Augmented Reality Services based on Embedded Metadata

Byoung-Dai Lee

Department of Computer Science, Kyonggi University, Suwon, Korea
blee@kgu.ac.kr

Abstract. Feature extraction and tracking is one of the core AR technologies. In particular, the markerless AR provides natural synthesis of a real-world and virtual objects, as it is able to identify objects directly within the video and obtain relevant information. However, the downside of the markerless AR is that it may not be appropriate for resource-constrained devices such as mobile phones, due to considerable amount of computations. In this paper, we propose a method to address the problem by putting into multimedia content the metadata necessary to provide AR services, such as virtual object information and supplementary information required for on-screen display.

Keywords: Augmented Reality, Feature Extraction and Tracking, ISO Base Media File, Metadata.

1 Introduction

Augmented Reality (AR) is a technology that provides augmented information services by synthesizing real-time image/voice and virtual objects or supplementary information. Recently, mobile devices with various built-in sensors such as cameras and GPS have been widely distributed, presenting diverse convergence services using high-speed mobile Internet and rapidly spreading mobile AR services.

Most of the existing AR services use real-time image recognition results to provide virtual object or supplementary information. That is, a played video is analyzed in real-time and the area where virtual object will be rendered is identified. However, technology that extracts features by accurately recognizing the object within the image on a real-time basis requires considerable amount of computations, which indicates that the quality of AR services depends on the complexity of feature extraction algorithms as well as the resource capability of the device. Along with the difficulty of extracting features, there is another weakness in the existing AR services: tightly-coupledness of AR application programs and the augmented information shown to the users. For instance, in the case of a service that shows a corporate logo image in the middle of a video to promote a product, when the logo will be presented to the users is determined by the logic of the application program. Thus, there is a possibility that a corporate logo unrelated to the actual video content that is being played appears on screen, resulting in a decreased advertising effect.

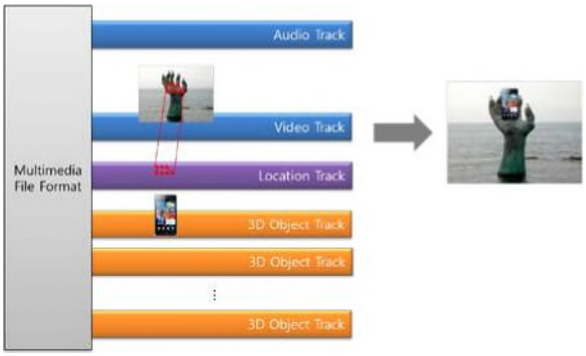

Fig. 1. Logical structure of the proposed media file with embedded metadata for AR services.

AR in multimedia services based on stored media is different from AR based on general real-time videos in that the editing of multimedia content by the service provider can be preceded. Therefore, this paper aims to solve the aforementioned problems by putting into multimedia content the metadata necessary to provide AR services, such as virtual object information and supplementary information required for on-screen display. In particular, this paper proposes a method to construct media files as the container that includes only the metadata to display AR contents on screen, so that the AR is not tightly coupled with certain AR technologies. Constructing media files by putting in metadata, as proposed by this paper, offers the following advantages. (1) Complicated processing is not required to extract features in a receiving device, thereby enabling easy use on a mobile device, which is typically resource constrained. (2) Deterioration of multimedia contents can be prevented by determining in advance the most suitable location from each scene of the video in which the virtual object will be displayed. (3) The image and virtual object are not tightly coupled mutually and can provide the AR services most appropriate for the user context (e.g., user location, performance of device).

2 Metadata for Augmented Reality

Fig. 1 shows the logical structure of the media files proposed in this paper. The media files include Audio/Video Track as well as Location Track, storing the location information that displays the virtual object provided in AR, and 3D Object Track, storing the actual virtual object information to be displayed at the relevant location.

The main role of Location Track is to save the location information of the space that displays the virtual object within the image. In particular, for natural synthesis between the image and virtual object, the virtual object must sequentially move with the image based on the time scale when the image is played. For the purpose, this paper defined AR region on the time scale based on the rate of movement of the virtual object. Fig. 2 depicts the example of defining AR regions on the time scale based on the rate of movement of the virtual object.

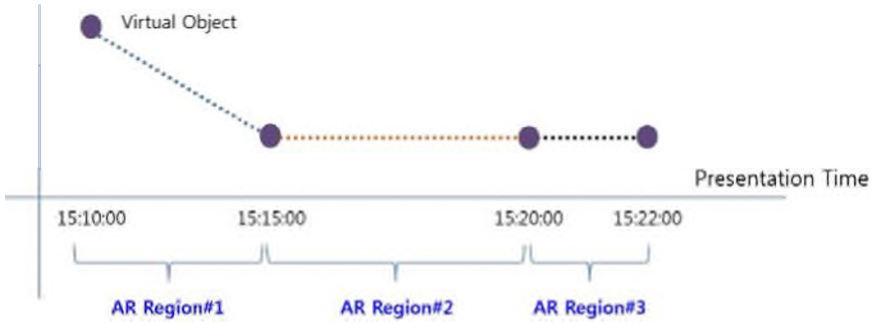

Fig. 2. An example of Location Track.

According to Fig. 2, the virtual object is displayed on screen from 15:10 to 15:22 after the image began playing and moves from the upper-left to lower-right corner of the screen. In AR Region #1 (15:10:00-15:15:00), it moves from the upper-left to the lower-right corner at a constant rate (e.g., α m/sec); in AR Region #2 (15:15:00-15:20:00), it moves from the lower-left to the lower-right corner at the same rate as in AR Region #1, However, in AR Region #3 (15:20:00-15:22:00), the traffic line of the object is the same as that of AR Region #2 (that is, from the lower left to the lower right), whereas the rate of movement is different (e.g., β m/sec); thus, it is defined as a different AR region, and the location information of the object is to be saved. As stated above, AR regions save the location information of the object that moves in a straight line at an equal speed based on the time scale. Therefore, as the range of AR region is narrower, the superimposition of the virtual object and image becomes softer.

3D Object Track describes the actual virtual object to be displayed at the location within the image clarified in Location Track. Multiple 3D Object Tracks can exist in the media files to support various AR contents, and the 3D Object Track most suitable for user conditions is selected by using reference data on the virtual object included in Location Track. Another characteristic of 3D Object Track is that it provides neutrality in certain representation techniques of virtual objects. Virtual objects used in AR can be represented using various techniques. However, 3D Object Track is not tightly coupled with certain representation, but it plays the role of a container regardless of the internal representation, therefore providing neutrality in the representation technique of virtual objects.

3 Conclusion

Feature extraction and tracking, which is one of the core AR technologies, is critical in determining the area on screen in which the virtual object is to be displayed, and it is one of the most difficult fields of study. Currently, many related technologies exist, but they have limitations for use on resource-constrained mobile devices, because they require significant amount of computations in most cases. Moreover, a great number of computations ultimately require the consumption of many batteries, making such technologies more difficult to apply to mobile devices. In this paper, we proposed a method to construct AR-enabled media files by putting into multimedia content the metadata necessary to provide AR services, such as virtual object information and supplementary information required for on-screen display.

Acknowledgments. This work was supported by Kyonggi University Research Grant 2012.

References

1. Klein, G., Murray, D., Parallel Tracking and Mapping for Small AR Workspace, Proc. of the 6th IEEE/ACM Int'l Symposium on Mixed and Augmented Reality (2007)
2. Lee, W., Woo, W., Real-time Color Correction for Marker-based Augmented Reality Applications, Proc. of Int'l Workshop on Ubiquitous Virtual Reality (2009)
3. Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T., Schmalstieg, D., Pose Tracking from Natural Features on Mobile Phones, Proc. of the 7th IEEE/ACM Int'l Symposium on Mixed and Augmented Reality (2008)
4. ARhrrr!, http://ael.gatech.edu/lab/research/hanheld-ar/arhrrr/
5. ETSI 3GPP TS 26.44, Transparent End-to-End Packet Switched Streaming Service (PSS): 3GPP File Format (3GP), The 3rd Generation Partnership Project (2009)
6. ISO/IEC 14496-12:2008(E), Information Technology-Coding of Audio-Visual Objects-Part 12: ISO Base Media File Format, ISO/IEC (2008)
7. ISO/IEC 14496-14, Information Technology-Coding of Audio-Visual Objects-Part 14: MP4 File Format, ISO/IEC (2003)