Optical Engineering 45(11), 117006 (November 2006)

Effects of image restoration on acquisition of moving objects from thermal video sequences degraded by the atmosphere

Oren Haik, MEMBER SPIE
Yosef Lior
Daniel Nahmani
Yitzhak Yitzhaky, MEMBER SPIE
Ben-Gurion University
Department of Electro-Optics Engineering
P.O.B. 653
Beer-Sheva 84105, Israel
E-mail: itzik@ee.bgu.ac.il

Abstract. Remotely sensed videos, captured by high-resolution imagers, are likely to be degraded by the atmosphere. In still images, the degradation sources, which include turbulence and aerosols, mainly cause blur. In video sequences, however, spatiotemporally varying distortions caused by turbulence also become important. These atmospheric degradations reduce image quality and therefore the ability of target acquisition by the observers. The effects of image quality and image restoration (deblurring) on target acquisition in still images were examined previously in several studies. Nevertheless, results obtained in static situations may not be appropriate for dynamic situations (with moving targets), which are frequently more realistic. This work examines the effect of image restoration on the ability of observers to acquire moving objects (such as humans and vehicles) in video sequences. This is done through perception experiments that compare acquisition probabilities in both restored and nonrestored video sequences captured by a remote-sensing thermal imaging system. Results show that image restoration can significantly improve the acquisition probability. These results correspond to the static case. However, unlike the static case, considerably smaller differences were obtained here between the probabilities of target detection and target recognition. [DOI: 10.1117/1.2388933]

Subject terms: target acquisition; moving-target acquisition; atmospheric degradation; video restoration.

Paper 060083R received Feb. 1, 2006; revised manuscript received Apr. 15, 2006; accepted for publication Apr. 21, 2006; published online Nov. 20, 2006.

1 Introduction

Acquisition of moving objects (targets) by human observers from long-distance video sequences is a fundamental task in many applications such as surveillance and reconnaissance. Consequently, a major effort has been made, in order to fabricate advanced imaging systems with low noise and better resolution. However, while the resolution of imaging hardware may be limited by aperture size or by the camera modulation transfer function (MTF), it is the atmosphere that usually limits image quality for long ranges.[1] The main atmospheric degradation sources are optical turbulence and particulates (aerosols). Turbulence results in random spatiotemporal variations in the atmospheric refractive index. Turbulence causes image blur and time-varying image shifts (caused by wavefront tilts) either of the whole image or of parts of it, depending on the isoplanatic patch.[1,2] Aerosols in the atmosphere cause attenuation and image blur, which result from absorption and scattering, respectively.[1,3] Both of these degradation sources may reduce the ability of acquiring moving objects by human observers in long-range imaging. Two main reasons for that are:

1. The objects in the video frames are distorted (blurred).
2. The time-varying image shifts caused by the turbulence induce additional movements in the scene (temporal clutter), which may distract the observer's attention from the moving objects.

Considerable effort has been made to set criteria for target acquisition probabilities for still images. The most basic and widely known are the *Johnson criteria*,[4,5] which consider the number of line pairs included in the target's narrow dimension that is required for different acquisition tasks (detection, orientation, recognition, and identification). Similar criteria for marine acquisition have also been developed.[1] Since Johnson, much work has been done extending the Johnson criteria to incorporate such aspects as multiple targets,[6] multiple observers,[7] obscurants,[8] atmospheric blur,[9] and clutter.[10–13] A more recent target acquisition model considered also properties of the visual system and combined noise and blur degradations in an image.[14]

The effect of image restoration on target acquisition performance by observers was analyzed for still images only.[15–17] In these studies, different amounts of atmospheric blur and Gaussian noise were added synthetically to the images. The images were then restored using Wiener filtering.[18] Perception experiments were performed, and it

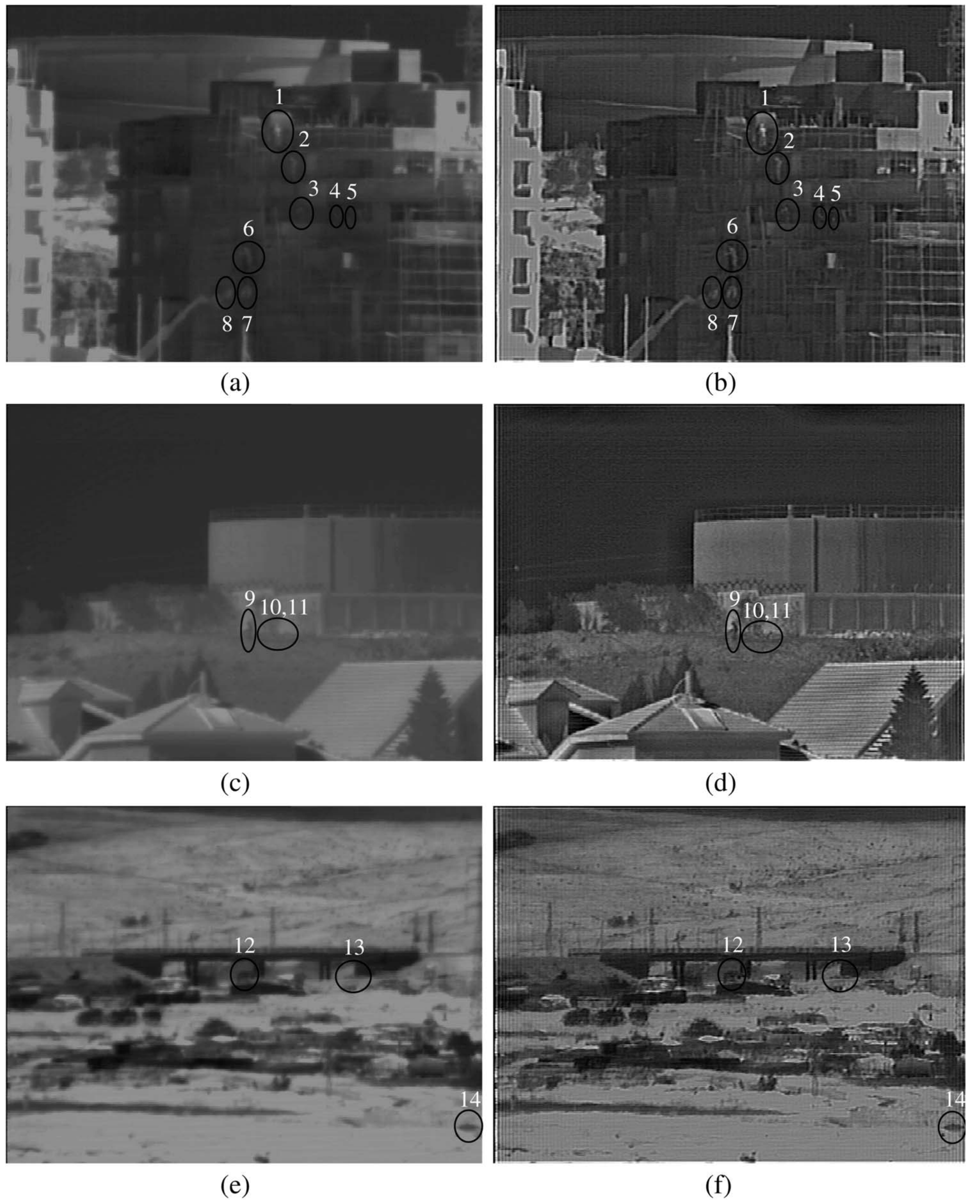

Fig. 1 Sample frames from each degraded and restored video sequence used in the first experiment. Recorded frames are shown on the left, and the corresponding restored frames are shown on the right. The moving objects (described in Table 1) are numbered and marked by circles. The restored versus the degraded videos are available on the Web.[25]

was concluded that restoration is beneficial for contrast-limited imaging (high levels of atmospheric blur and low levels of noise).

Nevertheless, conclusions and target acquisition models obtained from studies with still images are not necessarily appropriate for the dynamic situation (video sequence) where objects are moving (which may often be a more realistic case). The main reasons are the additional temporal information contained in the dynamic situation, and the special (highly sensitive) motion detection mechanism employed in the human visual system.[19,20] For higher target acquisition levels (such as recognition and identification) the visual system can employ dynamic information obtained from the detected target (for instance, the difference between motion properties of humans and vehicles can be employed). Therefore, the effect of image restoration on target acquisition in the static case may not be relevant to the dynamic case.

This research aims mainly to determine how restoration of video sequences degraded by the atmosphere can affect the ability of human observers to acquire moving objects. For this goal, thermal video clips were recorded, in which moving objects (a few kilometers away) such as humans and vehicles could hardly be seen. The frames of these clips were restored, and a comparative experimental procedure was carried out to assess the ability of observers to detect and recognize moving objects from restored and nonrestored clips. A secondary comparative experiment included the specific task of determining whether a moving person carries a thin pole (resembling a rifle) or not. The blurring

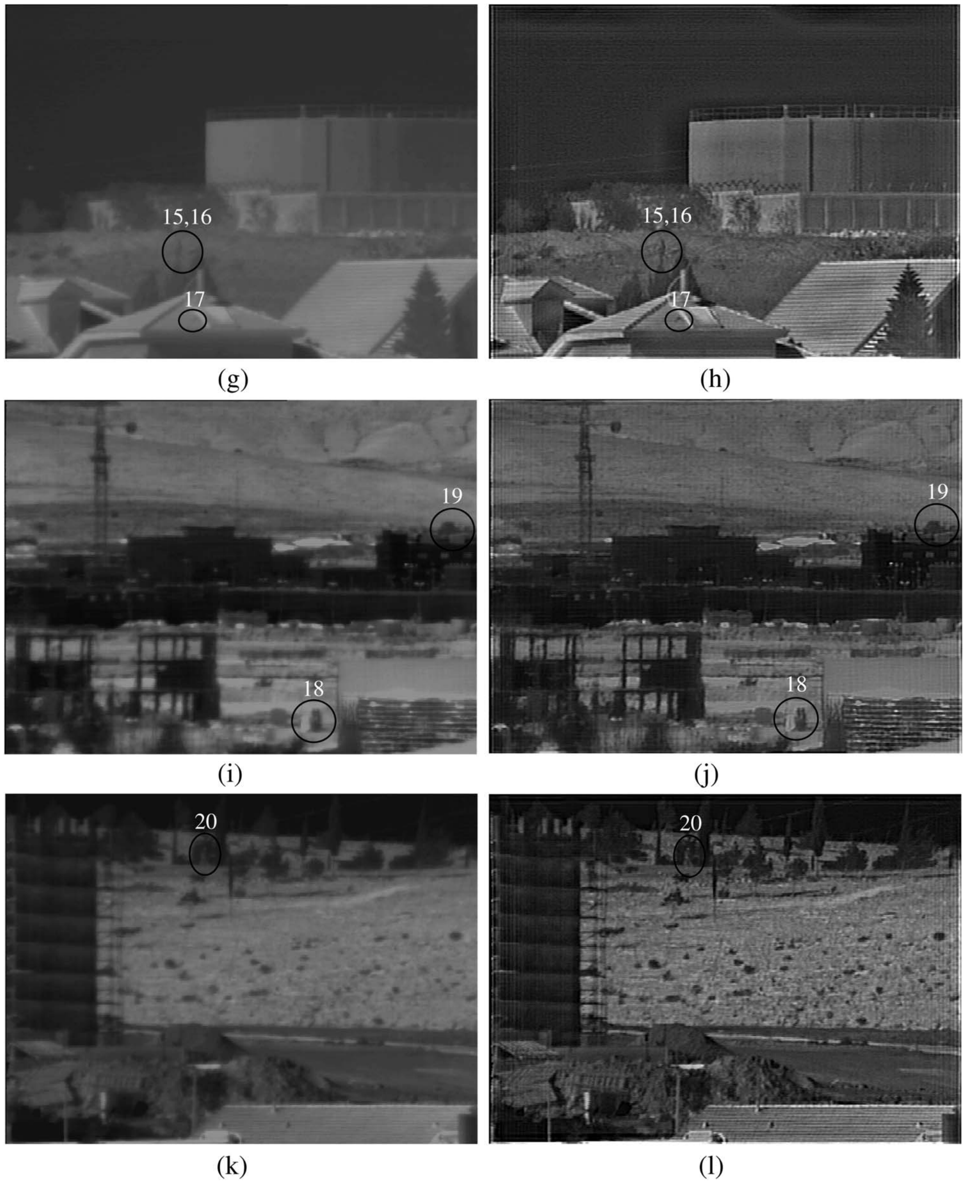

Fig. 1 (*Continued*).

(atmospheric-dominated) MTF required for the restoration filtering process was obtained for each clip using a step-edge response extracted from a static object.[21–23] The blur resulting from the motion of the moving objects themselves was neglected because their velocity in the image plane was low.

The remainder of this paper is organized as follows. Section 2 presents the procedure that was implemented here for restoration of the degraded video sequences. Section 3 describes two perception experiments that were performed in order to measure quantitatively the restoration's effect on the ability of observers to acquire moving objects. The results are presented in Sec. 4. Discussion and conclusions are in Sec. 5.

2 Restoration of the Atmospherically Degraded Video Sequences

Images degraded by the atmosphere are frequently modeled as a convolution between the original (unblurred) image and the atmospheric point spread function (PSF). If the PSF is known, simple deconvolution techniques can be employed in order to restore the image.[1,18] The advantages of these digital techniques are their simplicity, low cost, and practicality of application. The most common deconvolution technique is the Wiener filter, given by the following expression:[1,18]

$$\text{Wiener}(u,v) = \frac{1}{H(u,v)} \frac{|H(u,v)|^2}{|H(u,v)|^2 + \gamma} , \tag{1}$$

where u and v are the spatial frequency coordinates; $H(u,v)$ is the blurring MTF, which is the magnitude of the Fourier transform of the PSF, dominated in our case by the atmospheric effects; and γ is the relation between the spectra of the noise and the original image. Usually this relation is not known and γ is assumed to be a constant inversely proportional to the signal-to-noise ratio (SNR). Restoration

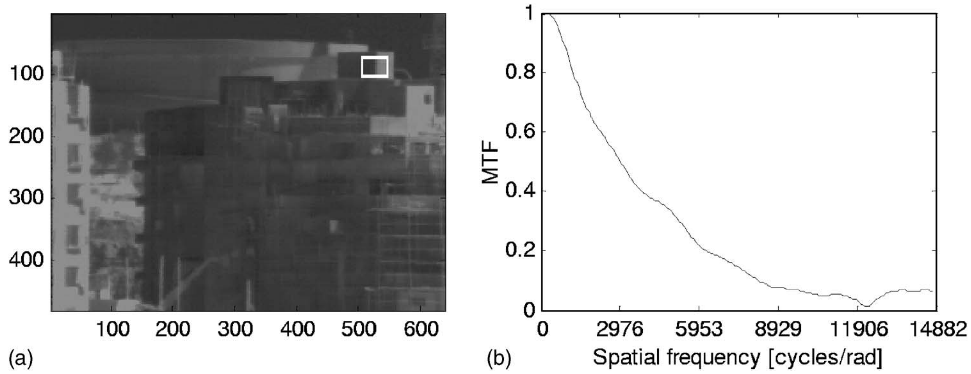

Fig. 2 (a) The degraded step edge selected in the image of Fig. 1(a) (marked by a white rectangle); (b) the MTF extracted from the step edge.

of the original (unblurred) image $\hat{f}(m,n)$ is achieved by multiplying the Fourier transform of the degraded image $G(u,v)$ by the Wiener filter (and then performing inverse Fourier transformation $\Im^{-1}$):

$$\hat{f}(m,n) = \Im^{-1}[G(u,v)\ \mathrm{Wiener}(u,v)]. \tag{2}$$

The atmospheric MTF depends on the imaging and atmospheric conditions, such as the distance to the target, the time of the day, the temperature, the wind speed, the solar flux, and the relative humidity.[1] Models for the weather-predicted atmospheric MTF have been developed previously[1,21] and implemented in image restoration.[22]

Since the atmospheric MTF did not change significantly during each video clip, the MTF required for the restoration filtering was estimated here once for each degraded video sequence from its first image. A single step edge (of a static object) was selected from the image, and an edge spread function was obtained from it. Then, the edge response derivative (line response) was computed in a direction perpendicular to the edge direction. Averaging several line responses along the edge (for noise reduction) produces an estimate of the line spread function (LSF), whose Fourier transform is the estimated overall MTF (assuming isotropic degradation properties).[21–23] The estimated MTF was then used to restore the whole image set in the video clip, according to Eq. (2).

3 Experiment Description

The effect of restoration on the ability of observers to detect and recognize various common moving objects was determined by comparing detection and recognition probabilities from restored and nonrestored videos clips. A secondary comparative experiment included a specific task of determining whether a moving person does or does not carry a thin pole (resembling a rifle).

3.1 *Experiment 1: General Moving-Object Acquisition*

For this experiment, six degraded videos were captured by a staring thermal camera (FOX 720), in the 3- to 5-μm wavelength range, manufactured by CONTROP Ltd.[24] The videos were taken in urban and suburban areas. The climate was semiarid. The line of sight was horizontal, over path lengths of several kilometers (between 2.5 and 6 km), and at an average elevation of 15 m (above the ground). The camera field of view was 0.76 deg horizontal by 0.57 deg vertical. Each video sequence contained 100 frames of 640×480 pixels, at a frequency of 30 Hz (frames per second). Possible moving objects in the video sequences were cars, trucks, bicycles, persons, heavy machines, birds, and dogs. The SNR of each frame in the sequence was around 30 dB.

Figure 1 presents sample frames from each recorded (degraded) and restored video sequence. Recorded frames are shown on the left, while the corresponding restored frames are shown on the right. The moving objects are numbered and marked in those frames by circles. A description of each of them is presented in Table 1. The restored videos were created using the restoration procedure described in Sec. 2 (with $\gamma=0.01$). The step edge, selected in the image of Fig. 1(a), and the MTF extracted from it are shown in Fig. 2(a) and 2(b), respectively. This MTF was

Table 1 A description of the 20 moving objects that appear in the recorded video sequences. These objects are numbered and marked by circles in the sample frames shown in Fig. 1.

Object number	Object type	Object number	Object type
1	Construction worker	11	Dog
2	Construction worker	12	Truck
3	Construction worker	13	Truck
4	Construction worker	14	Car
5	Construction worker	15	Walking person
6	Construction worker	16	Dog
7	Construction worker	17	Bird
8	Construction worker	18	Forklift
9	Walking person	19	Truck
10	Dog	20	Bicycle rider

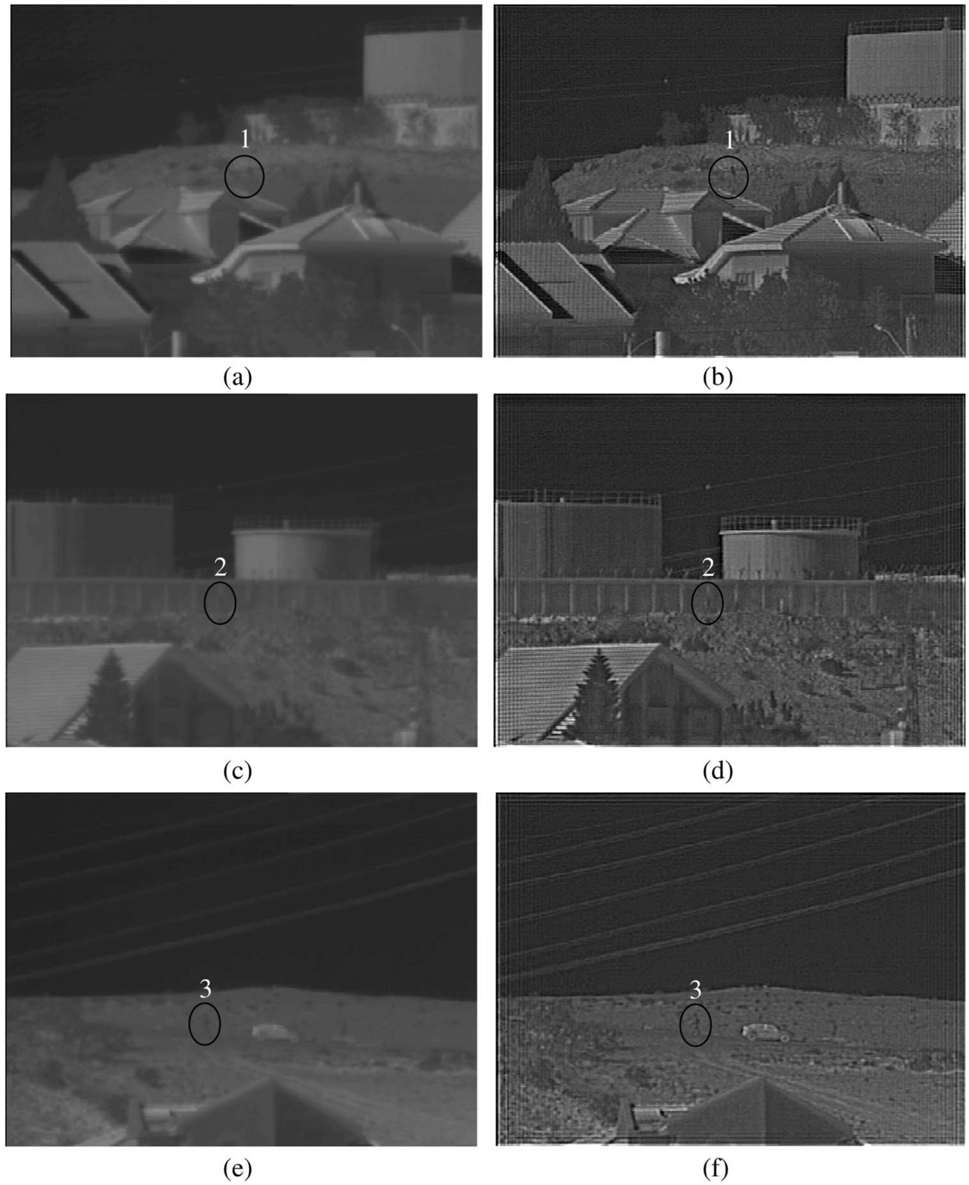

Fig. 3 Sample frames from each degraded and restored video sequence used in the second experiment. Recorded frames are shown on the left, and the corresponding restored frames are shown on the right. The moving persons are numbered and marked by circles. The persons in the top and bottom videos did carry the pole, whereas the person in the middle video did not carry it. The restored versus the degraded videos are available on the Web.[25]

used to restore the video sequence whose sample frame is shown in Fig. 1(a). The imaging and atmospheric conditions that affect this MTF are presented in the first row of Table 2. The condition in which the other videos were recorded are presented in the remainder of this table. The full restored versus degraded videos are available on the Web.[25]

The experiment involved 50 participants (observers). All of them were students (aged 20 to 30 years old), visually acute, and healthy. The observers were seated in a comfortable office chair in front of a computer screen. They were divided arbitrarily into two groups of 25 people each. The participants in each group viewed six videos; some of them were degraded and some of them were restored. In order to prevent a learning effect, a video that was shown degraded to the first group was shown as restored to the second group, and vice versa (i.e., only one version of each video was shown to each observer). The observers were asked to detect and recognize moving objects in the video clips. The observers were not told whether the video contained one, several, or no moving objects. Each video was shown ten times in order to give the observer sufficient time to examine it (the viewing duration was determined through pilot testing).

Each trial began when the experimenter displayed a video. Subjects searched the scene for moving objects. If a moving object was detected, they pointed to it on the display, and verbally described it if they could also recognize it. The experimenter checked the observer's response, and

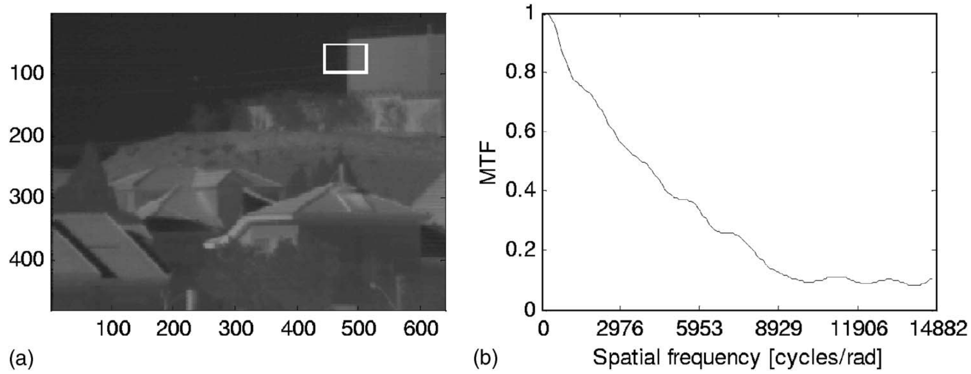

Fig. 4 (a) The degraded step edge selected in the image of Fig. 3(a) (marked by a white rectangle); (b) the MTF extracted from the step edge.

recorded the data as missed or falsely acquired if no object or a wrong object was found. Preliminary explanations as well as a few examples were given to each observer before starting the experiment. The experiment was conducted in a dimly lighted room.

3.2 *Experiment 2: Forced-Choice Detection of a Pole Carried by a Moving Person*

The procedure of the second experiment was similar to that of the first experiment. However, the goal here was more specific. The observers were asked to determine whether a moving person carries a thin pole (resembling a rifle). Three video clips were used here. Sample frames from each degraded and restored video clip are shown in Fig. 3. The persons in the first and in the third videos did carry the pole, whereas the person in the second video did not carry it. The full restored versus the degraded video sequences, in this case, are also available on the Web.[25]

An example of an MTF extraction in this experiment is shown in Fig. 4. The step edge selected in the image of Fig. 3(a), and the MTF extracted from it, are shown in Fig. 4(a) and 4(b), respectively. This MTF was used to restore the video sequence whose sample frame is shown in Fig. 3(a). It can be seen that this MTF is somewhat better (wider) than the MTF shown in Fig. 2(b). This may be explained by the different meteorological conditions (mainly, the lower relative humidity and solar flux, leading to weaker turbulence and aerosol effects).[1,22]

4 Results

4.1 *General Moving-Object Acquisition*

The experimental results, for each of the 20 moving objects and for each case (restored and nonrestored), contain two measures: (1) the *detection probability*, defined as the percentage of observers that detected the moving object; (2) the *recognition probability*, defined as the percentage of observers that recognized the moving object.

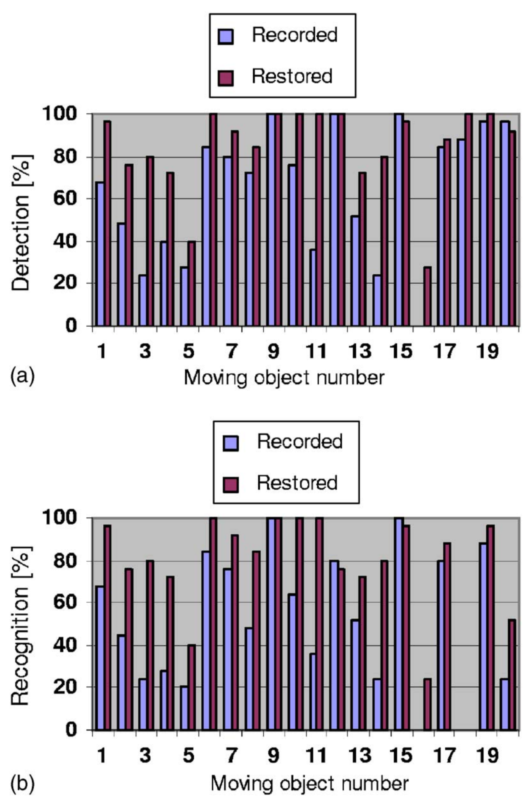

Fig. 5 (a) Detection probability for each of the 20 moving objects marked in Fig. 1, obtained from the recorded and the restored video sequences; (b) same as (a), but for recognition probability.

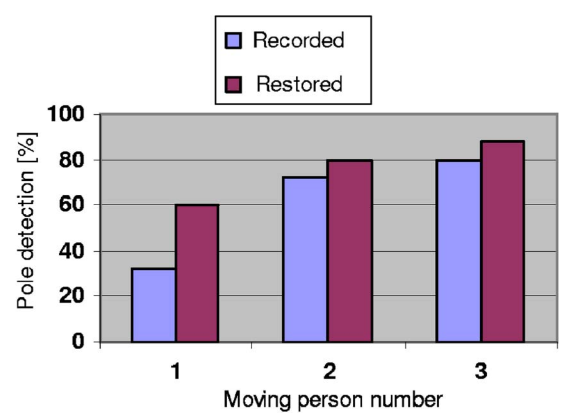

Fig. 6 Results on the pole detection success for each of the three moving persons marked in Fig. 3, obtained from the recorded and the restored video sequences.

Table 2 The imaging conditions and weather parameters measured by a meteorological station when the video sequences were recorded.

Fig.	Date, time	Distance (km)	Temperature (°C)	Relative humidity (%)	Wind speed (m/s)	Solar flux (kW/m^2)
1(a)	20.07.04,08:59	2.5	29.57	53.94	3.97	0.75
1(c)	01.07.04,17:04	3	30.89	36.5	5.21	0.35
1(e)	25.04.04,13:32	6	25.47	33.92	4.27	0.87
1(g)	01.07.04,17:03	3	30.89	36.5	5.21	0.35
1(i)	25.04.04,12:38	5.5	24.77	29.77	1.61	1
1(k)	18.07.04,15:52	2.5	31.21	42.65	4.94	0.59
3(a)	01.07.04,17:23	3	30.52	38.77	5.07	0.28
3(c)	01.07.04,17:17	3	30.52	38.77	5.07	0.28
3(e)	01.07.04,17:09	2.5	30.89	36.5	5.21	0.35

The results are summarized in Fig. 5. In Fig. 5(a), for each of the twenty moving objects marked in Fig. 1, the detection probability from the recorded video is compared with the detection probability from the restored video. Figure 5(b) is the same, but for the recognition probability. In most of the cases, restoration of the video sequences significantly improved the target acquisition probabilities. The average improvement was around 20 percentage points for detection (65% before restoration and 85% afterward), and 24 percentage points for recognition (52% before restoration and 76% afterward).

4.2 *Forced-Choice Detection of a Pole Carried by a Moving Person*

In this experiment, for each of the three moving persons, and for each case (restored and nonrestored), the *pole detection success* was defined as the percentage of observers that correctly determined whether a moving person carried a thin pole.

The Pole detection results are summarized in Fig. 6. It can be seen that the restoration significantly improved the detection success. The average improvement here was around 15 percentage points (61% before restoration and 76% afterward).

5 Discussion and Conclusions

The effect of restoration (deblurring) on the ability of observers to acquire moving objects from thermal video clips was examined here for video sequences degraded by the atmosphere. The video comparisons (recorded versus restored) appear on a Web site.[25] The experimental results indicate that image restoration can significantly improve the ability of observers to acquire moving objects (averages of 20-percentage-point improvement in detection, and 24-percentage-point improvement in recognition). The detection improvement results here agree with results obtained in a previous study,[16] which demonstrated that restoration of (synthetically) degraded static thermal images improves the detection probability by an average of 22 percentage points. Another study with static images showed that restoration may improve the identification probability of (synthetically) degraded characters.[15]

However, although the effect of image restoration on target acquisition capabilities in the static case resembles the effect of restoration in the dynamic case, the difference between the probabilities of the two tasks (detection and recognition) in the dynamic case is quite small. Such a small difference contradicts results obtained with static (still) images. For example, consider moving objects 1 and 3 in Fig. 1(b). The numbers of line pairs included in their narrow dimension are approximately 5 and 3, respectively, and their signal-to-clutter ratios (SCRs) are 5 and 0.9, respectively, where the SCR (for positive contrast targets) is defined as[10–13]

$$\mathrm{SCR} = \frac{\text{maximum target value} - \text{background mean}}{\text{clutter}}, \qquad (3)$$

where

$$\text{clutter} = \left(\frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \right)^{1/2}. \qquad (4)$$

In Eq. (4), N is the number of contiguous cells (blocks) in the scene (the size dimensions of each block were twice the size dimensions of the target), and σ_i is the standard deviation within the i'th block. Therefore, according to the Johnson criteria,[4,5] the detection probabilities for moving objects 1 and 3 in Fig. 1(b) have to be approximately 100% and 65%, respectively, while their recognition probabilities have to be 70% and 3%, respectively (it was suggested[10] to multiply the numbers in the Johnson chart by 2.5 for $\mathrm{SCR} < 1$, which represents high clutter). The results for the dynamic case obtained here (from the restored video) for moving object 1 were 96% for both detection and recognition, while those of moving object 3 were 80%. In another example, the dogs (objects 10 and 11) in Fig. 1(d) have about 6% recognition probability (and 86% detection probability) according to the Johnson criteria (assuming high clutter), while in the dynamic case shown here, all the observers successfully detected and recognized them from the restored video. It is very likely that the motion of the objects, together with their rough shape, contributed to the high recognition probabilities, which would be much lower given only the shapes in static images. This supports the idea that for higher acquisition levels (such as recognition) the visual system employs dynamic information obtained from the detected object, which considerably improves object recognition performance. Contextual cues (such as roads and buildings) may also assist in higher-level target acquisition tasks; however, in most of the video clips used here, such cues are ambiguous, and therefore may hardly be employed.

In another experiment intended to demonstrate a possible practical problem of determining whether a thin pole (resembling a rifle) is or is not carried by a moving person, an average of 15 percentage points improvement was achieved as a result of the restoration of the sequences. The video comparisons (recorded versus restored) in this case are also available on the Web.[25]

Acknowledgment

We greatly appreciate the support of the Wolfson Foundation regarding the thermal imaging equipment.

References

1. N. S. Kopeika, *A System Engineering Approach to Imaging*, 2th ed., SPIE Optical Engineering Press, Bellingham, WA (1998).
2. D. L. Fried, "Optical resolution through a randomly inhomogeneous medium for very long and very short exposures," *J. Opt. Soc. Am.* **56**(10), 1372–1379 (1966).
3. N. S. Kopeika, I. Dror, and D. Sadot, "The causes of atmospheric blur: comment on atmospheric scattering effect on spatial resolution of imaging systems," *J. Opt. Soc. Am. A* **15**(12), 3097–3106 (1998).
4. J. Johnson, "Analysis of image forming systems," in *Proc. Image Intensifier Symp.*, U.S. Army Research and Development Lab., pp. 244–273 (1958).
5. J. Johnson and W. R. Lawson, "Performance modeling methods and problems," in *Proc. IRIS Specialty Group on Imaging* (1974).
6. S. R. Rotman, E. S. Gordon, and M. L. Kowalczyk, "Modeling human search and target acquisition performance: I. First detection probability in a realistic multi-target scenario," *Opt. Eng.* **28**(11), 1216–1222 (1989).
7. S. R. Rotman, E. S. Gordon, and M. L. Kowalczyk, "Modeling human search and target acquisition performance: II. Simulating multiple observers in dynamic scenarios," *Opt. Eng.* **28**(11), 1223–1226 (1989).
8. S. R. Rotman, E. S. Gordon, and M. L. Kowalczyk, "Modeling human search and target acquisition performance: III. Target detection in the presence of obscurants," *Opt. Eng.* **30**(6), 824–829 (1991).
9. R. Melamed, Y. Yitzhaky, and N. S. Kopeika, "Experimental comparison of three target acquisition models," *Opt. Eng.* **37**(7), 1902–1913 (1998).
10. D. E. Schmieder and M. R. Weathersby, "Detection performance in clutter with variable resolution," *IEEE Trans. Aerosp. Electron. Syst.* **19**(4), 622–630 (1983).
11. J. M. Cathcart, T. J. Doll, and D. E. Schmieder, "Target detection in urban clutter," *IEEE Trans. Syst. Man Cybern.* **SMC-19**(5), 1242–1250 (1989).
12. S. R. Rotman, E. S. Gordon, and M. L. Kowalczyk, "Modeling human search and target acquisition performance: IV. Detection probability in the cluttered environment," *Opt. Eng.* **33**(3), 801–808 (1994).
13. K. Brunnstrom, B. N. Schenkman, and B. Jacobson, "Object detection in cluttered infrared images," *Opt. Eng.* **42**(2), 388–399 (2003).
14. R. Vollmerhausen, E. Jacobs, and R. Driggers, "New metric for predicting target acquisition performance," *Opt. Eng.* **43**(11), 2806–2818 (2004).
15. R. Succary, M. Maltz, O. Hadar, S. R. Rotman, and N. S. Kopeika, "Relative effects of distortion and noise on target acquisition: the advisability of image restoration," *Opt. Eng.* **37**(7), 1914–1922 (1998).
16. G. Aviram, S. R. Rotman, and R. Succary, "Evaluating and predicting human detection performance of targets embedded in distorted and restored infrared images," *Infrared Phys. Technol.* **43**(1), 1–15 (2002).
17. N. S. Kopeika, S. R. Rotman, A. Vander, and I. Taig, "Effects of image restoration on target acquisition," *Opt. Eng.* **42**(2), 534–540 (2003).
18. A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, Englewood Cliffs, NJ (1989).
19. J. P. H. van Santen and G. Sperling, "Temporal covariance model of human motion perception," *J. Opt. Soc. Am. A* **1**(5), 451–473 (1984).
20. J. P. H. van Santen and G. Sperling, "Elaborated Reichardt detectors," *J. Opt. Soc. Am. A* **2**(2), 300–321 (1985).
21. I. Dror and N. S. Kopeika, "Experimental comparison of turbulence MTF and aerosol MTF through the open atmosphere," *J. Opt. Soc. Am. A* **12**(5), 970–980 (1995).
22. Y. Yitzhaky, I. Dror, and N. S. Kopeika, "Restoration of atmospherically-blurred images according to weather-predicted atmospheric modulation transfer function (MTF)," *Opt. Eng.* **36**(11), 3064–3072 (1997).
23. B. Cohen, V. Avrin, M. Belitsky, and I. Dinstein, "Generation of a restored image from a video sequence recorded under turbulence effects," *Opt. Eng.* **36**(12), 3312–3317 (1997).
24. CONTROP Precision Technologies Ltd, "Innovative solutions for surveillance and reconnaissance," http://www.controp.co.il (accessed 10 Apr. 2006).
25. Y. Yitzhaky, Home page at Ben-Gurion University, Israel, http://www.ee.bgu.ac.il/~itzik/VideosOE06/ (accessed 10 Apr. 2006).

Oren Haik received his BS degree in physics in 2002 from Ben Gurion University, Israel, where he is currently a PhD student in the Electro-Optics Unit. His research interests are superresolution image reconstruction and acquisition of moving objects.

Yitzhak Yitzhaky received his BS, MS, and PhD degrees in electrical and computer engineering from Ben Gurion University, Israel, in 1993, 1995, and 2000, respectively. From 2000 to 2002 he was a postdoctoral research fellow at the Schepens Eye Research Institute, Harvard Medical School, Boston, Massachusetts. He is currently with the Electro-Optics Unit at Ben Gurion University. His research is mainly in the fields of image restoration, image enhancement, and computer vision.

Biographies and photographs for the other authors not available.