is the form of a person's life history used in the center method, and questions are asked that incorporate the upbringing and past life of people with dementia to direct the conversation [5]. In addition, we use LOD to dynamically discover words related to the topic from the Web and develop the content of the conversation.

C. Technical Challenges

The method of generating topics using life history and Linked Open Data (LOD) made it possible to deepen the topic based on one topic. However, in the previous study, it was necessary to create a scenario based on the items of the life history in advance, and there was a limit to the generation pattern of the topic to be used as a trigger.

In addition, for the contents of the life history, such as "How old are you?", "Where are you from?", and "What is your favorite color?", which are expected to have a unique answer, it is possible to create a scenario that develops the conversation for that answer in advance. However, there are also problems that it is difficult to decide the direction of the conversation in advance for things that do not have a single answer, such as

"How do you spend your day?", "What have you been familiar with for a long time?", "What are you good at?", and "What are your hobbies?".

On the other hand, in dementia counseling, where it is necessary to continue daily conversation, it is desirable to have many triggers for the topic. Furthermore, when generating interactions, it is necessary to dynamically search LOD based on the obtained words, but it is difficult to obtain an appropriate data set and apply it appropriately to the development of the conversation, so it is not possible to provide topics about nostalgic things, life, and hobbies that do not appear explicitly in the life history.

Moreover, the system in the previous study is a passive conversation that listens to the answer to the question and, if possible, develops it. The material of the conversation is information about the individual, and it depends on the information that only the individual knows. There is also a problem that some people may have difficulty answering questions.

D. Image Recognition AI

Image Recognition AI is a technology that uses AI to determine and discriminate features such as people's faces and characters in images. By loading a huge number of image data into AI and repeating learning, various information in the photo can be extracted. The Image Recognition AI used in this paper is the Recognize Anything Model (RAM) [6], which combines object detection technology that extracts tags and characteristics of objects in an image and image caption technology that explains the situation of objects in an image.

E. Documentation and summarization techniques

The development of technology in the field of natural language processing has been progressing for a long time in terms of "keyword extraction", "sentiment analysis", and "refinement of content" for Japanese text sentences. Regarding "keyword extraction", there is a morphological analysis technology that divides Japanese into parts of speech (e.g., nouns, adjectives, verbs) by morphemes (i.e., the smallest unit of expression elements with meaning). Representative morphological analysis technologies include Kuromoji [7] [8], MECAB [9], JUMAN [10], and JANOME [11].

On the other hand, in Japanese language processing methods for "sentiment analysis", there are three approaches: a general simple emotion dictionary (Japanese evaluation polarity dictionary, word emotion polarity correspondence table) [12] base, supervised learning by deep learning [13], and existing cloud services and Application Programming Interface (API) [14]. As a result, it is possible to analyze fine emotions such as negapoji judgment [15] and happy/sad results [16].

In addition, in Japanese language processing methods for "refinement of content", there are sentence correction [17] using Word2vec and sentence summarization [18] using Markov Chain. In this paper, we consider using these technologies to summarize the tags and captions generated by the Image Recognition AI. To reduce the burden on caregivers and enable

continuous support, it is desirable to have a mechanism that can aid memory through natural behavior in daily life. As a result, it is possible to save daily life in an easy-to-understand document and aid the memory of the elderly.

III. PROPOSED METHOD

A. Goal and Key Idea

The purpose of this paper is to verify whether the technology for developing services that support and guarantee the memory of the elderly and people with dementia by digital technology is practical. To this end, we use a travel video taken with a Wearable Camera as a sample, extract still images from the video at regular intervals, and use RAM, an Image Recognition AI described in II-D, to describe the objects moving to the still images. We extract *Tags* that characterize the text and extract *Caption*. By arranging these in chronological order, we summarize the video into text by analyzing it.

B. Overall Architecture

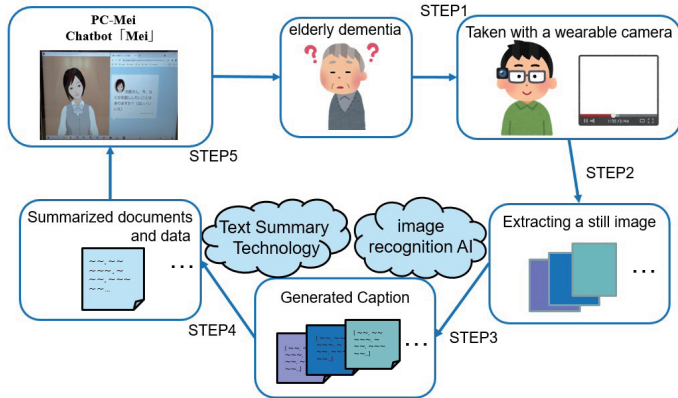


Fig. 1. Overall Architecture

The overall architecture of this research is to have the elderly and people with dementia shoot the original video of daily life behavior, which is the target of documentation, and use the method in this paper to save them as text. In addition, using the document, we use VAUI described in II-B to confirm with the person who took the video, and the final goal is to use it as a memory aid for the elderly and people with dementia. More specifically, the proposed method consists of the following five steps. In this paper, we implemented steps 1 to 4.

STEP 1: Daily life recording

STEP 2: Extraction of image data

STEP 3: Generation of *Tags* and *Caption* by generation AI

STEP 4: Summary and formatting of generated *Tags* and *Caption*

STEP 5: Confirmation of generated documents by the person

C. STEP 1 : Daily Life Recording

To aid the memory of the elderly and people with dementia, it is necessary to have a video that is the basis for saving data. Therefore, daily life recording tailored to each person is necessary. The method of recording daily life videos assumes recording with a Wearable Camera, and it is assumed that it is acquired by recording in a non-invasive manner in daily life.

D. STEP 2 : Extraction of Image Data

In STEP 2, still images are extracted from the daily life recording obtained in STEP 1 at regular intervals. The regular interval here is the interval at which still images can be extracted from the still images while suppressing duplication of the data of the still images when extracting information from the still images in STEP3.

E. STEP 3 : Generation of Tags and Caption by Image Recognition AI

In STEP 3, still images extracted in STEP 2 are characterized by Image Recognition AI, and *Tags* and *Caption* are extracted. This converts the image data of the still image into text data and makes it easy to save.

F. STEP 4 : Summary of Generated Tags and Caption

In STEP 4, *Tags* and *Caption* extracted in STEP3 are summarized using APIs. The method of summarization is to save it in a form that can be treated as a diary.

G. STEP 5 : Confirmation of Generated Documents by the Person

In STEP 5, in order to ensure the reliability of the document in STEP 4, the generated document is confirmed by the person. The method of confirming the document to the person is to use VAUI as described in the previous study in II-B and to do it by talking with the person. Therefore, the generated document is formatted and saved as a new document with high reliability. As a result, these technologies can be used as memory aid for the elderly and people.

IV. CASE STUDY

A. Overview

Based on the proposed method, the technologies used in this paper are as follows.

- Development language: python
- Image Recognition AI: Recognize Anything Model
- Document summarization technology (algorithm): Markov Chain
- Daily life recording: Daily life recording of “going out, shopping, etc.”

In the case study in this paper, as a feasibility of the documentation technology for daily life videos, we first used recordings of daily life “going out, shopping, etc.” using a Wearable Camera as the target daily life videos.

B. STEP 1 : Daily Life Recording

In the case study in this paper, as a feasibility of the documentation technology for daily life videos, we first used recordings of daily life “going out, shopping, etc.” using a Wearable Camera as the target daily life videos. The length of the videos used this time is a 5-minute videos.

C. STEP 2 : Extraction of Image Data

Still images were extracted from the daily life videos prepared in STEP 1. In this paper, still images were extracted at 10-second intervals, and a total of 30 still images were extracted.

D. STEP 3 : Generation of Tags and Caption by Generation AI

Tags and *Caption* that characterize the images were generated by RAM, an image generation AI, for the still images extracted in STEP 2. An example is shown in Figure 2 for a total of 30 still images.

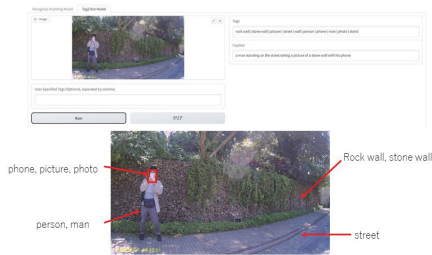


Fig. 2. Example of *Tags* and *Caption* generation by Image Recognition AI.

E. STEP 4 : Summary of Generated Tags and Caption

Tags and *Caption* for 30 image data generated in STEP 3 were summarized. Table I show tables summarizing the tags in the form of a daily life diary.

Below is a document 1 summarizing the *Caption* by Markov Chain.

A woman in a robe wearing a face mask walking on the balcony of her home. A woman in a white coat walks down a long hallway, the sidewalk adjacent to the street where the door is located. Standing in the park with his finger in his mouth. A man walking on the balcony of his home with a woman adjacent to the street with a door on the other side of the white wall from the balcony. A man standing on the sidewalk in a park with a stick and a skateboard. A woman in a mask standing on the street talking on a cell phone. A woman in a white coat is walking on the balcony of her house. A man in a robe standing in front of a door on the other side of a white wall from a balcony. Look up at the sky through

a window in a stone wall with trees on both sides of the road and a pathway. A man with a closed elevator door. Stone wall. Stone walls lit up at night and jungle trees and vegetation lush with paths and boardwalks.

Below is a document 2 summarizing the *Caption* by Markov Chain.

Rocks along the road. Standing in his hand. Stone door. A long walkway leading to a large open room. With a courtyard. A long hallway with a wall, a cell phone, and a number of people standing on a wooden porch leading to another room on both sides. Illuminated rocks at night. The man with the white coat in his mouth. Stone wall and garden between the elevators. A man between a walled house and a large room at night lit up and in a tie wearing a robe with the walls closed within the walls. A woman wearing a man standing in the street standing in a long hallway.

F. Discussion

It was verified that it is possible to extract still images at regular intervals from actual daily life recordings based on the proposed method, extract *Tags* and *Caption* that characterize the still images by Image Recognition AI, and summarize them. In this case study, the *Tags* and *Caption* generated by the Image Recognition AI determine most of the accuracy of this service. Therefore, in this paper, which uses only one Image Recognition AI it is difficult to measure how accurate the generated *Tags* and *Caption* are. In addition, since RAM generates English, it is necessary to translate it when creating a service for Japanese people, and it is necessary to consider a better translation method. In that case, it was found that the translation into Japanese was unnatural, so it is necessary to establish a criterion for judging whether the translation into Japanese is natural or not. Other than that, in the summary method in STEP4, the document was summarized in the form of a daily life diary or a Markov Chain, but it is necessary to consider other methods than these two methods and consider which method is the most effective as a memory aid. It can be considered as a task.

V. CONCLUSION

In this paper, we concluded that it is possible to verify whether the shooting video obtained by the Wearable Camera can be summarized into text. Specifically, *Tags* and *Caption* were generated using RAM, an Image Recognition AI, using actual daily life recordings, and a summary document for one day was created. As a future task, it is necessary to consider multiple technologies used in each architecture and select better technologies. In addition, it is necessary to verify that the document generated in STEP5, which could not be experimented in this paper, is confirmed by the person himself. In addition, it is necessary to consider the evaluation method of the created document as a future task.

TABLE I
DIARY OF A DAY IN THE LIFE BY TAG.

Date	Caption	Tags
10:28:21	a woman wearing a face mask walks down the balcony of a home	mask woman home house face mask balcony person man walk wear stand
10:28:31	looking up at a white wall from a balcony	building house balcony wall white
10:28:41	looking up at the sky through a window in a white wall	building house window wall sky roof white
10:28:51	a woman wearing a mask talking on the phone standing in an elevator	elevator mask woman door person phone man walk wear stand talk
10:29:01	a woman in a white coat walks down a long hallway looking at a man standing at the door	building room coat woman hallway door girl corridor person man hall walk stand look long white
10:29:11	the door to the elevator is closed	elevator door lift wall
10:29:21	a man wearing a suit standing in an elevator	elevator door suit lift wall person man wear stand
10:29:31	a long hallway with white walls and a door that leads to another room	building room hallway door light corridor wall walkway long white
10:29:41	a man in a robe standing in an elevator looking at a painting on the wall	elevator priest robe painting door mural wall person man stand look
10:29:51	a man wearing a shirt and tie standing in an elevator	elevator mask tie face mask shirt mannequin person man wear stand
10:30:01	a long hallway with a door on the far side of the wall	building hallway door corridor wall walkway long
10:30:11	a long walkway to a covered porch of a house	building house balcony porch walkway long
10:30:21	the corner of a room with a brick wall, a bench and a chair on the floor	building corner room floor hallway brick wall chair bench wall white
10:30:31	a couple of people standing in a large open room	building room people restaurant couple walk stand large
10:30:41	a covered walkway in a building with statues	building statue garden pavilion walkwaybuilding statue garden pavilion walkway
10:30:51	a walkway between a house and a garden	building garden house porch plant walkway
10:31:01	a long corridor with wooden floors and a wooden porch that leads to a large room	building room floor hallway house corridor porch walkway long wooden large
10:31:11	a house with a patio and motorcycles parked outside	building motorcycle house patio wall park
10:31:21	a sidewalk next to a street with a stone wall covered in vines and plants	stone wall sidewalk road street plant tree wall vine cover grow
10:31:31	a man standing in a park with his finger in his hand	hand driveway park finger person man stand red
10:31:41	a man standing on a sidewalk with a stick and a skateboard in a park	stick sidewalk cane pole park skateboard person man walk stand
10:31:51	a man holding a cigarette and a black pen in his hand	cigarette hand camera scissor pen person phone man hold black
10:32:01	a man standing on the street taking a picture of a stone wall with his phone	rock wall stone wall picture street wall person phone man photo stand
10:32:11	a man standing next to a wall made of stone	stone rock wall stone wall wall person man walk stand
10:32:21	a wall made of stone with trees on both sides of the road and a path	stone rock wall stone wall path road street tree wall
10:32:31	a stone wall lit up at night with rocks in the foreground	stone night rock wall stone wall rock light wall
10:32:41	a narrow path in the jungle with a stone wall and trees	jungle rock wall stone wall path road tree wall narrow
10:32:51	a garden with a stone wall, trees, and plants along a path	rock wall garden stone wall path road plant tree wall
10:32:59	a path or walkway through a tropical forest or jungle with trees and plants	jungle trail pathway garden path forest plant tree walkway tropical

VI. ACKNOWLEDGMENT

This research was partially supported by JSPS KAKENHI Grant Numbers JP19H01138, JP20H05706, JP20H04014, JP20K11059, JP22H03699, JP19K02973, and Young Scientists (No.23K17006).

REFERENCES

- [1] L. Ministry of Health and Welfare, "Future estimates of people with dementia," <https://www.mhlw.go.jp/content/001061139.pdf>, visited on 2023-10-21.
- [2] M. P. Team, "Mmdagent-tool kit for building voice interaction systems," <https://www.mhlw.go.jp/content/001061139.pdf>, visited on 2023-10-21.
- [3] S. Chen and M. Nakamura, "Designing an elderly virtual caregiver using dialogue agents and webtrc," in *2021 4th International Conference on Signal Processing and Information Security (ICSPIS)*. IEEE, 2021, pp. 53–56.
- [4] H. Igaki, M. Nakamura, and K.-i. Matsumoto, "Design and evaluation of the home network systems using the service oriented architecture," in *e-Business and Telecommunication Networks*. Springer, 2006, pp. 116–123.
- [5] D. C. Research and T. Center, "Care management for dementia center method related, *Center Method*," <https://www.dcnnet.gr.jp/study/centermethod/>, visited on 2023-10-21.
- [6] Y. Zhang, X. Huang, J. Ma, Z. Li, Z. Luo, Y. Xie, Y. Qin, T. Luo, Y. Li, S. Liu *et al.*, "Recognize anything: A strong image tagging model," *arXiv preprint arXiv:2306.03514*, 2023.
- [7] Atilica, "kuromoji," <https://www.atilika.com/ja/kuromoji/>, visited on 2023-10-21.
- [8] S. Chen and M. Nakamura, "Generating personalized dialogues based on conversation log summarization and sentiment analysis," in *The 23rd International Conference on Information Integration and Web Intelligence*, ser. iiWAS2021. New York, NY, USA: Association for Computing Machinery, 2022, p. 217–222. [Online]. Available: <https://doi.org/10.1145/3487664.3487695>
- [9] K. Taku, "Mecab : Yet another part-of-speech and morphological analyzer," <https://taku910.github.io/mecab/>, visited on 2023-10-21.
- [10] H. Morita, D. Kawahara, and S. Kurohashi, "Morphological analysis for unsegmented languages using recurrent neural network language model," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015, pp. 2292–2297.
- [11] T. Uchida, "Janome: Japanese morphological analysis engine," <https://mocabeta.github.io/janome/>, visited on 2023-10-21.
- [12] H. Takamura, T. Inui, and M. Okumura, "Extracting semantic orienta-

tions of words using spin model,” in *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL ' 05)*, 2005, pp. 133–140.

- [13] D. K. Devi, C. Akalya and S. Soundarya, “A survey based on human emotion identification using machine learning and deep learning,” https://www.researchgate.net/publication/328622102_A_Survey_Based_on_Human_Emotion_Identification_Using_Machine_Learning_and_Deep_Learning, visited on 2023-10-21.
- [14] T. A. . M. Azure, “Azure ai language;” <https://azure.microsoft.com/en-us/services/cognitive-services/text-analytics/>, visited on 2023-10-21.
- [15] E. Becker, R. Arora, S. Phan, J. K. Vinjumur, and F. Makedon, “Extending event-driven experiments for human activity for an assistive environment,” in *Proceedings of the 3rd International Conference on PErvasive Technologies Related to Assistive Environments*, 2010, pp. 1–8.
- [16] H. Zhoushao, “Withdrawn: Emotion recognition simulation of japanese text based on fpga and neural network,” 2020.
- [17] Y. Ogawa and K. Yamamoto, “Correcting misuse of japanese visually similar characters,” in *2017 International Conference on Asian Language Processing (IALP)*, 2017, pp. 57–60.
- [18] M. MURATA and Y. ABE, “Using machine learning for automatic estimation of emphases in japanese documents,” *IEICE Transactions on Information and Systems*, vol. E100.D, no. 10, pp. 2669–2672, 2017.