CameraMatch: Automatic Recognition of Subjects using Smartphones—toward Entertaining Photo Sessions

Kohei Matsumura
School of Systems Information Science,
Future University Hakodate
116-2 Kamedanakano-cho, Hakodate,
Hokkaido 041-8655, Japan.
matsumur@acm.org

Yasuyuki Sumi
School of Systems Information Science,
Future University Hakodate
116-2 Kamedanakano-cho, Hakodate,
Hokkaido 041-8655, Japan.
sumi@acm.org

CHI 2013 Extended Abstracts, April 27–May 2, 2013, Paris, France.
ACM 978-1-4503-1952-2/13/04.

Abstract
We seek to increase both the enjoyment in a photo session and the amount of information of a photo by appending Manga iconography to photos automatically. By reducing the number of dimensions in a video to that of a photo, a user can easily understand events and the context of the video in less time. However, it is not easy to realize such a system. In this paper, toward automatic Manga iconography generation, we report on our feasibility study in which we attempted to associate subjects in a video with their smartphones.

Author Keywords
Camera; Smartphones; Pattern Matching; Manga Iconography;

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction
The difference between a video and a photo can be viewed in terms of dimension. A video has a timeline while a photo does not. The timeline increases the amount of information a video has and enables a user to have a richer experience—for example, the context

of the video and the movement of the object in the video—than a photo. However, it takes a certain amount of time (equal to the length of the video) to watch a video. Video abstraction or video summarization research efforts focus on reducing this viewing time.

Since smartphones have recently become increasingly popular, the users easily have a small photo session and the needs for the enjoyment in a photo session are increasing.

Annotating a photo is one way in which both the amount of information and the enjoyment it contains can be increased. We propose a method that has two features. One feature endows the photo with rich information similar to a video, while the other feature enables the photo to be comprehended at a glance. Our proposed method inserts Manga iconographies into a photo to increase the en it contains with the aid of a sensor network consisting of smartphones. Sensors embedded in the smartphones are used to generate the Manga iconographies. As the first step toward fruition, this paper describes a subject recognition system that uses both a camera and a sensor network system consisting of smartphone.

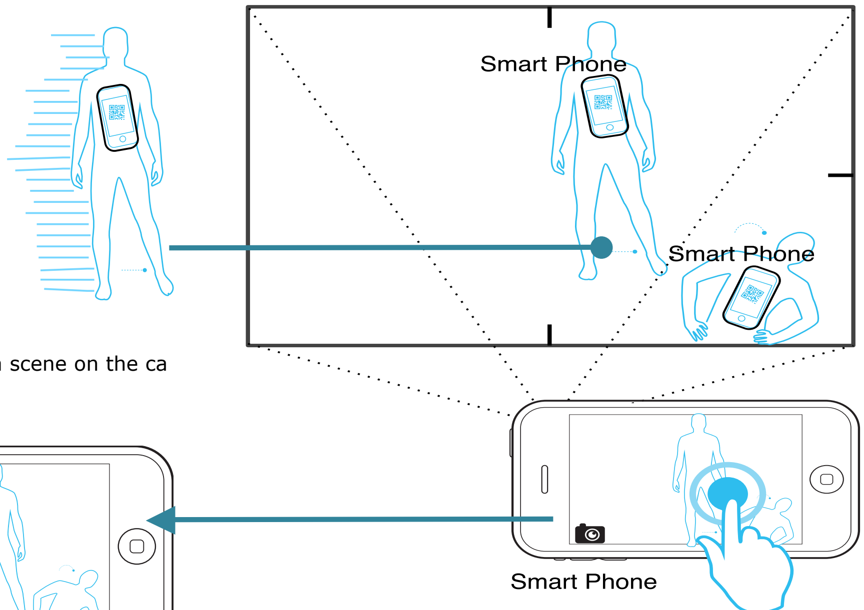
Related Work

Video summarization has been investigated in many studies. DeMenthon et al. [1] proposed an algorithm that extracts key-frames and demonstrated a video player that automatically makes chapters using it. Using their player, a user is able to make chapters dynamically by simply changing the summarization level. Uchihashi et al. [2] proposed a summarization tool that automatically lays out still images of key-frames in a manner similar to that of comics. The aim of these studies is to process commercial video such as that of movies and sports, which include scene changes. In contrast, our target is homemade videos that have no scene changes of such.

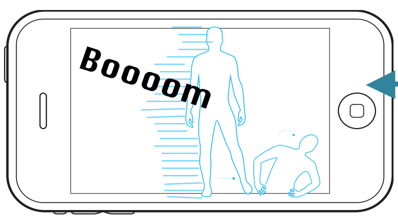
Nguyen et al. [3] developed a system that summarizes a video as a 3D object by stacking video frames on the time axis. Our aim is not to summarize a video as a 3D object to increase the amount of information, but to append annotations in the manner used in comics, such as speed lines and onomatopoeias, onto a photo.

In some studies, effects are appended onto the photo to increase the amount of information it contains and its power of expression. Masuch et al. [4] proposed a system in which speed line effects are appended to a photo using video features extraction. Kim et al. [5] developed a system that is able to append some effects semi-automatically to a scene by tracking the movement of objects in the video. Bennett et al. [6] proposed a system that is not only able to make movie digests, but is also capable of making motion blur effects using a virtual shutter. In these studies, all the effects are created from the results of video feature extraction. In contrast, we aim to create visual effects by calculating the movement of a subject using sensors embedded in a smartphone in the subject’s pocket. This makes it possible to append effects created from movements performed outside the field of view of the camera.

Hwang et al. [7] proposed a system that transforms a movie into a still image that is enhanced in a manner similar to that of comics. They not only transform the video frames, but also transform scripts and sounds to onomatopoeias in order to increase both the amount of information and the power of expression. In like fashion, we also aim to use a number of modalities and several of the sensors possessed by smartphones, such as



A person jumps into a scene on the ca mera.



Manga iconographies, such as onomatopoeia (illustrated as "Boooom") and speed lines are automatically generated in the picture.

Figure 1: When a photographer touches the screen on his/her smartphone, the shooting event is sent to the smartphones around him/her and each smartphone (except for the photographer's) sends its sensing data to the photographer's phone.

accelerometer, gyroscope, and microphone.

Manga Expressions in Photography

We use Manga iconographies as effects to increase the enjoyment in a photo session. As outlined above, the objective of our approach is not to extract features from a video but to create Manga iconographies based on data acquired from sensors embedded in smartphones. Figure 1 shows an example in which a photo is being taken based on our concept. The advantage of our approach over video feature extraction is that it is able to create Manga iconographies from movements performed outside of the view of the camera.

Several challenges currently exist on the path to realization of our proposed approach. The first step toward fruition, however, is to associate subjects on the camera with their smartphones because our aim is to create Manga iconographies of each subject. We have developed a camera application that communicates with sensor networks consisting of smartphones. Our technique is analogous to the technique proposed by Rofouei et al. in which the movement of users is associated with their smartphones using Microsoft Kinect [8]. We also try to associate subjects on the camera with their smartphones. Our progress is outlined in the next section.

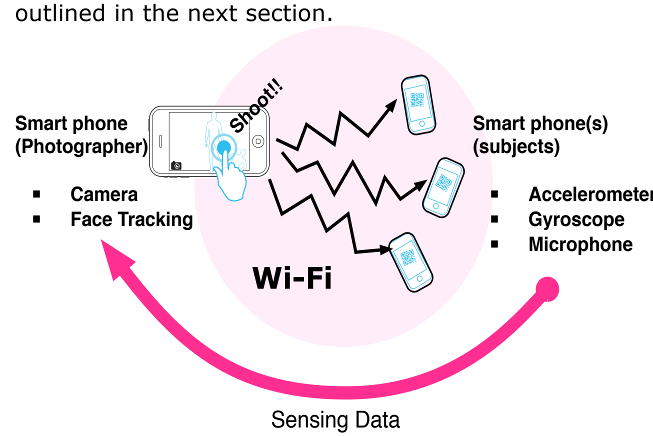


Figure 2: Smartphones are connected to each other

via Wi-Fi. The shooting event and the sensing data are broadcast over Wi-Fi.

Prototype: Camera App with Sensor Network

As the first step toward appending Manga iconographies onto a photo using data from the sensors embedded in smartphones, we developed a camera application for Android smartphones. Figure 2 shows the configuration of the application. The camera application has two modes—camera mode and sensor node mode—between which users can dynamically switch. When the user starts the application, the mode is set to sensor node mode. In sensor node mode, the mode is changed to camera mode by tapping the camera button on the display. Detailed descriptions of the two modes are given below.

Camera Mode

While the application is in camera mode, the user can take a photo by tapping the screen on the smartphone. In order to associate a subject's movement with the sensor data from his/her smartphone, the video is recorded for 10 s after s/he taps the screen. The resolution of the video is 1280 × 720 px. In addition, when the user taps the screen, the shutter notification (tapping event) is broadcast over the Wi-Fi network as a UDP packet. Thus, other smartphones are able to receive the notification.

Sensor Node Mode

When the application is in sensor node mode, the smartphone operates as a network of sensors. When the application receives the shutter notification, the smartphone acquires sensor data from a three-axis accelerometer, a three-axis gyroscope, and a microphone for 10 s. The sampling rates are 200 Hz for both the accelerometer and the gyroscope and 16 kHz for the microphone. The acquired data is then sent to the smartphone that sent the shutter notification as a TCP packet via Wi-Fi.

This prototype was implemented on a Samsung Galaxy Nexus smartphone. Using this prototype, we attempted to associate a subject with his/her smartphone. We were able to capture the subject's movements performed outside of the view of the camera and to create Manga iconographies from the movements.

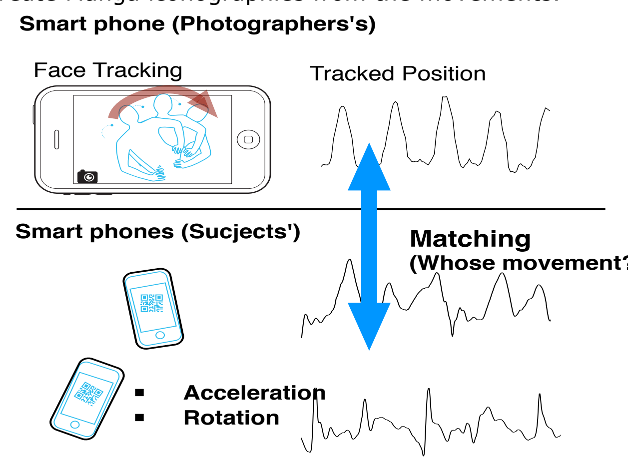
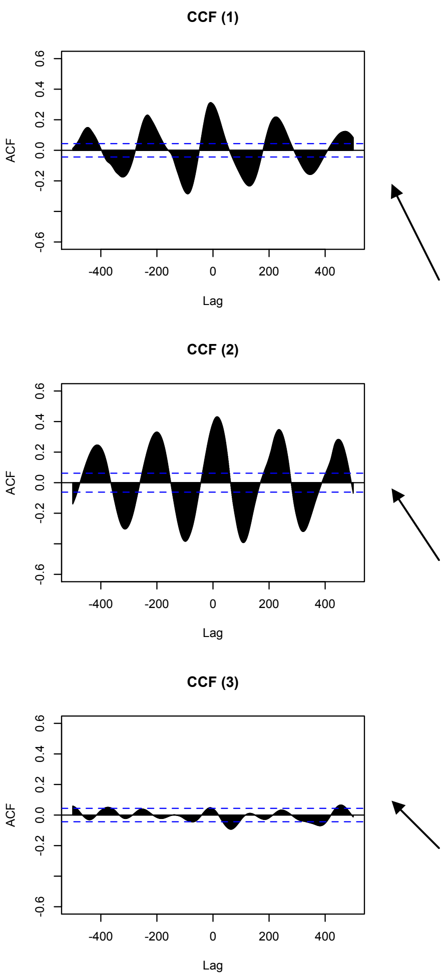


Figure 3: Configuration of the feasibility study. The photographer's smartphone tracks the subject's face. The others send sensing data.

Feasibility Study

We conducted a feasibility study using our prototype. As shown in Figure 3, the aim of the feasibility study was to associate the subject on camera with his/her smartphone. One healthy 29 year-old male participated in the feasibility study. We conducted six trials in which his movements were recorded using the prototype in camera mode. Each trial had a recording time of 10 s. A smartphone set to sensor node mode was placed in the left back pocket of the participant's pant, while another smartphone set to camera mode was fixed on the table. A white projector screen was used as the background of the scene for stable face tracking.



Cross Correlation Functions of each face tracking position and acceleration (y-axis) of the subject's smartphone.

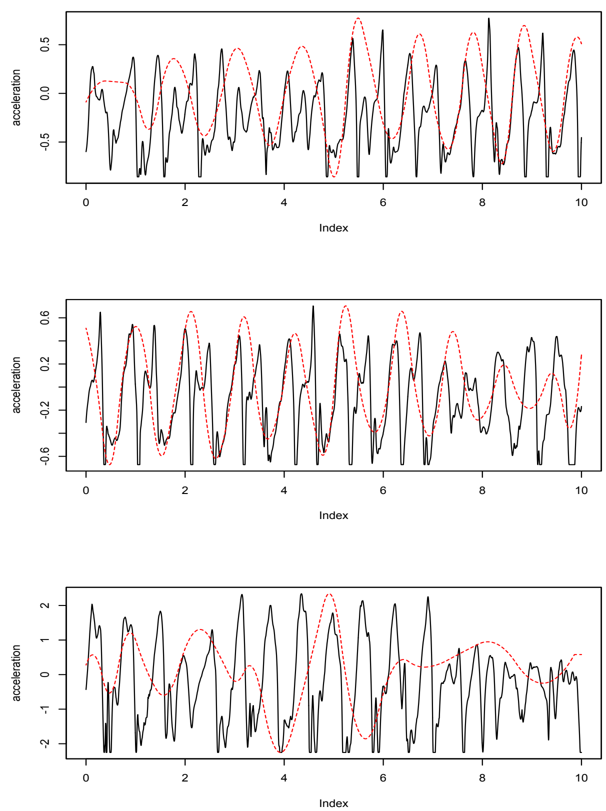


Figure 4: Time series data for three trials. The black lines show the acceleration of the smartphones. The red dotted lines show the face tracking data.

Face tracking technique

We processed six videos taken by the smartphones in camera mode for face tracking of the subject in post-processing activities. We employed the Continuously Adaptive Mean Shift (Camshift) [9] for face tracking and implemented it on the OpenFrameworks platform in the C++ language. We stored the tracked face positions as time-series at a sampling rate of 20 Hz.

Data Association

The time-series data for the face positions acquired from the result of Camshift was associated with the time-series data for the acceleration of the smartphone that was in sensor node mode as follows:

1. First, we associated the acceleration data with the face tracking data by calculating the difference between the time-series data of the face tracking at time t and the same data at time $t+1$.

2. We then interpolated the time-series data for the face tracking data, which was obtained at 20 Hz, to 200 Hz using spline interpolation.

Figure 4 shows the time-series data for three of the six trials conducted. The black lines show the acceleration of the smartphone, while the red lines show the pre-processed time-series data for the face tracking data. We calculated Cross Correlation Functions (CCFs) between the two time-series data for each trial. The left side of Figure 4 shows the corresponding CCFs for each trial. Except for the last trial, the CCFs all had high coefficients of correlation ($r > 0.4$). Because it failed to track the position of the face, the last trial had a poor correlation between the two time-series data. The tracking failure was only found in the last trial, however.

Five of the six trials had high coefficients of correlation (r = 0.394 on average, n = 5). For comparison, we calculated three CCFs between the time-series data for acceleration on the first, third, and fifth trials and the time-series data for face tracking on the second, fourth, and sixth, respectively. These coefficients of correlation were nearly zero ($r < 0.1$) at lag = 0. These results show that it is possible to associate a subject on camera with his/her smartphone by calculating the CCF between the time-series data for acceleration and the time-series of the face tracking data.

Concluding Remarks

In this paper, we proposed an approach that increases both the enjoyment in a photo session and the amount of information of a photo by appending Manga iconographies using smartphones. We developed a prototype application for smartphones that utilizes sensor networks on Wi-Fi to record videos that are associated with subjects' movements. We also conducted a feasibility study to associate a subject with his/her smartphones using our prototype. The results of our feasibility study indicate that it is possible to associate a subject that is in the view of the camera with his/her smartphone by calculating the CCF between the time-series data for acceleration of the smartphone and the time-series of the face tracking.

We plan to develop a robust method that associates subjects with their smartphones and a method that creates Manga iconographies from subjects' movements in the future.

References

[1] D. DeMenthon, V. Kobla, and D. Doermann. 1998. Video summarization by curve simplification. In *Proc. of the sixth ACM international conference on Multimedia* (MULTIMEDIA '98), 211-218.

[2] S. Uchihashi, J. Foote, A. Girgensohn, and J. Boreczky. 1999. Video Manga: generating semantically meaningful video summaries. In *Proc. of the seventh ACM international conference on Multimedia (Part 1)* (MULTIMEDIA '99), 383-392.

[3] C. Nguyen, Y. Niu, and F. Liu. 2012. Video summagator: an interface for video summarization and navigation. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12), 647-650.

[4] M. Masuch, S. Schlechtweg, and R. Schulz. 1999. Speedlines: depicting motion in motionless pictures. In *ACM SIGGRAPH 99 Conference abstracts and applications* (SIGGRAPH '99), 277-.

[5] B. Kim and I. Essa. 2005. Video-based nonphotorealistic and expressive illustration of motion. In *Proc. of the Computer Graphics International 2005* (CGI '05), 32-35.

[6] E. P. Bennett and L. McMillan. 2007. Computational time-lapse video. *ACM Trans. Graph.*26, 3, Article 102 (July 2007).

[7] W. Hwang, P. Lee, B. Chun, D. Ryu, H. Cho. 2006. Cinema comics: Cartoon generation from video stream. In Proc. of the First International Conference on Computer Graphics Theory and Applications (GRAPP 2006), 299-304.

[8] M. Rofouei, A. Wilson, A.J. Brush, and S. Tansley. 2012. Your phone or mine?: fusing body, touch and device sensing for multi-user device-display interaction. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12), 1915-1918.

[9] G. R. Bradski. 1998. Computer vision face tracking as a component of a perceptual user interface. In *Proc. of IEEE Workshop on Applications of Computer Vision*, 214–219.