IN-17
110793

P.18

NASA
Technical
Paper
3239

July 1992

The Effects of Video Compression on Acceptability of Images for Monitoring Life Sciences Experiments

Richard F. Haines
and Sherry L. Chuang

(NASA-TP-3239) THE EFFECTS OF VIDEO COMPRESSION ON ACCEPTABILITY OF IMAGES FOR MONITORING LIFE SCIENCES EXPERIMENTS (NASA) 18 p

N92-33933

Unclas

H1/17 0110793

NASA

NASA Technical Paper 3239

1992

The Effects of Video Compression on Acceptability of Images for Monitoring Life Sciences Experiments

Richard F. Haines
Foothill-DeAnza Community College District
Los Altos Hills, California

Sherry L. Chuang
Ames Research Center
Moffett Field, California

National Aeronautics and Space Administration
Office of Management
Scientific and Technical Information Program

SUMMARY

Future manned space operations for Space Station Freedom will call for a variety of carefully planned multimedia digital communications, including full-frame-rate color video, to support remote operations of scientific experiments. This paper presents the results of an investigation to determine if video compression is a viable solution to transmission bandwidth constraints. It reports on the impact of different levels of compression and associated calculational parameters on image acceptability to investigators in life-sciences research at Ames Research Center. Three nonhuman life-sciences disciplines (plant, rodent, and primate biology) were selected for this study. A total of 33 subjects viewed experimental scenes in their own scientific disciplines. Ten plant scientists viewed still images of wheat stalks at various stages of growth. Each image was compressed to four different compression levels using the Joint Photographic Expert Group (JPEG) standard algorithm, and the images were presented in random order. Twelve and eleven staffmembers viewed 30-sec videotaped segments showing small rodents and a small primate, respectively. Each segment was repeated at four different compression levels in random order using an inverse cosine transform (ICT) algorithm. Each viewer made a series of subjective image-quality ratings. There was a significant difference in image ratings according to the type of scene viewed within disciplines; thus ratings were scene dependent. Image (still and motion) acceptability does, in fact, vary according to compression level. The JPEG still-image-compression levels, even with the large range of 5:1 to 120:1 in this study, yielded equally high levels of acceptability. In contrast, the ICT algorithm for motion compression yielded a sharp decline in acceptability below 768 kb/sec. Therefore, if video compression is to be used as a solution for overcoming transmission bandwidth constraints, the effective management of the ratio and compression parameters according to scientific discipline and experiment type is critical to the success of remote experiments.

INTRODUCTION

A large number of life sciences experiments is planned for the 30-yr mission of the Biological Flight Research Laboratory (BFRL) on board Space Station Freedom (SSF) (Anon., 1990). Nonhuman life sciences experiments will be performed in the BFRL, which, for the purposes of this paper, consists of two Habitat Holding Units, a Service Unit Rack, a Life Sciences Glovebox Rack, and a 2.5-m-diameter centrifuge which will house the control experiments. Two distinct types of activities for this facility have been identified. The first type includes the collection, storage, distribution, analysis, and management of engineering and scientific data from the habitats, glovebox, and centrifuge. The second includes a broad range of remote experiments to be performed in the glovebox and habitat chambers in communication with the remotely located investigator. These remote activities require extensive video coverage, viewing, and/or recording and distribution to video displays on board SSF and on the ground. This paper concentrates on this second type of activity.

The study was performed using the payload architectural requirements of the BFRL. This facility will require an extensive video capability to permit remote monitoring of crew procedures and animal subject activity. Each of the two BFRL habitat racks is designed to be configurable for six rodent habitats or four plant habitats, or a combination of the two. Two cameras will be installed in each habitat, and there is a spare attachment for a third camera when needed. Therefore, a video system that can accommodate 12 to 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 nonhuman life sciences research can be conducted without contaminating the rest of the interior of SSF. Typical procedures to be performed in the GB may include administration of anesthesia or other animal restraint, collection of samples, dissection of specimens, and performance of video-based microscopy tasks. Two dedicated television 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 and the GB attached habitats, and transmissions from the ground or other external sources. Up to four simultaneous camera outputs must therefore be supported from the GB when an experimental procedure is in progress.

The centrifuge provides a selectable-gravity environment (between 0.01 g and 2 g) to house biological test subjects. Subjects exposed to constant acceleration equivalent to 1 g act as controls for the plants, rodents, squirrel monkeys, and other organisms subjected to the microgravity on board the BFRL. The centrifuge can hold at least six plant habitats, eight rodent habitats, six squirrel monkey habitats, or some combination of these on the perimeter ring. An additional four rodent habitats can be accommodated on the inner ring of the centrifuge. Each of these habitats will also have two video cameras installed, and a spare attachment for a third camera when needed. From 24 to 36 camera outputs must be supported in the centrifuge system.

Problem

The number of individual cameras in the entire BFRL ranges from 50 to 74 (when all components are in use), all simultaneously collecting NTSC quality image data. Thus quite a large amount of data could be transmitted by just one payload facility on SSF. In addition, there will be a limited-transmission bandwidth that will (likely) be available between SSF and the ground. The large quantity of video data alone raises many questions and concerns about operations and throughput on a networked data system.

Approach

One approach to these problems is through the use of video compression technologies. Video compression is a computerized method of eliminating "unneeded" information bits (taking into account the perceptual capabilities and limitations of the human visual system). Various algorithms incorporating subsampling, luminosity and chrominance frequency reductions, motion compensation, and other reduction methods may prove to be a means of reducing video file sizes on board the BFRL. See Haghiri and Denoyelle (1990) for a detailed discussion of such an approach.

In order to study the feasibility of video compression as a solution, however, we must understand the objectives and scientific judgment criteria of individual principal investigators (PIs) when viewing their video data. The PIs' criteria are even more important when we consider that currently there are no objective standards to reliably compare one level of compression with another, one compression algorithm (and its components) with another, or one recording-display system with another using the same compression approach. Nevertheless, there are numerous evaluative approaches available with which to reliably compare one multimedia system with another under operational conditions (Haines, 1990). We have employed an approach that involves the subject extensively in each step of this study: (1) interview the subjects and discuss with them the experiments that they perform in their own facilities; (2) tape the actual experiment that they have in progress with precalibrated recorders; (3) review the tapes with the subjects to select scenes that have scientific interest relevant to their physiological, neurovestibular, or behavioral research; (4) compress the selected images or video segments using precalibrated hardware; (5) present the compressed image or video segments to each subject in random order without comment other than instructions on the use of the rating forms; (6) have subjects rate the scenes according to their own scientific objectives; (7) conduct a post hoc interview concerning various subtle scientific details of the imagery. The PI-in-the-loop approach contributed to the accuracy and credibility of the results, and also differentiated our subjects' responses from those in past conventional survey-type studies.

A Typical Life Science Operations Scenario

It is helpful to have a general understanding of how a typical experiment might be carried out on SSF in a GB that includes multimedia support capability (i.e., digital computer, video cameras, voice channels). A prototype GB of this nature is described elsewhere (Haines and Jackson, 1990). Let us assume that the mission specialist's (MS's) job is to carry out a microscopic examination of a wheat stalk that has been growing in space within a plant-growing chamber for 40 days of its nominal 60-day growth cycle. The MS must do the following:

1. Adjust the focus, angle, pan, and zoom settings of the camera(s) to ensure that the field of view includes the part of the subject that is of interest.
2. Set the video compression level and transmission intervals, if an NTSC signal is not going to be used.
3. Set the video recording intervals, if recording is not continuous.
4. Establish the viewing reference with the ground (if the PI is involved in this particular procedure).
5. Obtain the plant modular habitat containing the wheat stalk specimen from the habitat rack.
6. Transport the habitat to the GB and attach it.
7. Transport a module containing laboratory equipment to the GB and attach it.
8. Insert arms and hands into the flexible gloves mounted on the transparent front of the GB.
9. Open the door to the module that contains the laboratory equipment and remove all necessary items.
10. Position all lab support equipment within the GB as desired.
11. Obtain the wheat stalk specimen from the habitat and visually inspect it for color, size, and other evidence of abnormalities, then cut it at its base.
12. Open the door to the plant habitat.
13. Place the wheat stalk on a clean glass microscope slide and position the slide on the microscope stage. Visually examine the image at a magnification of X15.
14. Make other TV camera image adjustments for best image color, resolution, and position in the field of view, and ensure that the various remote participants are receiving good images on the ground.
15. Carry out the required surgical procedures according to the experimental protocol (a checklist is on the computer screen inside the GB). If necessary, look at the

image transmitted from the ground-based PI's workstation camera. (Past training experiences and verbal instructions from the PI play an important role here.)

16. Label all test subjects or specimens appropriately.

17. Verify that all required steps have been successfully completed.

18. Remove all equipment from the GB by opening the equipment module door and replacing the equipment inside the equipment module.

19. Clean up the interior surfaces of the GB and stow all trash in labeled containers.

20. Remove arms and hands from the gloves and replace all remaining items in their proper storage places.

21. Announce to ground support crew that the experiment is completed.

Some of the above procedures (highly simplified here) can become very complex, yet all flight crewmembers and ground support personnel must still work effectively as a team. This is one of the goals of a well designed remote experimental system that involves video compression. But many questions arise. What will happen if the video imagery that is transmitted from SSF to the ground is compressed in such a manner that the ultrafine spatial detail or accurate color that is important for the PI's scientific judgment is lost? Will the remotely located PI be able to spot incorrect procedures or sample defects that would disqualify use of the sample? What level of video compression can be achieved before the PI's ability to make accurate scientific judgments based on transmitted images is impeded? The present study addressed these questions.

METHODOLOGY AND RESULTS

In general, the research approach used here was described in Haines and Jackson (1990). Video images were viewed by a variety of subjects including PIs and other professional personnel already familiar with the scene content. Each person made judgments of the quality and acceptability of each compressed scene, and the data were analyzed statistically. Three independent experiments were conducted, one using static images of plants and two using moving animals. The experiments are described below.

Experiment 1. Still-Image Compression of Three Plant Scenes

Procedure and test instructions– The subjects were told about the nature of the study and what they were expected to do, and were shown the apparatus and the scoring sheet and rating criteria sheet which were posted nearby. There were three scenes per experiment and three primary subjective judgments to be made per scene. In this first experiment there were two additional judgments to be made on each scene, as described below.

Guilford's (1954) method of pair comparisons was employed, in which each of four levels of image compression were presented in all possible pairs. The 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, in a random manner. Each image was 6.75 in. square (45.5 sq. in.). The following written instructions were presented:

> The study in which you are about to take part is designed to find out whether compression of video images to certain levels will influence your subjective judgments of their quality. You will be shown two screen images side by side for as long as you need to study them. No detailed information will be provided about the compression levels used. Here are the questions you will be asked about each pair of images:
>
> 1. Which image do you think has the best overall quality to support you in carrying out your scientific research?
>
>   Mark an **X** in the left or right column.
>
> 2. Next, refer to the *"Compressed Image Rating Details"* sheet posted to your left. Give the *numeric quality rating* to the image you just selected in question 1. Do *not* rate both images!
>
>   Insert this number in the numeric rating column.
>
> 3. Was the image you just selected acceptable to answer the kinds of questions you would normally ask of this particular image? Answer "yes" or "no."
>
>   Insert "Y" or "N" in the next column.
>
> 4. What specific image details led you to choose the image you selected? (Use the criteria symbols from the "Criteria Scoring Key" on the bottom of the "Compressed Image Rating Details" sheet).

The first subjective judgment required was to carefully inspect both images on the screen and select the one with the best overall quality to support the subject in carrying out scientific research. After this decision was made the other image was to be ignored.

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. Intermediate decimal values could be used if deemed necessary.

The third subjective judgment required was whether or not the image chosen would be acceptable to provide the kinds of answers to questions the subject would normally ask of this particular image. An image might be judged to have high overall quality and still not be acceptable because, for instance, the angle of video photography was not correct, some detail was not visible, the lighting was poor, the degree of magnification (zoom) was wrong, or some other such reason not directly related to image compression. However, we noted that the subject's written comments did not mention many of these kinds of factors. Subjects often cited low image brightness and contrast and poor resolution as reasons for not accepting an image.

The fourth subjective task was to list which image characteristics were used in making the numeric rating of image quality. The following image characteristics were provided (on a sheet of paper posted near the subject for ready reference): C_1 = color was of most importance; C_2 = color was important, but so were other features; B_1 = image brightness was the most important feature; B_2 = image brightness was important, but so were other features; R_1 = image resolution or sharpness was the most important feature; R_2 = image resolution was important, but so were other features; O = the other details that were important were _____ (subject inserted these as necessary). The C_1, B_1, and R_1 choices always stood alone: if any one was selected, no other characteristic(s) could be chosen. On the other hand, if C_2, B_2, and/or R_2 were selected, all other relevant characteristics were noted and the relative order of importance was also recorded.

The fifth and final requirement was to circle those parts of the image (on a black and white copy of the screen image) at which the subject had looked to make his or her judgments. This was done on a trial-by-trial basis for later analysis.

Apparatus– The apparatus, diagrammed in figure 1, 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 that yields a video image with over 360 TV lines horizontally (for scenes 2 and 3) when imaged through a microscope and then videotaped. An image capture board ("Moonraker," by Workstation Technology, Inc.) and a Joint Photographic Expert Group (JPEG) standard compression board ("Picture Press," by Storm Technology, Inc.) were installed in a Macintosh II with 8 mb of RAM and a 1.04-gb hard disk. Software image control was accomplished using image-manipulation software ("Photoshop," Adobe, Inc.). The images were displayed on a 20-in. (diagonal), high-resolution (1024- × 768-pixel) color monitor (Mirror Technologies, Inc., model C/T 20HA, Rev. G).

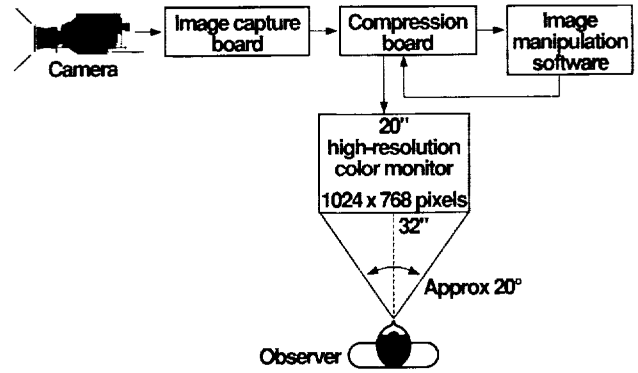


Figure 1. Image compression and viewing apparatus.

Image-compression details– The compression board used in the computer provided the JPEG standard encoding scheme in which groups of 8 by 8 pixels are processed as a unit. The processing includes subsampling, discrete cosine transfer, quantization matrix calculation, and Huffman encoding. This intraframe process (partially) removes image information that is not as likely to be perceived by human observers.

Table 1 shows the four image-compression levels investigated using the JPEG standard. The number shown for each level is referred to in the results section; 1 = Excellent, 2 = High, 3 = Good, and 4 = Fair. The table presents selected information on the four levels of compression derived from the scenes studied here, which are described later.

In this study, the scene details given above were constant. However, the JPEG standard algorithm may produce different values than those shown in table 1, depending upon the nature of the scene to be compressed. For instance, for scenes that are relatively monochromatic, significantly fewer bits per pixel (and consequently a smaller file size) are required; however, the compression time remains almost the same as in the table.

The subject sat with his or her eyes 32 in. (±2 in.) from the screen of the monitor so that the angular width of the two images subtended approximately a 20-deg arc. The subject was prevented from seeing the experimenter who controlled the compression generation procedures at a keyboard nearby. Another experimenter sat beside the subject to explain the procedures and ask and answer questions. All on-screen text information in the border of the image concerning image compression values was concealed. Screen brightness and contrast were adjusted to a midsetting (detent) and never varied. The room's ceiling fluorescent lights were left on at all times. There

Table 1. Compression details associated with test scenes

	Scene 1 "Wheat stalk"				Scene 2 "Wheat kernel cluster"				Scene 3 "Magnified single wheat kernel"			
Compression level	Exc. 1	High 2	Good 3	Fair 4	Exc. 1	High 2	Good 3	Fair 4	Exc. 1	High 2	Good 3	Fair 4
Size	157kb	56kb	23kb	11kb	133kb	47kb	21kb	11kb	121kb	39kb	15kb	7kb
Bits/pixel[a]	4.2	1.5	0.6	0.3	3.5	1.2	0.5	0.3	3.2	1.0	0.4	0.2
Subsampling ratio	1:1:1	2:1:1	2:1:1	2:1:1	1:1:1	2:1:1	2:1:1	2:1:1	1:1:1	2:1:1	2:1:1	2:1:1
Compression ratio	5:1	16:1	40:1	80:1	6:1	20:1	48:1	80:1	7:1	24:1	60:1	120:1
Compression time	1.60s	0.80s	0.78s	0.60s	1.58s	0.80s	0.78s	0.78s	1.55s	0.90s	0.78s	0.76s

[a]Pixels make up an image on the screen of a video display. Each pixel is generated by "n" bits of information, where a bit refers to a single digit (either a 0 or a 1). JPEG starts with 24 bits per pixel and compresses it to the "n" bits indicated here. The lower the "n," the more efficient the compression coding.

were no screen-face reflections of ceiling or other lights visible to the subject.

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 subject per scene.

Test scenes– Figure 2 shows a subject viewing the "wheat stalk" scene during the paired-comparison testing. Figure 3 shows the "wheat kernel cluster" and the "single wheat kernel," each under excellent and fair levels of compression.

Test subjects– A total of ten volunteer subjects, eight males and two females, took part. While most were senior-level NASA investigators, contractors, or visiting

Figure 2. *Subject viewing "wheat stalks" scene.*

faculty working in such fields as plant physiology and biology, closed-environment life support research and development, and plant nutrition, three were graduate students working at Ames in plant growth dynamics for the SSF program. As a group, these subjects were considered to be PIs.

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

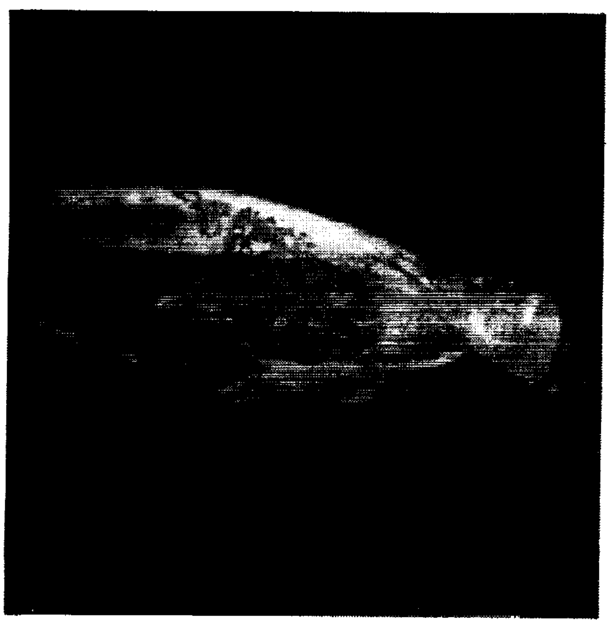
1. Image accuracy judgment: Since (1) the same level of compression was never compared against itself, and (2) the subject had to select which image in each pair possessed the *best* overall quality, a percentage correct of well over the 50-percent guess rate would be required to indicate accurate perception of each image pair. Table 2 shows, for each of the three scenes, the proportion of correct image judgments for each compression level. A cutoff value of 75 percent or higher was selected as the threshold for a "reliable" judgment. Boldface numbers indicate unreliable ("guessing") data according to this criterion. Each cell contains the results for the best image quality presented on the left (uppermost value) and on the right (lower value) of the screen.

Note that (1) as expected, the larger the difference in compression levels between the two paired images, the greater the accuracy of judgment; (2) for the three pair-comparison conditions that are one compression level apart (i.e., 1,2; 2,3; 3,4), only 28 percent were reliably rated, i.e., more than 75 percent of the subjects' ratings were correct; (3) for the two pair-comparison conditions that are two compression levels apart (i.e., 1,3 and 2,4), 42 percent were reliably rated; (4) of the single pair

(A) "Excellent"

(B) "Fair"

Scene 2 "wheat kernel cluster"



(A) "Excellent"

(B) "Fair"

Scene 3 "single magnified wheat kernel"

Figure 3. Examples of scenes 2 and 3 under two levels of compression.

Table 2. Percentage of subjects who selected correct image (plants)

	Compression levels by pairs*					
	1,2 2,1	1,3 3,1	1,4 4,1	2,3 3,2	2,4 4,2	3,4 4,3
Scene 1	**70**	80	90	**50**	100	**70**
"Wheat stalk"	80	**60**	80	**50**	**60**	90
Scene 2	**50**	**70**	100	80	100	80
"Kernel cluster"	**20**	**50**	80	**50**	90	90
Scene 3	**50**	**70**	**50**	**60**	**70**	**60**
"Single kernel"	**60**	**50**	**60**	**50**	80	**70**

*1, 2 refers to compression level 1 on the left, 2 on the right, etc.

comparison that was three compression levels apart (i.e., 1,4), 67 percent were reliably rated; (5) for over half of the image pairs, the percentage of correct judgments was below the established threshold for reliability; (6) there are differences in the proportion of unreliable judgments according to the kind of scene presented; and (7) there is only a relatively small difference in reliability according to whether the best image was located on the left or the right side of the screen, as would be expected. In nine of the eighteen cells (50 percent) the judgments were only ten percent apart, which indicates a generally good level of agreement. In five of the 18 cells (28 percent) the judgments were 20 percent apart, in two (11 percent) they were 30 percent apart, in one (5 percent) they were 40 percent apart, and in one (5 percent) they were identical. For the single cell in which there was perfect agreement, the judgments were at the 50 percent "guessing" level, and is therefore not considered reliable.

2. Image numeric ratings: After each subject chose which of each pair of images appeared to have the best overall quality, they assigned that image a number to indicate its relative image quality—from completely unacceptable image quality (1) to maximally clear and acceptable image quality (5). The mean results for each scene and four JPEG compression levels are given in table 3.

An analysis of variance on the above data showed no significant differences among the four JPEG compression levels, but there was a significant difference for the three scenes tested (F = 8.25; df = 2; p = 0.009), with scene 2 producing the highest subjective quality ratings for each compression level (mean = 3.70). These mean differences are considered to be highly significant and very likely the result of the influence of the visual features making up the scene. The present results may be extended to other scenes only to the extent that the new scenes correspond in general to the present ones in terms of their range of color and brightness, resolution, general judged usefulness, and identifiability of content.

It may be of value to note that for two of the three scenes (scenes 1 and 2), the Fair compression level produced the largest variance of ratings about the mean. This could become an important consideration if the total

Table 3. Mean image ratings (plants)

Scene #		JPEG compression level				
		Excellent 1	High 2	Good 3	Fair 4	Row mean
	N	47.0	32.0	34.0	9.0	
1	Mean	3.16	3.25	3.05	2.74	3.05
	S.D.	0.71	1.07	0.81	1.72	
	N	38.0	48.0	28.0	6.0	
2	Mean	3.49	3.56	3.68	4.08	3.70
	S.D.	0.71	0.67	0.60	1.02	
	N	34.0	35.0	29.0	17.0	
3	Mean	3.36	3.11	2.62	2.62	0.93
	S.D.	0.93	1.10	1.18	1.10	
	Column mean	3.34	3.31	3.12	3.15	3.23

number of separate viewing trials (of remotely located specimens) is limited, in which case the Excellent JPEG compression level would be desirable because of its relatively smaller response variance.

3. Image acceptability judgment: Mean acceptability of each of the four JPEG compression levels presented is shown in table 4. Analysis of variance showed that the differences among these means were not significant, but the means for the three scenes were ($F = 9.73$; $df = 2$; $p = 0.006$). These results parallel (in magnitude) the above mean image rating results across the four compression levels and three scenes.

4. Image characteristics selection: Each subject was asked to indicate which image characteristics were used in making a judgment. Color, brightness/contrast, and resolution were used in almost every combination. The subjects' replies were tallied and the characteristics were ranked from most frequently used to least frequently used for the three scenes. It was found (see table 5) that resolution 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.

Conclusions– There was no clearly perceptible difference in the ratings of image quality between any of the four JPEG compression levels studied here for any of the three scenes presented. There was a significant difference found between the scenes studied. The magnified image of scene 3 was significantly darker than the other two, and it was difficult to identify the specific tissues. Scene 3 elicited the lowest mean rating of all three scenes across the 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 kb/image) appears to be sufficient.

Table 4. Mean image acceptability (plants)

	JPEG compression level				
	Excellent	High	Good	Fair	Row mean
Scene 1 "Wheat stalk"	88	82	88	92	88
Scene 2 "Kernel cluster"	95	96	96	100	97
Scene 3 "Single kernel"	88	85	80	71	81
Column mean	90	88	88	88	89

Table 5. Frequency-ranked image characteristics (plants)

	Scene 1	Scene 2	Scene 3
Highest frequency	R_1	R_1	R_1
	R_2	R_2	B_2
↓	C_2	B_1	B_1
	B_2	B_2	R_2
	C_1	C_1	C_2
Lowest frequency	B_1	C_2	C_1

R = resolution, M = motion, B = brightness/contrast, and C = color. Meaning of subscript numbers was described earlier.

Experiment 2. Motion-Image Compression of Three Rodent Scenes

Procedure and test instructions– The subject read a printed instruction sheet and was told, "You will be shown a videotape showing two white rats in a small enclosure that has just returned to 1 g after two weeks on an Ames animal centrifuge rotating at 2 g. You will see three separate scenes. Each 15-second-long scene will be repeated four times, each time at a different level of video compression. Of course I can't tell you what compression level was used. I will stop the tape immediately after each scene so that you can make your ratings without haste. First, assign the scene a number from 1 to 5 indicating its image quality, where 1 = completely unclear and unacceptable image quality, 3 = average image quality, and 5 = maximally clean and acceptable image quality. You will find a copy of this numeric rating scale posted to your left for ready reference. You may use intermediate decimal values such as 4.4, 2.1, etc., if necessary. We realize that you have not seen these images or compression levels before so it will be difficult to judge what is average quality. What most people find useful is to use a number near the middle of the numeric scale until you have had an opportunity to see all compression levels, and then go back, if necessary, and modify your earlier judgments.[1] Your second judgment for each scene is to answer yes or no to the question of whether that scene would be acceptable to you in order for you to answer the kinds of questions you would normally ask of this particular image. Finally, using the image characteristic key that is posted to your left,[2] please write down which specific image details led you to select the numeric rating (1 to 5) you chose."

Three scenes were selected from the original videotape. Scene 1 consisted of general animal movement inside the enclosure where one rodent jumps away from the camera to the far left corner of the enclosure and then walks to the far right corner behind the other animal. This "Jump" scene was felt to be important from a neurovestibular dynamics point of view. Scene 2 showed both rodents in highly dynamic "Play" activity in which they rolled on top of each other and chased each other around the enclosure. This play scene was selected because the motion compression algorithm showed image blurring at some compression levels. The third scene was of a subtle "Fall-over" behavior which was of interest to several subjects. In this fall over scene both animals were generally sedentary, and were located in the center and far right corner of the enclosure. At one point one animal moves to the center of the enclosure where the second animal is standing. The second animal suddenly falls over for no apparent reason. This scene was of interest because of the possibility that it illustrates changes in neuromuscular control resulting from the prolonged period of adaptation to 2-g acceleration. Of course, no commentary or explanation concerning any of the three scenes was given to the subject before or during data collection. In addition, the three scenes provided a wide variety of image brightness and animal head-region (eyes, nose, mouth) and coat-coloring details as well as different types of body motions ranging from high-frequency (limb) scratching behavior to (whole body) running and jumping.

[1] It was found that only one subject modified earlier numeric judgments. Most subjects started at the middle of the scale to permit them maximal latitude in changing their ratings as they saw all the compression levels.

[2] This sheet was located about 26 in. away at eye level.

Apparatus– The original rodent behavior imagery was recorded on the NASA centrifuge using a Panasonic CCD (model WV-CD-110A) camera with 16-mm fixed-focus lens located outside the transparent animal cage. This signal was routed to a betacam recorder through a slip-ring assembly on the rotational axis of the centrifuge. The betacam medium was chosen as a reference recording medium because of its ability, in composite mode with a high-quality CCD camera, to record about 400 lines of imagery. The recorded betacam tape was then re-recorded onto both super VHS (SVHS) and standard VHS tape for later use.

Scenes selected from the SVHS tape were compressed to four levels (384, 448, 768, and 1540 kb/sec) using a Compression Labs, Inc., "Rembrandt" model codec and then re-recorded on a new SVHS tape in random order. This hardware employs a proprietary inverse cosine transform (ICT) compression algorithm. A brief scene label was also inserted before each scene, numbered one through four by scene and compression level. In order to further randomize scene presentation order, two separate tapes were made with the scenes in different order, one or the other of which was presented to different subjects. The subjects did not know which compression level was being shown, nor were they shown NTSC broadcast-quality imagery before testing.

The display apparatus consisted of SVHS color tapes (only), on which the compressed images were recorded; an SVHS tape recorder for playback; and an NEC (model PR-2000S) color TV monitor with 20-in. (diagonal) screen. The subjects sat with their eyes 32 in. from the TV screen and with their line of sight normal to the screen.

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, or animal behavior. Four people were animal-care technicians who were familiar with the health and status characteristics of white rats. It should be emphasized that the participants rated each scene from the point of view of their own discipline, each looking for somewhat different features of behavior and appearance. This fact was expected to contribute to the intersubject response variability. Table 6 provides selected information about the subjects, including their stated areas of interest in the imagery.

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

1. Image ratings: The mean numeric ratings for each bandwidth and scene are presented in table 7.

These means were subjected to analysis of variance. It was found that the differences related to variation of compression level were not statistically significant. However, as expected, the mean ratings among the three

Table 6. Selected information about participants—Experiment 2

Subject no.	Sex	Position	Stated interests
1.	F	NASA PI	"Precise motions during jumping, all body movements, timing of limb movements."
2.	F	Lab assistant	"General behavior and alertness signs, health and status of animals."
3.	M	NASA PI	"Stride dynamics (duration, distance), stance, forelimb/body placements, limb extension dynamics."
4.	M	NASA PI	"Motion and clarity of image, smoothness of motion."
5.	M	NASA PI	"Hair/coat quality, cleanness of eyes, lack of signs of poor health, alertness."
6.	M	NASA PI	"Clarity during motion, image sharpness of small (limb/head) motions, limb position, stability of behavior, grooming behavior."
7.	M	Physiologist	"General health, motion of animals."
8.	M	NASA PI	"Overall health signs (coat, cleanliness, etc.), excitation level, signs of bleeding, feeding and grooming behavior."
9.	F	Animal care technician	"Head movement, eyes/ears, details of health, overall activity and condition of cage bedding, drinking behavior."
10.	F	Animal care technician	"Coat condition, activity, exudate around eyes, social interaction (play, fighting), alertness of animals."
11.	M	Animal care technician	"Quality of coat, abnormal discharge from eye region, normal activity, playing or fighting behavior, facial features."
12.	M	Veterinarian	"Signs of alertness, responsiveness, color of coat, facial motions and small details."

Table 7. Mean image ratings (rodents)

		Bandwidth (kb/sec)				
		384	448	768	1540	Row mean
Scene 1	Mean	2.33	2.42	2.67	3.04	2.62
"Jump"	S.D.	0.78	0.67	0.91	0.72	
Scene 2	Mean	2.39	2.58	2.64	3.18	2.70
"Play"	S.D.	0.77	1.06	1.03	1.02	
Scene 3	Mean	1.58	1.95	2.36	2.42	2.08
"Fall over"	S.D.	0.85	0.97	0.70	0.82	
Column mean		2.1	2.32	2.56	2.88	2.47

scenes were significantly different (F = 5.01; df = 2; p = 0.0025). Differences in 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 or decompressed and visually analyzed plays a very significant role in the image rating. Whereas the participants in Experiment 1 were all involved in plant research and those in Experiment 2 were in rodent-related disciplines, the magnitude of the differences in the scenes detected by each group illustrate the fact that there are intradiscipline image-evaluation differences.

2. Image acceptability: As expected, the percentage of the twelve participants who rated each scene as acceptable tended to increase as bandwidth increased for all three scenes, as shown in table 8. These findings must be qualified because of the different requirements that each viewer places upon each scene. Whereas a vestibular physiologist may be interested in subtle indications of neuromuscular coordination, an animal care technician might be more interested in the appearance of the animal's coat or exudate accumulation around the eyes. Each scene contained a wide variety of detail from which the subject extracted relevant information.

Table 8. Mean image acceptability (rodents)

	Bandwidth (kb/sec)				
	384	448	768	1540	Row mean
Scene 1 "Jumping"	58	42	92	92	71
Scene 2 "Playing"	58	42	67	83	62.5
Scene 3 "Fall over"	17	33	67	75	48
Column mean	44.3	39	75.3	83.3	60.5

These data suggest that the largest gain in acceptance is between 448 and 768 kb/sec. Whether or not the mean acceptance level at 768 kb/sec is adequate depends upon the individual experimental situation (probably including scientific discipline, level of expertise, and other subjective factors).

3. Image characteristics selection: Table 9 presents the rank-ordered image characteristics from highest to lowest frequency of occurrence for each of the three scenes.

Table 9. Frequency-ranked image characteristics (rodents)

	Scene 1	Scene 2	Scene 3
Highest frequency	R_2	R_2	R_2
	M_2	M_2	B_2
↓	B_2	B_2	M_2
	C_2	C_2	C_2
	R_1	B_1	R_1
Lowest frequency	B_1	M_1	B_1
		R_1	

(Note: Meaning of symbols was given earlier.)

In table 9 it can be seen that resolution, motion, and brightness/contrast (combined) are the most frequently selected image characteristics across all three scenes. Color by itself was never selected as an important characteristic.

Conclusions– The four compression levels did not yield a significant difference in mean image ratings, whereas the mean ratings for the three test scenes were significantly different. The largest difference in image acceptability across the four bandwidths occurred between 448 and 768 kb/sec, averaging 36 percent increase in mean acceptance (cf. column mean).

Experiment 3. Motion Image Compression of Three Primate Scenes

Procedure and test instructions– The same procedures and instructions were used as those described for Experiment 2, except that (1) a mature squirrel monkey was the animal subject, and (2) 576 kb/sec was substituted for 448 kb/sec in order to provide an additional data point on possible curve plots of the combined results. The three test scenes selected were as follows. Scene 1 was a moderately close-up view with the entire face of the monkey filling the field of view. The animal was fed a raisin during the 15-sec scene. This scene showed chewing behavior, eye movements, and very fine detail of hair of varying colors. Another useful dynamic image feature was the relatively high-angular-velocity head movements (estimated at 100–200 deg/sec). 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 interest to the human subjects were the color of the skin surrounding the cap, fur color and matting, and the experimental apparatus. The last 5 sec of this scene consisted of a slow zoom out to show the entire body of the animal and the restraining enclosure.

The third scene was a close-up of the eyes. Referred to as "Eye Close-Up," this scene produced a high-resolution, high-color image of the animal's right eye, which filled the screen, then the animal's left eye, which also filled the screen, and finally, both eyes (and forehead) together. Of importance to the subjects was eye coloring, evidence of tearing or other exudate around the eyes, fur matting, binocular muscular coordination, and other fine details. Each scene was about 15 sec long.

Apparatus– The hardware was identical to that used in Experiment 2 except that the CLI compression device was preset to 384, 576, 768, and 1,540 kb/sec for each of the two SVHS video test tapes made.

Test subjects– Eleven volunteers took part (9 males, 2 females). Table 10 lists the stated interests of each participant.

Results– As for Experiment 2, the results are presented in three sections: *image ratings, image acceptability,* and *image characteristics.*

1. Image ratings: The mean numeric ratings for each bandwidth and scene are presented in table 11.

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 showed 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 $p < 0.01$).

2. Image acceptability: The percentage of the eleven subjects who rated each image as being acceptable

Table 10. Selected information on participants—Experiment 3

Subject no.	Sex	Position	Stated interests
1.	M	NASA PI	"Head movement, feeding, alertness, fine detail, skin color."
2.	M	NASA PI	"Eye and head movement, eyeball motion and spatial details, implant area condition."
3.	M	Physiologist	"Animal well-being, health, instrumentation status, color and eye definition."
4.	M	NASA PI	"General feeding behavior, fur and eye region condition, implant condition, facial skin condition."
5.	M	Physiologist	"Clarity/resolution of the eyes and body movements."
6.	F	Primate trainer/ technologist	"Condition of the skin of face, level of alertness and activity, eye clarity, electrode and headcap status, scabs."
7.	M	Animal care technician	"Abnormal (eye) discharges, head/eye movement(s), general health and status, electrode/skin separation status."
8.	M	Veterinarian	"Level of alertness/responsiveness and health, ocular movements, pupillary dynamics, eyelid/conjunctival coloring."
9.	M	NASA PI	"Head/eye movements, visibility and color of shaved head, implant status, and image sharpness."
10.	M	Veterinarian	"Anatomical and physiological details, pathological abnormalities, skin condition, eye details and movements."
11.	F	NASA PI	"Structural abnormalities, edema from incisions, scleral spots, implant status."

Table 11. Mean image ratings (primates)

		Compression level (kb/sec)				
		384	576	768	1540	Row mean
Scene 1 "Face"	Mean	2.59	3.35	3.70	3.71	3.34
	S.D.	0.86	0.61	0.77	0.72	
Scene 2 "Rear of head"	Mean	2.27	3.28	3.53	3.73	3.20
	S.D.	1.21	0.83	0.88	0.90	
Scene 3 "Eye close-up"	Mean	2.32	3.05	3.45	3.64	3.12
	S.D.	1.03	0.85	1.04	0.77	
Column mean		2.39	3.23	3.56	3.69	3.22

increased regularly with increasing bandwidth, as expected (see table 12).

3. Image characteristics selection: The image characteristics cited as being important to each participant's judgments were rank ordered by frequency of occurrence. The results are presented in table 13.

Resolution and color were considered to be the most important image characteristics across all three scenes. Neither color nor brightness/contrast was selected as being an important image characteristic by itself.

Conclusions– The most significant observation that can be made from Experiment 3 is that for each of the three scenes, there was a clearly significant increase in judged image quality with increasing bandwidth. The largest increase in image acceptability occurred between 384 and 576 kb/sec: acceptability increased from 45 to 79 percent (mean) acceptance level (cf. column mean).

Table 12. Mean image acceptability (primates)

	Bandwidth (kb/sec)				
	384	576	768	1540	Row mean
Scene 1 "Face"	73	100	100	100	93.3
Scene 2 "Rear of head"	18	73	82	82	63.8
Scene 3 "Eyes close-up"	45	64	100	100	77.3
Column mean	45.3	79	94	94	78.1

Table 13. Frequency-ranked image characteristics (primates)

	Bandwidth (kb/sec)		
	Scene 1	Scene 2	Scene 3
Highest frequency	R_2	C_2	R_2
↓	C_2	R_2	C_2
	B_2	B_2	M_2
	M_2	M_2	B_2
Lowest frequency	M_1	R_1	R_1

DISCUSSION

The degree to which a still or moving image can be compressed or decompressed and remain acceptable and useful depends upon numerous hardware, software, and human "factors." With regard to motion image compression, the present study has shown that for the four dynamic 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. 1(a)) in which the same codec and compression levels were used as well as the same type of white rats within enclosures. When the slight difference in rating scale indices used in these two studies is taken into account, the mean ratings are seen to be almost identical. For example, the 384, 448, and 768 kb/sec compression levels result in a slightly above-average mean rating of 3.5, whereas the 1,540 kb/sec compression level resulted in a mean rating a full point higher.

A comparison of the mean acceptance data from Experiment 2 with that from Haines and Jackson (1990) (fig. 2(a)) shows that the present mean image acceptability increases significantly with a decrease in compression, whereas in the previous study it did not. This difference is probably due to a difference in rating instructions used in the two studies. In the present study the subject was instructed to accept or not accept the scene according to whether the scene answered the kinds of questions he or she would typically ask when viewing this particular scene. In the earlier study they accepted or rejected the scene according to whether they were able to judge the overall health and status of the animal. Lower mean acceptance levels across the compression levels would be expected from the earlier study because of the highly constrained judging criteria that were to be used. In effect, all participants had to look for far fewer behavioral details in the earlier study, while in the present study the subjects were free to accept or reject the scene on whatever scientific criteria they chose. This resulted in a significant relationship between image acceptance level and compression level.

We believe that these results can be extended to other experimental procedures and subjects. One of the reasons for drawing subjects at random from the larger pool of subjects is to improve the probability that the experimental results will be representative. To the extent that the people tested here were representative of all PIs, we believe this objective was met. Except that a small percentage of the subjects were students or technicians, we do not know of any special exclusionary factors that would suggest that they are nonrepresentative.

The fact that the subjects recommended the types of scenes that were evaluated also played a significant role in image quality assessment. In addition to the selection of scenes based on representative experiments, scenes were intentionally selected to sample all of the basic visual perception domains involved in anticipated Space Station Freedom life sciences operations. Domains such as high and low visual resolution were included as well as a wide range of colors, brightness, and contrast.

It is true that future Space Station Freedom life sciences 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 features they will look at will be similar to those studied here. The kinds of specimens may also be different, but the size range 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 pseudocoloring, 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 infrared since it is the final display that the subject looks at that determines the ultimate image contrast. For these reasons, the results presented here are reasonably representative of those that would be found if other life sciences specimens (which exhibit different dynamic behavior) were substituted.

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 the most important image characteristic for the still-frame imagery, followed by resolution combined with color or brightness/contrast. For moving imagery using a discrete cosine transform compression algorithm, a transmission bandwidth of about 768 kb/sec or one-half of T-1 was found to provide high mean acceptability for 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 were resolution, color, and image motion, in some combination. The present testing methodology, which involved individual subjects 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 depend upon complex, interrelated hardware, software (video architecture), and human factors. In order to optimize this imagery and related scientific procedures, advanced simulations using representative flight end-to-end hardware should be conducted, and studies should be done of the role of infrared imagery, and of switching and scheduling algorithms in order to optimize the use of available transmission bandwidths.

REFERENCES

Anon., Space Exploration's Centrifuge Facility Project Communications Plan, "Draft 2/27/90." Communications and Network Development Branch, NASA Ames Research Center, 1990.

Guilford, J. P.: The Method of Pair Comparisons, *Psychometric Methods,* McGraw-Hill, New York, 1954, pp. 154-177.

Haghiri, M.; and Denoyelle, P.: A Low Bit Rate Coding Algorithm for Full Motion Video Signal. *Signal Processing: Image Communication*, vol. 2, Aug. 1990, pp. 187-199,

Haines, R. F.: Human Performance Measurement-Validation Procedures Applicable to Advanced Manned Telescience Systems. NASA CR-185447, 1990, and RIACS TR-90-10, 1990.

Haines, R. F.; and Jackson, R. W.: Television Image Compression and Small Animal Remote Monitoring. NASA CR-186614, 1990, and RIACS TR-90-19, 1990.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1992	Technical Paper

4. TITLE AND SUBTITLE

The Effects of Video Compression on Acceptability of Images for Monitoring Life Sciences Experiments

5. FUNDING NUMBERS

476-14-03

6. AUTHOR(S)

Richard F. Haines (Foothill-DeAnza Community College District, Los Altos Hills, CA 94022-4599) and Sherry L. Chuang

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Ames Research Center
Moffett Field, CA 94035-1000

8. PERFORMING ORGANIZATION REPORT NUMBER

A-92040

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

NASA TP-3239

11. SUPPLEMENTARY NOTES Point of Contact: Sherry L. Chuang, Ames Research Center, MS 269-4, Moffett Field, CA 94035-1000; (415) 604-3376 or FTS 464-3376

The material in this report was presented at the IEEE Computer Society Data Compression Conference, March 24–26, 1992, Snowbird, Utah.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified — Unlimited
Subject Category 17

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)* Future manned space operations for Space Station Freedom will call for a variety of carefully planned multimedia digital communications, including full-frame-rate color video, to support remote operations of scientific experiments. This paper presents the results of an investigation to determine if video compression is a viable solution to transmission bandwidth constraints. It reports on the impact of different levels of compression and associated calculational parameters on image acceptability to investigators in life-sciences research at Ames Research Center. Three nonhuman life-sciences disciplines (plant, rodent, and primate biology) were selected for this study. A total of 33 subjects viewed experimental scenes in their own scientific disciplines. Ten plant scientists viewed still images of wheat stalks at various stages of growth. Each image was compressed to four different compression levels using the Joint Photographic Expert Group (JPEG) standard algorithm, and the images were presented in random order. Twelve and eleven staffmembers viewed 30-sec videotaped segments showing small rodents and a small primate, respectively. Each segment was repeated at four different compression levels in random order using an inverse cosine transform (ICT) algorithm. Each viewer made a series of subjective image-quality ratings. There was a significant difference in image ratings according to the type of scene viewed within disciplines; thus ratings were scene dependent. Image (still and motion) acceptability does, in fact, vary according to compression level. The JPEG still-image-compression levels, even with the large range of 5:1 to 120:1 in this study, yielded equally high levels of acceptability. In contrast, the ICT algorithm for motion compression yielded a sharp decline in acceptability below 768 kb/sec. Therefore, if video compression is to be used as a solution for overcoming transmission bandwidth constraints, the effective management of the ratio and compression parameters according to scientific discipline and experiment type is critical to the success of remote experiments.

14. SUBJECT TERMS

Video compression, Compression acceptability, Life sciences monitoring

15. NUMBER OF PAGES 16

16. PRICE CODE A03

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified		

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

NASA-Langley, 1992