Metadata Generation for Digital Photograph by using Smartphone's Sensors

Yoh Shiraishi*, Yuki Kato*, Yoshitaka Nakamura* and Osamu Takahashi*

* School of Systems Information Science, Future University Hakodate, Japan
{siraisi, y-nakamr, osamu}@fun.ac.jp

Abstract - Recently, taking photographs by using camera function of a smartphone for many purpose as well as a digital camera only for photo shooting becomes common. Metadata of digital photographs are add to the photographs at the shooting behavior. There are various kinds of services using these metadata. The existing photographs shot by a digital camera and smartphone's camera function include many metadata based on the Exif format. By using smartphone's sensors, it is possible to generate new kind of metadata for a photograph. For the purpose of sightseeing, persons such as photographers and travelers browse and share many photographs and the metadata.

This paper proposes a method for generating photograph's metadata by using multiple sensors equipped with a smartphone such as an acceleration sensor, direction sensor and camera function. In this study, we focus on sightseeing applications, and consider metadata available for the purpose. Our method uses an acceleration sensor, a direction and a face recognition function for generating metadata: a photographer's action, direction and number of persons in a photograph respectively. We implemented the prototype system on a smartphone and conducted the preliminary experiments.

Keywords: metadata, digital photograph, smartphone, sensor and action estimation.

1 INTRODUCTION

Many people use photographs to leave with memories. Recently, we often take photographs by smartphones as well as digital cameras. A smartphone has not only a camera function but also various functions. Meta data such as shooting data, shooting model, resolution and location are recorded in a digital photograph data by exchangeable image file format (Exif). There are many social networking services (SNSs) and applications that use the Exif metadata.

A smartphone has an acceleration sensor and a direction sensor. These sensors is not built-in the existing digital camera. There is a possibility that we can generate new kinds of metadata by using these smartphone's sensor.

Meta data are not data itself but are the related information with the data. When we take a photograph by digital camera, metadata such as shooting date, time and models are annotated to the digital photograph with Exif format. There are famous social networking services such as Facebook [1] and 4travel [2]. The Facebook supports uploading and sharing photographs and movies among people all over the world. The uploaded contents are arranged and classified by using metadata such as location information (geo-tagged contents) and keywords annotated manually by users. The 4travel is a service for supporting travels. Travelers are users of the service and can post photographs, the location information and the related comments. The users can evaluate and share the contents in the portal site for supporting travel. In this way, many social networking services (especially, ones of travel supporting) have many photographs and use the related metadata for sharing and communication of the uploaded contents. Generating new kinds of metadata promotes the popularization and expansion of the services, and develops new kinds of services with digital photographs.

This paper organizes as follows. In section 2, we refer to the related works and point out some problems. In section 3, we mention to our approach for the problems, and describe a new method for generating metadata for digital photographs by using smartphone's sensors. In section 4, we show the results of the preliminary experiments. Finally, we conclude in section 5.

2 RELATED WORK

Digital cameras support many kinds of Exif formats, but smartphones don't support some of camera parameters such as shutter speed, aperture of the lens, ISO sensitivity, focal length and color space. A digital camera for only shooting is superior to a smartphone for multiple functions (including shooting) in order to grasp the camera state.

On the contrary, equipping new kinds of sensors with a camera provides the potential for using these sensor data as metadata of the digital photographs. Watanabe at al. use some sensors that are not equipped in the existing digital cameras and acquire new kinds of metadata that are not included in the Exif format [3]. These metadata are provided for the browsing users to support understanding of the content of the photograph such as where the photograph was taken. However, this related work does not discuss concrete application examples that use such metadata.

There are many situations for sightseeing and traveling that use metadata acquired at shooting. Location information (namely, latitude and longitude) are often used for such applications. Kurata et al. developed the prototype system to support action planning of travelers [4]. The system can plot the POI (point of interested) uploaded from many travelers and visualize a distribution map of points with frequent shooting.

There are other studies that use location information as photograph's metadata [5]. These studies pay attention to the shooting points, classifies the points to some categories for visualization and extracts high spots for sightseeing.

There are lots of studies for action estimation by using an acceleration sensor and a smartphone in the field of ubiquitous and mobile computing. By using an acceleration sensor, we can estimate various kinds of actions: walking,

running, stopping, standing and so on. A smartphone has not only an acceleration sensor but also other sensors including a direction sensor, GPS, microphone and camera. A smartphone with these sensors has a great potential to estimate various kinds of user actions.

We can acquire time series data from an acceleration sensor of smartphone's sensor, and estimate the user state by using such sensor value. A microphone and camera can be used for grasping visually the state. Acquiring new kinds of metadata from smartphone's sensors when a photographer takes a photograph may cause the state and condition of the photographer.

Metadata about digital photographs are often used to support sightseeing. However, the existing works cannot estimate what is the aim of the tour (what is the aim of shooting) by using photograph's metadata. Realizing of estimating the shooting aim by using photograph's metadata is a challenge for effective sightseeing support.

Therefore, we propose a method for generating new kinds of metadata of digital photographs by using smartphone sensors in order to estimate what is the aim of a photographer at shooting. We target sightseeing as a concrete application of the proposed method.

3 PROPOSED METHOD

3.1 Approach

Digital camera is superior to a smartphone in grasping the camera state at shooting. However, we can use various kinds of sensor data from multiple sensors of a smartphone. For example, we can acquire time series of sensor data from smartphone's sensor such as an acceleration and direction sensor, and extract maximum and minimum values of such sensor data. It is a possibility to estimate typical actions of a photographer by using these smartphone sensors because these sensors are often used for action estimation in the existing works.

Table 1: Metadata of digital camera and smartphone

Meta data	Digital camera	Smartphone
Shooting date / time	supported	supported
Shooting point	supported	supported
Focal length	supported	not supported
Diaphragm	supported	not supported
Max/min of acceleration sensor	not available	available
Shooting direction	not available	available
Time series of acceleration data	not available	available
Time series of direction sensor data	not available	available
Number of times of continuous shooting	not available	available
Time to shoot	not available	available
Face recognition	available	available

Our research aims at generating new kinds of metadata of photographs to support sightseeing. In this study, we consider photograph's metadata for grasping the following situations:

(a). The target of a photographer at shooting
(b). Feeling of a photographer at shooting
(c). Location of a photographer at shooting

In order to grasp the target at shooting, we classify photographs into some categories. Each category represents the type of the photograph. The type of a photograph is metadata to estimate what kinds of the photograph and the shooting situation. If we can arrange photographs by using the type of the photograph (namely, sightseeing), a user (not the photographer) to browse these photographs grasp the aim of sightseeing of the photographer based on the type of the photograph.

A survey report by Net Mile suggests that "who goes together", "what to do in the tour", "what cost of travel" and "what a traveler wants in the tour" are important when sightseeing. We pay attention to "who goes together" and "what to do" in the tour. In order to estimate such situations, we assume four categories ("person", "group", "landscape" and "dish") as the photograph types.

"Person" indicates a travel alone or in two some. "Group" indicates a travel that more than three persons go on such as family, school and company trip. "Landscape" indicates a travel that a traveler goes around sightseeing spots and takes photographs many times. "Food" indicates a travel for visiting famous restaurants and gourmet tour. We classify a photograph shoot by a smartphone into each of these four categories. In realistic situation, these photographs cannot be classified in such simple way. As the first step of our research, we try to make rough classification by using multiple sensors of a smartphone.

We consider two kinds of metadata ("the number of times of continuous shooting" and "time to shoot from starting the camera application") to express feeling (for example, "fan", "surprise") of a traveler at shooting.

The number of times of continuous shooting is how many times a traveler take photographs continuously with running the camera application. When a traveler finds a famous sight and spot, he or she will take photographs of the sight and spot not once or twice. The proposed system starts to count the number of times of continuous shooting after activated, and resets the number to zero when it is stopped. The time to shoot is the time from activating the camera function to shooting. When a traveler is surprised at the sight, he or she will start shooting once activating the camera function. These metadata may inform the feeling of the photographer to other travelers on a web site for traveling. Currently, a user for such site posts his or her comments and evaluation points for recommendation manually. Generating such metadata automatically will reduce the posting cost after traveling.

By recording location information of a photographer at shooting, it can display these points on a map. It can show the type of photographs, the number of times of continuous shooting and the time to shoot related to the shooting point.

These metadata will be useful for realizing a service with good quality. If we can know type of a photograph at the shooting point, we will estimate the target of the shooting and selects the photographs with the specific target. If we can know the number of times of shooting and the time to shoot, we will estimate the feeling of the photographer and use these metadata as materials for travel planning.

3.2 Acquisition of sensor data

In the proposed system, metadata are generated at different phases as follows:

- Phase 1: time from invoking the application to first shooting
- Phase 2: time at each shooting
- Phase 3: time after shooting (off-line processing)

In phase 1, the system acquires 3-axis acceleration data by using the built-in acceleration sensor, direction data such as azimuth, roll and pitch by using the direction sensor and time to shoot. In phase 2, it acquires roll value of direction sensor data at each shooting, the number of times to push the shooting button and whether the macro mode is on or not. We use the macro function to shoot nearby object such as dishes and flowers. In phase 3, it detects whether persons are in the photograph, and not and counts the number of the persons in the photograph by using face recognition as an image processing technique.

3.3 Generation of metadata by using sensor data

We use actions of a photographer, camera direction at shooting and the number of faces in a photograph to generate the type of the photograph. We use the time to shoot and the number of times of continuous shooting to generate metadata related with the feeling of the photographer.

Our system classifies the shooting photograph into the following group: “person”, “group”, “landscape”, “dish”. In this paper, we define “person” as one person or two or three person, and define “group” as more than four persons.

Our system executes the following steps for classifying photographs:

- Step [1]
  It examines whether persons (faces) are in the photograph by using face recognition. If it detects persons, go to the step [2]. Otherwise, go to the step [3].
- Step [2]
  It counts the number of the faces in the photograph. If the number is one or two, it classifies the photograph into “person”. Otherwise (namely, more than three faces), it classifies the photograph to “group”.
- Step [3]
  It examines the camera direction. If the direction is downward, go to the step [4]. Otherwise, it classifies the photograph into “landscape”.
- Step [4]
  It checks the macro mode. If the mode is on, go to the step [5]. Otherwise, it classifies the photograph into “landscape”.
- Step [5]
  It estimates the photographer’s action by using the acceleration and direction sensor of the smartphone. If the estimated action is included in action patterns that a user does not take at shooting dishes (for example, squatting down, looking over from side to side, rising the camera), it classifies the photograph into “landscape”. Otherwise, it classifies the photograph into “dish”.

We use the acceleration sensor for estimating “squatting down” and “rising the camera” and the direction sensor for estimating “looking over from side to side”. In the latter case, we examine the difference of directions between at starting the application and at shooting.

4 EXPERIMENTS

4.1 Implementation of the prototype system of the proposed method

We implemented the proposed method as the prototype system on a smartphone (Xperia acro, Android 2.3.4). Fig.1 shows the interface of the prototype system.

Figure 1: The interface of the prototype system

4.2 Experimental results

First, we examined typical actions at shooting. We gave an examinee the instruction to do the following actions: (1) move the smartphone forward, (2) move it backward (return to initial position), (3) move it to left, (4) move it to right (return to initial position), (5) move it backward, (6) move it forward (return to initial position), (7) move it to right and (8) move it to left (return to initial position). Fig.2 shows time series data of 3-axis acceleration sensor when the examinee acting along the instruction.

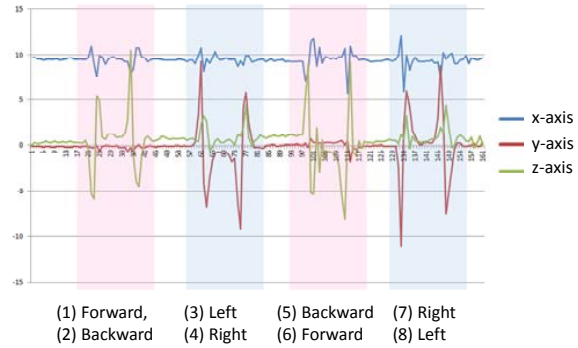

Figure 2: The change of 3-axis acceleration sensor

Fig.2 suggests that y-axis data are highly changed when an examinee move the smartphone from side to side (left to right or right to left), and z-axis data are highly changed when moving it forward and backward. These actions have typical pattern of change of time series data. If we extract such pattern, we can estimate the actions: move from side to side, and move forward or backward.

Figure 3 shows time series data of a direction sensor on a smartphone. An examinee makes the following action: (1) turn the smartphone upper, (2) turn it to left, (3) turn it down and (4) turn it to right. After doing each action, he returns it to initial direction.

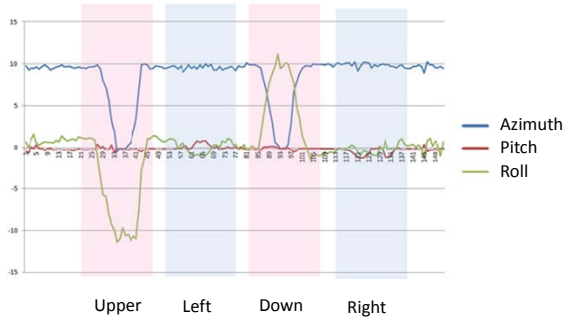

Figure 3: Time series data of a direction sensor

In this study, we use y-axis and z-axis acceleration sensor data for estimating actions that a user does not make when shooting a dish, and use roll value of a direction sensor for estimating the camera direction at shooting.

5 CONCLUSION

In this paper, we proposed a method for generating photograph's metadata by using sensors equipped with a smartphone such as an acceleration sensor and camera function. In this study, we focus on sightseeing, and consider metadata available for the purpose. This method generates some kinds of metadata such as photographer's behavior, direction and number of persons in a photograph for classifying photographs shot by a smartphone. Our method uses an acceleration sensor, a direction sensor and a face recognition function for generating metadata: a photographer's action, direction and number of persons in a photograph respectively. We implemented a prototype system of the proposed method on a smartphone and conducted the preliminary experiments.

REFERENCES

[1] Facebook, http://www.facebook.com/facebook.

[2] 4travel, 4travel, Inc., http://4travel.jp (*in Japanese*).

[3] K. Watanabe and K. Tsukada, "WillCam: a digital camera visualizing users' interest," Information Processing Society of Japan, Interaction 2008, pp.193-194, 2008 (*in Japanese*).

[4] Y. Kurata, K. Sugimoto and N. Yabe, "Onsite Tourist Guide without Detail Information: Extraction and Application of Point-of-Interest Data," Annual Conference on GIS Association of Japan, 2010 (*in Japanese*).

[5] K. Miwa, K. Oku and F. Hattori, "Extraction of a Viewable Point from the Photograph Data with Location Information for Route Recommendation," The 2nd Forum on Data Engineering and Information Management (DEIM), I-9, 2010 (*in Japanese*).