Towards Automated Quality Curation of Video Collections from a Realistic Perspective

Todd Goodall*, Maria Esteva†, Sandra Sweat‡ and Alan C. Bovik *
*University of Texas at Austin, Austin, Texas 78712
†Texas Advanced Computing Center, Austin, Texas 78758
‡Athena Health, Austin, Texas 78701

***Abstract*—We investigate the use of automated Video Quality Assessment (VQA) algorithms to evaluate digital video collections. These algorithms are driven by well-defined natural scene statistics (NSS), which capture the behavior of natural distortion-free videos. Because human vision has adapted to these real-world statistics over the course of evolution, quality predictions delivered by these NSS-based VQA algorithms correlate well with human opinions of quality. In particular, we expect these algorithms to accurately predict quality on sizable and diverse video collections. To test this hypothesis, we gathered a testbed of video clips that represent a larger video art collection. Next, we conducted a human study in which users scored the quality of the clips. Enabled by the human study, we trained three VQA algorithms (Video BLIINDS, BRISQUE, and VIIDEO) using our testbed collection to assess a real-world digital video art collection from our university museum. Two of the algorithms provided good automatic predictions of the quality of the videos. These same algorithms also highlighted limitations that arise when assessing artistic collections. We present current research progress and discuss future directions for testbed and algorithm improvement. Our ongoing effort furthers the field of Computational Archival Science by applying computational models of human perception to video appraisal and preservation tasks.**

1. Introduction

Museums, libraries and archives are amassing significant digital video collections for which appraisal and curation decisions need to be made. Among other criteria, these decisions concern video format, content, provenance, authorship, and condition. Evaluating video condition is a quality assessment task which includes understanding the types and degrees of distortions present throughout a single video. Any diverse video collection can be expected to represent a plethora of different encoding formats with individual videos containing perhaps multiple types of distortions. The presence of these distortions could be inherent to the file format/encoding in which the video was created, or they could be acquired through reformatting and transcoding processes across a video's lifecycle. Understanding the condition of videos informs activities related to video appraisal, selection, preservation, and access.

However, collecting institutions have difficulty knowing the condition of their digital video collections since manual assessment is time-consuming and nuanced. Typically, condition assessment involves visual inspection by individuals who should be very knowledgeable regarding degradations inherent to analogue formats, modern digital compressions, and digitization activities. Because collected videos may include a wide variety of distortion types, there could be inconsistencies and ambiguities in the condition report done by humans. Furthermore, while there are software tools that detect individual distortions in video objects, they still require humans to decide on the overall quality of a piece. In turn, individual video quality assessment should be normalized in order to identify the condition of an entire collection in a consistent fashion. Importantly, assessment solutions need to scale with the growth of video collections.

In other application domains, such as digital television, automated assessment of video condition is accomplished using perception-based Video Quality Assessment (VQA) algorithms. Video streaming services use these algorithms to inspect incoming videos, and to later assess their post-compression quality as they are delivered over the Internet. We are interested in exploring how these kinds of algorithms, or variations of them, can be used to assess the quality condition of large video collections for long-term archiving purposes. Specifically we study general no-reference (NR) VQA algorithms: the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [1], Video BLIINDS [2], and the Video Intrinsic Integrity and Distortion Evaluation Oracle (VIIDEO) [3]. No-reference refers to methods that do not have a (presumably pristine) original source video available for direct comparison. The most successful NR algorithms predict motion picture quality using perceptually relevant Natural Scene Statistics (NSS) models, which describe statistical regularities arising in images and videos of real-world scenes. To predict a final quality score of an image or video, the algorithms use 'quality-aware' features that capture statistical departures from known models that accurately characterize pristine images and videos. These departures are strongly correlated with the presence of visual distortions. Quality-aware features are designed to be sensitive to a large or even unknown set of distortions. Common digital video distortions include, among many others, blur, noise, and blocking. In addition, these quality-aware features correlate well with human mean opinion scores (MOS) of image and video quality, and thus we expect that these algorithms may be successfully deployed to accurately predict qualities of the contents of sizable digital video collections.

To conduct automated NR VQA, digital videos are fed to a VQA algorithm that has been pre-trained on a controlled set of videos impaired by varying types and degrees of distortion and on their associated MOS or difference MOS (DMOS). Typical existing video quality controlled datasets contain a small collection of unique content, each subjected to a single synthesized distortion. As such, these datasets do not capture the entire spectrum of distortions that are observed in real-world video collections, nor do they contain mixtures of distortions, which often occur together in complex combinations. We are motivated to bring to bear recent significant advances in this field towards finding a scalable, automated, and reliable way to measure video quality for use in video archival appraisal and preservation decision tasks. Traditionally, these tasks are done by humans in a pen and paper fashion. In the context of Computational Archival Science (CAS), we are interested in the possibility of instead using efficient and accurate perception-based VQA models to ultimately conduct these archival processing tasks. This could offer the possibility of greatly improved throughput, consistency, accuracy, and scalability.

Towards these goals, we evaluated a video art collection using algorithms trained on a set of videos containing real-world distortion types. Toward this end, we assembled a testbed video collection, as described in Sec. 3. In Sec. 4, we describe a study wherein we obtained a large number of human opinions of perceived video quality. This human data was used as "golden" ground truth on which researchers can seek to understand and model how the distortions existing in the testbed impact perceived video quality. Section 5 provides a brief overview of the evaluated VQA algorithms. The video art collection and project deliverables are described in Sec. 6. In Sec. 7, we evaluate a set of VQA algorithms with the aid of High Performance Computing (HPC) resources, which allow efficient and timely testing on the larger video corpus. Finally, we present conclusions and future work in Sec. 8.

2. Previous Work

Automated video quality assessment solutions are developed to meet the needs of curators who are increasingly concerned with managing large quantities of digital video. The open source quality control tool for video preservation, VCQ, enables automated objective analysis of digitized video through multiple indicators, the results of which have to be interpreted by collection curators [4]. A different approach was undertaken by Esteva and collaborators, who tested the ability of the VQA perceptual model BRISQUE to rate the quality of individual videos on a large video art collection using training data from a controlled dataset [5]. This latter approach provides a path by which a summary condition score can be predicted based on the perceptual mechanisms that drive human opinions of quality. Thus far, VQA algorithms have only been trained on controlled collections, neglecting the realistic case of co-occurring distortions that happen in video art collections. This work furthers this direction of research by training relevant VQA algorithms on data that is representative of a video art collection.

Video editing software uses quality indicators based on certain video properties. Even though these individual measurements may be helpful in the context of editing tasks, the continuous range of video quality cannot be inferred accurately from contrast, brightness, and camera motion alone. Comparing motion information between a distorted and pristine version of a video has been quite helpful in capturing video quality [6]. However, motion has had limited application to no-reference (NR) methods due to the difficulty in capturing generalizable motion statistic regularities [7], although some specific measures may be partially indicative of video quality [8]. The trade-off between brightness and contrast can be one powerful indicator of image quality by taking into account how the human eye is sensitive to spatial frequencies [9]. However, combining spatial contrast masking with motion is nontrivial for NR quality assessment. In other words, perceptual pooling mechanisms regarding how to combine local spatial-temporal measurements when forming a single score of subjective quality are not yet well modeled.

3. Methodology

The ultimate goal of this research is to rank videos contained in a real-world video art collection on a continuous quality scale, e.g., from worst to best quality. To develop an automated system for ranking these videos, we must be able to map each video content to a single quality score. This mapping is enabled by leveraging perceptual statistics, by modeling how the human visual system encodes spatio-temporal masking [9], frequency regularities [10], scale invariance [11], and the high order structural [12], [13] properties in a given video. VQA algorithms employ various methods for measuring these statistics. Machine learning algorithms learn the mapping from all of the extracted statistics to final subjective scores. This necessarily requires a representative collection of videos with corresponding subjective quality data.

Our real-world art collection, henceforth known as the corpus, comprised of 18 videos gathered from the local university museum. We also compiled a testbed collection of 120 ten second video clips, each representative in quality and content of a typical academic video art collection. The testbed is comprised of 16 clips obtained from the corpus, 80 clips were obtained from video art students, and 24 clips from free videos discovered through the Internet Archive. These 120 clips were rated by human subjects by following standard guidelines provided in [14].

We avoided evaluating the performance of VQA algorithms trained using our testbed on controlled collections like the CSIQ [15], LIVE VQA [16], or MCL-V [17]

Figure 1. Continuous sliding scale used by subjects.

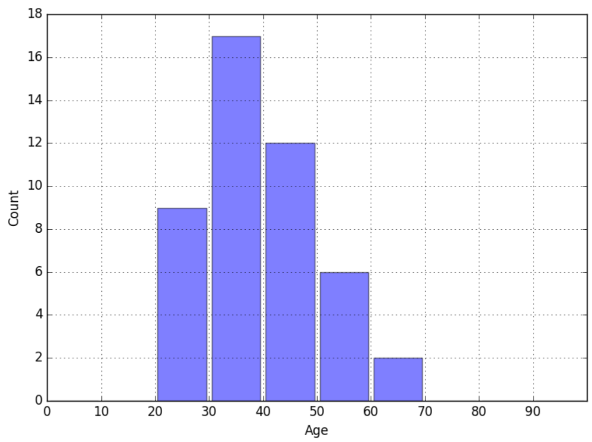

Figure 2. Distribution of ages determined from the survey.

databases. First, we found that those databases only include a small subset of the types of distortions represented in the testbed and in our corpus collections. Second, we determined that these databases and associated opinions were collected under different conditions of display distance, display type, and display resolution. Third, we realized that subjective opinions gathered for these databases assume that a pristine reference video exists, and the provided subjective scores are normalized with respect to these videos. According to Section 8 of ITU-T P.800.2 [18], we should not compare opinion scores between databases that were not designed to be compared, which validates our reasoning. For our evaluations, we used standard cross-validation procedures involving only the testbed collection.

4. Gathering Human Opinions about Video Quality

We conducted a human study to quantify video quality following the same experimental procedures used in similar studies [15], [16], [17]. Asking human observers to rate video quality mainly requires that they have normal (or corrected) vision, and that these observers understand video quality is not based on the favorability of video content. Participants were recruited via email to attend a 30-minute video rating session, in which they were instructed to rate video quality of the 10 seconds clips on a continuous sliding scale with range 0-5. This sliding scale, as depicted in Fig. 1, contains indicators marking qualitative points of "Worst," "Fair," and "Excellent." The clips were viewed on a Google Nexus 9 tablet, since tablets are an increasingly popular choice when conducting human subjective studies [19]. We also elected to use a tablet since it strikes a balance between the physical size and visual quality [20] of displayed videos

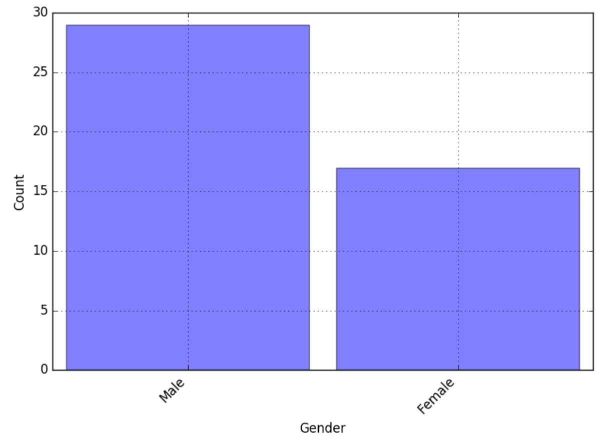

Figure 3. Distribution of genders determined from the survey.

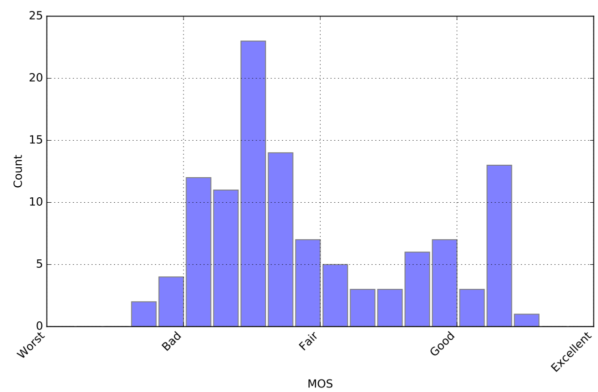

Figure 4. Distribution of mean quality scores determined from the subjective study.

which may be viewed by typical consumers of archival video content.

The study consisted of four steps administered by a researcher: pre-survey, training session, test session, and a post-survey. The pre-survey asked for demographic information, participant vision impairment, and video viewing habits. All participants then received a short training session of 6 clips to understand the rating process before starting the testing session. Researchers prepared the test room for viewing videos by dimming the lights to reduce glare, requesting that participants silence cellular devices, and removing other visual and auditory sources of distraction. After the test session, each participant completed a post-survey, containing questions about lighting, noticing distracting video content, and whether they became disengaged during the session.

A total of 45 subjects participated in the study. This is three times more than the minimum recommended number of subjects for this type of study [14]. The distribution of participant ages and genders are provided in Figs. 2 and 3 respectively. The median participant age was 37. Participants were not selected based on background knowledge, since we desired to collect ratings from naive observers, as recommended in Section 2.5 of ITU-R BT.500-13 [14]. Most subjects, 27 total, had some type of visual impairment. Among these, two subjects admitted lacking the necessary

TABLE 1. MEDIAN RESULTS FROM 1000 TRIALS OF 80%/20% TRAIN/TEST SPLITS FOR THE BRISQUE ALGORITHM.

SVR			RF		
LCC	SRCC	MSE	LCC	SRCC	MSE
0.8186	0.7771	0.3291	0.8538	0.7737	0.2605

TABLE 2. MEDIAN RESULTS FROM 1000 80%/20% TRAIN/TEST SPLITS FOR VIDEO BLIINDS AND ITS INDIVIDUAL FEATURE GROUPS. THE SYMBOL * DENOTES THAT ALL FEATURE GROUPS ARE USED.

Feature Group	SVR			RF		
	LCC	SRCC	MSE	LCC	SRCC	MSE
*	0.8014	0.7281	0.3697	0.8412	0.7688	0.2795
NIQE	0.7918	0.7253	0.4029	0.8422	0.7757	0.2774
DC Features	0.0465	0.0825	1.0571	-0.0293	0.0086	1.1863
NVS Ratios	0.3733	0.4653	0.8808	0.1649	0.1455	1.0461
Coherency	0.4218	0.4652	0.9234	0.4506	0.3845	0.8783
Global Motion	0.2688	0.3216	1.0458	0.3081	0.2662	1.0416

corrective lenses, hence they were excluded from the study. In the post-survey, all subjects agreed that the background lighting of the test room was adequate, but some participants complained about screen glare. Even though ITU-R BT.500-13 [14] recommendations were followed, 9 of the subjects were distracted during the task, and a total of 14 lost focus at some point during the test. We decided not to reject those who admitted to getting distracted, assuming any variance from distraction would be averaged out given the randomized video load order, and since we applied standard subject rejection protocols.

Given the 114 videos used for the human study session along with the 43 remaining subjects, we studied the distribution of subject scores. Based on ITU-R BT.500-13 Annex 2.2 [14] recommendation, we rejected subjects whose scores were statistically irregular. In the end, we rejected 8 subjects, leaving a final total of 35 acceptable subjects. The scores from these acceptable subjects were normalized then averaged per video to compute Mean Opinion Scores (MOS). The distribution of these normalized scores as depicted in Fig. 4 shows a broad spread of video qualities, with a larger representation of “Bad” and “Good” quality videos.

5. VQA Algorithms

No-Reference (NR) algorithms are used to measure the perceived quality of images and videos when there is no original or pristine version available for comparison [21] [22]. We propose that NR algorithms could be useful to understand a collection’s quality without the need for humans to review each video. We used three different algorithms developed at the Laboratory for Image and Video Engineering: the Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE), Video BLIINDS, and the Video Intrinsic Integrity and Distortion Evaluation Oracle (VIIDEO). These algorithms take very different approaches to quality assessment, but each fundamentally relies upon Natural Scene Statistics (NSS) modeling.

BRISQUE measures the spatial NSS of images and predicts a score per frame. It first computes Mean-Subtracted Contrast Normalized (MSCN) coefficients, which have been observed to follow a Gaussian distribution for natural images. When natural images are subjected to distortion, the distribution of these MSCN coefficients deviates predictably. To produce a final rating on a video, BRISQUE perceptual features are averaged across frames before applying a machine learning method to predict a subjective quality score.

Video BLIINDS measures spatial-temporal statistics to predict a score for each video by incorporating multiple feature groups. The first feature group, “Spatial Naturalness,” is computed using the Naturalness Image Quality Evaluator (NIQE) [23], which measures picture “naturalness.” The second feature group, “DC features,” detects irregular changes in average luminance across frames, often caused by compression-induced flicker. The “NVS-shape parameter ratios” group measures the regularities arising from frame differences. The “Coherency Measure” group is a statistical measurement of the dominance of motion over local patches. Lastly, the “Global Motion Measure” captures ego-motion that may affect user perception of quality.

VIIDEO uses frame differencing and empirical space-time correlation features computed over blocks to measure inter and intra-frame difference statistics. The output of VIIDEO is a single quality score, which, unlike previous algorithms, is produced with no machine learning for mapping features to human scores. This, in principle, increases its generality, albeit at the cost of reduced predictive power.

6. The Corpus and Open Data

Most of the videos in our corpus were produced between the years 2000 and 2008 and were distributed using DVD/MPEG-2 format. The remainder of the video art was developed and stored using VHS formats, implying that these pieces had gone through various digitization and transcoding processes before they were distributed on DVD. Most of the videos contain obvious distortions. Some of them contain segments that do not correspond to natural scenes.

Due to copyright conditions, we are unable to provide the clips as open data, as is customary with other video databases. However, we are making the features of Video BLIINDS and BRISQUE that were computed on the database available online along with associated MOS scores [24]. This information will allow others to make use of our work.

7. Results

Before evaluating the algorithms, all videos in the corpus and testbed were fit to the 2048x1536 native display resolution of the Nexus 9 tablet. After this preprocessing step, our evaluation involved three components. First, in Subsec. 7.1, we study overall testbed performance. Second, in Subsec. 7.2, we compute automatic video quality predictions by applying the algorithms to be compared on our video corpus collection. Third, in Subsec. 7.3, we study model prediction outliers.

7.1. Algorithm Evaluation on the Testbed

We evaluated the algorithms and their feature groups, which allowed us to identify those regularities that best captured departures from naturalness. We also identified which feature groups contributed the most to the accuracy of the final predictions. Since the content in our video collection was compiled from unique scenes, we followed a standard cross-validation procedure of randomly splitting the dataset into a training portion of 80 percent, and a testing portion of 20 percent. We repeated this randomized training/test split technique 1000 times, collecting prediction results on each test set per trial. The repeated trials ensured that we obtained overall measures of performance that were not affected by a poorly represented training or testing subset. We report the median Linear Correlation Coefficient (LCC), Spearman's Rank Correlation Coefficient (SRCC), and Mean-Squared Error (MSE) over these trials. To formulate predictions from the feature groups, two standard machine learning models were used, Support Vector Regression (SVR) [25] and Random Forest Regression (RF) [26]. Note that VIIDEO does not need a learning model, since it provides a direct score.

To provide an upper bound on algorithm performance, we correlated half of the human mean opinion scores against the other half. The SRCC between these two groups was found to be 0.9667, which is high subject agreement. The best performing algorithm would be expected capture the average subject's opinion, which should correlate highly with the average opinions collected in our study given this high subject agreement. Among the three algorithms, we observed good performance from both BRISQUE and Video BLIINDS, and poor performance from VIIDEO, which was found to yield an SRCC of -0.2386. Therefore, we did not further evaluate this model. Tables 1 and 2 compare BRISQUE and Video BLIINDS, the feature groups, and the machine learning models. From these results, we first conclude that spatial information, i.e. features computed on single frames, contributed the most to the overall successful prediction. In fact, the temporal information (computed on multiple frames, such as motion) provided by the "DC Features," "Coherency Measure," and "Global Motion Measure" feature groups provided little contribution to the prediction accuracy. Since this has not been observed on other video datasets, which are typically motion-intensive, this may reflect the type of used video content in the museum collection. When comparing between SVR and RF machine learning models, RF provided the more accurate model.

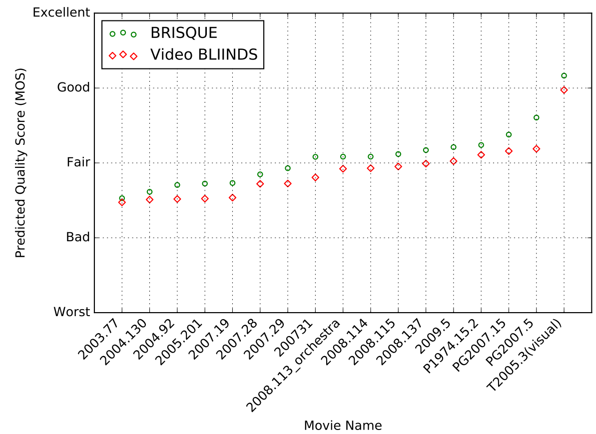

Figure 5. BRISQUE and Video BLIINDS final predictions for the corpus collection using the RF regressor.

7.2. Algorithm Evaluation on Museum Corpus

By training BRISQUE and Video BLIINDS on the video art testbed, we were able to obtain automatic predictions of video quality on it. Since most real-world videos are longer than 10 seconds, quality was predicted on 10-second segments with the final score being computed based on the average of these intermediate predictions. We utilized high-performance computing (HPC) resources to compute quality predictions within an acceptable period of time, since most of the videos were of lengths on the order of 50,000 frames. Figure 5 depicts these final predicted scores when using the RF regression model on the museum collection. We can see that the two algorithms agree well, achieving an SRCC of 1.0 and an MSE of 0.0627 against each other. This agreement makes sense, since the dominant feature group in Video BLIINDS is NIQE, which uses the same kind of features as BRISQUE. These results may be used to allow a curator to understand the quality of each video art piece in relation to the rest of the collection.

7.3. Evaluation of Outliers

To determine which videos in the video art testbed collection of 114 videos were not predicted well by the algorithms, we used leave-one-out cross-validation, training on 113 clips, then predicting the quality of the left-out clip. This yields the best possible prediction on the left-out video, providing information about the types of video distortions that are not captured, due to limitations that may exist in either BRISQUE or Video BLIINDS.

By measuring the differences in predicted quality vs. assessed quality, we found that 20 predictions differed from the MOS scores by at least 0.625, the distance halfway between two qualitative points on the quality scale. After visually inspecting these outliers, we noted that they tend to include non-natural artistically rendered scenes, heavy VHS artifacts, dark scenes, color distortions, out-of-focus regions, and black borders. These types of distortions are not well accounted for in the tested algorithms. In addition, these algorithms cannot capture color distortions since they operate only on pixel luminance.

8. Conclusions and Future Work

We achieved our research goal of ranking the quality of individual video art pieces and of producing a normalized collection assessment, as shown in Fig. 5. However, we also found that despite training on a realistic testbed, the selected algorithms did not capture the entirety of distortions present in the video art collection. This has motivated us to work towards collecting a more representative testbed for training VQA algorithms for archival applications. We are gathering a publicly releasable testbed, to avoid licensing restrictions, and we expect that the videos in this new testbed will demonstrate more diversity of content and qualities. In addition, distribution of this testbed along with the collected human opinions will allow other researchers to make progress on this complex problem.

From the gap between algorithm performance and subject agreement, we learned that existing VQA algorithms still need further improvement. At a minimum, these algorithms must account for the outliers observed in Subsec. 7.3. We plan to utilize the naturalness features that demonstrated good performance in Table 2 to produce better VQA algorithms suitable for this task. We also plan to research new features that describe general distortions by employing feature-learning paradigms [27], [28] on large video collections. From a CAS perspective, we demonstrated progress toward automating a real-world archival problem, but we need to conduct more research to further address this problem.

Acknowledgement

The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources that have contributed to the research results reported herein. URL: http://www.tacc.utexas.edu

References

[1] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, 2012.

[2] M. A. Saad, A. C. Bovik, , and C. Charrier, "Blind prediction of natural video quality," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1352–1365, 2010.

[3] A. Mittal, M. A. Saad, and A. C. Bovik, "A completely blind video integrity oracle," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 289–300, 2016.

[4] B. Turkus, "Drop video file(s) here: the emergence of free quality control tools for video preservation," 2015. [Online]. Available: https://www.bavc.org/sites/default/files/resource/QCTools_WhitePaper2.pdf

[5] M. Esteva, A. Bowen, T. Goodall, A. Bovik, and Z. Abel, "Evaluation of non-reference quality assessment algorithms to curate born-digital video collections," *IS&T Archiving Conf.*, vol. 2015, no. 1, pp. 124–129, 2015.

[6] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 335–350, Feb 2010.

[7] D. W. Dong and J. J. Atick, "Statistics of natural time-varying images," *Network*, vol. 6, no. 3, pp. 345–358, 1995.

[8] M. A. Saad and A. C. Bovik, "Natural motion statistics for no-reference video quality assessment," in *QoMEx 2009*. IEEE, 2009, pp. 163–167.

[9] D. J. Swift and R. A. Smith, "Spatial frequency masking and weber's law," *Vision Res.*, vol. 23, no. 5, pp. 495–505, 1983.

[10] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am.*, vol. 4, no. 12, pp. 2379–2394, 1987.

[11] ——, "Scale-invariance and self-similar 'wavelet' transforms: an analysis of natural scenes and mammalian visual systems," *Wavelets, Fractals, and Fourier Transforms: New Developments and New Applications*, 1993.

[12] R. Baddeley, "The correlational structure of natural images and the calibration of spatial representations," *Cogn. Sci.*, vol. 21, no. 3, pp. 351–372, 1997.

[13] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 49–70, 2000.

[14] "ITU-BTR13," 2017. [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.500-13-201201-I!!PDF-E.pdf

[15] P. V. Vu and D. M. Chandler, "Vis3: an algorithm for video quality assessment via analysis of spatial and spatiotemporal slices," *J. Electron. Imaging*, vol. 23, no. 1, pp. 013 016–013 016, 2014.

[16] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1427–1441, 2010.

[17] J. Y. Lin, R. Song, C.-H. Wu, T. Liu, H. Wang, and C.-C. J. Kuo, "MCL-V: A streaming video quality assessment database," *J. VIS. COMMUN. IMAGE R.*, vol. 30, pp. 1–9, 2015.

[18] "Recommendation P.800.2," 2017. [Online]. Available: https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-P.800.2-201607-I!!PDF-E

[19] A. Bovik, C. Bampis, and T. Goodall, "Perceptual issues of streaming video," *SID*, May 2017.

[20] H. Mller, J. Gove, and J. Webb, "Understanding tablet use: a multi-method exploration," *MobileHCI 2017*, pp. 1–10, 2012.

[21] A. C. Bovik, "Automated prediction of image and video quality," *Proceedings of IEEE*, vol. 101, no. 9, pp. 2008–2024, Sept. 2013.

[22] A. K. Moorthy and A. C. Bovik, "Visual quality assessment algorithms: What does the future hold?" *Multimed Tools Appl.*, vol. 51, no. 2, pp. 675–696, 2011.

[23] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, 2013.

[24] "Automated quality curation of video collections." [Online]. Available: http://live.ece.utexas.edu/research/automated_video_collections/index.html

[25] D. Basak, S. Pal, and D. C. Patranabis, "Support vector regression," *NIP-LR*, vol. 11, no. 10, pp. 203–224, 2007.

[26] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.

[27] W. Zhou, W. Qiu, and M.-W. Wu, "Utilizing dictionary learning and machine learning for blind quality assessment of 3-D images," *IEEE Trans. Broadcast.*, 2017.

[28] J. Kim, H. Zeng, D. Ghadiyaram, S. Lee, L. Zhang, and A. C. Bovik, "Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 130–141, 2017.