We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



3

Quantifying Interpretability Loss due to Image Compression

John M. Irvine[1] and Steven A. Israel[2]
[1]*Draper Laboratory, Cambridge, MA,*
[2]*Scientist,*
USA

1. Introduction

Video imagery provides a rich source of information for a range of applications including military missions, security, and law enforcement. Because video imagery captures events over time, it can be used to monitor or detect activities through observation by a user or through automated processing. Inherent in these applications is the assumption that the image quality of the video data will be sufficient to perform the required tasks. However, the large volume of data produced by video sensors often requires data reduction through video compression, frame rate decimation, or cropping the field-of-view as methods for reducing data storage and transmission requirements. This paper presents methods for analyzing and quantifying the information loss arising from various video compression techniques. The paper examines three specific issues:

- **Measurement of image quality**: Building on methods employed for still imagery, we present a method for measuring video quality with respect to performance of relevant analysis tasks. We present the findings from a series of perception experiments and user studies which form the basis for a quantitative measure of video quality.
- **User-based assessments of quality loss**: The design, analysis, and findings from a user-based assessment of image compression are presented. The study considers several compression methods and compression rates for both inter- and intra-frame compression.
- **Objective measures of image compression**: The final topic is a study of video compression using objective image metrics. The findings of this analysis are compared to the user evaluation to characterize the relationship between the two and indicate a method for performing future studies using the objective measures of video quality.

1.1 Information content and analysis

Video data provides the capability to analyze temporal events which enables far deeper analysis than is possible with still imagery. At the primitive level, analysis of still imagery depends on the static detection, recognition, and characterization of objects, such as people or vehicles. By adding the temporal dimension, video data reveals information about the movement of objects, including changes in pose and position and changes in the spatial

configuration of objects. This additional information can support the recognition of basic activities, associations among objects, and analysis of complex behavior (Fig. 1).

Fig. 1 is a hierarchy for target recognition information complexity. Each box's color indicates the ability of the developer community to assess the performance and provide confidence measures. The first two boxes on the left exploit information in the sensor phenomenology domain. The right two boxes exploit extracted features derived from the sensor data.

To illustrate the concept, consider a security application with a surveillance camera overlooking a bank parking lot. If the bank is robbed, a camera that collects still images might acquire an image depicting the robbers exiting the building and show several cars in the parking lot. The perpetrators have been detected but additional information is limited. A video camera might collect a clip showing these people entering a specific vehicle for their getaway. Now both the perpetrators and the vehicle have been identified because the activity (a getaway) was observed. If the same vehicle is detecting on other security cameras throughout the city, analysis of multiple videos could reveal the pattern of movement and suggest the location for the robbers' base of operations. In this way, an association is formed between the event and specific locations, namely the bank and the robbers' hideout. If the same perpetrators were observed over several bank robberies, one could discern their pattern of behavior, i.e. their *modus operandi*. This information could enable law enforcement to anticipate future events and respond appropriately (Gualdi *et al.* 2008; Porter *et al.* 2010).

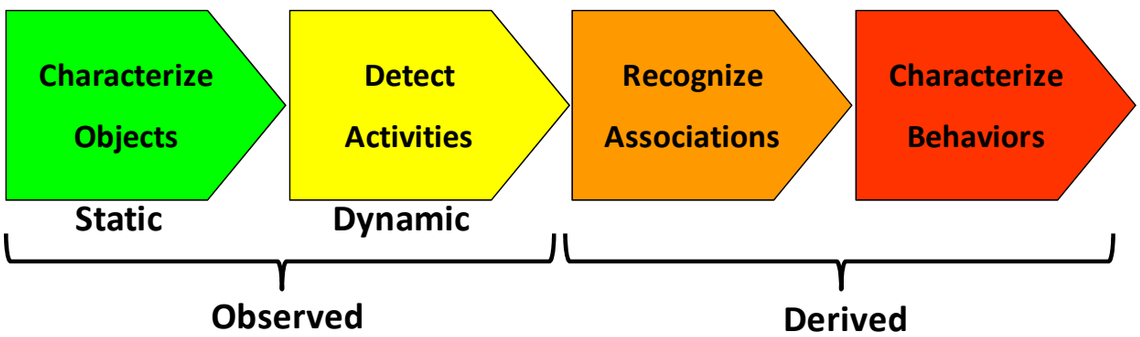


Fig. 1. Image Exploitation and Analysis

1.2 Image interpretability

A fundamental premise of the preceding example is that the imagery, whether a still image or a video clip, is of sufficient quality that the appropriate analysis can be performed (Le Meur *et al.* 2010; Seshadrinathan *et al.* 2010; Xia *et al.* 2010). Military applications have led to the development of a set of standards for assessing and quantifying this aspect of the imagery. The National Imagery Interpretability Rating Scale (NIIRS) is a quantification of image interpretability that has been widely applied for intelligence, surveillance, and reconnaissance (ISR) missions (Irvine 2003; Leachtenauer 1996; Maver *et al.* 1995). Each NIIRS level indicates the types of exploitation tasks an image can support based on the expert judgments of experienced analysts. Development of a NIIRS for a specific imaging modality rests on a perception-based approach. Additional research has verified the relationship between NIIRS and performance of target detection tasks (Baily 1972; Driggers *et al.* 1997; Driggers *et al.* 1998; Lubin 1995). Accurate methods for predicting NIIRS from the

sensor parameters and image acquisition conditions have been developed empirically and substantially increase the utility of NIIRS (Leachtenauer *et al.* 1997; Leachtenauer and Driggers 2001).

The NIIRS provides a common framework for discussing the interpretability, or information potential, of imagery. NIIRS serves as a standardized indicator of image interpretability within the community. An image quality equation (IQE) offers a method for predicting the NIIRS of an image based on sensor characteristics and the image acquisition conditions (Leachtenauer *et al.* 1997; Leachtenauer and Driggers 2001). Together, the NIIRS and IQE are useful for:

- Communicating the relative usefulness of the imagery,
- Documenting requirements for imagery,
- Managing the tasking and collection of imagery,
- Assisting in the design and assessment of future imaging systems, and
- Measuring the performance of sensor systems and imagery exploitation devices.

The foundation for the NIIRS is that trained analysts have consistent and repeatable perceptions about the interpretability of imagery. If more challenging tasks can be performed with a given image, then the image is deemed to be of higher interpretability. A set of standard image exploitation tasks or "criteria" defines the levels of the scale. To illustrate, consider Fig. 2. Several standard NIIRS tasks for visible imagery appear at the right. Note that the tasks for levels 5, 6, and 7 can be performed, but the level 8 task cannot. The grill detailing and/or license plate on the sedan are not evident. Thus, an analyst would assign a NIIRS level 7 to this image.

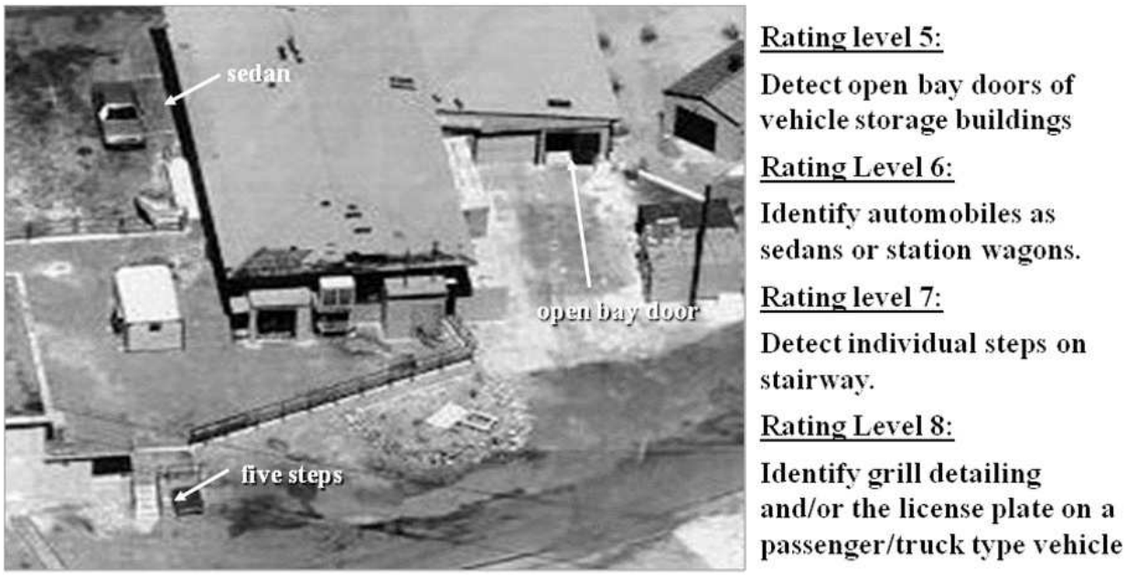


Fig. 2. Illustration of NIIRS for a still image

Recent studies have extended the NIIRS concept to motion imagery (video). In exploring avenues for the development of a NIIRS-like metric for motion imagery, a clearer understanding of the factors that affect the perceived quality of motion imagery was needed (Irvine *et al.* 2006a; Young *et al.* 2010b). Several studies explored specific aspects of this

problem, including target motion, camera motion, and frame rate, and the nature of the analysis tasks (Hands 2004; Huynh-Thu *et al.* 2011; Moorthy *et al.* 2010). Factors affecting perceived interpretability of motion imagery include the ground sample distance (GSD) of the imagery, motion of the targets, motion of the camera, frame rate (temporal resolution), viewing geometry, and scene complexity. These factors have been explored and characterized in a series of evaluations with experienced imagery analysts:

- Spatial resolution: Evaluations shows that for motion imagery the interpretability of an video clip exhibits a linear relationship with the natural log of the ground sample distance (GSD), at least for clips where the GSD is fairly constant over the clip (Cermak *et al.* 2011; Irvine *et al.* 2004; Irvine *et al.* 2005; Irvine *et al.* 2007b) .
- Motion and Complexity: User perception evaluations assessed the effects of target motion, camera motion, and scene complexity on perceived image quality (Irvine *et al.* 2006b). The evaluations indicated that target motion has a significant positive effect on perceived image quality, whereas camera motion has a barely discernable effect.
- Frame Rate: These evaluations assessed object detection and identification and other image exploitation tasks as a function of frame rate and contrast (Fenimore *et al.* 2006). The study demonstrated that an analyst's ability to detect and recognize objects of interest degrades at frame rates below 15 frames per second. Furthermore, the effect of reduced frame rate is more pronounced with low contrast targets.
- Task Performance: The evaluations assessed the ability of imagery analysts to perform various image exploitation tasks with motion imagery. The tasks included detection and recognition of objects, as might be done with still imagery and the detection and recognition of activities, which relies on the dynamic nature of motion imagery (Irvine *et al.* 2006b; Irvine *et al.* 2007c). Analysts exhibited good consistency in the performance of these tasks. In addition, dynamic exploitation tasks that require detection and recognition of activities are sensitive to the frame rate of the video clip.

Building on these perceptions studies, a new Video NIIRS was developed (Petitti *et al.* 2009; Young *et al.* 2009). The work presented in this paper quantifies video interpretability using a 100-point scale described in Section 3 (Irvine *et al.* 2007a; Irvine *et al.* 2007b; Irvine *et al.* 2007c). The scale development methodologies imply that each scale is a linear transform of the other, although this relationship has not been validated (Irvine *et al.* 2006a; Irvine *et al.* 2006b). Other methods for measuring video image quality frequently focus on objective functions of the imagery data, rather than perception of the potential utility of the imagery to support specific types of analysis (Watson *et al.* 2001; Watson and Kreslake 2001; Winkler 2001; Winkler *et al.* 2001).

2. Image compression

A recent study of compression for motion imagery focused on objective performance of target detection and target tracking tasks to quantify the information loss due to compression (Gibson *et al.* 2006). Gibson *et al.* (2006) leverage recent work aimed at quantification of the interpretability of motion imagery (Irvine *et al.* 2007b). Using techniques developed in these earlier studies, this paper presents a user evaluation of the interpretability of motion imagery compressed under three methods and various bitrates. The interpretability of the native, uncompressed imagery establishes the reference for comparison (He and Xiong 2006; Hewage *et al.* 2009; Yang *et al.* 2010; Yasakethu *et al.* 2009).

2.1 Data compression

The dataset for the study consisted of the original (uncompressed) motion imagery clips and clips compressed by three compression methods at various compression rates (Abomhara *et al.* 2010). The three compression methods were:

- Motion JPEG 2000 – intraframe only
- MPEG-2 – intraframe and interframe
- H.264 – intraframe and interframe

All three were exercised in intraframe mode. Each of the parent clips was compressed to three megabits per second, representing a modest level of compression. In addition, each parent clip was severely compressed to examine the limits of the codecs. Actual bitrates for these severe cases depend on the individual clip and codec. The choice of compression methods and levels supports two goals: comparison across codecs and comparisons of the same compression method at varying bitrates. Table 1 shows the combinations represented in the study. We recorded the actual bit rate for each product and use this as a covariate in the analysis.

The study used the Kakadu implementation of JPEG2000, the Vanguard Software Solutions, Inc. implementation of H.264, and the Adobe Premiere's MPEG-2 codec. In each case, the 300 key frame interval was used for interframe compression unless otherwise noted. Intraframe encoding is comparable to interframe encoding with 1 key frame interval.

The study used progressive scan motion imagery in a 848 x 480 pixel raster at 30 frames per second (f/s). Since most of the desirable source material was available to us in 720 P HD video, a conversion process was employed to generate the lower resolution/lower frame rate imagery. We evaluated the conversion process to assure the goals of the study could be met. The video clips were converted using Adobe Premiere tools.

Bitrate	**Uncompressed**	**H.264 (VSS)**		**JPEG 2000 (KDU)**	**MPEG 2 (Premiere)**	
		Inter-frame	**Intra-frame**	**Intra-frame**	**Inter-frame**	**Intra-frame**
Native	X					
3 MB/sec		X	X	X	X	X
Severe		X		X	X	

Note: the severe bitrate represents the limit of the specific codec on a given clip.

Table 1. Codecs and Compression Rates

2.2 Experimental design

The study consists of two parts. Both parts used the set of compression products described above. The first part was an evaluation in which trained imagery analysts viewed the compressed products and the original parent clip to assess the effects of compression on interpretability. The second part of the study implemented a set of computational image metrics and examined their behavior with respect to bitrate and codec. The typical duration of each clip is 10 seconds. Ten video clips were used for this study.

3. User-based evaluation of compression

To quantify image interpretability, subjective rating scale was developed by Irvine *et al.* (2007c), based on consistent ratings by trained imagery analysts. The scale assigns the values 0 to a video clip of no utility and 100 to clips that could support any of the analysis tasks under consideration (Fig. 3). Three additional clips identified in this study formed markers to evenly divide the subjective interpretability space. Thus, reference clips were available at subjective rating levels of 0, 25, 50, 75, and 100.

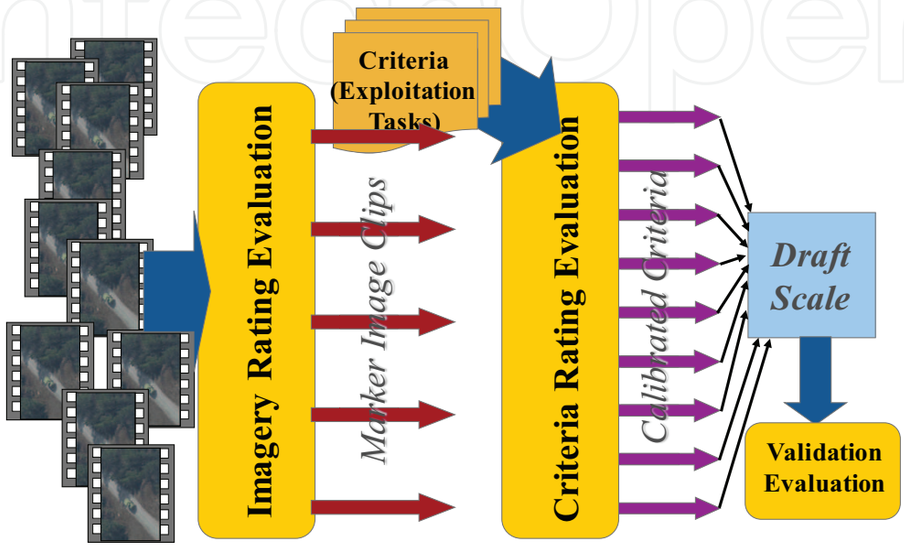


Fig. 3. NIIRS Development Functional Decomposition

A set of specific image exploitation tasks were reviewed by imagery analysts and rated relative to these marker video clips. In this way, these analysis tasks were calibrated to the subjective rating scale. A subset of these "calibrated" analysis tasks were used to evaluate the compressed video products (Table 2). Note that some of these tasks do not require analysis of temporal activity and could be performed with still imagery. We label these as "static" tasks. A second set of tasks are "dynamic" because they require direct observation or inference about movements of objects.

Image analysts rated their confidence in performing each image exploitation task with respect to each compression product, including the original (uncompressed) clip. We calculated an overall interpretability rating from each analyst for each clip.

3.1 Approach

For each parent clip, three criteria (image exploitation tasks) were assigned. The considerations for selecting the criteria were:

- The criteria should "bound" the interpretability of the parent clip, i.e. at least one of the three should be difficult to do and one should be easy
- The criteria (or at least some of the criteria) should reference objects and activity that are comparable to the content of the clip
- The criteria should have exhibited low rater variance in the previous evaluations

Static Criteria	Detect large buildings (e.g. hospitals, factories)
	Identify individual rail cars (e.g. gondola, flat, box) and locomotives by type (e.g. steam, diesel)
	Identify vehicles by general class (civilian, military tracked, artillery, construction)
	Detect presence of freight in open-bed trucks
	Detect gatherings of 5 or more people
Dynamic Criteria	Determine motion and direction of large vehicles (e.g. trains, barges, 18-wheelers)
	Track a civilian-size moving vehicle (e.g., car, truck, SUV)
	Distinguish between an orderly assembly and a panicked crowd
	Distinguish among a person walking, a person running, and a person riding a small vehicle (bicycle, moped, or motorcycle)
	Determine whether a person is getting into or out of a vehicle

Table 2. Video Analysis Tasks

3.2 Analysis and findings

The data analysis progresses through several stages: verification and quality control, exploratory analysis to uncover interesting relationships, and statistical modeling to validate findings and formally test hypotheses of interest. The initial analysis examined the data for anomalies or outliers. None were found in this case.

Next, we calculated an overall interpretability rating from each analyst for each clip. The method for calculating these ratings was as follows: Each of the three criteria used to rate each clip was calibrated (on a 0-100 scale) in terms of interpretability, where this calibration was derived from an earlier evaluation (Irvine *et al.* 2007c). Multiplying the interpretability level by the IA's confidence rating produces a score for each criterion. The final interpretability score (Equation 1) was the maximum of the three scores for a given clip.

$$\text{Interpretability Score}(j, k) = \max \{C_{i,j,k}\, I_{i,k} : i=1,2,3\} / 100 \quad (1)$$

Where $C_{i,j,k}$ is the confidence rating by the jth IA on the kth clip for the ith criterion and $I_{i,k}$ is the calibrated interpretability level for that criterion. All subsequent analysis presented below is based on this final interpretability score. The remaining analysis is divided into two sections: interframe compression and intraframe compression. Ultimately, we compared the

analyst derived utility measures, as a NIIRS surrogate, to the automated computational values.

All three codecs yielded products for the evaluation. However, MPEG-2 would not support extreme compression rates. Bitrate was the dominant factor, but pronounced differences among the codecs emerged too (Fig. 6 and Fig. 7). At modest compression rates, MPEG-2 exhibited a substantial loss in interpretability compared to either H.264 or JPEG-2000. Only JPEG-2000 supported more extreme intraframe compression. A computational model was developed to characterize the significance among codec, scene, and bitrate's effect on data quality. There were systematic differences across the clips, as expected, but the effects of the codecs and bitrates were consistent. When modeled as a covariate, the effects of bitrate dominate. The effect due to codec is modest, but still significant. As expected, there is a significant main effect due to scene, but no scene-by-codec interaction.

3.3 Interframe compression

Analysis of the interframe ratings shows a loss in image interpretability for both MPEG-2 and H.264 as a function of bitrate (Fig. 4). The initial compression from the native rate to 3 MB per second corresponds to a modest loss in interpretability. This finding is consistent with previous work. At extreme compression levels (below 1 MB per second), the interpretability loss is substantial. H.264 generally supported more extreme compression levels, but the interpretability degrades accordingly. Although the exact compression level

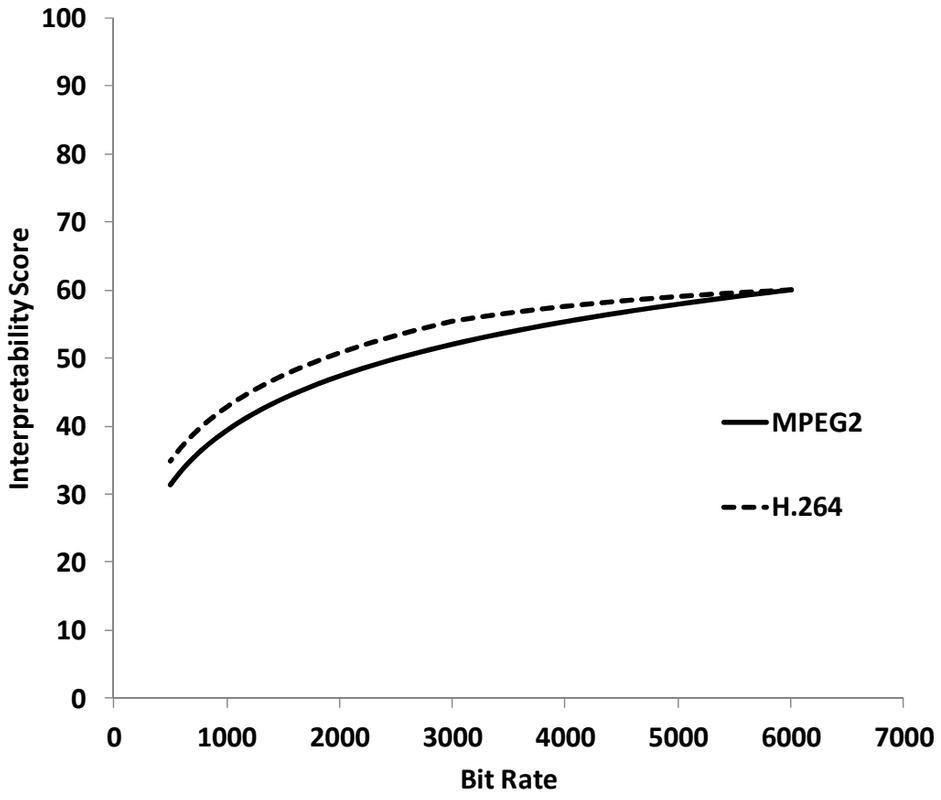


Fig. 4. Summary Comparison Across Codec and Bitrate

varies by clip, the pattern is clear for all clips. Statistical analysis shows that bitrate, modeled as a covariate, is the primary factor affecting interpretability. For interframe compression, the differences between H.264 and MPEG-2 are small but statistically significant (Table 3). The pattern holds across all scenes, as indicated by the lack of a codec-by-scene interaction effect.

Source	Deg. Of Freedom	F-statistic	Significance <0.025
Intercept	1	6.2	0.0033
BitRate	1	59.7	0.00003
Codec	2	5.7	0.015
Scene	4	15.4	0.00028
Codec * Scene	8	0.5	0.82

Table 3. Analysis of Covariance for Interframe Comparisons

3.4 Intraframe compression

In the case of intraframe compression, all three codecs yielded products for the evaluation, although MPEG-2 would not support extreme compression rates. The findings in this case are slightly different than interframe compression. Bitrate remained the dominant factor, but more pronounced differences among the codecs emerged (Figure 4). At modest compression rates, MPEG-2 exhibited a substantially loss in interpretability compared to either H.264

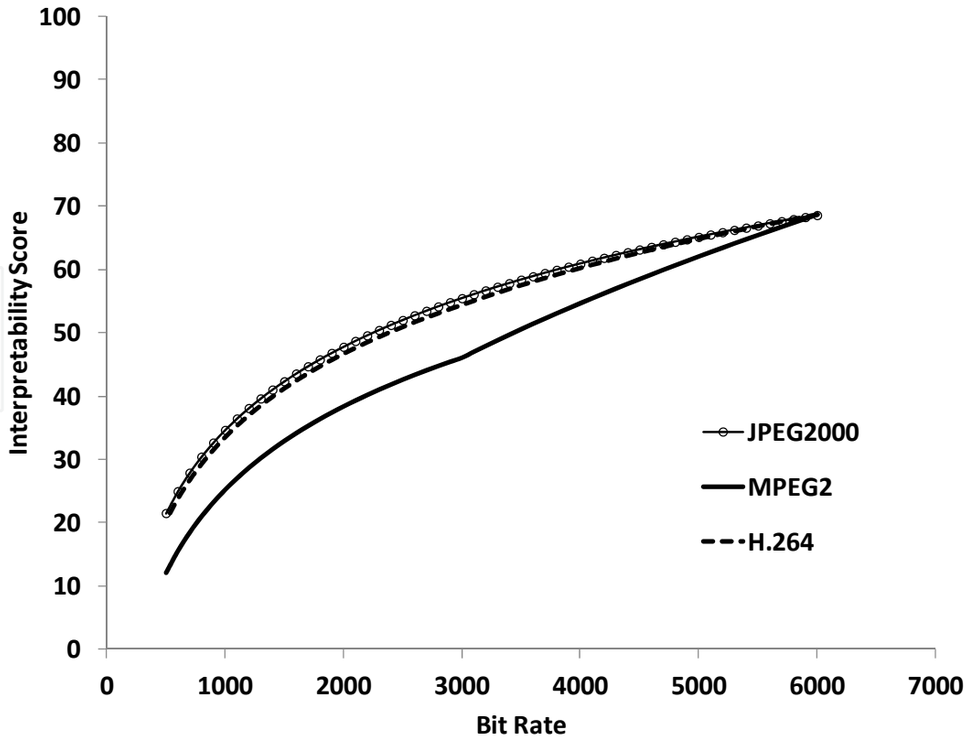


Fig. 5. Summary Comparison Across Codec and Bitrate

or JPEG-2000. Only JPEG-2000 supported more extreme intraframe compression and highly compressed renditions were produced from all of the parent clips. As with the interframe comparisons, there were systematic differences across the clips, as expected, but the effects of the codecs and bitrates were consistent. The analysis of covariance confirms these statistical effects (Table 4). When modeled as a covariate, effects of bitrate dominate. The effect due to codec is modest, but still significant. As expected, there is a significant main effect due to scene, but no scene-by-codec interaction.

Source	Deg. of Freedom	F-statistic	Significance < 0.025
Intercept	1	0.9	0.0038
BitRate	1	24.4	0.0078
Codec	3	5.6	0.001
Scene	4	26.2	0.00001
Codec * Scene	12	0.5	0.84

Table 4. Analysis of Variance for Interframe Comparisons

4. Computational measures and performance assessment

In the previous section, analyst assessments of image quality were characterized. This section identifies computational attributes for image quality that can be extracted from video clips. Performance measures will evaluate the computational image quality metrics and provide an understanding of how well they compare to codec, bitrate, and scene parameters.

4.1 Computational image metrics

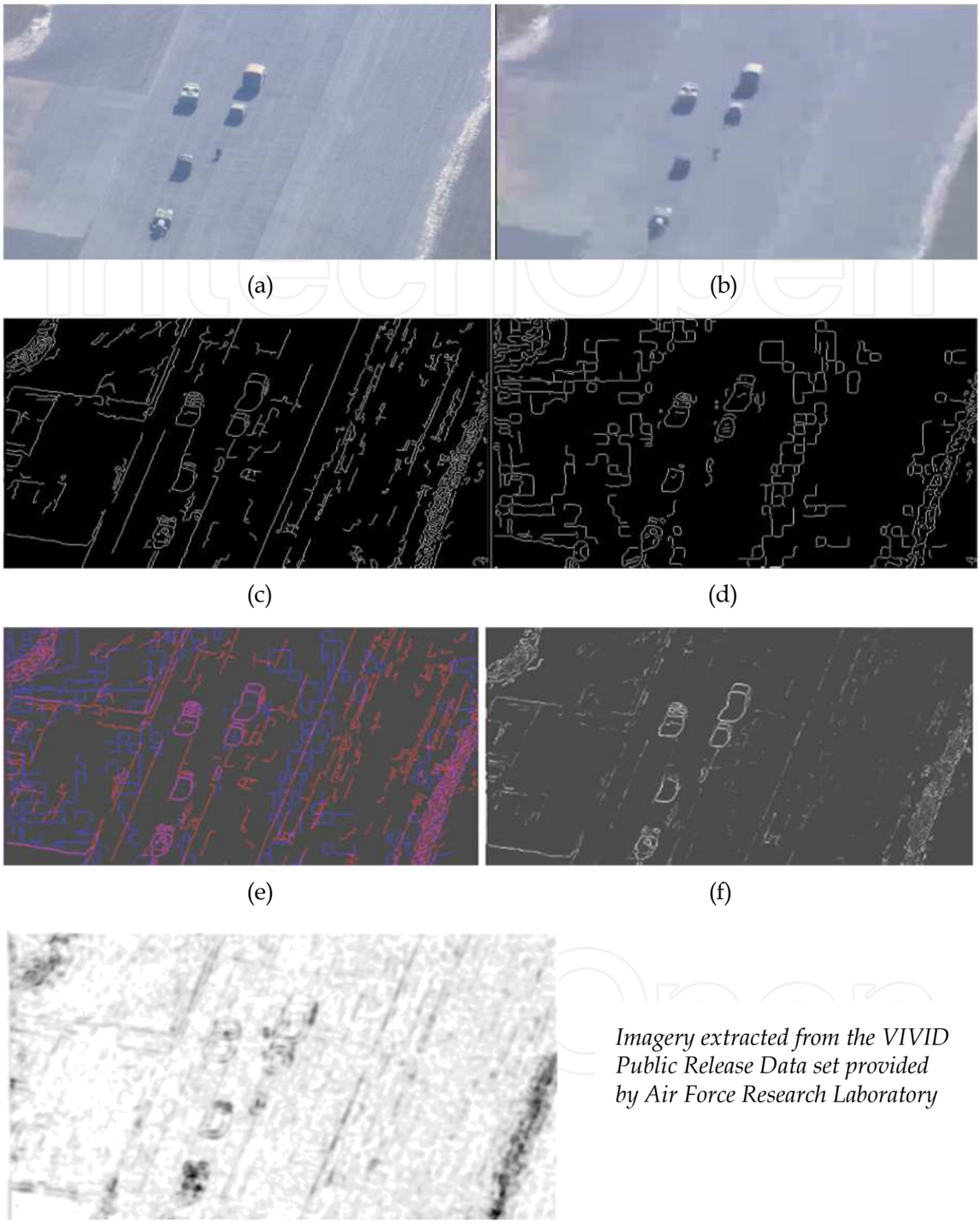
We reviewed a variety of image metrics to quantify image quality (Bhat *et al.* 2010; Chikkerur *et al.* 2011; Culibrk *et al.* 2011; Huang 2011; Sohn *et al.* 2010). Based on a review of the literature and assessment of the properties of these metrics, we selected four measures for this study: two edge-based metrics, structural similarity image metric (SSIM), and SNR. SSIM and edge metrics are performed at each pixel location. The resultant can be viewed as an image (Fig. 6). SNR metrics deal with overall information content and cannot be visualized as an image. These metrics were computed for the original (uncompressed) clips and for all of the compressed products. We will present the computation methods and the results.

The color information was transposed into panchromatic (intensity) using either a HSI transformation or luminance. Intensity was computed using (2):

$$I = \frac{R + G + B}{3} \tag{2}$$

Luminance was computed (3)

$$Y = 0.299R + 0.587G + 0.114B \tag{3}$$



Imagery extracted from the VIVID Public Release Data set provided by Air Force Research Laboratory

Fig. 6. (a) Original, (b) Compressed version, (c,d) original edge images, (e) edge images displayed together where Red, Blue, Magenta are from the original, the compressed, both edge images respectively (f) edge intensities, and (g) is the SSIM image darker areas represent more noticeable differences.

4.1.1 SSIM

The first metric for image quality is the Structural Similarity Image Metric (Wang *et al.* 2004). SSIM quantifies differences between two images, I_1 and I_2, by taking three variables into consideration, luminance, contrast, and spatial similarity. For grey level images, those variables are measured in the images as the mean, standard deviation, and Pearson's correlation coefficient between the two images respectively. For our application, the RGB data was converted to grey level using the standard Matlab function. Let:

$$\mu = mean(I_1)$$
$$\sigma_1 = \text{standard deviation}(I_1)$$
$$\sigma_{12} = \frac{1}{N-1}\sum_{i.j}(I_1(i,j)-\mu_1)(I_2(i,j)-\mu_2) : \text{covariance}$$

then

$$\text{SSIM}(I_1, I_2) = \frac{2\mu_1\mu_2}{\mu_1^2+\mu_2^2} * \frac{2\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2} * \frac{\sigma_{12}}{\sigma_1^2\sigma_2^2} \quad (4)$$

Equation 4 is modified to avoid singularities, e.g., when both means are 0. SSIM is computed locally on each corresponding MxM sub-image of I_1 and I_2. In practice, the sub-image window size is 11x11, implemented as a convolution filter. The SSIM value is the average across the entire image.

4.1.2 Edge metrics

Two edge metrics were examined. The first is denoted by CE for Common Edges and the second is denoted SE for strength of edges (O'Brien *et al.* 2007). Heuristically, CE measures the ratio of the number edges in a compressed image to the number of edges in the original; whereas SE measures a ratio of the strength of the edges in a compressed version to strength of the edges in the original.

Given two images I_1 and I_2 CE(I_1, I_2) and SE(I_1, I_2) are computed as follows. From the grey level images, edge images are constructed using the Canny edge operator. The edge images are designated as E_1 and E_2. Assume that the values in E_1 and E_2 are 1 for an edge pixel and 0 otherwise. Let "*" denote the pixel wise product. Let G_1 and G_2 denote the gradient images of I_1 and I_2 respectively. G(m,n) was approximated as the maximum of absolute value of the set {I(m,n) - I(m+t1,n + t2) | -6 < t1 < 6 and -6< t2 < 6}, i.e. the maximum difference between the center value and all values in a 5x5 neighborhood around it. With that notation,

$$CE(I_1, I_2) = \frac{2 * \sum(E_1 * E_2)}{\sum E_1 + \sum E_2} \quad (5)$$

where the sum is taken over all the pixels within a given frame.

$$SE(I_1, I_2) = \frac{\sum(E_2 * G_2)}{\sum E_1 * G_1} \quad (6)$$

where the sum is taken over all the pixels within a given frame.

An additional set of edge operators were also applied. These operators are called edge strength (ES) metrics. Let Y_1 be the luminance component of an original frame from an clip and let Y_2 be the corresponding frame after compression processing, also in luminance. We apply a Sobel filter, S, to both Y_1 and Y_2, where for a grayscale frame F:

$$S(F) = \sqrt{(H * F)^2 + (V * F)^2} \tag{7}$$

The filters H and V used in the Sobel edge detector are:

$$H = \begin{matrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{matrix} \tag{7.1}$$

$$V = H^T \tag{7.2}$$

We define two metrics, one for local loss of edge energy (EL) (thus finding blurred edges from Y_1 in Y_2) and the other for the addition of edge energy (thus finding edges added to Y_2 that are weaker in Y_1). Each metric examines the strongest edges in one image (either Y_1 or Y_2) and compares them to the edges at the corresponding pixels in the other (Y_2 or Y_1).

For the grayscale image F, let I(F,f) be the set of image pixels, p, where F (p) is at least as large as f * max(F). That is:

$$I(F, f) = \{Pixels : F(p) \geq f * \max(F)\} \tag{8}$$

Using the definition of Y(F,f), the two edge metrics are:

$$BlurIndex = \frac{mean(S(Y_2))}{mean(S(Y_1))} \tag{9}$$

where the means are taken over the set $I(S(I_1), 0.99)$

$$AddedEdgeEnergy = \frac{mean(S(Y_1))}{mean(S(Y_2))} \tag{10}$$

where the means are taken over the set $I(S(Y_2), 0.99)$.

4.1.3 SNR

Finally, we examined the peak signal to noise ratio (PSNR). The PSNR is defined for a pair of m×n luminance images, Y_1 and Y_2. Let MSE be defined by,

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \left\| Y_1(i,j) - Y_2(i,j) \right\|^2 \tag{11}$$

The PSNR is defined as:

$$PSNR = 10\log_{10}\frac{MAX_I^2}{MSE} = 20\log_{10}\frac{MAX_I}{\sqrt{MSE}} \quad (12)$$

where MAX_I is maximum pixel value of the image. In our case, MAX_I is taken to be 255.

4.2 Metrics and performance

The image metrics were plotted (Fig. 7). The image metrics are all highly correlated across both bitrate and codec, for both intraframe and interframe compression techniques. For the set of clips with every 300 key frame interval, the correlation was greater than 0.9. In each case, the lower information content is indicated by lower position on the Y axis, quality. The X axis is the target bitrate. Due to the high correlation a single computational metric was chosen for more detailed analysis to quantify the relationship between image quality and bitrate. SSIM was selected because it generates an image to diagnose unexpected values and the computation is based upon perceptual difference of spatial similarity and contrast.

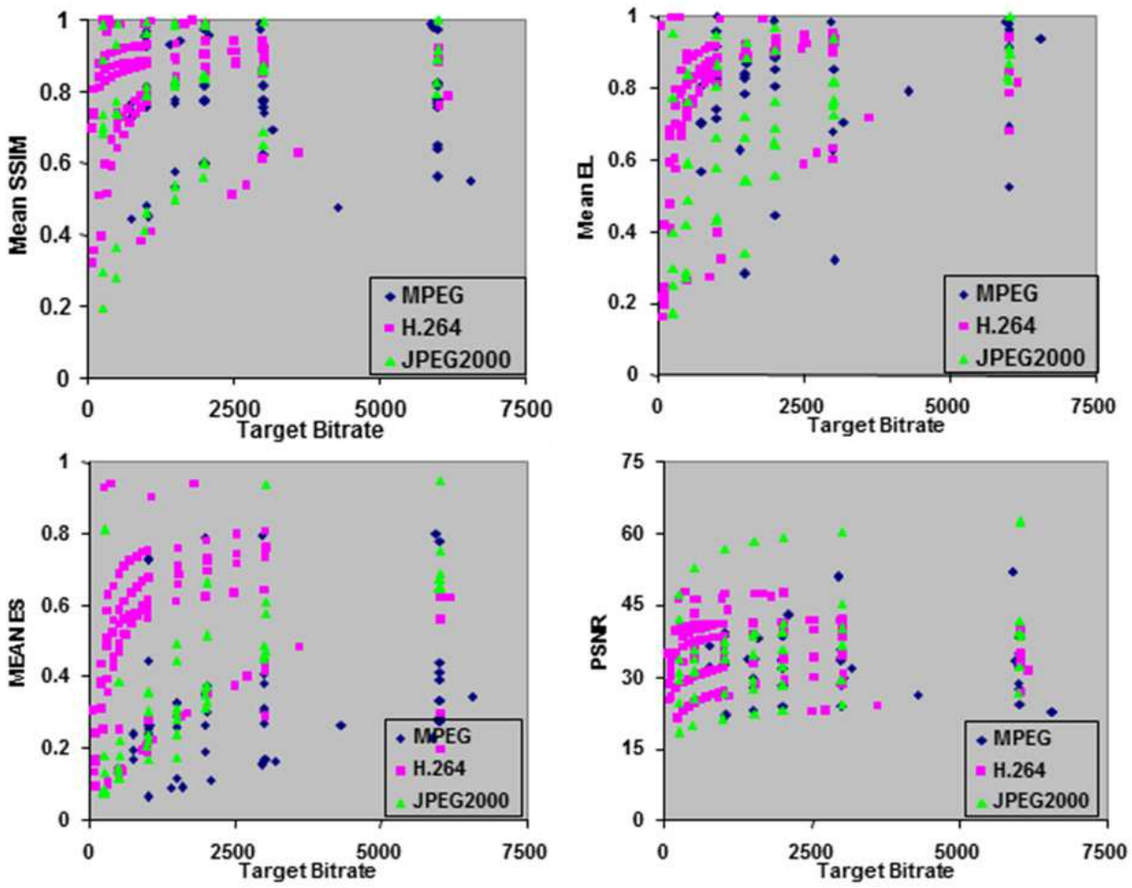


Fig. 7. Target Bitrate (k bps) versus Image Metric: SSIM, EL, ES, and PSNR

Fig. 7 indicates that SSIM, CE, and SE that measures separate image quality based on bitrate. H.264's asymptotic quality improvement observed in the rise in the graph from the initial frames (Fig. 7). This corresponds to exactly where the algorithm is increasing its fidelity of

the compressed frames to the original frames. Along this initial portion of the clip the metrics agree with human perception of the image quality increasing.

Fig. 8 plots SSIM versus frame at differeing bitrates for the H.264 codec, which is an interframe codec. The saw-tooth nature of the graph is the result of the group of pictures (GOP) sequence. The peak and trough differences are between bilinear interpolation between key frames (B) and predicted (P) encoded frames.

The observations for the metrics listed above for H.264 were also visually evident in the case of MPEG compression. Close inspection of the clips shows the quality to be lower in the case of MPEG than for H.264. The example in Fig. 9 is taken from a clip that was compressed to 2 Mbits/second using both codecs. While discernable in both the original and the H.264 compressed versions, some of the individuals' heads seem to be nearly totally lost in the MPEG version.

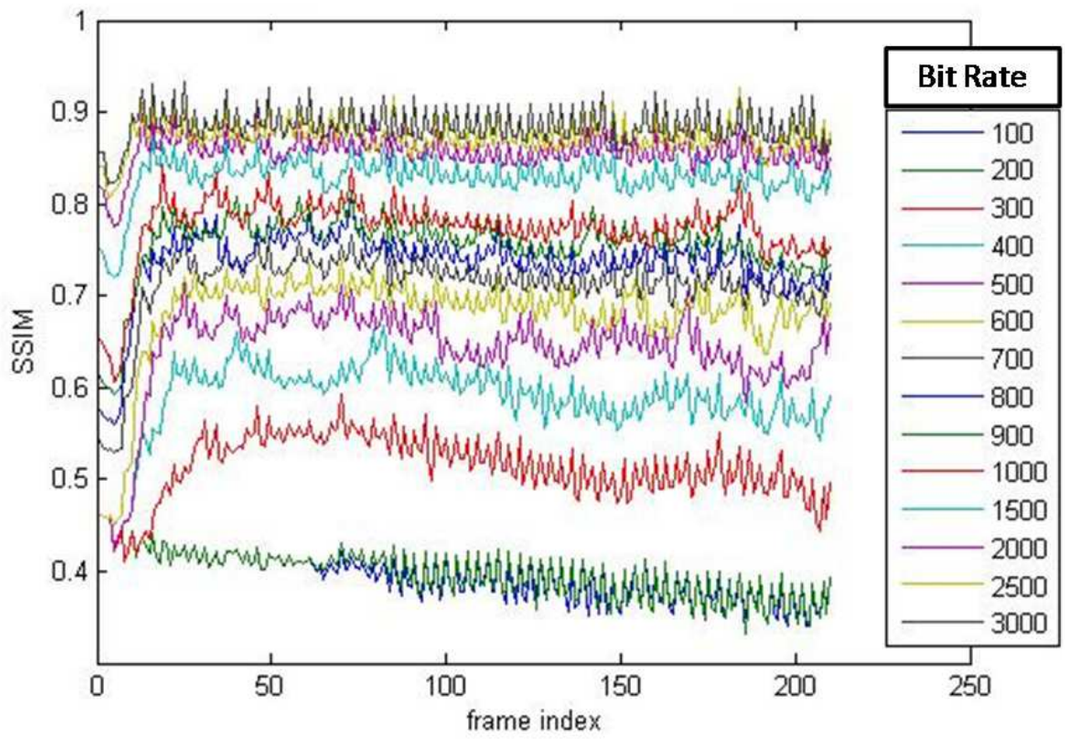
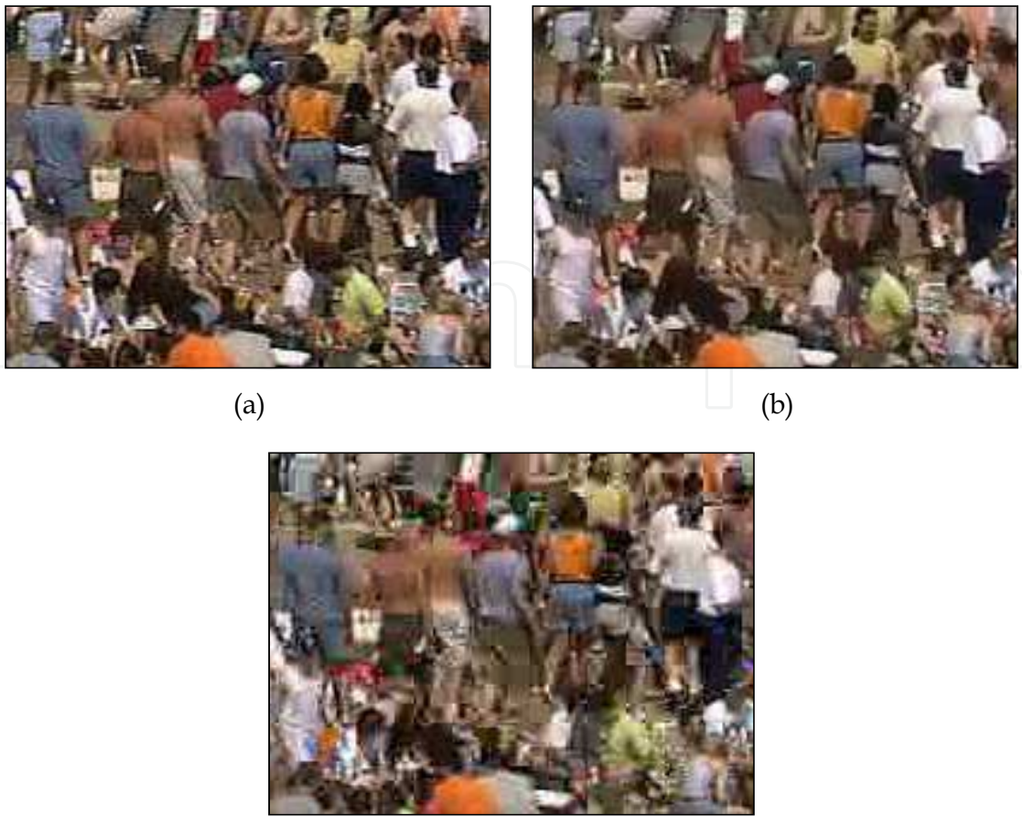


Fig. 8. Plot of the SSIM evaluated on each frame for 11 different bit rates.
Each clip was compressed using H.264 with the key frames 1 every 300 frames.

5. Discussion

These experiments demonstrate the existence of several metrics that are monotonic with bitrate. The metrics showed considerable sensitivity to image quality that matched the authors' observations. Specifically, the MPEG quality was considerably less than H.264 at the same bitrate. The knee of the quality curves exist between 500k and 1000k bps. In addition, the metrics were sensitive to the encoded structure of the individual frames as the saw tooth differences between the B and P frames were readily observable.



(a) (b)

(c)

Fig. 9. (a) Original frame and compressed using (b) H.264 and (c) MPEG.

A qualitative comparison of the objective metrics to the user assessment of interpretability shows strong consistency. Compression of these video products to bitrates below 1,000k bps yields discernable losses in image interpretability. The objective metrics shows a similar knee in the curve. These data suggest that one could estimate loss in interpretability from compression using the objective metrics and derive a prediction of the loss in Video NIIRS. Development of such a model would require conducting a second user experiment to establish the relationship between the subjective interpretability scale used in this study and the published Video NIIRS. The additional data from such an experiment would also support validation of a model for predicting loss due to compression.

6. Conclusion

The evaluations and analyses presented in this Chapter characterize the loss in perceived interpretability of motion imagery arising from various compression methods and compression rates. The findings build on previous studies (Irvine *et al.* 2007a; O'Brien *et al.* 2007). The findings are consistent with other evaluations of video compression (Gibson *et al.* 2006; Young *et al.* 2010a). Evaluation of image compression for motion imagery illustrates how interpretability-based methods can be applied to the analysis of the image chain. We present both objective image metrics and analysts' assessments of various compressed products. The results show good agreement between the two approaches. Further research

is needed to validate a model-based relationship that could predict Video NIIRS loss due to compression using the objective image metrics presented here.

7. References

Abomhara, M., Khalifa, O.O., Zakaria, O., Zaidan, A.A., Zaidan, B.B., and Rame, A. 2010. Video Compression Techniques: An Overview, *Journal of Applied Sciences,* 10(16): 1834-1840.

Baily, H.H. 1972. Target Acquisition Through Visual Recognition: An Early Model, *Target Acquisition Symposium,* Orlando, 14-16 November.

Bhat, A., Richardson, I., and Kannangara, S. 2010. A New Perceptual Quality Metric for Compressed Video Based on Mean Squared Error, *Signal Processing: Image Communication,* 25(7): 588-596.

Cermak, G., Pinson, M., and Wolf, S. 2011. The Relationship Among Video Quality, Screen Resolution, and Bit Rate, *IEEE Transactions on Broadcasting,* 57(2): 258-262.

Chikkerur, S., Sundaram, V., Reisslein, M., and Karam, L.J. 2011. Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison, *IEEE Transactions on Broadcasting,* 57(2): 165-182.

Culibrk, D., Mirkovic, M., Zlokolica, V., Pokric, M., Crnojevic, V., and Kukolj, D. 2011. Salient Motion Features for Video Quality Assessment, *IEEE Transactions on Image Processing,* 20(4): 948-958.

Driggers, R.G., Cox, P.G., and Kelley, M. 1997. National Imagery Interpretation Rating System and the Probabilities of Detection, Recognition, and Identification, *Optical Engineering,* 36(7): 1952-1959.

Driggers, R.G., Cox, P.G., Leachtenauer, J., Vollmerhausen, R., and Scribner, D.A. 1998. Targeting and Intelligence Electro-Optical Recognition and Modeling: A Juxtaposition of the Probabilities of Discrimination and the General Image Quality Equation, *Optical Engineering,* 37(3): 789-797.

Fenimore, C., Irvine, J.M., Cannon, D., Roberts, J., Aviles, I., Israel, S.A., Brennan, M., Simon, L., Miller, J., Haverkamp, D., Tighe, P.F., and Gross, M. 2006. Perceptual Study of the Impact of Varying Frame Rate on Motion Imagery Interpretability, *Human Vision and Electronic Imaging XI,* SPIE, San Jose, 16-19 January, 6057: 248-256.

Gibson, L., Irvine, J.M., O'Brien, G., Schroeder, S., Bozell, A., Israel, S.A., and Jaeger, L. 2006. User Evaluation of Differential Compression for Motion Imagery, *Defense and Security Symposium 2006,* SPIE, Orlando, 17-21 April, 6209: paper number 6209-03.

Gualdi, G., Prati, A., and Cucchiara, R. 2008. Video Streaming for Mobile Video Surveillance, *IEEE Transactions on Multimedia,* 10(6): 1142-1154.

Hands, D.S. 2004. A Basic Multimedia Quality Metric, *IEEE Transactions on Multimedia,* 6(6): 806-816.

He, Z. and Xiong, H. 2006. Transmission Distortion Analysis for Real-Time Video Encoding and Streaming Over Wireless Networks, *IEEE Transactions on Circuits and Systems for Video Technology,* 16(9): 1051-1062.

Hewage, C.T.E.R., Worrall, S.T., Dogan, S., Villette, S., and Kondoz, A.M. 2009. Quality Evaluation of Color Plus Depth Map-Based Stereoscopic Video, *IEEE Journal of Selected Topics in Signal Processing,* 3(2): 304-318.

Huang, S.C. 2011. An Advanced Motion Detection Algorithm with Video Quality Analysis for Video Surveillance Systems, *IEEE Transactions on Circuits and Systems for Video Technology,* 21(1): 1-14.

Huynh-Thu, Q., Garcia, M.N., Speranza, F., Corriveau, P., and Raake, A. 2011. Study of Rating Scales for Subjective Quality Assessment of High-Definition Video, *IEEE Transactions on Broadcasting,* 57(1): 1-14.

Irvine, J.M. 2003. *National Imagery Interpretability Rating Scale,* 1442-1456. In: Driggers, R.G. *Encyclopedia of Optical Engineering*, Marcel Dekker, Inc., New York.

Irvine, J.M., Fenimore, C., Cannon, D., Roberts, J., Israel, S.A., Simon, L., Watts, C., Miller, J.D., Avilés, A.I., Tighe, P.F., and Behrens, R.J. 2004. Feasibility Study for the Development of a Motion Imagery Quality Metric, *33rd Applied Imagery and Pattern Recognition Workshop: Image and Data Fusion,* IEEE Computer Society, Washington, 13-15 October: 179-183.

Irvine, J.M., Fenimore, C., Cannon, D., Roberts, J., Israel, S.A., Simon, L., Watts, C., Miller, J.D., Avilés, A.I., Tighe, P.F., and Behrens, R.J. 2005. Factors Affecting Development of a Motion Imagery Quality Metric, *Defense & Security Symposium,* SPIE, Orlando, Florida, 29-31 March, 5817: 116-123.

Irvine, J.M., Cannon, D., Miller, J., Bartolucci, J., Gibson, L., Fenimore, C., Roberts, J., Aviles, I., Brennan, M., Bozell, A., Simon, L., and Israel, S.A. 2006a. Methodology Study for Development of a Motion Imagery Quality Metric, *Defense and Security Symposium,* SPIE, Orlando, Florida, 17-21 April.

Irvine, J.M., Fenimore, C., Cannon, D., Haverkamp, D., Roberts, J., Israel, S.A., Simon, L., Miller, J., Avilés, A.I., and Brennan, M. 2006b. Development of a Motion Imagery Quality Metric, *American Society for Photogrammetry and Remote Sensing (ASPRS) Annual Meeting,* ASPRS, Reno, Nevada, 1-5 May.

Irvine, J.M., Avilés, A.I., Cannon, D.M., Fenimore, C., Israel, S.A., O'Brien, G., and Roberts, J. 2007a. Quantifying Interpretability for Motion Imagery: Applications to Image Chain Analysis, *10th International Conference on Information Fusion (Fusion 2007),* IEEE, Quebec, 9-12 July.

Irvine, J.M., O'Brien, G., Israel, S.A., Fenimore, C., Roberts, J., Brennan, M., Bartolucci, J., and Cannon, D.M. 2007b. Perceived interpretability of Motion Imagery: Implications for Scale Development, *SPIE Defense and Security Symposium,* Orlando, 9-13 April.

Irvine, J.M., O'Brien, G., Miller, J., Israel, S.A., Fenimore, C., Roberts, J., Bartolucci, J., and Cannon, D.M. 2007c. User-oriented Evaluation of Compression for Motion Imagery, *SPIE Defense and Security Symposium,* Orlando, 9-13 April.

Le Meur, O., Ninassi, A., Le Callet, P., and Barba, D. 2010. Overt Visual Attention for Free-Viewing and Quality Assessment Tasks: Impact of the Regions of Interest on the Video Quality Metric, *Signal Processing: Image Communications,* 25: 5.

Leachtenauer, J. 1996. National Imagery Interpretability Rating Scale: Overveiw and Product Description, *Annual Meeting,* ASPRS, Baltimore, April.

Leachtenauer, J.C., Malila, W., Irvine, J.M., Colburn, L., and Salvaggio, N. 1997. General Image-Quality Equation: GIQE, *Applied Optics,* 36(32): 8322-8328.

Leachtenauer, J.C. and Driggers, R.G. 2001. *Surveillance and Reconnaissance Imaging Systems: Modeling and Performance*, Artech House, Inc., Boston.

Lubin, J. 1995. *A Visual Discrimination Model for Imaging System Design and Evaluation*. In: Menendez, A.R. and Peli, E. *Vision Models for Target Detection and Recognition*, World Scientific.

Maver, L.A., Erdman, C.D., and Riehl, K. 1995. Image Interpretability Rating Scale, *Digest of Technical Papers: International Symposium Society for Information Display,* Society for Information Display, XXVI: 117-120.

Moorthy, A.K., Seshadrinathan, K., Soundararajan, R., and Bovik, A.C. 2010. Wireless Video Quality Assessment: A Study of Subjective Scores and Objective Algorithms, *IEEE Transactions on Circuits and Systems for Video Technology,* 20(4): 587-599.

O'Brien, G., Israel, S.A., Irvine, J.M., Fenimore, C., Roberts, J., Brennan, M., Cannon, D.M., and Miller, J. 2007. Metrics to Estimate Image Quality in Compressed Video Sequences, *SPIE Defense and Security Symposium,* Orlando, 9-13 April.

Petitti, F., Young, D., Butto, R., Bakir, T., and Brennan, M. 2009. Video-NIIRS RP 0901, *http://www.gwg.nga.mil/misb/rrpubs.html*: 1-13.

Porter, R., Fraser, A.M., and Hush, D. 2010. Wide-Area Motion Imagery: Narrowing the Semantic Gap, *IEEE Signal Processing Magazine,* (September): 56-65.

Seshadrinathan, K., Soundararajan, R., Bovik, A.C., and Cormack, L.K. 2010. Study of Subjective and Objective Quality Assessment of Video, *IEEE Transactions on Image Processing,* 19(6): 1427-1441.

Sohn, H., Yoo, H., De Neve, W., Kim, C.S., and Ro, Y.M. 2010. Full-Reference Video Quality Metric for Fully Scalable and Mobile SVC Content, *IEEE Transactions on Broadcasting,* 56(3): 269-280.

Wang, Z., Bovik, A.C., Sheikh, H.R., and Simoncelli, E.P. 2004. Image Quality Assessment: Form Error Visibility to Structural Similarity, *IEEE Transactions on Image Processing,* 13(4): 600-612.

Watson, A.B., Hu, J., and McGowan III, J.F. 2001. DVQ: A Digital Video Quality Metric Based on Human Vision, *Journal of Electronic Imaging,* 10(1): 20-29.

Watson, A.B. and Kreslake, L. 2001. Measurement of Visual Impairment Scales for Digital Video, *Human Vision, Visual Processing, and Digital Display IX,* SPIE, San Jose, 4299: 11 pages.

Winkler, S. 2001. Visual Fidelity and Perceived Quality: Towards Comprehensive Metrics, *Human Vision and Electronic Imaging,* SPIE, San Jose, California, 21-26 January, 4299: 114-125.

Winkler, S., Sharma, A., and McNally, D. 2001. Perceptual Video Quality and Blockiness Metrics for Multimedia Streaming Applications, *International Symposium on Wireless Personal Multimedia Communications*: 547-552.

Xia, T., Mei, T., Hua, G., Zhang, Y.D., and Hua, X.S. 2010. Visual Quality Assessment for Web Videos, *Journal of Visual Communication and Image Representation,* 21(6): 826-837.

Yang, F., Wan, S., Xie, Q., and Wu, H.R. 2010. No-Reference Quality Assessment for Networked Video via Primary Analysis of Bit Stream, *IEEE Transactions on Circuits and Systems for Video Technology,* 20(11): 1544-1554.

Yasakethu, S.L.P., Fernando, W.A.C., Kamolrat, B., and Kondoz, A.M. 2009. Analyzing Perceptual Attributes of 3D Video, *IEEE Transactions on Consumer Electronics,* 55(2): 864-872.

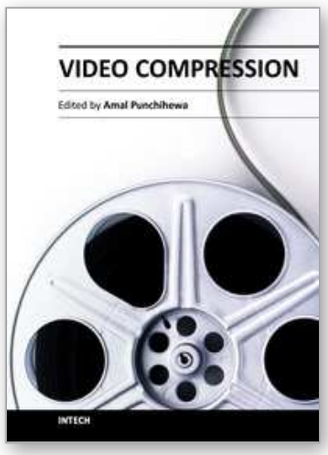
Young, D., Yen, J., Petitti, F., Bakir, T., Brennan, M., and Butto, R. 2009. Video National Imagery Interpretability Rating Scale Cirterial Survey Results, *Proceedings of the SPIE,* San Diego, 7307.

Young, D., Bakir, T., Butto, R., Duffield, C., and Petitti, F. 2010a. Loss of Interpretability Due to Compression Effects as Measured by the New Video NIIRS, *Proceedings of the SPIE*, 7529.

Young, D., Bakir, T., Butto, R., and Petitti, F. 2010b. Factors Related to Low Slant Angle Affecting Airborne Video Interpretability, *Proceedings of the SPIE,* San Diego, 7668.

IntechOpen

IntechOpen



Video Compression

Edited by Dr. Amal Punchihewa

ISBN 978-953-51-0422-3
Hard cover, 154 pages
Publisher InTech
Published online 23, March, 2012
Published in print edition March, 2012

Even though video compression has become a mature field, a lot of research is still ongoing. Indeed, as the quality of the compressed video for a given size or bit rate increases, so does users' level of expectations and their intolerance to artefacts. The development of compression technology has enabled number of applications; key applications in television broadcast field. Compression technology is the basis for digital television. The "Video Compression" book was written for scientists and development engineers. The aim of the book is to showcase the state of the art in the wider field of compression beyond encoder centric approach and to appreciate the need for video quality assurance. It covers compressive video coding, distributed video coding, motion estimation and video quality.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

John M. Irvine and Steven A. Israel (2012). Quantifying Interpretability Loss due to Image Compression, Video Compression, Dr. Amal Punchihewa (Ed.), ISBN: 978-953-51-0422-3, InTech, Available from: http://www.intechopen.com/books/video-compression/how-much-to-compress-quantifying-intrepretability-loss-from-compression

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen