Monitoring of User Generated Video Broadcasting Services

Denny Stohr
Distributed Multimedia Systems
Technische Universität Darmstadt, Germany
dstohr@cs.tu-darmstadt.de

Stefan Wilk
Distributed Multimedia Systems
Technische Universität Darmstadt, Germany
swilk@cs.tu-darmstadt.de

Wolfgang Effelsberg
Distributed Multimedia Systems
Technische Universität Darmstadt, Germany
effelsberg@cs.tu-darmstadt.de

ABSTRACT

Mobile video broadcasting services offer users the opportunity to instantly share content from their mobile handhelds to a large audience over the Internet. However, existing data caps in cellular network contracts and limitations in their upload capabilities restrict the adoption of mobile video broadcasting services. Additionally, the quality of those video streams is often reduced by the lack of skills of recording users and the technical limitations of the video capturing devices. Our research focuses on large-scale events that attract dozens of users to record video in parallel. In many cases, available network infrastructure is not capable to upload all video streams in parallel. To make decisions on how to appropriately transmit those video streams, a suitable monitoring of the video generation process is required. For this scenario, a measurement framework is proposed that allows Internet-scale mobile broadcasting services to deliver samples in an optimized way. Our framework architecture analyzes three zones for effectively monitoring user-generated video. Besides classical Quality of Service metrics on the network state, video quality indicators and additional auxiliary sensor information is gathered. Aim of this framework is an efficient coordination of devices and their uploads based on the currently observed system state.

Categories and Subject Descriptors

C.2.3 [**COMPUTER-COMMUNICATION NETWORKS**]: Network Operations—*Network monitoring; Network management*; I.4.8 [**IMAGE PROCESSING AND COMPUTER VISION**]: Scene Analysis—*Sensor fusion*; C.2.4 [**COMPUTER-COMMUNICATION NETWORKS**]: Distributed Systems—*Client/server; Distributed applications; Network operating systems; Distributed Systems*

WISMM'14, November 7, 2014, Orlando, FL, USA.

ACM 978-1-4503-3157-9/14/11 ...$15.00.
http://dx.doi.org/10.1145/2661714.2661726.

Keywords

Network monitoring, Measurement, Video composition, Mix, Cellular networks, Mobile, Video broadcast

1. INTRODUCTION

Large, real world events motivate dozens to record video using their mobile recording devices. At the same time, there is a huge interest in watching these events remotely. As an example, a politically motivated demonstration has attracted thousands of viewers on the mobile broadcasting platforms bambuser[1], uStream[2] and JustinTv[3] as shown in Figure 1.

Many of the users recording videos with mobile devices would like to instantly share their video recordings to meet the demand for such live content. However, high quality content can not be shared instantly as cellular networks suffer from limited upload capacities, especially when multiple users upload their recordings in parallel. Also, data caps in cellular network contracts limit the possible amount of such mobile video broadcasting services. To efficiently broadcast video in such large-scale events with multiple recorders at the same time, is a resource-intensive task that involves intelligent allocation of uptime and thus upload capacities.

In this paper, we propose a monitoring and metric framework and distribution system for such mobile video broadcasting services. The proposed component-based architecture allows measurements in large-scale environments. The remainder of this work is structured as follows. First, we will introduce a classification of factors that influence the QoE (Quality of Experience) for live streaming of user-generated video. Next, we briefly introduce a director service that manages multiple input streams for broadcasting as well as a monitoring framework that collects and distributes parameters related to the discussed QoE factors. Last, we discuss relevant publications in the field of user-generated content and conclude our work.

2. PREDICTION OF QUALITY OF EXPERIENCE

The main goals of the proposed monitoring framework are (1) to provide the basis for an improved QoE for the users of the mobile video broadcasting service and (2) improve

[1]`http://bambuser.com/`
[2]`http://ustream.tv/`
[3]`http://justin.tv/`

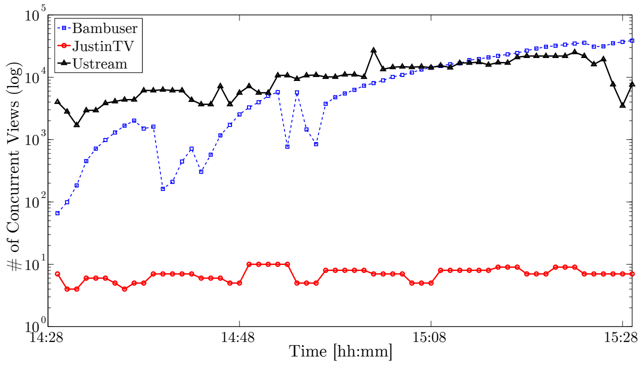

Figure 1: Aggregated concurrent views of mobile video recordings from smart phones broadcast from Ukraine, Odessa on the 4th May 2014, grouped by platform.

the efficiency of collaborative creation of an high quality live video stream. We focus on identifying and measuring aspects that affect the QoE during creation of content and aim at supporting an automatic reduction of parallel video streams, e.g. by the composition of one single video. Thus, effects degrading the QoE such as stalling or dropped video segments are commonly investigated video streaming systems. This work concentrates on effects of uploading video to the CaOS (Composition and Orchestration Service) server.

We define four components that provide indicators for the predicted QoE during the creation of a collaborative video stream: (1) Sensor information (2) Network status (3) Video quality metrics (4) Content relevance. These metrics differ in terms of their source, computational complexity and relevance. As illustrated in Figure 2, sensor information

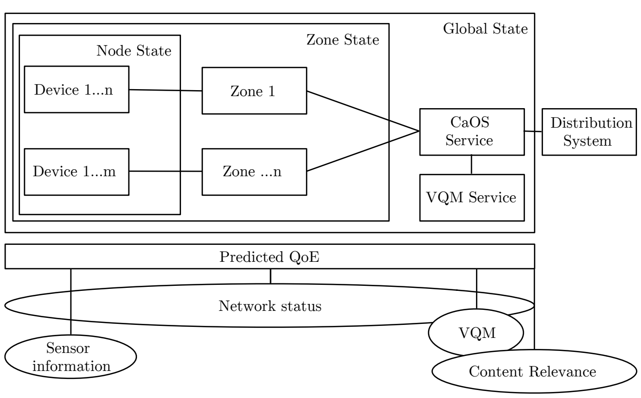

Figure 2: Architecture of the monitoring framework for the user generated video broadcasting service

is solely measured on the mobile device during recording. This information is used to annotate the generated video stream with the motion, environmental information such as the position of the device[4]. Based on this data, the detection of user behavior that leads to degraded video quality, such as camera shaking or wrong orientation of the phone, can also be observed directly on the users' device without analyzing (or transmitting) the video data itself [5]. In previous work [10], we found that degradations such as camera shakes have a huge impact on the QoE.

In contrast, the network status investigates the transmission from end devices to the receiving CaOS server, spanning across all components of the system. By observing the distribution networks metrics like bandwidth, jitter and package loss it is possible to evaluate factors that influence the quality loss during the transfer of the recorded video [2].

Next, no-reference video quality metrics - such as by Keimel et al. [7] - are provided by a VQM (Video Quality Metric) service. As this process requires higher processing power, processing on mobile devices would be unfeasible or leading to an undesirably high energy consumption, therefore this service is ideally being hosted on a central server.

Last, the content relevance describes the abstract measure of the relevance for a given video stream. This can be based on high-level knowledge of CaOS, e.g., the GPS location of the video providers, or information tracking the user engagement in the distribution system. In [11] the relevance of a recording position on the attractiveness of a user-generated video is being illustrated.

Gathered information as illustrated above is taken into account by the CaOS service when selecting which views to use for the generated live video. This process will be discussed in the following section.

2.1 Composition and Orchestration Service

Relevant recorded video streams are gathered at the central CaOS server, as is would be infeasible to directly distribute the content to thousands of viewers from resource-capped mobile phones. As shown in Figure 1, parallel recording of the same event occurs regularly. Our aim is to improve the QoE for both the recorders as well as the viewers by providing a composition of different incoming streams to a single, composed video mix. CaOS is creating this composition on the basis of feature extraction from the video representing: compression artifacts such as blocking and temporal distortions such as frame drops as provided from the monitoring framework. We aggregate this information in the so-called Video Quality Metric (VQM).

Besides video quality, the behavior of the user may influence the perceived QoE. Camera Shaking or wrong positioning while recording a video may reduce the quality of the video stream. CaOS avoids such video degradations by introducing shot boundaries. A currently broadcasted video sequence is replaced by another, if such degradations occur. Those approaches proved suitable, as shown e.g. by Saini [8]. Analysis of the users' behavior while they are interacting with the mobile device is possible based on the sensors integrated into the devices. Typical sensors integrate the location and motion of users. Also, software-based sensors indicate the current video recording properties or energy level of the device. Here, patterns can be used to identify recording degradations based on sensor readings - as shown by CriCri [5].

CaOS learns about shot diversity by avoiding recording degradations and additionally providing view diversity to increase the watching experience. The determination of view

[4]http://developer.android.com/guide/topics/sensors/sensors_overview.html

switches as well as identifying the best available view are complex tasks as computational complex algorithms are used to estimate human visual experience. Analyzing visual properties of images and videos in order to determine video quality already achieves high correlations with the subjective impressions of humans, but identification of content relevance and aesthetical appeal is currently under investigation. Therefore, integrating those approaches into CaOS for large-scale events requires an appropriate metric classification in conjunction with suitable monitoring system.

3. MONITORING SYSTEM

The main challenge for generating a live stream of user generated content is a limited upload capacity for the users. This especially applies for cellular networks. Therefore, an efficient coordination of the streams that are to be uploaded is required while minimizing the overhead introduced by the coordination of the system.

To address the specific needs for monitoring parameters that influence the QoE for user-generated video we propose a dedicated monitoring framework. The main design goals for this framework are to achieve the best trade-off between the freshness of information and the level of detail for required orchestration and composition of video streams while minimizing necessary overhead. The system design is based on in three zones: (1) node state, (2) zone state (3) and global state. As illustrated in Figure 2, each of the zones derives different kinds of information that is then used for QoE prediction. The first zone describes processes which are executed independently on each participating device. Here, sensor information is utilized and aggregated to a quality indicator for the video. Each new node that joins the system contacts the central CaOS server and transfers information regarding its zone, network parameters and sensor information related to the current recording. The device then provides periodic messages updating this information.

The zone state classifies a collection of devices registered at the same access point. This mapping is achieved by uploading the current SSID in case of a wireless network connection or the cell ID for devices being connected in a cellular network. Using this information, devices are categorized by zone and the over-utilization of upload capacities can be mitigated by setting a limit for parallel uploads allowed for each zone based on the its type. This limit can be higher for newer generation mobile data access points (LTE) and wireless networks and lower for low performance connections like 3G. Therefore the influence on the availability of network resources between each other can be reduced.

The global state includes aggregated views of the node and zone states as well as indicators for video quality metrics and content relevance which are derived from external services. The CaOS service holds two lists of devices for streaming: (1) Active devices, that currently transfer a live stream to the CaOS and can be selected for the resulting live feed. (2) Standby devices, that periodically upload measurements of sensor and zone information. Here, the first list is based on the currently available number of joined devices and their allocation in zones. Based on this data, CaOS selects the most relevant video stream and coordinates the devices that are selected for uploading their live stream. Devices in the second list can be activated for active streaming by CaOS based on the predicted relevance of their content. The proposed system is currently being developed in Java, based on the ØMQ[5] framework.

Message Type	*Data*	*Source*
JOIN	Device type Cell-ID AP-ID Device-ID GPS coordinates	Node
JOIN-ACK	Update interval	Node
UPDATE	Device-ID Recording status Network status Sensor information	Node
UPDATE-REQ	Single video frame (opt.)	Server
SET	Start/Stop upload Update interval	Server
LEAVE	Device-ID	Node

Table 1: Overview of messages used for system coordination

The system uses a set of messages for node registration, data retrieval, update and coordination as listed in Table 1. First, the JOIN message is sent by each node in order to register at CaOS. This message contains data regarding the network based location of the node, the physical location, a generated unique Device-ID as well as device capabilities for video recording (e.g. max resolution). Upon receiving, the server replies with a JOIN-ACK message that contains the initial update interval for the node. Based on this interval, the node updates its current recording status annotated with the related sensor information by sending an UPDATE message. In case more recent information is required by CaOS, an UPDATE-REQ message is sent, which follows an immediate UPDATE message from the addressed node. By an optional field in the UPDATE-REQ message, a frame of the current recording can be requested by the server. Further, the server can initiate and stop the upload of currently recorded video by each device using a SET message. Last, leaving nodes may sent a LEAVE message to unregister from the CaOS. The ungraceful exit of nodes is assumed if three successive expected UPDATE message were not received or the same amount of UPDATE-REQ were ignored.

4. RELATED WORK

The 3GGP [1] investigates different scenarios for media broadcasting from mobile devices including mobile video broadcasting. Mobile video broadcasting allows smartphone users to record video and directly cast it to remote viewers. With the increasing number of smart phones and platforms for receiving live broadcast video such as bambuser and rose, researchers such as Engstrom concentrate on the collaborative creation of video compositions from smart phones,

[5] `http://zeromq.org/`

in this case music video jocking [6]. Parallel recordings of events such as concerts or sports could be leveraged to create superior videos, by composing a video mix as proposed with Shrestha's virtual director [9]. Here, the authors introduced a completely automatic algorithm based on an objective quality function. MoviMash [8] is the most advanced approach existing for mobile broadcasting services. The system combines a no-reference video quality metric and aesthetical guidelines. Composition thus relies in many approaches on increasing the video quality. A no-reference video quality metric is required for predicting video quality. Yang [12] proposes a combination of spatial distortions in all images in a video. To integrate temporal defects in two adjacent images weighting of structural distortions is applied. Besides classical video quality, content relevance is of interest to our application. Bao's work [4, 3] contributes content relevance estimation by using user's reactions on watching a video. In MOVI an approach is described that generates video summaries based on collaborative sensing in mobile phones. With Smile the front camera of modern mobile phones is recording user's reactions on a video being displayed, allowing the ranking of individual segments of a video according to their importance and their quality.

All prior work concentrates on the usage of video analysis algorithms, which is computationally expensive. CriCri [5] shows a first step towards replacing video analysis with mechanisms that leverage different sensors in the recording devices to compose the video. We propose a monitoring and metric framework for the increasing need of mobile broadcasting systems that rely on video collection of multiple streams and composition of one superior video.

5. CONCLUSION

In this work we describe a classification of metrics, a monitoring and distribution framework that can be leveraged for monitoring internet-scale mobile video broadcasting services. We illustrated that for video quality, network measurement as well as content relevance that the metric composition is aggregating the information on correct scope.

Our future work includes the realization and prototypical as well as simulative evaluation of this approach.

6. ACKNOWLEDGMENTS

This work has been funded by the German Research Foundation (DFG) in the Collaborative Research Center (SFB) 1053 MAKI – Multi-Mechanism Adaptation for the Future Internet.

7. REFERENCES

[1] 3GPP. TR 22.947 - Study on Personal Broadcast Service (PBS) (Release 11). Technical report, 3GPP, Sophia Antipolis Valbonne, France.

[2] J. Apostolopoulos, R. A. Guérin, S. Tao, S. Member, and R. Gu. Real-Time Monitoring of Video Quality in IP Networks Real-Time Monitoring of Video Quality in IP Networks. *IEEE/ACM Transactions on Networking*, 16(5):1052–1065, 2008.

[3] X. Bao, S. Fan, A. Varshavsky, K. Li, and R. Roy Choudhury. Your reactions suggest you liked the movie: Automatic content rating via reaction sensing. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp*, pages 197–206, 2013.

[4] X. Bao and R. Roy Choudhury. Movi: Mobile phone based video highlights via collaborative sensing. In *ACM International Conference on Mobile Systems, Applications, and Services - MobiSys*, pages 357–370, 2010.

[5] F. Cricri, K. Dabov, I. D. D. Curcio, S. Mate, and M. Gabbouj. *Multimodal extraction of events and of information about the recording activity in user generated videos*, volume 70. 2012.

[6] A. Engström, M. Esbjörnsson, and O. Juhlin. Mobile collaborative live video mixing. In *ACM International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI*, pages 157–166, 2008.

[7] C. Keimel, M. Klimpke, J. Habigt, and K. Diepold. No-reference video quality metric for hdtv based on h.264/avc bitstream features. In *IEEE International Conference on Image Processing (ICIP)*, pages 3325–3328, 2011.

[8] M. K. Saini, R. Gadde, S. Yan, and W. T. Ooi. MoViMash: Online Mobile Video Mashup. In *ACM International Conference on Multimedia - MM*, pages 139–148, 2012.

[9] P. Shrestha, P. H. de With, H. Weda, M. Barbieri, and E. H. Aarts. Automatic mashup generation from multiple-camera concert recordings. In *ACM International Conference on Multimedia - MM*, pages 541–550, 2010.

[10] S. Wilk and W. Effelsberg. The influence of camera shakes, harmful occlusions and camera misalignment on the perceived quality in user-generated video (accepted for publication). In *IEEE International Conference on Multimedia and Expo - ICME*, 2014.

[11] S. Wilk and W. Effelsberg. Systematic assessment of the video recording position for user-generated event videos (accepted for publication). In *ACM International Conference on Multimedia - MM*, 2014.

[12] F. Yang, S. Wan, Y. Chang, and H. Wu. A novel objective no-reference metric for digital video quality assessment. *IEEE Signal Processing Letters*, 12(10):685–688, 2005.