Watching the Watchers: Resource-Efficient Mobile Video Decoding through Context-Aware Resolution Adaptation

Octavian Machidon
octavian.machidon@fri.uni-lj.si
University of Ljubljana
Ljubljana, Slovenia

Tine Fajfar
tf5442@student.uni-lj.si
University of Ljubljana
Ljubljana, Slovenia

Veljko Pejović
veljko.pejovic@fri.uni-lj.si
University of Ljubljana
Ljubljana, Slovenia

ABSTRACT

Mobile computing evolution is critically threatened by the limitations of the battery technology, which does not keep pace with the increase in energy requirements of mobile applications. A novel approach for reducing the energy appetite of mobile apps comes from the approximate Computing field, which proposes techniques that in a controlled manner sacrifice computation accuracy for higher energy savings. Building on this philosophy we propose a context-aware mobile video quality adaptation that reduces the energy needed for video playback, while ensuring that a user's quality expectations with respect to the mobile video are met. We confirm that the decoding resolution can play a significant role in reducing the overall power consumption of a mobile device and conduct a user study with 22 participants to investigate how the context in which a video is played modulates a user's quality expectations. We discover that a user's physical activity and the spatial/temporal properties of the video interact and jointly influence the minimal acceptable playback resolution, paving the way for context-adaptable approximate mobile computing.

CCS CONCEPTS

• **Human-centered computing** → **Mobile computing**; **Empirical studies in ubiquitous and mobile computing**.

KEYWORDS

mobile computing, approximate computing, video decoding, context inference, spatial information, temporal information

ACM Reference Format:
Octavian Machidon, Tine Fajfar, and Veljko Pejović. 2020. Watching the Watchers: Resource-Efficient Mobile Video Decoding through Context-Aware Resolution Adaptation. In *17th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous), November 9-11, 2020, Darmstadt, Germany*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Mobile computing field underwent an exponential growth in the last few decades – invented only slightly more than a decade ago, the smartphone is already owned by more than three billion people in the world and the variety of mobile applications have fully transformed the way we communicate, do business, navigate in space, or find social contacts.

The change in the way we consume information via mobile devices is particular staggering, moving from traditional voice and text media to video. The amount of content seen through mobile video is more than doubling every two years [10]. Surveys show that already 90% of the owners watch videos on their mobile devices and that more than 70% of all YouTube content is consumed via mobile devices [15]. In 2019, mobile video traffic accounted for half of the total mobile data traffic, and the forecast indicates that almost 80% of the worldwide mobile data traffic will be video traffic by 2022 [10]. The recent COVID-19 pandemic further exacerbated the growth, with the fields as diverse as the education, remote work, and healthcare, rapidly jumping on the mobile video bandwagon [7].

However, the proliferation of mobile video, and, indeed, mobile computing in general, is hindered by the physical constraints and limitations of the underlying hardware. One key issue in this regard is related to one of the most critical resources of a mobile devices – its battery. The battery technology is experiencing a disproportionally slower growth – practically a stagnation — compared to the other mobile resources including the CPU speed and computing power, storage space, and wireless transmission speed [25]. The lack of a revolutionary solution for modest battery capacity calls for further efforts towards the efficient use of limited resources available on mobile devices.

Building upon the philosophy of Approximate Mobile Computing (detailed in Section 2), in this work we investigate the opportunity for improving the energy efficiency of mobile video playback by adjusting the playback resolution according to the actual context-depended needs of a mobile user. Two key premises guide our work: 1) the energy requirements of video playback vary with the selected resolution and 2) users requirements vary with the context in which the playback is seen. Through fine-grain measurements we confirm the former, and we conduct a 22-user study, described in Section 4, to examine the latter hypothesis. We discover that contextual situations, such as whether a user is still, running, walking, or riding in a car, significantly impact the minimum playback resolution the user is satisfied with. Findings, further examined in Section 5, also uncover other aspects that can play a role in the user's tolerance with lower video quality, such as the video's spatial and temporal complexity. Based on the collected dataset, we observe the opportunity to extending the battery life by 23.2% in our particular use-case, should the context-dependent resolution requirements be perfectly matched by the playback app. Noting that the information on a

MobiQuitous, November 9-11, 2020, Darmstadt, Germany

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

user's activity practically comes "for free" in modern mobile operating systems, we believe that context-dependent video decoding quality adaptation represents a viable alternative to the existing work (summarized in Section 6) for preserving the battery charge and enabling further proliferation of mobile computing. In Section 7 we draw final conclusions and highlight future research directions in approximate mobile video playback.

The work presented in this paper was performed with reproductibility in mind. Thus, in accordance to the best practices [27], the collected experimental data is publicly available to the research community at https://gitlab.fri.uni-lj.si/lrk/approximate_video_study/.

2 TOWARDS APPROXIMATE MOBILE COMPUTING

Approximate computing (AC) is a resource-efficient computing paradigm grounded in the observation that the result of a computation often need not be perfectly accurate in order to satisfy the end-user's needs [22]. Opportunities for AC frequently arise in situations where the computation inputs are noisy (e.g. sensor data), or when the output is further manipulated and interpreted by the user (e.g. 3D graphics rendering). In such situation, approximate computation can deliver a perfectly satisfactory result while reducing the energy use. AC techniques have already proven their efficiency in various desktop scenarios, with approaches ranging from speeding up code execution through compiler-level optimizations that omit certain lines of code [21] to performing neural-network based approximations instead of complex function calculations [14], demonstrating significant energy savings while maintaining acceptable result accuracy.

Building upon the idea of AC, approximate mobile computing (AMC) introduces approximation on mobile devices [26]. The core difference from the conventional AC being the context of use, which in mobile computing tends to vary over time. A user's physical activity, her location and collocation with other users, the outside brightness, and numerous other factors may vary throughout the day and impact the user's requirements with respect to mobile computation. Significant challenges lay ahead before the full potential of AMC can be exploited: 1) practical means of enabling approximation in mobile applications need to be provided; 2) the benefits of approximate execution need to be quantified; 3) opportunities for approximation need to be identified and profiled, and 4) lightweight context recognition relevant for the AMC application needs to be implemented.

In this paper we present our initial efforts towards enabling AMC. We focus on mobile video as one of the most prominent and most energy hungry aspects of mobile computing. We hypothesize that the context in which a mobile video is played may impact a user's perception of the content. The context can be represented by a potentially unlimited number of dimensions, thus, backed by the prior work [29, 31, 33] here we focus on the two most relevant and intuitive dimensions – a user's physical activity and the characteristics of the mobile video. Consequently, we formulate the following research questions (RQ) that our study aims to answer:

- *RQ_1: Is the physical activity the user is engaged in when watching a video on a mobile device influencing the user's quality expectations/requirements?*
- *RQ_2: Does the video content (its spatial and temporal characteristics) impact the user's satisfaction with the lower video quality and how this relates to the physical mobility state of the user?*

In addition, should we prove that a link between the context and the quality expectations exist, we are interested in the potential of enabling energy savings by adjusting video playback according to the current context. Thus, we also aim to answer:

- *RQ_3: How much energy can be saved, if the video playback quality is adjusted to the minimal level that still satisfies the user's context-dependent quality expectations?*

First, however, for AMC to be practical we require a straightforward means of adjusting approximation. Moreover, the reduction in computation should lead to a gradual decrease in the end-result accuracy, without the loss of correctness (i.e. the result is usable at all times, and the approximation "knob" always gives a correct result). Furthermore, the reduction in computation should result in reduced resource usage. In our work we settle on *video decoding resolution* adjustment. Virtually all video distribution frameworks (e.g. Youtube, Vimeo), as well as mobile video players, support playback resolution adaptation. Furthermore, setting video resolution always leads to correct execution and the loss of quality is gradual as we dial down the resolution. In the following section we also confirm that the loss of quality corresponds to lower resource usage making video decoding resolution a suitable technique for approximate computing adaptation.

3 PRELIMINARY: ENERGY VS. QUALITY TRADE-OFF IN MOBILE VIDEO DECODING

A monotonically increasing relationship between the computation accuracy and the resource consumption is at a core of approximate computing. In this section we chart the relationship between the video decoding quality and the mobile consumption. When performing the energy measurements, we use a popular video decoding software VLC Player [4] running on a Samsung Galaxy S3 (I9300) Android smartphone. Despite being released eight years ago, the phone supports both hardware and software video decoding and, importantly, has a detachable battery that allows us to connect the phone to a high-frequency power meter. The VLC Player was chosen for the energy measurements due to its flexibility in allowing rapid enabling/disabling of hardware accelerated decoding.

We used Monsoon [2], a high sampling frequency platform commonly used for power measurements in mobile computing [28]. The experimental setup for measuring energy consumption relies on measurements from the Monsoon High Voltage Power Monitor (HVPM), which generates energy readings at a sampling frequency of $5kHz$. Each sample contains a timestamp in ms, voltage in mV and electrical current in mA. The HVPM is directly attached to the battery interface of the mobile device, which is powered solely by the power supplied through the HVPM.

During the energy measurements, the HPVM output voltage was set to $4.2V$, which corresponds to an almost full battery voltage. The same 1-minute video was downloaded from YouTube on the

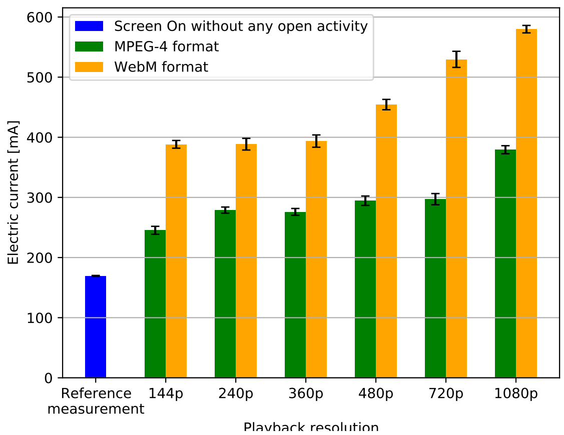

Figure 1: Smartphone average current consumption during video playback at different resolutions together with the standard deviation of the measurements. A monotonically increasing relationship between the video decoding resolution and the current consumption is evident for both software (WebM) as well as hardware (MPEG 4) decoding.

device in the following resolutions: 144p, 240p, 360p, 480p, 720p and 1080p, in both WebM and MPEG-4 formats. The baseline for comparison was a reference energy measurement performed with just the phone screen turned on, without other apps/services running. For each resolution, the video was played 10 times using VLC Player, and the energy readings were averaged over the 10 runs. During the measurements, the screen brightness was set to the minimum, all non-essential services running on the smartphone that could interfere with the energy measurements were shut down, and the smartphone's Airplane mode was turned on, in order to minimize the effect of on-device communication modules (e.g. GSM, Wi-Fi, Bluetooth, etc.).

3.1 Energy measurement results

The results of the energy measurements for video playback on the mobile device at different resolutions are shown in Figure 1. We observe a significant difference in power consumption for playing videos using MPEG-4 vs. WebM decoding. This is expected since MPEG-4 decoding is hardware-accelerated in modern smartphones, while WebM decoding is performed in software. With both formats we see a generally increasing trend – the higher the decoding quality (resolution), the higher the consumption is. Interestingly, in the WebM case the lower resolutions (144p, 240p and 360p) have similar average current consumption, while the consumption increases considerably as we move to higher resolutions (480p, 720p and finally 1080p). Since there are no significant differences between the lower three resolutions, from the energy efficiency point of view lowering the resolution under 360p would have no positive impact on energy savings, moreover it would only potentially decrease a user's satisfaction.

In Figure 1 we also show the reference measurement performed with just the screen turned on indicating that a large part of the energy consumption belongs to the screen itself, apart from the energy required by the actual video decoding. This also confirmed by other studies on smartphone energy consumption [9].

4 USER STUDY: CONTEXT-DEPENDENT ACCURACY REQUIREMENTS

In Section 3 we confirmed that video decoding resolution represents a suitable knob for approximate mobile computing. In this section we investigate the opportunities for lowering the decoding resolution, thus reducing the energy usage, while still satisfying the users' needs.

Perception of video playback is shaped by limited capabilities of human attention and sensory systems [8]. These are in turn affected by numerous factors [29]. For instance, the movement of a device on which a video is watched impacts the ability to focus and interpret the picture; outside brightness impacts the contrast of the OLED display preventing a user from discerning details in the picture; and the properties of the video, such as the dynamics at which the content changes, require more or less attentional resources from the viewer [23, 34].

The influencing factors collectively form the context which, we hypothesize, impacts users' requirements with respect to the video playback resolution. In this work we focus on a user's *physical activity* as the most prominent dimension of the outside context and the one that can be acquired with the minimal use of the mobile's energy. In Android OS coarse-grained physical activity (e.g. "running", "walking", "in vehicle", "still", etc.) can be acquired using Google Play Services' internal classifier that is jointly maintained for all apps on the device. Having in mind that activity detection is used across a range of apps, from navigation, over exercise tracking, to health and wellbeing apps, and that an average user has more than thirty apps installed on her phone [1], there is a high probability that activity recognition pipeline would anyway be active and routinely queried by other apps. Consequently, querying this classifier for our purpose would likely incur negligible additional energy cost, which makes the physical activity context perfectly suited for our goal of reducing the energy use. Besides the physical activity, we also hypothesize that *the content of the video* impacts a user's decision to require a higher or a lower resolution decoding. This information, too, can be acquired with very little cost as no additional device components need to be powered on. Therefore, we further calculate a video's spatial and temporal information and inspect their role on a user's desired video playback resolution.

For video rendering during the user experiments we use NewPipe – an open source YouTube-streaming frontend for Android [3] – which allows both online and offline video playing. For the experiments in this study, the videos were preloaded in order to avoid any networking effects that might impact the user perception when watching the videos. We add logging functionalities to the app, thus in each experiment we record the initial resolution, physical activity state[1], the video played, and each event of a user

[1] In this paper we describe controlled experiments, where the users were instructed to perform a certain activity at a certain time, thus, we do not use on-device classifier for recognising activities, but log them manually. Nevertheless, we have implemented

changing the resolution. For resolution change events we record the new resolution and the timestamp marking the moment the change took place.

The volunteers in our study were 22 students from our institution with both technical and non-technical backgrounds. The group consisted of 13 male and 9 female participants. We select 12 one-minute-long YouTube videos to be watched by the users. The video content The content varied among the videos from music, sports, outdoor/indoor activities, and others, resulting in various spatial and temporal characteristics of the videos.

Each of the participants in the study group watched all 12 videos in different activity states (three videos per state): still, walking, running, and traveling as a passenger in a vehicle. All the experiments were performed on the university campus: in the same laboratory room when still, on the same hallway when walking and running, and on the same route on the campus when traveling as a passenger in a vehicle (the same driver and vehicle for all tests/subjects). The following smartphones were used during this study for watching videos by the participants: Samsung Galaxy S3, Samsung Galaxy S4 and Nexus 6.

To ensure the obtained results were comparable and relevant, all participants were instructed to follow the same protocol during the experiments. Hence, the following instructions were given to the participants:

- The users were instructed about the resolutions available and the process of changing the resolution when watching a video clip. They were asked to switch the resolution to a higher one only when dissatisfied with the quality;
- They were asked to keep the device horizontal at all times in order to ensure the video is played in full-screen;
- The brightness was pre-set to 80% and the participants were asked not to change it;
- Before each experiment the users were informed about what video and in what starting resolution they should watch; the starting resolutions presented a pseudorandom distribution. We choose this approach to avoid the situation where always starting from a low resolution might artificially reduce the user expectations due to indolence in changing to a higher resolution.

5 RESULTS AND DISCUSSION

We examine how the user's satisfaction and quality expectations are impacted by the physical activity by analyzing the resolutions that were found acceptable when watching videos in each of the four mobility states. Furthermore, we perform a statistical investigation to determine other factors impacting the user's tolerance to lower video quality. Such investigation reveals that the video content, i.e. its spatial and temporal characteristics, also play an important role. In addition to these two elements (activity context and video content) we discuss the impact the AMC approach has on the energy savings.

automatic activity recognition and plan to run an in-the-wild study as a part of our future work.

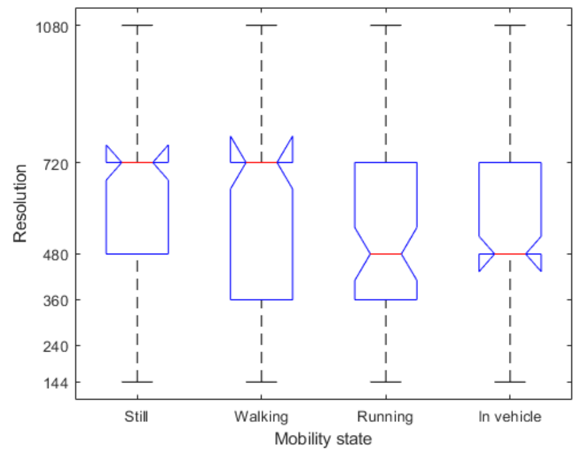

Figure 2: Boxplot depiction of the distribution of resolutions in which users completed watching videos in each activity state. Central line in each box: median; edges of the boxes: 25th and 75th percentiles of the distributions; Whiskers: most extreme data points not considered outliers.

5.1 Activity impact on user satisfaction

The distribution of the final resolutions in which users completed watching videos while in each of the activity states is depicted in Figure 2.

The results are in favor of the *RQ1* hypothesis that the activity context of the user impacts her perception of the video quality, and ultimately her satisfaction with the viewing experience. Thus, the data shows users are satisfied with higher resolutions when they watch the video while still. This is expected, since in such situations a user can fully concentrate on the video. The next highest average resolution is found in case the users are walking. In this state the distribution tails are more prominent, and while the median of distribution remains as high as it was with users being still (i.e. 720p), the 25^{th}-percentile of distribution is at 360p (c.f. 480p for still users). Riding as a passenger in a vehicle induces further tolerance towards lower resolutions, with the median of the acceptable resolution dropping to 480p, yet the distribution becomes more "compact" than it is the case with the distribution observed when the users are in the walking state. We suspect that the effect stems from varying abilities of our users to simultaneously walk and pay attention to the video. For some such multitasking may be a routine endeavor, thus, they require a higher resolution, whereas others might find it difficult to pay attention to the videos and regard the resolution unimportant. Finally, the running state leads to a further drop of resolution distribution, with the the 25^{th}-percentile at 360p and the median at 480p. This is not surprising since when engaged in a intense physical activity the user is less likely to be focused on the screen for anything but brief periods of time. By having to divide the attention between the video and the surroundings, the users find lower resolutions acceptable since they do not have the time to notice imperfectly rendered details.

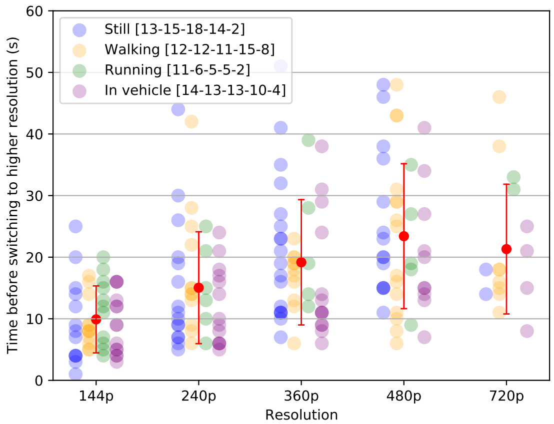

Figure 3: Time elapsed before users switching to a higher resolution for different activity states. A colored circle marks the moment in time the user increased resolution while watching the video. The red dot is the average represented with relation to two standard deviations (the red segment's extremities). In the legend, the values indicate the number of changes performed by the users in each of the resolutions.

To help understand user behavior in each activity state, Figure 3 shows all the changes in resolution performed by the users in the four activity states and the time elapsed before each change was made. In the legend the number of changes in each resolution for each mobility state can be observed. These results confirm that users had the lowest quality expectations (or highest tolerance to lower quality) while running, since in this state they made the lowest number of switches to higher resolution (the green circles on the chart). Also, the figure shows that the highest resolutions are encountered in the still state, which is also the state where the users were more likely to change the resolution for a higher one, confirming that when in this activity state, users have the highest quality expectations. Finally, irrespective of the physical activity, we observe a slight increase in the time to switch to a higher resolution, as opposed to a lower one. Since we note only the last resolution a user settles on, this confirms that the users complied with the protocol, i.e. switched the resolution only when not satisfied with the current one.

In addition to the above, we performed the statistical analysis of the results. A Kruskal-Wallis test shows that there is a significant difference in the acceptable resolution depending on the activity state: $H(3) = 14.139, p < 0.003$. This confirms the hypothesis that the activity state influences the user's video quality requirements. In order to assess the strength of the relationship between the investigated variables (context and resolution) we computed the effect size estimate for the Kruskal-Wallis result. Estimates of effect size allow the assessment of the strength of the relationship between the investigated variables, fostering an evaluation of the magnitude and importance of the result obtained [32]. In our case, we computed the eta-squared measure (η^2) using the following formula [11]:

$$\eta_H^2 = \frac{H - k + 1}{n - k} \tag{1}$$

where H is the value obtained in the Kruskal-Wallis test (the Kruskal-Wallis H-test statistic), k is the number of groups and n the total number of observations.

Eta-squared estimate assumes values from 0 to 1 and multiplied by 100 indicates the percentage of variance in the dependent variable explained by the independent variable [32]. For our experiment the computed eta-squared was 0.04; in the related scientific literature [11] eta-squared values less than 0.06 account for a small (weak) effect. Thus, while there is a statistically significant relationship between the activity state and the resolution, this relationship is shown to be weak.

5.2 Video content impact on user satisfaction

In light of the above statistical results, which indicate that other factors might influence a user's satisfaction with lower resolutions in different mobility states, we analyzed the impact of the video content on a user's receptivity to different video resolutions. The Kruskal-Wallis test shows that there is a statistically significant relationship between *the actual video content being played* and user's quality expectations (resolution found acceptable): $H(11) = 65.328, p < 0.001$. For evaluating the strength of this relationship we computed the same eta-squared effect size measure using Equation 1, with the result being 0.20. Based on the related literature [11], values higher than 0.14 indicate a large effect. This confirms RQ_2, i.e. that there is a strong relationship between the video content and the user's quality expectations when watching the video in specific mobility states.

To further assess the influence of video content on user satisfaction in different mobility states, for each video we computed two metrics: the average Spatial Information (SI) and the average Temporal Information (TI) indexes [18]. SI represents the spatial detail in a video frame (complexity) while TI relates to the amount of temporal changes in a video scene (motion), and the two metrics can be used for objective video quality prediction [13].

SI is based on the Sobel filter. Each video frame (luminance plane) at time n (F_n) is first filtered with the Sobel filter $[Sobel(F_n)]$. The standard deviation over the pixels (std_{space}) in each Sobel-filtered frame is calculated. This step is repeated for each frame in the video sequence and results in a time series of spatial information of the scene. The maximum value in the time series (max_{time}) is chosen to represent the spatial information content of the scene [18]. This process is described by the following equation:

$$SI = \max_{time} \left\{ std_{space} \left[Sobel(F_n) \right] \right\} \tag{2}$$

TI measures temporal changes (motion) in a sequence of video frames [18]. TI is based on motion differences between the pixels in the luminance plane of two consecutive frames $F_n(i, j)$ and $F_{n-1}(i, j)$, i.e., discrete time n and $n - 1$, at pixel position (i, j):

$$M_n(i, j) = F_n(i, j) - F_{n-1}(i, j) \tag{3}$$

TI is defined as the maximum value of the standard deviations obtained for the sequence of motion differences in the spatial domain [18]:

Table 1: Spatial information (SI) and Temporal information (TI) indices for the videos used in our study.

Video ID	Average SI	Average TI
1	55.51	19.45
2	117.26	26.58
3	52.59	7.77
4	61.69	15.39
5	59.32	29.42
6	29.05	11.52
7	56.65	9.72
8	46.14	8.81
9	39.77	11.41
10	80.04	19.03
11	126.88	13.85
12	36.38	8.60

Table 2: Pearson correlation coefficient between the final selected resolution and the average video SI/TI when a user is in a particular mobility state.

	Resolution vs. SI	Resolution vs. TI
Still	-0.05	0.21
Walking	0.31	0.54
Running	0.86	0.23
In vehicle	-0.28	-0.17

$$TI = \max_{time} \left\{ std_{space}[M_n(i,j)] \right\} \tag{4}$$

We computed the average SI and TI for all 12 videos, and the results are shown in Table 1. These numbers illustrate the heterogeneity in the video content with regard to their spatial and temporal features. To evaluate how this relates to the user quality perception of the videos in each mobility state we analyzed the link between the average resolution of the videos viewed in each state versus their SI and TI scores. We computed the Pearson correlation coefficient between the resolution and average SI and TI values for each mobility states, and the results are shown in Table 2.

The strongest link between the selected playback resolution and the SI is observed when a user is running (a Pearson correlation of 0.86). Running activity is of a particular interest to this study since it is the mobility state where one would expect the user's satisfaction requirements to drop the most. This strong link shows that when a user is engaged in an active physical activity (e.g. running), the required video quality and the spatial complexity of the video being played exhibit a strong positive linear correlation (i.e. the higher the spatial complexity of the video, the higher the required resolution). Out of the videos watched by the users while running, for videos 10 and 11, which have the highest SI scores, the users required the highest resolutions.

With regard to the link between average resolution and TI index in each mobility state, the Pearson correlation analysis indicates that a moderate positive linear correlation is present when the user is in mobility states requiring moderate physical movement, such as walking, where the coefficient is 0.54. While walking the users requested the highest average resolution for video number 5, which has the highest TI index among the videos watched while walking.

To better illustrate how the spatial and temporal characteristics of a video influence the user's quality perception in different activity states, Figure 4 shows how a selection of videos are perceived by the users when standing still vs. running (a subset of videos which users watched in both activity states: videos 6, 8, 9 and 11). The plot displays the average resolution for each video in each of the two activity states, and it is noticeable that videos 6, 8 and 9 show a similar behavior, i.e. they score similar average resolutions when still (between 650 and 550p) and their average resolutions drop considerably during running (between 350 and 500p). Video 11 however has a different behavior: while it also has an average resolution of about 650p while standing still, it does not decrease while running, on the contrary it slightly increases. The reason behind this phenomenon is that video 11 has the highest spatial information index among all 12 videos, and thus users perceptually require higher resolutions when running and viewing this video, compared to the other videos with lower spatial complexity.

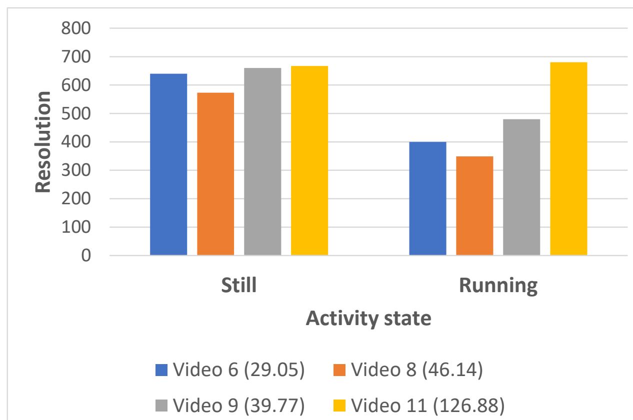

Figure 4: Average resolution in still vs. running for selected videos (and their corresponding SI values). While the general trend is that a running user is satisfied with a lower resolution than a walking user, a high SI video (Video 11) leads to higher resolution requirements when a user is running.

To statistically examine the interplay between the physical activity and the video content and its role on a user's expectations we created a linear regression model where the dependent variable is the resolution and the explanatory variables are the activity states, the spatial and temporal scores, and the cross-products representing the interaction effects between the activity states and the SI/TI scores. The results of this linear regression are presented in Table 3.

The regression shows the impact of a particular activity and the specific spatial and temporal complexity of a video on the required resolution. When users are walking or running, they require a lower resolution as indicated by the strong negative coefficients and low p-values; the effect is less pronounced when in-vehicle. The effects of the spatial and temporal complexity of a video on the required resolutions are not relevant by themselves (non-significant values

Table 3: Linear regression results for the resolution as the dependent variable.

Variable	Coefficient	p-value
Intercept	647.37	<2e-16***
walking	-247.89	0.02**
running	-394.62	<0.01***
in_vehicle	-65.83	0.48
spatial	0.31	0.73
temporal	-3.72	0.41
walking:spatial	0.25	0.87
running:spatial	2.93	0.05*
in_vehicle:spatial	-1.16	0.37
walking:temporal	13.15	0.05*
running:temporal	5.53	0.45
in_vehicle:temporal	6.03	0.30
Multiple R-squared: 0.1094		
Adjusted R-squared: 0.0705		

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

for "temporal" and "spatial"), only in interaction with certain activities. As such, high temporal information videos require higher resolution when a user is walking (as indicated by the low p-value of 0.05 and thus confirming the correlation illustrated in Table 2). In addition, higher spatial information videos require higher resolutions when a user is running (the low p-value of 0.05 confirms the correlation also illustrated in Table 2).

In addition to the above, however, the linear regression R-squared value is low, indicating that the model does not fully explain the data. This may stem from the limited data collected in our user study. More specifically, not all videos where watched in all activity states and not all videos were watched by all users. Furthermore, low R-squared value is likely an indicator that other contextual variables not considered in our study (e.g. outside noise, a user's interest in the video content, etc.) may impact quality expectations.

5.3 Discussion and potential energy savings

Our study reveals that the mobile user's quality expectations when watching videos on a mobile devices are influenced by both her mobility state and the video content (the spatial and temporal complexity of the video). Corroborated with the significant differences in energy consumption for different video playback resolutions, this confirms the feasibility of context- and content-driven approximate mobile computing for mobile multimedia apps and its potential to yield significant energy savings.

For an illustration of the potential energy savings we consider a hypothetical case of continuous video playback on a 3000mAh battery smartphone. In this scenario, watching videos in 720p vs. 1080p using hardware-enabled decoding (MPEG-4) would enable increasing battery life by 27% and for 360p vs. 1080p the increase is even higher, up to 37%, according to our measurements presented in Section 3. Even more significant savings can be achieved when using software-based decoding (WebM) – up to 47% longer battery life when watching a video in 360p vs. 1080p. The results answer our RQ_3 showing that significant energy savings can be achieved by lowering video resolution. While these numbers represent the upper bound of potential energy savings, certain devices, such as smart glasses, are often used in dynamic settings where a person is on a move and are likely to significantly benefit from context-aware video adaptation. However, for realistic energy savings we have to match our approach to a real use case.

We now provide a realistic energy savings computation for a particular use case – i.e. the scenario examined in our study. We first compute the total amount of energy required, if all users in this study watch all the videos at the highest possible resolution (1080p) with hardware decoding. The total adds up to 21629.2 J. Examining at the the collected dataset from this study we apply a model that accurately approximates (downgrades) the viewing resolution to the level found satisfying by the users. The total energy consumption now drops to 16608 J for all users to watch all the videos. This translates to 23.2% potential energy savings when applying the context-aware resolution adaptation model. Note that the savings achieved in another scenario would requires us to have the information of the users' activity and the accuracy expectations in that scenario.

6 RELATED WORK

Our research brings the concept of approximate computing (AC) to a particular application of mobile computing – mobile video playback. While relatively novel, the concept of AC has been examined on different levels of both software and hardware. The breadth of the proposed AC techniques prevents us from providing a detailed taxonomy in this article, thus we invite an interested reader to [22] where such a taxonomy is presented. Instead, in this section we focus on the role of video playback in the energy consumption of a mobile device, techniques for making video playback more efficient, as well as on the factors affecting the perception of mobile video.

6.1 Energy efficient mobile multimedia

The limited battery charge became the key pressing issue preventing further growth of mobile computing [25] and exacerbating the need for utilizing the available resources as efficiently as possible. Pang et al. conducted a survey of mobile app developers and users, and confirmed that energy-inefficient programming leads to negative app-store reviews and poor user satisfaction [24]. Among the services consuming the largest amount of energy in a mobile device, multimedia apps [12, 30] stand out, together with network traffic [35] and machine learning [20]. Yet, the high popularity of mobile multimedia makes addressing the energy consumption of such apps a pressing issues. A recent Atos study [6] reveals that mobile multimedia apps are the second most intensively used applications (based on average time spent by the user) and consequently also rank second in impact on the average daily energy consumption of a mobile device.

Solutions for reducing the energy consumption of mobile video apps include the work by Shin et al. [30], where the authors present an approach for reducing the energy consumption of RNC (random network coding) based media streaming applications on smartphones by manipulating the frequency controllers in the smartphone's operating system. Another solution proposed by Hu and

Cao [17] introduces an energy-aware CPU frequency scaling algorithm for mobile video streaming, which selects the CPU frequency that can achieve a balance between saving the data transmission energy and CPU energy. Ahmad et al. [5] developed a battery-aware rate adaptation for extending video streaming playback time which adapts to the appropriate bit rate to prolong the battery lifetime. An energy efficient video decoding for the Android operating system is proposed by Liang et al. [19], based on dynamic voltage and frequency scaling (DVFS). Hamzaoui et al. in their work [16] propose a measurement-based methodology for modeling the energy consumption of mobile devices and use video decoding tasks (both on-device and remote streaming) for the experimental power measurements.

Most of the above-mentioned energy-saving solutions focus on optimizations at the hardware and network layer for video streaming; by comparison, our approach is hardware-agnostic and adapts the video resolution according to the user's context, which influences his quality requirements. In addition, this context- and content-aware adaptation strategy has the advantage of being applicable for both network video streaming and on-device playback.

6.2 Mobile video quality perception

Dynamic viewing environment makes mobile video strikingly different from the conventional TV or Desktop PC viewing experience. Contextual factors, such as whether a viewer is indoor or outdoor, walking, running or riding a bus, and others, may change even during a single viewing session [34]. Research in this field identified several factors that influence mobile video quality perception, such as the display size, viewing distance from the display, environmental luminance, and physical activity of the user and showed that environment-aware video rate adaptation can enhance mobile video experience while reducing the bitrate requirement by an average of 30% [34]. Another study shows that in the mobile environment, sensory experience is a significant factor for enjoyment and engagement with the video as outside interruptions decrease the user's video quality experience on a mobile device [29]. This might be the reason for heavy tailed distributions of selected resolutions when users are walking or running, observed in our dataset. It is possible that, while generally too distracted to pay attention to fine video details, at certain occasions, users select a higher resolution to counter the effect of environmental disruptions.

The correlation between video content and user perceptual satisfaction is underlined by the existing research focused on this phenomena. Trestian et al. demonstrate a low spatial information video watched in low quality is likely to be found more acceptable/satisfying by the user than watching a high spatial and temporal complexity video the same quality [33]. The research findings also support the theory that one can expect significant differences in the user satisfaction at the same quality level depending on the particularities of the video. More specifically, the authors observed 20% average user satisfaction level difference between two videos watched in the lowest quality setting. We can see this in our study as well: from a subset of videos watched by users in "still" and "running" states, the video with a very high spatial complexity stands out as requiring a substantially higher resolution from the users when running, compared to the other videos in the subset which had lower SI scores (Figure 4). This indicates that the the video's spatial information feature influences the user's quality expectations in physically active states, such as running.

Song et al. identify a stronger relationship between acceptability and content type at a relatively low bitrate range of 200 - 400kbps [31]. The paper also concludes that the acceptability rate is influenced by the video content type: at higher resolutions, such as 480x320 pixels and 640x480 pixels, acceptability higher than 60% can be achieved, if the bitrate is greater than 300 Kbps for news, 400 Kbps for animation, movie, and music, and 800 Kbps for sports videos. The video content directly impacts the video's spatial and temporal information scores, e.g. animations usually have lower SI/TI, while sport videos have much higher scores. Our study confirms this: the videos with the highest SI and TI are either sport videos (basketball match – video 2, car dashboard camera recording – video 11 or body camera recording of mountain bike trail – video 5).

7 CONCLUSIONS, LIMITATIONS, AND FUTURE WORK

In this paper, inspired by the concept of approximate mobile computing, we investigated the foundations for dynamic energy-efficient context-aware video playback adaptation. Our measurements show that, as we have initially hypothesised, higher quality videos require more energy for decoding on mobile devices. To assess where exactly the opportunities for energy savings lie, we conducted a user experience study with 22 people, which involved watching videos during different physical activity states and monitoring the resolution users found satisfying in each case. This study revealed that the minimum video quality found acceptable by users is different for each physical activity; however, aside for the mobility state the user is engaged in when watching the video, the content, in particular its spatial and temporal complexity, also impacts the user quality requirements.

Diving in details, we calculated the Temporal and Spatial Information indexes of each video and studied how these factors correlate with the average resolution required for each video in each mobility state. This analysis revealed that videos exhibiting a high temporal complexity (high TI score) require higher resolution when the user is standing still or walking, while videos with high spatial complexity (high SI score) require higher resolution when the users is engaged in more intense physical activities, such as running.

While we confirm our findings through statistical tests, other factors that were not examined in the study may impact a user's perception of a mobile video playback. For instance, the outside brightness, screen quality, user fatigue and interests, and other factors could all steer a user towards desiring a higher or a lower playback resolution. In this study we control for most of these factors by strictly defining the indoor locations of the experiments, using the same devices throughout the study, and keeping the length of the videos to the minimum. Increasing the dimensionality of the considered context calls for more measurements, which in turns requires either more time from a single user or more users to be recruited. In future we plan to harness automatic context sensing integrated with our custom mobile video playback app in order to collect ecologically valid data over a large number of users.

Moreover, in our future research we plan on examining the other factors that could influence user quality expectations, such as the user's personality traits and internal motivation.

Our future research will also focus on implementing real-time AMC adaptation by integrating both context- and content-based approximations in a mobile multimedia application. The availability of the necessary contextual data, such as a user's physical activity, through common mobile programming APIs will allow us to proactively adapt video viewing resolution and quantify energy savings in real-world environments. Should these prove promising, we plan to explore other avenues for approximate mobile computing adaptation, such as 3D rendering, and augmented and virtual reality.

ACKNOWLEDGMENTS

The research presented in this paper was funded by Slovenian Research Agency (grants no. P2-0098 and N2-0136 "Bringing Resource Efficiency to Smartphones with Approximate Computing"). The authors would like to thank Ivan Majhen for his help with the energy measurements and the user study, and to all the volunteers who participated in the study.

REFERENCES

[1] Google/Ipsos report, 2016. Average Number of Apps Installed on Smartphones. https://www.thinkwithgoogle.com/marketing-strategies/app-and-mobile/average-number-of-apps-on-smartphones/

[2] Monsoon Solutions high voltage power monitor. http://msoon.github.io/powermonitor/HVPM.html

[3] NewPipe - A libre lightweight streaming frontend for Android. GitHub repository.. https://github.com/TeamNewPipe/NewPipe/

[4] VideoLAN VLC Media Player. https://www.videolan.org/vlc/

[5] Hasnah Ahmad, Navrati Saxena, Abhishek Roy, and Pradipta De. 2018. Battery-aware rate adaptation for extending video streaming playback time. *Multimedia Tools and Applications* 77, 18 (2018), 23877–23908.

[6] Atos and Greenspector Report. 2019. *Top 30 energy consumption of the world's most popular mobile apps*.

[7] Mark Beech. 2020. COVID-19 pushes up Internet use 70% and streaming more than 12%, first figures reveal. *Forbes* (2020). https://www.forbes.com/sites/markbeech/2020/03/25/covid-19-pushes-up-internet-use-70-streaming-more-than-12-first-figures-reveal

[8] Sebastian Bosse, Klaus-Robert Müller, Thomas Wiegand, and Wojciech Samek. 2016. Brain-Computer Interfacing for multimedia quality assessment. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 002834–002839.

[9] Aaron Carroll, Gernot Heiser, et al. 2010. An analysis of power consumption in a smartphone.. In *USENIX annual technical conference*, Vol. 14. Boston, MA, 21–21.

[10] CISCO. 2019. *Cisco visual networking index: global mobile data traffic forecast update, 2017–2022*. Technical Report. https://davidellis.ca/wp-content/uploads/2019/12/cisco-vni-mobile-data-traffic-feb-2019.pdf.

[11] Barry H Cohen. 2008. *Explaining psychological statistics*. John Wiley & Sons.

[12] John Elliot, Ah-Lian Kor, and Oluwafemi Ashola Omotosho. 2017. Energy consumption in smartphones: an investigation of battery and energy consumption of media related applications on Android smartphones. In *International SEEDS Conference*. http://eprints.leedsbeckett.ac.uk/4703/

[13] Majed Elwardy, Hans-Jürgen Zepernick, Thi My Chinh Chu, and Veronica Sundstedt. 2019. Objective perceptual video quality prediction using spatial and temporal information differences. In *2019 19th International Symposium on Communications and Information Technologies (ISCIT)*. IEEE, 436–441.

[14] Hadi Esmaeilzadeh, Adrian Sampson, Luis Ceze, and Doug Burger. 2012. Neural acceleration for general-purpose approximate programs. In *2012 45th Annual IEEE/ACM International Symposium on Microarchitecture*. IEEE, 449–460.

[15] YouTube for Press. 2020. *YouTube in numbers*. Technical Report. YouTube, https://www.youtube.com/intl/en-GB/about/press/.

[16] Khalil Ibrahim Hamzaoui, Mohammed Berrajaa, Mostafa Azizi, Giuseppe Lipari, and Pierre Boulet. 2020. Measurement-based methodology for modelling the energy consumption of mobile devices. *International Journal of Reasoning-based Intelligent Systems* 12, 1 (2020), 4–16.

[17] Wenjie Hu and Guohong Cao. 2017. Energy-aware CPU frequency scaling for mobile video streaming. In *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. 2314–2321.

[18] International Telecommunication Union (ITU). 2008. P.910 : Subjective video quality assessment methods for multimedia applications. *ITU-T Telecommunication standardization section of ITU* (2008).

[19] Wen-Yew Liang, Ming-Feng Chang, Yen-Lin Chen, and Chin-Feng Lai. 2013. Energy efficient video decoding for the android operating system. In *2013 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 344–345.

[20] Andrea McIntosh, Safwat Hassan, and Abram Hindle. 2019. What can android mobile app developers do about the energy consumption of machine learning? *Empirical Software Engineering* 24, 2 (2019), 562–601.

[21] Sasa Misailovic, Stelios Sidiroglou, Henry Hoffmann, and Martin Rinard. 2010. Quality of service profiling. In *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering-Volume 1*. 25–34.

[22] Sparsh Mittal. 2016. A survey of architectural techniques for near-threshold computing. *J. Emerg. Technol. Comput. Syst.* 12, 4, Article 46 (Dec. 2016), 26 pages. https://doi.org/10.1145/2821510

[23] Anush Krishna Moorthy, Lark Kwon Choi, Alan Conrad Bovik, and Gustavo De Veciana. 2012. Video quality assessment on mobile devices: Subjective, behavioral and objective studies. *IEEE Journal of Selected Topics in Signal Processing* 6, 6 (2012), 652–671.

[24] Candy Pang, Abram Hindle, Bram Adams, and Ahmed E Hassan. 2015. What do programmers know about software energy consumption? *IEEE Software* 33, 3 (2015), 83–89.

[25] Joseph A Paradiso and Thad Starner. 2005. Energy scavenging for mobile and wireless electronics. *IEEE Pervasive computing* 4, 1 (2005), 18–27.

[26] Veljko Pejović. 2019. Towards Approximate Mobile Computing. *GetMobile: Mobile Comp. and Comm.* 22, 4 (May 2019), 9–12. https://doi.org/10.1145/3325867.3325871

[27] Joelle Pineau, Philippe Vincent-Lamarre, Koustuv Sinha, Vincent Larivière, Alina Beygelzimer, Florence d'Alché Buc, Emily Fox, and Hugo Larochelle. 2020. Improving Reproducibility in Machine Learning Research (A Report from the NeurIPS 2019 Reproducibility Program). *arXiv preprint arXiv:2003.12206* (2020).

[28] Andreas Schuler and Gabriele Anderst-Kotsis. 2019. Examining the energy impact of sorting algorithms on Android: an empirical study. In *Proceedings of the 16th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services* (Houston, Texas) *(MobiQuitous '19)*. Association for Computing Machinery, New York, NY, USA, 404–413. https://doi.org/10.1145/3360774.3360808

[29] Eric W.K. See-To, Savvas Papagiannidis, and Vincent Cho. 2012. User experience on mobile video appreciation: How to engross users and to enhance their enjoyment in watching mobile video clips. *Technological Forecasting and Social Change* 79, 8 (2012), 1484 – 1494. https://doi.org/10.1016/j.techfore.2012.03.005

[30] Heehoon Shin and Joon-Sang Park. 2019. Reducing energy consumption of RNC based media streaming on smartphones via sampling. *Multimedia Tools and Applications* 78, 20 (2019), 28461–28475.

[31] Wei Song, Dian Tjondronegoro, and Michael Docherty. 2011. Saving bitrate vs. pleasing users: where is the break-even point in mobile video quality?. In *Proceedings of the 19th ACM International Conference on Multimedia* (Scottsdale, Arizona, USA) *(MM '11)*. Association for Computing Machinery, New York, NY, USA, 403–412. https://doi.org/10.1145/2072298.2072351

[32] Maciej Tomczak and Ewa Tomczak. 2014. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. (2014).

[33] Ramona Trestian, Arghir-Nicolae Moldovan, Cristina Hava Muntean, Olga Ormond, and Gabriel-Miro Muntean. 2012. Quality utility modelling for multimedia applications for Android mobile devices. In *IEEE international symposium on broadband multimedia systems and broadcasting*. IEEE, 1–6.

[34] Jingteng Xue and Chang Wen Chen. 2014. Mobile video perception: New insights and adaptation strategies. *IEEE journal of selected topics in signal processing* 8, 3 (2014), 390–401.

[35] Ming Yan, Chien Aun Chan, André F Gygax, Jinyao Yan, Leith Campbell, Ampalavanapillai Nirmalathas, and Christopher Leckie. 2019. Modeling the total energy consumption of mobile network services and applications. *Energies* 12, 1 (2019), 184.