Characterization of 360-degree Videos

Shahryar Afzal, Jiasi Chen, K. K. Ramakrishnan
University of California, Riverside

ABSTRACT

Online streaming of Virtual Reality and 360° videos is rapidly growing, as more and more major content providers and news outlets adopt the format to enrich the user experience. We characterize 360° videos by examining several thousand YouTube videos across more than a dozen categories. 360° videos, at first sight, seem to pose a challenge for the network to stream because of their substantially higher bit rates and larger number of resolutions. However, a careful examination of video characteristics reveals that there are significant opportunities for reducing the actual bit rate delivered to client devices based on the user's field of view. We study the bit rate and the motion in 360° videos, and compare them against regular videos by investigating several important metrics. We find that 360° videos are less variable in terms of bit rate, and have less motion than regular videos. Our expectation is that variability in the bit rates due to the motion of the camera in regular videos (or switching between cameras) is now translated to responsiveness requirements for end to end 360° streaming architectures.

CCS CONCEPTS

• **Networks** → *Network measurement*;

KEYWORDS

360° videos, Measurements, Video delivery

ACM Reference format:
Shahryar Afzal, Jiasi Chen, K. K. Ramakrishnan. 2017. Characterization of 360-degree Videos. In *Proceedings of VR/AR Network '17, Los Angeles, CA, USA, August 25, 2017,* 7 pages.
https://doi.org/10.1145/3097895.3097896

1 INTRODUCTION

New rich multimedia experiences are being introduced to the online user, including virtual reality (VR) and 360° video formats. VR video content is currently served by several major platforms, with both professional content providers such as NBC (who broadcast the 2016 Summer Olympics in VR), news outlets (e.g., CNN, New York Times) and user-generated content platforms such as Facebook and YouTube. [1]

[1]In the paper, we use the terms "360° " and "VR" interchangeably, as also the terms "regular" and "non-360° ".

VR/AR Network '17, August 25, 2017, Los Angeles, CA, USA

ACM ISBN 978-1-4503-5055-6/17/08…$15.00
https://doi.org/10.1145/3097895.3097896

In spite of the considerable work being done with regard to video delivery over communication networks, challenges continue to be presented as content creation and consumption evolve at a rapid pace. New forms of videos require new ways of transmitting the data efficiently so as to reduce the burden on the network and the user. While improved video encoding [1], bit rate adaptation [2], and multicast [3] have all previously been proposed and well-studied to reduce data transfers from video streaming, it is important to understand and address the challenges presented by emerging forms of video streaming. In this paper, we focus on enhancing our understanding of the characteristics of 360° videos and their requirements on streaming.

With the advent of 360° video, users finally have the opportunity to actively engage with the video. However, 360° videos typically consume more bandwidth than regular videos, since they require more data to cover all spatial directions. For example, YouTube VR recommends a resolution of 3840x2160 pixels, compared to usual 1080p ratio of 1920x1080 [4]. This four-fold increase in resolution will certainly contribute to the growth of video traffic, especially for wireless networks such as cellular and even WiFi. Current wireless networks may have trouble to meet this growing demand. While it is important to deliver high quality video to help achieve a reasonable level of viewer quality of experience (QoE), the high capacity requirements imposed by such rich video formats require us to rethink how we can satisfy user QoE while finding a practical and acceptable way of limiting bandwidth consumption.

Video delivery methods have evolved over time. Similarly, content providers have adapted the production and encoding of video based on both consumption patterns and the variation of available bandwidth to the user. In particular, the development of DASH enables the delivery of content at a quality most suitable to the available bandwidth (e.g., to minimize stalling of stream). Many of these enhancements come from an understanding of the video characteristics, delivery environment and user consumption, through measurements. We believe similar understanding of 360° video characteristics will greatly enable the evolution of efficient protocols over both wired Internet and wireless channels. These new video streams have inherent characteristics that are related to human factor considerations. For example, a user may not view a VR video for as long of a duration as a regular video, due to feeling uneasy or experiencing motion-sickness. So, understanding how 360° content has been created and their duration is also useful. Additionally, possibly due to expected bandwidth limitations or camera capabilities, the resolutions at which 360° video is encoded is also different than what is being used for regular videos. While these are expected challenges, we also examine the bit rate and motion in the videos in greater detail to see if there are ways to mitigate these challenges. In particular, we look at the variability in the bit rate and the possible underlying cause for that variability (e.g., motion) of 360° videos in contrast to regular videos. Understanding all these aspects will help in future design and continual improvement of video delivery mechanisms.

The contributions of this paper are as follows:

- We collect a dataset of YouTube 360° videos and report on their aggregate statistics, including bit rates, resolutions, and durations.
- We develop a technique to estimate the effective resolution of 360° videos based on the typical "field of view" of a VR headset. This helps us to understand the strategies that may be available for mitigating the bandwidth requirements for 360° videos.
- We compare the intra-video variability of bit rates of 360° and regular videos. Our main observation is that the bit rate variability of the 360° videos is considerably lower than for the regular videos. We hypothesize that the underlying cause is the reduced motion in 360° videos, and examine the motion characteristics of several representative 360° and regular videos.

2 CURRENT PRACTICE

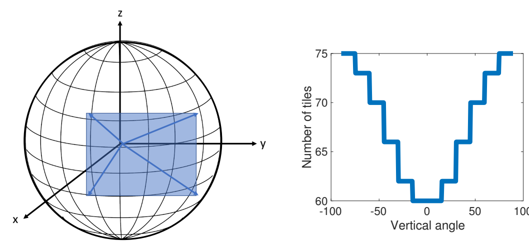
Background: Videos are recorded using specialized 360° recording setups, consisting of between 2-6 individual cameras (e.g., 6 GoPros arranged in a cubic arrangement). The videos are synchronized in time and stitched in space using specialized software to create a spherical view. The spherical coordinates are converted to rectangular format (similar to the process of flattening the globe of the Earth into a 2-dimensional map), then compressed to H.264 format or another format of the user's choice. Once processing is complete, the user uploads the video to the server, optionally adding metadata to alert the server of the 360° format and that special processing is required. The server may perform additional processing to convert to video to a streaming format (e.g., MPEG-DASH). The video is then ready to stream to viewers using the content provider's choice of streaming protocol. Viewers can play back the video using a custom video player provided by the content provider. The user can rotate the field of view of the video using her mouse/finger (if she is using a web browser/mobile app) or through head motions (if she is wearing a VR headset).

Current Practice: We examined packet traces to see how several popular content providers (Facebook, YouTube) stream 360° videos in desktop players. We find that Facebook treats 360° video streaming as a progressive file transfer, transferring the entire video (using byte ranges), no matter where the user is looking at in the scene. Although Facebook has publicly described spatial video chunking [5], we were unable to find evidence of this in the desktop 360° video player, as the byte range requests increase sequentially even if we rotate the view. This both wastes network resources and may result in high costs for the user in the current usage-based charging environment, especially for a user who is receiving video data that she does not even view. NBC uses a similar tactic to Facebook where only one file is transferred progressively.

YouTube treats 360° videos similar to regular MPEG-DASH videos: it encodes these videos at different resolutions and streams the video chunks at the quality appropriate for the available network bandwidth. CNN also uses similar DASH-like chunking at different resolution encodings.

Category	# of Videos	Category	# of Videos
All	2285	Roller coaster	325
Animals	216	Scenery	315
Cartoon	197	Shark	24
Concert	67	Skydiving	70
Documentary	122	Space	126
Driving	176	Sports	139
Horror	180	Video game	197
Movie trailer	131		

Table 1: YouTube video dataset



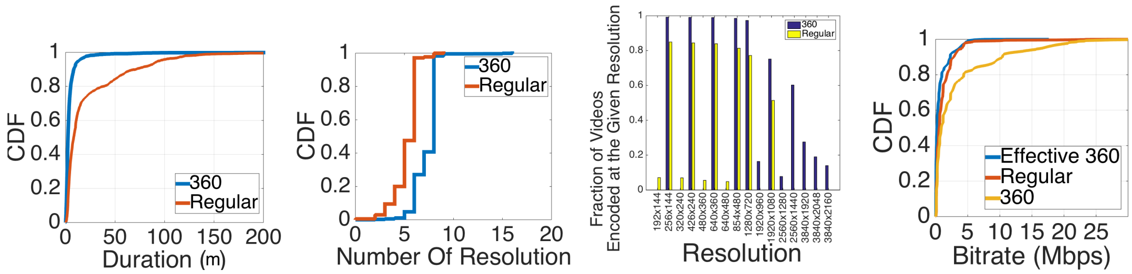
(a) (b)

Figure 1: Number of tiles downloaded as user's head rotation changes.

3 YOUTUBE MEASUREMENTS

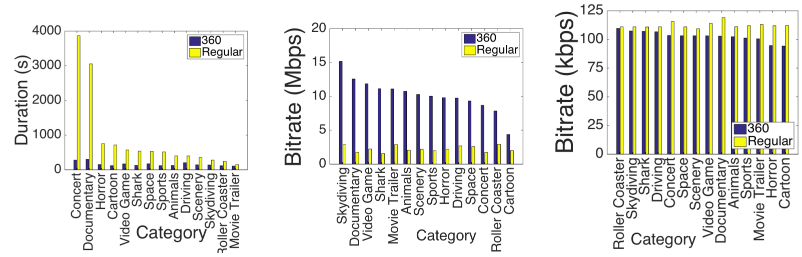
In this section, we describe our measurements of YouTube 360° videos in order to understand their characteristics in-depth. We first selected a number of videos from YouTube, and then classified them based on their genre, using the following methodology. We obtained an initial set of 360° videos by searching YouTube using the keyword "360", and including a filter for 360° videos. This gave us approximately 700 of the 360° videos. To expand this set of videos and find their non-360° counterparts, we categorize the videos, and use the category names as new search keywords, resulting in a total of 4570 videos (2285 regular and 2285 360° Videos). Specifically, we calculate the term frequency across all the video titles, removing filler words such as "the" and "and") and map them to the appropriate category (e.g., basketball and soccer map to "Sports" category). We stop collecting videos from a category once the search results are deemed less relevant (according to YouTube's relevance ranking) (i.e., after the n^{th} page of search results, with n determined qualitatively per-category). The number of videos we collected in each category is shown in Table 1.

Summary statistics: A simple first question that arises is: how does the duration of 360° videos differ from non-360° videos? We expect that 360° videos, being a relatively new format, will mainly be used by early adopters who are still determining the best way to use the new format, and therefore are still experimenting with shorter clips. In addition, longer videos are also unlikely due to users experiencing fatigue with current VR headsets. To verify this, we plot the CDF of the duration of the 360° and non-360° videos in Fig. 2a. The default soft limit on the uploaded video duration is 15 minutes, but users can extend the maximum allowable duration by verifying their account information. We observe that 360° videos have a shorter duration than non-360° videos (median duration 143s vs 490.5s, respectively), and also have a shorter tail, matching our expectation.



(a) Duration (b) Number of Resolutions (c) Resolution Distribution (d) Bit rates

Figure 2: Summary statistics of 360° versus non-360° videos.



	Median	90th %ile
Driving 360	7558	16273
Driving reg	2031.8	4627.2
Horror 360	4185.4	14078
Horror reg	1648.1	3848.2
Roller 360	5516.5	20115
Roller reg	2377.3	5681.9

(a) Duration across categories (b) Bit rate of Maximum Resolution (c) Bit rate of Minimum Resolution (d) Median and 90th %ile Bit rates (kbps)

Figure 3: Per-category measurements of 360° vs non-360° videos.

We then examine the resolution of 360° videos. YouTube recommends a minimum resolution of 7168x3584 for uploads [4]. YouTube transcodes the videos into various preset resolutions, e.g., to support adaptive streaming. We wish to understand what are the typical preset resolutions, and how they differ from the resolutions of regular videos. By examining packet traces, we found that the set of the encodings available for desktops and mobile devices (Android YouTube app) were identical in the vast majority of cases, so we focus on the videos encoded for desktops for the remainder of this section. In Fig. 2c, we plot the fraction of the total set of videos which are encoded at each resolution (filtering out uncommon resolutions with <100 videos). We observe that 360° videos tend to have higher maximum resolutions than regular videos, but similar minimum resolutions. The minimum resolution we observe across all videos is 82x144, and the maximum we observe is 8192x8192. Also, we also observe that a small number of videos (55) have a square resolution (e.g., 4096 by 4096), which is due to the left eye and right eye being encoded separately and stacked on top of each other (steroscopic encoding). 360° videos also tend to have a larger number of distinct resolutions, as shown in Fig. 2b, where we plot the CDF of the number of resolutions encoded for each video.

Effective resolution based on field of view: We next wish to compare the bit rates of 360° and regular videos. However, note that only a portion of the 360° video is seen by the user, since VR headsets have a limited field of view. Therefore, comparing the advertised bit rate of 360° videos with the advertised bit rate of regular videos may be an unfair comparison. To address this, we introduce the notion of "effective resolution" of the 360° videos: namely, based on a typical field-of-view of a VR headset (likely to be similar, if not slightly higher than with a Flash player on a browser), we calculate the fraction of the video pixels viewed. This calculation is based on YouTube's equirectangular projection, which converts the rectangular video streamed by the server to a spherical representation that is viewed by the user in the VR headset.

As an example, assume that the VR headset has a 110° horizontal field of view and a 90° vertical field of view. A naive calculation of the fraction of the video viewed would be: $\frac{90^\circ}{180^\circ} \times \frac{110^\circ}{360^\circ} = 15\%$. To improve this estimate, we can take into account the user's vertical head position. When the user faces upwards or downwards, the equirectangular projection results in a larger number of tiles being downloaded, as in Fig. 1a. Assuming that the user typically looks ahead, left, right or behind, but occasionally glances upwards/downwards, we can calculate the average fraction of tiles downloaded. The details can be found in a technical report [6]. In Fig. 1b, we plot the number of tiles downloaded as the user rotates her head vertically (still assuming a 110° horizontal and a 90° vertical field of view, and that the video is divided into 288 tiles at a 15° angle. .) The resulting percentage of tiles downloaded is thus slightly higher than the naive calculation, coming to 22% (64 tiles in Fig. 1b).

Using this definition of effective resolution, in Fig. 2d, we plot the CDF of the bit rate for 360° and regular videos. Clearly, the bit rates of the 360° videos are generally higher. However, by factoring in the effective resolution experienced by the user, we can see that the bit rate of the 360° videos experienced by the user in her limited field of view is similar to that of the regular videos.

Per-category comparison of 360° videos: We next discuss the characteristics of regular and 360° videos across different categories. We first measure the median duration of videos in different categories (Fig. 3a). The categories are sorted by descending duration of regular videos. While regular videos have long duration in certain expected categories (namely concerts and documentaries), most of the 360° videos are relatively short. This may be because 360° is a new medium, and durations may increase in the future and become similar to the regular videos, as 360° technology and cameras are more widely adopted.

We also examine the distribution of the video bit rates across categories. In Fig. 3c, we plot the bit rate of the minimum resolution available for each video, and in Fig. 3b, we plot the bit rate of the maximum resolution available for each video. In both plots, the categories are sorted in descending order of the bit rate of the 360° videos. The bit rates of the minimum resolution encodings tend to agree with our intuitive categorization; e.g., roller coasters and skydiving, with lots of motion and complex scenery, have higher minimum bit rates than horror and cartoons. For regular videos, however, the variance of the bit rate across categories is less (the heights of the bars are similar).

For the bit rates of the maximum resolution, the results are slightly harder to interpret. Firstly, we observe in Fig. 3b that the 360° bit rates are generally much higher than the bit rates of regular videos, as expected. The categories with the highest bit rates are skydiving (expected due to high motion and changing scenes) and documentary (unexpected!). Upon digging deeper, our qualitative impression is that more 360° documentaries are topics with active subjects, such as war and refugees, while the regular video documentaries tend to be more nature or biography focused. On the other hand, the categories with the lowest bit rates at the maximum resolution are cartoons (expected) and roller coasters (unexpected!). The roller coasters are particularly unexpected since these videos tend to have lots of motion as the camera moves on the roller coaster track, and even the regular roller coaster videos show higher bit rate. Overall, the 360° videos tend to exhibit more variability across categories (e.g., the skydiving category has $\sim 3\times$ the bit rate of the cartoons), while variability of the regular videos across categories is much smaller.

4 BIT RATES OF 360° & REGULAR VIDEOS

We now delve deeper into the characteristics of YouTube 360° videos by analyzing the intra-video bit rate, and compare with the bit rates of regular videos, to understand the network load they impose for transporting the video. Our methodology is as follows. For the highest resolution version of each video, we look at a 10-minute segment in the middle of the video, and compute the per-second bit rate, which we believe is representative of the bit rate of the video. We first report on the aggregate bit rate statistics across a few of the categories (we look at 30-200 videos of each category) for both 360° and regular videos. Table 3d shows the median and 90th percentile for the three categories, for the two classes of videos. We observe, as expected, the median bit rate (in kbps) for the 360° videos is much higher (by a factor of 2-4) than regular videos. Categories like driving and roller coasters have a higher median rate than for the horror category.

We then look at the bit rate variability/burstiness across the videos for each of the categories. To account for the different encoding rates of each video, we normalize the per-second bit rate by the average bit rate (over the 10-minute interval). Fig. 5 shows the box plot of the bit rates. The (normalized) median, as expected, for 360° and regular videos is similar. Within a particular category (e.g., Driving), the bit rate variability for the regular videos is much higher compared to the 360° videos. This is especially true for the outliers bit rates (the red crosses). For example, in the driving category, the maximum is more than 17 times the average for the regular videos, while it is less than 7 times the average for the 360° videos. Thus, the variability for the 360° videos is lower than for regular videos. This is important, especially if the network needs to provision close to peak capacity (e.g., cellular networks).

We zoom in on the detail of the percentiles in Fig. 4. The 75th and 90th percentile for the majority of categories are indeed higher for the regular videos. This matches with the intuition that we have that the camera motion (panning the view to capture the overall scene) with the regular videos results in more motion and higher variability in the bit rate. In particular, the "high motion" categories such as driving, skydiving, sports, rollercoaster are all less variable for 360° videos compared to their regular counterparts. The exceptions are the cartoon, video game, horror, and movie trailer categories, where the 360° videos have higher variability. While we cannot definitively pinpoint the root cause, for cartoons and video games, we believe that artists' creating/rendering of a scene may influence how much motion there is. Similarly, for the horror and movie trailer genres, we speculate that the reason is that the movie creator induces variability to have the desired effect (e.g., selecting short scenes of high action in a trailer to catch the viewer's attention).

We expect that with regular videos, the camera itself has to move its field of view to capture the entire scene. This means that even within a fixed, unchanging environment, the movement of the camera results in motion which has to be represented properly and transmitted. In fact, consider for example the video taping of a movie with multiple regular video cameras with a limited field of view. To capture the overall scene (say of multiple people having a conversation), the video will have to frequently switch from one camera to another. On the other hand, with 360° video, the camera does not move (it is one overall scene captured with multiple lenses and stitched together). Thus, only the inherent movement in scene has to be captured. This inherent motion may be what causes the bit rate to change in 360° videos, but it appears to be less than if the camera were to move or if the video shifted from one camera to another. We believe that if what we see here is observed to be generally true across more (if not all) of the 360° videos, it has an important consequence for transporting such videos. Provisioning can be easier as the variability in the videos is lower. The challenge then becomes finding techniques to reduce the overall/average bandwidth requirement for 360° videos based on such concepts as focusing on the field of view of the viewer and encoding or transcoding the remaining regions (outside the field of view) at a different, lower resolution.

This benefit of the reduced variability of the bit rates for the network comes at a price, however. Instead of the camera moving to capture activities of interest in the regular videos, now the user (e.g., a viewer with a VR headset or using a mouse while viewing in a browser) brings that movement to the time of viewing the video. Therefore, this requires the system to be much more responsive to the user requests for changing the field of view - and latency now becomes a much more important criterion for the network.

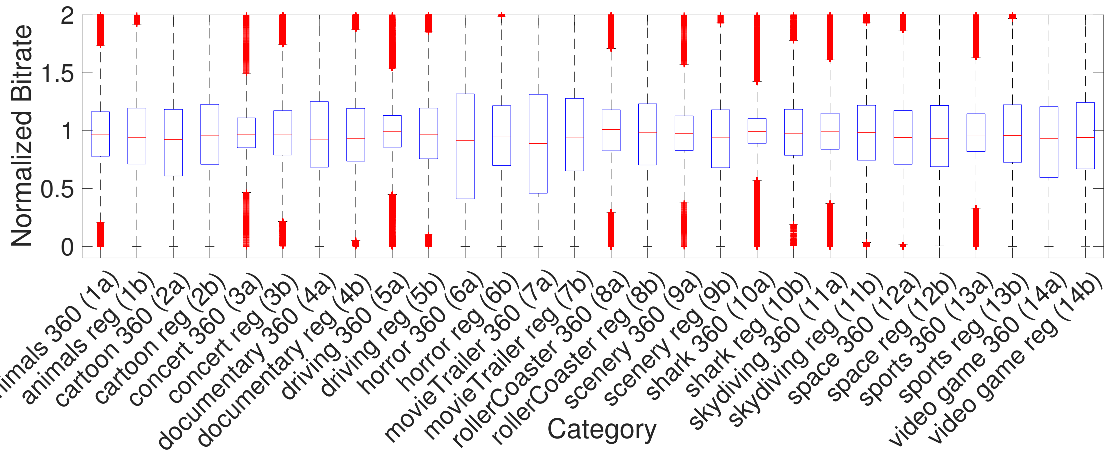


Figure 4: Variable bit rate across video categories: "Zoomed-in" from Fig. 5 for detail.

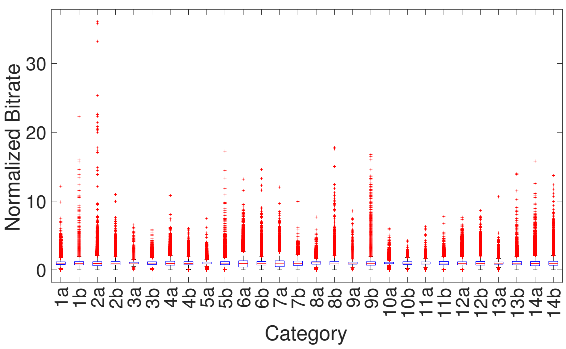


Figure 5: Variable bit rate across video categories

Motion characteristics: We next investigate the motion characteristics of 360° versus regular videos. We wish to understand whether the motion information can explain the higher bit rate variability, as previously hypothesized. During recording of regular videos, the camera can move in 3D space, as well as rotate (yaw, pitch, and roll). When these videos are encoded, the motion vectors must capture both the direction and magnitude of these 3D translations as well as the rotations. 360° videos, on the other hand, are invariant to rotation, by definition, because they capture all 360° information from the current 3D position. Therefore, we hypothesize that 360° videos will have smaller motion vectors than regular videos, since their motion vectors only need to encode 3D translations, and not rotations. On the other hand, the motion vectors of regular videos will also change more rapidly over time, since any movement of the camera results in more non-zero motion vectors.

Our methodology is as follows. Using `ffprobe` [7], we extract the motion vectors from two representative videos in low-motion and high-motion categories: horror and skydiving. We compute the magnitude of the motion vector for each macroblock at position (x, y) in a frame, take the average across x or y, and plot the magnitude over time.

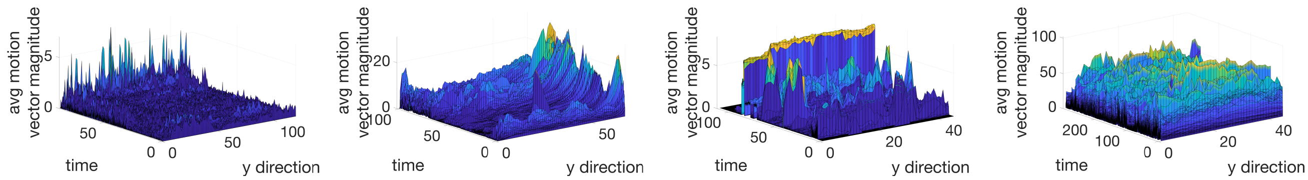
The results are shown in Fig. 6 macroblocks at different y locations. In the spatial domain, the 360° skydiving video has some spatial locality (at $y = 50$), while the regular skydiving video does not. We believe that this is because a regular video camera used in skydiving is apt to rotate randomly due to loss of user control, and so experiences a high degree of motion across all regions of the video, while a 360° video camera is invariant to rotation. In the case of the 360° skydiving video, $y = 50$ corresponds to the ground, which is the changing part of the scene. The horror videos do not show spatial locality because both the 360° video and the regular video are relatively stationary, similar to a surveillance camera.

Now consider the change in motion over time in Fig. 6. Comparing the 360° horror and skydiving videos, we can see that the average motion vector magnitude is higher for the high-motion skydiving video, which makes sense. Comparing the 360° and the regular videos, we can see that per-category, the 360° videos have lower motion vector magnitude. The regular videos also seem to exhibit greater variability over time, while the 360° videos are more uniform. To compare the bit rates, in Fig. 7, we plot the per-frame bit rate for the same four videos evaluated in Fig. 6. Comparing the 360° horror and skydiving videos, the bit rate variability is higher for the high-motion skydiving video. Comparing regular and 360° videos, we can see that the mean bit rate of the regular videos varies more over time, while the 360° videos seem to have a more constant mean. This is consistent with our observations above concerning the motion vector variability across time. We plan to look at a larger number of videos, across more categories, to confirm this hypothesis.

5 RELATED WORK

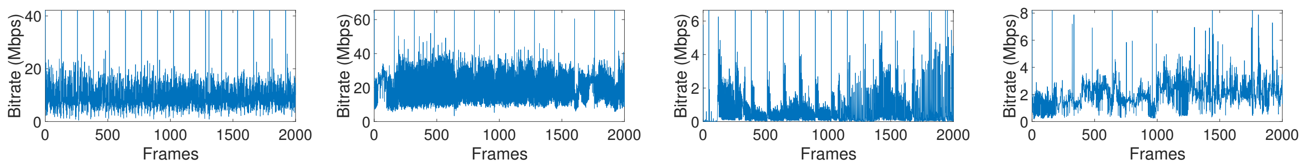
Video compression: Recent versions of MPEG (e.g., H.264/MPEG-4, H.265) enable regions of interest to be encoded at a higher quality than background objects [8]. Specifically, each frame can be divided into slices (spatial regions), and each slice encoded as I, P, or B. Our work is complementary to this and tries to understand the differences between spatial variations in regular and 360° videos, which encoders could then leverage.

Video adaptation: Recent work has focused on adapting the video quality based on the estimated network bandwidth [2, 9, 10]. Our goal is complementary: to understand how much bit rate variability a 360° video has, and therefore how high of a bit rate can be



(a) Horror 360° **(b) Skydiving 360°** **(c) Horror regular** **(d) Skydiving regular**

Figure 6: Motion vector magnitude over time.



(a) Horror 360° **(b) Skydiving 360°** **(c) Horror regular** **(d) Skydiving regular**

Figure 7: Variable bit rate across time.

supplied under the given bandwidth. Recently, spatially-adaptive chunk selection schemes have also been proposed [11, 12]. Our work can help guide these adaptation schemes by understanding the bit rate variability across spatial chunks.

Adaptation for 3D videos based on object QoE, redundant information, and user perception has been studied [13–15]. The 360° videos we consider in this work are omnidirectional 2D videos, as opposed to 3D videos with a fixed field of view in which the camera view must shift to show different parts of the same scene. The ideas of prioritizing objects in 3D and teleimmersive video, however, could be useful in our future work of performing in-network manipulation of 360° videos.

360° videos: [16] studies characteristics of YouTube videos, but does not consider more recent 360° videos. [17] performs a measurement study of the head motions of a small set of users viewing 360° videos, and [18, 19] study 360° projection schemes. Our work aims to add to this body of measurement studies by understanding motion and bit rate characteristics of 360° videos.

6 CONCLUSIONS

This paper characterized 360° videos from the point of view of network streaming, and compared them to regular videos that have been the popular media format until now. Our comparison shows that 360° videos have substantially higher bit rates and larger number of resolutions; however, after more careful examination, we find that the bit rates for the 360° videos have less variability than the regular videos, which can be highly beneficial for the network due to the network provisioning for the peak rates.

To explain lower bit rate variability, we demonstrated that the average motion for the 360° video is less than that for a comparable regular video. We believe that this is because the motion described in a 360° video is that which is inherently in the scene, rather than the rotation or panning of the camera in space.This implies that the panning now occurs at the time of user viewing the video. Thus, the new requirement on the network is that it needs to be more responsive to the user changing the field of view. We believe these aspects have deep implications on networked video streaming systems for both capacity and latency requirements.

We gratefully acknowledge funding from Huawei Technologies Co., Ltd. Innovation Research Program.

REFERENCES

[1] Alan C. Bovik. *The Essential Guide to Video Processing*. Academic Press, 2nd edition, 2009.
[2] Mpeg-dash. http://dashif.org/.
[3] Jiasi Chen, Mung Chiang, Jeffrey Erman, Guangzhi Li, K.K. Ramakrishnan, and Rakesh K Sinha. Fair and optimal resource allocation for LTE multicast (eMBMS): Group partitioning and dynamics. *IEEE INFOCOM*, 2015.
[4] Upload 360-degree videos - youtube help - google help. http://support.google.com/youtube/answer/6178631.
[5] Next-generation video encoding techniques for 360 video. https://code.facebook.com/posts/1126354007399553/next-generation-video-encoding-techniques-for-360-video-and-vr/.
[6] Characteristics of 360-degree videos (technical report). http://www.cs.ucr.edu/~jiasi/360_tr.pdf.
[7] Vadim Kantorov and Ivan Laptev. Efficient feature extraction, encoding and classification for action recognition. In *IEEE CVPR*, 2014.
[8] Thomas Wiegand, Gary J Sullivan, Gisle Bjontegaard, and Ajay Luthra. Overview of the h. 264/avc video coding standard. *IEEE Trans. Circ. Sys. for Video Tech.*, 13(7):560–576, 2003.
[9] Junchen Jiang, Vyas Sekar, and Hui Zhang. Improving Fairness, Efficiency, and Stability in HTTP-based Adaptive Video Streaming. *ACM CoNEXT*, 2012.
[10] Te-Yuan Huang, Ramesh Johari, Nick McKeown, Matthew Trunnell, and Mark Watson. A buffer-based approach to rate adaptation: Evidence from a large video streaming service. *ACM SIGCOMM*, 2014.
[11] Feng Qian, Lusheng Ji, Bo Han, and Vijay Gopalakrishnan. Optimizing 360 video delivery over cellular networks. In *All Things Cellular Workshop*, 2016.
[12] Xavier Corbillon, Alisa Devlic, Gwendal Simon, and Jacob Chakareski. Viewport-adaptive navigable 360-degree video delivery. *CoRR*, abs/1609.08042, 2016.
[13] Pengye Xia and K. Nahrstedt. Object-level bandwidth adaptation framework for 3d tele-immersive system. *2013 IEEE ICME*, pages 1–6, July 2013.
[14] Zhenyu Yang, Bin Yu, Klara Nahrstedt, and Ruzena Bajscy. A multi-stream adaptation framework for bandwidth management in 3d tele-immersion. *ACM NOSSDAV*, 2006.
[15] Wanmin Wu, Ahsan Arefin, Gregorij Kurillo, Pooja Agarwal, Klara Nahrstedt, and Ruzena Bajcsy. Color-plus-depth level-of-detail in 3d tele-immersive video: A psychophysical approach. *ACM MMSys*, 2011.
[16] Michael Zink, Kyoungwon Suh, Yu Gu, and Jim Kurose. Characteristics of youtube network traffic at a campus network - measurements, models, and implications. *Computer Networks*, 53(4):501–514, 2009.
[17] Yanan Bao, Huasen Wu, Albara Ah Ramli, Bradley Wang, and Xin Liu. Viewing 360 degree videos: Motion prediction and bandwidth optimization (poster). *IEEE ICNP*, 2016.
[18] Chao Zhou, Zhenhua Li, and Yao Liu. A measurement study of oculus 360 degree video streaming. *ACM MMSys*, 2017.
[19] M. Yu, H. Lakshman, and B. Girod. A framework to evaluate omnidirectional video coding schemes. In *2015 IEEE International Symposium on Mixed and Augmented Reality*, pages 31–36, Sept 2015.

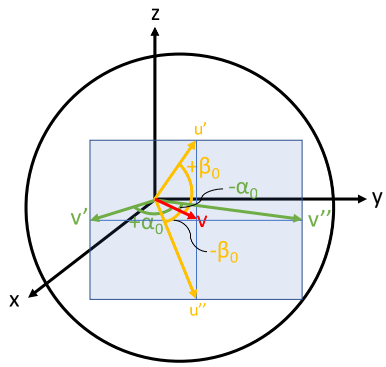


Figure 8: Calculating the number of tiles needed to cover the shaded area.

A CALCULATING EFFECTIVE RESOLUTION OF 360° VIDEOS

Consider Fig. 8. Let the spherical coordinates be denoted by (θ, ϕ), and the 3D Cartesian coordinates be denoted by (x, y, z). Suppose the center of the user's field-of-view is v, which is given in spherical coordinates (yaw and pitch) from the head tracking software. The user's field-of-view extends by α_0° horizontally and β_0° vertically from v. The number of tiles viewed is determined by the boundary of the user's field-of-view. Therefore, we need to determine the (x, y, z) coordinates of the boundary, perform a 2D projection (e.g., equirectangular), and count the number of tiles covered by the projected area.

Without loss of generality, assume that $v = (\theta, 0^\circ)$. Let v', v'', u', u'' be points on the boundary of the field-of-view, as shown in Fig. 8, and let $\beta = \frac{\pi}{2} - \theta$.

To convert to Cartesian coordinates, we have that:

$$x = \cos\alpha_0 \cos\beta \tag{1}$$

$$y = \sin\alpha_0 \tag{2}$$

$$z = \sin\beta \tag{3}$$

From there, to convert to spherical coordinates, we have that:

$$\theta = \arccos z = \frac{\pi}{2} - \beta \tag{4}$$

$$\phi = \arctan\frac{y}{x} = \arctan\frac{\tan\alpha_0}{\cos\beta} \tag{5}$$

Now let us consider the equirectangular projection. Let the coordinate system of the 2D video be denoted by (X, Y). The equirectangular projection is defined as $X = \phi, Y = \theta$. Let the angular width of each tile be G_X, G_Y in the horizontal and vertical directions, respectively. We assume that there is no roll (rotation) in the user's field of view (we leave this calculation to future work). Then the number of tiles of video spanned by the horizontal field-of-view $[-\alpha_0, +\alpha_0]$, centered at v, can be found from the difference of the X-coordinates (ϕ=coordinates) of v' and v'':

$$N_X(\beta) = \left\lceil \frac{2\arctan\frac{\tan\alpha_0}{\cos\beta}}{G_X} \right\rceil \tag{6}$$

The number of tiles spanned by the vertical field-of-view $[-\beta_0, +\beta_0]$, centered at v, can be found from the difference of the Y-coordinates (θ-coordinates) of u' and u'':

$$N_Y = \left\lceil \frac{2\beta_0}{G_Y} \right\rceil \tag{7}$$

To calculate the number of tiles, we can scan over β between vertical tile i and $i+1$ and count the maximum number of tiles needed:

$$N(\beta) = \sum_{i=0}^{N_Y-1} \max_{B(i)\le\beta\le B(i+1)} N_X(\alpha_0, \beta) \tag{8}$$

$$B(i) = \left(\left\lfloor \frac{\beta_0}{G_Y} \right\rfloor + i \right) G_Y \tag{9}$$

where $B(i)$ is defined as the lower angle of the i^{th} vertical tile.

Finally, to get the average number of tiles used, we assume a probability distribution function on the user's pitch; i.e., we assume knowledge of PDF f_θ. Then the average number of tiles viewed is:

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} N\left(\frac{\pi}{2} - \theta\right) f_\theta(\theta) d\theta \tag{10}$$

We set $\alpha_0 = \frac{110^\circ}{2}$ and $\beta_0 = \frac{90^\circ}{2}$, which are typical horizontal and vertical field-of-views, respectively, for VR headsets [11].