

NEW QUALITY METRICS FOR MULTIMEDIA COMPRESSION USING FAULTY HARDWARE

In Suk Chong, Hye-Yeon Cheong and Antonio Ortega

Signal and Image Processing Institute
Department of Electrical Engineering-Systems,
University of Southern California

ABSTRACT

In this paper, we study the problem of evaluating the impact of hardware faults on decoded image/video quality. Our goal is to define quality evaluation tools for fault acceptability of hardware; acceptable faults are those that lead to a degradation in decoded quality that is deemed “acceptable”, according to perceptual criteria. For faults arising in both the discrete cosine transform (DCT) and Motion Estimation (ME) hardware, we focus on the distortion between faulty and fault-free decoded outputs. Our results show that, in the DCT case (applicable to JPEG or MPEG coders), suitable metrics can be found by combining tools developed for image quality ranking with a notion of error rate and significance. Instead, in the ME case, we show that the effect of faults can be well quantified by using standard quality metrics. The difference between the two cases can be explained by the different characteristics of errors in each case (error effects remain local in the DCT case, while they have a global impact in the ME case).

1. INTRODUCTION

Multimedia compression hardware systems are designed under the assumption that only fault free systems will be used. However, our recent work on error tolerant image and video compression has shown that in some cases faulty hardware can be used, as long as only “acceptable errors” are introduced [7, 8]. The motivation for this work is an increasing awareness of the importance of Error Tolerance (ET) techniques; a relaxation of the requirement of 100% correctness for devices and interconnects. ET is motivated as a tool to dramatically reduce costs for manufacturing, verification, and testing [5]. In our work to date we have studied the impact of hard (deterministic) errors, likely to be originated by real defects on chips. However, other sources of hardware errors are worth considering, such as soft errors due to Deep Submicron (DSM) noise and voltage scaling which cause probabilistic and input-dependent errors in multimedia systems [11, 17]. For any of these scenarios, determining which errors are acceptable should be done by evaluating the resulting image/video quality and determining whether the degradation in perceptual quality is “acceptable”.

Mean Squared Error (MSE) is a widely used objective quality metric for multimedia compression. While its limitations are

well-known [10], it is frequently used to provide a rough comparison between different coding techniques or to drive efficient rate control algorithms. In this paper we show that for certain systems suffering from computation errors, MSE is *not* a suitable performance measure. In particular we will show that alternative metrics are needed for errors in the DCT, while MSE might be appropriate when errors occur in motion estimation.



Fig. 1. Left: Higher MSE but Acceptable, Right: Lower MSE but Non-Acceptable

To get some intuition about why this may be true in some cases, consider Figure 1. The image on the left is obtained with a fault-free JPEG encoder and has higher MSE ($PSNR = 37.9db$). This image, however, is perceptually better than the one on the right, which has lower MSE ($PSNR = 38.9db$), and was obtained by simulating the presence of faults in the JPEG encoder. In this second image we can observe a regular error pattern in various areas. In a typical compression scenario, quantization noise tends to affect all components (e.g., *all blocks, all frequencies within a block*). Instead, we are now considering hardware errors that can lead to unevenly distributed artifacts, because they only manifest themselves for certain inputs. We have observed that this is indeed true for hardware errors in the DCT in the context of JPEG coding: only *certain blocks* and *certain frequency components* within those blocks can be seen to be affected, as seen in Figure 1. Clearly, MSE would then not be a suitable metric, since a large error in a only a few blocks (or frequencies), while clearly visible by the end user, may still lead to a low overall MSE.

More specifically, here we consider two types of commonly used deterministic faults¹: i) Single Stuck At 0 (SSA0), ii) Single

¹We focus our discussion on these faults but the quality assessment tools we propose could also be applied in other scenarios, e.g., when dis-

This paper is based upon work supported in part by the National Science Foundation under Grant No. 0428940. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Stuck At 1 (SSA1) [12]. In SSA0 faults a specific bus line (p_{th}) is modeled to be always set to 0, so that whenever that line was meant to have a value of 1, an error with magnitude 2^p is added to or subtracted from the original value. The same holds for SSA1 case except that error occurs when the bus line is set to 0. Clearly, if one of these faults affects the computation of a frequency term in the DCT, two similar blocks may suffer very different errors (i.e., no error vs. 2^p error) if they differ in the specific faulty bus line.

Thus, in general, hardware faults in image/video compression systems can generate errors that are unevenly distributed over i) frequency components, ii) blocks or iii) frames (in the case of video). Therefore, in our work we consider quality metrics in terms of *error significance*, which measures degradation in individual components of the signal (e.g., a block), and *error rate*, measuring the probability that unacceptable degradation occurs in an individual component.

Note that here we are considering the additional degradation introduced by hardware faults on decoded outputs. Thus we will typically evaluate the error between pairs of decoded images or frames, namely those produced by faulty and fault-free systems. This is in contrast to standard quality evaluation in the context of compression, where often a decoded image is compared with the original uncompressed one. Our problem is in effect an image ranking problem (where two different decoded versions of an image are compared) [16]. Unlike typical ranking scenarios, however, in our case a single encoder (with fixed coding parameters) is considered and the image generated by the faulty system cannot be better in quality than that generated by the fault free system.

In this paper we focus on two very common components of multimedia compression systems, DCT and motion estimation (ME). The DCT is used in video coders, e.g., MPEG [14], and image coders, e.g., JPEG [13], and similar linear block transforms are used in emerging compression systems, such as ITU-T H.264/MPEG4 AVC [2]. Note that in all these systems the transform is followed by quantization. Thus, while we consider faults in the transform operation, our analysis considers the impact of faults *after* quantization. ME is a key component of video compression systems such as ITU-T H.264/MPEG4 AVC [2]; it is critical to achieve efficient overall rate-distortion (RD) performance.

We will show that faults in the DCT block cause numerical errors that are unevenly distributed over both blocks and frequency components (and frames in the video case). Instead, faults in the ME process have a more indirect impact in quality; they result in higher energy in the motion compensated residual signal and thus somewhat worse overall RD performance. Unlike the DCT case, the quality degradation is spread out over multiple blocks and frames.

The paper is organized as follows. In Section 2 we consider faulty DCT in a JPEG system and propose to use block based error rate and error significance metrics. Our proposed error significance metric is based on Watson’s work on DCT basis [15, 3], and takes into account the fact we are comparing two decoded images. In Section 3 we discuss the impact of faulty DCT in a MPEG system, and in particular show how some of the errors introduced in a frame are in fact attenuated for SSA0 case and similar or accumulated for SSA1 case in successive frames. In Section 4, we present and analyze the behavior of ME faults, which suggests that measures such as MSE or PSNR may still be sufficient for evaluation of the quality degradation due to faults.

2. FAULTY DCT IN A JPEG SYSTEM

We start by introducing some notation. We consider three images, the original image, \mathbf{I} , the decoded image obtained from a fault-free encoder, $\mathbf{Q}(\mathbf{I})$, and that produced by a faulty encoder, $\mathbf{Q}(\mathbf{I}')$. Moreover much of our discussion will consider blocks within these three images, denoted $\bar{\mathbf{Y}}$, $\mathbf{Q}(\bar{\mathbf{Y}})$ and $\mathbf{Q}(\bar{\mathbf{Y}}')$, respectively, as well as individual frequency coefficients within one such block, denoted $Y(u, v)$, $Q(Y(u, v))$, $Q(Y'(u, v))$, respectively. Whenever possible we drop the (u, v) index to simplify the notation. We will consider distortion metrics at the image, block or frequency component level, which we will denote D_I , D_B , and D_C , respectively. This is summarized in Table 1.

We consider three possible distortions (introduced here at the image level, but also used at block and component level with appropriate subscript): $D_I^1 = D(\mathbf{I}, \mathbf{Q}(\mathbf{I}))$, $D_I^2 = D(\mathbf{I}, \mathbf{Q}(\mathbf{I}'))$, $D_I^3 = D(\mathbf{Q}(\mathbf{I}), \mathbf{Q}(\mathbf{I}'))$, where D is a “basic” distortion metric applied to the two images (or blocks or frequency components). D will be discussed in Section 2.1. D_I^1 and D_I^2 quantify the degradation introduced by quantization (with fault-free and faulty systems, respectively) with respect to the original image. D_I^3 evaluates the difference between the two decoded images.

Object/Level	image	block	frequency comp.
original	\mathbf{I}	$\bar{\mathbf{Y}}$	Y or $Y(u, v)$
fault free decoded	$\mathbf{Q}(\mathbf{I})$	$\mathbf{Q}(\bar{\mathbf{Y}})$	$Q(Y)$ or $Q(Y(u, v))$
faulty decoded	$\mathbf{Q}(\mathbf{I}')$	$\mathbf{Q}(\bar{\mathbf{Y}}')$	$Q(Y')$ or $Q(Y'(u, v))$
Distance	D_I	D_B	D_C or $D_C(u, v)$

Table 1. Summary of Notation

In our previous work [7] we studied the test design for fault acceptability of DCT hardware based on the assumption that tools are available to determine what is perceptually acceptable. Here we analyze some of the characteristics such a metric should have and propose a concrete objective metric, specifically for hardware testing in a faulty scenario.

We wish to compare $\mathbf{Q}(\mathbf{I}')$ and $\mathbf{Q}(\mathbf{I})$ by ranking the quality of these two images, i.e., by computing D_I^1 and D_I^2 and comparing these two distortions. As will be discussed later, $D_I^2 \geq D_I^1$. Thus if $D_I^2 = D_I^1$, the fault degradation will certainly be acceptable, while in cases where $D_I^2 > D_I^1$ degradation may be acceptable if the difference is small.

2.1. Basic Distortion Metric Selection

Since our ultimate goal is to introduce quality acceptability criteria in automated hardware testing, we consider only objective metrics that take into account the human visual system (HVS) [9]. The faults we consider occur within the DCT block and thus produce errors in individual frequency components. For this reason we take Watson’s [18, 3] techniques as a starting point to define D . In [3], visibility thresholds for each DCT frequency component ($Th_w(u, v)$) are proposed. Using these thresholds, a perceptual objective metric is introduced in [3, 18], which evaluates distance between two images: reference and test image (typically \mathbf{I} and $\mathbf{Q}(\mathbf{I})$ in our notation). The DCT is performed on non-overlapping 8×8 blocks in both images and the absolute differences ($L1$ distance) between corresponding coefficients in the two images are computed. Each resulting difference or error component is

ortion is added by soft errors

weighted by the corresponding perceptual threshold ($Th(u, v)$). Then pooling is performed first over the frequency components in one block and then over all blocks in the image pair, so as to obtain a single distortion metric value. In both cases Minkowski pooling is used. For frequency pooling this can be written as

$$D_k = \left(\sum_{u,v} D_k(u, v)^b \right)^{\frac{1}{b}}$$

where D_k is the metric for block k , $D_k(u, v)$ is the weighted absolute difference for frequency (u, v) , and b is a parameter that controls how much emphasis is given to the largest values. Note that in some cases the $Th(u, v)$ can be modified so as to incorporate luminance compensation (LC) and contrast masking (CM).

For our purposes we modify Watson's approach as follows. First, since our goal is to incorporate these thresholds into a testing strategy *we do not use either LC or CM modified thresholds*, as these would be image dependent. Second, we have noted that faults could lead to large errors in some blocks and no errors in others. For this reason *we do not perform spatial pooling* of the metric and will measure block by block error (see Section 2.3 for further details). Finally, because faults in the DCT operation can lead to errors being added to only certain frequencies in a given block we use Minkowsky pooling setting $b = \infty$, i.e., the *metric for a block is the largest weighted absolute error in the block*.

2.2. Discrete Cosine Transform (DCT) and Quantization

We now provide a detailed analysis of faulty DCT operation followed by quantization. The input to the system (see Fig. 2) is a vector \bar{X} , which we assume drawn from a vector distribution that can be statistically characterized, e.g., by its covariance structure. We can define the set of possible faults, or fault space, \mathbf{F} , by analyzing the architecture of the system. Assume there is a single fault $f_i \in \mathbf{F}$ in the transform and denote its faulty (vector) output \bar{Y}' . Denote \bar{Y} the output of the fault-free system when the input is \bar{X} .

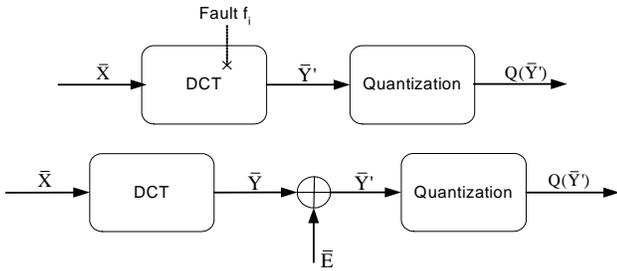


Fig. 2. DCT and Quantization

To analyze the effect of fault f_i , we first simplify the problem by viewing its effect as an error term \bar{E} added to the fault-free output \bar{Y} . Clearly, \bar{E} is a deterministic function of f_i , \bar{X} , and the structure of the DCT. Since we consider invertible transforms, there is a 1-to-1 mapping between \bar{X} and \bar{Y} , and thus, \bar{E} is not independent of \bar{Y} .

Note that scalar quantization is normally used, so that each component of \bar{Y} (or \bar{Y}') is independently quantized. Denote $Y(u, v)$, $E(u, v)$, $Y'(u, v)$ the (u, v) -th component of the vectors \bar{Y} , \bar{E} , and \bar{Y}' , respectively, with $u, v = 1 \dots N$, with N is the vector

dimension. When considering individual components it is reasonable to assume that $E(u, v)$ will be independent of $Y(u, v)$, even though \bar{Y} and \bar{E} are dependent. To see why, note that for a specific value of $Y(u, v)$ there are many possible values of $E(u, v)$, which depend on the $Y(k, l)$, $(u, v) \neq (k, l)$. In what follows we make the assumption that $E(u, v)$ is a random additive error term independent of $Y(u, v)$. We have verified that this is a reasonable assumption for typical systems. For convenience, in what follows we focus on one component and drop the frequency index (u, v) .

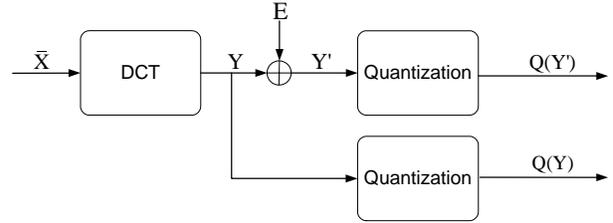


Fig. 3. Quantization Analysis

Our focus now turns to analyzing how the error (E) leads to error *after quantization*. Let E and Y be discrete and continuous random variables, respectively, with known pmf/pdf. Let the quantization step size be Δ , and define $D_C^2 = |Q(Y') - Y|$, $D_C^1 = |Q(Y) - Y|$ (we use $L1$ as discussed in the previous section).

We make the following observations. First, because of the structure of DCT computation hardware, individual faults do not affect all frequency components (see [7] for details). Second, in some cases, even though $Y' \neq Y$, we will have that both Y and Y' fall in the same quantization bin, so that $Q(Y') = Q(Y)$. In these cases the error is masked by the quantization operation, and thus errors are not uniformly distributed over the image blocks. Another consequence of this observation is that, as quantization becomes coarser, fewer blocks are likely to be different in $Q(\mathbf{I})$ and $Q(\mathbf{I}')$. However, the magnitude of those differences that are present after quantization will increase as the quantization becomes coarser. To understand why, recall that the same quantization step-size is applied to both images, thus:

$$Q(Y') = Q(Y) \pm K\Delta,$$

for an integer K . Therefore, as Δ grows it is more likely that K will be zero, but when K is not zero the corresponding error $K\Delta$ will be larger.

Finally we observe that $D_C^2 \geq D_C^1$, i.e., the image produced by the faulty system can only have larger distortion than that produced by the fault-free one. First recall the definitions: $D_C^1 = |Q(Y) - Y|$ and $D_C^2 = |Q(Y) - Y \pm K\Delta|$. Then noting that $|Q(Y) - Y| \in \{0, \frac{\Delta}{2}\}$ it is clear that $D_C^2 \geq D_C^1$, since for $K \neq 0$ we have that $|Q(Y) - Y \pm K\Delta| \geq \Delta/2 \geq D_C^1$ (for $K = 0$, $D_C^1 = D_C^2$). This obviously also holds at the block and image levels ($D_B^2 \geq D_B^1$ and $D_I^2 \geq D_I^1$).

2.3. Error rate and Error significance

As can be seen in Section 2.2 errors are unevenly distributed over blocks. This justifies our initial assertion that a global averaging metric (such as MSE) would not be suitable for our problem. Thus we propose to quantify error significance on a block-by-block basis and then quantify error rate. The error significance will be a

blockwise objective metric (to be introduced in the next section) that quantifies the difference between the two decoded images. The error rate will be the probability that the error significance exceeds a certain threshold in any one block. Acceptability can then be defined (for a particular application) in terms of the percentage of blocks (in an image or a video sequence) for which visible differences between the images can be tolerated.

2.4. Block based Distortion and Acceptability Issue

So far we have discussed distortion metrics that are appropriate for our problem and can be applied to pairs of images. However, unlike traditional image quality assessment problems we now have three images to consider.

One possible blockwise metric is $D_B^{new} = D_B^2 - D_B^1 \geq 0$, since $D_B^2 \geq D_B^1$. We can think of D_B^{new} as *perceptual distance between $\mathbf{Q}(\bar{\mathbf{Y}})$ and $\mathbf{Q}(\bar{\mathbf{Y}}')$* when using the original image $\bar{\mathbf{Y}}$ as a common reference.

Another candidate metric is D_B^3 , which directly evaluates the distance between $\mathbf{Q}(\bar{\mathbf{Y}}')$ and $\mathbf{Q}(\bar{\mathbf{Y}})$. This alternative metric is problematic. When ranking we would expect that when $D_B^2 = D_B^1 + \delta$ (with $\delta > 0$ small) the two decoded image will have almost the same perceptual quality. However, it is possible for D_B^3 to be large relative to δ when the quantization parameter is large. To see why, consider an example where a single frequency $Y(u, v)$ in a block is affected by a *small* error $E(u, v)$ and $Y(u, v)$ is near the boundary of a quantization bin. Assume the error is sufficiently large so that $Y'(u, v)$ is now in the immediately neighboring quantization bin. Then $D_C^1(u, v) \simeq D_C^2(u, v) \simeq \Delta/2$ and the same will be true for $D_B^1(u, v)$ and $D_B^2(u, v)$ so that D_B^{new} will be small. However $D_C^3(u, v) = \Delta$, which can be large. Based on this, we choose D_B^{new} as our distortion metric between two decoded image blocks. Alternatively, we could use D_B^3 when Δ is small, but for simplicity we use D_B^{new} for all quantization settings.

We now define an acceptability threshold for D_B^{new} . In [16], it was shown that if Watson's distance between two images (in their case, original and decoded) is less than or equal to 1, there are no visible artifacts. This was based on setting the Minkowski frequency pooling parameter b to ∞ (as we do, albeit for different reasons). Equivalently, this condition states that two blocks cannot be visually differentiated as long as $D_C(u, v) \leq Th(u, v)$ for all u, v . We believe it is reasonable to use the same condition for our problem, because D_B^{new} is attempting to quantify perceptual distance between $\mathbf{Q}(\bar{\mathbf{Y}})$ and $\mathbf{Q}(\bar{\mathbf{Y}}')$. In our preliminary simulations, this assumption looks reasonable, but more thorough perceptual experiments are needed for a more complete validation of this assumption. In summary, we define acceptability for an individual block as follows: $\mathbf{Q}(\bar{\mathbf{Y}}')$ is acceptable if $D_C^{new}(u, v) \leq Th(u, v)$ for all u, v . In our previous project, using this approach we proposed a test for faulty DCT in a JPEG system, with a simple modification to reduce number of test vectors (see [7] for details).

3. FAULTY DCT IN A VIDEO ENCODER

Our discussion thus far has centered on DCT computation in the context of an image coder. We now consider the impact of DCT faults in a video encoder, e.g., an MPEG-2 encoder. I frames are individually coded and thus are already covered by our previous analysis, however, errors occurring in these frames may actually propagate over time. Moreover, since the DCT computation system will be used for both I and P frames, errors will manifest them-

selves in both types of frames. We are currently developing an analysis to capture the behavior of standard video systems in terms of temporal error propagation, which would also be applicable to P frames.

We performed some preliminary experiments using MPEG2 TM5 [1] (IPPP structure) with the Foreman and Stefan sequences. We simulated the faulty DCT in the architecture analyzed in [7], and introduced various kind of faults. Our experimental observations indicate that the temporal behavior of errors depends on the fault type and that errors can accumulate or be attenuated over time.

Consider first as an example a SSA0 fault affecting a relatively high significance bit line. This fault is more likely to manifest itself as an error in an I frame, where we can expect larger frequency values. In successive P frames, the fault will not affect blocks for which the residual energy after estimation is low. Erroneous blocks in the I frame will be stored in the frame memory and used as potential reference for the next P frame. However, because ME attempts to match correct blocks (in the new frame) it will often not pick as best matches those blocks affected by error, thus increasing the chances that these errors will be eliminated from the video sequence. Our preliminary experiments confirm that this behavior can be observed and errors generated at the I frame gradually disappear.

As alternative example, consider the SSA1 case with errors also occurring at high significance bits. Here, errors are less likely to appear in I frames, but they may lead to errors in encoding the P frames. Because these errors manifest themselves when the energy of the prediction residual is low, they tend to appear frequently. Roughly speaking, if the error is large enough that an intra block is sent to refresh that area, the next prediction will be better and the error will again appear. These preliminary results are based on limited observations and we are in the process of developing a more systematic approach to tackle DCT faults in the video context.

4. FAULT IMPACT ON MOTION ESTIMATION

A key factor in video compression efficiency is how well the temporal redundancy is exploited by motion compensated prediction. The ME process comprises a search strategy of the motion displacement offset (motion vector, MV) and a matching metric computation. A searching strategy aims at selecting a set of candidate MVs and then proceeds to compute the matching metric for the candidates and select the one that minimizes a relevant metric, such as a Lagrangian cost. $J = D + \lambda * R$, where the term D (distortion) corresponds to the prediction error, i.e. Sum of Absolute Differences (SAD) or Sum of Squared Differences (SSD). R corresponds to the actual or estimated bitrate required for encoding the current MV while λ is a Lagrangian multiplier. λ can be selected using well-known methods such as [4, 19]. After the encoder selects the MV that minimizes cost J, it encodes the difference block (prediction residual) between the original and motion compensated blocks. Each residual block is transformed, quantized, and entropy coded. For simplicity, in this paper we only consider the case where $\lambda=0$.

There are several types of hardware implementation architectures [8], which we will refer to as Matching Metric Computation (MMC) architectures, with different levels of parallelism that are used to compute D . Most MMC architectures can be represented as a binary tree graph, where each inner node represents

an adder, leaf nodes represent absolute difference or squared difference computations, and edges connecting two inner nodes represent a data bus. Our work is focused on the interconnect faults within a MMC architecture that affect the data transfer between processing elements (PEs), assuming that each PE is error-free, using the SSA fault model. More detailed description on MMC architecture and SSA fault model analysis can be found in [8].

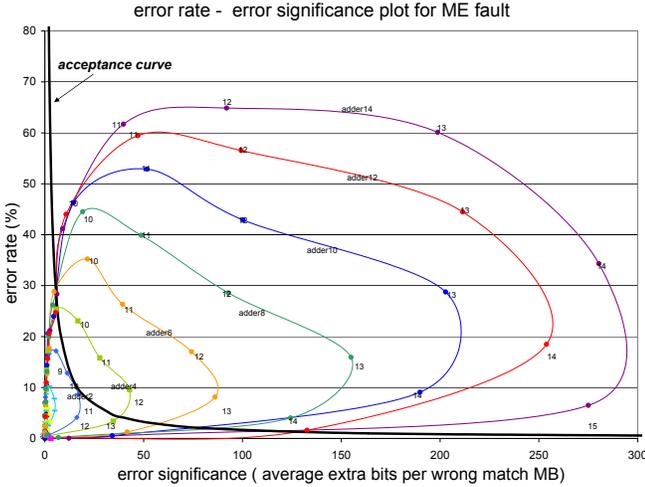


Fig. 4. Error Rate and Significance for ME fault

Faults in a MMC architecture would imply that it is likely that the MV selected for Motion Compensated Prediction (MCP) (MV_f) may not be equal to the best MV (MV_{min}) corresponding to ME without faults. We define a block being in *Error* if $MV_f \neq MV_{min}$ due to a fault. An error does not occur for all blocks but only occurs if certain conditions are met. For example, for the Lagrangian cost function mentioned above, an error occurs iff, a) if a fault is in the p -th data line, then the input to p must be 0 for MV_{min} and 1 for MV_f , and b) $0 < D(MV_f) - D(MV_{min}) \leq 2^p$. Therefore, our focus is on how often these errors occur (error rate $P_e = \text{prob}(MV_f \neq MV_{min})$) and how much additional quality degradation is introduced (error significance $S_e = D(MV_f) - D(MV_{min})$), representing the level of inaccuracy of MCP). Error rate and significance depend highly on the fault position with a certain variation due to the input sequence characteristics. Figure 4 demonstrates clearly how error rate and significance values change with faults at different positions. Points connected with the same line are faults in the same interconnect data bus with a different data bit line. Points shown in outer lines indicate faults in the data bus which has 32 more leaf nodes connecting towards that data bus than the adjacent inner line. Also note that SSA0 and SSA1 faults at the same positions produce identical results in both error rate and significance. Proof of this and further analysis on this concept of error rate and significance of ME fault are provided in [6].

While faults in the DCT block tend to have a rather direct impact on visual quality degradation and the type of artifacts introduced, ME faults have a more indirect impact on overall quality. When a ME error occurs the residual signal energy increases. In most cases there are constraints imposed on the bit rate of the system such that a certain local average bit rate needs to be maintained over time by using a rate control (RC) algorithm. Thus, if addi-

tional bits are required to encode a particular residual block/frame (because of its increased energy due to the error in ME), this extra rate will be compensated by a reduction in the number of bits used for other blocks/frames. This will lead to an increase in distortion for other blocks/frames. In this case, artifacts tend to appear in the form of standard quantization artifacts, rather than, as was the case for DCT, error-specific artifacts. Therefore, the problem of measuring the visual quality degradation due to fault introduction within the ME can be equally seen as the problem of measuring the impact on picture quality of a compression process. Thus, any visual quality metric which captures the video compression impairment reliably can be also applied to this case.

Distortion variation metrics can be particularly meaningful for our study of ME faults since we observed that introducing a SSA fault in ME always increases both the temporal and spatial quality variation on the video output. Note that the level of variation increase depends on the RC scheme. Typically RC process performs bit allocation by selecting the encoder's quantization step size (QP) for the residual block/image. Most modern implementations of RC, a constraint is employed on the increment/decrement of the quantizer, which results in distributing distortion throughout the picture. Therefore errors occurring with certain rate are subdued and smoothly spread out over the picture after the rate controlled quantization process. Therefore distinguishing errors into two measures of error rate and significance becomes no longer necessary. We have observed that spatial distortion variations tend to be imperceptible, on the other hand temporal variations can be significant. Spatial variation was measured by computing the variance of Qp values for each frame and by averaging them. To evaluate temporal variations within a video sequence we defined the measure temporal quality variation, TQV , as:

$$TQV = \frac{\sum_{i=0}^{N-2} |MSE_i^{fault} - MSE_{i+1}^{fault}|}{\sum_{i=0}^{N-2} |MSE_i^{no-fault} - MSE_{i+1}^{no-fault}|} \quad (1)$$

where MSE_i^{fault} and $MSE_i^{no-fault}$ are the frame MSE values of the decoded images with and without ME faults respectively, and N is the total number of frames considered.

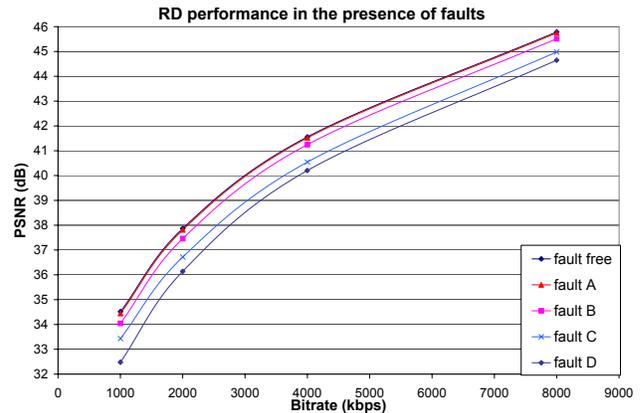


Fig. 5. Rate-Distortion Impact of Faults

Figure 5 depicts the RD performance for the Foreman sequence, at CIF resolution, in the presence of ME faults within an MPEG-2 encoder. Similarly, TQV for the same faults are presented in

Figure 6. Since error rate and significance measurements are no longer useful after RC Quantization for ME fault case, PSNR with additional measure of temporal quality variation would be able to represent well the quality degradation introduced by ME fault. However, we observed that temporal variation increase is relatively small compared to PSNR change and roughly proportional to the PSNR degradation, so that PSNR by itself may be sufficient to capture the behavior of the system. Figures 5 and 6 illustrate well this point, as large quality variations in Figure 6 occur when significant drops in PSNR are observed in Figure 5. In Figure 4, a PSNR based quality threshold T (e.g. $T = 0.1dB$) can be used to classify faults. This threshold essentially defines an *acceptance curve*, according to which faults below this curve are considered as acceptable, while faults above are considered as unacceptable.

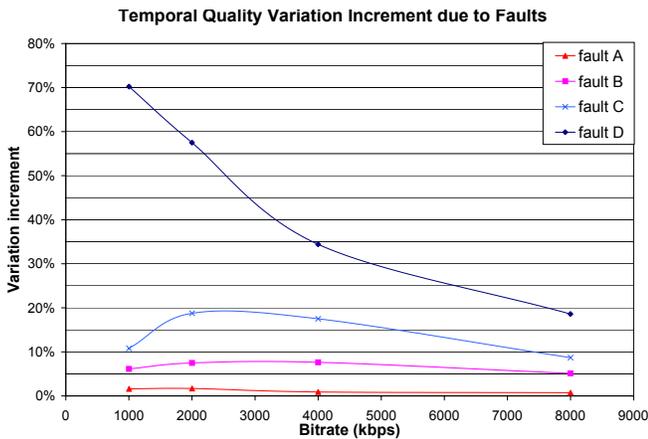


Fig. 6. Temporal Quality Variations due to Faults

5. CONCLUSION

We studied quality measurement for images/video encoded with faulty hardware. We propose perceptually-based objective metrics at the block level, and a notion of error rate and error significance for the DCT computation case. For the ME case we showed that standard quality metrics could be used.

6. REFERENCES

- [1] MPEG software simulation group, MPEG2 video codec version 1.2.
- [2] Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264/ISO/IEC 14 496-10 AVC in Joint Video Team(JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-G050, 2003.
- [3] A. J. J. Ahumada and A. Watson. An improved detection model for DCT coefficient quantization. In *Human Vision, Visual Processing, and Digital Display IV*. Allebach ed., SPIE, 1993.
- [4] A.Ortega and K. Ramchandran. Rate-distortion methods for image and video compression. *IEEE Signal Processing Magazine*, pages 23–50, Nov 1998.
- [5] M. A. Breuer, S. K. Gupta, and T. M. Mak. Defect and error tolerance in the presence of massive numbers of defects.

- IEEE Design & Test of Computers*, 21:216–227, May–June 2004.
- [6] H. Y. Cheong and A. Ortega. System level fault tolerant motion estimation algorithms & techniques. Technical report, Signal and Image Processing Institute, Univ. of Southern California, 2006.
- [7] I. Chong and A. Ortega. Hardware testing for error tolerant multimedia compression based on linear transforms. In *Proc. of IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems, DFT'05*, pages 523–534, 2005.
- [8] H. Chung and A. Ortega. Analysis and testing for error tolerant motion estimation. In *Proc. of IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems, DFT'05*, pages 514–522, 2005.
- [9] A. M. Eskicioglu. Quality measurement for monochrome compressed images in the past 25 years. In *Journal of Electronic Imaging*, volume 10(1), pages 20–29, 2001.
- [10] B. Girod. *What's wrong with mean-squared error*. MIT Press, United States, 1993.
- [11] R. Hedge and N. R. Shanbhag. Soft digital signal processing. In *IEEE Trans. on very large scale integration (VLSI) systems*, volume 9, 2001.
- [12] N. Jha and S. Gupta. *Testing of Digital Systems*. Cambridge University Press, United Kingdom, 2003.
- [13] J. L. Mitchell and W. B. Pennebaker. *JPEG Still image data compression standard*. Van Nostrand Reinhold, New York, 1993.
- [14] J. L. Mitchell, W. B. Pennebaker, C. E. Fogg, and D. J. LeGall. *MPEG Video Compression Standard*. Chapman and Hall, New York, 1997.
- [15] H. A. Peterson, H. Peng, J. H. Morgan, and W. B. Pennebaker. Quantization of color image components in the DCT domain. In *Human Vision, Visual Processing, and Digital Display II., SPIE*, volume 1453, 1991.
- [16] R. Rosenholtz and A. B. Watson. Perceptual adaptive jpeg coding. In *Proceedings of ICIP*, volume 1, pages 901–904, 1996.
- [17] K. L. Shepard and V. Narayanan. Noise in deep submicron digital design. In *ICCAD*, pages 524–531, San Francisco, CA, Nov. 1996.
- [18] A. B. Watson, J. Hu, and J. F. McGowan. Digital video quality metric based on human vision. In *Journal of Electronic Imaging*, volume 10(1), pages 20–29, 2001.
- [19] T. Wiegand and B. Girod. Lagrange multiplier selection in hybrid video coder control. In *Proceedings of the 2001 IEEE International Conference on Image Processing (ICIP'01)*, volume 3, pages 542–545, 2001.