I.J. Intelligent Systems and Applications, **2012, 7, 37-42**
Published Online June 2012 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2012.07.05

Design of Tabletop Interface for Adding Tags to Non-Annotated Image Collections through Natural Discussion

Kazuma Mishimagi
CyberAgent, Inc., Tokyo, Japan
Email: lonely.two.mn@gmail.com

Masashi Toda
Center for Multimedia and Information Technologies, Kumamoto University, Kumamoto, Japan
Email: toda@cc.kumamoto.ac.jp

Toshio Kawashima
School of Systems Information Science, Future University Hakodate, Hokkaido, Japan
Email: kawasima@fun.ac.jp

Abstract—Many media forms can be stored easily at present. Photographs, for example, can be easily stored even though most of them have not been edited. This means they will gradually lose their value and become essentially unusable. To make better use of photographs, we tried to make use of information provided by viewers who had seen and commented on them. We felt that analyzing this information would enable us to make maximum use of photographic data. To do this, we defined a "tag propagation" model and relationships between photos. We also proposed a system that uses image processing to analyze viewers' handling of photos and how the photos are relevant to each other. We then validated our model by using it.

Index Terms—annotation, image collections, viewer's action, propagation of tags

I. Introduction

Digital cameras have become popular in recent years and disc capacity has increased at the same time. This has enabled people to take and store photos easily. Most households, including those listed in the Hakodate[6] digital archive project that holds hundreds of thousands of old photographs, keep a large number of photos. However, most of them have not edited at all so it is impossible to search for them. To enable these photos to be used in an effective way, they must be edited.

Looking at photographs is a much more common activity than editing them. As an example, families viewing marriage ceremony photos tend to simply look at them and not edit them in many or even most cases. Logging and analizig above situation would enable pepole to tag and edit their photos in an effective manner.

The issues we addressed in this research were 1), What actions did viewers take pertaining to photographs that would yield what information? and 2), How could these actions be detected?. We first report on how this research model interprets viewers' actions, then describe how it detects and edits them by processing images from videos under discussion, and finally review and discuss the results obtained.

II. Approach

We propose a system that records the status of tabletop discussions, analyzes viewers' handling of them, and automatically adds information to photos.

With this system, viewers first spread a large number of photos on a table and discuss them. The viewers' handling of them is captured by a camera over the table and the system analyzes the way the photos are handled and adds that information (tags) to photos (Figure1). For example, one viewer takes a photo labeled "such-and-such university" and talks about it. During this time, another viewer notices other photo about the university and joins the discussion. At that time, a "Univ."-tag is added to the second photo from the first photo that was temporarily tagged "Univ.". As the conversation continues, a photo tagged "Festival" is put nearby these photos and a tag is added to it in a similar way (Figure2).

This system enables information to be added to photos and for them to be edited through a natural process, i.e., through discussions among persons who have some knowledge of the photo contents.

However, it is not yet known precisely what actions yield what information and what would be the best way to translate these actions into useful data. Therefore, we addressed this problem by defining a model of relationships between photos and tag-propagation. This led to the development of a system that analyzes videos of discussions, defines relationships between photos and tag-propagation, and adds information as model-based tags. This report discusses the results we have obtained and their validity.

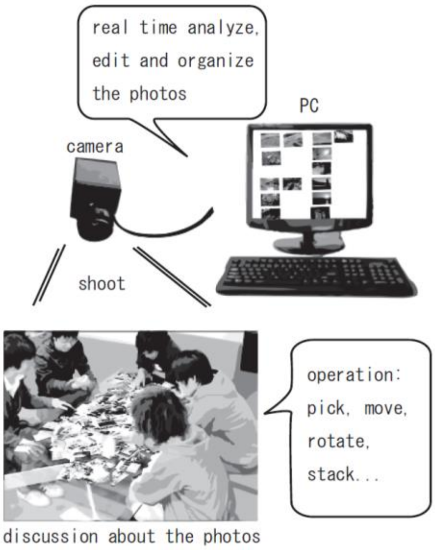

real time analyze, edit and organize the photos
PC
camera
shoot
operation: pick, move, rotate, stack...
discussion about the photos

Figure 1. Overview image of proposed system

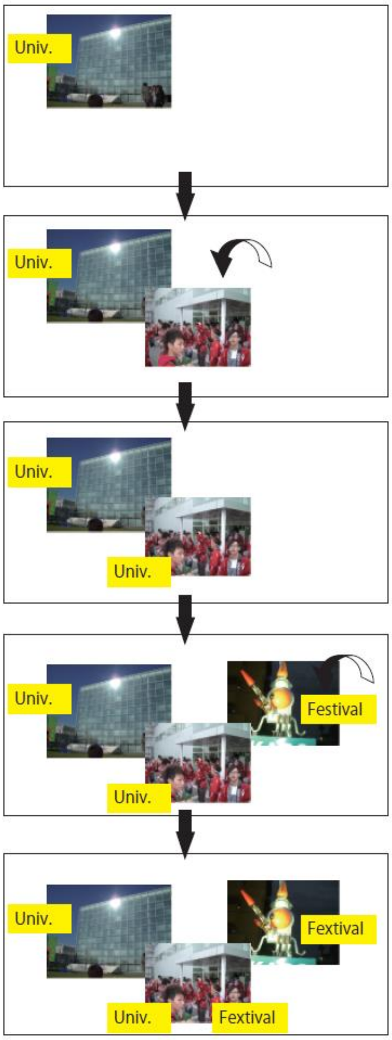


Figure 2. Adding tags by tag propagation

III. Related work

A considerable amount of research has been done in the area of editing image information and tabletop systems.

In researching automatic clustering, Hilliges et al. created a system that clusters similar images and picks out the best of them[7]. In addition, research on PhotoTOC[4] resulted in a proposed system that picks out representative images using timestamp and color histogram information.

A DigitalDesk[9] tabletop system was proposed in 1993. It is a system that bridges the physical and electronic worlds. Koike et al. attempted to solve the information direction problem with the concept of a rotary table[2]. They described layouts that are both sequential and spiral. In addition, Microsoft Surface, a commercially available surface computer developed by Microsoft Corp., takes advantage of its table-like form to interact with objects on it.

Although this research has addressed clustering by image features and editing using information like timestamps, it has not discussed clustering or editing by adding information.

Since the ultimate goal is similar to Flickr-like tagging, our study is closely related to Flickr[3]. Many people view, tag, comment on, and annotate photos, and these activities create searchable large-scale photo archives. However, since each of these archives are result of creation of information and archiving, the Flickr-like way would not work well for photos that are already archived without any supporting information. Furthermore, even if we were to add tags or comments to them, they would not provide much helpful information. Additionally, adding information to such large archives would consume a great deal of time and energy. We therefore consider that new structures are needed to add information to photos more naturally.

A research work has been published that compares manipulation of physical and digital media, in which it was concluded that physical metaphors and input methods are quite different in the digital realm[8].

J. Kim et al. logged all statuses of the desk and documents on it with object tracking by image processing[5]. Their work enabled them to use a PC to search for buried documents on the desk by tracking all of the documents on the desk continuously. This method, however, only made it possible to establish a connection between digital documents in the PC and physical ones on the desk. To our knowledge, no study has yet been conducted that adds new information to objects by analyzing the way the objects are handled.

IV. Models

A. Information of Photos and Tags

Photos include a great deal of information, not only image-processable information such as colors, but also background information such as who were there, where the location was, why the photo was taken, and under what circumstances situation it was taken. In photo-sharing services such as Flickr, such information is added as "tags". Consequently, our research goal was to add information to photos as Flickr-like tags.

B. Viewers' Actions and Status of Photos

In this work, we use the word "viewers" to refer to viewers who have some slight knowledge about the photos. Thus, our usage of the word does not include viewers who have no knowledge at all of them.

Important information about photos can come up in conversation when a large number of photos of mixed quality are being viewed.

If one of a group of viewers shows interest in a particular photo, he or she will usually draw it closer and look at it (Figure3). If he or she is not interested in the photo, it stays there without being moved further. However, if someone is greatly moved by it or has a comment about it, he or she will usually ask others for their opinions. Then, if others are also interested in it, the photo will be moved in front of them to be viewed (Figure4). Viewers pointing with their fingers at a photo are considered a sign of interest (Figure5). If many viewers point to the same photo at the same time, the photo could be considered to be the main topic of discussion. If viewers align or stack photos methodically, we can presume that they may be comparing or grouping them (Figure6). If one of the viewers is looking at a photo and other one notices it, then he or she will usually draw the related photo closer. These are common actions.

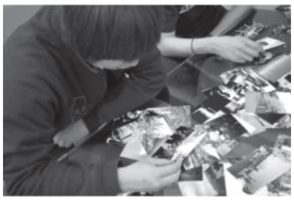


Figure 3 Interest in a photo

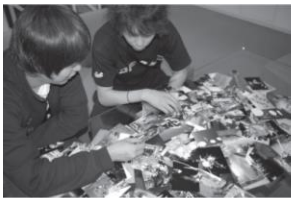


Figure 4. Two viewers are interested in a photo

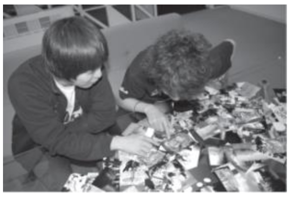


Figure 5. Pointing to a photo

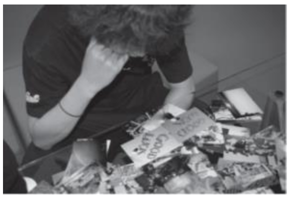


Figure 6. Aligning and comparing

There are a number of possible signal-level behaviors relevant to photos in the above-described situation that can be detected, which can be detected in Figures 7-10, respectively.

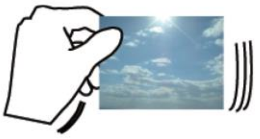


Figure 7. Position / moving.

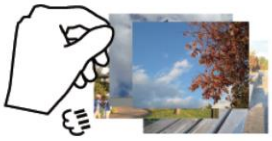


Figure 8. Overlapping

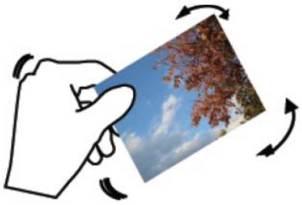


Figure 9. Rotation (angle).

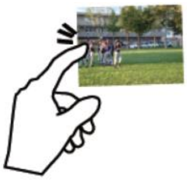


Figure 10. Pointing direction

C. Relationships between Photos and Propagation of Tags

Thus, much information can be extracted from viewers' actions. We focused on actions related to the gathering of related photos such as stacking and aligning. Therefore, we defined that if some photos were put nearby each other, they would be related to each other in some way.

We assumed that photos related to each other would share some common information and thus would have subliminal common tags.

When a photo *P1* relates to another photo *P2* through information *I*, if *P1* has *I* as a tag *T* preliminarily, we can add *I* to *P2* from *P1* by adding *T* of *P1* to *P2*. Thus we describe the adding of tags between photos related to each other as "tag-propagation". Using this method enables information to be added to photos that have no attached information.

In this way, large numbers of many kinds of tags will be added to photos through the discussion, and thus some of them may be incorrect. However, weighting the tags of these by discussing many time with other photo-set and with other persons makes it possible to increase the accuracy of information that the tags provide.

The extent of propagation is defined through three types: spatial extent, temporal extent, and hop count. This extent is represented as the weight of tags. The spatial extent weight is calculated based on the distance from one photo to another. Similarly, the temporal extent weight is calculated based on elapsed time. Here, hop count means the number of times that a tag is allowed to propagate. Each of the weights calculated for the three types are multiplied and added to the tags.

V. System

To validate the above-described models, we developed a system that analyzes videos of discussions, defines relationships among photos, and adds information as tags based on the models.

A. Required Features

Analyzing the situation referred to in Section2 requires the following features:

- Detecting/identifying photos on the table
- Understanding implications of photos
- Tracking photos
- Detecting photo-tagging relationships by tag-propagation

B. Process

The process overview is shown in Figure11.

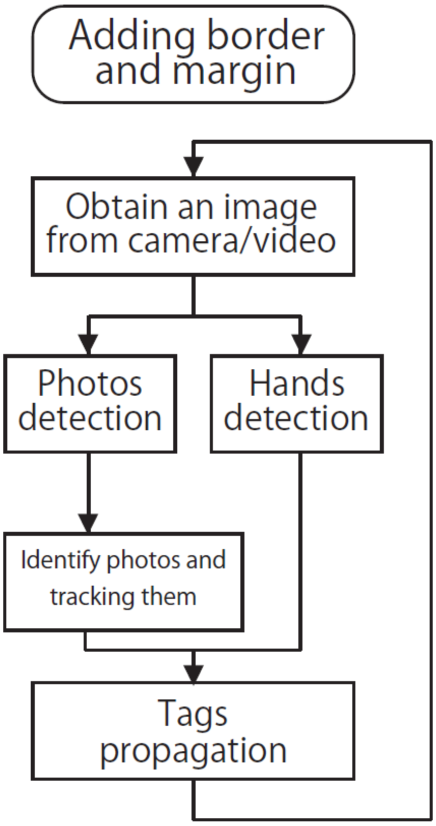


Figure 11 Process overview

1) Preparation of Photos

White margins and border lines around the photos were added to simplify the detection. This margin size is 10% of the length of the photo's short side from each edge and the border-line thickness is 2.5% of the length of its short side. These photos are printed out and used for discussion.

2) Detection and Identification of Photos

The system obtains tabletop images by means of a camera above the table. In the current system, it analyzes and binarizes recorded video images and performs contour definition on them to detect photo position and rotation.

To identify the images, we use SURF[1] feature descriptors. The SURF system includes scale, rotation, and translation-invariant 64-dimensional features. It is able to match the features of each photo and identify those on the table by matching them with preliminarily extracted features of stored photos in PC (Figure12).

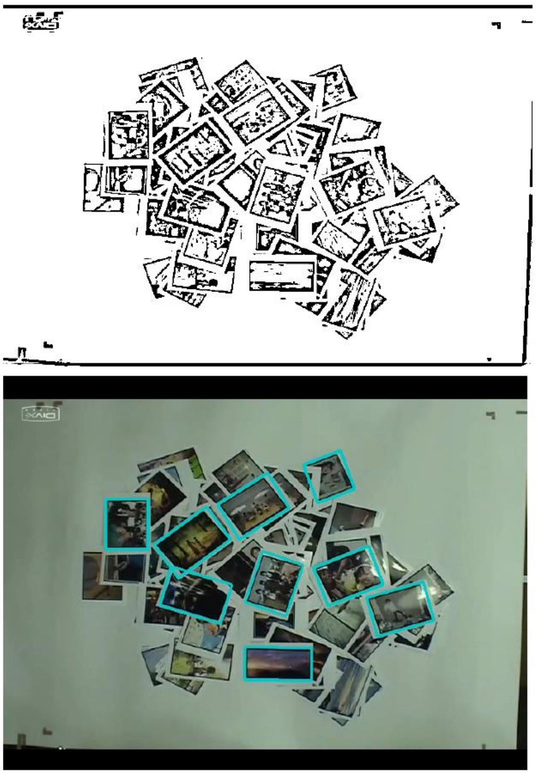


Figure 12 Detection of photos

3) Propagation of Tags

When a photo is identified, the system updates its current status, such as the position and angle used when the subject was photographed and the SURF descriptors used. If the photo subject has moved more than 30 pixels from its previous position, the system takes it as the beginning of a move. After one second, the system checks its position again. If it has left the current position, the system takes it as the ending of a move.

When the system detects that a photo subject's movement has ended and if the moved subject has some tags, the tag weights are calculated based on their physical relationships and the distance over which they have been propagated.

To distinguish the subject of a photo held in one’s hand from one that put the position, the system performs simple skin detection by color filtering. When the ending of a photo subject's move is detected and skin area overlaps it, no tags are propagated.

C. Hardware

To capture images, we used a common video camera that can capture 720x480-pixel images at a speed of 30 fps. The camera was fixed at 120 cm above the tabletop to capture the whole table. It was affixed to a tripod stand, had a self-produced arm, and its frame rate was set to 30 fps. The table was white in color and was covered with diffuse reflective material. The computer we used for processing was a common laptop PC with a Core2 Duo 2.2 GHz CPU and 2 GB RAM.

VI. Experiments

A. Overview

We conducted an experiment to confirm the validity of our proposed models and system. The experiment conditions were as follows:

- Subjects: Two students who were mutual friends
- Time: Five minutes
- Photos: Fifteen photos related to subjects, five of which were tagged
- Instructions to subjects: "View these photos freely for five minutes."

Five of the photos contained a total of ten tags, while the remaining ten photos contained no tags.

In this experiment, video images were analyzed by the system. Weights for spatial extent and temporal extent were defined by a step function. If the distance from one photo to another was within an 80-pixel range, the weight of its spatial extent was 1; otherwise, it was 0.

Similarly, the temporal extent was also limited by the step function. When the system detects that a photo subject's movement has ended and if less than 30 seconds elapsed since photos relevant to the one had been leaved the position, the temporal extent weight was 1.

Hop counts in the current system are limited to one hop. This means that the weights for tags that were preliminarily added to photos are 1 but that those for tags newly added in the discussion are 0.

B. Results

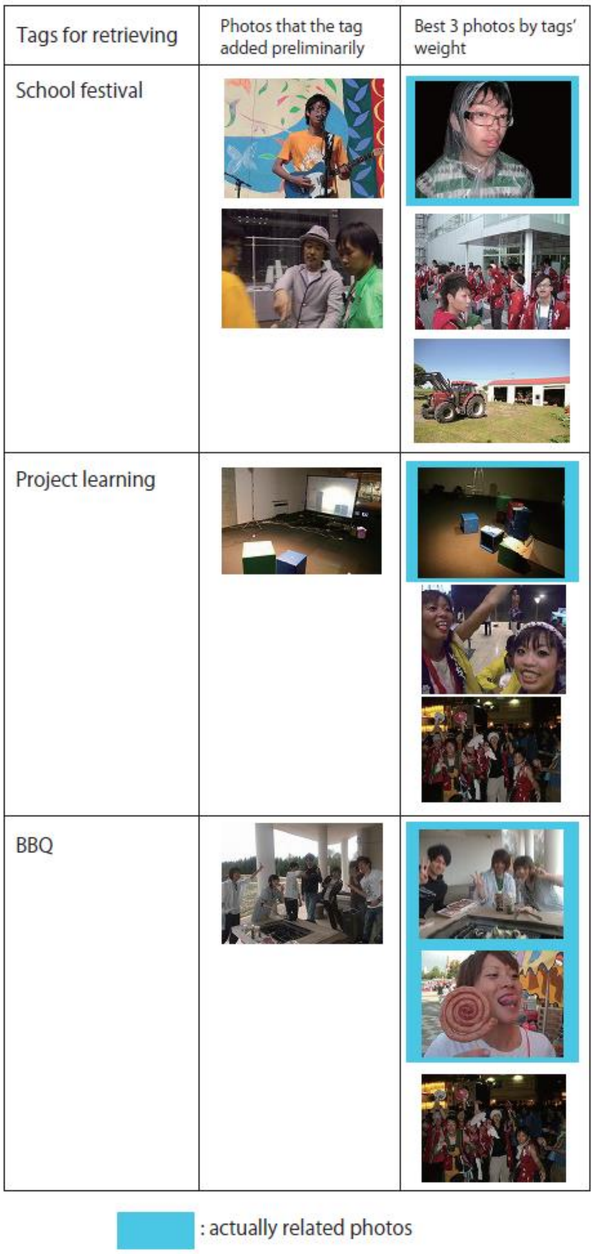
As experimental results, we retrieved photos used in the experiment by tags. An example of the results is shown in Figure13.

This table shows the results obtained in retrieving photos by several tags. The first column shows tags for retrieving, the second shows photos to which a tag was added preliminarily, and the third shows the best three photos retrieved by the tag. The retrieving rank is defined by the weight so that the photo that has the most heavily weighted retrieving tag is ranked at the top.

At the beginning of the experiment, five of the 15 photos had had 10 tags (i.e., 10 types of tags) added to them and at the end, a total of 67 tags (again, 10 types of tags) had been added to the 15 photos.

As indicated in the table, three separate tags were affixed to appropriately relevant photos. Of course all of them contained a lot of noise, but many useful results were derived from the weighted tags. Overall, eight of the 10 tags derives appropriately photos in the top three.

It is also important to note that some of the added tags did not have any image-processable information, meaning that the information was derived from "discussion". Thus these results show the usefulness of our proposed models system.



Tags for retrieving	Photos that the tag added preliminarily	Best 3 photos by tags' weight
School festival		
Project learning		
BBQ		

🟦 : actually related photos

Figure 13 Example of results

VII. Conclusion and Future work

Our research modeled viewers' actions and through it we developed a system that analyzes situations regarding the discussion of photos and that tags the photos. The results we obtained showed that useful information could be derived from 80% of the existing tags. These results verified the validity of our system.

To improve our system we are conducting many experiments using hundreds of photos and implementing a system that runs in real-time. In particular, we will attempt to achieve improved tag-propagation models. A more continuous weighting function should be developed as a means for decreasing noise. Improving the image detection ability of the system and refining tag-propagation models should ensure that this system will be a very useful one for editing large-scale image archives.

References

[1] H. Bay, T. Tuytelaars, and L. J. V. Gool, "Surf: Speeded up robust features", ECCV (1), pages 404-417, 2006

[2] K. F. Hideki Koike, Shintaro Kajiwara, and Y. Sato, "Information layout and interaction on virtual and real rotary tables", Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer System, pages 95-102, 2007

[3] Yahoo! Inc. flickr. http://www.flickr.com/.

[4] B. A. F. John C. Platt and Mary Czerwinski, "Phototoc: Automatic clustering for browsing personal photographs", Technical report, Microsoft Research, 2002.

[5] J. Kim, S. M. Seitz, and M. Agrawala, "Video-based document tracking: unifying your physical and electronic desktops", Proceedings of the 17th annual ACM symposium on user interface software and technology, pages 99-107, Santa Fe, NM, USA, 2004. ACM

[6] H. C. C. Library. Hakodate city central library digital archive collections. http://www.lib-hkd.jp/digital/.

[7] A. P. Otmar Hilliges and Peter Kunath, "Browsing and Sorting Digital Pictures Using Automatic Image Classification and Quality Analysis", pages 882-891, 3. Springer-Verlag, 2007.

[8] L. Terrenghi, D. Kirk, A. Sellen, and S. Izadi, "Affordances for manipulation of physical versus digital media on interactive surfaces", Proceedings of the SIGCHI conference on human factors in computing systems, pages 1157-1166, San Jose, California, USA, 2007. ACM.

[9] P. Wellner, "Interacting with paper on the digitaldesk", Communications of the ACM, 35(7):87-96, July 1993

Kazuma Mishimagi received M.S. in Systems Information Science from Future University Hakodate. His research interest in schooldays is Computer Vision.

Masashi Toda was born in Hamamatsu, Shizuoka Pref., Japan in 1969. He received B.S. from the University of Tokyo, Japan in 1993, and M.S. and Ph.D. in electro-informatics engineering from Hokkaido University, Japan in 1995 and 1998. From 1998 to 2001, he was a researcher in IS Labo., SECOM Co., Ltd., Japan. From 2001 to 2005, he was an assistant professor in School of Systems Information Science, Future University Hakodate, Japan. From 2005 to 2012, he was an associate professor in Future University Hakodate, Japan. Since 2012, he has been a professor in Center for Multimedia and Information Technologies, Kumamoto University, Japan. His main research interest is in image processing technology. He is also interested in human interface technology, wearable computing, ubiquitous computing, information retrieval technology, and educational information system. He is a member of the Information Processing Society of Japan.

Toshio Kawashima is currently a professor in School of Systems Information Science, Future University Hakodate.

How to cite this paper: Kazuma Mishimagi,Masashi Toda,Toshio Kawashima,"Design of Tabletop Interface for Adding Tags to Non-Annotated Image Collections through Natural Discussion", International Journal of Intelligent Systems and Applications(IJISA), vol.4, no.7, pp.37-42, 2012. DOI: 10.5815/ijisa.2012.07.05