WillCam: a digital camera visualizing users' interest

Keita Watanabe
Graduate School of Media and Governance,
Keio University
5322 Endo Fujisawa
Kanagawa 252-8520, JAPAN
100kw@sfc.keio.ac.jp

Koji Tsukada
National Institute of Advanced
Industrial Science and Technology
1-18-13 Sotokanda, Chiyoda-ku,
Tokyo 101-0021 Japan
tsuka@acm.org

Michiaki Yasumura
Faculty of Environmental Information,
Keio University
5322 Endo Fujisawa
Kanagawa 252-8520, JAPAN
yasumura@sfc.keio.ac.jp

CHI 2007, April 28 – May 3, 2007, San Jose, USA
ACM 1-xxxxxxxxxxxxxxxxxx.

Abstract

With the increased usage of digital cameras and camera-enabled mobile phones in recent years, large numbers of photos are being taken. Many of the photos that are taken are used little, if at all. Researchers and companies have developed systems where photos can be annotated or tagged to facilitate storage and retrieval. However, people are often unwilling to spend the time and effort to carry out annotation.

To solve this problem, we focus on real-time annotation where photographs are annotated at the time when they are taken. We propose a novel digital camera, 'WillCam', which enables users to capture various information, such as location, temperature, ambient noise, and photographer facial expression, in addition to the photo itself. WillCam also helps users express their interest -what object or information in the picture/scene is most important for them- visually.

Keywords

Context awareness, Users' interest, Digital photography, Visualization, Meta-data

ACM Classification Keywords

H5.2. [Information interfaces and presentation]: User Interfaces -*Input devices and strategies*

Introduction

The widespread availability of digital cameras with features such as auto-focusing and de-blurring have allowed relatively unskilled people to take pictures quickly and easily. Freed from the costs of analog film and the development of that film, people can take many pictures without caring about the cost of poor shots (which can be quickly deleted). The remaining effort in digital photography for the consumer user largely resides in organizing, sharing, and printing the images after they have been shot. Often, digital photos end up being stored on computers in large and organized collections that are difficult to search and browse. Photographs become lost in the resulting large and amorphous collections and in some cases may never again be viewed after being stored.

There are many services and research projects to utilize stored photographs by annotating or tagging them [1, 10]. Annotations are not only useful for browsing and searching, but also for adding to the value of the photographs by providing contextual information (e.g. locations, situations, or stories) relating to the photographs (much as labels enhance the value of artifacts displayed in museums). However, there is an asymmetry in effort and reward between the author/photographer and the viewer which stands in the way of annotation. In the current situation, the photographer does the work, while the viewer gets the benefit.

To solve this problem, we focus on real-time automated annotation techniques that require little or no additional effort from the photographer, while continuing to provide benefits in terms of ease of search and browsing, and in terms of viewing the photos in a meaningful context. Annotation can be carried out by capturing various information from sensors, such as location, temperature, ambient noise, and photographer facial expression, at the time when photographs are taken.

We propose a novel digital camera, 'WillCam' (Figure 1) that provides the automated annotation capabilities introduced above. WillCam helps users easily annotate their photographs without troublesome operations, such as inputing annotations or tags for each photograph. In addition, WillCam enables users to express their ‘interest’ in taking the photo (e.g. what object or sensor information is most important for them) visually.

In the following section we describe the contextual information that WillCam associates with photographs.

Context of photographs

WillCam collects contextual information of three types.

1. Author/photographer information: "who" takes a photograph. Most users are more interested in photographs taken by their families or friends than photographs by strangers.

2. Situations: Information about how and where photographs are taken. Photographs taken at famous places or emblematic situations (e.g. "hot" summer, and "noisy" street) may attract users' attention.

3. Targets: "what" object or situation in the scene portrayed in a photo is most important for the photographer.

WillCam Functions

This paper focuses on three core functions of WillCam: VisualEXIF, RealFocus, and MetaFocus.

Figure 1. Prototype system of WillCam

VISUALEXIF

VisualEXIF is a technique for visualizing various types of information concerning the situation and who photographed it: *how/where/by whom* photographs are taken. Digital cameras usually attach EXIF information (e.g. product name, company name, and date) to captured photographs as meta data. VisualEXIF generates metadata that is composited into photographs (Figure 2). The VisualEXIF functionality focuses on capturing further information with various sensors: temperature, ambient noise, brightness, location, facial expression and posture of the photographer, and so on (Figure 3) . Consequently, VisualEXIF helps users understand various situations - how/where/by whom photographs were taken- at a glance.

Figure 2. Sample photograph taken with WillCam

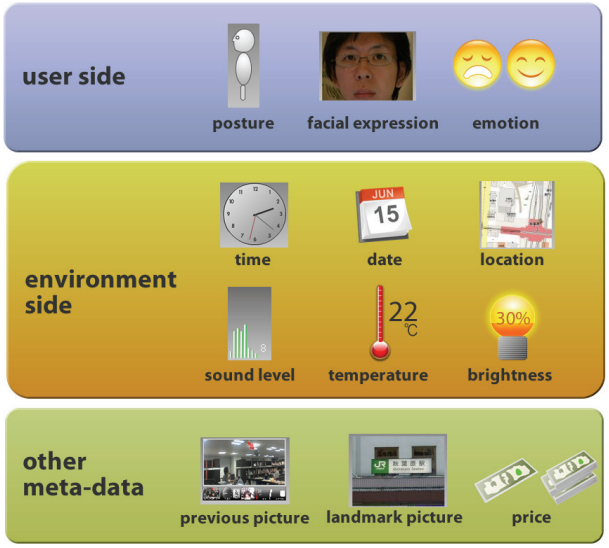

Figure 3. Variety of VisualEXIF Metadata Options.

REAL FOCUS

RealFocus is a technique for emphasizing focused objects in photographs. RealFocus extends the existing focus point with special pointers, such as arrows or circles, overlaid on preview images (e.g., the pointer shown in Figure 2). Users can move the RealFocus pointer icons with a trackpoint, and can change the size of them with a shutter button that has an associated pressure sensor. When a user becomes particularly interested in an object, he can express that interest by pressing the shutter button more strongly.

META FOCUS

MetaFocus is a technique for emphasizing focused VisualEXIF icons. The most important target doesn't always exist inside photographs. For example, people may take photos which seem meaningful to them at the time because of extraneous factors. As a result, there original intentions may be difficult to decipher from what can be seen in the photo. To signal this type of associated interest, the author can adjust the temperature icon associated with the RealFocus pointer. Thus, authors can convey their sense of how important a particular photo is quickly, using MetaFocus.

IMPLEMENTATION

HARDWARE

The prototype of WillCam consists of a mobile computer, cameras, and several sensors. The WillCam prototype uses a SONY VGN-UX90PS as the computer (Figure5). The screen resolution is 1024 * 600. A user can move the RealFocus icon using a trackpoint mounted on the right side of the computer. Two cameras are used: a built-in camera on the front of the computer (300,000 pixels) for capturing a photographer's face, and a USB camera attached on the back of the computer (1,300,000 pixels, Logitech QVX-13NS camera) for capturing pictures.

Figure 4. Meta focus: focusing on date icon

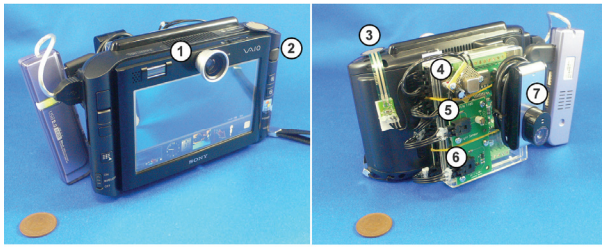

Figure 5. Hardware: (1) Camera and wide-angle lens, (2) Track point, (3) Pressure sensor, (4) Acceleration sensor, (5) Luminance sensor, (6) Temperature sensor, (7) USB camera.

Third, we attached a Phidgets Interface Kit and three sensors (an acceleration sensor, a temperature sensor, and a luminance sensor) on the back of the computer. We also set a pressure sensor on the top as a shutter button. Moreover, we used the USB camera's built-in microphone to detect the surrounding sound level.

SOFTWARE

In this section, we explain the software configuration of the WillCam prototype. The software mainly consists of two components: (1) a front-end GUI displaying preview images and VisualExif/RealFocus icons, and (2) a back-end program that handles input from sensors as well as capturing the current image displayed on the front-end. We developed the front-end with Adobe Flash and the back-end with Microsoft Visual C#. Captured images are saved in JPEG format with a resolution of 1024*600. The sensor data, displayed as VisualEXIF icons, is also saved as EXIF information in each of the resulting JPEG images.

Discussion

Our design of the WillCam was influenced by an informal survey of the various genres of existing photos. While it was often possible to discern the photographers' interest in viewing photos, there were still many photos, where our information couldn't decipher what objects or situation was of interest to the photographer. For example, some pictures were taken when people bought high-priced goods for review and compared options. However, this interest could not be discerned simply from viewing the photo. In general, interests that are strongly related to individual motivations and situations will be difficult to recognize automatically. The situation is somewhat like a person watching home movies shot by a stranger and not being able to appreciate the value of many of the clips because of not knowing the situations and motivations that governed the decisions to shoot them.

We carried out an informal user analysis of the WillCam prototype. Participants found the inclusion of the photographer's face in the photo to be an interesting feature, particularly when the photo involved a social interaction between the photographer and the people being photographed. We found that adding the user's facial expression and background helped to inform viewers of the situation in which the picture was taken. Since WillCam combines the original photograph with visualized meta-data in JPEG format, the annotated photos can be displayed easily on a variety of platforms. However, further research is needed to develop a method of browsing annotated WillCam images. We also identified a number of problems with the current WillCam prototype based on our evaluation. Since the current WillCam saves all captured information - a picture and VisualEXIF/RealFocus icons - in JPEG format, users can't separate these icons after taking pictures. To solve this problem, we are considering a new capture method: saving all information in a SWF (Shockwave Flash) file. SWF files are easily played not only in personal computers (e.g. Windows/Mac/Linux), but also in various game machines (e.g. PSP, Wii, and XBOX360) and mobile phones. Moreover, pictures in SWF format have more options for manipulation than usual images. For example, users may change the layout and visibility of VisualEXIF icons after taking pictures.

Related Work

Prior research has addressed the problem of storing photographs more effectively, including the use of efficient annotation techniques [1, 10] and visualization strategies [2]. In contrast to this earlier work which has emphasize a separate editing process, we focus on real-time automated annotation techniques.

Research has been carried out on automated annotation for digital photographs. There are two

approaches to solving this technical problem: automated methods based on feature analysis of images [6, 7], and the use of context information from sensors [3, 4, 5, 8, 9]. For instance, WayMarkr [3] utilized a camera-enabled mobile phone to take continuous photographs from the vantage point of the wearer (e.g. mounted on a backpack strap) with location information calculated from cell tower information. Davis et al [4, 9] proposed a system for inferring location information for pictures taken with camera phones. Capturing the Invisible [5] designed real-time visual effects for digital cameras using simulated sensor data. ContextCam [8] proposed a context-aware video camera that provides time, location, person presence and event information using several sensors and machine learning techniques. Our approach is different in visualizing more context information and users' interests with RealFocus/MetaFocus icons.

Conclusion

We propose a novel digital camera, 'WillCam', which helps users easily annotate their photographs without troublesome operations, such as inputing annotations or tags for each photograph. In addition, WillCam enables users to express their 'interest' in taking the photo (e.g. what object or sensor information is most important for them) visually. We plan to evaluate the efficiency of VisualEXIF/RealFocus/MetaFocus functions with a field test. We expect that WillCam annotation may new ways of experiencing and enjoying photographs.

RFERENCES

[1] Abowd, G. D., Gauger, M., and Lachenmann, A. 2003. The Family Video Archive: an annotation and browsing environment for home movies. In Proceedings of MIR '03. ACM Press, New York, NY, 1-8.

[2] Bederson, B. B. 2001. PhotoMesa: a zoomable image browser using quantum treemaps and bubblemaps. In Proceedings of UIST '01. ACM Press, New York, NY, 71-80.

[3] Bukhin, M. and DelGaudio, M. 2006. WayMarkr: acquiring perspective through continuous documentation. In Proceedings of MUM '06, vol. 193. ACM Press, New York, NY, 9.

[4] Davis, M., King, S., Good, N., and Sarvas, R. 2004. From context to content: leveraging context to infer media metadata. In Proceedings of MULTIMEDIA '04. ACM Press, New York, NY, 188-195.

[5] Hakansson, M., Ljungblad, S., and Holmquist, L. E. 2003. Capturing the invisible: designing context-aware photography. In Proceedings of DUX '03. ACM Press, New York, NY, 1-4.

[6] Jeon, J., Lavrenko, V., and Manmatha, R. 2003. Automatic image annotation and retrieval using cross-media relevance models. In Proceedings of SIGIR '03. ACM Press, New York, NY, 119-126.

[7] Jiebo Luo Savakis, A. 2001. Indoor vs outdoor classification of consumer photographs usinglow-level and semantic features. In Proceedings of International Conference on Image Processing, vol. 2, 745-748.

[8] Patel, S.N. and Abowd, G.D. 2004. The ContextCam: Automated Point of Capture Video Annotation. In Proceedings of UbiComp 2004: Lecture Notes in Computer Science, pp. 301-318.

[9] Sarvas, R., Herrarte, E., Wilhelm, A., and Davis, M. 2004. Metadata creation system for mobile images. In Proceedings of MobiSys '04. ACM Press, New York, NY, 36-48.

[10] Shneiderman, B., and Kang, K. 2000. Direct Annotation: A Drag-and-Drop Strategy for Labeling Photos. In Proceedings of IV2000. 88-95