Real-time Metadata Extraction from UGC Video

Marcus Thaler, Andras Horti, Albert Hofmann, Werner Bailer
JOANNEUM RESEARCH, DIGITAL
Steyrergasse 17, 8010 Graz, Austria
{firstname.lastname}@joanneum.at

Wolfram Hofmeister, Jameson Steiner, Reinhard Grandl
Bitmovin
Lakeside B01, 9020 Klagenfurt, Austria
{firstname.lastname}@bitmovin.com

ABSTRACT
User generated content (UGC) is a valuable source for improving the coverage of events, such as concerts, festivals or sports events. When using UGC in live productions, content needs to be checked for quality, and metadata captured by the mobile device and extracted from the content are relevant for filtering the UGC streams. We demonstrate a system for capturing live audio and video streams on a mobile device, performing automatic metadata extraction in real-time and indexing the metadata for access by a production system. The system receives an audio, video and metadata stream from the mobile device, and creates additional metadata from the audiovisual content. All metadata are provided as a stream, indexed in a metadata store, and visualised in an HTML5-based user interface.

Author Keywords
user generated content, mobile, content analysis, live

ACM Classification Keywords
H.5.1 Information Interfaces and Presentation: Multimedia Information Systems

INTRODUCTION
User generated content (UGC) is a valuable source for improving the coverage of events, such as concerts, festivals or sports events. In order to integrate user generated content into existing production workflows, both the quality of UGC needs to be checked and metadata needs to be extracted. Such metadata, together with sensor information from the mobile device, will help the production team to assess the context and relevance of the user contribution. Recently, apps like Meerkat[1] or Periscope[2] have raised the interest for live UGC. However its integration in production systems is still challenging. In particular, live UGC needs to be filtered in order not to overwhelm the production team. Quality is one important criteria for filtering, and the location of the recording is another. Due to the time constraints, it is not feasible to perform the filtering task manually, but one must rely on automatic tools. The metadata needed by these tools must either be captured with the content (e.g. using the sensors of the mobile device), or extracted from the content.

In this paper we demonstrate a system for capturing live audio and video streams on a mobile device, performing automatic metadata extraction in real-time and indexing the metadata for access by a production system. The system receives an audio, video and metadata stream from the mobile device, and creates additional metadata from the audiovisual content. All metadata is available as a stream (with low latency from the extraction), and is indexed in a metadata store. Metadata needed in the real-time process can be read directly from the stream, and earlier metadata can be queried from the store. We decided to build on an existing framework with many standard components and able to handle the decoding of commonly used media formats. The GStreamer[3] open source multimedia framework was chosen for this purpose.

The rest of this paper is organised as follows. Section Analysis system describes a system for capturing content and metadata, and performing quality analysis on the mobile device and further analysis on the server side. The visualisation of the resulting metadata is discussed in Section Data visualisation followed by a description of the demonstration setup.

ANALYSIS SYSTEM
The system consists of a dedicated capture app, which sends video, audio and metadata as separate streams. This saves the muxing/demuxing effort and also facilitates processing different modalities on different machines in the cloud. All data are provided as RTP streams. The processing system performs the necessary decoding and transformation for the content, and also includes a set of interconnected analysis modules. These modules may only use the content as input, but may also use metadata from the device or from other modules. All extracted metadata are provided as streams again, and a logging module listens to these streams and indexes data in the metadata store. The audiovisual streams can be connected to viewers or to an editing system. A web-based metadata viewer displays the audiovisual data together with the extracted metadata.

The integrated capture application for Android enables users to perform quality analysis while capturing sensor data and streaming captured video. The main features are: (a) audio and video recording, via the built-in microphone and camera respectively, (b) metadata capturing from different sensors available on the device, (c) on-device analysis of captured essence to meet quality constrains, (d) en-/transcoding and packaging of recorded content and (e) the up-streaming functionality to servers for processing. Raw video and audio data is captured through the camera and microphone of the device and encoded using Android's *MediaCodec API*, while at the same time, the quality of video frames is analysed. Once the encoding for a frame has finished it is committed into the

[1] https://meerkatapp.co
[2] https://www.periscope.tv
[3] http://gstreamer.freedesktop.org

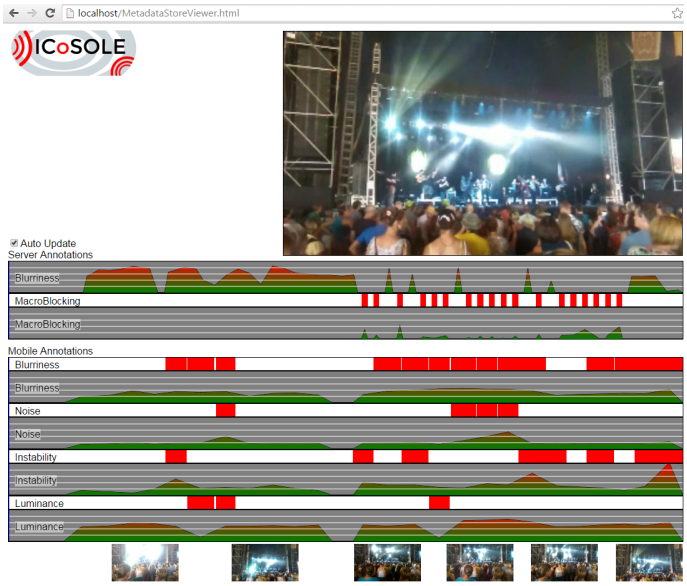

Figure 1. Web-based content and metadata visualisation.

buffer and/or sent to an RTP packager. Synchronization is done by keeping track of the latest PTS for each stream.

The server-side processing system is based on GStreamer, using existing decoding modules, and implementing new modules for metadata extraction, synchronisation and metadata handling. The metadata extraction functionality is provided by existing content and quality analysis algorithms, which have been integrated into the real-time framework. The functionalities include cut detection, key frame detection, detection of MPEG-7 visual descriptors, sharpness estimation and macroblocking detection. A GStreamer module at the end of the processing chain consumes all metadata streams and indexes data in the metadata store.

DATA VISUALISATION

The data exchange between the analysis platform and the operator's backend is realised via a metadata store. This metadata store is a persistent repository accessible over a REST interface. To support fast in- and output transactions the database is a Redis[4] in-memory data structure store. Metadata for the current time window is kept in Redis for efficient insert and querying of recent metadata. For live productions, all required metadata can be kept in this store.

The extracted metadata are used for automatic selection from the UGC streams, e.g., discarding streams based on quality metadata or their location. By querying the metadata store with the appropriate criteria, the relevant streams can be selected and provided to the live editing system for review by the editor.

The visualisation of a live video stream and time-aligned metadata is done by the HTML5 metadata viewer (see Figure 1). The metadata store is polled in defined intervals for recent data and the UI is updated accordingly. Both quality annotations done at the mobile client and server are shown. For each annotation type, a chart with the continuous quality measure is shown, and an additional event view to spot segments that do not meet predefined quality standards. Since the metadata store demonstrator is implemented as HTML5 viewer the incoming media stream is re-streamed by the analysis platform. This can be done as RTP stream with very low latency (requiring a plugin) or providing a stream for consumption by an HTML video player, with possibly higher latency.

[4] http://redis.io

Figure 2. Screenshot of the capture app, including quality feedback.

DEMO SETUP

For the demonstration, mobile devices running the integrated capture and streaming application for Android are provided. The application package is also available for download to enable users to perform quality analysis while capturing sensor data and streaming captured video using their own devices. The application continuously measures sharpness, noise, luminance, exposure and detects the use of brightness compensation before streaming captured video. For each quality measure an overlay including a related icon and message is displayed to immediately notify the user, if the quality of the captured content is not within the expected limits (see Figure 2). In this way, users can interactively experiment on how various factors contribute to objective quality degradations by getting immediate feedback on the device. For example, by shaking the mobile phone or generating bad lighting conditions visual quality impairments can be forced. For visualisation of the live video stream and the time-aligned quality measures a HTML5 metadata viewer is used.

ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 610370, ICoSOLE ("Immersive Coverage of Spatially Outspread Live Events", http://www.icosole.eu).