

# VULNERABILITY ASSESSMENT OF IMAGE AND VIDEO QUALITY ESTIMATORS

*William Cheswick, David Kormann, and Amy R. Reibman*

AT&T Labs – Research

## ABSTRACT

In this paper, we perform vulnerability assessment of image and video quality estimators (QEs). Our goal is to bring awareness of the vulnerabilities of QEs to those who design, implement, and deploy QEs. We hope this knowledge will provide motivation to improve the accuracy of quality estimators. We focus on the two applications of QEs for which money is exchanged, and explore situations in which a system that uses an image or video QE can be “gamed”, such that one party takes advantage of another financially. We describe two specific “attacks” on QEs, and demonstrate using subjective tests that some existing QEs are vulnerable to these attacks.

## 1. INTRODUCTION

In analyzing systems, security researchers often use an adversarial model, positing a “black-hat” adversary having certain motivations and capabilities; a system’s strength can be expressed in terms of its ability (or failure) to resist the attacks such an adversary could feasibly mount.

In the image processing community, such an adversarial approach has been effectively used in watermarking and forensics applications. The typical cycle is that “proponents” design algorithms and “attackers” find ways to get around these algorithms, which leads the proponents to design better algorithms.

The first watermarking algorithms [1] were claimed to be “robust against common signal processing algorithms and geometric distortions when used on some standard images” [2]. Subsequently, attacks on these watermarks were created [2, 3], which then inspired the design of new watermarking algorithms that were more robust to these attacks [4].

The goal of image forensics is to detect when an image has been tampered or altered. Tamper-detection algorithms include [5, 6]. Next, Gloe et al. [7] illustrate ways educated image-counterfeiters can alter images so that they are not detected as tampered by these algorithms. Subsequently, new image forensics algorithms will be introduced that will be robust to these improved tampering strategies.

We believe a similar cycle can be effective to improve the design of objective image and video quality estimators (QEs). However, while the incentives for both adversaries (the proponents and the attackers) are clear for the watermarking and image forensics applications, the incentives for such an arms

race in QE design is less clear. In this paper, we use vulnerability assessment to explore the adversarial positions of QE. A brief overview of vulnerability assessment is presented in Section 2. Our goal is to bring awareness of the vulnerabilities of QEs to those who design, implement, and deploy QEs. We hope this knowledge will motivate improvements in both the accuracy of quality estimators and the robustness of systems deployed using QEs.

Next, in Section 3, we focus on two applications of QEs in which money may exchange hands: content acquisition, delivery and the associated legal agreements, and benchmarking for product sales and marketing. For these two applications, we describe incentives that may tempt any party in these negotiations to become “attackers” (i.e., to take unfair monetary advantage of another party). We elaborate on a variety of scenarios in which vulnerabilities in QEs may be exploited.

While one can envision many ways to cheat, we focus here on ways that exploit weaknesses in an objective quality estimator  $Q_{obj}$ . These weaknesses occur when the objective QE has output greater than a threshold  $T$  ( $Q_{obj} > T$ ) but the actual subjective quality is less than the same threshold ( $Q_{subj} < T$ ), or vice versa. While any objective QE likely has some instances in which this occurs, those QEs with *systematic* weaknesses are most vulnerable to exploitation.

Section 4 describes several specific attacks on a subset of existing QEs, whose effectiveness are verified in section 5 using subjective tests. We conclude with some recommendations for those designing, implementing, and deploying QEs.

## 2. VULNERABILITY ASSESSMENT

The security community has long used adversarial models as tools to assess and strengthen systems. Such analysis generally starts with a system or protocol having certain desirable security properties. The analysis proceeds by considering whether an attacker, given reasonable resources (for example, time and money) could violate those properties and, if so, how the system might be improved so that the attack is infeasible. The goal of such analysis is not generally to prove theoretical correctness or stringent bounds on the characteristics of an algorithm; instead, it is to assess how the system might perform in a real-world environment, where an attacker’s capabilities and motivations are typically bounded by practical limits. Thus, we make two fundamental assump-

tions throughout: everyone wants more money (either greater income or smaller outflow) and no one will pay more to exploit a weaknesses than they would gain as a result. It is useful to express this tension between the value of exploitation and the cost of protecting against attack in economic terms [8].

The techniques used in analyzing such systems are broadly referred to as “vulnerability assessment” or “vulnerability analysis” [9]. The means to perform such assessments vary, but generally involves finding a class of input which an attacker could use to cause the system to behave in a way unforeseen by the designers of the system and beneficial to the attacker. The system is then said to be *vulnerable* to an attack using those inputs. The best-known example of such an attack is the buffer overflow: a buffer of length  $n$  accepts input from the attacker, who supplies a value of length  $n + m$ ; the overflowing value is used to overwrite the program’s frame or stack pointer, allowing the attacker to take control of the program.

While this assessment is performed against a specific system (say, a particular web server implementation), the vulnerabilities exposed generally fall into broad classes, allowing general design and implementation guidelines to be deduced. The buffer overflow, for example, is a class of input validation failure; such flaws can be mitigated by compile- and run-time checks, improved programmer education, and limitations on the degree of control an attacker has over the input to a program. By the same token, while we consider flaws in specific objective QEs in this paper, we believe applying these adversarial approaches to the problem of quality estimation will aid in the development of stronger systems.

### 3. APPLICATIONS WITH MONETARY EXCHANGE

A variety of applications for a QE are described in [10], including algorithm optimization, writing product benchmarks, system provisioning, content acquisition and delivery, outage detection, and system troubleshooting. Of these, two are likely to involve an exchange of money between two or more parties: service-level agreements (SLAs) for content acquisition and content delivery, and benchmarking for product sales and marketing. We describe these here, along with descriptions of some scenarios for exploiting potential vulnerabilities in an objective QE.

#### 3.1. Service Level Agreements and other legal contracts

Service level agreements (SLAs) are legal contracts between entities for content acquisition, content delivery, or both. There may be up to three types of entities in an SLA: the service provider S, the network provider N, and the customer C. Contracts may be implicit or explicit. In general, money flows from C to N, from N to S, or directly from C to S. It is also possible to consider contracts between only two parties, for example, between S and C or between two different network providers, N1 and N2.

A QE for *content acquisition* and *content delivery* determines if either the incoming material or the outgoing material has sufficient quality. For acquisition and delivery of images, a QE must [10]

Alarm when  $Q_{subj}(v) < T$  for more than  $\alpha\%$  of images. (1)

Similarly, a video QE must

Alarm when  $Q_{subj}(v) < T$  more than  $N$  times in  $t$  seconds. (2)

A legal contract for content acquisition and delivery will specify an objective  $Q_{obj}()$ , and the parameters  $T, N, t$ , and  $\alpha$ .

With 3 types of entities, and possibly multiple networks, many options to obtain monetary advantage are possible for SLAs. A subset of scenarios for potential exploitation are:

**Scenario SLA1:** S provides N video to be sent to C. N saves bandwidth (and money) by reducing the bit-rate of the video while satisfying its contract with S that  $Q_{obj} > T$ . However, the video may not satisfy  $Q_{subj} < T$ .

**Scenario SLA2:** Video is sent from S to network provider N1, to network provider N2, to C. N1 is both a service provider and an access provider (i.e., a competitor to S and N2). N1 degrades the video transported to N2 such that  $Q_{subj} < T$ , while satisfying contractual obligations with  $Q_{obj} > T$ . If C buys the video directly from N1, N1 provides undegraded video at the same  $Q_{obj}$ . A customer then has incentive to switch from N2 or S to competitor N1.

**Scenario SLA3:** N charges S to deliver video to C using one QE,  $Q_{obj1}$ . S uses a second more accurate QE,  $Q_{obj2}$ , to obtain better subjective quality for the same  $Q_{obj1}$  (and hence cost).

**Scenario SLA4:** S and C collude to take advantage of N. S alters  $v$  to create  $Q_{obj}(v) \ll T$ , but with  $Q_{subj}(v) > T$ . C demands a refund from N, even though the quality was acceptable.

While these are only a few examples, they illustrate several important aspects of effective contracts using QE. First, monetary gain can be obtained without being malicious. Some situations correspond to a loss of *potential* revenue, as opposed to the loss of *actual* revenue. Second, N must ensure its SLAs measure  $Q_{obj}$  using the same function for both acquisition and delivery, otherwise it will be vulnerable to exploits similar to the final one above. While this may seem obvious, it is important to emphasize, since early systems have been deployed without this realization.

#### 3.2. Product marketing and sales

The goal of a product marketer, M, is to convince the purchaser, P, of the superiority of one product over another. Statements about subjective quality for product marketing may take the form [10]

$Q_{subj}(p_1(v)) > T$  for  $\alpha\%$  of  $v \in \mathcal{V}$ , (3)

or

$$Q_{subj}(p_1(v)) > Q_{subj}(p_2(v)) + \delta \text{ for } \beta\% \text{ of } v \in \mathcal{V}, \quad (4)$$

where  $p_1$  and  $p_2$  are the two products being compared, and they produce processed content  $p_1(v)$  and  $p_2(v)$ , respectively.  $T$  is a quality threshold, and the marketer, M, will carefully select the set of sources  $\mathcal{V}$  for which  $\alpha$ ,  $\beta$ , and  $\delta$  are as large as possible. When a QE is used for product marketing,  $Q_{obj}()$  is substituted for  $Q_{subj}()$ .

We describe three scenarios for this application in which vulnerabilities of a QE may be exploited for marketing and sales to consumers (not professionals). The first and third rely on P trusting the quality estimator  $Q_{obj}()$ , and the first two rely on M showing at least one actual visual comparison to P prior to purchase.

**Scenario M1:** M scares P into buying a more expensive product. The purchaser P has done research about his required specifications, and knows he wants  $Q_{subj} > T$ . M has two products,  $p_1$  and  $p_2$  for sale, each with  $Q_{subj} > T$  for almost all original videos. However, the more expensive  $p_1$  actually has  $Q_{subj}(p_1(v)) \gg T$  for almost all  $v$ . The specifications for each product state that  $Q_{obj}(p_i(v)) > T$  for most  $v$ , so either product should satisfy the purchaser P.

However, M knows a systematic weakness in  $Q_{obj}$ . M visually demonstrates the products to P using a specially selected  $v^*$ , an original video for which  $Q_{subj}(p_1(v^*)) > T$ ,  $Q_{subj}(p_2(v^*)) \ll T$ , and  $Q_{obj}(p_2(v^*)) \approx T$ . M shows P both products with  $v^*$ . P can see that  $p_1$  looks better than  $p_2$  and that  $p_2$  does not have the desired quality for *this specific video*. P spends more money to purchase  $p_1$ , even though  $p_2$  satisfies his requirements.

**Scenario M2:** M convinces P to buy an inferior product from M instead of a better product from M's competitor. Let  $p_1$  be the product from M and  $p_2$  be the product from M's competitor. The purchaser P knows she wants  $Q_{subj} > T$ . The benchmarks show that  $Q_{obj}(p_1(v)) \leq Q_{obj}(p_2(v))$  for a substantial fraction of  $v \in \mathcal{V}$ , and that  $Q_{obj}(p_1(v)) < T$  for most  $v$ . Thus,  $p_1$  does not have sufficient quality for P's needs. M demonstrates  $p_1$  using several  $v^*$  (which are different than that in the previous scenario), all with  $Q_{subj}(p_1(v^*)) > T$ . (These may also have  $Q_{obj}(p_1(v^*)) < T$ .) P is (incorrectly) convinced that  $p_1$  is adequate and buys it from M.

**Scenario M3:** M lies to P about the superiority of M's product relative to a competitor. Again, let  $p_1$  be the product from M and  $p_2$  be the product from M's competitor. M knows that for almost all  $v$ ,  $Q_{subj}(p_1(v)) < Q_{subj}(p_2(v))$ . However, M carefully chooses a set  $\mathcal{V}$  of videos for which  $Q_{obj}(p_1(v)) > Q_{obj}(p_2(v))$ , and reports specifications for this set of videos only. This can be achieved easily if M has a second more accurate QE,  $Q_{obj2}$ , whose use would allow M to screen potential  $v$ .

### 3.3. Discussion

In both product marketing and SLAs, there is often an asymmetry in power between attackers and defenders. Attacker M has more time to develop strategies than defender P, for example. Also, inside a network N, the available inputs to a QE are typically constrained due to cost and system constraints. Such QEs often rely on bitstream or even parametric information about the bitstream [11] (i.e. they may use only bit-rate or packet loss rate). On the other hand, S and C often have no such constraints and may use a pixel-based QE; thus they may take advantage of systematic and *unavoidable* weaknesses in a parametric QE.

## 4. "ATTACKS" ON QUALITY ESTIMATORS

In this section, we provide concrete examples of systematic weaknesses in some selected QEs, to illustrate how one might attack existing QEs. First, we describe an attack that systematically processes an arbitrary video to create  $Q_{obj}(p(v)) > T$  when  $Q_{subj}(p(v)) < T$ , for a given QE. This attack is appropriate for Scenario SLA2. Second, we present an attack that processes a video to create  $Q_{obj}(p(v)) < T$  when  $Q_{subj}(p(v)) > T$ , which is appropriate for Scenario SLA4. Third, we briefly discuss the adversarial approach first introduced in [12] that uses one more accurate QE to systematically identify weaknesses in a less accurate QE.

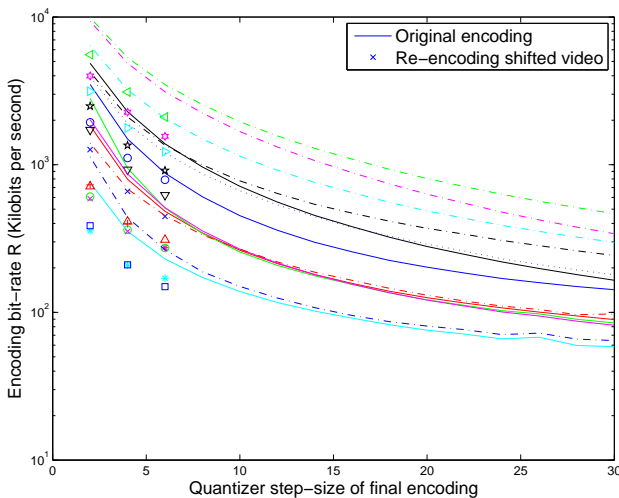
### 4.1. Re-encoding attack

The re-encoding attack is a systematic way to create a class of videos for which  $Q_{obj} > T$  when  $Q_{subj} < T$ , for vulnerable QEs. While it is most effective against parametric QE, it also demonstrates vulnerabilities in other pixel-based QEs.

In the re-encoding attack, the attacker first heavily compresses the video  $v$ , using a coarse quantizer. The coarsely-quantized video is decoded, and the pixels are shifted slightly to the right or left. The shifted video is then re-encoded with a fine quantizer. The final video has a relatively high bit-rate because the second encoding faithfully retains the poor quality caused by the first encoding. Thus, a parametric QE that uses both quantizer step-size and bit-rate (or either one individually) fails to characterize the poor subjective quality.

We applied this attack to nine 10-second CIF sequences: *Foreman*, *Silent*, *News*, *Akiyo*, *Coastguard*, *Container*, *Mother and Daughter*, *Table Tennis*, and *Hall monitor* and four 5-second CIF sequences: *Mobile and Calendar*, *Football*, *Bus*, and *Bicycles*. Each video is first encoded using MPEG-2 with a fixed quantizer step-size of  $QP_{attack} = 28$ . Each video is then decoded and shifted by two pixels to the right or left. Video whose original has a black bar on the left are shifted left; all others are shifted right. Finally, we re-encode with MPEG-2 using a fixed  $QP=2,4$ , or 6.

Figure 1 shows the bit-rate,  $R$ , as a function of quantizer step-size,  $QP$ . The lines correspond to a single encoding at the indicated  $QP$  and the symbols correspond to the re-encoded



**Fig. 1.** Bit-rate as a function of quantizer step-size. Lines are for a single encoding. Symbols correspond to re-encoded, shifted videos.

video. Individually, each attacked video has lower  $R$  for the same QP than without the attack. However, because of the large variation between  $R$  and QP in the thirteen original encodings, the pair  $(R, QP)$  of all except two of the “attacked” videos are within the range of those without the attack. Note that the shift is necessary for the re-encoding attack to be effective. Re-encoding using a fine quantizer without first shifting the image produces substantially lower bit-rates than when the shift is present.

#### 4.2. Pillarboxing attack

The pillarboxing (or equivalently letterboxing) attack is a systematic way to create a class of videos for which  $Q_{obj} < T$  when  $Q_{subj} > T$ , for vulnerable QEs. The most vulnerable QEs to this attack are those that measure blocking by averaging on-grid edge strength. Letterboxing is the black pixels on top and bottom of an image; pillarboxing is black pixels on the left and/or right of an image.

In the pillarboxing attack, the attacker adds pillarboxing (or letterboxing) to the original image prior to encoding with a fine quantizer. To be most effective, the attack forces the number,  $L$ , of black pixels on the image edge to be exactly  $L = 8$ . In our implementation of this attack, we first add pillarboxing and compress using MPEG-2 with a fixed QP=4. The effectiveness of this attack is evaluated below.

#### 4.3. Adversarial attacks

Adversarial attacks enable an attacker to use one (better) QE to systematically identify weaknesses in another QE. The notion of using one QE as an adversary to another has already been proposed by Wang and Simoncelli [12]. Their Maximum Differentiation (MAD) Competition synthesizes image stimuli which maximize (and minimize) the response of one

QE while holding the response of the other QE fixed. The optimization strategy they present relies on the fact that the two QEs they consider (MSE and SSIM [13]) are differentiable. However, their method demonstrates the feasibility of using such an optimization for scenarios SLA1 and SLA3. In addition, such an approach could be used to assist in screening for selected  $v^*$  in Scenarios M1-M3.

### 5. SUBJECTIVE EVALUATION OF ATTACKS

We performed a subjective test of image quality using frame 30 of four sequences: *Foreman*, *Silent*, *Hall Monitor*, and *Coastguard*. A set of 15 images was generated for each sequence using ten images of a single encoding, with QP=2, 4, 6, 10, 14, 18, 22, 26, 28, 30; two images of the re-encoding attack with first QP<sub>attack</sub> = 28 and second QP=2, 4; and three images for the pillarboxing attack with QP=4 and  $L = 0, 7, 8$ . When  $L = 0$ , any black pixels on the image edges are replaced with the closest non-black pixels.

Figure 2 show a sampling of these images, specifically frame 30 (an I-frame) of *Silent*, without and with attacks. Fig. 2(a) and (c) both have QP=4, but the latter has been attacked with the pillarboxing attack with  $L = 8$ . Fig. 2(b) is encoded with QP=28, and Fig. 2(d) has been attacked with the re-encoding attack applying a shift of 2 pixels and a second encoding with QP=4. It is evident that the visual quality of Fig. 2(d) is much worse than that of Fig. 2(a), although both have identical quantizer step-size in the final encoding. In addition, with the exception of the black bar on the right, the visual quality of Fig. 2(c) is quite similar to that of Fig. 2(a).

We conducted a subjective test using paired comparison among the 15 images in each set for the four sequences. Twenty viewers were shown a series of pairs of images obtained by processing the same original image, and asked to select the image with “better quality” using a viewing distance of 24 inches. The pairs to test within each set are chosen adaptively for each viewer using a binary-sort strategy [14], which ensures that more similar pairs are compared more often. Finally, each viewer compared those pairs that were separated by one ranking in the resulting sorted list, thus obtaining additional comparisons of similar images.

We use the Bradley-Terry model [15] to analyze the subjective test results. In the Bradley-Terry model, the relative quality between two images is  $\log \pi_i - \log \pi_j$ , where the probability that image  $i$  is preferred to image  $j$  is modeled by  $\pi_i / (\pi_j + \pi_i)$ , with  $\sum_i \pi_i = 1$ . The maximum likelihood estimate of  $\pi_i$  can be found iteratively [15]. This model only defines the relative quality between tested images, not absolute quality. We define a Just Noticeable Difference (JND) unit to be a change in subjective quality that can be identified by 75% of viewers. Then we express the *Relative Subjective Quality*,  $\Delta Q_s$  in JND units relative to the image of the same content compressed with QP=2.

Figure 3 shows Marziliano’s Blur [16] as a function of



(a) QP=4 one encoding, Blur=6.2, GBIM=1.4,  $\Delta Q_s = 0.020$  JNDs



(b) QP=28 one encoding, Blur=18.9, GBIM=6.6,  $\Delta Q_s = 12.46$  JNDs



(c) Pillarbox attack,  $L = 8$ , QP=4, GBIM=4.2,  $\Delta Q_s = 0.22$  JNDs



(d) Re-encoding attack: QP=28 shifted and re-encoded with QP=4. Blur=8.5,  $\Delta Q_s = 12.50$  JNDs.

**Fig. 2.** *Silent*, frame 30 with and without attacks. (a) and (c) have similar subjective quality; however, the black pixels on the edge of (c) result in a substantially worse rating from GBIM. (b) and (d) have similar subjective quality; although, (d) is rated by a vulnerable QE similarly to (a).

relative subjective quality  $\Delta Q_s$  for the re-encoding attack, and Figure 4 shows the Generalized Block Impairment Metric (GBIM) [17] as a function of relative subjective quality  $\Delta Q_s$  for the pillarboxing attack. Both QEs increase as  $\Delta Q_s$  increases from zero. The minimum GBIM is defined as one, while the Blur is non-negative. In each figure, the lines correspond to the images without attack, while the isolated symbols correspond to the images produced from an attack.

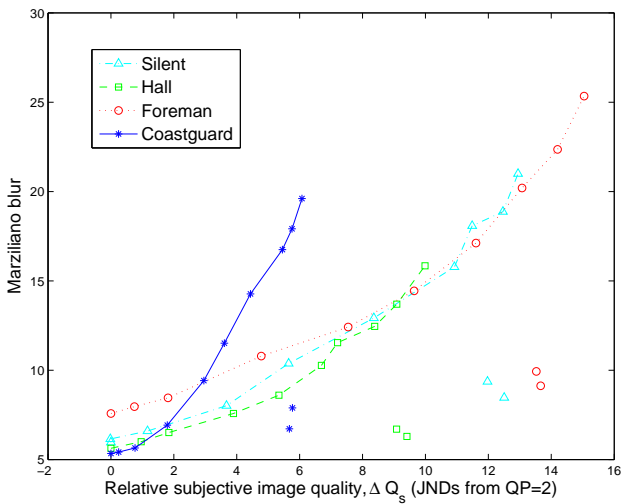
When there is no attack, Blur and GBIM increase with increasing  $\Delta Q_s$  for each sequence, as expected. The results for *Silent*, *Foreman*, and *Hall Monitor* are closely aligned, while the objective QE values for *Coastguard* increase more quickly than for the other three sequences. This highlights another drawback of these QE; they are not always accurate at quantifying quality *between* different source content.

However, Figures 3 and 4 illustrate that both attacks are highly successful. For all sequences, the re-encoding attack

dramatically degrades the subjective quality without substantially altering the Blur value. Similarly, the pillarboxing attack dramatically alters the output of GBIM without significantly affecting the subjective quality. Further, pillarboxing can be easily hidden during display, either coincidentally (with overscan in a TV) or intentionally (by an attacker who also controls the software for display, e.g. Scenario SLA4), further lessening the visual impact of such an attack.

## 6. CONCLUDING THOUGHTS

In this paper we applied vulnerability assessment to objective image and video quality estimators. We described incentives for scenarios in which attackers might exploit QEs. Both the re-encoding and the pillarboxing attacks we presented are highly successful on the QEs studied. However, these attacks are primarily a proof of concept; for example, not all blocking QEs will be vulnerable to the pillarboxing attack. These at-



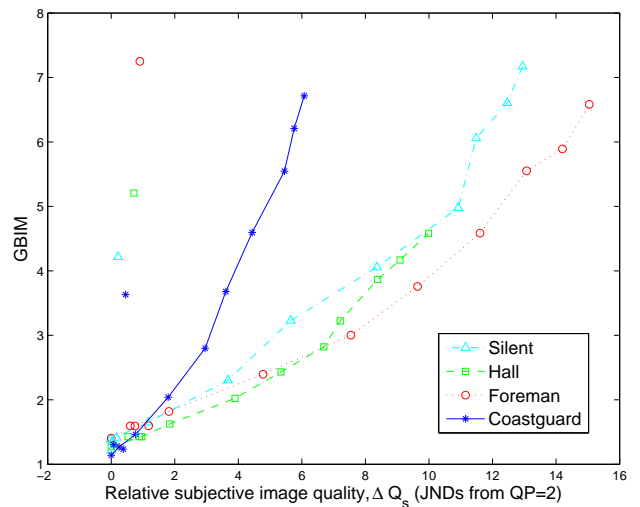
**Fig. 3.** Blur as a function of subjective quality. Symbols-only correspond to the re-encoding attack with both QP=2 and 4.

tacks do, however, illustrate that systematic weaknesses exist in several of today's QEs. In particular, bitstream or parametric QEs (that use only bit-rate, packet loss rate, and other "parameters") will almost always be at a disadvantage relative to QEs that rely on additional pixel information.

The attacks described here illustrate several important general principles about the design and deployment of QEs. In particular, an adversarial approach is useful to supplement the standard subjective verification of objective QEs. Exploitable vulnerabilities should be quantified, if not eliminated. For example, it is easy to envision simple modifications to increase the robustness of the vulnerable QEs presented here. When a QE is selected for deployment, it should be evaluated based not only on its cost, computational efficiency, accuracy and applicable content types, but also on the threat environment it will encounter in its application.

## 7. REFERENCES

- [1] I.J. Cox et al., "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Proc.*, pp. 1673–1687, Dec 1997.
- [2] F. A. P. Petitcolas et al., "Attacks on copyright marking systems," in *Information Hiding: Second International Workshop*, 1998, vol. 1525, pp. 218–238.
- [3] S. Voloshynovskiy et al., "Attacks on digital watermarks: Classification, estimation-based attacks and benchmarks," *IEEE Communications Magazine*, August 2001.
- [4] Xiangui Kang et al., "A DWT-DFT composite watermarking scheme robust to both affine transform and JPEG compression," *IEEE Trans. CSVT*, pp. 776–786, Aug. 2003.
- [5] A. C. Popescu and H. Farid, "Exposing digital forgeries by detecting traces of resampling," *IEEE Trans. Sig. Proc.*, vol. 53, pp. 758–767, February 2005.
- [6] J. Lukas and J. Fridrich, "Estimation of primary quantization



**Fig. 4.** GBIM as a function of subjective quality. Symbols-only correspond to pillarboxing attack with  $L = 8$ .

matrix in double compressed JPEG images," in *Proc. of the Digital Forensic Research Wkshp.*, 2003.

- [7] T. Gloe et al., "Can we trust digital image forensics?," in *ACM Int. Multimedia Conf.*, Sept. 2008, pp. 78–86.
- [8] R. Anderson, "Why information security is hard-An economic perspective," in *ACSAC '01: Proc. of the 17th Annual IEEE Comp. Security Applications Conf.*, 2001, p. 358.
- [9] I. V. Krsul, *Software vulnerability analysis*, Ph.D. thesis, Purdue University, West Lafayette, IN, USA, 1998.
- [10] S. S. Hemami and A. R. Reibman, "No-reference image and video quality estimation: Applications and human-motivated design," *Signal Processing: Image Communication*, to appear.
- [11] S. Winkler and P. Mohandas, "The evolution of video quality measurement: From PSNR to hybrid metrics," *IEEE Trans. Broadcasting*, vol. 54, no. 3, pp. 660–668, Sept. 2008.
- [12] Z. Wang and E. Simoncelli, "Maximum differentiation (MAD) competition: A methodology for comparing computational models of perceptual quantities," *J. of Vision*, pp. 1–13, 2008.
- [13] Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Proc.*, vol. 13, no. 4, pp. 600–612, April 2004.
- [14] D. A. Silverstein and J. E. Farrell, "An efficient method for paired comparison," *Journal Elec. Im.*, vol. 10, no. 2, pp. 394–398, April 2001.
- [15] R. A. Bradley, "Paired comparisons: Some basic procedures and examples," in *Handbook of Statistics, Vol. 4*, P. R. Krishnaiah and P. K. Sen, Eds., pp. 299–326. Elsevier Science Publishers, 1984.
- [16] P. Marziliano et al., "Perceptual blur and ringing metrics: Application to JPEG2000," *Sig. Proc.: Image Comm.*, pp. 163–172, February 2004.
- [17] H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," *IEEE Sig. Proc. Letters*, vol. 4, no. 11, pp. 317–320, November 1997.