Context-aware Recommendation System using Content Based Image Retrieval with Dynamic Context Considered

Yuta Miyazawa, Yukiko Yamamoto, Takashi Kawabe
School of Information Environment
Tokyo Denki University
Inzai, Japan
{10jk241, 10jk267, kawabe}@mail.dendai.ac.jp

Abstract— **This paper conceptually proposes a context-aware recommendation system that gives optimal information for users based on 1) a content-based image retrieval (CBIR) mechanism to search the similar images aiming to extract the detailed information to the text-inexpressible images 2) the contextual information of such similar images searched from the Web, and 3) user's dynamic context or situation considering time-variant factors as well as space factors. It is expected to increase the precision or optimality of recommendation by matching and fusing the context of similar images obtained by CBIR with textual and signal information about user's situation or dynamic context.**

Keywords- context-aware; image retrieval; recommendation;

I. INTRODUCTION

As one of innovative information systems, the significance of context-aware systems/services has gathered a lot of attention and their demand has increased since anyone can access any necessary information at any time [1], [2].

Further, mobile devices also have been remarkably developed; especially, smart phones and tablets equipped with a high-definition camera as well as high-speed Internet access have become widely available, which has made it much easier to handle multimedia contents such as pictures and movies than before. This also has contributed to the improvement of usability in information search.

The necessity of information retrieval by images: conventional information retrievals by texts cannot always fulfill our expectations as suggests information irrelevant to what we want, and what is worse, we are often unable to describe images that come up in our mind correctly or every little detail with exact, appropriate words.

The necessity of similar image search: to acquire information based on images, the content of the images must be identified. There is a number of research in query by image content, also known as Content-Based Image Retrieval (CBIR) [3]-[5].

The necessity of taking account in the metadata of similar images: we, however, still have to question the accuracy of CBIR in terms of its approaches; it does not mean the similar images are suggested technically, not semantically. Thus that would be very risky to consider the similar images have the same content of the query-targeted picture. Here, we shall take an example. Suppose that a user inquires information with the following image presented: a red circle on a big white cloth. It would be recognized not only as Japanese flag but also as an apple on a big white cloth. If the user, however, implies something "fruit" to the presented image simultaneously, we can recognize that the user actually requires information of an apple.

This paper conceptually proposes a context-aware recommendation system that gives optimal information for users based on 1) CBIR engines to search the similar content image, aiming to extract the detailed or relevant information to the text-inexpressible images 2) the contextual information of the similar images extracted from the Web pages and 3) user's dynamic context or situation considering time-variant factors as well as space factors. As time-variant factors, we focus on time and seasonal information.

This paper is organized as follows: Section II describes context-aware image and text fusion; Section III explains the conceptual idea of the proposed context-aware question answering system; Section IV discusses the significance of considering dynamic context; and finally Section V concludes the paper.

II. CONTEXT-AWARE IMAGE AND TEXT FUSION

Firstly, we shall give an example of image and text fusion type context-aware recommendation. A senior citizen presents the photograph of a certain iron tower, and asks a question by the natural language text, saying "Isn't there any sightseeing place such as this tower that I can climb?"

The system retrieves relevant image information using a similar image search employing content-based information retrieval (CBIR) techniques. Once the image information is retrieved, information is extracted, and answered in consideration of user's situation or context. In this example, the rule that the range of action is somewhat limited because it is assumed that the user is a senior citizen and information of a spot near the user's present place is returned. Moreover, potential answers such as transmitter towers that cannot be climbed are excluded since it is deduced from the context that the user wants to climb the tower for sightseeing.

A. Content-based Image Retrieval

Users rarely annotate images for retrieval. Namely, textual information is not available for retrieval purposes. Luckily, a lot of research has focused on the area of content-based image retrieval (CBIR). CBIR extracts features directly from images and uses these as an index that can consequently be employed for retrieval. This is particularly useful for query-by-example search where an image is given by the user and visually similar images are retrieved [5]. Using the metadata (not for retrieval) given

978-1-4799-3211-5/13 $31.00 © 2013 IEEE
DOI 10.1109/SITIS.2013.126

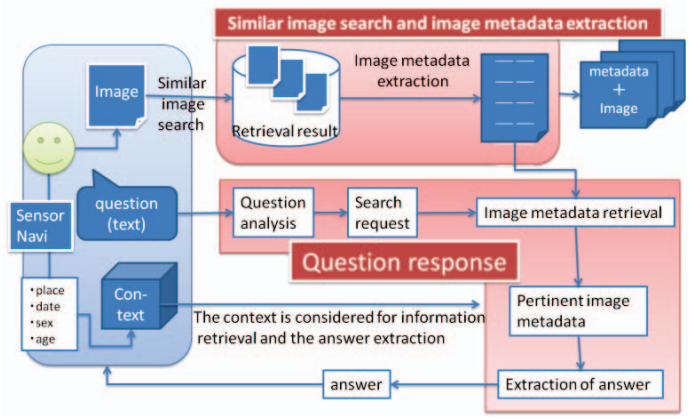

Figure 1. The overall architecture of the proposed system.

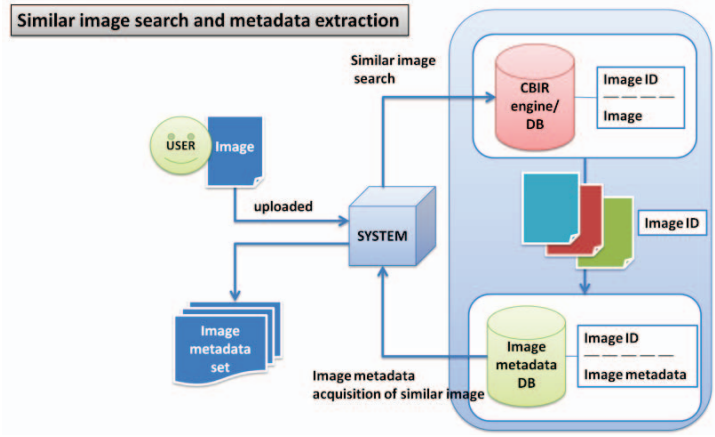

Figure 2. Image retrieval: image (meta)data extraction.

to a similar image, the content of an image related to it can be recognized.

B. *Context-awareness*

The information that fits user's context is retrieved. Information treated as belonging to the user's context includes items such as individual information specific to the user (user profile), GPS information, and so forth. These are obtained from the followings:

- the information terminals to the user,
- sensor information, and questions asked by the user in natural language and serve as a source of such information

III. PROPOSED SYSTEM

Fig. 1 illustrates the overall architecture of our proposed system.

There are three sources of information coming from the user, namely, the image, the textual questions, and the context (user's situation).

The system consists of: 1) the similar image search and metadata extraction and 2) the question answering.

A. *Similar Image Search and Metadata extraction*

The image data and its metadata extraction using CBIR is illustrated in Fig. 2.

Image metadata is extracted in the following order:

1) The image from a user is uploaded to the system,
2) A similar image is retrieved by the similar image search engine such as Google Image Search [6] and MiPai [7]. The reason we exploit such image search engines is we are not focusing on the CBIR techniques themselves, but a recommendation system by making use of such widely, easily available techniques.
3) The database is consulted to extract and manage the image metadata from the ID of the image that the engine selected as a candidate,
4) The acquired image metadata is associated to the image ID and stored as a dataset of image ID and image metadata.

B. *Question answering*

Fig. 3 shows the question answering.

The Question answering follows typical factoid type questions coined by Norman Mailer.

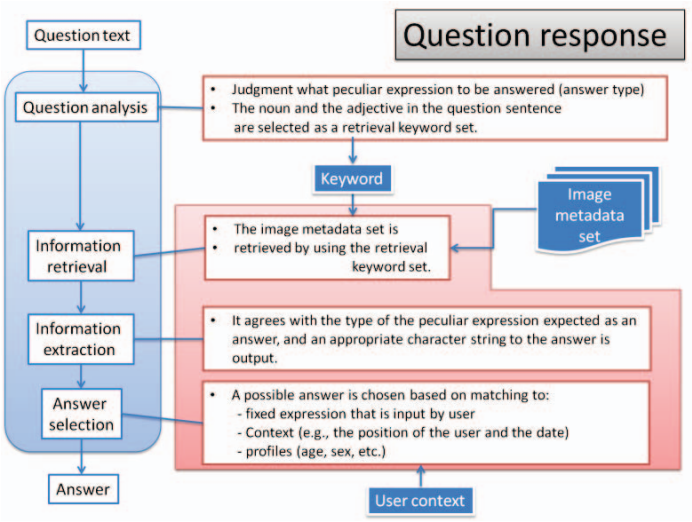

Figure 3. The procedure of the question answering.

The question answering is proceeded as follows:

1) user's question is analyzed in question analysis. Specificallly, the question type is determined, and then nouns and adjectives in the question are stored as a set of search keywords of the image.

2) The set of the search keywords are checked if they are matched or not with each metadata set of the similar image metadatabase.

3) The information retrieval part retrieves the image metadata set from the image part by using the input retrieval keyword set.

4) In the information extraction part, a character string of the image metadata matching the type of the peculiar expression expected as an answer and appropriate to the answer is output.

5) In the answer selection part, the inputted peculiar expression including keywords in natural language texts, user context, and profile information are matched in the answer selection part, and one of the answer candidates is selected and returned to the user.

6) Information used for matching is kept as temporal (time series) context in retrieved metadata. The question and answering, namely, steps 1) - 5) can be repeated, and the temporal context is used for such cases and for the recommendation to the same (kind) of users.

C. *Example of use*

We shall consider an example of the use of the system as shown in Fig. 4. First, a user presents a picture of an Asian skunk cabbage, and simultaneously, asks the name of the flower as the user does not know what it is. The system, according to the extracted information related to the image such as the content of the image, responds with the appropriate answer, which is the name of an Asian skunk cabbage.

Subsequently, the user asks a place where he can see it. In this case, the system introduces spot information according to the user context such as user's current location and the time.

D. *Expected Results*

Using the situation awareness of the system, recommendation precision increases through fusing CBIR (content-based image retrieval/matching) with the partial context obtained in time-series from other types of data such as textual information. For such fusing, the system needs collections of image, each annotated with a sufficient amount of metadata used for recommendations in multi-media (image, natural language text, etc) and frequently time-series dialogue. Indeed, manual annotation to enable image retrieval is expensive, time consuming and error-prone, and CBIR is very useful [5]. However, the method proposed in this paper tries to use the image metadata just for recommendations or explanations to use the object or to sell the product that the image represents but never for search. People are willing to annotate such information to image. Indeed, in the Web, there is a sufficient amount of natural language texts and images fused as data to explain each of products, sightseeing places etc. The problem for the automatic generation of such kind of metadata annotated to image is just how to, what to, or to what images to extract such image metadata. However, this can be realized by specifying templates or knowledge and ontology depending on each recommendation system and each image (e.g. illegal collection etc). Further, even if collections or links of image data annotated by metadata are difficult legally, online extraction and creation of metadata from natural language textual explanation of images can be done usually within interactive time. Thus, the proposed method is expected to be a promising one. Of course, dependent on recommendation systems, efficiency and precision of such

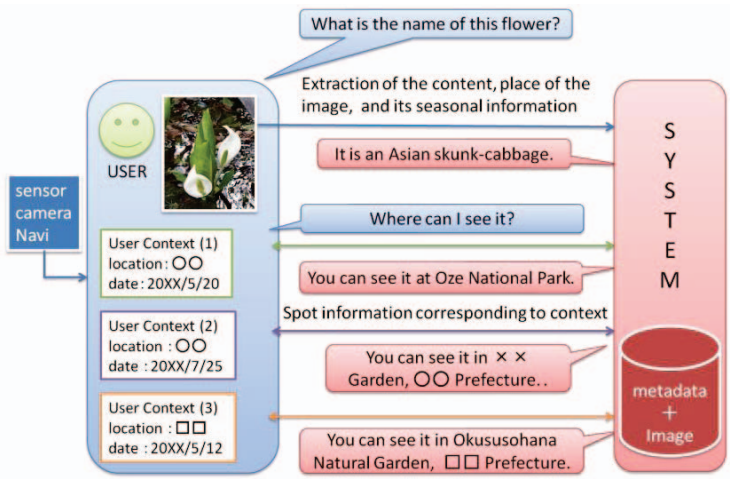

Figure 4. Context-aware image query response.

metadata extraction or creation by natural language processing have sometimes problems. For example, some images have not sufficient explanation texts in the same site and languages are different etc. They are future issues.

IV. Consideration on Dynamic Context

Generally, the things around us change in their conditions as time passes; the degree of information's freshness can be considered as a significant factor for the credibility of the information itself. Accordingly, more accurate answer to users can be achieved considering the update frequency of the databases.

A. *The influence of information's freshness on the search accuracy*

The freshness of information influences the search accuracy. The degree is represented as the parameters of speed and frequency. For example, a few hours previous information about the sports game situation is useless. Meanwhile, the state of the autumn leaves neither changes so quickly nor so often. However, they change every year. Furthermore, the names of buildings and creatures do not change until some events occur.

The next problem is how to get the latest information. Information on the Web is not always close to the latest one, especially for the image. In case of narrowing the applicant site through similarity match using a similar image search, the results change depending on the priority level of information freshness to similarity. In addition, it is very difficult to tell when the image or photo was taken. As a matter of fact, when the image namely photo is taken is different from the time when the photo image is uploaded on the Web. Therefore, accurate information cannot be obtained in many cases.

B. *Limitation of up-to-date information acquisision*

However, as long as image similarity is prior to the information freshness, there is no guarantee that we can acquire up-to-date information from the Web, especially concerning information on images. Technically, it would be difficult to identify the date when photos on the Web pages were actually taken without the annotation.

These problems can be solved by showing the date of the information source to the user and checking if the user wants to search again. Firstly, content based image retrieval is done by weighting the similarity more than the information freshness; here, the date of the article updated is acquired. If the user wishes re-search in response to the first answer with the date of the information source, the next search considers the update of information.

Or if we find the right weight between similarity and freshness, the system can be improved in its responses by automatically selecting Website(s) for use according to similar image searches.

V. Conclusion and Future Work

In this paper, we proposed a context-aware recommendation system that gives optimal information for users based on 1) a content-based image retrieval (CBIR) mechanism to search the similar images aiming to extract the detailed information to the text-inexpressible images 2) the contextual information of such similar images searched from the Web, and 3) user's dynamic context or situation considering time-variant factors as well as space factors.

Using the situation awareness technology considering time-variant information or the freshness factor in information, recommendation precision is expected to increase through fusing the partial context obtained by CBIR in time-series with other types of data such as (con) textual information, considering time-variant factor.

However, our proposed system is described just at a conceptual level; therefore, as a next step for the research, the prototype system will be developed based on more detailed implementation design. Then the usefulness and efficiency of the proposed idea and its implementation will be validated through the evaluation experiment using the prototype system.

References

[1] M. Okumura, H. Isozaki, J. Miyazaki, and D. Uemura, "The Context-Aware information recommendation system based on situation-dependent user taste modeling," Information Processing

Society of Japan, (in Japanese) Journal of Database by Information Processing Society of Japan, Information Processing Society of Japan (in Japanese), 2007.

[2] M. Okumura, H. Isozaki, P. Higashinaka, M. Nagata, and T. Kato, "Question answering system," CORONA PUBLISHING CO., LTD.(in Japanese), 2009.

[3] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," ACM Computing Surveys, vol.40, pp.1–60, 2008.

[4] G. Schaefer, Content-Based Image Retrieval:Some Basics, 2nd Int. Conference on Man-Machine Interactions, pp.21-29, 2011.

[5] G. Schaefer, Content-Based Image Retrieval: Advanced Topics, 2nd Int. Conference on Man-Machine Interactions. pp. 31–37, 2011.

[6] Google Image Search, https://www.google.co.jp/imghp?hl=ja&tab=wi&ei=5pd0Up_PNc6QkQWdkoGIBw&ved=0CAQQqi4oAg

[7] MiPai similiarity search engine, http://mipai.esuli.it/