Augmented Reality Technology Empowering Mobile News Clients: A Preliminary Exploration (Postprint)

Authors: Li Lei

Date: 2023-10-08T00:00:00+00:00

Abstract

In September 2020, the General Office of the CPC Central Committee and the General Office of the State Council issued the "Opinions on Accelerating the In-depth Integrated Development of Media", which explicitly stipulates leveraging advanced technology to guide and propel integrated development. This paper explores how augmented reality technology can be utilized to empower mobile news clients, thereby achieving the refinement, interactivity, and immersion of news content, and facilitating the in-depth integration of media.

Full Text

A Preliminary Exploration of Empowering Mobile News Clients through Augmented Reality Technology

Li Lei
Tianmu Media Co., Ltd., Hangzhou, Zhejiang 310039

Abstract

In September 2020, the General Office of the CPC Central Committee and the General Office of the State Council issued the "Opinions on Accelerating the In-depth Integrated Development of Media," which explicitly called for using advanced technology to lead and drive integrated development. This paper explores how augmented reality technology can empower mobile news clients to achieve refined, interactive, and immersive news content, thereby facilitating deep media integration.

Keywords: Augmented Reality (AR); virtual scenes; immersive experience; deep media integration

Augmented Reality (AR) technology integrates virtual scenes with real-world environments [1]. Its origins trace back to the 1950s and 1960s when American photographer and inventor Morton Heilig, drawing on his filmmaking experience, created the "Sensorama Stimulator," a machine that used imagery, sound, and vibration to immerse users in the sensation of riding a motorcycle at high speed. This invention marked the beginning of AR' s development trajectory. Today, AR technology is characterized by virtual-real fusion, real-time interaction, and environmental perception, with 3D tracking and 3D registration being its most critical features. Currently, AR is widely applied in gaming, cultural tourism, healthcare, industry, and military contexts.

Several years ago, the globally popular mobile game *Pokémon GO* introduced the general public to the wonders of AR on personal mobile devices and familiarized them with the concept of augmented reality. In this context, AR refers to the visual enhancement of reality through virtual content, displayed via smartphone screens or specialized wearable devices such as Microsoft HoloLens, Magic Leap, and Google Glass. The key to AR technology lies in the device' s ability to perceive and understand its surroundings. Fundamentally, this involves determining the device' s spatial position; more advanced capabilities include real-time environmental reconstruction (SLAM), while the most sophisticated level encompasses recognition, cognition, and interaction. Positioning is foundational—only when the device' s location is precisely determined can virtual content be seamlessly and realistically integrated with the real scene in real time. AR primarily enhances rather than completely virtualizes reality, embedding virtual objects within real-world scenes and anchoring them in three-dimensional space such that they remain stationary regardless of viewer movement.

1. AR Technology Principles

Currently, AR requires external devices such as smartphones or professional AR equipment like glasses and headsets to function. These devices capture real-world scenes through cameras and overlay virtual objects onto the imagery [2]. AR technology can be categorized into two major approaches based on its implementation principles.

1.1 Computer Vision-Based AR

Computer vision-based AR establishes a mapping relationship between real-world scenes and screens (such as smartphone displays or AR glasses) through CV algorithms, enabling virtual 3D scenes to be superimposed onto reproduced real-world imagery. This approach can be further divided into two types.

The first is Marker-Based AR, which requires a predefined marker—such as a business card or poster—placed at any location in the real scene. This marker establishes a planar reference point in the physical environment. When a smartphone or AR glasses camera recognizes the marker, CV algorithms use the marker' s center as the origin of a template coordinate system, creating a mapping between this template coordinate system and the screen coordinate system. Based on this mapping, virtual 3D scenes displayed on the screen appear to attach to the marker.

[Figure 1: see original paper] Marker-Based AR Implementation Principle

The second approach, Marker-Less AR, operates without requiring specially prepared markers. Instead, it utilizes any object with sufficient feature points—such as human bodies, faces, gestures, or vehicles—as reference points. Machine learning algorithms (e.g., SURF, ORB, FERN) extract and memorize these object features. When the camera scans a scene, it extracts feature points from objects in the environment and compares them with the memorized features. If the number of matching feature points meets or exceeds a predetermined threshold, the system recognizes the object and establishes a mapping relationship, allowing virtual scenes to attach to various objects. This is how three-dimensional facial animation stickers are implemented in beauty camera applications.

1.2 LBS-Based AR

Location-Based Service AR determines the user' s geographic coordinates, orientation, and device tilt angle through mobile device sensors including GPS, compass, and gyroscope. It then retrieves information about nearby points of interest (POIs)—such as scenic spots, schools, restaurants, and stations—from data sources like online maps. Using this information, the system establishes planar references for target objects in the real scene and creates mapping relationships through coordinate system transformations, enabling the combination and superposition of virtual and real scenes on screen. This approach leverages mobile device sensors, eliminating dependence on fixed markers and avoiding real-time marker recognition and feature point calculation, thereby delivering better system performance and user experience on mobile devices compared to Marker-Based and Marker-Less AR.

2. Mainstream AR Technical Solutions

A complete mobile AR technical solution primarily comprises four business modules: AR engine (SDK), AR content creation, AR content recognition, and AR content management. The AR engine is the most critical component, determining whether the AR system can overlay virtual scenes onto real space, perform SLAM (motion tracking and localization), estimate lighting and shadows, and intelligently analyze and cognize real scenes through AI technologies such as facial recognition, image recognition, and motion recognition to achieve augmented reality effects.

Most AR engines are built upon Google ARCore and Apple ARKit, which support Android and iOS devices respectively. These foundations are further encapsulated and extended to adapt to more mobile devices and application systems. The mainstream AR engines include:

(1) **ARCore**: Developed by Google, this AR engine is exclusively for Android platform development and does not support iOS devices. It primarily implements motion tracking, plane tracking, point clouds, cloud anchors, lighting estimation, environment probes, facial tracking, 2D image tracking, and human occlusion.

(2) **ARKit**: Developed by Apple, this engine exclusively supports iOS devices for AR application development on iPhone and iPad platforms. Like ARCore, it enables motion tracking, 3D object tracking, plane tracking, point clouds, cloud anchors, lighting estimation, environment probes, facial tracking, and motion capture [3].

(3) **Vuforia**: Developed by Qualcomm, this engine incorporates both ARCore and ARKit, supports Unity3D-created virtual content, and works across Android and iOS devices. The business data generated by this engine can be deployed in cloud environments.

(4) **EasyAR**: Developed by SightPlus Information Technology, this engine includes ARCore and ARKit support, accommodates Unity3D-created content, and features a proprietary SLAM system for various application scenarios. The complete AR system supports both local and cloud deployment.

(5) **Ali AR**: Developed by Alibaba, this engine focuses on 2D image recognition and 3D object tracking, commonly applied in Alipay' s annual "Scan the Fu" campaign.

(6) **LandMarkAR**: Developed by ByteDance, this engine is primarily used within the Douyin app to add AR special effects to city landmarks in short videos.

In summary, mainstream AR engine solutions each possess distinct capabilities while facing certain adaptation and application scenario limitations. For mobile clients, selecting a complete AR technical solution should consider: (1) engine universality and device compatibility—simultaneous support for both ARKit and ARCore platforms, compatibility with most Android and iOS devices, and Unity support for gaming requirements; (2) functional richness—support for surface tracking, planar image tracking, 3D object tracking, motion tracking, facial tracking, lighting estimation, cloud anchors, and point clouds; (3) system integration ease—SDK-based integration into mobile applications or web-based implementation for scanning recognition images and objects to present 3D animated models, videos, images, text, UI buttons, and support 3D model interaction; (4) content management and distribution—support for cloud storage, processing, and review of AR files, plus CDN distribution; and (5) AI extensibility—intelligent analysis and cognition of real scenes and algorithm training for expressions, gestures, body postures, etc., to achieve corresponding augmented displays.

Using the Tianmu News client as an example, after thorough technical research and considering content positioning and operational requirements, the EasyAR comprehensive technical solution was adopted. This solution comprises four main components: AR engine (mobile SDK), content scripting production system, cloud recognition management system (Cloud Recognition Service), and cloud content management system (Operation Center).

The AR engine (mobile SDK) supports both iOS and Android platforms, adapting to mainstream device models on the market with particularly friendly adaptation for numerous Android models. The engine' s primary functions include: (1) planar image recognition and tracking—real-time recognition and tracking of textured planar objects such as books, business cards, or even graffiti walls, dynamically generating tracking targets from standard images while recognizing and tracking multiple targets simultaneously; (2) 3D object recognition and tracking—real-time recognition and tracking of richly textured 3D objects in natural scenes, accommodating various shapes and structures with simultaneous multi-object recognition; (3) dense map spatial localization and tracking—using smartphone or AR glasses cameras to perform 3D dense reconstruction of the surrounding environment to obtain dense point cloud and mesh maps, enabling virtual objects to better integrate with real environments for proper occlusion and collision effects; (4) sparse map spatial localization and tracking—using cameras to scan surroundings and construct 3D environmental point clouds, where each 3D point records local visual information, ultimately generating a 3D visual map of the environment for visual localization and tracking, suitable for developing persistent AR applications or multi-user interactive AR experiences; and (5) motion tracking—continuously tracking the mobile device' s position and orientation in space, aligning virtual and real scenes in real-time within the same coordinate system for AR displays, games, videos, or photography applications, creating an experience where virtual and real scenes merge. For persistent AR applications, this can be combined with EasyAR' s sparse spatial map localization and tracking.

Beyond these core functions, the engine provides multiple programming language interfaces, particularly native development-friendly Java and Objective-C interfaces, supports video recording, includes a built-in video player with H.264 and H.265 hardware decoding, supports mainstream 3D scene creation tools such as 3ds Max, Maya, and Unity, and implements multiple compression algorithms for mobile network conditions to enhance user experience.

The content scripting production system supports online content editing and updates, including model editors, scene editors, and script editors. It supports universal JavaScript scripting, enabling everything from simple model displays to complex mini-games through the script editor. Its greatest advantage is that AR content updates require only content modification without client repackaging and redistribution, avoiding cumbersome release processes.

The Cloud Recognition Service sends image information as recognition requests to the cloud, where servers retrieve matching target images from associated galleries and invoke the engine to load 3D virtual models for rendering corresponding AR effects. The system supports multiple AR triggering methods including planar target recognition, gesture recognition, and pose recognition. It comprises recognition, monitoring, and statistics subsystems, with the recognition system being the core component of CRS, including Targeter and Searcher systems for managing recognition target APIs and a CV feature-based retrieval system called Retriever. Additionally, it provides monitoring and statistical services for scan hit analysis and active AR scanning device tracking.

[Figure 2: see original paper] System Deployment Diagram (Relationship between CRS System and Dependent Systems)

The cloud content management system (Operation Center) provides storage management, recognition image management, version management, content distribution, and CDN acceleration for AR content packages compatible with the SDK. It works in coordination with the cloud recognition management system to offer a complete cloud private deployment solution for users with content security management requirements.

[Figure 3: see original paper] System Deployment Diagram (Relationship between OC System and Dependent Systems)

3. AR Technology for Content Integration Innovation

The "Opinions on Accelerating the In-depth Integrated Development of Media" issued by the General Office of the CPC Central Committee and the General Office of the State Council in September 2020 explicitly stated that advanced technology should lead integrated development, urging the utilization of information technology achievements such as 5G, big data, cloud computing, Internet of Things, blockchain, and artificial intelligence, and strengthening forward-looking research and application of new technologies in news communication [4].

AR technology brings an unprecedented sense of immersion to the converged media era by combining virtual content with real environments. This approach expands news content information volume and enhances user interaction, enabling users to directly perceive and access news scenes. This represents the most distinctive feature of AR news content.

AR news delivers strong interactivity and enhances news 趣味性 (interest/engagement). While text engages users visually and television engages both hearing and vision, AR technology directly engages multiple senses including sight, hearing, and touch during content delivery. This innovation realizes a multi-sensory 联动 (coordinated) perception of content in news communication, immersing users in news scenes to experience the "most authentic" sense of presence. Although AR news provides partially virtual scenes, the multi-sensory perception it evokes is profoundly real. Furthermore, AR-empowered content leverages its strong interactivity to increase engagement through gamification, making younger users more willing to actively consume AR news and thereby enhancing content dissemination power.

AR news content is no longer a one- or two-dimensional narrative but restores news scenes in three- or even four-dimensional ways. Rather than retelling news scenes and event details, AR news directly transmits the news scene to users, allowing them to experience and feel the scene firsthand. This creates understanding and sensation of news events that is far more authentic and immediate than traditional news reporting.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.