Optimization of Video Viewing System with Real-time Video Transcoding

Leon Dragić
Faculty of Electrical Engineering and Computing
University of Zagreb, Croatia
leon.dragic@fer.hr

Abstract — In the past years the rapid development of smartphones and mobile telecommunications technologies (3G, 4G) enabled users to easily access multimedia content using mobile devices. High quality videos which are streamed from the Internet are power consuming and demand significant amounts of network resources. Viewers access video content with various devices which differ in screen resolution and screen size. Current video streaming systems don’t provide the optimal video quality which depends on the device that is accessing the video and is limited with the network conditions at the time of streaming. In this paper, the overview of current video viewing system is made. In order to provide the best video quality based on network conditions and device properties while saving power and network resources, video transcoding system is proposed.

Keywords—transcoding; quality of experience; mobile video; power consumption; video streaming

I. INTRODUCTION

Multimedia consumption using mobile devices has experienced rapid growth with the development of smartphones. Over 6 billion hours of video are watched each month on YouTube and 40% of that is done using mobile devices [1]. Streaming of 6 billion hours of video per month requires huge amount of network resources for video transmission and power resources [2] for video decoding and playing. Because of the great variety of devices which are accessing the multimedia content under adverse network conditions, current video streaming systems in most cases don’t provide the optimal video quality and thus waste valuable resources or unnecessarily lower the Quality of Experience (QoE). Most of the current video streaming systems offer videos previously encoded to standard video resolutions e.g. 360p, 480p, 720p, 1080p. For every resolution, a new copy of video sequence is stored on the server and that demands huge storage resources.

II. VIDEO VIEWING SYSTEM

Multimedia viewing system consists of 3 parts as shown in Fig. 1.

A. Server

The process of video viewing starts at the server where video is encoded to a specific spatial, temporal and amplitude resolution (STAR) using specific video codec. The combination of STAR is the key factor when it comes to Quality of Experience (QoE), power consumption and network requirements. With the increase of spatial and temporal resolution and decrease of quantization step, the video quality enhances. Higher video quality leads to better QoE but also increases network bandwidth consumption and power consumption of mobile device. It is important to minimize power consumption of video decoding and network requirements without drastically affecting QoE.

In order to estimate the video quality, Video Quality Assessment (VQA) models are used. There are many VQA models, but one of the most popular is the Video Quality Experts Group (VQEG) model. This model uses a combination of objective and subjective methods to assess the quality of a video. The objective methods use computer algorithms to measure the differences between the original and decoded video, while the subjective methods involve human observers who rate the quality of the video.

Keywords—video transcoding; video quality assessment; mobile video; power consumption; video streaming

II. VIDEO VIEWING SYSTEM

Multimedia viewing system consists of 3 parts as shown in Fig. 1.

A. Server

The process of video viewing starts at the server where video is encoded to a specific spatial, temporal and amplitude resolution (STAR) using specific video codec. The combination of STAR is the key factor when it comes to Quality of Experience (QoE), power consumption and network requirements. With the increase of spatial and temporal resolution and decrease of quantization step, the video quality enhances. Higher video quality leads to better QoE but also increases network bandwidth consumption and power consumption of mobile device. It is important to minimize power consumption of video decoding and network requirements without drastically affecting QoE.

In order to estimate the video quality, Video Quality Assessment (VQA) models are used. There are many VQA models, but one of the most popular is the Video Quality Experts Group (VQEG) model. This model uses a combination of objective and subjective methods to assess the quality of a video. The objective methods use computer algorithms to measure the differences between the original and decoded video, while the subjective methods involve human observers who rate the quality of the video.
models and their main goal is to automatically assess average viewer's opinion on quality of video. Combining VQA models with other models or processes [3], such as power or bandwidth consumption model, it is possible to determine video properties that will provide optimal video quality for certain conditions. Xin Li et al. [4] proposed an analytical power consumption model for H.264/AVC video decoding using a hardware accelerator on popular mobile platforms. The model is expressed as a product of the power functions of video spatial resolution and temporal resolution. Extensive simulations showed that quantization doesn't affect the decoding power for hardware codec solutions. Such model can be applied in power-rate constrained mobile adaptive streaming for estimating video decoding power and determining the proper video resolutions, considering the receiver's remaining battery life and limited network bandwidth.

B. Network

The second step of video viewing process is transmission of the encoded video through the network. The network has its own parameters which also affect the QoE. Low bandwidth, jitter and packet loss can lead to increased power consumption, increased video loading time and video distortion [5]. Junlun Ma et al. [6] conducted a measurement study on power consumption of mobile video streaming apps under various network conditions. The study showed that power consumption of wireless data transmission which accounts for 30% of total power consumption, increases significantly under adverse network conditions and that power consumption is mostly determined by the duration of busy period of the Wireless Network Interface Card (WNIC) and not the download rate. In order to make the wireless data transfer power efficient, the busy period has to be shortened by increasing the available bandwidth and thus enabling WNIC to enter power saving mode. Packet loss and large propagation delay are cause for frequent data retransmission, which prolongs WNIC's busy period and increases the power consumption which is inversely proportional to the available bandwidth.

C. Device, user and surrounding influences

After the transmission, encoded video is decoded on the user's device. Andrew Catellier et al. made an exploratory experiment [7] using five different mobile devices in two testing environments to understand how the perceived quality of video sequences differed among many devices. The mobile devices ranged from a smartphone to a broadcast quality LCD television. Subjects were shown 8 video sequences with 3 video quality levels on all of the devices. Each video sequence was encoded to match the native resolution of each device with the aspect ratio of the video sequences left unchanged. The results showed that iPhone 4's display (960x640 px) didn't seem to have an effect on MOS (mean opinion scores) when compared to the iPod Touch's display (480x320 px). Because the iPhone 4 and the iPod Touch have the same screen size and statistically similar MOS, it is possible to deliver iPod Touch resolution video sequences to the iPhone 4 with no negative consequences and significant bandwidth savings.

The later studies [8] showed that this conclusion was not correct. The measurements showed that the QoE at a given image resolution on iPhone4 is much lower than that of similar resolution on iPhone3GS. This indicates that people need a higher video quality for watching videos on a mobile device with a higher display resolution, more so than if they were using a device with a lower display resolution. The paper also proposed a set of novel acceptability-based QoE models which were able to predict users' acceptability and pleasantness in various mobile video usage scenarios. Acceptability-based QoE models are very useful in designing adaptive video streaming systems unlike MOS-based models which are unable to indicate whether video quality is acceptable or not.

Video player can play a big role in draining the battery power. By selecting software video decoder instead of hardware video decoder, the power consumption of video decoding will drastically increase. This makes the choice of video player an important part of video viewing process.

QoE is also affected by the environment in which the video is being viewed. Movement of the user, viewing distance and ambient luminance are some of the important factors that can impact QoE, which was shown in [9]. The viewers will notice less video signal distortion and have lower expectations on the quality of mobile video in the presence of environmental visual interference. This can be used in building adaptive video streaming system in which videos would be encoded to lower quality when the environmental visual interference increases.

III. VIDEO TRANSCODING

A. General

Video transcoding is a process of converting a video from one encoding to another. In this process, video properties, such as spatial, temporal and amplitude resolution or video codec can be changed. There are many uses for video transcoding, but in this paper, special interest is put on real-time video transcoding for maximizing QoE while keeping the resource requirements as low as possible. This video transcoding system will provide the optimal video quality for a specific device and network conditions without demanding huge storage resources like the current video streaming systems.

The video transcoding process is extremely compute-intensive and the implementation of an efficient transcoding system is a challenging task. In order to implement transcoding system, servers capable of processing big amount of data in real-time are needed. Low power, high performance processing can be achieved using modern multi and many-core architectures and streaming models as shown in [10, 11], high performance SOC [12,13] or by distributing the process in the cloud [14, 15]. Current real-time software implementation of the H.265 4Kp30 video encoder requires four 8-core Intel Xeon microprocessors [16]. However, the same encoder can also be implemented on one Virtex-7 690T FPGA which shows that FPGA performs the same work with substantial power reduction.
B. Specific video properties

Determining the specific video properties is one of the key parts of video transcoding system. The right selection of these properties enables the transcoding system to provide the best possible video quality under certain conditions to a specific device. There are many conditions that can be introduced into the model for determining the specific video properties such as power consumption, available network resources, the limits of human eye, etc. Some of the conditions can vary while the streaming process is lasting. The advantage of real-time transcoding system is that it can meet those changes by adapting the specific video properties on the fly.

C. Video transcoding system

The first step in transcoding is to check the available resources for video transmission and to get the information about the device that is accessing the video content. Based on this data, the system determines the specific video properties for video streaming such as spatial, temporal and amplitude resolution or video codec. The video is then encoded in real-time to a format that meets the determined specific video properties and transmitted to user’s device through network. During the transmission, network conditions are periodically checked. If the conditions change during the transmission, video properties may change in order to improve QoE.

D. The impact of transcoding

Similar to transcoding, there are other techniques for adapting video content to improve QoE [17]. One of the techniques is SVC – Scalable Video Coding [18]. SVC is an extension to a very popular video codec H.264 [19] and its successor H.265 [20]. The idea behind SVC is encoding of a video in 3 scalabilities: spatial, temporal and amplitude scalability. Each of those scalabilites consists of base layer and multiple enhancement layers. Each enhancement layer increases the quality of the video. The enhancement layer of spatial scalability increases the video resolution. In temporal scalability it increases the frame rate while in amplitude scalability it decreases the quantization step. In order to adapt the bitstream to the needs of end users as well as the capabilities of the receiving device or the network conditions, the selection of layers which are going to be transmitted is done. Depending on the varying conditions such as network status, the selection of layers can be dynamically adjusted. The SVC extension solves the problem of storing multiple versions of the same video on server and it enables dynamic adaptation of video while streaming it to a user’s device. The downside of the SVC is that layers are still pre-encoded to some standard video parameters which don’t have to be optimal for the device or the conditions under which the video is being streamed.

The experiment was conducted [21] to find out how much network traffic and power can be saved by delivering video encoded to device-specific video resolution instead of delivering video encoded to standard video resolutions. 9 video testing sequences with length between 10 and 20 seconds were used. Every video sequence was encoded to 5 standard resolutions and 1 device-specific resolution represented in Table I. All the resolutions had an aspect ratio of 16:9 except 640p which had 3:2. The specific device resolution was calculated using 3 parameters: device display pixel density, viewing distance and the sensitivity of the human eye. There are many discussions regarding the upper limit of pixel density after which human eye is unable to discern individual pixels at a typical viewing distance. According to Apple, the upper limit is 326 PPI while holding device on a standard viewing distance which is around 30 cm [9]. Other scientists [22] put upper limit of pixel density at 900 PPI. In this experiment the limit of 326 PPI was used as the pixel density after which the improvement in video quality is no longer visible by the user. The videos were viewed on HTC One smartphone with pixel density of 469 PPI. Pixel density can be lowered by increasing the display size or by decreasing the display resolution. To determine the device-specific resolution we calculated the spatial resolution at which the pixel density would be 326 PPI which resulted with the spatial resolution of 1632x918 px. In the experiment the priority was put on lowering spatial resolution while keeping quantization parameter (QP) fixed to 24 for all video sequences and resolutions. The paper [4] showed that quantization parameter doesn’t affect the power consumption unlike spatial resolution. The acceptability-based model and the paper [23] showed that PAcc and MOS drop faster while increasing OP over 24 than they drop while reducing spatial resolution. The experiment resulted in 108 values which represent average power consumption and average bitrate for 9 sequences at 6 video properties.

<table>
<thead>
<tr>
<th>Resolution name</th>
<th>Resolution (width x height, in px)</th>
<th>Aspect ratio</th>
<th>YouTube</th>
<th>Vimeo</th>
</tr>
</thead>
<tbody>
<tr>
<td>360p</td>
<td>640x360</td>
<td>16:9</td>
<td>Y,V</td>
<td></td>
</tr>
<tr>
<td>480p</td>
<td>854x480</td>
<td>16:9</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>640p</td>
<td>960x640</td>
<td>3:2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>720p</td>
<td>1280x720</td>
<td>16:9</td>
<td>Y,V</td>
<td></td>
</tr>
<tr>
<td>918p</td>
<td>1632x918</td>
<td>16:9</td>
<td></td>
<td>specific</td>
</tr>
<tr>
<td>1080p</td>
<td>1920x1080</td>
<td>16:9</td>
<td>Y,V</td>
<td></td>
</tr>
</tbody>
</table>

* Used as a streaming resolution on YouTube or Vimeo

<table>
<thead>
<tr>
<th>Resolution (width x height, in px)</th>
<th>Average bitrate (kb/s)</th>
<th>Average power consumption (mW)</th>
<th>PAcc</th>
</tr>
</thead>
<tbody>
<tr>
<td>640x360</td>
<td>2681</td>
<td>402</td>
<td>0.554</td>
</tr>
<tr>
<td>854x480</td>
<td>4509</td>
<td>443</td>
<td>0.670</td>
</tr>
<tr>
<td>960x640</td>
<td>6373</td>
<td>491</td>
<td>0.737</td>
</tr>
<tr>
<td>1280x720</td>
<td>9169</td>
<td>537</td>
<td>0.843</td>
</tr>
<tr>
<td>1632x918</td>
<td>15404</td>
<td>635</td>
<td>0.929</td>
</tr>
<tr>
<td>1920x1080</td>
<td>22702</td>
<td>719</td>
<td>0.966</td>
</tr>
</tbody>
</table>
resolutions which are shown in Fig. 2 and Fig. 3. The results showed that specific resolution 1632x918 px in average consumed almost 12% less power with the bitrate values lower for 32% as shown in Fig. 4. If the power consumption of network streaming which is described in [6] is taken into consideration, the reduction of bitrate will save additional 32% of power in respect to the baseline model of idle network state. This will be done by decreasing the amount of time spent in the wireless network interface controller fetching state. Using the acceptability-based model PAcc was calculated for all resolutions for which power consumption and bitrate was measured. PAcc is calculated using the ratio of video resolution and device display resolution and it represents the possibility of a video quality making viewers pleasant or comfortable. PAcc values for selected spatial resolutions viewed on 1920x1080 px displays are shown in Table II. The model showed that device-specific resolution 1632x918 px has PAcc value of 92.9% which is only 3.7% lower than PAcc value of 1920x1080 px which is 96.6%. In other words, bitrate and power consumption of transmission process can be lowered by 32% (Fig. 5), power consumption of video decoding process by 12% while reducing the QoE by only 3.7%.

IV. CONCLUSION

With the development of smartphones and mobile telecommunications technologies, multimedia consumption has increased significantly. Video streaming services such as YouTube stream huge amounts of video content on a daily basis and the pressure on resources is building up. Videos are previously encoded to standard video parameters and are not optimized to a great variety of the devices that are accessing it or the network conditions at the time of the streaming. In order to minimize the resource requirements and maximize the QoE, video transcoding system is proposed. The experiment showed that by transcoding videos to the specific video resolution instead of using standard resolutions, significant savings of network traffic and power can be made. The challenging part of transcoding is that it is a very compute-intensive process and additional research of the transcoding process should be made. Special focus should be put on optimizing transcoding process by selecting appropriate technologies, parallelizing the algorithms, distributing tasks among many cores and other techniques that can improve the efficiency of transcoding process.
REFERENCES


