Perceptual study of the impact of varying frame rate on motion imagery interpretability

Charles Fenimore, John Irvine, David Cannon, John Roberts, Ivelisse Aviles, Steven Israel, Michelle Brennan, Larry Simon, James Miller, Donna Haverkamp, Paul F. Tighe, Michael Gross (CF, JR, IA; NIST, Gaithersburg MD, 20899-8940) (JI, DC, SI, LS, JM, DH; SAIC, Burlington, MA, 01803) (MB, Moriarty; PFT, MG, Booz, Allen, and Hamilton; Reston VA 20191)

ABSTRACT

The development of a motion imagery (MI) quality scale, akin to the National Image Interpretibility Rating Scale (NIIRS) for still imagery, would have great value to designers and users of surveillance and other MI systems. A multi-phase study has adopted a perceptual approach to identifying the main MI attributes that affect interpretibility. The current perceptual study measured frame rate effects for simple motion imagery interpretation tasks of detecting and identifying a known target. By using synthetic imagery, there was full control of the contrast and speed of moving objects, motion complexity, the number of confusers, and the noise structure. To explore the detectibility threshold, the contrast between the darker moving objects and the background was set at 5%, 2%, and 1%. Nine viewers were to detect or identify a moving synthetic “bug” in each of 288 10-second clip. We found that frame rate, contrast, and confusers had a statistically significant effect on image interpretibility (at the 95% level), while the speed and background showed no significant effect. Generally, there was a significant loss in correct detection and identification for frame rates below 10 F/s. Increasing the contrast improved detection and at high contrast, confusers did not affect detection. Confusers reduced detection of higher speed objects. Higher speed improved detection, but complicated identification, although this effect was small. Higher speed made detection harder at 1 Frame/s, but improved detection at 30 F/s. The low loss of quality at moderately lower frame rates may have implications for bandwidth limited systems. A study is underway to confirm, with live action imagery, the results reported here with synthetic.

Keywords: colored noise, confusers, frame rate, imagery interpretability, motion imagery, motion complexity, NIIRS, quality scale, speed, target

1. INTRODUCTION

The National Imagery Interpretability Rating Scale (NIIRS) has been embraced by the intelligence community for quantifying the interpretibility of still imagery[1, 2, 3]. Each NIIRS level indicates the types of exploitation tasks an image can support based on the judgments of experienced analysts. Development of a NIIRS for a specific imaging modality rests on a perception-based approach[4]. Accurate methods for predicting NIIRS from the sensor parameters and image acquisition conditions have been developed empirically and substantially increase the utility of NIIRS[2, 5]. In exploring avenues for development of a similar metric for motion imagery, a clearer understanding of the factors that affect the perceived quality of motion imagery is needed. An initial study explored the relationship between perceived quality of motion imagery and scene content (such as target motion)[6]. It was found that target motion had a significant effect on perceived image quality. Motion imagery clips in which the targets were moving, were consistently rated higher. This result is not surprising, since motion increases target salience.

The effects of target motion and camera motion, which were not factors for still imagery, are expected to influence user perceptions of image interpretability for motion imagery. Although scene complexity does not appear to be a major factor affecting perceived quality of still imagery, it could be important for motion imagery. In particular, an interaction between target motion and scene complexity has been hypothesized for motion imagery. Platform motion changes the camera position, affecting obscuration, masking, and perception of three-dimensional information.

The goal of this project was to develop an understanding of motion imagery quality by isolating these fundamental issues affecting perceived quality. The findings of the initial evaluation provided a first step in developing a quality metric for motion imagery. Based on the findings of that evaluation, a series of additional, focused evaluations were conducted to investigate other factors that could affect image quality. The first of these was a study of color and the interaction between color and motion. Two further studies addressed the effects of user perceptions and the frame rate (the current study) on the user’s ability to perform specific image exploitation tasks.

Human Vision and Electronic Imaging XI, edited by Bernice E. Rogowitz, Thrasyvoulos N. Pappas, Scott J. Daly, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6057, 60570P, © 2006 SPIE-IS&T · 0277-786X/06/$15

The frame rate study focuses on the quality effects of acquiring imagery at less than 30 frames per second. Reasons for considering lower frame rates include the reduction in buffer size and data storage, improved quality of single frames associated with longer integration times, and reduced burdens for data transmission. The effects of reduced frame rates are closely coupled with the effects of motion and the types of activity to be observed. For periodic activity, the Nyquist principle applies, so the frame rate must exceed twice the periodicity to offer any hope of observing the activity. In general, the ability to observe relevant activity at reduced frame rates depends on the type and complexity of the target and camera motion and the effects of target obscuration.

The plan for investigating frame rate effects was to conduct the present initial investigations with synthetic imagery that would permit close control of the primary factors. The findings gleaned from the study with synthetic imagery are to be validated with a follow-on evaluation using camera-capture imagery. For synthetic imagery the exploitation task to be performed was to detect a target (either stationary or moving), when confusers were also in the scene. For real imagery the tasks may be expanded to include identification, tracking, and other higher level forms of imagery interpretation.

In this report of the synthetic frame rate study, we describe our approach and the design elements. These include the development of the evaluation clip set in which we varied several factors in addition to frame rate, including target to background contrast. The perceptual character of the study required care in providing analysts with optimal viewing conditions as described in the Study Design section. We found a significant effect of frame rate on interpretability (Analysis section), particularly below 10 frames per second. The complexity of the study permitted us to identify significant effects of contrast and target motion. Some detection tasks for moving objects may be similar to the detection of a flickering stationary target[7], particularly for our slower moving targets.

2. APPROACH

The frame rate evaluation used imagery analysts to address fundamental issues of the impact of frame rate on interpretability for motion imagery. Imagery analysts were asked to determine whether a target sprite was located on the right-hand-side or on the left-hand-side at the end of a 10-second motion image clip. In addition, analysts were asked to rate their level of confidence for each response. Analysis of these ratings provided estimates of the statistically significant effects.

Fundamental perceptual task for viewers

The frame rate evaluation used a panel of 9 viewers. The subjects were recruited from the staff of NIST and from government imagery analysts, each of whom had normal visual acuity (possibly with correction). All imagery was in grey scale so it was unnecessary to evaluate for color blindness. The authors were not subjects. This mix of experienced and naïve viewers was acceptable because of the simple character of the interpretation tasks - to view 288 motion imagery (MI) clips and at the end of each clip to detect or identify the location of a particular moving synthetic *target* sprite. In asking the viewer/analyst to identify the target location, the study employed a forced choice method. In the absence of confusers, the task was detection; with confusers the task was identification. A typical frame from one of the clips is shown in Figure 1.

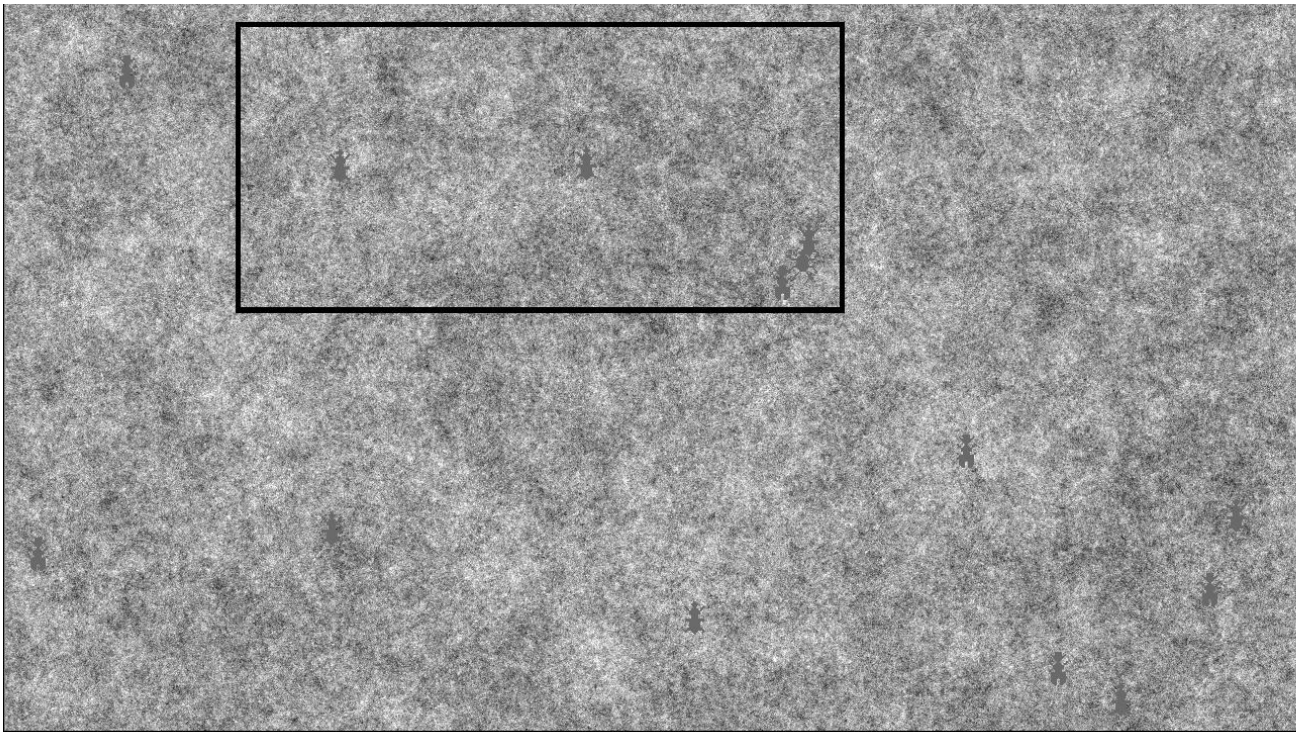

Figure 1 A single frame from a clip used in the Frame Rate Study. in this clip, a total of 13 confusers are present. The boxed area is shown at full resolution in Figure 2. The contrast between the sprites and the background is enhanced for presentation. The background varies from clip to clip; it is static within each clip.

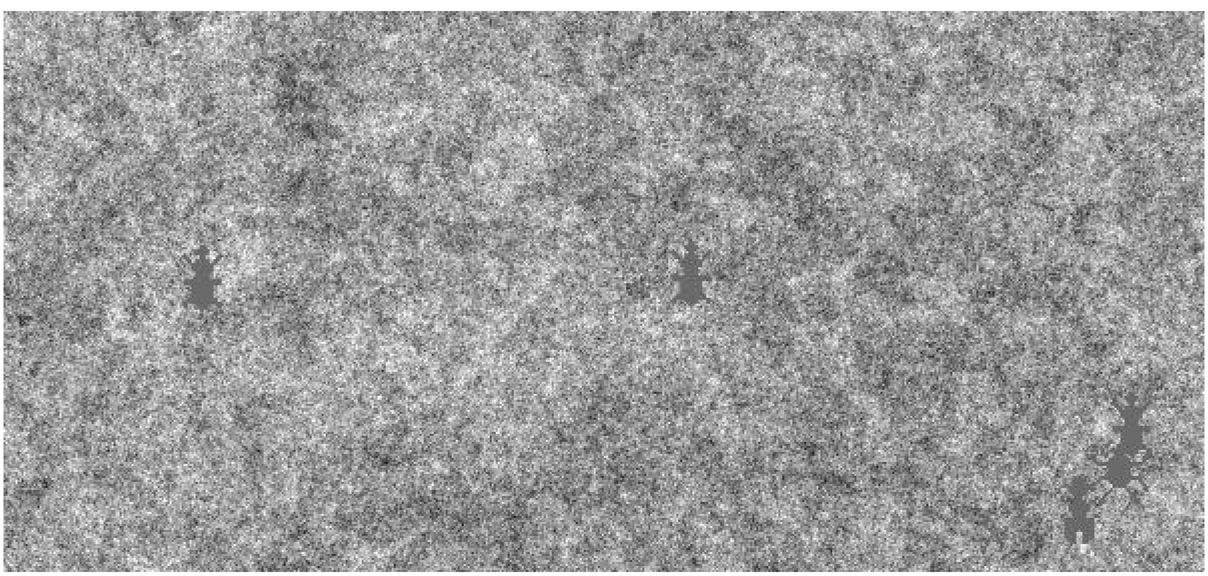

Figure 2. Detail of frame in Figure 1. The Spider sprite can be seen near the lower right corner, although it is partially obscured between nearby Ant and Fly sprites. Two other Ants are in the detail. The contrast between the sprites and the mean background is enhanced for presentation. In motion the sprites obscure each other only briefly.

In the initial training phase, each viewer was familiarized with the three specific bugs shown in the clips, viewed a number of training clips, and received feedback to questions about the evaluation. Each clip had a *spider* as the target sprite. In addition, half the clips had confuser-sprites which were *flies* and *ants*. The fundamental task for the viewer was to record the location of the spider on either the left or right half of the image at the end of the 10-second clip. At the end of each clip the analyst was presented with a split screen. Fig. 3 presents a split screen with text and interior boxes showing the “safe area” in which the sprites lay at clip's end.

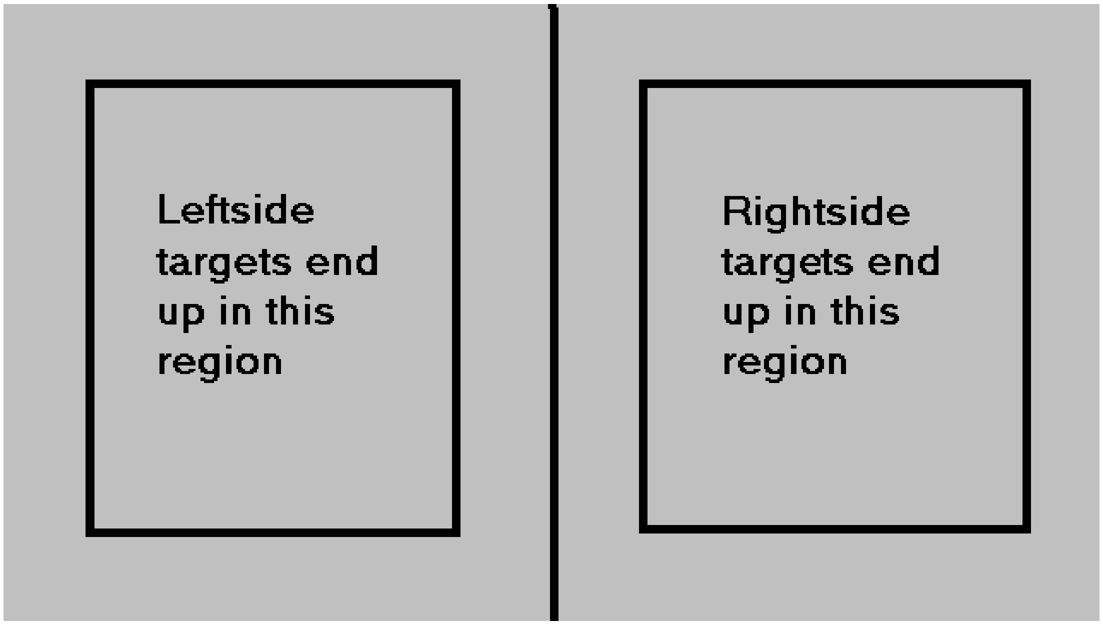

Figure 3: After viewing a 10-second clip, the subject reported the location of the target as either on the left or right side of the image area. The interior boxes indicate a schematic “safe area” for target final location. The safe area was intended to eliminate ambiguity in target location; for a width equal to 5% of the screen height, it succeeded. Boxes were not displayed in the viewing evaluation.

The subjects’ responses were recorded manually. The response template is displayed in Fig. 4. The confidence scores were converted to an integer in the range 0 - 8.

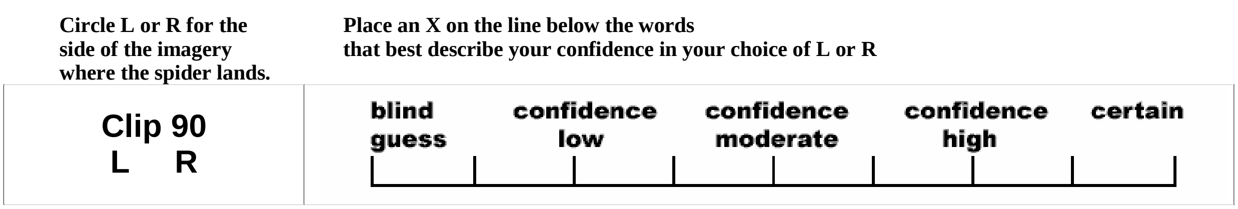

Figure 4. The subjects completed the response template by circling either “L” or “R” in the left-hand box and marking an X to report confidence in the detection task in the right-hand box.

In each clip the ***Spider*** sprite was the target. Half the clips also had confusers, the ***Ant*** sprites and the ***Fly*** sprites (Fig 5). All sprites were 33 X 33 pixels in a 720 X 1280 frame. This corresponded to an angular size of ~ 45 arcmin.

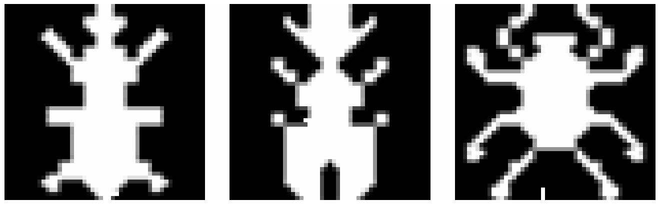

Figure 5a - c: The "bug" sprites were Ant, Fly, and Spider, (5a) – (5c) seen from left to right. Each motion imagery clip had a single Spider. Half the clips had Ants and Flies as confusers. Moving bugs are superimposed on a static colored noise background to form a clip.

3. STUDY DESIGN

The frame rate (FR) study employed a full-factorial design for the parameters.

- *Frame Rate* had six levels: 30, 15, 10, 5, 1 and 0 Frames/s, where 0 F/s was a still image. The
- *Confusers* had two levels. Each clip contained a Spider sprite. Half the clips contained the single Spider, while the other half contained the Spider and 13 additional Ants and Flies in nearly equal number.
- *Contrast* had three levels. Within each clip all the bugs had a fixed gray level and traveled on a colored noise background (*Background*, below). The contrast between the bug and the background was determined as the percent-contrast between the gray-level of the sprites ($\mathbf{Y_S}$) and the mean gray-level of the background ($\mathbf{Y_B}$.). The contrast levels, C, were 5.0%, 2.5%, and 1.0%, where the contrast was defined as

$$\mathbf{C = (Y_B - Y_S) / (Y_B + Y_S)} \qquad \mathbf{(1)}$$

  Each of the sprites was darker than the mean background.
- *Speed*: In each clip all bugs traveled either at 30 pixels/second (Px/s) or at 240 Px/s.
- *Motion Complexity* was either low or high: Each bug stayed in the image frame for the entire clip and at clip's end landed in one of the safe areas [Fig.3.]
  - In the clips with low motion complexity, each bug followed a randomly chosen straight-line trajectory. At a boundary the motion of the bug was reflected with the angle of reflection equal to the angle of incidence.
  - In the clips with high motion complexity, each bug followed a straight-line trajectory until it was randomly perturbed. The direction of motion was randomly forward scattered at randomly selected time intervals. The scattering times were uniformly distributed in the range from 1.0 to 2.0 seconds. The forward-scattering angle fell in the cone [$-\pi/2$, $\pi/2$] centered on the direction of motion.
- *Background* had two levels. The colored-noise background was generated as a fractional Brownian noise[8] with two levels for the Hurst (correlation) parameter, H, of 0.60 and 0.75.

Evaluation materials

The motion imagery clips were generated as a sequence of synthetic uncompressed stills having resolution 1280x720 pixels. The background fractional Brownian noise and the final composite imagery is generated using software developed for the project. The various frame rates were achieved by suitable frame decimation and repetition. Each frame rate was an integral divisor of the 60 Hz frame rate of the playback system.

In an initial dry run of the evaluation, it was found that viewers learned the trajectory of the target sprite when clips were reused at more than one frame rate. As a consequence, separate video clip were generated for each of the 288 combinations of test parameters, including the six levels of frame rate. The viewing sequence of the clips was randomized.

Test Method – Viewing conditions

The evaluation was conducted in the National Institute of Standards and Technology (NIST) Motion Imagery Quality Lab. Illumination levels in the theater followed the specifications of ITU-R BT.500 [9] to provide for optimal viewing. The evaluation used the Lab's 3-chip projection system, which uses deformable mirror devices. The mean screen luminance was 37 cd/m^2 A low level of overhead ambient light was provided to assist the analysts in manually scoring the clips.. The picture height was 0.9 m and the viewing distance was 3.0 m, so the sprites subtended a viewing angle of about 45 arcmin.

The basic test cell was 25 seconds in duration as detailed in Table 1. After preliminary titling, a sequence of viewing/scoring cells followed. Each viewing session was conducted with a single subject. The evaluation was organized into half-hour segments; in addition, the analysts were permitted to break at any time.

Beginning-Ending Times	**imagery Displayed in the Interval**
0.0 – 5.0 s	Brief title and shift in gray level screen signals new clip
5.0 – 15.0 s	Clip
15.0 - 25.0 s	Scoring template

Table 1: Sequencing and duration of imagery in the 25 s basic test cell.

4. ANALYSIS

Analysis of the frame rate data employed logistic regression of the correct response on the study parameters. The partial t-test indicated that confusers, frames rate (FR), and contrast were significant in our generalized linear model. Motion and Background were not significant. The overall correct response rate (interpretability for both detection and identification) was ~ 86%. For **FR**, there was a notable decrease in correct responses at about 5 F/s (Fig. 6).

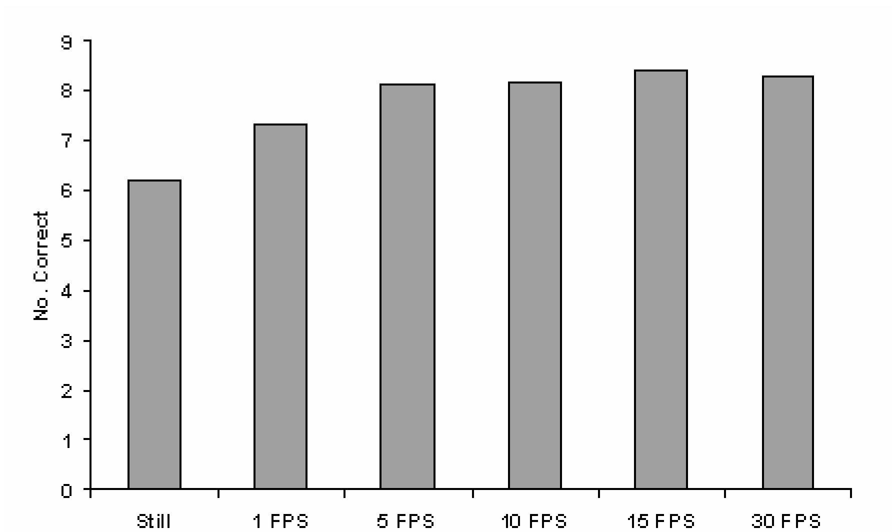

Figure 6: The mean number of correct responses from the 9 subjects displayed as a function of frame rate. The aggregate interpretability fell off at about 5 F/s.

Figure 7 shows the separate correct responses for the detection task (without confusers) and the identification task (with). Note the slight increase in correct detection with decreasing frame rate; decreasing detection is seen below 5 F/s. This result is surprising. It may be related to the geometrically larger changing image area surrounding each sprite that results as the frame rate decreases [7]. By contrast, the correct identification below 15 F/s appears to fall steadily. The presence of confusers affects correct interpretation at lower frame rates (FR=10, 5, and 1)

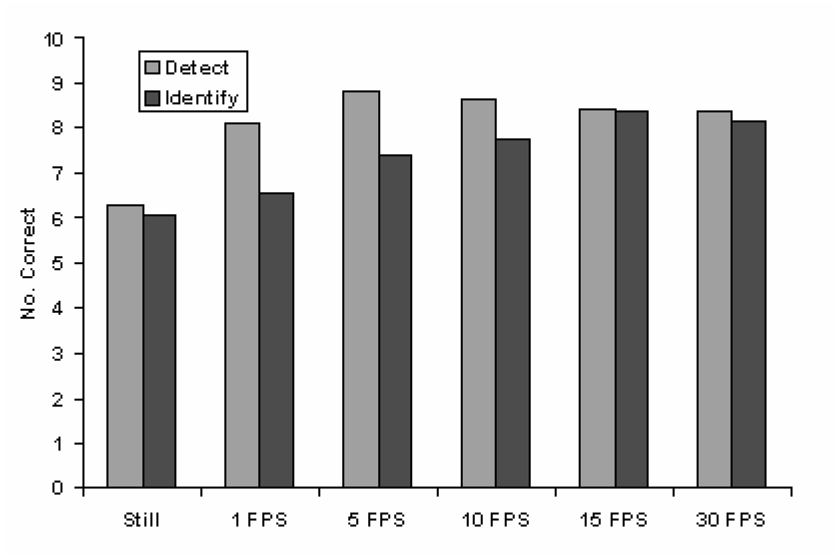

Figure 7: Correct detection does not decrease with decreasing FR to about 5 F/s. It is possible that detection may improve as FR decreases over the same range, but in this study the effect does not achieve statistically significance at the 95% level. Correct identification decreases below 15 F/s.

At the highest contrast, the Confusers did not affect detection and identification differently (Fig 8). Interpretibility increased with the contrast; at lower contrast, Confusers affected both correct detection and identification. Confusers affected interpretibility at higher speeds. At the speed of 30 Px/s (40 arcmin/s) each sprite moved a distance equal to its own width in a second, while at 240 Px/s (5 deg/s) the sprite moved its own width in an eighth of a second. Higher speed made the detection task easier, but complicated the identification task, although this effect was small (Fig 9). Speed affected detection for FR = 1and 30: Higher Speed made detection harder for FR=1, but increased correct detection for FR=30.

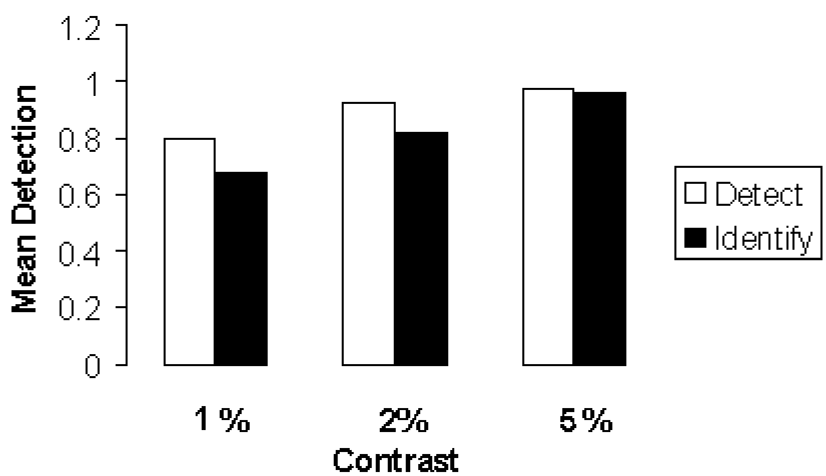

Figure 8: At the highest contrast, the confusers had a comparable effect on correct detection and identification. Both detection and identification, as measured by the fraction of correct responses, decreased with the contrast over the contrast range of the study.

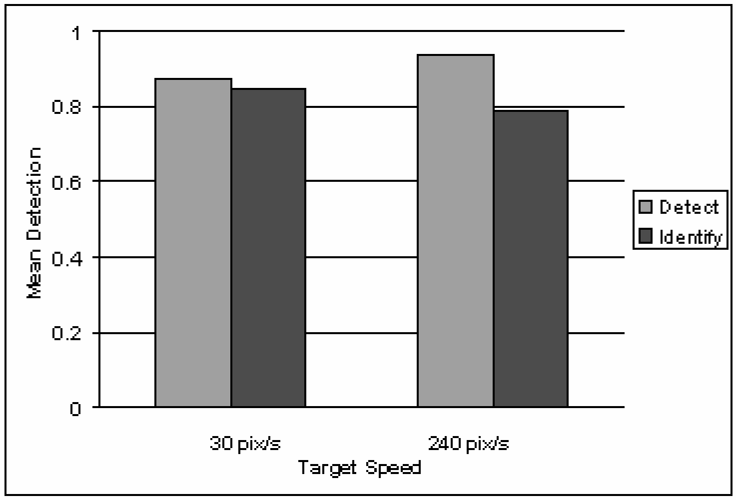

Figure 9: Higher Speed made the detection task easier, but complicated the identification task. Here mean detection was measured as the fraction of correct responses.

In addition to the effects of FR, Contrast, and Confusers, we observed significant effects of Contrast and Background on detection in still images (FR=0).

5. SUMMARY

This perceptual frame rate study found significant effects of Frame Rate, Contrast, and Confusers on the interpretibility of motion imagery. The range of frame rates was selected to span commonly used frame rates in motion imagery systems. The data seems to suggest that the effects of lowered fame rate may have been to enhance some interpretation tasks, but this suggestion is not statistically significant at the 95% level. Pappas and Hinds[10] have found a marked decrease in acceptability of teleconferencing imagery below 5 F/s. One of the main limitations of the present study was the exclusive use of synthetic imagery. While this permitted the careful control of interpretation elements such as target size and scene motion, the use of camera-captured imagery would provide a more reliable view of the magnitude of the frame rate effects in realistic applications. The significance of this study to the development of a Motion Imagery Quality Scale was to provide an estimate of main effects to be included in the scale. The study found a robust effect of Frame Rate, Contrast, and Confusers. The inclusion of Confusers presents a special challenge in the development of a useful scale as the presence of confusers may not be known before staging a collection and, unlike contrast, compensating for it may not be possible.

6. REFERENCES

[1] J.M. Irvine 2003. *"National Imagery Intelligence Rating Scale (NIIRS)" in Driggers, R.G.* The Encyclopedia of Optical Engineering, Marcel Dekker., October 2003.

[2] C. Leachtenauer "National Imagery Interpretability Rating Scales: Overview and Product Description," *Proceedings of the American Society of Photogrammetry and Remote Sensing Annual Meetings*, April 1996.

[3] L.A. Maver, C.D. Erdman, K. Riehl, "Imagery Interpretability Rating Scales", *Digest of Technical Papers, International Symposium Society for Information Display*, Society for Information Display, Vol. XXVI, pp: 117-120, 23 May, 1995.

[4] J.M. Irvine,. "National Imagery Interpretability Rating Scales (NIIRS): Overview and Methodology" *Proceedings of the International Society for Optical Engineering (SPIE),* 29-30 July, 1997, Volume 3128, pp.93-103.

[5] J. C. Leachtenauer, W. Malila, J. M. Irvine, L. Colburn, and N. Salvaggio, "General Image-Quality Equation: GIQE," *Applied Optics*, vol. 36, pp. 8322-8328, 1997.

[6] J. M. Irvine, C. Fenimore, D. Cannon et al., Factors Affecting Development of a Motion Imagery Quality Metric, *Proceedings of the International Society for Optical Engineering (SPIE)*, 29-30 July, [2005], Volume 3128.

[7] T. Cornsweet, *Visual perception*, Harcourt, [1970].

[8] B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, NY [1982].

[9] Recommendation ITU-R BT.500, *Methodology for the Subjective Assessment of the Quality of Television Pictures*.

[10] T. Pappas and R. Hinds, "On video and audio data integration for conferencing", *Proceedings of the Society for Optical Engineering (SPIE)*, Volume 2411, 120- 127 February 1995