

Augmented Reality Services of Photos and Videos from Filming Sites Using Their Shooting Locations and Attitudes

Yoon-Seop Chang*, Seong-Ho Lee*, Joonmyun Cho*

*ETRI (Electronics and Telecommunications Research Institute), Republic of Korea

ychang76@etri.re.kr, sholee@etri.re.kr, jmcho@etri.re.kr

Abstract— Usually lots of photos and videos are recorded additionally to capture the scenes of many filming sites of dramas and movies. When people visit the sites later, however, they can rarely experience these contents again. The objective of this study is to exploit these photos and videos from filming sites together with their shooting locations and attitudes in order to provide users with AR (Augmented Reality) tourism services. Users will be able to experience as if they are at the site just at the time of filming. We designed and made an external attachable device for DSLR (Digital Single-Lens Reflex) cameras to log shooting locations and attitudes. We also implemented syncing software and defined data formats to integrate these data with photos and videos. We implemented mobile AR tourism apps for iOS and Android platforms. We adopted FAST (Features from Accelerated Segment Test) algorithm with additional modification for AR visualization of photos in mobile environment. We adopted Ferns algorithm, histogram specification and OpenGL texture mapping for AR visualization of videos with better performance. These apps and the results of this study were evaluated at actual filming sites of several dramas, and the survey of many filming staffs was conducted for feedbacks. The results of this study could be applied many other fields concerned with multimedia contents.

Keywords— Photo, Video, Location, Attitude, Filming Site, Augmented Reality

I. INTRODUCTION

All photos and videos have their own spatial circumstances, i.e. shooting locations and attitudes such as longitude, latitudes, azimuth, tilt, roll, field of view when they are taken. If these shooting location and attitude data can be acquired and integrated with photos and videos, they will be useful to solve many problems in various fields.

Usually lots of photos and videos are recorded additionally to capture the scenes of actors, staffs and the site itself at filming sites of dramas and movies. When people visit the sites later, however, they can rarely experience these contents again. If these photos and videos are used together with their shooting locations and attitudes, AR tourism services could be provided for past filming sites [1]. Users will be able to experience as if they are at the site just at the time of filming. The objective of this paper is to exploit these photos and

videos from filming sites with their shooting locations and attitudes to provide AR tourism services.

The rest of this paper is organized as follows. Some related studies are reviewed in section 2. Section 3 describes our approach and methods to solve problems. Section 4 presents results of this study. We conclude with a discussion of future works in section 5.

II. RELATED WORKS

There were studies which tried to improve existing methods to classify and search using shooting location and attitudes. Epshtein et al classified photos based on contents using not only camera positions but also view frustums [2]. Lee and Rhee improved the accuracy of feature based search of photos using locations and attitude data together [3]. Hao et al analysed FOVs (Field of Views) of videos to extract hotspots and visualize key frames on maps [4-6].

There were also studies adopting SfM (Structure from Motion) to photos from internet and crowdsourcing. These studies reconstructed 3D spaces around cities and attractions, explore through those spaces, and estimated individual photos locations. Most well-known example is 'Photo Tourism' by Snavely et al [7]. Frahm et al tried to improve their accuracy using geotags and panorama images together [8].

There were studies using Google Street View panorama images [9-13]. They estimated shooting locations and attitudes of input photos using locations and attitudes of panorama images around input photos. They could estimate 3D positions of POIs in panorama images. They also could reconstruct 3D spaces and estimate locations.

III. APPROACH AND METHODS

A. Attachable External Device

Shooting location and attitude data should be measured and logged to be integrated with photos and videos. Nowadays mobile devices such as smartphones and tablets usually have various useful built-in sensors. Many of these sensors are designed to measure locations and attitudes of mobile devices. GPS (Global Positioning System), gyroscope, magnetometer and accelerometer are examples of them.

DSLR cameras are commonly used at filming sites rather than mobile devices. DSLR cameras are used to take not only photos and also videos with full-HD resolution. For example, canon 6D cameras are popular at filming sites of many dramas in South Korea. DSLR cameras, however, are not equipped with sensors mentioned above. So shooting locations and attitudes are not able to be acquired and integrated with photos or videos only using these cameras. In this study, we designed and made an attachable small device which logs shooting locations and attitudes to assist DSLR cameras without sensors.

GPS and AHRS modules were embed in the device. GlobalTop GMS-G9 module was adopted as a GPS module which measured locations with the accuracy of 2.5 m CEP (Circular Error Probability). MEMS (Microelectromechanical Systems) based module, InvenSense MPU-9150 was adopted as an AHRS (Attitude and Heading Reference System) module. MPU-9150 is a 9-axis motion tracking module which combines a 3-axis gyroscope, 3-axis accelerometer and a 3-axis digital compass. We could acquire shooting attitudes with the accuracy of 1.5 degree using this module.

Figure 1 shows the appearance and its parts of the device made in this study. It could be attached to the camera tripod hole and the antenna was located slightly over the camera.

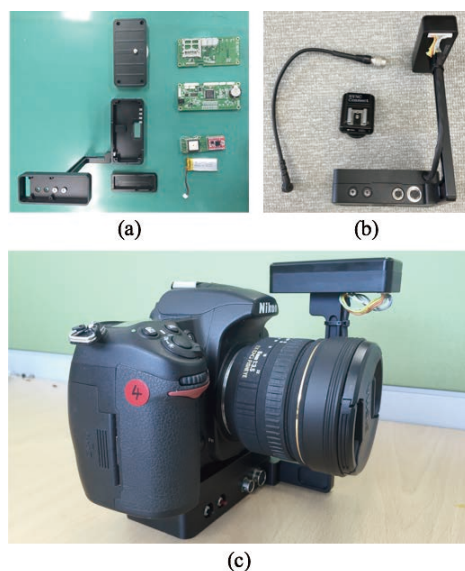


Figure 1. (a) Boards and the frame, (b) the assembled device, and (c) the attached device to the Canon 6D DSLR camera

B. Data Acquisition and Synchronization

Figure 2 shows the flow to measure and log shooting location and attitude data. Data were stored to the SD card in the attached device. When photos were taken, i.e. the shutter was clicked, shooting locations and attitudes were logged synchronously using the camera's strobe signals. For videos, we used the IR (Infrared) signal to synchronize with the start and end time of video recording because Canon 6D's video shooting didn't trigger strobe signals. Shooting locations and attitudes were logged for every frame of videos. As a result,

we could acquire shooting locations and attitudes successfully whenever we took photos and videos.

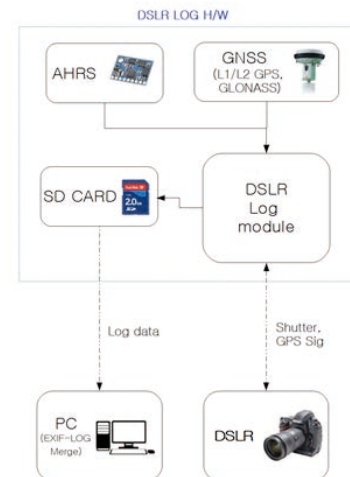


Figure 2. The flow to measure and log shooting location and attitude data

We defined data formats to store and exchange shooting location and attitude data as shown in Figure 3. Formats of logs from the attached data were identical for photos and videos except their logging frequencies. Shooting location and attitude data were converted to JSON (JavaScript Object Notation) formats. JSON data includes GPS location, AHRS altitude and camera sensor information. These were stored into Exif user comment tags for photos and individual files for videos. Data formats for videos store shooting locations and attitudes for all frames.

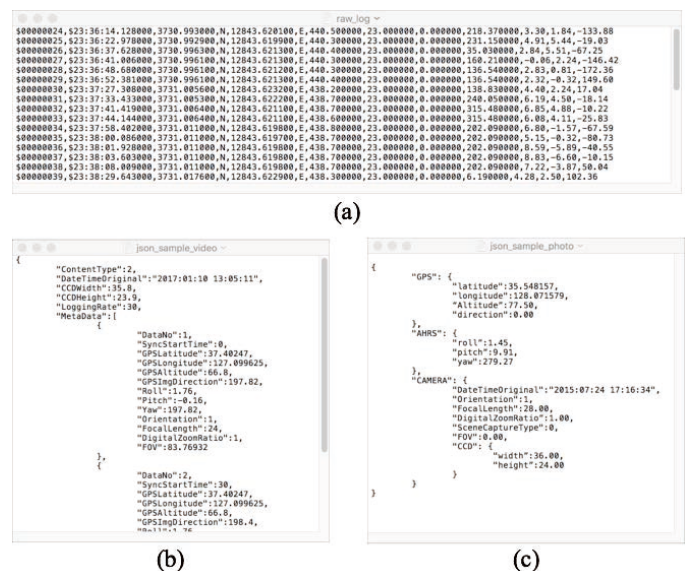


Figure 3. (a) Raw data logs, (b) json data for photos, and (c) for videos

C. Augmented Reality Visualization

When users, i.e. tourists, visit filming site later, the mobile app in this study can visualize photos and videos which were taken before at the time of filming. These photos and videos

were registered and visualized on the real time screen of mobile devices.

Figure 4 shows the flow of AR visualization of photos from filming sites. Photo images were resized and high pass filtered to improve the performance during feature points detection and matching in mobile environment. Feature points were detected using FAST algorithm with non-maximum suppression option and threshold value 40. Number of feature points were limited up to 500 and 300 each for reference and input images. Input images are real time images on the screen of mobile devices. On other hand, reference images are those which were taken before at time of filming. Reference images were registered and visualized over the input images in real time.

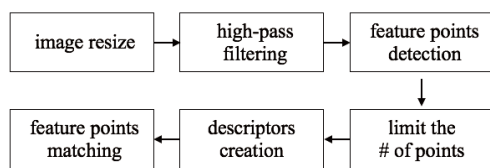


Figure 4. The flow of AR visualization of photos from filming sites

BRIEF (Binary Robust Independent Elementary Features) descriptors were created for feature points, then feature points matching were proceed comparing these descriptors. One of disadvantages of FAST algorithm is that it does not include orientation invariant information of images. We calculated descriptors while rotating reference images by 45 degrees to complement this disadvantage as shown in Figure 5. Then feature points matching were carried out using OpenCV DescriptorMatcher with BruteFore-Hamming algorithm.

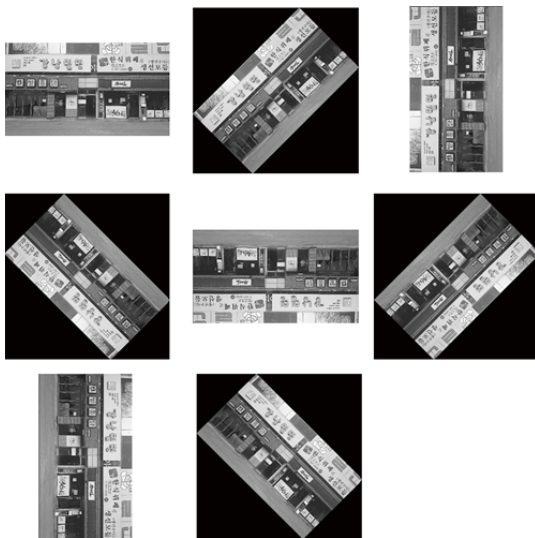


Figure 5. Calculate descriptors for feature points rotating by 45 degree

We adopted Ferns algorithm, histogram specification and OpenGL texture mapping technique to improve the performance of AR visualization of videos on mobile devices. Ferns algorithm proceed the learning of rotation and scale invariant information of images for each feature point. Then it

deals with feature points matching as a classification problem. Feature points matching using Ferns algorithm is faster than other existing algorithm such as SIFT (Scale Invariant Feature Transform) and SURF (Speed-Up Robust Features), so it is rather suitable for usage on mobile devices. One of disadvantages of Ferns algorithm is that it takes much time during learning process. We coped with this problem by proceeding the learning process in offline. In order to render videos more fast on mobile devices, we adopted OpenGL texture mapping rather than dealing with video frames directly. Histogram specification technique diminish problems from the variation of weather and illumination condition between input and reference images. It enhanced the histogram of input images by using histogram information of reference images.

The flow for AR visualization of videos from filming sites is like shown in Figure 6. Learning process of Ferns algorithm was proceeded in offline. Input videos were enhanced through histogram specification for feature point detection. Feature points were detected up to 200 points. Feature points matching were proceeded using Ferns classification algorithm. If feature matching was successful, tracking process was followed by. While tracking was successful, videos were registered and rendered in the appropriate region on the screen using OpenGL texture mapping.

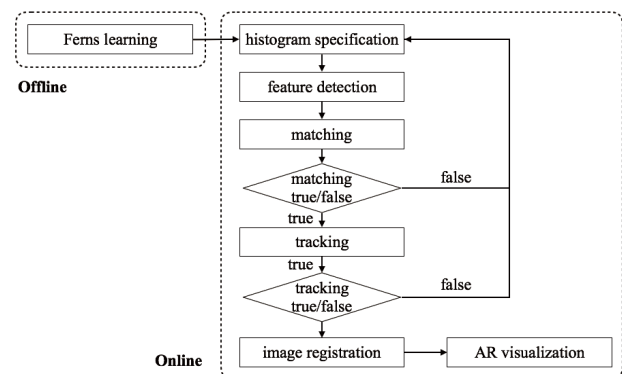


Figure 6. The flow of AR visualization of videos from filming sites

Figure 7 shows the learning process in this study using Fern algorithm. Frame images were extracted from videos and user set ROI (Region of Interest) on the selected image for learning. Feature points were detected within the region. Learning were proceeded rotating patches by $-45 \sim 45$ degree and changing the scale by $0.5 \sim 1.5$. Learned result data were stored on the server and sent to clients, i.e. mobile devices, at runtime. Histogram data for the selected frame were also stored together for later histogram specification process.

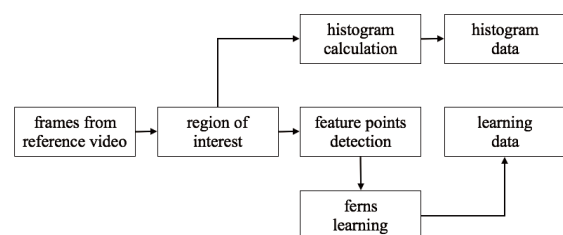


Figure 7. The flow of leaning process based using Ferns algorithm

IV. RESULTS

A. Visit to Filming Sites of dramas

We visited actual filming sites of three dramas including a famous Korean drama 'Uncontrollably Fond'. Photos and videos were taken for tests and demonstrations using a Cannon 6D DSLR camera with our attachable device. Photos and videos captured scenes of filming sites such as actors, actresses, staffs, filming sets and equipment as shown in Figure 8. These photos and videos were used as service contents for users, i.e. tourists, who would visit the site again with interest.

**Figure 8.** Visit to filming sites of several dramas for data acquisition

B. Mobile AR Tourism Apps

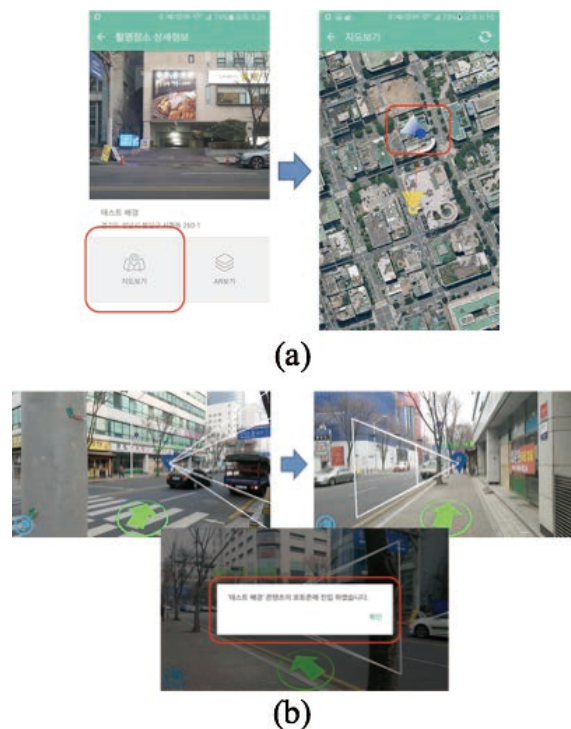
Photos and videos from filming sites were visualized in forms of AR on mobile devices. Mobile AR tourism apps were implemented on the Android and iOS platforms for demonstration. Figure 9 shows the screen captures of AR visualization of photos and videos in mobile apps. These photos and videos had been taken before at the time of filming, and were augmented and visualized over the real time screen of mobile devices. So users could feel as if they were there at that time as a filming staff.

Users were guided to viewpoints where they could experience AR visualization of photos and videos. Users could be guided in 2D map and AR modes as shown in Figure 10. It was possible by analysing locations and attitudes of mobile devices, photos and videos. If users were within the appropriate range of viewpoints then mobile apps popped alarms to inform it.

C. Tests and Evaluation

Several tests were carried out in advance to determine parameters for optimal performance of AR visualization on mobile devices, especially for video data. As a result, mobile AR tourism apps in this study showed the performance of image registration error within 10 pixels and rendering speed up to 24 fps for video data.

The survey of about 40 filming staffs, mainly program directors, working in the field was conducted by interviews to evaluate the usability and potential of this study. We got many positive responses from the survey.

**Figure 9.** The screen captures of AR visualization of photos and videos at filming sites**Figure 10.** Users are guided to viewpoints in (a) 2D map and (b) AR modes

V. CONCLUSIONS

In this study, we exploited photos and videos from filming sites with their shooting locations and attitudes to provide AR tourism services. We designed and made an attachable small device which logs shooting locations and attitudes to assist cameras without sensors. We also defined data formats to store and exchange shooting location and attitude data. We adopted FAST algorithm with additional modification for AR visualization of photos, and adopted Ferns algorithm, histogram specification and OpenGL texture mapping for AR visualization of videos on mobile devices. We visited actual filming sites of three dramas, and Photos and videos were taken for tests and demonstrations. As a result, mobile AR tourism apps were implemented on the Android and iOS platforms. Several tests were carried out in advance to optimize the performance of AR visualization. The survey of filming staffs in the field was conducted to evaluate the usability and potential of this study. The results of this study could be applied many other fields concerned with multimedia contents. Further works on leveraging geospatial information for improving the accuracy of AR visualization need to be continued.

ACKNOWLEDGMENT

This research is supported by Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Contents Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2017.

REFERENCES

- [1] Y. S. Chang, S. H. Lee, Y. H. Seo, and J. Cho, "Integrating shooting location and posture information into photos and videos from film sites of movies and dramas," in *Proc. International Symposium on Mobile Mapping Technology*, 2015.
- [2] B. Epshtein, E. Ofek, Y. Wexler, and P. Zhang, "Hierarchical photo organization using geo-relevance," in *Proc. Annual ACM International Symposium on Advances in Geographic Information Systems*, 2007.
- [3] Y.H. Lee and S.B. Rhee, "Efficient Photo Image Retrieval System Based on Combination of Smart Sensing and Visual Descriptor," *Intelligent Automation & Soft Computing*, Vol. 21, pp. 39-50, 2015.
- [4] J. Hao, G. Wang, B. Seo, and R. Zimmarmann, "Keyframe Presentation for Browsing of User-generated Videos on Map Interfaces," in *Proc. International Conference on Multimedia*, 2001, pp. 1013-1016.
- [5] B. Seo, J. Hao, and G. Wang, "Sensor-rich video exploration on a map interface," in *Proc. ACM International Conference on Multimedia*, 2011, pp. 791-792.
- [6] J. Hao, G. Wang, B. Seo, and R. Zimmarmann, "Point of Interest Detection and Visual Distance Estimation for Sensor-Rich Video," *IEEE Transactions on Multimedia*, Vol. 16, pp.1929-1941, 2014.
- [7] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: Exploring Photo Collections in 3D," *ACM Transactions on Graphics*, Vol. 25, pp. 835-846, 2006.
- [8] J. Frahm, J. Heinly, E. Zheng, E. Dunn, P. Fite-Gergel, and M. Pollefeys, "Geo-registered 3D models from crowdsourced image collections," *Geo-spatial Information Science*, Vol. 16, pp. 55-60, 2013.
- [9] P. Gronat, M. Havlena, J. Sivic, and T. Pajdla, *Building streetview datasets for place recognition and city reconstruction*, Research Reports of CMP, Czech Technical University in Prague, 2011.
- [10] M. Park, J. Luo, R. T. Collins, and Y. Liu, "Beyond GPS: Determining the Camera Viewing Direction of A Geotagged Image," in *Proc. International Conference on Multimedia*, 2010.
- [11] V. Tsai and C. Chang, "Feature positioning on Google street view panoramas," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 305-309, 2012.
- [12] S. Verstockt, M. Gerke, and N. Kerle, "Geolocalization of Crowdsourced Images for 3-D Modeling of City Points of Interest," *IEEE Geosciences and Remote Sensing Letters*, Vol. 12, pp. 1670-1674, 2015.
- [13] K. Lee, S. Lee, W. J. Jung, and K. T. Kim, "Fast and Accurate Visual Place Recognition Using Street-View Images," *ETRI Journal*, Vol. 39, pp. 97-107, 2017.



Yoon-Seop Chang received his BS, MS, and PhD degrees in Geographic Information System from Seoul National University, South Korea, in 1999, 2001 and 2005, respectively. He joined ETRI (Electronics and Telecommunications Research Institute), South Korea, in 2005 and is currently working as a principal researcher. Since 2008, he has also been a faculty member of University of Science and Technology, South Korea, as an associate professor. His research interests include geographic information system, web mashup, augmented reality and virtual reality.



Seong-Ho Lee received the BS and MS in computer science from Chungbuk National University, South Korea, in 1997 and 2000, respectively. Since 2000, he has been a senior member of research staff with ETRI, South Korea, and he is also working toward the PhD in computer science Chungbuk National University. He is currently working in Location-based Smart Content Platform project as a senior researcher. His research interests are spatio-temporal database systems, geographic information systems, and location-based services.



Joonmyun Cho received his B.S., M.S. and Ph.D. degrees in mechanical engineering from KAIST (Korea Advanced Institute of Science and Technology) in 1993, 1995 and 2006, respectively. He joined ETRI (Electronics and Telecommunications Research Institute), South Korea in 2007 and was involved with the URC (Ubiquitous Robotic Companion) project until 2011 and Beyond Smart TV project until 2015. He is currently working in Intelligent IoT SW Platform project as a senior researcher. His research interests include knowledge based systems, intelligent agent systems and machine learning.