Building Intelligent Video Capabilities for Short-Video News Clients in the 5G Era: Postprint

Authors: Li Lei

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

Abstract

With the advent of the 5G era, short videos have become the "focal point" of mobile internet, effectively capturing users' fragmented time. Short videos are gradually emerging as an important form of content expression in news client applications. This paper explores how to construct intelligent video capabilities and empower the deep integration of media through "short video+" artificial intelligence technology.

Full Text

Preamble

Title: How Short Video News Clients Can Build Intelligent Video Capabilities in the 5G Era
Author: Zhejiang Online Technology Center, Hangzhou, Zhejiang 310039

Abstract: With the advent of the 5G era, short videos have become the "trend" of mobile internet, firmly capturing users' fragmented time. Short videos are gradually emerging as a crucial form of content presentation for news clients. This paper explores how to build intelligent video capabilities and empower deep media convergence through the integration of "short video +" artificial intelligence technologies.

Keywords: 5G; artificial intelligence; short video; news client; media convergence
Classification: G212
Document Code: A
Article ID: 1671-0134(2021)02-036-04
DOI: 10.19483/j.cnki.11-4653/n.2021.02.007

Citation Format: Li Lei. How Short Video News Clients Can Build Intelligent

Video Capabilities in the 5G Era [J]. China Media Technology, 2021(02): 36-38+110.

1. Current Status of Short Video News Clients

In recent years, short video social platforms such as Douyin, Kuaishou, and Weishi have developed rapidly, while information-based short videos have also risen quickly. According to QuestMobile' s "China Mobile Internet 2020 Semi-Annual Report,"the monthly active user scale of the short video industry reached 852 million in June 2020 [1]. As audiovisual technology advances and 5G networks achieve widespread coverage, short video content under one minute has firmly captured users' fragmented time in the high-speed mobile internet era, keeping users engaged for extended periods and becoming a phenomenal "screen-dominating" product. The content is characterized by distinct personas, brevity, and engaging qualities that resonate emotionally with users, directly triggering strong interactivity and high user stickiness.

Compared with traditional reporting formats such as text, images, and broadcast television, short video news innovates by integrating visual and auditory elements, ushering video into the "seconds" era. Its fragmented, mobile, and social features cater to users' content consumption habits of fragmentation, shallow reading, and strong interaction in the mobile internet age, while delivering the most exciting, emotionally expressive, and resonant content directly to users. As media convergence develops in depth, with the goal of building "four-all media," short videos have gradually become an important means of communication for information content. Short video news serves as a key entry point for traditional media to achieve deep convergence, providing new driving force for media transformation and upgrading.

Given this trend in content dissemination, major media organizations have launched short video news services and channels. In October 2016, The Beijing News launched the "We Video" project, and in November of the same year, the Pear Video client went live, forming the early prototype of information-based short video platforms. In the past two years, with further promotion of deep media convergence, both central and local media have invested significant resources to build short video clients. Examples include the "People' s Daily +" client, a mainstream short video PUGC aggregation platform launched by People' s Daily, the "Yangshipin" client, a comprehensive audiovisual new media flagship platform launched by China Media Group, and "Tianmu News," a short video news client built by Zhejiang Online to serve the national strategy for integrated development of the Yangtze River Delta. These developments mark short video news clients as a new competitive highland in media convergence and an important track for media innovation.

Against this backdrop, how to creatively build intelligent video capabilities for short video news clients has become a significant subject requiring exploration and practice.

2. Building Intelligent Video Production Capabilities

Current mainstream short video news clients operate under two production models: PGC (Professionally Generated Content) and UGC (User Generated Content). These two content production scenarios require different supporting capabilities. From a research and development perspective, the first consideration is the applicability of content production capabilities, followed by the reusability of capabilities across both scenarios to avoid redundant development and reduce R&D costs.

By integrating artificial intelligence, big data, 5G, VR, AR, and MR technologies to empower short video news clients, an "AI+" technology ecosystem can be formed to build intelligent audiovisual capabilities in the following key areas:

2.1 Intelligent Converged Media Asset System

Under converged media requirements, there is a rigid demand for using rich media resources across platforms, media types, and terminals. It is necessary to leverage cloud computing and artificial intelligence to build a cloud-based intelligent media asset system that integrates content aggregation, storage management, processing, channel distribution, and analysis mining. This system mainly includes the following capabilities:

2.1.1 Intelligent Media Asset Storage Based on cloud computing' s elastic scaling capabilities, powerful middleware, and rich database services, the underlying infrastructure for converged media assets can be guaranteed. The system supports storage of various media data types, accommodates multiple data sources, enables scenario-based management, and employs different storage types to support diverse business needs. For example, production and distribution content media assets use standard storage in object storage, which provides highly reliable, highly available, and high-performance object storage services that support frequent data access. This is suitable for business scenarios such as social and sharing-based images and audiovisual applications. Additionally, the system provides multiple storage types including infrequent access, archival storage, and cold archival storage, which not only support high-frequency, high-concurrency access scenarios in mobile internet but also reduce usage costs.

2.1.2 Intelligent Media Asset Processing Built upon cloud storage, this capability performs a series of multimedia data processing operations on audiovisual content in the cloud, transcoding it into formats suitable for playback across all platforms. It must support common internet audiovisual encoding capabilities such as H.264 and H.265 encoding, and support overlaying images and text watermarks on output videos to enhance product recognition. It can separate audio or video from video files independently. For long videos, it supports parallel transcoding of video segments, which can significantly improve transcoding speed. It can capture JPG format images from video files stored

in object storage at specified times, supporting both single and multiple screenshots. By understanding video content and combining it with visual aesthetics, it can select optimal keyframes as video cover images or intelligently extract the most representative set of screenshots from video content to form a GIF as a video summary. It extracts sound, image, and temporal features from videos to generate video fingerprints, enabling functions such as video fragment traceability, which can be applied to video deduplication, infringing video filtering, and original video protection.

[Figure 1: see original paper]

2.1.3 Intelligent Structured Management Audiovisual content consists of unstructured files, which present challenges for querying, analyzing, and mining during management and reuse. Traditional approaches require substantial human resources for cataloging management, often with multi-level cataloging requirements. Since manual operations are time-consuming, labor-intensive, and produce cataloging results that vary due to different operators' understanding of content, a better approach is needed. Based on deep learning, computer vision technology, and massive datasets, videos can be analyzed across multiple modalities including content, text, speech, and scenes. First, it can automatically output news "5W" element tags, video classification, and other multi-dimensional content tags, applicable to personalized recommendation and video search scenarios. Second, it recognizes faces in videos and supports facial keypoint positioning, facial attribute analysis, and rapid face clustering, applicable to machine editing, desensitization risk control, character correlation, and knowledge graph scenarios. Third, it converts audio to text and extracts keywords, which serve as sources for tag content and dimensions for desensitization risk control. Finally, OCR detection identifies text in multimedia data, accurately recognizing subtitles, titles, bullet comments, and other key content in video frames.

2.1.4 Intelligent Risk Control Management In the converged media era, the dissemination of massive rich media audiovisual content means traditional text-based review systems can no longer meet the demand for reviewing vast amounts of content in the mobile internet age. Through a combination of AI machine review and manual review, content review labor costs can be effectively reduced while providing higher technical guarantees for audio, image, and video content security. This approach simultaneously improves review efficiency and accuracy. This capability mainly provides multi-modal content risk control for politically sensitive, pornographic, terrorist, violent content, as well as for characters, scenes, and objects. Particularly for UGC content production, where data volume is large and growing rapidly, manual review is slow and costly, posing significant compliance risks. Content security through AI deep learning algorithms can automatically and intelligently identify non-compliant content, substantially reducing labor costs, improving review efficiency, and effectively meeting risk control management requirements.

2.1.5 PCDN Acceleration Based on P2P technology, this capability builds a low-cost, high-quality content distribution network service by mining and utilizing massive fragmented idle resources at the edge network. It is suitable for video on-demand, live streaming, and other business scenarios. After integrating the PCDN SDK, it can significantly improve distribution quality compared with ordinary CDN, providing basic network guarantees for achieving "instant loading" effects for videos while also reducing distribution costs to a certain extent. In terms of security protection mechanisms, it employs encryption and authentication systems for anti-hotlinking and DDoS attack defense, uses high-strength encryption for node caching to prevent content tampering, and ensures content remains under control [2].

2.2 Intelligent Short Video Production Platform

Utilizing artificial intelligence and computer vision technology, this cloud-based online production platform integrates material management, online editing, post-production packaging, rendering export, and publishing. The system performs structured processing of video materials in the cloud, extracting metadata from videos for automated tagging. Specifically, it analyzes uploaded videos through scene classification, character recognition, speech recognition, and text recognition to form hierarchical and refined classification tags, thereby enabling precise search of video materials. AI technologies such as intelligent effects, intelligent subtitles, and intelligent speech enhance online video production efficiency. Even script text can be quickly transformed into short videos through preset production templates based on NLP. The commonly used cloud-based rapid editing function allows editors to quickly clip videos, add transition effects, rapidly add subtitles through AI speech-to-text conversion, utilize CV technology for virtual anchor dubbing, and render and export with one-click publishing to clients and various video platforms. Simultaneously, the platform can integrate 5G live streaming feeds in real time, enabling visual operations to slice live videos into short videos for immediate output and publication through timestamp marking during broadcast. In summary, the online collaborative production approach for editorial teams significantly shortens the entire video production workflow, substantially lowering the threshold and cost of video content production. AI empowerment of video production achieves breakthroughs in video processing productivity, transforming traditional video production workflows.

[Figure 2: see original paper]

3. 5G Mobile Cloud Live Streaming System

Under large-scale commercial deployment of 5G networks, the high speed, low latency, and universal connectivity characteristics of 5G provide basic network guarantees for mobile live streaming to enter an era of high definition, strong interactivity, and full scenarios. First, under 5G high-speed conditions, high video bitrates can be achieved, enabling high-quality signal transmission based

on high-definition picture quality, thereby meeting users' increasingly high visual experience demands for converged media live streaming. Second, the low-latency characteristic of 5G enables "simultaneous occurrence" between news scenes and user experience, enhancing users' desire for interaction with media and among users themselves. Third, 5G enables universal connectivity, allowing everything from professional cameras, drones, 360° VR gimbals, traffic security monitors to any mobile phone to serve as stable live streaming signal sources, thereby achieving full-scenario, immersive live streaming content.

In summary, the complete 5G mobile cloud live streaming system mainly includes the following components: (1) Live project management: supporting live project creation, streaming address allocation, broadcast time setting, and live cover image production; (2) Live material management: centralized management of various materials used during live broadcasts to enable rapid search and quick deployment; (3) Real-time directing management: performing directorial switching of multiple live signal streams to achieve cross-spatiotemporal, multi-camera, multi-scene picture management; (4) Replay file management: viewing, downloading, archiving, replay replacement, replay segment selection, and rapid editing of replay files; (5) Real-time risk control management: implementing broadcast control of live content while reducing manual content identification and improving intelligent identification and early warning capabilities; and (6) Data statistics management: analyzing collected data to achieve multi-dimensional user behavior analysis and output analytical reports.

[Figure 3: see original paper]

4. Mobile Video Application Capabilities

For short video client users, the most direct experience is video playback. Whether in on-demand or live streaming scenarios, the video player' s loading method, interactive operations, video clarity, playback stability, and streaming file decoding compatibility all relate to user experience quality. Second, short video clients with UGC scenarios must provide short video shooting, production, and special effects functions to enhance the fun of user short video creation and strengthen user social experience.

4.1 Video Player

The mobile player must support the following technical requirements: (1) Support for mainstream video formats including MP4, M3U8, FLV, and MKV, MP3 audio format, H.264 and H.265 hardware video decoding, and AAC audio encoding, with iOS support for AC3 audio encoding, providing seamless switching of multi-bitrate HLS; (2) Playback control functions including start, end, pause, resume, replay, and loop playback, with support for both on-demand and live streaming functions and network video playback via URL; (3) Support for switching between multiple clarity streams for on-demand and transcoded content, providing live time-shifted video stream playback and video caching while

playing, suitable for short video loop playback scenarios; (4) Support for first-frame instant loading for both on-demand and live streaming, dynamic frame chasing for live streaming to reduce latency, automatic live reconnection, and quick dragging without clearing buffered content; (5) Support for personalized data collection based on player user behavior information; and (6) Support for real-time metric monitoring, root cause analysis, and full-link problem tracking capabilities.

The short video shooting function must include the following basic features: (1) Real-time watermarking, camera switching, resolution setting, real-time audio mixing, and speed adjustment; (2) Support for selecting videos from albums, cropping by duration and frame, and importing and splicing multiple videos, multiple photos, or mixed photo-video content with configurable transition modes and durations; (3) Provision of advanced beauty, makeup, micro-shaping, body shaping, and gesture recognition AR capabilities based on face and human body CV algorithms; and (4) Support for adding filters, subtitles, dynamic/static stickers, background music, and doodles in the editing interface.

[Figure 4: see original paper]

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