ANALYZING THE SPATIAL QUALITY OF INTERNET STREAMING VIDEO

Amy R. Reibman, Subhabrata Sen, and Jacobus Van der Merwe

AT&T Labs - Research
{amy,sen,kobus}@research.att.com

ABSTRACT

The quality of video over the Internet is determined both by the original encoded video quality and by any impairments caused by the network. In this paper, we consider the first of these factors. We observe streaming video sessions over a one-week period at a cable headend of a large cable provider. All traffic flows for the widely used Windows Media streaming application were collected. We then apply several no-reference pixel-based blocking and blurring quality metrics to the decoded video. We present a statistical characterization of the spatial quality, considering the impact of spatial resolution, temporal frame-rate, bit-rate, and compression algorithm.

1. INTRODUCTION

As broadband access connectivity becomes more prevalent, more users are able to stream video over the Internet. This trend is in turn propelling the use of Internet streaming video by a host of applications such distance education, e-commerce, corporate presentations, sports, entertainment, and news. The wider audience brings greater potential revenue – for content providers, hosting centers, and network service providers.

Network service providers would like an online performance monitoring tool in order to create a real-time understanding of video streaming applications. This can aid in monitoring compliance of service-level agreements (SLAs) between Internet Service Providers (ISPs), hosting centers, and content providers; alert operators to potential performance problems; and help in root-cause analysis and debugging.

In this paper, we consider the problem of predicting application-level video quality by monitoring video packets within the network. Our final goal is to obtain accurate, real-time, end-to-end estimates of per-stream quality for all videos observed from our in-network monitoring location. Our previous work focuses on identifying when the encoded bitstream has been modified (by either server or network) enough to impair the displayed quality [1]. In fact, the vast majority (over 86%) of videos in our observation set have no network-induced impairments. Therefore, in this work, we focus on characterizing the quality of the original encoded videos, that is, the quality without network impairments. Moreover, while video quality depends on both spatial *and* temporal quality, we focus here only on the spatial quality of the still frames that comprise the video.

In-network monitoring of video quality imposes structural and processing constraints. First, Full-Reference (FR) [2] or Reduced-Reference (RR) [3] metrics are not useful, because the original video is never available. Further, No-Reference Pixel (NR-P) methods that require decoded video [4, 5] may not be practical, because they require a decoder inside the network for every video stream, and may not be viable for proprietary compression algorithms or encrypted video payloads. No-Reference Bitstream (NR-B) methods do not require a complete decoder for every stream [6]. Full-Parse NR-B methods use variable length decoding to obtain source and compression parameters. QuickParse NR-B methods require the least processing among the options described, but are still capable of providing valuable information regarding networked video quality [1].

Ideally, we would like to have an accurate mapping from QuickParse parameters to the overall perceptual quality. However, there are many challenges that must be overcome before this problem can be solved. For example, due to privacy restrictions, the videos in this study can *not* be viewed visually. Therefore, the typical method of obtaining "ground truth" regarding video quality, by using subjective tests, is not possible. Further, while the raw parameters that can be extracted using QuickParse techniques are clearly defined, choosing the proper set of derived parameters to best characterize quality is an open problem.

In this paper, we make a first step by presenting an initial exploration into the relationship between QuickParse parameters and objective measures of spatial quality. We characterize spatial quality using available objective NR-P methods, to provide a ground truth. In particular, we use the GBIM blocking metric [4] and the blurring metric from [5].

This paper is organized as follows. Section 2 describes the parameters we extract from the video using both QuickParse and NR-P methods. Section 3 describes our measurement study using these extracted parameters to characterize the video data we observed on the Internet during a one-week interval. We start by using QuickParse parameters to characterize the observed videos in terms of their (a) compression algorithm, or codec, (b) spatial resolution, (c) temporal frame rate, and (d) bits per pixel. We then apply NR-P blocking and blurring metrics to a subset of the videos, with the goal of obtaining a deeper understanding of their spatial quality. Section 4 concludes with a discussion of future work.

2. QUICKPARSE AND NR-P PARAMETERS

In this section, we describe the parameters that can be extracted from TCP/IP packets containing the popular Windows Media streaming format, using different processing depths. Here, the payload of the TCP stream consists of MMS packets [7] which in turn carry ASF packets [8] as payload. We begin by describing NR-B QuickParse parameters: information that can be obtained from the ASF packets without analyzing the media objects (or frames) therein. We then describe NR-P parameters that may be useful to characterize spatial quality.

2.1. QuickParse

Each ASF transmission, or file, starts with a header describing file information (including the file GUID, overall file duration, and overall file bit-rate), the number of available audio and video streams within the file, and summary information for each audio stream (codec, sample rate, etc.) and video stream (codec, spatial resolution, etc.) [8].

We can also obtain the following information associated with each media-object inside an ASF data packet: $i, j, b_{ij}, k_{ij}, s_{ij}$, p_{ij}, m_{ij}, a_{ij}. Media-object j from stream number i has a total of b_{ij} bytes. If $k_{ij} = 1$, then this media-object is a key-frame (or I-frame). The remaining four variables are timing information associated with this media-object. The ASF packet contains the target send time of the packet, which is the time the server is supposed to transmit this ASF packet. We denote s_{ij} as the send time of the last ASF packet which contains any part of media-object j from stream i. The target presentation time, p_{ij}, for media-object j in stream i is the time intended by the encoder for the client to present, or display this media-object. The time the monitor observes the last received ASF packet containing any part of this media-object is m_{ij}. The arrival time at which the client itself receives all information about this media-object, a_{ij}, may not be exactly observable, depending on the location of the monitor in the network. Because we are interested only in the quality intended by the encoder, we do not consider s_{ij}, a_{ij} or m_{ij} in this paper.

Once these basic parameters have been extracted from the ASF packets for each frame, we next obtain derived parameters which can be used to characterize spatial quality. We believe that, in addition to the spatial resolution (height and width of each image) and the compression algorithm, some other essential parameters for spatial quality are the encoded temporal frame rate, and the bit-rate, expressed both in bits per second (bps) and the compressed bits per coded pixel, or more simply, the bits per pixel (*bpp*).

Bit-rates used to compress video are typically measured in *bps*, while bit-rates used to compress images are typically measured in *bpp*. All things being equal, higher *bps* means higher quality. However, since all things typically are not equal, we believe *bpp* provides a potentially useful supplemental indicator of spatial quality. Intuitively, larger spatial resolution improves quality for the same *bpp*; better, more recent codecs have better quality per *bpp*; higher temporal frame rate increases temporal correlation and hence improves quality for the same *bpp*.

We derive the encoded temporal frame rate and both bit-rates *bps* and *bpp* using a causal, sliding one-second interval, where the timing is based on the presentation times, p_{ij} for each video frame. As information for each frame is extracted, we compute the frame rate and bit-rates using information from all frames with presentation times in the one-second interval ending at the current presentation time. The encoded temporal frame rate is the number of frames to be presented in this one-second interval. To estimate the bit-rates, we sum all received bits in this one-second interval, and for *bpp* divide by the number of coded frames and the number of pixels per frame. We estimate separately the bits per coded I-pixel, or *bpIp* using all I-frames in the one-second interval.

Summarizing, we directly extract sequence-level information consisting of stream ID, number of pixels, and number of lines from the ASF header. We directly extract frame-level information, comprising frame ID, presentation time, and key-frame flag. Further, using extracted parameters we derive the following frame-level estimates based on the past one second: frames per second, bits per second, bits per coded I-pixel *bpIp*, and bits per coded pixel *bpp*.

2.2. No-Reference Pixel metrics

To implement NR-P metrics, we use MPlayer [9] with Windows binaries to obtain video pixels for all codecs. However, we reiterate that these pixels are never displayed for privacy reasons. Simple source descriptions that can be estimated from pixels include: mean, variance, horizontal and vertical spatial correlation, and temporal correlation. Also, we apply the GBIM blocking metric [4] and Marziliano's blurring metric [5]. We choose these two metrics because these two performed the best among a number of NR-P metrics examined in [10]. In particular, they are able to show that when a coarser quantizer is used for compressing I-frames, both blocking and blurring increase. Therefore, we expect they will be useful here, where we have the intuition that, everything else being equal, a smaller *bpIp* should increase both blockiness and blurriness.

The GBIM blocking metric [4] computes a weighted sum of the squared differences between adjacent pixels across a block boundary. The weights account for masking by the human visual system, and are derived from averages and standard deviations of the pixel values in the local neighborhood. Larger GBIM values imply stronger visual artifacts aligned with coding block boundaries.

Marziliano's blurring metric [5] measures the spatial extent of each edge in an image using inflection points in the luminance to demarcate the start and end of an edge. The final metric averages the edge extent of each detected edge. As presented in [5], this metric does not take into account the spatial resolution of the image. However, a blur of 8 pixels extent in a 160×120 image is clearly not the same as a blur of 8 pixels extent in an image twice as large. Therefore, here we scale the blur metric of [5] by the horizontal extent of the image. As a result, the blur metric we use has the straightforward interpretation that a value of, say 0.1, means that the average extent of all edges in the image is 1/10th of the horizontal image size.

3. MEASUREMENTS OF SPATIAL QUALITY

In this section, we analyze the spatial quality of Internet streaming videos. We first describe how we obtain our data set. Next, we give a high-level characterization of the 6215 video streams, using QuickParse. Finally, we apply NR-P methods to a subset of the streams to gain further understanding of their spatial quality.

3.1. Video Data Set Description

Our trace data was obtained from the cable head-end of a large cable provider over a one-week interval in early 2004. We captured traces of all Windows Media traffic using the GigaScope monitoring appliance [11]. The raw trace data was processed off-line for our study. First the raw trace data were converted into libpcap [12] files, the format on which our tool operates. By parsing what is known of the Microsoft Media Server (MMS) protocol [7], these files were then partitioned into a single file per streaming session that contains all associated IP packets. The per-session files were then processed to emulate the bitstream as seen by the client. In

Codec	Description	Number videos (%)
WMv3	Windows Media 9	2987 (48.06)
WMv2	Windows Media 8	2432 (39.13)
WMv1	Windows Media 7	576 (9.27)
MP43	MPEG4-v3	212 (3.41)
Other		8 (0.13)

Table 1. Breakdown of codec usage

Spatial Resolution	Number videos (%)
320×240	4701 (75.64)
240×180	802 (12.90)
160×120	116 (1.87)
400×300	113 (1.82)
352×240	91 (1.47)

Table 2. Top 5 spatial resolutions

particular, we performed TCP reassembly (involves removing duplicate packets, reordering packets received out-of-order, and extracting the application level byte stream), and estimated the time at which a certain TCP byte-rate was received by the client.

In this paper, we consider only a subset of the total data observed. First, we only consider streams which use TCP as the transport protocol. Second, we only study streams for which the client received at least 10 seconds of video, because it is unlikely that a human observer would be able to judge the video quality for a shorter period. Third, we restrict our analysis to those videos which played continuously once started. Thus we do not consider any streams in which the client interrupted playback with a pause, replay or skip command. We do not consider live streams. And finally, we do not consider any videos in which the server adapted the transmission rate in response to network conditions. As a result, we have a total of 6215 video streams observed over 7 days.

3.2. Analysis using only QuickParse

In this section we present a high-level characterization of the data using QuickParse, focusing particularly on aspects that impact the video quality. Video from seven different codecs were present in our streaming data. Table 1 shows the breakdown of codecs used in our data set. Not surprisingly, most of the content was encoded in more recent versions of Windows Media (i.e., WMv2 and WMv3).

Our data shows a relatively large number, namely 74, of spatial resolutions being used. This is somewhat surprising given the fact that most encoding software has a small set of preset options. Table 2 shows the breakdown for the top 5 spatial resolutions. These account for 93.7% of all video streams, with the remaining 69 resolutions being spread over 392 video streams.

We now focus on the 4 most frequently seen codecs, and examine the cumulative distribution function (CDF) of the video bit-rate for these 6207 streams, shown in Figure 1. Two observations are immediate. First, the two newer codecs, WMv2 and WMv3, have lower bit-rates, on average, than the two older codecs, WMv1 and MP43. Second, for each codec, the mass of the distribution (corresponding to the steps in the CDFs) is concentrated around a few values. All distributions show a peak around $25 - 28$ Kbps; these are probably streams coded for dial-up connections. The largest mass for each distribution is in the neighborhood of $175 - 250$Kbps, depending on the codec. Approximately 62% of the WMv3 streams have bandwidths between 175 and 200 $Kbps$, while approximately 50% of the WMv2 streams have bandwidths between 175 and 220 Kbps. Further, WMv2 has almost 10% more streams with bit-rate below 100 Kbps than WMv3 does.

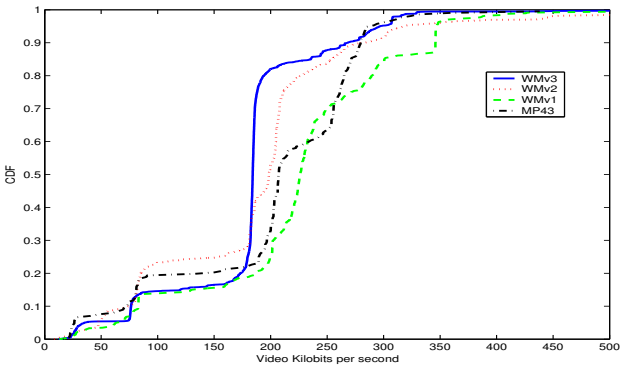

Fig. 1. CDF of video bits per second for 6207 streams.

Next, we examine the CDF of the video frame rates, shown in Figure 2. A few frame rates dominate the distributions, viz. around 6 fps, 15 fps, 24 fps and 30 fps. MP43 is the only codec whose distribution has significant mass below 10 fps, with about 15% of streams at 6 fps. MP43, WMv1, WMv2 and WMv3 each have $41, 58, 23$, and 29% of their respective masses between $14 - 16$ fps. Similarly, each have $25, 24, 42$, and 54% of their respective masses above 28 fps. WMv2 is the only codec with sizeable mass around 24 fps: about 10% of these streams have 24 fps. Clearly, the conclusion is that the more recent WMv2 and WMv3 codecs typically use higher frame-rates than the older codecs.

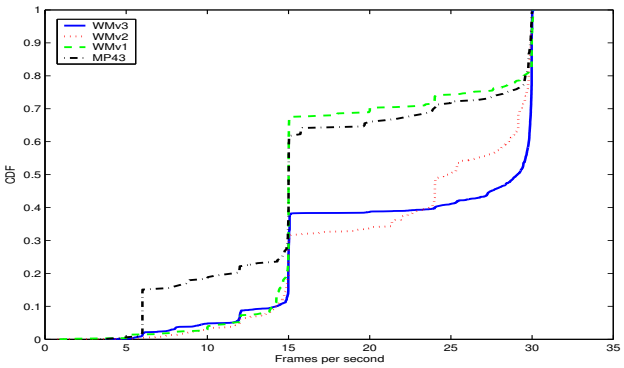

Fig. 2. CDF of coded frames per second for 6207 streams.

The CDFs of the bits per pixel (Figure 3) show that overall, the distributions for WMv2 and WMv3 favor smaller values than either MP43 or WMv1. As an illustration, while 90% of WMv3 streams have less than 0.2 bpp, only 70% of MP43 streams and 57% of WMv1 streams have bpp this low. The most likely explanation of this behavior is that the newer codecs can achieve similar quality with a lower bpp; however, without a "ground truth" characterizing the video quality, this is only a hypothesis.

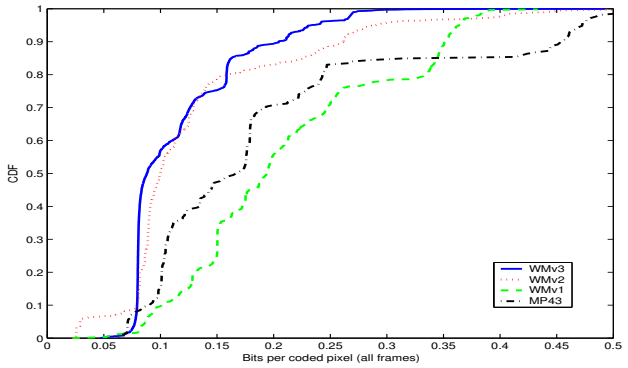

Fig. 3. CDF of bits per coded pixel (all frames) for 6207 streams.

We next examine the distribution of the stream durations for the different codecs (Figure 4). This metric measures how many seconds of video was transmitted to a client for each stream. More than 40% of streams (overall) were streamed for 30 seconds or less, but approximately 25% of streams were streamed for over 3 minutes. The newest codec, WMv3, has more longer duration streams than the other 3 codecs. This is in part because the WMv3 videos are also typically longer than videos produced by the other codecs - we can check this by examining the distributions of the file durations in the ASF header. However, WMv3 streams may also have higher quality, so people may be more likely to watch them longer. This would need to be explored in more depth.

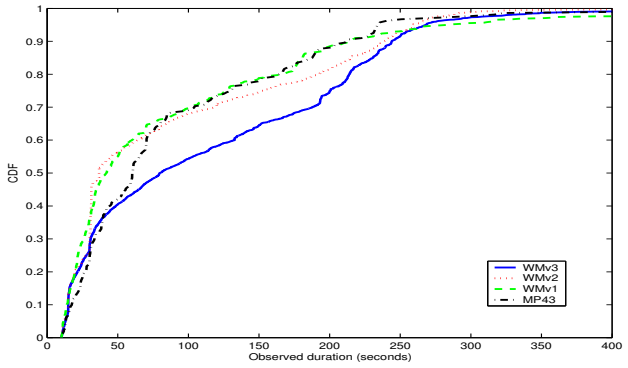

Fig. 4. CDF of duration of 6207 streams.

3.3. Analysis using QuickParse and NR-P metrics

We now apply the NR-P metrics to a subset of the data, focusing on a few specific combinations of codec, spatial resolution and frame rate. We explore the relationship between bit-rate (*bps* and *bpIp*) and the NR-P blocking and blurring metrics, for fixed codec, spatial resolution, and frame rate. Our expectation is that higher *bpIp* will reduce both blocking and blurring.

In this section, we first report on the trends we see in the data as we vary codec, spatial resolution, and/or frame rate. We begin in Section 3.3.1 by exploring the impact of changing both temporal and spatial resolution for 2 specific codecs. Next, we explore in Section 3.3.2 the impact of frame-rate for fixed spatial resolution. Finally, in Section 3.3.3, we explore trends in bit-rate or quality across the 4 codecs. In all cases, if the trends do not match our expectations, we use additional NR-P measures to gain a deeper understanding of the source itself, to explain the observations.

Resolution		Bit-rate		NR-P metrics	
Spatial	Temporal	*Kbps*	*bpIp*	GBIM [4]	Scaled Blur [5]
192 × 144	7.5	34.4	0.317	2.670	0.0540
240 × 180	15	86.4	0.393	2.647	0.0479
320 × 240	30	205	0.343	2.742	0.0378

Table 3. Median bit-rates (*bps* and *bpIp*) and NR-P metrics (GBIM and blur) as function of spatial and temporal resolution for WMv3.

Resolution		Bit-rate		NR-P metrics	
Spatial	Temporal	*Kbps*	*bpIp*	GBIM [4]	Scaled Blur [5]
160 × 120	7.5	26.9	0.340	3.126	0.0646
240 × 180	15	98.3	0.373	3.073	0.0435
320 × 240	30	220	0.330	3.283	0.0308

Table 4. Median bit-rates (*bps* and *bpIp*) and NR-P metrics (GBIM and blur) as function of spatial and temporal resolution for WMv2.

We do not apply the NR-P metrics to all 6215 streams. In some cases, we use all streams available for a given codec and resolution. In other cases, we choose a subset of the available streams due to the processing complexity of fully decoding each video stream. We focus only on I-frames, and partition the I-frames from each codec into categories, based on their spatial resolution and local frame rate. Local frame rates of 6-9 fps are mapped into 7.5 fps; 14-16 fps are mapped into 15 fps, and frame rates above 28 fps are mapped into 30 fps.

3.3.1. Impact of joint spatial resolution and frame rate

We start by exploring the impact of changing both temporal and spatial resolution for 2 specific codecs: WMv3 and WMv2. For each codec, we consider the three most frequent combinations of spatial resolution and temporal frame rate in our data set. Tables 3 and 4 show the median values for *Kbps*, *bpIp*, and the GBIM blocking and blurring metrics for the different resolutions for WMv3 and WMv2, respectively. In what follows, we explore the median values in the table one column at a time, and use additional graphs to gain a deeper understanding of the distributions when necessary.

For each of the 6 cases of bit-rate (*Kbps*) in Tables 3 and 4, the distribution is very narrow, so the median is a good characterization. The larger resolutions have higher median *Kbps*. One can expect that these combinations are chosen specifically to accommodate different networking configurations. Additionally, from the CDFs (not shown), we see that the bit-rates for nearly 100% of the lowest-resolutions streams for both codecs are less than 100 Kbps, while the bit-rates for approximately 90% of the highest-resolution streams for both codecs are less than 300 Kbps.

Next we explore the *bpIp* for the two codecs, and supplement the median values in the tables with the CDFs, shown in Figures 5 and 6. Interestingly, for WMv3 the CDF of the *bpIp* is nearly invariant for the three different spatial and temporal resolutions. It would be interesting to know if this is intentional. Examining

the figures shows that the middle-resolution WMv2 streams have a higher *bpIp* than the other 5 cases. For example, the fraction of WMv2 streams with 240 × 180 and 15 fps that have over 1 *bpIp* is more than twice that of any of the other 5 combinations in this comparison. The additional *bpIp* likely results in better quality if the statistics of the video scenes being compressed are equivalent. If the video scenes being compressed are more difficult, then the additional *bpIp* may be necessary to achieve the same quality.

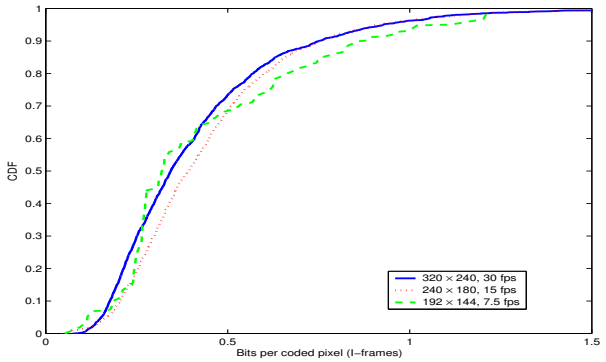

Fig. 5. CDF of bits per coded I-pixel for WMv3: Table 3.

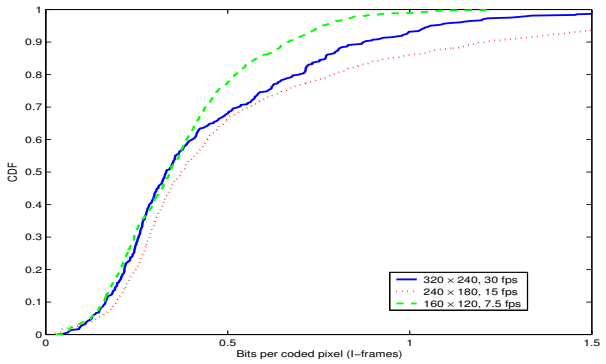

Fig. 6. CDF of bits per coded I-pixel for WMv2: Table 4.

We explore this next, by considering the median GBIM values in Tables 3 and 4, along with CDFs for the GBIM values shown in Figures 7 and 8. For GBIM values below about 2.5, all 6 distributions appear similar. However, the tail of the GBIM CDF for WMv3, 320 × 240 and 30 fps shows significantly higher GBIM values than the other 5 cases. For example, the 90-th percentile for WMv3 320 × 240 and 30 fps has a GBIM value of 8 while the next lowest 90-th percentile value is 5.

However, an explanation for this is not obvious. A closer examination shows that GBIM values in this set larger than 5 appear only in 9 (out of 50) distinct video streams. The frames with large GBIM also have large spatial variance, although many other I-frames have similarly large spatial variance with much lower GBIM values. We believe this is an artifact of the source content of these 9 streams, in which very strong edges coincidentally align with coding block boundaries.

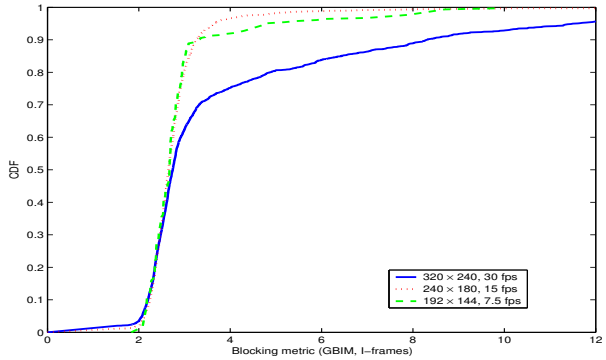

Fig. 7. CDF of blocking metric GBIM for WMv3: Table 3.

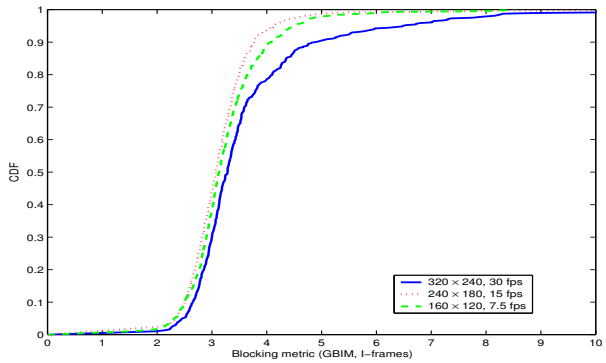

Fig. 8. CDF of blocking metric GBIM for WMv2: Table 4.

We now consider the medians of the blurring metric, given in Tables 3 and 4. We see that using the simple scaling of the blurring metric as described in 2.2, smaller spatial resolutions are blurrier, which is to be expected. The CDFs (not shown) show a similar trend. For example, for WMv2, the 85-th percentile for spatial resolution of 160 × 120 is 0.1, meaning that 85% of these I-frames have an average edge-extent that is less than or equal to 10% of the image extent. Meanwhile, for WMv2 and a spatial resolution of 320 × 240, 85% of the I-frames have an average edge-extent that is less than or equal to 5% of the image extent.

3.3.2. Impact of frame rate for fixed spatial resolution

Now, we explore the impact of changing frame rate for a fixed spatial resolution of 320 × 240. Table 5 shows the median values for *Kbps*, *bpIp*, GBIM, and blurring for streams with a temporal resolution of either 15 or 30 fps. In this section, we compare medians in each column while holding codec and spatial resolution fixed, but varying the frame rate from 15 fps to 30 fps.

Examining first the *Kbps*, we see that interestingly, both WMv3 and WMv2 have a lower median *Kbps* for frames with 30 fps compared to 15 fps. On the other hand, both WMv1 and MP43 have

Codec	Frame rate	Bit-rate		GBIM [4]	Scaled Blur [5]
		Kbps	*bpIp*		
WMv3	15	243.1	0.701	2.595	0.0314
WMv3	30	205.4	0.343	2.742	0.0378
WMv2	15	252.0	0.224	2.616	0.0322
WMv2	30	220.7	0.330	3.283	0.0308
WMv1	15	282.9	0.465	4.067	0.0325
WMv1	30	294.5	0.349	4.370	0.0361
MP43	15	274.5	0.583	3.579	0.0264
MP43	30	294.6	0.524	4.275	0.0293

Table 5. Median bit-rate, *bpIp*, GBIM, and blur as function of spatial resolution, codec, and frame rate.

a higher median *Kbps* for frames with 30 fps compared to those with 15 fps. Considering next the *bpIp* we see that for WMv3, the median *bpIp* for 15 fps is more than twice the median value for 30 fps. We expect this additional *bpIp* would reduce the amount of blocking and blurring, and indeed, we see that both NR-P metrics have a lower median value for 15 fps than for 30 fps. Similar behavior is seen for WMv1 and MP43, although the reduction is not so dramatic.

On the other hand, the relationship among *Kpbs, bpIp*, blocking and blurring for WMv2 streams with 15 fps is not the same as the other codecs. In particular, these streams have a lower median *bpIp* yet also a lower median blocking. To understand this behavior, we examined an additional variable: the pixel variance (computed using the decoded pixels). We see that these streams have a significantly smaller pixel variance than other streams. Thus, for this source content, a small *bpIp* is sufficient to create reasonable-quality video. The low *bpIp* relative to the *Kbps* can be explained by noticing that these streams have a significantly higher proportion of I-frames than the other streams. For these streams with 15 fps, 25% of the WMv2 frames are I-frames, while less than 4% of the other codec's frames are I-frames. Since P-frames typically take fewer bits per pixel than I-frames, streams with a higher fraction of I-frames for a given *Kbps* will be forced to use lower *bpIp*.

3.3.3. Impact of codec for fixed spatial resolution and frame rate

We can also use the data in Table 5 to examine the impact of the codec for a fixed spatial resolution of 320×240 for a fixed frame rate of either 15 fps or 30 fps.

If we compare WMv3 and MP43 streams at both 15 and 30 fps, we see that WMv3 streams are simultaneously significantly blurrier and significantly less blocky than the corresponding MP43 streams. One conclusion is that WMv3 streams achieve lower blockiness at the expense of somewhat increased blurriness. This is interesting, in light of the conclusion by Farias et al. [13], that if artifacts are perceived to have equal strength, users perceive blurriness as more annoying than blockiness.

4. CONCLUSIONS

We presented a statistical characterization of the spatial quality of streaming Internet video, considering the impact of spatial resolution, temporal frame-rate, bit-rate, and compression algorithm. We took a first step by presenting an initial exploration into the relationship between QuickParse parameters and objective measures of spatial quality. Our eventual goal is to develop a mapping from these parameters to subjective quality, using NR-P methods to provide an accurate ground truth.

Despite the large number of no-reference blocking and blurring metrics available in the literature, more work must be done in this area before NR-P metrics can be effectively used for ground truth. First, the metrics should be easily comparable across multiple contexts (e.g., different spatial or temporal resolutions). Second, a metric is a useful tool only if there exists a precise non-linear mapping between the metric and subjective quality. To use the metric, we need to know, for example, how different a value of 3.0 is compared to 3.2. Unfortunately, many existing metrics do not have an associated mapping. Third, the metric and the mapping function should be robust to a variety of source content. For instance, strong sharp edges in the image content that happen to be aligned with coding block boundaries (which might appear with "letter boxing") should not be counted as blocking artifacts. We are pursuing research on NR-P based subjective video quality estimation along these directions.

5. REFERENCES

[1] A. R. Reibman, S. Sen, and J. Van der Merwe, "Network monitoring for video quality over IP," in *Picture Coding Symposium*, December 2004.

[2] M. A. Masry and Sheila S. Hemami, "A metric for continuous quality evaluation of compressed video with severe distortions," *Signal Processing: Image Comm.*, vol. 19, pp. 133–146, February 2004.

[3] A. Webster et al., "An objective video quality assessment system based on human perception," in *Proc SPIE, vol 1913*, pp. 15–26, 1993.

[4] H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," *IEEE Sig. Proc. Let.*, November 1997.

[5] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proc. IEEE ICIP*, vol. 3, pp. 57–60, 2002.

[6] A. R. Reibman, V. Vaishampayan, and Y. Sermadevi, "Quality monitoring of video over a packet network," *IEEE Transactions on Multimedia*, vol. 6, pp. 327–334, April 2004.

[7] Streaming Download Project, "MMS streaming protocol." Available from: `http://sdp.ppona.com/`.

[8] Microsoft Corporation, "Advanced systems format (ASF) specification, revision 1.20.02." ASF_Specification.doc available from: `http://download.microsoft.com`, June 2004.

[9] http://www.mplayerhq.hu/: MPlayer 1.0-pre5try2 software.

[10] A. Leontaris and A. R. Reibman, "Comparison of blocking and blurring metrics for video compression," in *Proceedings ICASSP*, March 2005.

[11] C. Cranor, T. Johnson, O. Spatscheck, and V. Shkapenyuk, "Gigascope: A stream database for network applications," in *Proc ACM SIGMOD*, June 2003.

[12] V. Jacobson, C. Leres, and S. McCanne, "tcpdump/libpcap." Available from: `http://www.tcpdump.org/`.

[13] M. C. Q. Farias, S. K. Mitra, and J. M. Foley, "Perceptual contributions of blocky, blurry and noisy artifacts to overall annoyance," in *Proc. IEEE ICME*, 2003.