Influence of severe vibrations on the visual perception of video sequences

Adrian Stern, MEMBER SPIE
Etan Fisher
Stanley R. Rotman, MEMBER SPIE
Norman S. Kopeika, FELLOW SPIE
Ben-Gurion University of the Negev
Department of Electrical and Computer Engineering
P. O. Box 653
84105 Beer-Sheva, Israel
E-mail: stern@ee.bgu.ac.il

Abstract. There are two kinds of video image sequence distortions caused by vibration of the camera. The first is the vibration of the line-of-sight, causing location changes of the scene in successive frames. The second distortion is the blur of each frame of the sequence due to frame motion during its exposure. The relative effects of these two types of degradations on the ability of observers to recognize targets are investigated. This study is useful for evaluating the amount of effort required to compensate each effect. We found that the threshold contrast needed to recognize a target in a vibrating video sequence under certain conditions is more affected by the motion blur of each frame than the oscillation of the line-of-sight. For digital sequence restoration methods, this study determines the required precision of the deblurring and registration processes. It shows that the deblurring process should not be neglected, as it often is. [DOI: 10.1117/1.1367256]

Subject terms: image sequence vibrations; image restoration; motion perception; target acquisition.

Paper 200336 received Aug. 25, 2000; revised manuscript received Jan. 18, 2001; accepted for publication Jan. 29, 2001.

1 Introduction

Vibration may hide vital information in a video sequence captured by a camera subject to mechanical vibration. Image sequence vibration is common when the imaging system is located on aircraft, ships, helicopters, and other vehicles with active motors or mechanical structural resonances. Vibration can be minimized by proper design; in practice, however, it is often the main influence on the visual perception despite the best attempts at stabilization.

Much effort has been undertaken to understand the way the human eye-mind system perceives moving and vibrating objects. Several models were developed to explain, characterize, and predict the perception of moving stimuli.[1–6] It is obvious that moving stimuli perception models have to consider the temporal behavior of the visual system in addition to the spatial behavior. One of the most common spatio-temporal models of the visual system is that presented by Kelly.[2] Kelly measured the contrast sensitivity of the visual system for sinusoidal gratings crossing the field of view with stabilization fixation control.[2,7] He found that the human eye acts as a bandpass filter both in the spatial and temporal frequency domains. In other studies the temporal response of the eye was measured by stimulating flickering light sources, moving targets, or vibrating objects.[7] It was found that the visibility of oscillatory movements is best at temporal frequencies between 4 to 12 Hz and depends on the oscillation amplitude.

The purpose of this work is to investigate the perception of the image sequences captured by a vibrating camera. An observer viewing a sequence of images captured with a vibrating camera senses that the scene is oscillating. Therefore one could assume that perception of an object captured by a vibrating camera is similar to that of the object itself vibrating as viewed with the unaided eye. However, the optical signal captured by a vibrating camera differs significantly from that of an oscillating object because of the capturing process. In the capturing process, the optical signal is temporally integrated during the exposure and then discretized. Because of the sampling, the signal loses the temporal continuity. The integration during the exposure together with the vibration of the camera causes different blur in each frame. This is explained in Sec. 2. Therefore the capturing process cannot be modeled as a simple linear time invariant (LTI) process. This, together with the nonlinear temporal response of the eye[3,8] prevents formulation of a straightforward relation between the perception of vibrated image sequences and the unaided eye perception models. Consequently, the current study is based on experimental results rather than on previous models developed for unaided eye viewing of the object.

Two effects distort an image sequence captured by a camera subject to vibration. The first is the oscillation of the line-of-sight (LOS), causing the sensation that the scene is vibrating. This effect obligates the observer's eye to follow the vibrating objects to track them. The physiological tracking mechanism is complicated and depends mainly on motion velocity.[9] The second distorting effect is the blur in each frame caused by the nonzero exposure time. Relative motion between the camera and the object during the exposure smears the images and decreases the spatial resolution, which in turn makes the perception of detail more difficult. The severity of these effects depends on the vibration conditions (amplitude and frequency) and on the exposure time. Large vibration amplitude and high frequency increase both effects, while the exposure time affects only the blur extent. Principally, shortening the exposure time can reduce the amount of blur. However, often this is not practical because this decreases the signal to noise ratio. The dependence of the blur on the vibration amplitude, frequency, and exposure time is discussed in Sec. 2.

Understanding the influence of these two effects on visual perception is important for designing postcapture digital sequence processing techniques. Image sequence restoration techniques (see for example Refs. 10 and 11) perform basically two processes: registration and frame restoration. In the registration step the frames are realigned to compensate for the undesired motion of the LOS. This step is often referred to as the motion compensating step. The vibrations also induce significant motion blur in each frame. The motion blur together with blur from other sources (optical, electronic, environmental, etc.) is recovered in the frame restoration step. In addition to deblurring, the noise is suppressed at this step. The study in this work gives an indication of the relative importance of these two restoration steps from the visual perception point of view. It will attempt to answer the question: Given the characteristics of the camera vibrations, on which restoration step do the sequence restoration algorithms concentrate?

Most digital motion compensation algorithms implemented today only stabilize the LOS by performing the registration step (see for example Refs. 12, 13, and 14). This approach assumes that it is more important to overcome the oscillation of the blur than to recover the blur. Indeed, if the purpose of the digital process is to obtain a video sequence that is pleasant to the eye, this assumption is correct since scene vibrations are the main cause of eye fatigue. However, we demonstrate in this work that if extraction of fine information and details is required, the above assumption is not correct. It is shown that the threshold contrast (TC) needed to recognize targets in a vibrating video sequence under certain conditions is more affected by the motion blur of each frame than the oscillation of the LOS. This means that digital image sequence restoration algorithms should emphasize motion blur restoration rather than precise recognition of the frames.

2 Vibration Distortion

We assume that the camera is subject to low-frequency vibrations. Low-frequency vibrations are defined as those vibrations for which the exposure time t_e is less than the vibration period T_0 (Fig. 1). Low-frequency vibrations are more common and more complicated to deal with by mechanical stabilization and postcapture processing. Image degradation caused by low-frequency vibration has been previously analyzed in detail.[15–18] We will summarize a few characteristics of low-frequency vibration degradations essential for this work.

The blur extent depends on the vibration amplitude and the ratio between the exposure time and vibration period. For example, the average blur is given[16] by $\overline{d} = 3.57A\,t_e/T_0$, where A and T_0 are the vibration amplitude and period respectively, and t_e is the exposure time.

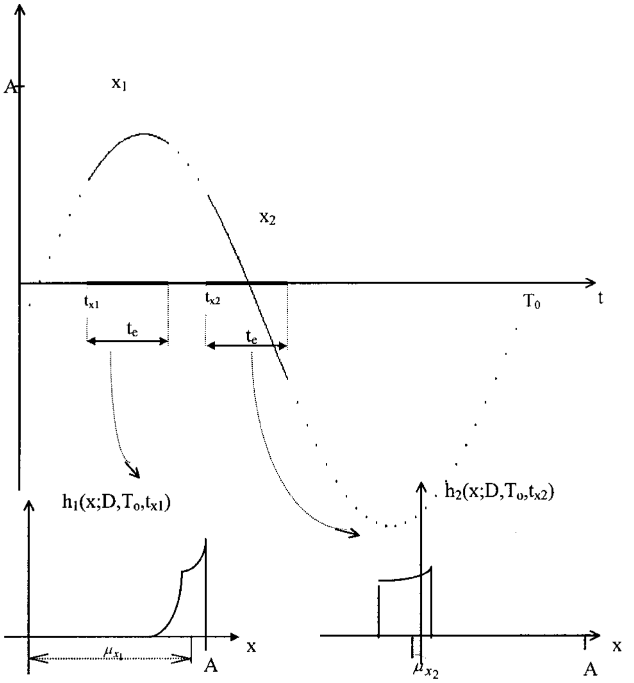


Fig. 1 Simple motion functions during exposure of a low frequency vibrating image platform with amplitude A, vibration period T_0, and given exposure time t_e. The PSF h_1 and h_2 depend on the instant of exposure t_{x1} and t_{x2}. The PSFs have different shapes and location (μ_1 and μ_2 represent PSFs in the center of mass locations).

The point spread function (PSF) due to low-frequency vibrations exhibits a strong random behavior dependent on the start time of the exposure. This is illustrated in Fig. 1. It can be seen that at different instants of exposure (t_x) the PSF exhibits different shapes causing different blur in each frame. Images captured near the vibration extreme locations, i.e., h_1 in Fig. 1, are less blurred than those captured near the zero crossing of the motion function, i.e., h_2 in Fig. 1.

From Fig. 1 it can also be seen that the low-frequency vibration PSFs have also different displacements (μ_1 and μ_2 denote the center of mass of h_1 and h_2, respectively). This causes oscillation of the scene.

3 Experiment Description

To evaluate the relative effects of the two degradation factors (LOS vibration and motion blur) on human eye recognition capability, we simulated in each experiment three types of movies. The first type is a simulation of movies fully degraded by vertical low-frequency vibrations. Two effects distorted the image sequence: field-of-view (FOV) oscillation and motion blur of each frame. This type of image sequence is referred to as the ''F movie.'' An illustration of this type of image sequence is shown in Fig. 2(b). The second type of movies were similar to F movies but without motion blur in each frame. This kind of image sequence is distorted by FOV oscillation only, as illustrated in Fig. 2(c). We denote this type of movies as ''V movies.'' The V movies represent image sequences captured by a vibrating camera having zero exposure time, or F movies that were ideally deblurred (the motion blur of each frame was completely restored). The third type of image sequence, denoted ''S movies,'' are similar to F movies but without FOV oscillation [Fig. 2(d)]. In the context of postcapture image sequence restoration, S movies represent precise registration of the frames.

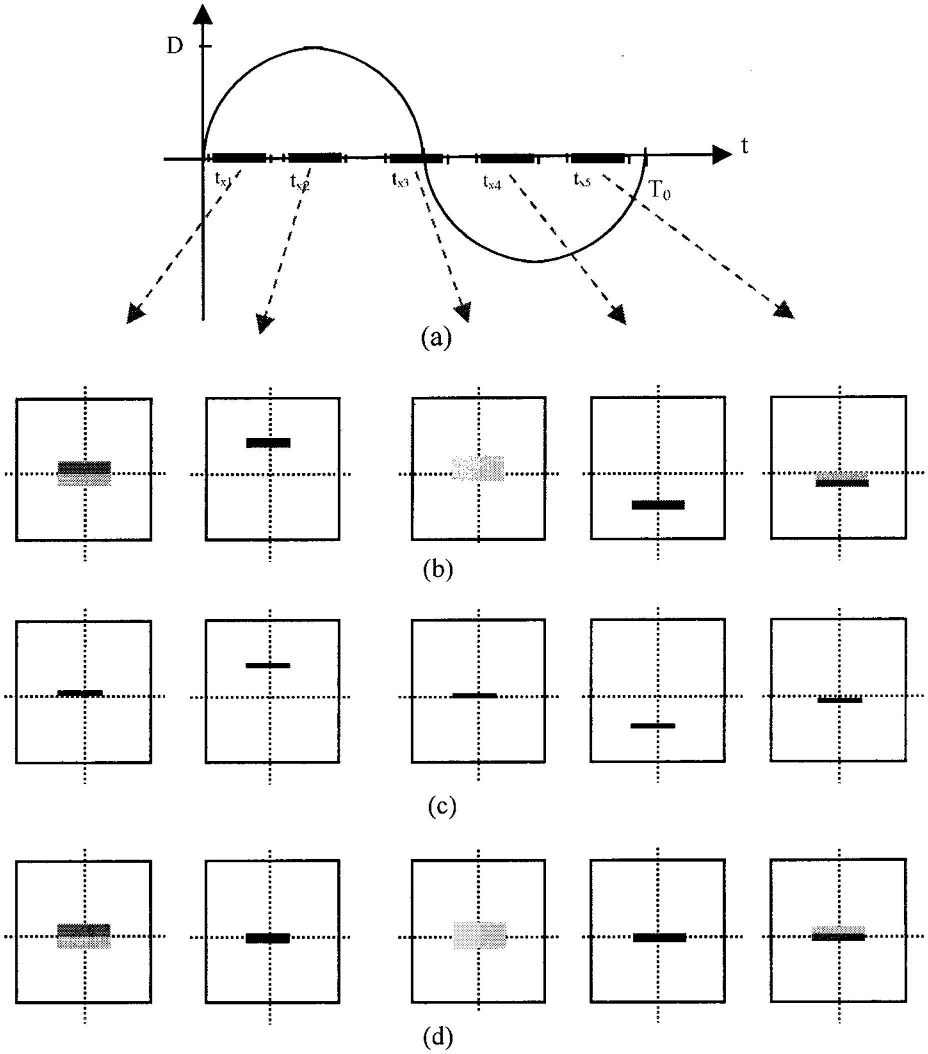


Fig. 2 Illustration of three types of simulated movies. The object is a horizontal bar filmed by a vertically vibrating camera with the motion function (a) and captured at exposure instants $t_{x1}, \ldots t_{x5}$. (b) F movie—fully degraded movies distorted both by scene oscillation and motion blur; (c) V movie—movies degraded only by scene oscillation; and (d) S movie—stabilized movies degraded only by motion blur.

Each movie contained one letter of the English alphabet. The minimal dimension of the letter was 1 mm over an angle of 1.4 mrad (viewer seated 70 cm from screen) and the maximal dimension of each letter was over an angle of 14 mrad. We note that the TC for static scenes is least[15] at an angle of approximately 4 mrad (spatial frequency of 2 cycles per degree). This angle is within the range of angles of minimal and maximum dimensions of the letters. This means that small details can be resolved even with minimum contrast.

Another type of target we have considered, besides the characters, is bar targets. In general bar targets are very convenient because they permit defining the precise spatial frequency for which the experiment is carried out. However, we found bar targets to be inconvenient in our case. The perception of a movie with a periodical target performing periodical motion may be influenced artificially because of aliasing. A well-known example is the apparent backward motion of coach wheels. At certain rotational frequencies, above the temporal bandwidth the wheel bars are better recognized than at other lower rotational frequencies. This would not happen to wheels with nonperiodical bars. Therefore we choose to use nonperiodical targets as characters.

The parameters that were tested were the temporal frequency f and amplitude A of the vibrations. The purpose was to discover the range of values in which it was possible to identify a change in the relative effect of FOV oscillation and blurring. Table 1 contains the five values of frequency and amplitude used. These values are typical of severe vibrations of a nonstabilized video system after optical magnification. The vibration amplitude ranges between 5 and 20% of the FOV.

Two parameters chosen as constants were exposure time and playback frequency. The exposure time of each frame was chosen as 1/60 msec. The movies were played at a frequency of 50 Hz (frames per second).

Table 1 Frequency and amplitude values tested in the experiment.

Condition	Temporal frequency [Hz]	Pitch angular amplitude [mrad]
1	5	14
2	10	28
3	13	35
4	15	42
5	20	56

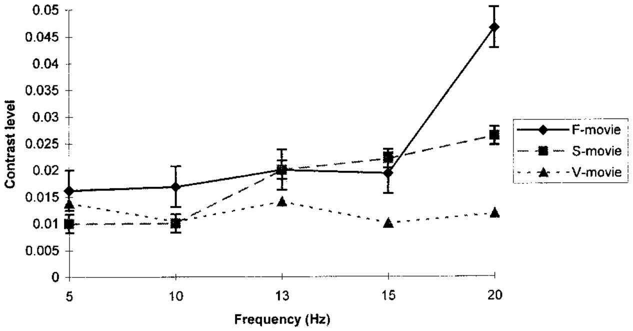


Fig. 4 The TCC as function of vibration frequency at vibration amplitude of 35 mrad.

In this experiment we measured the threshold contrast curve (TCC) at which each subject identifies the letter in each movie. Movies were built for the three states of degradation previously mentioned (fully degraded, vibration only, and blurring only), and for all the values of amplitude and frequency stated before. Each movie was also created for different levels of contrast. Movies with different vibration frequency and amplitude were played randomly to prevent a learning effect. At the beginning, movies with low contrast were displayed. If a movie was not identified at a lower contrast level, the interface immediately chose randomly another movie and the nonidentified movie was played afterward at a higher contrast level.

The experiment involved 30 participants. The subjects were asked to recognize the letter on the movie screen for 3 sec. All subjects were to have 6/6 vision or vision corrected with eyeglasses. Special care was taken to maintain uniform experimental conditions for each observer (such as lighting, spatial frequency, and brightness) and to nullify psychological factors as learning, fatigue, and interface errors. The subjects were seated at a constant distance from the screen to maintain a constant spatial frequency. The distance chosen as normal and comfortable was 70 cm. To maintain constant spatial resolution, the screen resolution was calibrated at the beginning of the experiment. The constant size of the display, together with the constant observation distance, maintained the angular scaling of the vibration as described in Table 1 and the angular size of the letters as described before. The computer screen brightness, contrast levels, and the lighting in the lab were maintained as constant. To prevent fatigue and confusion the experiment was limited to half an hour.

This experimental setup is designed to give a typical example of relative importance of both effects distorting the image sequence. Many parameters in the experiment can be varied further to match specific imaging conditions. The vibration conditions (amplitude, frequency, and direction) can be varied, targets with different characteristics can be chosen (spatial frequency spectrum, background, clutter, orientation, and color) and different filming and environmental conditions can be used (exposure time, illumination, and display).

4 Results

The average TCC was calculated as a function of vibration frequency and amplitude. Figures 3, 4, and 5 show typical examples of TCC behavior as a function of vibration frequency at different constant amplitudes. It can be seen that as the vibration frequency grows, the average TC rises. This is expected since fast vibrations make recognition more difficult. At low vibration frequencies (up to 10 to 13 Hz), the S (stabilized and blurred) and V (deblurred but nonstabilized) movies have a similar affect on image recognition. This means that oscillation of FOV and the motion blur within the frames have similar effects on recognition. At higher vibration frequencies the TC of the V movies is significantly lower than that of the S movies, indicating that blur affects recognition more than oscillation of the scene. Figures 6, 7, and 8 show typical examples of TCC behavior as functions of vibration amplitudes at different constant frequencies. It can be seen that as the vibration amplitude grows, the average contrast level at which the subjects identify the S and F movies increases more than that in the V movies. For a frequency of 5 Hz, the average TC for S and V movies is similar and almost

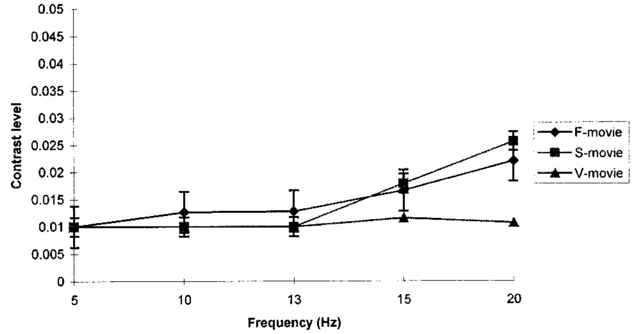


Fig. 3 The recognition TCC as function of vibration frequency at vibration amplitude of 14 mrad.

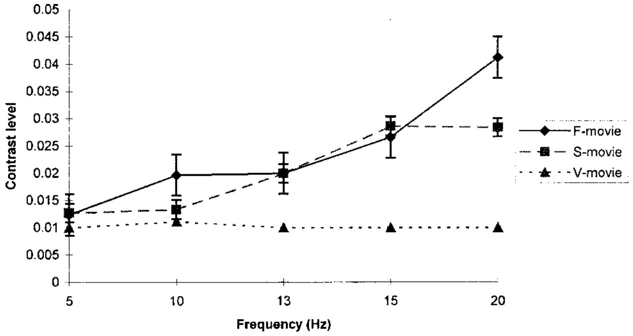


Fig. 5 The TCC as function of vibration frequency at vibration amplitude of 42 mrad.

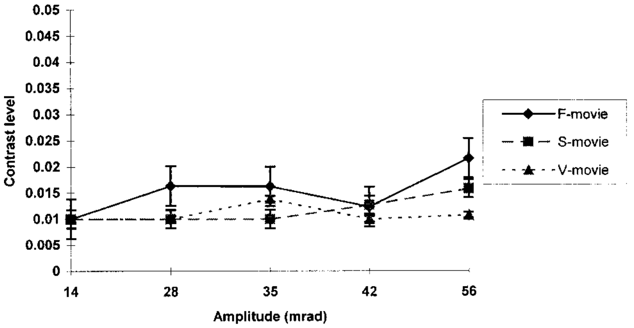


Fig. 6 The TCC as function of vibration amplitude at vibration frequency of 5 Hz.

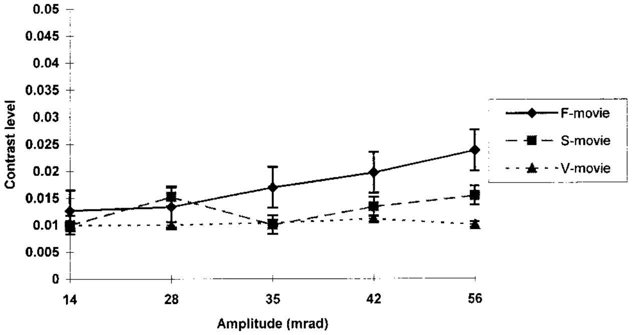


Fig. 7 The TCC as function of vibration amplitude at vibration frequency of 10 Hz.

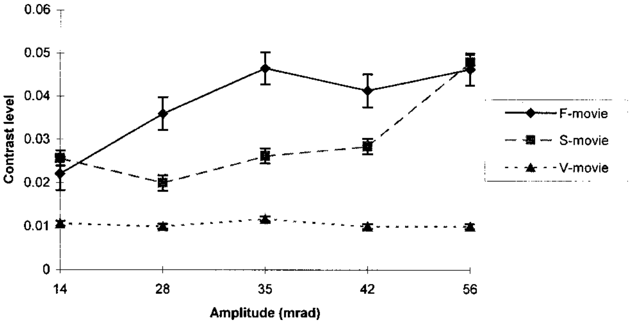


Fig. 8 The TCC as function of vibration amplitude at vibration frequency of 20 Hz.

Table 2 Comparison between F and S movie data, and F and V movie data, as functions of frequency.

Frequency [Hz]	*P* value for data sets F and S	*P* value for data sets F and V
5	0.16	0.04
10	0.16	0.0004
13	0.67	0.1
15	0.77	0.0002
20	0.55	0.0001

Table 3 Comparison between F and S movie data, and F and V movie data, as functions of amplitude.

Amplitude (mrad)	*P* value for data sets F and S	*P* value for data sets F and V
14	0.77	0.001
28	0.1	~0
35	0.15	0.001
42	0.26	1.26e-5
56	0.78	2.53e-5

independent of the vibration amplitude. For a frequency of 10 Hz (Fig. 7), the contrast level for the S and V movies are both slightly lower than that for F movies. The contrast level for the S and V movies are still close, and that of S movies is marginally higher than that for the V movies. Beginning at 13 Hz, the behavior of the S movie TCC grows toward the behavior of the F (fully degraded) movie TCC, meaning that stabilization improves less than deblurring.

To examine statistically the similarity between the elements of degradation (S and V) and the full degradation (F), a two-factor analysis of variance[19,20] (ANOVA) was performed.

Tables 2 and 3 show the *P* values (Prob$\{F[\alpha<0.05;(a-1)(b-1),ab(n-1)]\}\geqslant F$ where $a=2$, $b=5$, $n=30$) for the TC values measured in the F and S movies, and F and V movies as functions of vibration amplitude and frequency, respectively. The *P* value (also "observed significance value"[20]) is a common measure of disagreement to a null hypothesis.[19] In our case, *P* values smaller than the significance value $\alpha=0.05$ indicate that there is sufficient evidence that the TCC of the two movies tested differ. On the other hand, large *P* values mean that differences between the TCCs of the two movies can be explained as chance.

This test indicates that there is no statistical significant difference between the TC levels in S and F movies ($p>0.05$). It can be seen also that as frequency and amplitude values rise, the *P* value rises, meaning the difference between the TC levels in S and F movies decreases. However, for V (blurred but nonstabilized) movies, the opposite happens. The TC levels in the V movies are significantly different from those in F movies ($p\ll 0.05$). As values of vibration frequency and amplitude rise, the *P* value decreases for V movies. This means that the difference between the TC levels in V and F movies increases. In other words, as the amplitude and the frequency rise, the deblurring of each frame has more of an effect on recognition than does the stabilization of the blurred scenes.

5 Summary and Conclusions

The relative influence of the two image sequence degradation factors (FOV oscillation and motion blur) on the human eye recognition capability was investigated. The inquiry was carried out by comparing the contrast level required to recognize letters in image sequences fully degraded by vibrations (F movies) to that degraded only by motion blur of each frame (S movies) and that distorted only by the oscillation of the LOS (V movies). The following summarizes our observations from our measurements.

For low values of amplitude and frequency, the effect of the motion blur and oscillation of the FOV on the TCC are similar, meaning that both degradation factors have similar influences on recognition capability.

As frequency rises the TCC for stabilized (but blurred) movies and fully degraded movies rises. The TCC for vibrated only (but not blurred) movies remains at a low contrast level.

Beginning at a vibration amplitude value of approximately 28 mrad and a frequency value of 13 Hz, the dominant degradation factor is the blurring, and the stabilized (but blurred) movie TCC behaves similarly to the fully degraded movie TCC. This indicates that at high amplitude and frequency the blur within each frame is the main cause of recognition degradation.

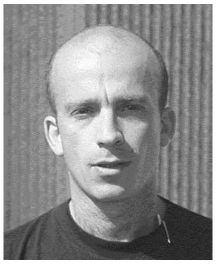
As mentioned in Sec. 1, it is difficult to directly relate our results of images captured on camera and displayed on a monitor respectively to perception models of the unaided eye. A possible interpretation of our experimental results is the following. At the vibration frequencies the experiment was performed (above 5 Hz), the eye is unable to track the target.[9] It is more probable that the eye adapts itself to the vibration by selecting the ‘‘good shots’’ (least distorted frames) together with a motion extrapolating mechanism. In the F and S movies the ‘‘good shots’’ appear at the vibration extrema because, as mentioned in Sec. 2, there the blur is smaller. We assume that the eye recognizes the target mainly from these frames. However, the blur of the ‘‘good shots’’ depends also on the vibration conditions. Therefore, as the vibration amplitude and frequency increase, the F and S movies are more difficult to recognize. This is not true for the V movies that consist of nonblurred frames. Therefore, the recognition of V movies is almost independent of the vibration conditions. The relative high recognition ability of the V movies indicates that the eye integration time cannot be much larger than the display period of one frame (20 ms), otherwise a multitarget image would be seen. If during the eye integration period more than one frame of the V movies are displayed, the image perceived by the eye is of several letters having same quality and located at different positions, and eventually overlapping. This would make the recognition more difficult and very dependent on the vibration amplitude.

The main conclusion from this work, which is that in certain conditions the blur is the main cause of recognition degradation, should be taken into account when designing dynamic imaging systems and restoration techniques. For example, digital image restoration should not concentrate on precise registration of the images and neglect the motion blur. On the contrary, in cases of severe vibrations above certain amplitude and frequency values that may vary with imaging conditions, the emphasis should be on deblurring the images rather than on precise registration.

We note that these results hold for the specific setup of the experiment. It is expected that for a different setup (different exposure time, vibration amplitudes and frequencies, direction of vibration, target and background contrast, and clutter) these results may differ somewhat quantitatively, but not in their essential physical interpretation.

References

1. D. H. Kelly, ‘‘Motion and vision. I. Stabilized images for stationary gratings,’’ *J. Opt. Soc. Am.* **69**, 1266–1274 (1979).
2. D. H. Kelly, ‘‘Visual processing of moving stimuli,’’ *J. Opt. Soc. Am.* **2**, 216–225 (1985).
3. A. B. Watson and A. J. Ahumada, ‘‘Model of human visual-motion sensing,’’ *J. Opt. Soc. Am. A* **2**, 322–341 (1985).
4. E. H. Adelson and J. R. Bergen, ‘‘Spatiotemporal energy models for perception of motion,’’ *J. Opt. Soc. Am. A* **2**, 284–299 (1985).
5. S. J. Anderson and D. C. Burr, ‘‘Spatial and temporal selectivity of the human motion detection system,’’ *Vision Res.* **25**(8), 1147–1154 (1985).
6. A. T. Smith, ‘‘Correspondence-based and energy-based detection of second order motion in human vision,’’ *J. Opt. Soc. Am. A* **11**(7), 1940–1948 (1994).
7. R. W. Dithcburn, *Eye-Movements and Visual Perception*. Chapters 7.42, 10.9, 10.10, Clarendon Press, Oxford (1973).
8. P. A. Buser and M. Imbert, *Vision*, Chapter 5, MIT Press, Cambridge (1992).
9. A. J. Martins, E. Kowler, and C. Palmer, ‘‘Smooth pursuit of small-amplitude sinusoidal motion,’’ *J. Opt. Soc. Am.* **2**, 234–242 (1985).
10. M. Elad and A. Feuer, ‘‘A restoration of a single superresolution image from several blurred, noisy, and undersampled measured images,’’ *IEEE Trans. Image Process.* **6**(12), 1646–1658 (1997).
11. V. Avrin and I. Dinstein, ‘‘Restoration and resolution enhancement of video sequences,’’ *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing* **4**, 2549–2553 (1997).
12. K. Ratakounda, ‘‘Real time digital video stabilization for multimedia applications,’’ *Proc. 1998 IEEE Int. Symp. Circuit Systems* **4**, pp. 98-72, IEEE Press, New York (1998).
13. S. J. Ko, S. H. Lee, S. W. Jean, and E. S. Kang, ‘‘Fast digital image stabilizer based on Gray-coded bit plane matching,’’ *IEEE Trans. Consum. Electron.* **45**(3), 598–903 (1999).
14. C. Monomoto and R. Chellappa, ‘‘Fast electronic digital image stabilization for off road navigation,’’ *Real-Time Imag.* **2**(5), 289–296 (1996).
15. N. S. Kopeika, *A System Engineering Approach to Imaging*, Chap. 18, pp. 517–524, SPIE Press, Bellingham, WA (1998).
16. D. Wulich and N. S. Kopeika, ‘‘Image resolution limits resulting from mechanical vibration,’’ *Opt. Eng.* **26**, 529–533 (1987).
17. O. Hadar, I. Dror, and N. S. Kopeika, ‘‘Image resolution limits resulting from mechanical vibration. Part IV: Real time numerical calculation of optical transfer functions and experimental verification,’’ *Opt. Eng.* **33**(2), 566–578 (1994).
18. A. Stern and N. S. Kopeika, ‘‘Analytical method to calculate optical transfer function for image motion and vibration using moments,’’ *J. Opt. Soc. Am. A* **14**, 388–396 (1997).
19. R. V. Hogg and J. Ledolter, *Applied Statistics for Engineers and Physical Science*, 2nd ed., Chapters 7 and 8, Macmillian Publishing Company, New York (1992).
20. W. Mendehnall and T. Sincich, *Statistics for Engineering and the Sciences*, 4th ed., p. 445, 789–919, Prentice Hill, New Jersey (1995).



Adrian Stern received his BSc and MSc (cum laude) degrees in electrical and computer engineering from Ben-Gurion University of the Negev, Israel, in 1988 and 1997, respectively. He is now a PhD student in the Electrical and Computer Engineering Department and part of the electro-optics research group. His current research interests include effects of motion and vibrations on sequences of images, influence of video camera vibration on the human visual system, image restoration, target aquisition, image compression, and imaging through atmosphere. He is a member of SPIE.

Etan Fisher received his BSc in electrical engineering from Ben-Gurion University of the Negev, Beer-Sheva in 1999. He is currently studying toward his MSc and is involved in research in the fields of audio and video engineering and music technology.

Stanley R. Rotman received his BS, MS, and PhD degrees from the Electrical Engineering and Computer Science Department of the Massachusetts Institute of Technology in 1979, 1980, and 1985. In 1985 he joined the staff of the Science and Technology Division of the Institute for Defense Analyses, where he did research on solid-state lasers, IR target acquisition, and electronic materials characterization. In 1987 he became a senior lecturer with the Department of Electrical and Computer Engineering at Ben-Gurion University of the Negev, Israel. His present research interests are characterization of optical materials for solid-state lasers, modeling human target acquisition in IR images and digital image processing. Prof. Rotman is a member of OSA, IEEE, and SPIE.

Norman S. Kopeika received the BS, MS, and PhD degrees in electrical engineering from the University of Pennsylvania, Philadelphia, in 1966, 1968, and 1972, respectively. In 1973 he joined the Department of Electrical Engineering, Ben-Gurion University of the Negev, Beer-Sheva, where he is currently a professor and incumbent of the Reuben and Frances Feinberg Chair in Electrooptics. During 1978-1979 he was a visiting associate professor in the Department of Electrical Engineering, University of Delaware, Newark. His research interests include atmospheric optics, effects of surface phenomena on optoelectric device properties, optical communication, electronic properties of plasmas, laser breakdown of gases, the optogalvanic effect, electromagnetic-wave-plasma interaction in various portions of the electromagnetic (EM) spectrum, and utilization of such phenomena in EM-wave detectors and photopreionization lasers. He has published over 120 reviewed journal papers in these areas. From 1989 to 1993 he served two terms as department chair. Dr. Kopeika is a senior member of the IEEE and a member of SPIE, the Optical Society of America, and the Laser and Electrooptics Society of Israel.