Evaluating Perceptual Video Quality for Mobile Clients in 802.11n WLAN

Victor Omwando, Amit Pande, Yunze Zeng, Prasant Mohapatra
Department of Computer Science, University of California, Davis
{voomwando, pande, zeng, pmohapatra}@ucdavis.edu

ABSTRACT

In this paper, we characterize the performance of HD video streaming in 802.11n WLANs under user mobility. We conducted experiments in QuRiNet, a large-scale outdoor wireless testbed that experiences little electromagnetic interference. We observe the variation in video quality with the variance of both speed of a mobile user and his distance from access point (AP). Using subjective scores and objective video quality assessment metrics, we build a non-linear regression model to estimate video quality based on user speed and distance. An ensemble machine learning kernel, bagging, is used in conjunction with Reduced Error Pruning Decision Trees to build a non-linear prediction model that scores 69% correlation with video quality. Overall, we find that distance has larger impact on video quality than speed. However, the physical factors such as speed and distance cannot be used in isolation to estimate video quality accurately.

Categories and Subject Descriptors

C.4 [**Performance of systems**]: Performance attributes

General Terms

Measurement, Performance

Keywords

Video quality, measurement, WiFi

1. INTRODUCTION

In recent years, there has been an explosive growth in Internet usage fueled by ubiquitous availability of smartphones, tablets and other Internet-connected mobile devices. Video makes up a significant proportion of the data consumed by users. Indeed, the Cisco Visual Networking Index reports that mobile video traffic accounted for 52% of total mobile data traffic by the end of 2012. This figure is expected to grow to about 66% by 2017 [7]. At the same time, network providers are faced with increased infrastructure costs and restricted spectrum allocations. This has triggered the introduction of data caps on cellular service subscriptions and an almost total elimination of unlimited data service plans. Carriers have also sought an alternative solution: heterogeneous wireless networks that allow cellular network data traffic to be offloaded to a wireless LAN in public places such as coffee shops, at home and at the office.

However, the performance of mobile video delivery in the WiFi standard is not yet clearly understood in an experimental context. Much work has been done to characterize video quality on mobile devices, but few, if any, have been undertaken in a live WiFi network with mobile clients. Most efforts have focused on scenarios where packet drops and distortions are injected to videos. Even then, these attempts do not take into consideration the speed at which the mobile user is traveling as well as his distance from the access point.

We created a database of forty five videos streamed over 802.11n WLAN. Three high-definition video files were transmitted in different mobility conditions at QuRiNet, a wide-area wireless outdoor testbed [27]. We switched off the extra WiFi transmitters at this ecological reserve to obtain an interference-free network. In addition to application layer data, we recorded network packet traces, MAC-layer statistics from the wireless driver and speed and distance information. This rich source of information provides valuable insight on how various network and physical characteristics affect video quality.

We then obtained both subjective scores (mean opinion score or MOS of users) and established reduced-reference (RR) and no-reference (NR) video quality metrics to estimate the video quality using distance and speed information. We used three quality metrics: the Temporal Variation Index (TVI) [6], the blocking metric [26] and the Natural Image Quality Evaluator (NIQE) [18]. First, we modeled the impact of user speed on subjective video quality and obtained a 33% correlation. Whereas subjective scores are available for the video as a whole, not on a per-frame basis, the distance of the user from the access point changes on a per-frame basis. Since TVI has finer granularity and achieved the highest correlation with subjective quality (81%) we used it to develop a regression model that estimated perceptual quality from user speed and distance.

In our experiments, we varied user velocity values from zero to 50 kilometers per hour (30mph), with no Access Point handover, since the IEEE 802.11 WLAN standard does not support it. This conforms to use cases involving a group of people moving relatively slowly while requesting the same data, for example in stadiums, malls and campuses. In such locations, service providers can offload traffic to reduce congestion on their last-mile networks.

For regression, we used both linear regression models and non-linear regressions including machine learning algorithms such as

WiNTECH'13, September 30 2013, Miami, Florida, USA

http://dx.doi.org/10.1145/2505469.2505476.

Reduced Error Pruning Decision Tree (REPTree) [22] and bagging [3], an advanced ensemble classifier.

The main contributions of this paper are as follows:

- Our experiments demonstrate that a linear model outperforms various non-linear models when predicting the effect of speed on MOS, even though the correlation (ρ) with the actual MOS is only 33%.

- We build non-linear regression models to predict the effect of speed and distance on the TVI metric in an idealized environment with little electromagnetic interference and no background traffic. We use the bagging ensemble machine-learning classifier over REP decision trees to obtain a 69% correlation accuracy on our full training set. However, the correlation drops to 40% when we apply 10-fold cross-validation [14], an indication that speed and distance cannot be used in isolation to achieve an accurate estimate of video quality.

- Using various prediction models, we also show that distance has a bigger impact on video quality than speed.

The paper is organized as follows: Section 2 gives an overview of related works in this area. Section 3 presents the experiment setup, followed by a brief summary of the objective and subjective quality metrics in Section 4. In Section 5, we describe the prediction models used in the paper. Section 6 presents the experiment results and Section 7 concludes and discusses directions for future work.

2. RELATED WORK

A number of studies have been conducted to evaluate 802.11 behavior in infrastructure mode and vehicular clients, but most of them do not have any emphasis on video traffic [10, 4, 20, 16, 28, 8]. They also do not consider the newer 802.11n standard. Notably, Mahjan et al. [16] use 802.11 base-station (BS) beaconing messages as the underlying traffic to the fundamental characteristics of WiFi-based connectivity between BS and vehicles in urban settings. They find that intermittent periods of poor very connectivity are not caused by vehicular motion *per se* but by the variability of the environment combined by the vehicle traversing locations that have poor coverage by the BS. In contrast, we aim to isolate and quantify the effect of mobility by performing our experiments in an interference-free environment with little obstruction and predefined movement patterns. In [15], the authors analyze the capabilities of 802.11 b/g/n for both unicast and multicast streaming transmissions directed to mobile devices in a simple testbed. However, they do not consider clients in motion.

On the other hand, much work has done to develop and evaluate objective video quality metrics in WiFi networks, for example in [5, 6], but their endeavors assume non-moving clients. Moorthy et al. [19] generate an extensive database of distorted videos and corresponding scores, and conduct subjective, objective and behavioral assessment of the videos. Even so, as in [25] their database does not contain videos that have been transmitted over live wireless links. Instead, they simulate errors in wireless environments using bit error patterns and software available from the VCEG [24].

To our knowledge, the impact of client motion on perceptual video quality has not been quantified in any prior work.

3. EXPERIMENT SETUP

We first explain the experimental setup.

(a) Video A Sample Frame

(b) Video B Sample Frame

(c) Video C Sample Frame

Figure 1: Example frames from the videos used in the study

3.1 Source Sequences

We collected data using three HD video sources. The first two videos were captured in interlaced format (1080i) using a Sony video camera at 30 interlaced frames per second. They were stored in a MTS (MPEG Transport stream) container at a high bit-rate to maximize the fidelity of the video. We were unable to detect any differences in quality from the originals. Both videos were approximately sixty seconds long. The third video was captured in progressive RAW YUV format with a chroma sub-sampling of 4:2:0 and a bit-rate of 18 Mbps. It has a frame rate of 30 frames per second. Similar to the two videos described above, it was then compressed using the MPEG-4/H.264 Advanced Video Coding (AVC) codec via ffmpeg, and encapsulated in a MPEG Transport stream, maintaining the original bit-rate. As with the first two, the video was 60 seconds long. A more detailed description of the three videos follows:

- *Highway (Video A)* - Filmed by a passenger on the front seat of a car.The video has a bitrate of 5Mbps, 30 interlaced frames per second, with a resolution of 1080i (1440x1080).

- *Park Scenery (Video B)* - Filmed at a local park in Davis, California. It has a bitrate of 7Mbps, 30 interlaced frames per second, with a resolution of 1080i (1440x1080).

- *Park Run (Video C)* - Captures a person jogging in a park while holding an umbrella. Every ten seconds, the subject of the video momentarily comes to a complete stop before running again. The video has a bitrate of 18Mbps, 30 frames per second and has a resolution of 1080p (1920x1080).

Figure 1 shows sample frames from the various video sequences.

3.2 Location

We performed our experiments in the Quail Ridge Wireless Multihop Testbed (QuRiNet) [27]. QuRiNet is an outdoor, solar-powered wireless testbed deployed in the Quail Ridge Reserve at Lake Berryessa, California. A layout of the site is shown in 2(a). Since the area is uninhabited, it gave us an excellent chance to deploy large-scale experiments in an area that is largely free of electromagnetic interference and other sources of perturbation.

3.3 Device Setup

For our evaluation, we designated one site (site X), as the source of transmission. A laptop was connected to the site router, a Soekris net4826, via a CAT 5 Ethernet Cable. The streaming server was a

(a) QuRiNet layout with server location circled

(b) Trail Used in Experiments

(c) Server Setup

(d) Client Setup

Figure 2: Experiment setup

Dell laptop with an Intel Core i5-2520M CPU processor clocked at 2.5GHz, integrated Intel graphics card, and 4GB of main memory.

The client site Y had a similar Soekris board setup, with a Dell Latitude E5400 laptop that had an Intel Core 2 Duo CPU clocked at 2.0 GHz and 2.0 GB of main memory. Both server and client laptops ran the Linux-based Ubuntu 12.04 operating system. We placed the client set-up on an All-Terrain Vehicle (ATV).

Each of the routers had one 2x2 Microtik R52n-M MIMO IEEE 802.11a/b/g/n Mini-PCI form-factor wireless card with the Atheros AR9220 chipset and the ath9k open source wireless driver. The transmission power was fixed at 16dBm. The channel was set to 2.462GHz, to further isolate the experiments from other testbed routers set to the non-overlapping channel 1 (2.412GHz). Due to the isolation enjoyed by the testbed, we could not detect any other wireless transmissions on that frequency. The wireless interfaces on the two routers had an omni-directional antenna with 7.4dBi antenna gain. We attached only one antenna to each of the cards. The antennas at both sites were set 155cm above the ground. Despite the overall hilly terrain of the environment, the specific location of our experiments had relatively flat topology. In throughput tests, we achieved an average of 11Mbps on our link. However, we empirically realized transmission rates of up to 65Mbps, corresponding to Modulation and Coding Scheme (MCS) Index 7. The rate was dynamic owing to the use of the Ministrel rate control algorithm, which chooses the modulation scheme based on an Exponential Weighted Moving Average (EWMA) of packet delivery success rates and adjusts the sending rate accordingly. We drove the ATV over a range of approximately 300 meters and used the US GlobalSat BU-353 GPS receiver to log GPS coordinates. Fig. 2 shows our experimental setup. For clarity, we have circled the laptop, router and antenna on both the server and client side.

3.4 Data Collection Methodology

For each video, we considered three different mobility scenarios, and performed five runs in each of them, for a total of 45 runs for the HD videos:

- *Scenario 1* - A speed range of 0 mph to approximately 4 mph (0 km/h to approximately 6 km/h), mimicking motion from a standstill to a leisurely walk. Average speed of 3mph.
- *Scenario 2* - A speed range of 0 mph to 10 mph (0 km/h to 16 km/h). Average speed of 8mph.
- *Scenario 3* - A speed range of 10 mph to 50 mph (0 km/h to 50 km/h). Average speed of 20mph.

For Scenarios 2 and 3, we tried to prune the periods with little motion. This was limited to the start of the run, maneuvering around a slight curve in our path, or when we had to execute a turn.

Before starting the runs, we first synchronized the two routers and two laptops to within half a second of each other. At each run, we started driving the ATV at the same time as the scheduled transmission of the video, and concluded that run at the end of the video. Using tcpdump, we obtained packet traces from all four devices during this period. Additionally, at the routers, we logged various wireless parameters from the iw and iwconfig utilities, and enabled the debug filesystem in order to access MAC-layer statistics. The GPS device was directly connected to the client laptop, and at every run, gathered GPS location information on a per-second granularity. This was facilitated by the freely-available gpsd daemon which avails location information from the GPS device to a suitable client software[1]. In our case, we used gpxlogger, a tool developed in conjunction with gpsd that returns XML-formatted fixes every second. We calculated the distance from the AP using the Haversine formula[23] implemented in the Python-based gpxpy library [2], and derived speed from those values.

We used the VideoLan VLC software for video streaming. During video transmission, the video stream was routed from the laptop to the Soekris board, which then unicasts the video over the wireless link using the Real-time Transport Protocol (RTP) over UDP. We chose this set up because the routers are too computationally constrained to handle the video streaming by themselves. We saved the streamed video at the client, but we did not perform any additional transcoding during or after the stream.

4. VIDEO QUALITY METRICS

4.1 Objective Metrics

We used two no-reference spatial and one reduced-reference temporal video quality assessment (VQA) algorithm to evaluate objective video quality at the receiver. Spatial metrics attempt to quantify intra-frame video quality aspects such as blocking or blurring. Temporal metrics on the other hand capture inter-frame distortions like jitter and frame delay .

The two spatial metrics are the blocking metric [26] and the Natural Image Quality Evaluator (NIQE) [18]. The Temporal Variation Index (TVI) rounds out our objective VQA algorithms [6]. NIQE has been recently proposed in literature, and like TVI, has been found to have a high correlation with subjective video quality. We chose blocking as the additional spatial metric because unlike NIQE, it is not derived from the image characteristics of a set of pristine images, but from the video coding artifacts or the test video itself. TVI compares the motion between the source and received videos to identify frame losses and video freezing.

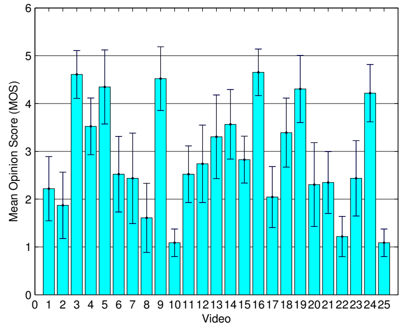

Figure 3: MOS and 95% confidence intervals for test videos

4.2 Subjective Experiments

We engaged 23 volunteers as observers to provide a subjective evaluation of 25 randomly chosen received video samples[1]. Our test volunteers had an age range of 20 to 45. There were six females (26%) and 17 males. Their occupation ranged from undergraduate students to laboratory technicians. Each subject was asked to grade the quality of videos on a five-point scale [13], once for each video, and enter the opinion on a prepared form. Score 1 represents a bad quality video with very annoying impairment, and Score 5 indicates excellent quality and imperceptible impairment.

The videos were displayed on a Dell Latitude E6420 laptop with 14 inch LED display and a screen resolution of 1366x768. All of the volunteers used the VLC media player software for playback.

We performed our test according to the single-stimulus (SS) method [13]. Standard videos with five different scores were shown at the beginning of the test. During the test, only the videos to be scored were shown, in arbitrary order, without any display of the standard videos. For each video clip, we averaged the opinion scores given by the observers to obtain the Mean Opinion Score (MOS) for that video.

Figure 3 shows that the MOS obtained for all videos ranged from 1.08 to 4.65. This shows that our test videos captured a wide range of quality. The average size of the 95% confidence interval among the videos is 0.65, indicating consensus among the observers.

5. PREDICTION MODELS

Typically, a regression or classification setting will have a learning or training set of $\mathcal{L}$ consisting of data $\{(y_n, \boldsymbol{x}_n), n = 1, ..., N\}$ where the y's are either numerical responses or class labels and $\boldsymbol{x}_n \in \mathbb{R}^d$ is the d-dimension attribute or Feature Vector (FV) for the nth instance. This training set is used as input to a machine learning algorithm to form a predictor $\varphi(\boldsymbol{x}, \mathcal{L})$. That is, given an input $\boldsymbol{x}$, we predict y using the function $\varphi(\boldsymbol{x}, \mathcal{L})$.

5.1 Linear Parametric Model

We considered a linear parametric model using ordinary least-squares regression for our investigation [17].

5.2 Reduced Error Pruning Tree

We used the Reduced Error Pruning Tree (REPTree) to build non-linear predictive models[22]. REPTree comes from the family of decision trees, which build classification or regression models in a tree-like structure. An internal node represents a test on an attribute, branches denote the outcome of the tests, and the leaf node corresponds to the numerical target. Decision trees use the ID3 algorithm at their core[21]. ID3 does a greedy space search over the possible branches without backtracking. Given an example set S, ID3 will choose the root node as the attribute that has the lowest entropy or highest information gain. This can be calculated using the formula $s_X - s_{X,Z}$, where X is the attribute, Z is the target variable and s is the standard deviation. The attribute that results in the lowest $s_{X,Z}$ is the root, since it results in the highest standard deviation reduction. For a continuous attribute A, a threshold c is picked such that it maximizes the information gain. Two branches are added, representing the two subsets of $A < c$ and $A > c$. The attribute selection and branch splitting at each of the new nodes is repeated, until the number of predicting attributes has been exhausted, or the stopping criteria has been met.

5.3 Bagging

Bagging, a sobriquet of bootstrap aggregating, is an ensemble machine learning meta-algorithm used to improve the stability and accuracy of machine learning methods [3]. It generates multiple versions of a predictor and then uses these to get an aggregate predictor. Each predictor is obtained by bootstrapping the learning set [9]. It works as follows: instead of using the training set as-is, generate M new training sets $\{\mathcal{L}_i\}$, each of size N, by sampling from $\{\mathcal{L}\}$ uniformly and *with* replacement. M is usually chosen as 50 or 100, and N is the size of the training data. Each of the $\{\mathcal{L}_i\}$ is a bootstrap sample. The bootstrapped estimator is then calculated via $\varphi(\boldsymbol{x}, \mathcal{L}_i)$. The resulting estimators can be aggregated by averaging for the regression case.

$$\varphi_B(\boldsymbol{x}) = \sum_{i=1}^{M} \varphi(\boldsymbol{x}_i, \mathcal{L}_i) \quad (1)$$

Finally, this random division of data is repeated over multiple iterations, say between 10 and 10,000, to provide robustness in the results.

Bagging has been shown to improve the accuracy of unstable procedures where a small change in $\mathcal{L}$ can lead to a large change in φ. This is the case for neural networks, classification and regression trees, and subset selection in linear regression [3]. In our work, both REPTrees and linear regressions are used to construct different bagging predictors.

We use the linear regression, bagging and REPTree implementations in Weka [11].

6. EXPERIMENTS AND RESULTS

6.1 Evaluation of Objective Metrics

In our analysis, we first consider the objective VQA metrics from nine different received videos, representing all three videos and all three mobility scenarios. For this foray, instead of taking the raw per-frame values of the metrics, we split them into bins of one second each and use the average value over each of these bins. Thus, whereas each metric was derived from one frame, we use the mean of thirty such metrics to represent the per-second objective image quality. In addition to making it convenient when pairing with speed and distance values, this approach works well to mitigate the impact of frame losses on reduced or full-reference metrics like TVI. It also reduces the inaccuracies introduced by frame misalignment, since the metric is now derived from the average of 30 consecutive frames instead of from the values of potentially mismatched frames. To further highlight the changes in blocking and NIQE after transmission, we calculate the difference between the per-second average values between the source and received videos, giving us delta-blocking and delta-NIQE. Figure 4 provides basic

[1]According to ITU-R BT.500-13 subjective assessment standard, [13], a minimum of 15 observers is needed for subjective quality evaluation

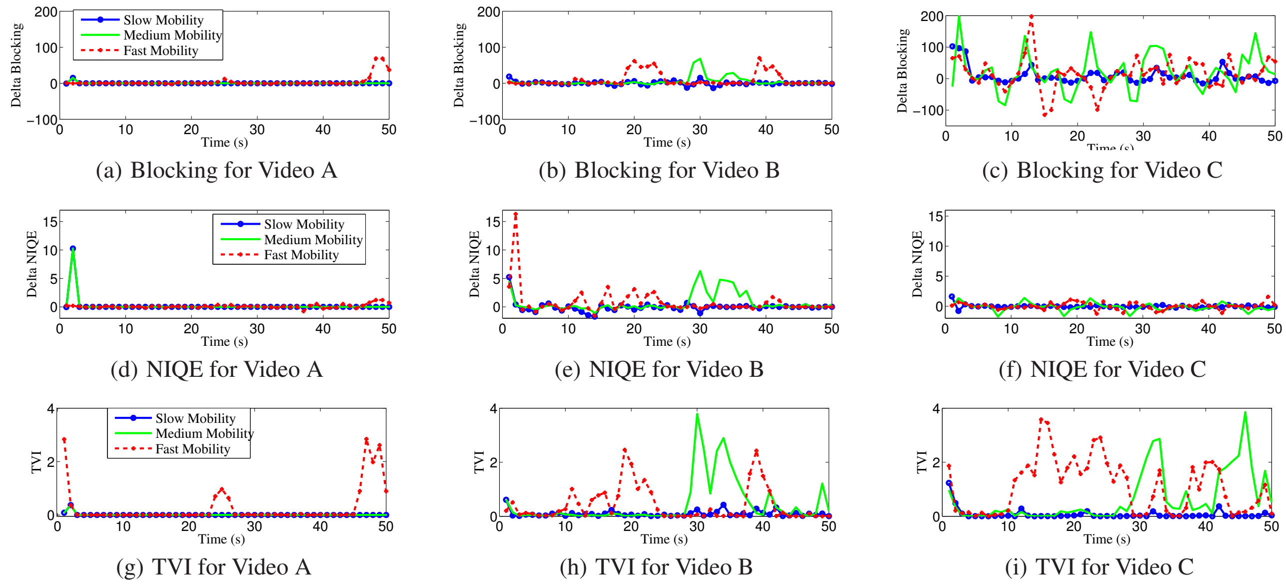

Figure 4: Blocking, NIQE and TVI metrics for three sample videos

Table 1: Spearman correlation coefficients of MOS and objective metrics

	NIQE	Blocking	TVI
MOS	-0.3366	-0.6769	-0.8151

comparisons on the effect of mobility on the video quality metrics for three randomly selected received videos.

A number of insights can be obtained from Figure 4, which was representative of the database. For all possible scenarios, there was an obvious degradation of quality at the very start of the stream, due to delays in video processing at the receiver during the initial RTP session setup. Since RTP runs over UDP, and the latter has no congestion or error control, it is the responsibility of the client RTP implementation to reorder packets using the sequence numbers contained in the RTP packet header. At the outset, the client may struggle to process the new burst of incoming packets, leading to dropping some packets in order to ensure a timely service. The dropped packets manifest themselves in an initial decrease in quality in all cases.

Another key observation is that the video quality metrics, and especially TVI, do not exhibit a linear relationship with the speed or distance. Instead, the metric degradation episode frequency increases with speed and distance; that is, losses in objective video quality are bursty, and the probability of this "burstiness" increases at higher speeds and distances.

Table 1 contains the correlation coefficients of MOS and the three objective quality indicators. MOS enjoys the highest correlation with TVI (-0.8151), followed by Blocking and NIQE at -0.6769 and -0.3366 respectively. Given the correlations, it is clear that TVI is much better at capturing quality degradation that is effected by variations in speed and distance. This also implies that temporal artifacts have a significant impact on the quality of delay-intolerant HD mobile video streaming with a non-scalable video codec.

6.2 Modeling the Effect of Speed on Subjective Video Quality

Table 2: Modeling the effect of speed on MOS with 10-fold cross-validation

Predictor Used	ρ	RMSE	MAE
Linear Regression	0.3356	1.0453	0.9666
REPTree	0.2045	1.1435	1.0152
Bagging with REPTree	0.3125	1.0523	0.938
Bagging with linear model	0.3339	1.0489	0.9726

Next, we estimate the effect of speed on MOS. We propose the following linear regression model:

$$\widehat{MOS} = \hat{\beta}_0 + \hat{\beta}_1 Speed \quad (2)$$

The equation above has speed as the predictor variable and $\widehat{MOS}$ as the predicted MOS.

Table 2 gives the performance results for the estimation of the relationship between MOS and Speed using linear regression, REPTree and bagging. For reasons explained in 6.3, we only include per-video average speed as our predictor and exclude distance information. We use 1000 iterations for bagging. Since the sample size is only 25, we perform 10-fold cross-validation for all predictors[14]. ρ is the correlation between the predicted output and the actual MOS values. RMSE is the Root Mean Square Error while MAE is the Mean Absolute Error. Based on the ρ, RMSE and MAE, it can clearly be seen that linear regression gives the best performance when predicting the impact of speed on MOS. Notably, bagging using a linear model does not provide any improvement. This is not unexpected because bagging is only useful with linear regression when the task is subset selection, that is, when one is considering adding other attributes to obtain the best model. In our case, we only have one attribute[3].

The linear equation for estimating MOS is:

$$\widehat{MOS} = 3.6311 - 0.0621 Speed \quad (3)$$

The 95% confidence interval for $\hat{\beta}_1$ is (-0.03572, -0.0884). Since this interval does not include zero, speed has a statistically significant correlation with MOS. Even so, the low ρ value of 0.3356 suggests that speed alone has a low impact on subjective video quality.

Table 3: Impact of speed and distance on TVI (10-fold cross-validation ρ in parentheses)

Model used	ρ	RMSE	MAE
Linear Regression			
Speed	0.1881 (0.1831)	0.8116 (0.8124)	0.5357 (0.5362)
Distance	0.202 (0.1959)	0.8093 (0.8103)	0.5406 (0.5413)
Speed + Distance	0.2394 (0.2327)	0.8023 (0.8037)	0.5274 (0.5282)
REPTree			
Speed	0.3483 (0.2376)	0.7746 (0.807)	0.4998 (0.5205)
Distance	0.3391 (0.275)	0.7774 (0.8003)	0.5105 (0.5175)
Speed + Distance	0.4184 (0.3258)	0.7585 (0.7883)	0.4705 (0.4985)
Bagging with REPTree Predictor			
Speed	0.5029 (0.2743)	0.7258 (0.7967)	0.4663 (0.5094)
Distance	0.5916 (0.3138)	0.6834 (0.7887)	0.4435 (0.5058)
Speed + Distance	0.6859 (0.4049)	0.6312 (0.756)	0.3959 (0.4709)

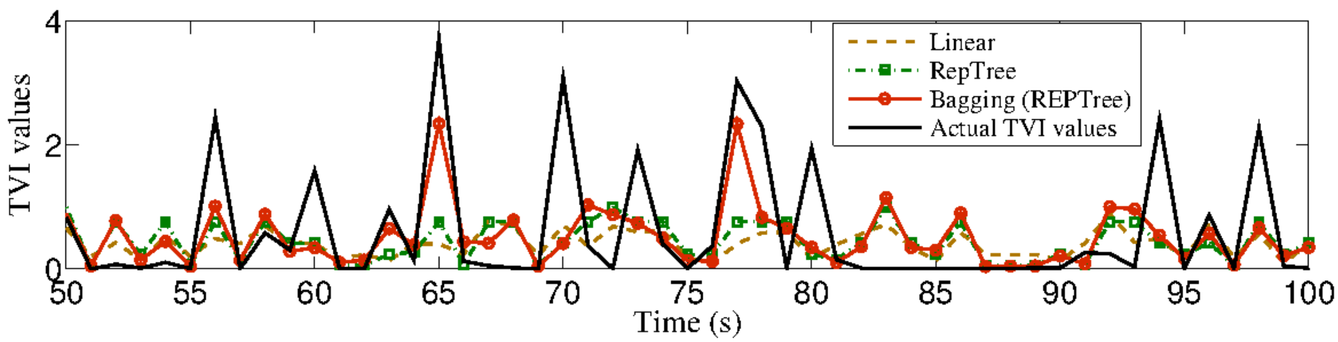

Figure 5: Plot of TVI against speed and distance using different predictors

6.3 Modeling the Effect of Speed and Distance on TVI

In section 6.1, our model assumes that the effect of speed on MOS is the same at all distances [17]. To remove this assumption, we have to include the per-video average of distance parameter as a predictor variable. However, the range of distance values is large, especially for the fast mobility scenarios. This has two challenges: first, using the per-video average distance as a predictor in this model will ouput unreliable results since the variation is unaccounted for. Secondly, the averages exclude a lot of information, since they tend to cluster around three values corresponding to the mobility scenarios.

To counter this problem, we substitute the MOS with an objective metric that is highly correlated with it, but at the same time has finer granularity. From the results above, per-second TVI is well suited for this.

In Table 3 we give the results of three different models formed by using only the speed attribute, the distance attribute and both the speed and distance attributes respectively. We took forty of the forty five received videos and found the per-second values of TVI, speed and distance for the first 50 seconds of each. This gave us a total of 2000 data points. We then used three different predictors for each model, linear regression, bagging, REPTree and bagging with REPTree. We report the values from using our model on the full training set, and include the result from 10-fold cross-validation in parentheses.

The results are markedly different from the model that was predicting MOS. The linear regression models have the worst performance; as seen in Figure 4, there is a non-linear relationship between the per-second objective metrics and motion. Bagging with the REPTree algorithm has the best performance in terms of the low ρ, RMSE and MAE. This is consistent with the assertion that bagging works well when the underlying algorithm is a regression tree [12].

All the models that only use distance as the feature vector demonstrate a higher correlation between predicted and actual TVI than those that only use speed, implying that the effect of distance is bigger than that of speed. Combining both of these feature vectors on the full training set gives the highest ρ of 0.6859 and the lowest RMSE and MAE of 0.6312 and 0.3959 respectively when bagging with REPTree is the predictor. Nevertheless, the 10-fold cross-validation ρ value of 0.4049 suggests a weak impact of both of these FVs on TVI. Figure 5 is representative of the relative performance of each predictor on the speed and distance model when we do a 70/30 training-test split on the full training set of 2000, train the predictors on the 1400 training samples, and then test them on the remaining test data.

7. CONCLUSION AND FUTURE WORK

In this work, we evaluated perceptual video quality in 802.11n wireless networks with mobile clients using both objective and subjective video quality metrics. We conducted our experiments in a live outdoor wireless testbed with little electromagnetic interference, and used three different mobility scenarios, effected by mounting the mobile clients on All-Terrain Vehicles and driving at average speeds of approximately 3mph, 8mph and 20mph for each of the scenarios. We used three different HD-quality videos encoded in MPEG-4/H.264 AVC format.

We found that the relationship between MOS and speed is best represented using a linear model, but the resulting model only had a 33% correlation and RMSE of 1.05. When both speed and distance were used to predict objective quality, bagging with REPTree was the best predictor, with a moderately high correlation of 0.69% and RMSE of 0.6312 on the full training set, but a lower correlation of 40% and higher RMSE of 0.756 when 10-fold cross-validation is applied. This indicates that although speed and distance have an effect on video quality, this effect is not overwhelming. Lastly, across all predictors, models that had distance as the lone feature vector performed better than those that only used speed.

Spurred by these results, we plan to identify other features that will provide more accurate models for video quality in a mobile WiFi setting. Another direction for future work is to develop loss-tolerant and computationally inexpensive objective VQA algorithms that more closely correlate with the MOS, especially in mobile scenarios.

8. ACKNOWLEDGMENTS

Carlo Vallati and Christopher Patton helped with conducting the experiments. An (Jack) Chan and Eilwoo Baik provided insightful feedback. We thank them all for their invaluable contributions.

9. REFERENCES

[1] gpsd – a GPS Service Daemon. `http://http://catb.org/gpsd/`. Accessed March-30-2013.

[2] gpxpy – GPX file parser. `https://github.com/tkrajina/gpxpy`. Accessed May-24-2013.

[3] BREIMAN, L. Bagging Predictors. *Machine learning 24*, 2 (1996), 123–140.

[4] BYCHKOVSKY, V., HULL, B., MIU, A., BALAKRISHNAN, H., AND MADDEN, S. A Measurement Study of Vehicular Internet Access Using *In Situ* Wi-Fi Networks. In *Proc. ACM MobiCom Conference, 2006*.

[5] CHAN, A., ZENG, K., MOHAPATRA, P., LEE, S.-J., AND BANERJEE, S. Metrics for Evaluating Video Streaming Quality in Lossy IEEE 802.11 Wireless Networks. In *Proc. IEEE INFOCOM, 2010*.

[6] CHAN, A. J., PANDE, A., BAIK, E., AND MOHAPATRA, P. Temporal quality assessment for mobile videos. In *Proc. ACM MobiCom, 2012*.

[7] CISCO, C. V. N. I. Global Mobile Data Traffic Forecast Update, 2012-2017. *Cisco Visual Networking Index (VNI) Forecast* (2013).

[8] DESHPANDE, P., HOU, X., AND DAS, S. R. Performance comparison of 3G and metro-scale WiFi for vehicular network access. In *IMC, 2010*.

[9] EFRON, B., AND TIBSHIRANI, R. *An Introduction to the Bootstrap*, vol. 57. Chapman & Hall/CRC, 1993.

[10] GASS, R., SCOTT, J., AND DIOT, C. Measurements of In-Motion 802.11 Networking. In *IEEE WMCSA, 2006*.

[11] HALL, M., FRANK, E., HOLMES, G., PFAHRINGER, B., REUTEMANN, P., AND WITTEN, I. H. The WEKA Data Mining Software: An Update. *SIGKDD Explor. Newsl. 11*, 1 (Nov 2009), 10–18.

[12] IHAKA, R., AND GENTLEMAN, R. R: A Language for Data Analysis and Graphics. *Journal of computational and graphical statistics 5*, 3 (1996), 299–314.

[13] ITU. Methodology for the Subjective Assessment of the Quality of Television Pictures. *International Telecommunication Union: Recommendation ITU-R BT.500-13* (2012).

[14] KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proc. IJACAI Conference, 1995*.

[15] KOSTUCH, A., GIERŁOWSKI, K., AND WOZNIAK, J. Performance Analysis of Multicast Video Streaming in IEEE 802.11 b/g/n Testbed Environment. In *Wireless and Mobile Networking*. Springer, 2009, pp. 92–105.

[16] MAHAJAN, R., ZAHORJAN, J., AND ZILL, B. Understanding WiFi-based Connectivity from Moving Vehicles. In *Proc. ACM SIGCOMM, 2007*.

[17] MATLOFF, N. From Algorithm to Z-Scores: Probabilistic and Statistical Modeling in Computer Science, 2009.

[18] MITTAL, A., SOUNDARARAJAN, R., AND BOVIK, A. Making a "Completely Blind" Image Quality Analyzer. *Signal Processing Letters, IEEE 20*, 3 (2013), 209–212.

[19] MOORTHY, A., CHOI, L. K., BOVIK, A., AND DE VECIANA, G. Video Quality Assessment on Mobile Devices: Subjective, Behavioral and Objective Studies. *Selected Topics in Signal Processing, IEEE Journal of 6*, 6 (2012), 652–671.

[20] OTT, J., AND KUTSCHER, D. Drive-thru Internet: IEEE 802.11b for "Automobile" Users. In *In Proc. IEEE INFOCOM, 2004*.

[21] QUINLAN, J. R. Induction of Decision Trees. *Mach. Learn. 1*, 1 (mar 1986), 81–106.

[22] QUINLAN, J. R. Simplifying Decision Trees. *Int. J. Man-Mach. Stud. 27*, 3 (sep 1987), 221–234.

[23] ROBUSTO, C. The Cosine-Haversine Formula. *The American Mathematical Monthly 64*, 1 (1957), 38–40.

[24] ROTH, G., SJÖBERG, R., LIEBL, G., STOCKHAMMER, T., VARSA, V., AND KARCZEWICZ, M. Common Test Conditions for RTP/IP over 3GPP/3GPP2. *ITU-T SG16 Doc. VCEG-M77, Austin, TX, USA* (2001).

[25] SESHADRINATHAN, K., SOUNDARARAJAN, R., BOVIK, A., AND CORMACK, L. Study of Subjective and Objective Quality Assessment of Video. *Image Processing, IEEE Transactions on 19*, 6 (2010), 1427–1441.

[26] WANG, Z., BOVIK, A., AND EVAN, B. Blind measurement of blocking artifacts in images. In *Proc. IEEE ICIP, 2000*.

[27] WU, D., GUPTA, D., AND MOHAPATRA, P. QuRiNet: A Wide-area Wireless Mesh Testbed for Research and Experimental Evaluations. *Ad Hoc Networks 9*, 7 (2011), 1221 – 1237.

[28] ZAHN, T., O'SHEA, G., AND ROWSTRON, A. Feasibility of Content Dissemination between Devices in Moving Vehicles. In *Proc. ACM CoNEXT, 2009*.