THE EFFECTS OF VIDEO COMPRESSION ON ACCEPTABILITY OF IMAGES FOR MONITORING LIFE SCIENCES' EXPERIMENTS

Richard F. Haines and Sherry L. Chuang
Spacecraft Data Systems Research Branch
Ames Research Center - NASA
Moffett Field, CA 94035

1. Introduction

Current plans indicate that there will be a large number of life science experiments carried out during the thirty year-long mission of the Biological Flight Research Laboratory (BFRL) on board Space Station Freedom (SSF) [1]. Non-human life science experiments will be performed in the BFRL. Two distinct types of activities have already been identified for this facility: (1) collect, store, distribute, analyze and manage engineering and science data from the Habitats, Glovebox and Centrifuge, (2) perform a broad range of remote science activities in the Glovebox and Habitat chambers in conjunction with the remotely located principal investigator (PI). These activities require extensive video coverage, viewing and/or recording and distribution to video displays on board SSF and to the ground. This paper concentrates mainly on the second type of activity. Each of the two BFRL habitat racks are designed to be configurable for either six rodent habitats per rack, four plant habitats per rack, or a combination of the above. Two video cameras will be installed in each habitat with a spare attachment for a third camera when needed. Therefore, a video system that can accommodate up to 12-18 camera inputs per habitat rack must be considered.

The present Glovebox (GB) design provides an enclosed, bioisolated workspace in which a wide variety of non-human life sciences research can be conducted without contaminating the rest of the interior of SSF. Two dedicated video cameras are installed in the walls of the GB to provide orthogonal views of the activity in the work volume. These cameras are in addition to others installed in the attached Modular Habitats. The user will be able to select and display video data from the work volume, GB attached habitats, transmissions from the ground or other external video sources.

The centrifuge provides a selectable gravity environment (between .01g and 2g) for biological specimens. In addition, four rodent habitats can be accommodated on an inner ring of the centrifuge. Each of these habitats will also have two video cameras installed with a spare attachment for a third camera when needed. Therefore, from 24-36 camera outputs must be supported in the centrifuge system.

2. Problem and Approach

The number of individual cameras in the entire BFRL can range between 50-74 cameras (when all of the components of the BFRL are in use), each simultaneously collecting NTSC quality image data. In addition, there will be a limited transmission bandwidth that will (likely) be available between SSF and the ground. An approach to these problems may be through the use of video compression technologies.

3. Methodology and Results

In general, the research approach described elsewhere [3] was used here. Video images were viewed by a variety of Ss consisting of PIs and other professional personnel already familiar with the scene content. Each made judgments of the quality and acceptability of each compressed scene. The data were then subjected to statistical analyses in order to be able to extend the parametric findings beyond the present groups of participants. Three independent experiments were conducted, one involving static images of plants and two involving moving animals. Each is described in detail below.

3.1 Experiment 1: Still Image Compression of Three Plant Scenes

3.1.1 Procedure and Test Instructions. The subject (S) was told about the nature of the study, what he or she was supposed to do, and shown the apparatus and scoring sheet and rating criteria sheet which were posted nearby. There were three scenes per experiment and three primary subjective judgments made per scene. In the first experiment there were two additional judgments made on each scene. The method of pair-comparison [2] was employed whereby each of the four levels of image compression were presented in all possible pairs. These twelve image pairs were displayed side-by-side on a high resolution color monitor with the highest resolution image located on the right or left side of the screen on a random basis. Each image measured 6.75" x 6.75" (45.5 square inches).

The *first* subjective judgment required was to carefully inspect both images on the screen and select the one having the best overall quality to support the S in carrying out scientific research. The *second* subjective judgment to be made [only with respect to the screen image chosen (above) as having the best overall quality] was a numeric rating from 1 to 5 where: 1 = completely unacceptable image quality, 3 = average image quality, and 5 = maximally clear and acceptable image quality. The *third* subjective judgment required was to either accept or not accept the image chosen in terms of whether it would be acceptable to provide the kinds of answers to questions the S would normally ask of this particular image. The *fourth* subjective judgment to be made was to list which image characteristics were used in making the numeric rating of image quality such as color, contrast, brightness, resolution, etc. The *fifth* requirement was to circle those parts of the image (on a B&W copy of the screen image placed in front of the S) at which the S looked in order to make his or her judgments. This was done on a trial by trial basis for later analysis.

3.1.2. Apparatus. The apparatus consisted either of a Panasonic color CCD (model WV-CD 110A) camera (for scene 1), or a Toshiba color CCD (model IK-M30) microminiature camera yielding over 360 TV lines horizontally (for scenes 2 and 3) taken through a microscope. An image capture board ("Moonraker" by Workstation Technology Inc.) and Joint Photographic Expert Group (JPEG) standard compression board ("Picture Press" by Storm Technology, Inc.) were installed in a Macintosh II with 8 megabytes of RAM and a 1.04 Gigabyte hard disk. Software image control was accomplished using image manipulation software ("Photoshop" by Adobe, Inc.). The images were displayed on a 20 inch (diagonal), high resolution (1024 x 768 pixel) color monitor (Mirror Technologies, Inc.; model C/T 20HA, Rev. G).

3.1.3 Image Compression Details. Following are the four image compression levels investigated using the JPEG standard. The number by each level is referred to in the results section. 1 = Excellent; 2 = High; 3 = Good; and 4 = Fair. Table 1 presents selected information on the four levels of compression derived from the scenes which were studied here.

Table 1. Compression Details Associated With Test Scenes Studied Here

	Scene 1 "wheat stalk"				Scene 2 "wheat kernel cluster"				Scene 3 "magnified single wheat kernel"			
comp. level	Excel	High	Good	Fair	Excel	High	Good	Fair	Excel	High	Good	Fair
size												
bits/pix.												
comp. ratio												
comp. time (sec.)	1.60	0.80	0.78	0.60	1.58	0.80	0.78	0.78	1.55	0.90	0.78	0.76

Notes: Excellent settings uses 1:1:1 subsampling ratio
The range of compression ratios is between 5:1 to 7:1
High setting uses 2:1:1 subsampling ratio
The range of compression ratios is between 16:1 to 24:1
Good setting uses 2:1:1 subsampling ratio
The range of compression ratios is between 40:1 to 60:1
Fair setting uses 2:1:1 subsampling ratio
The range of compression ratios is between 80:1 to 120:1

The S sat with his or her eyes 32" (+/- 2") from the screen of the monitor so that the angular width of the two images subtended approximately 20 degrees arc. Each of the three compressed test scenes was presented with each of the others in all combinations in random order for a total of twelve pairs per S per scene. The order in which the three scenes was presented also was randomized.

3.1.4 Test Scenes. Three static color images were investigated: (1) a "wheat stalk" cluster scene, (2) "wheat kernel cluster" scene and, (3) a magnified "single wheat kernel."

3.1.5 Test Subjects. A total of ten volunteer Ss took part. Most were senior level NASA investigators, contractors, or visiting faculty working in such fields as plant physiology and biology, closed environment life support research and development, plant nutrition. Three were graduate students working at Ames in the area of plant growth dynamics for the SSF program.

3.1.6 Results. The results are presented in four sections, each of which deals with the subjective judgments that were made on each compressed image.

Image Accuracy Judgement. As expected, (1) the larger the difference in compression levels between the two paired images the greater was the accuracy of judgement. (2) Of the three pair-comparison conditions that were one compression level apart (viz., 1,2; 2,3; 3,4), only 28 percent were judged accurately. (3) Of the two pair-comparison conditions that were two compression levels apart (viz., 1,3 and 2,4), 42 percent were reliably correct. (4) Of the single pair-comparison that were three compression levels apart (viz., 1,4), 67 percent were reliably correct. (5) There were differences in the proportion of unreliable judgements due to the kind of scene presented, and (6) There was only a relatively small difference in these judgements due to whether the best image was located on the left or the right side of the screen as would be expected. Nine of the eighteen cells (50 percent) were only ten percent apart indicating a generally good level of agreement. Five of the 18 cells (28 percent) were twenty percent apart, two (11 percent) were thirty percent apart, one (5 percent) was forty percent apart, and one (5 percent) was identical. The single cell in which there was perfect agreement was at the 50 percent "guessing" level and is, therefore, not considered reliable.

Image Numeric Ratings. There were no statistically significant differences (by analysis of variance) among the four JPEG compression levels. The three scenes tested did produce a significant difference ($F = 8.25$; df=2; $p=0.009$) with scene 2 producing the largest subjective quality ratings for each compression level (mean = 3.70).

Image Acceptability Judgement. There were small (statistically insignificant) differences among these means. An analysis of variance showed that the four JPEG compression level means were not significantly different from one another but that the three scenes were ($F = 9.73$; df = 2; $p = 0.006$). These mean acceptability results parallel (in magnitude) the above mean image rating results across these four compression levels and three scenes.

Image Characteristics Selection Results. Each S was asked to indicate which particular image characteristics were used in making a judgment. It was found that resolution by itself was the single-most important image characteristic regardless of scene content. This was followed in frequency of occurrence by resolution, color and brightness/contrast combined.

3.1.7 Conclusions. These results suggest that there is no clearly perceptible difference in the ratings of image quality among any of the four JPEG compression levels studied here for any of the three scenes presented. There was a significant difference found due to the type of scene studied. The magnified image of scene 3 was significantly darker than were the other two and was difficult to discern the identity of specific tissues. It elicited the lowest mean rating of all three scenes across these four JPEG compression levels. For scenes that are clearly familiar to the viewer and possess sufficient resolution, brightness, and contrast a FAIR JPEG compression level (i.e., average 10 kilobytes/image) appears to be sufficient.

3.2 Experiment 2: Motion Image Compression of Three Rodent Scenes

3.2.1 Procedure and Test Instructions. Each S viewed fifteen second-long dynamic scenes four times, each generated at a different level of video compression. First they had to assign the scene a number from one to five indicating its image quality where 1 = completely unclear and unacceptable image quality, 3 = average image quality, and 5 = maximally clear and acceptable image quality. Their second judgment was to answer yes or no to the question of whether that scene would be acceptable to them in order for them to answer the kinds of questions they would normally ask of this particular image. Finally, using the image characteristic key that is posted to their left[1], to write down which specific image details led them to select the numeric rating (1 to 5) they chose."

Three rodent scenes were selected. Scene 1 consisted of general animal movement including a (**Jump**) . Scene 2 showed both rodents in **Play** activity where they rolled on top of each other and chased each another around the enclosure. The third scene was of a subtle "fall-over" behavior which was of interest to several Ss. In this **Fall-Over** scene both animals were generally sedentary.

[1] This sheet was located about 26" away at eye level.

3.2.2 Apparatus. The original rodent behavior imagery was recorded on a NASA-Ames' animal centrifuge using a Panasonic CCD (model WV-CD-110A) camera with 16mm fixed focus lens located outside the transparent animal holding cage. This signal was routed to a BetaCam recorder through a slip-ring assembly on the rotational axis of the centrifuge and thence to a SVHS recorder.

Scenes selected from the SVHS tape were compressed to four levels (384, 448, 768, 1540 kbits/sec.) using a *Compression Labs Inc.'s* "Rembrandt" model Codec and then re-recorded on a new SVHS tape in random order. The S never knew which compression level was being shown nor was he or she shown NTSC broadcast quality imagery prior to testing. The display apparatus consisted of SVHS color tapes (only) on which the compressed images were recorded, a SVHS tape recorder for playback, and a NEC (model PR-2000S) color TV monitor with 20" (diagonal) screen.The S sat with his or her eyes normal to the TV screen and 32" away.

3.2.3 Test Subjects. Twelve people took part as volunteer subjects. Eight were senior level staff who regularly worked with rodents in such disciplines as physiology, neuromuscular dynamics, behavioral interactions, etc. Four people were animal care technicians who were very familiar with the health and status characteristics of white rats.

3.2.4 Results. The results are presented in three sections, *image rating, image acceptability, image characteristics.*

Image Rating Results. The differences in mean image ratings related to variation of compression level were not statistically significant. However, as expected, the mean ratings among the three scenes waeresignificantly different ($F = 5.01$; $df = 2$; $p = 0.0025$). The mean ratings for scene 3 "fall over" approached significance ($p=0.06$). As was found in Experiment 1, the type of experimental situation that must be compressed/decompressed and visually analyzed plays a very significant role in the image rating.

Image Acceptability Results. The percentage of the twelve Ss who selected each scene as being acceptable tended to increase as bandwidth increased for all three scenes. These mean data suggest that the largest gain in image acceptance is between 448 and 768 kilobits per second.

Image Characteristics Selection Results. Resolution, motion, and brightness/contrast (combined) are the most frequently selected image characteristics across all three scenes. In addition color, by itself, was *never* selected as being an important characteristic of these images.

3.2.5 Conclusions. Experiment 2 has shown that these four compression levels did not yield a statistically significant difference in mean image ratings while the mean ratings for the three test scenes were significantly different. The largest difference in image acceptability across these four bandwidths occurred between 448 and 768 kilobits per second, averaging 36 percent increase in mean acceptance.

3.3 Experiment 3: Motion Image Compression of Three Primate Scenes

3.3.1 Procedure and Test Instructions. The same procedures and instructions were used as just described for Experiment 2 except that, (a) a mature squirrel monkey was the animal of interest, and (b) 576 kilobits/sec was substituted for 448 kilobits/second in order to provide an additional data point on possible curve-plots of the joint results. The three test scenes selected were as follows: Scene 1 showed a relatively close-up view with the entire face of the monkey filling the field of view. This scene is referred to as **Face**.

The second scene, **Rear of Head,** showed the rear of the animal's head with a hard plastic electrode cap attached and the restraining device in which the animal sat. Of particular experimental monitoring interest by the Ss was the color of the skin surrounding the cap, fur condition and color, and other apparatus detail. The third scene was a close-up of the eyes - it is referred to as **Eye Close-up**.

3.3.2 Apparatus. The hardware was identical to that used in Experiment 2 except that the CLI compression device was pre-set to 384, 576, 768, and 1,540 kilobits per second for the SVHS video tape made.

3.3.3 Test Subjects. Eleven volunteers took part (9 males, 2 females). All were NASA personnel or contractors working with these animals routinely and all were different from the Ss used in the first two studies.

3.3.4 Results. The results are presented in three sections, *image ratings, image acceptability, and image characteristics..*

Image Rating Results. Separate analyses of variance were conducted on the data from each scene. The means of the four bandwidths were significantly different for each of the three scenes: Scene 1 ($F = 9.36$; $df = 3$; $p = 0.0001$); Scene 2 ($F = 4.88$; $df = 3$; $p = 0.0055$); and Scene 3 ($F = 4.35$; $df = 3$; $p = 0.0096$). Another ANOVA incorporating all of the data found that the scene main effect was also highly significant ($F = 11.1$; $df = 2$); $p = 0.0001$). While the bandwidth main effect in this analysis was statistically significant it was not as large an effect as found in the individual scene analyses (all $p < 0.01$).

Image Acceptability Results. The percentage of the eleven subjects who selected each image as being acceptable increased regularly with increasing bandwidth as expected.

Image Characteristics Selection Results. The image characteristics which were cited as being important to each participant's judgments were rank ordered by frequency of occurrence. Resolution and color were considered to be the most important image characteristics across all three scenes. Neither color nor brightness/contrast was ever selected (except once) as being an important image characteristic.

3.3.5 Conclusions. Perhaps the most general conclusion that can be drawn from Experiment 3 is that for each of the three scenes selected, this range of bandwidths produced a clearly significant increase in judged image quality. The largest increase in image acceptability occurred between 384 and 576 kilobits per second, increasing from 45 to 79 percent (mean) acceptance level.

4. Discussion: General Overview of Findings.

The degree to which a still or moving image can be compressed and decompressed and remain acceptable and useful depends upon numerous hardware, software, and humanware "factors". With regard to motion image compression, the present study has shown that for the four dynamic (inter-frame compression) bandwidths examined here using a proprietary algorithm, mean image ratings increase reliably as compression level decreases. The results from the second experiment involving rodents may be compared directly with previous data (Haines and Jackson, 1990; fig. 1a) in which the same codec and compression levels were used as well as white rats within enclosures. When the slight difference in rating scale indices used in these two studies is taken into account it is seen that the mean ratings are almost identical. For example, the 384, 448, and 768 kilobits per second compression levels result in a slightly above average mean ratings of 3.5 while the 1,540 kilobits per second compression level yields a full point higher.

The Ss' recommendation on the type of scene that was evaluated also played a significant role in judged image quality. In addition to the selection of scenes based on representative experiments, scenes were purposefully selected which would sample all of the basic visual perception domains involved in anticipated future Space Station Freedom life science operations. Domains such as high and low visual resolution were included as well as a wide range of colors, brightness, and contrasts.

It is true that future Space Station Freedom life science procedures will probably differ from the rather passive animal monitoring and plant examinations carried out here. Nevertheless, if PIs must visually inspect in-space specimens from the ground using video compression hardware and software, the fundamental image features they will look at will be similar to those studied here. The kinds of specimens also may be different but the range of sizes of critical detail to be inspected cannot be much greater (without the use of high definition TV). Likewise, unless digital image processing techniques are used that involve pseudo-coloring, edge enhancement, etc., the range of image brightness and contrast cannot be much greater than what was presented here. Of course this is also true for other types of TV sensors such as low light or infra-red since it is the final display which the S looks at which determines the ultimate image contrast and brightness. It is for these reasons that the presented results are reasonably representative of those that would be found if other life science specimens (which possess different dynamic behavior) were substituted.

5. Conclusions

The JPEG standard was found to provide acceptable still-frame imagery of plants at compressions as high as 120:1, depending upon particular scene content. Resolution by itself was most important for the still-frame imagery followed by resolution combined with color or brightness/contrast. For moving imagery using an inverse cosine transform compression algorithm, a transmission bandwidth of about 768 kb/s or one-half of T-1 was found to provide very high acceptability for the mean of the three scenes in which camera imagery was colorful and showed high detail. Finally, the visual judgment criteria that were selected most often as being important for evaluating dynamic imagery was resolution, color, and image motion in some combination. The present testing methodology which involved individual Ss evaluating their own data was effective in evaluating video compression effects.

A wide array of local (in space) and remote (on the ground) visual judgments will be made on plant and animal specimens on board Space Station Freedom in the future. While this and other studies have shown that carefully selected video compression techniques provide an acceptable solution to transmission bandwidth constraints, the final quality of the remote television imagery that is achieved will be dependent upon complex, interrelated hardware, software (video architecture), and humanware factors. Advanced pre-flight simulations using representative flight end-to-end hardware should be conducted in order to optimize this imagery and related scientific procedures, particularly the role that can be played by infra-red imagery, switching and scheduling algorithms in order to optimize the use of available transmission bandwidths.

References

[1] *Space Exploration's Centrifuge Facility Project Communications Plan*, Communications and Network Development Branch, Ames Research Center, NASA, (Draft 2/27/90) March 1990.

[2] J. P. Guilford, "The Method of Pair Comparisons," in *Psychometric Methods,* McGraw-Hill, New York, 1954.

[3] R. F. Haines, "Human Performance Measurement-Validation Procedure Applicable to Advanced Manned Telescience Systems," *Research Institute for Advanced Computer Science, Technical Report No. 90.10*, February, 1990.

[4] R. F. Haines and R.W. Jackson, "Television Image Compression and Small Animal Remote Monitoring," *Research Institute for Advanced Computer Science, Technical Report No. 90.19*, April 1990.