A MATLAB-BASED FRAMEWORK FOR IMAGE AND VIDEO QUALITY EVALUATION

Adithya V. Murthy and *Lina J. Karam,*
School of Electrical, Computer & Energy Engineering
Arizona State University, Tempe, AZ 85287-5706
{adithya.murthy,karam}@asu.edu

ABSTRACT

This work provides a MATLAB-based framework for image and video quality evaluation that integrates test content generation, subjective testing interfaces, objective quality metric evaluation and correlation analysis into a single application. It also provides an HTML-based help system and a Graphical User Interface that makes applications based on the framework user friendly. The framework supports a configurable interface that makes it easy to integrate existing or new quality metrics and to compare them against each other. The framework abstracts the user from underlying details like file format, report generation and test flow control.

Index Terms— Image quality, video quality, objective metrics, framework, Graphical User Interface

1. INTRODUCTION

The ubiquitous presence of multimedia has resulted in content being delivered and consumed in a wide variety of formats over a variety of media, channels and at different bitrates. Differing display resolutions have resulted in different frame dimensions. Bandwidth and storage space limitations have resulted in different compression schemes and bitrates. The techniques used to accommodate this diversity sometimes result in the degradation of the quality of the content due to lossy compression, loss of detail due to resizing and so on. Researchers in academia and industry are interested in measuring how a certain process can affect image or video quality. This information can be used to determine the Quality of Service (QoS) by in-service monitoring of transmission quality, or to assess the overall user experience when viewing multimedia content. It can also be used to measure the performance of a system or to optimize an algorithm [1].

Human beings are the ultimate consumers of the multimedia content. Hence, traditional techniques of evaluating image or video quality have involved human subjects. Subjective evaluation of image and video quality is the most reliable. However, it is a costly, cumbersome and time-consuming process because complex equipment must be used, the environment must be controlled, and subjects have to manually view and rate content. It is also not suitable for in-service monitoring.

Researchers have developed objective quality metrics to overcome some of the constraints of subjective image and video quality evaluation techniques. Evaluation of an objective metric requires one to have access to test content, subjective ratings of that content, access to objective quality metrics and access to methods to correlate the objective metric's prediction of quality and the subjective rating. Video sequences for video quality testing are publicly available at [2], [3] and [4]. Publicly available image databases are provided by [5], [6], [7] and [8]. If test content does not exist, it must be generated by simulating conditions appropriate to the application. Some databases like [4], [5], [6], [7] and [8] provide subjective scores. However, if subjective scores do not already exist, they must be generated by asking subjects to evaluate the quality of the test content. Reference implementations for some objective quality metrics are publicly available. However, comparing two metrics is not straightforward. This is because they might require data in different file formats (e.g., The MOVIE index [9] requires input to be YUV4:2:0 planar while the VQM metric [10] accepts AVI or interleaved UYVY files). Reference implementations for parsing different file formats and other operations already exist. However, it is time consuming to integrate all these implementations and to tweak the code to suit the research objectives. A framework that provides support infrastructure, libraries and other software to facilitate easier integration and development abstracts the researcher from these details, thus enabling them to become more productive.

This paper is organized as follows. Section 2 briefly discusses the existing frameworks. Section 3 discusses the proposed framework. Section 4 describes a visual blur assessment application built using the framework. Section 5 describes an image and video quality assessment application built using the framework. Section 6 concludes this paper and discusses areas for future work.

2. EXISTING FRAMEWORKS

This section provides an overview of existing attempts to develop image / video quality assessment applications.

A quality evaluation toolbox has been implemented by Sprljan [11]. This toolbox is command line based. It provides a few objective measures of quality. However, it does not have additional code for correlation analysis, subjective testing and test content generation. It is also difficult to integrate a new metric in this toolbox without modifying the source code.

Another package for visual quality assessment has been provided by Gaubatz [12]. This application is also command line based and does not have code for correlation analysis. This package includes error statistics based metrics and metrics based on the Human Visual System (HVS).

An image evaluation system for evaluating videos and images based on perceptual quality measures has been proposed by Bellofiore *et al* in [13]. It has the ability to integrate the user-specified codecs, noise generators, image data and quality metrics. It provides a graphical user interface (GUI). However, it is a proprietary system built in Visual C++ for Microsoft Windows and incorporating new metrics require modifying the C++ source code. Furthermore, it does not incorporate any of the recent image and

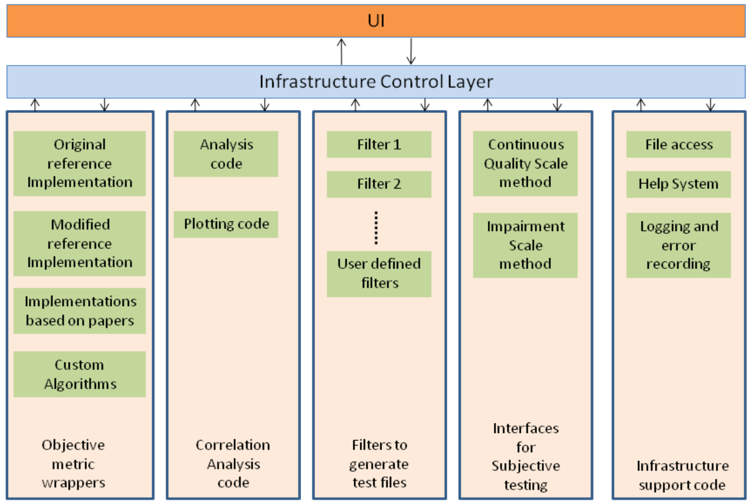


Figure 1. Block diagram of the proposed image and video quality evaluation framework.

Figure 2. User interface for subjective testing using the continuous quality scale. The 768 x 512 "bikes" image is taken from the LIVE database [5].

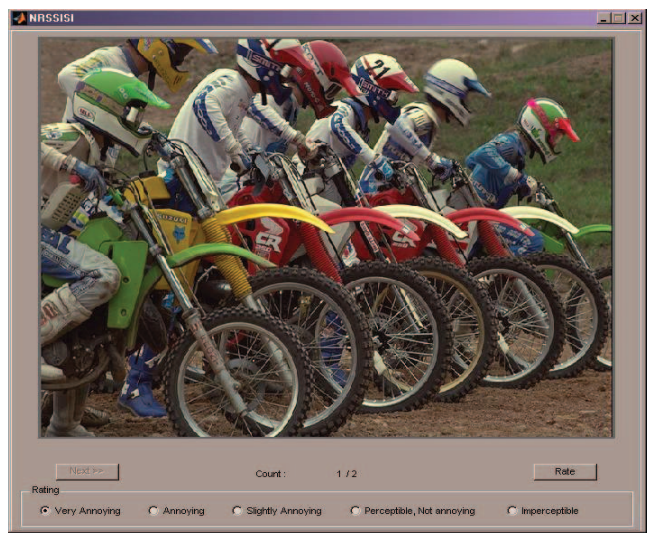


Figure 3. User interface for subjective testing using the impairment scale. The 768 x 512 "bikes" image is taken from the LIVE database [5].

video quality metrics and does not allow correlation analysis and analysis of raw subjective scores.

3. PROPOSED FRAMEWORK

This paper proposes a MATLAB-based framework for image and video quality evaluation. The framework includes a platform-independent user-friendly graphical user interface that abstracts the users from coding details and helps them in focusing on the image and video quality assessment tasks.

MATLAB was chosen because it is easy to program and many reference implementations of quality metrics are implemented in MATLAB. MATLAB enables the user to call C routines and libraries. Thus the framework can run MATLAB-based code as well as implementations in other languages.

The help system used in the framework is HTML-based. One can present screenshots, tables and animations in an HTML page. HTML also enables the creation of hyperlinks to other sections in the help system, external references and to the websites that provide the reference implementations of the work described in published literature. HTML pages can also be used on a website that provides applications built on this framework for download.

5.1. Block diagram of the framework

Figure 1 shows the block diagram of the framework. The framework consists of seven main modules. The first module consists of the code to control the GUI. The second module consists of code to control the infrastructure and program flow. This latter code is responsible for setting up loops for multi-file

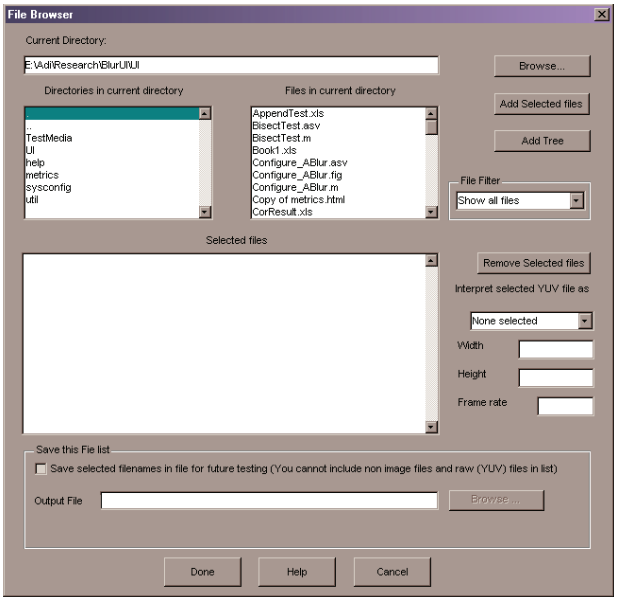


Figure 4. The file browser.

and multi-metric processing. The third module is the infrastructure support code, which consists of code to perform file access, format conversion, error logging and an HTML-based help system. The fourth module consists of interfaces for subjective quality testing. Both continuous-quality-scale and impairment-scale based subjective assessment methods are supported. Figure 2 and Figure 3 show the interfaces for subjective testing. The fifth module consists of code to perform correlation analysis of subjective and objective scores. It also contains code to generate scatter plots of the Mean Opinion Score (MOS) and the objective score. The sixth module consists of code to generate test content. It is able to cascade filters to generate test content. It also provides the ability to dynamically preview the output visual content at every stage of the cascaded filters. The seventh module is the objective visual quality assessment module. The objective quality metrics are not directly called by the control code but through a wrapper. The wrapper makes it possible to link libraries and to perform operations (e.g., format conversion) specific to a particular metric. The objective quality metrics are grouped as follows: 1) Original reference implementations, which include implementations downloaded from publicly available websites; 2) Modified reference implementations, which include implementations that have been modified to suit the framework infrastructure; example modifications include those made to file access and error logging; 3) Implementations based on information provided in published papers; these non-original implementations may not accurately match the results in the papers because of assumptions made about parameters not given in the papers; 4) Implementations of metrics that do not fall into any of the above mentioned categories. Users intending to test and compare their metrics against existing metrics can place their metrics here.

5.2. Framework features

The framework provides a powerful file browser that enables the user to select files for processing. The file browser may be used to include files from different directories. One can also mix file types. (e.g., JPEG and YUV files). The file browser enables the user to generate spreadsheets containing file names and associated file parameters (e.g., width, height and subsampling information for YUV files) for future testing. It can also read spreadsheets when used for batch processing of multiple files. The file browser is shown in Figure 4.

The framework can write the generated output into Microsoft Excel, Comma Separated Value (.csv) files, text files or HTML files. The support for spreadsheets enables the user to analyze the stored data further using formulae and plotting functionality provided by the spreadsheet. The text-based output is useful for users who wish to store results without further processing. HTML-based output is useful for users who might want to report the results on a website.

The framework is versatile because it is configurable. It uses text-based configuration files to specify the available metrics, filters for test media generation and supported file formats. In the configuration files, the first column is a user-given name that describes the metric, filter or file format. This description is seen by the user in the GUI. The next column has a three/four letter code that is a suffix to a MATLAB function that implements the required metric or filter. This is illustrated in Figure 5. In Figure 5, the name of the configuration file is "INRMetrics.csv". The contents of the first column are displayed in the list box of the GUI. The actual functions are implemented in functions named "INR_Metrics_var.m", "INR_Metrics_spectrum.m" and so on.

By editing the configuration file and ensuring that the relevant MATLAB function files or wrappers are present in the right locations, one can easily change the metrics and filters in the application. For example, by storing the descriptions and code names of blur assessment metrics in the configuration file that specifies the available metrics, the framework can be used to build a blur assessment application. To change the application to one that detects blockiness, one would have to replace the blur assessment metrics in the configuration file by the names of the blockiness metrics and save the relevant MATLAB functions that implement the blockiness metrics in the relevant folders.

The framework also enables analysis to be performed on the raw subjective scores that have been obtained after subjective testing. The framework allows one to calculate Z-scores, MOS and Differential MOS (DMOS) and standard deviations of the scores.

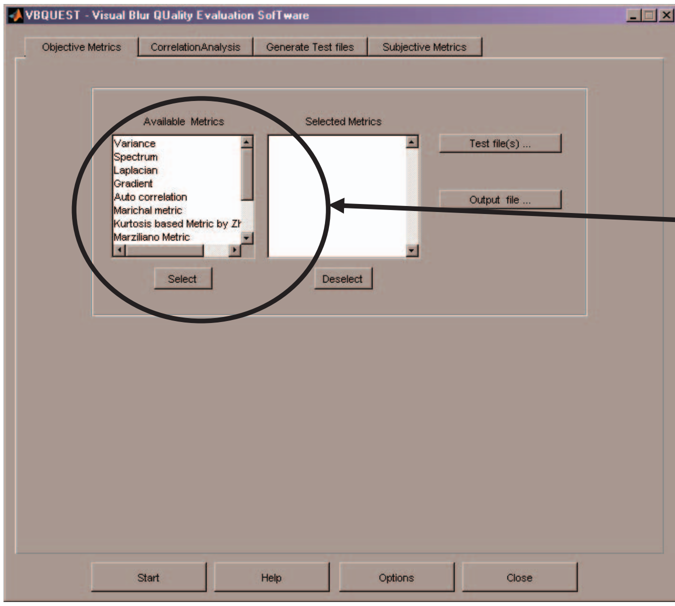
The framework also enables the user to analyze the correlation between the objective metrics score and the subjective ratings. It can calculate the Pearson correlation coefficient, Spearman rank order coefficient, Root mean square error, mean absolute error and the outlier ratio. It can also plot the scatter plot of the MOS and the objective score. It can also plot the predicted MOS on the scatter plot and highlight outliers.

The proposed framework also provides the user with several other options, such as the ability to specify a user-defined logistic function, progress reporting, error logging and so on.

4. VISUAL BLUR QUALITY EVALUATION SOFTWARE (VBQUEST)

This section describes a no reference image blur assessment application developed using the framework.

The VBQUEST application includes a Gaussian filter, averaging filter, pillbox filter and a motion blur filter for generating test content. It also enables the user to specify a custom filter through a spreadsheet file or a text file. The VBQUEST application includes most of the popular existing no-reference blur metrics such as the perceptual blur metric by Marziliano *et al*[14], the Just Noticeable



Contents of "INRMetrics.csv"

Variance,var
Spectrum,spectrum
Laplacian,lap
Gradient,grad
Auto correlation,auto
Marichal metric,marichal
Kurtosis based Metric by Zhang,zhang
Marziliano Metric,marziliano
HP metric,hp
Riemannian Tensor based metric,riem
Kurtosis of Wavelet Coefficients,wavelet
Just Noticeable Blur,jnbm
Cumulative Probability Blur Detection, cpbd

Figure 5. Using configuration files to set metrics, filters and file formats.

Table 1. Performance of some of the no reference blur metrics for distorted images from the LIVE database [5].

Metric	Distortion	Pearson Correlation coefficient	Spearman rank order coefficient	Root Mean Square Error	Mean Absolute Error
Variance [17]	Gaussian Blur	0.168	0.250	17.095	14.341
	JPEG2K Compression	0.088	0.215	17.174	15.444
Marziliano [14]	Gaussian Blur	0.840	0.853	9.266	6.714
	JPEG2K Compression	0.758	0.718	11.254	8.08
JNBM [15]	Gaussian Blur	0.839	0.839	9.304	7.189
	JPEG2K Compression	0.690	0.653	12.474	9.434
CPBD [16]	Gaussian Blur	0.928	0.945	6.384	4.731
	JPEG2K Compression	0.879	0.845	8.212	6.067

Table 2. Performance of some of the full reference image metrics for distorted images from the LIVE database [5].

Metric	Pearson Correlation coefficient	Spearman rank order coefficient	Root Mean Square Error	Mean Absolute Error
MSE	0.880	0.888	8.222	6.246
PSNR	0.563	0.698	14.326	11.327
SNR	0.563	0.698	14.326	11.327
SSIM [18]	0.905	0.901	7.384	5.491
MSSIM [18]	0.928	0.926	6.457	4.784
VSNR [19]	0.925	0.918	6.581	4.907
VIF [20]	0.964	0.95	4.632	3.549
VIFP [20]	0.948	0.937	5.512	4.148
UQI [21]	0.902	0.902	7.480	5.424
IFC [22]	0.936	0.926	6.124	4.593
NQM [23]	0.563	0.698	14.33	11.327
WSNR [24]	0.563	0.698	14.33	11.327

Table 3. Performance of the different video metrics for the LIVE video database [4].

	Pearson correlation coefficient	Spearman rank order coefficient	Root Mean Square Error	Mean Absolute Error
VQM [10]	0.73	0.714	7.223	5.898
PSNR	0.503	0.485	9.128	7.548
VSNR	0.663	0.643	7.904	6.711

Blur Metric (JNBM) by Ferzli and Karam [15], and the Cummulative Probability of Blur Detection (CPBD) metric by Narvekar and Karam [16]. It also provides interfaces for subjective testing and correlation analysis.

Table 1 shows the performance of the aforementioned perceptual quality metrics for distorted images from the LIVE database [5]. The perceptual quality based metrics perform better than error statistics based metrics (for e.g., variance). Results for the other databases are available in [25]. The results in Table 1 are consistent with the results available in [26]. The current version of the VBQUEST application is available for download at http://ivulab.asu.edu/Quality/VBQUEST.

5. IMAGE AND VIDEO QUALITY EVALUATION SOFTWARE (IVQUEST)

The previous section illustrated the application of the framework to visual blur quality evaluation. This section describes an application for image and video quality evaluation. The IVQUEST interface is similar to the VBQUEST interface. However, it enables the user to specify reference files in addition to test files. It also supports image and video quality metrics. It also includes more filters for test content generation.

The current version IVQUEST application includes full reference, reduced reference and no reference metrics for image quality assessment and full reference metrics for video quality assessment. Reduced-reference and no-reference video quality metrics can also be easily added by users due to the user-friendly reconfigurability feature provided by the proposed framework.

The IVQUEST application was tested by passing images from the databases [5], [6], [7] and [8] to the incorporated full reference, reduced reference and no reference metrics. It was also tested by passing the video sequences from [4] to the video metrics. Table 2 shows the performance of the full reference image metrics for distorted images from the LIVE database [5]. The results reiterate the fact that perceptual quality metrics (for e.g., SSIM [18], VSNR [19], VIF [20], UQI [21] and IFC [22]) perform better than the error statistics based metrics (for e.g., MSE and PSNR). The results in Table 2 are consistent with results reported in the literature. Table 3 shows the performance of the full reference video metrics for distorted videos from the LIVE database [4]. In Table 3, PSNR and VSNR are full reference image metrics that have been applied to each frame of the video. The resulting score has been obtained by taking the mean of the scores of the individual frames. Hence, temporal effects on visual quality have not been taken into account. VQM correlates better than PSNR or VSNR because it takes spatio-temporal regions into account and is therefore able to predict visual quality better. The results in Table 3 are consistent with the results reported in [27]. More results and details are available in [25].

The current version of the IVQUEST application is available for download at http://ivulab.asu.edu/Quality/IVQUEST.

7. CONCLUSION

This work described a framework for image and video quality evaluation. It supports single file and batch processing. The framework is also useful for education. A user learning about image or video quality, could use the filters to generate test content thus learning about the effect of different processes on perceived quality. The user could then see how different metrics assess the content and finally correlate the metrics' prediction of quality with the subjects' perception of quality.

This work may be extended by integrating common encoders like a JPEG2K encoder, JPEG encoder or H.264 encoder to generate test content. More metrics could also be included. The current implementation only includes interfaces for subjective testing of images. While MATLAB based players exist for video, they do not have an interface for instantaneous rating of video quality. It might be useful to develop and integrate a video player that provides an interface for instantaneous video quality rating. Another extension could be to develop a web-based objective quality evaluation application using the proposed framework. A compiled executable could be run on a web server. The front end GUI could be in HTML and hosted on a website. A user could then visit the website, specify the Uniform Resource Locator (URL) of the reference and/or test files and choose the metrics to be run. The server could then evaluate the metrics and return the results in the form of an HTML page or through email.

REFERENCES

[1] U. Engelke, H-J Zepernick, "Perceptual-based quality metrics for image and video services: a survey," 3rd EuroNGI Conference on Next Generation Internet Networks, May 2007, pp. 190-197.

[2] The Video Quality Experts Group database. [Online]. Available: ftp://vqeg.its.bldrdoc.gov/

[3] The Consumer Digital Video Library. [Online]. Available: http://www.cdvl.org/

[4] (2009) LIVE Video Quality Database. [Online]. Available: http://live.ece.utexas.edu/research/quality/live video.html

[5] H.R. Sheikh, Z. Wang, L. Cormack and A.C. Bovik, "LIVE Image Quality Assessment Database Release 2," [Online]. Available: http://live.ece.utexas.edu/research/quality

[6] P. L. Callet, and F. Autrusseau, Subjective quality assessment IRCCyN/IVC database. [Online]. Available: http://www.irccyn.ec-nantes.fr/ivcdb/

[7] Media Information and Communication Technology (MICT) Lab, University of Toyama, Japan, MICT Image Quality

Evaluation Database. [Online]. Available: http://mict.eng.u-toyama.ac.jp/mict/index2.html

[8] N. Ponomarenko, M. Carli, V. Lukin, K. Egiazarian, J. Astola, and F. Battisti "Color Image Database for Evaluation of Image Quality Metrics," in Proceedings of the International Workshop on Multimedia Signal Processing, Australia, Oct. 2008, pp. 403-408. [Online]. Available: http://ponomarenko.info/tid2008.htm

[9] K. Seshadrinathan and A.C. Bovik, "Motion Tuned Spatio-temporal Quality Assessment of Natural Videos," IEEE Transactions on Image Processing, vol. 19, no. 2, pp. 335-350, February 2010.

[10] M. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," IEEE Transactions on Broadcasting, vol. 50, no. 3, pp. 312-446, September 2004. [Online]. Available: http://www.its.bldrdoc.gov/vqm/

[11] Nikola Sprljan. [Online]. Available: http://www.sprljan.com/nikola/matlab

[12] MeTriX MuX Visual Quality Assessment Package. [Online]. Available: http://foulard.ece.cornell.edu/gaubatz/metrix_mux/

[13] S. Bellofiore, L. J. Karam, W. Metz and T. Acharya, "A Flexible and User-Friendly Image Quality Assessment System," IASTED International Conference on Signal and Image Processing, Las Vegas, Nevada, December 1997.

[14] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Applications to JPEG2000," Signal Processing: Image Communications., vol. 19, no. 2, pp. 163–172, Feb. 2004.

[15] R. Ferzli and L. J. Karam, "A No-Reference Objective Image Sharpness Metric Based on the Notion of Just Noticeable Blur (JNB)," IEEE Transactions on Image Processing, vol. 18, no. 4, pp. 717-728, April 2009. [Online]. Available: http://ivulab.asu.edu/Quality/JNBM

[16] N. Narvekar and L. J. Karam, "A No-Reference Perceptual Image Sharpness Metric Based on the Cumulative Probability of Blur Detection," First International Workshop on Quality of Multimedia Experience (QoMEX), pp 87-91, July 2009. [Online]. Available: http://ivulab.asu.edu/Quality/CPBD

[17] S. Erasmus and K. Smith, "An automatic focusing and astigmatism correction system for the SEM and CTEM," J. Microscopy, vol. 127, pp. 185–199, 1982.

[18] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004. [Online]. Available. http://www.ece.uwaterloo.ca/~z70wang/research/ssim/

[19] D.M. Chandler and S.S. Hemami, "VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images," IEEE Transactions on Image Processing, vol. 16, no. 9 pp. 2284-2298, Sept. 2007.

[20] H. R. Sheikh and A. C. Bovik, "Image Information and Visual Quality," IEEE Transactions on Image Processing, vol. 15, no. 2, pp. 430- 444, February 2006.

[21] Z. Wang and A. C. Bovik, "A Universal Image Quality Index," IEEE Signal Processing Letters, vol. 9, no. 3, pp. 81-84, March 2002. [Online]. Available: http://www.cns.nyu.edu/~zwang/files/research/quality_index/demo.html

[22] H. R. Sheikh, A. C. Bovik, and G. D. Veciana, "An Information Fidelity Criterion for Image Quality Assessment Using Natural Scene Statistics," IEEE Transactions On Image Processing, vol. 14, no. 12, pp. 2117-2128, December 2005.

[23] N. Damera-Venkata, T. Kite, W. Geisler, B. Evans and A. Bovik, "Image Quality Assessment Based on a Degradation Model," IEEE Transactions on Image Processing, Vol. 9, No. 4, Apr. 2000.

[24] T. Mitsa and K. Varkur, "Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 5, April 1993, pp. 301–304.

[25] A. V. Murthy, "A MATLAB-Based Framework for Image and Video Quality Evaluation," Master's thesis, Arizona State University, May 2010.

[26] N. D. Narvekar, "Objective No-Reference Visual Blur Assessment," Master's thesis, Arizona State University, Dec. 2009.

[27] K. Seshadrinathan, R. Soundararajan, A. Bovik and L. Cormack, "Study of Subjective and Objective Quality Assessment of Video," to appear in the IEEE Transactions on Image Processing, 2010.