

Research Article

Spatial, Temporal, and Interchannel Image Data Fusion for Long-Distance Terrestrial Observation Systems

Barak Fishbain, Leonid P. Yaroslavsky, and Ianir Ideses

Department of Physical Electronics, School of Electrical Engineering, Faculty of Engineering, Tel Aviv University, Ramat Aviv 69978, Israel

Correspondence should be addressed to Barak Fishbain, barak@eng.tau.ac.il

Received 14 May 2007; Revised 24 October 2007; Accepted 16 December 2007

Recommended by Michael A. Fiddy

This paper presents methods for intrachannel and interchannel fusion of thermal and visual sensors used in long-distance terrestrial observation systems. Intrachannel spatial and temporal fusion mechanisms used for image stabilization, super-resolution, denoising, and deblurring are supplemented by interchannel data fusion of visual- and thermal-range channels for generating fused videos intended for visual analysis by a human operator. Tests on synthetic, as well as on real-life, video sequences have confirmed the potential of the suggested methods.

Copyright © 2008 Barak Fishbain et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Long-distance terrestrial observation systems have traditionally been high-cost systems used in military and surveillance applications. Recent advances in sensor technologies (such as in infrared cameras, millimeter wave radars, and low-light television cameras) have made it feasible to build low-cost observation systems. Such systems are increasingly used in the civilian market for industrial and scientific applications.

In long-distance terrestrial observation systems, infrared sensors are commonly integrated with visual-range charge-coupled device (CCD) sensors. Such systems exhibit unique characteristics—thanks to the simultaneous use of both visible and infrared wavelength ranges. Most of them are designed to give the viewer the ability to reliably detect objects in highly detailed scenes. The thermal-range and visual-range channels have different behaviors and feature different image distortions. Visual-range long-distance observations are usually affected by atmospheric turbulence, which causes spatial and temporal fluctuations to the index of refraction of the atmosphere [1], resulting in chaotic geometrical distortions. On the other hand, thermal channels are less vulnerable to the turbulent effects [2–7] but usually suffer from substantial sensor noise and reduced resolution as compared to their visual-range counterparts [8]. One way to overcome those problems is to apply data fusion techniques.

In recent years, a great deal of effort has been put into multisensor fusion and analysis. Available fusion techniques may be classified into three abstraction levels: pixel, feature, and semantic levels. At the pixel level, images acquired in different channels are combined by considering individual pixel values or small arbitrary regions of pixels in order to make the fusion decision [9–12]. At the feature-level fusion, images are initially subjected to feature-driven segmentation in order to produce a set of regions with various properties that are used to determine which features from which image are to be included in the fused image [13–16]. Semantic-driven methods transform all types of input data into a common variable space, where the data is fused [17]. In developing new data fusion methods, it is possible to extract different types of features from different channels before fusing them—a concept that the above-mentioned methods fail to do since they apply the same fusion criteria to all input channels. Unlike previous methods, the method in this paper applies sensor-specific criteria to each channel before fusing the data.

The development of fusion algorithms using various kinds of pyramid/wavelet transforms has led to numerous pixel- and feature-based fusion methods [18–24]. The motivation for the pyramid/wavelet-based methods emerges from observations that the human visual system is primarily sensitive to local contrast changes, for example, edges and corners.

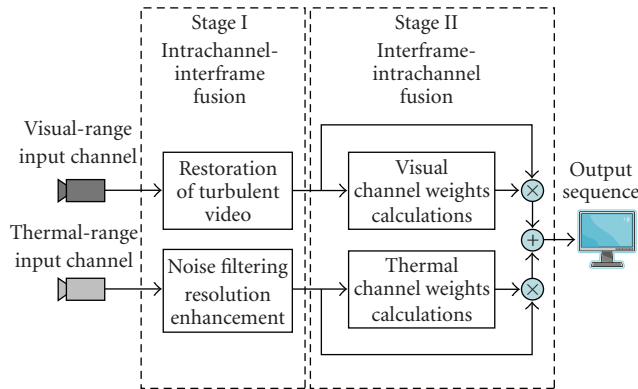


FIGURE 1: Fusion algorithm flow diagram: visual-range spatial and temporal image fusion for image stabilization and super-resolution (upper branch) and thermal-range spatial and temporal image fusion for image denoising and resolution enhancement (bottom branch).

However, observation systems are characterized by diversity in both location and size of the target of interest; therefore, rigid decomposition, which is characteristic for multiresolution fusion methods, turns out to be less suitable for long-range observation tasks [13, 16, 25].

This paper describes a video processing technology designed for fusion of thermal- and visual-range input channels of long-range observation systems into a unified video stream intended for visual analysis by a professional human operator. The suggested technology was verified using synthetic, as well as real-life, thermal and visual sequences from a dedicated database [26, 27]. The database contained video sequences acquired in the near vicinity of the camera as well as sequences of sites located as far as 25 km away.

2. SYSTEM DESCRIPTION

The proposed video processing technology is outlined in a schematic block diagram in Figure 1. The outlined method is based on a two-stage process. The first stage consists of *intrachannel-interframe* processing methods used to perfect each input channel independently. Each channel's processing method is designed according to the sensor's specific limitations and degradations. For the visual-range channel processing, spatial-temporal fusion is implemented for compensating turbulence-induced image geometrical distortions, as well as super-resolution above the visual sensor's sampling rate. For the thermal channel, spatial-temporal fusion is implemented for sensor noise filtering and resolution enhancement by means of 3D (spatial-temporal) local adaptive filtering. These visual- and thermal-range intrachannel fusion schemes are thoroughly described in Sections 3 and 4, respectively. The second stage is *interframe-intrachannel* fusion. At this stage, thermal- and visual-range channel image frames, corrected and enhanced, are fused frame by frame using a multiple-criteria weighted average scheme with locally adapted weights. The second stage is detailed in Section 5.

3. VISUAL-CHANNEL IMAGE FUSION FOR IMAGE STABILIZATION AND SUPER-RESOLUTION

3.1. Channel characterization and processing principles

In remote sensing applications, light passing long distances through the troposphere is refracted by atmospheric turbulence, causing distortions throughout the image in the form of chaotic time-varying local displacements. The effects of turbulence phenomena on imaging systems were widely recognized and described in the literature, and numerous methods were proposed to mitigate these effects. One method for turbulence compensation is adaptive optics [28, 29]. Classical adaptive optics, which uses a single deformable mirror, provides correction for a limited field of view (FOV). Larger FOV corrections can be achieved by several deformable mirrors optically conjugated at various heights [30–33]. In modeling images with distortion caused by atmospheric turbulence, light from each point in the acquired scene is assumed to possess a slightly different tilt and low-order aberration, and it can be modeled by convolving a raw image with a space-variant pseudorandom point spread function [34]. Therefore, multiconjugate adaptive optics techniques require complex structure and reconstruction processes, making them unsuitable for operational systems.

Other turbulence compensation methods use an estimation of modulation transfer function (MTF) of the turbulence distortions [35–38]. The drawback of those methods is that they require some prior knowledge about the observed scene, which is often unavailable.

Methods that require no prior knowledge are suggested in [3–7, 13, 39–42]. The principal idea is to use, for reconstructing distortion-compensated image frames, an adaptively controlled image resampling method based on the estimate of image local displacement vectors. Using those concepts, turbulence compensation algorithms which preserve genuine motion in the scene are suggested in [43–47].

In this paper, these techniques are further elaborated and improved upon in order to obtain super-resolution in addition to turbulence compensation. The new techniques are used as an interframe-interchannel fusion mechanism for the visual-range input channel. As shown in the flow diagram, presented in Figure 2, visual-range video processing consists of three processing stages: (i) estimation of the reference frames, (ii) determination of the motion vectors for all pixels in image frames and motion vector analysis for real-motion extraction, and (iii) generation of stabilized frames with super-resolution and preservation of the real motion. Those stages are thoroughly described, respectively, in Sections 3.2, 3.3, and 3.5.

3.2. Estimation of the reference frames

The reference images, which are the estimation of the stable scene, are obtained from the input sequence. The reference images are needed for measuring the motion vectors for each current video frame. One way to measure the motion vectors of each image frame is by means of elastic registration with

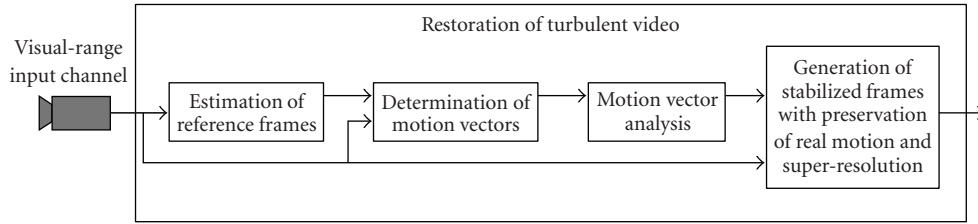


FIGURE 2: Flow diagram of video processing in visual-range channel.

the previous frame. However, this method does not allow reliable discrimination of real movements in the scene from those caused by the atmospheric turbulence. For this task, estimation of the stable scene is required. We adopt the approach of [48] and suggest using a pixelwise rank gray-level filtering of video frames in a temporal sliding window for generating such an estimation intended to serve as the reference frame. The use of rank smoothing filters such as median or alpha-trimmed-mean filters is substantiated in two ways. First, distribution of a light beam propagating through a turbulent atmosphere has a mean of zero. This means that the center of the deflection is located at the same point that the light beam would have hit if there was no turbulence present. Therefore, the statistical expectation of the gray-level values is relatively close to the mean of the trace of the same pixel's values over a long period of time. The reason for using a rank filter instead of a mean filter is the fact that for moving objects that accommodate a pixel for a short period of time, the gray-level distribution for this pixel is found to be tail-heavy. When applying rank filters, the distribution tails will be eliminated from evaluation of estimated values. Rank filtering might result in resolution degradation. This will be dealt with in subsequent processing stage, which suggests resolution enhancement (see Section 3.4). It was found experimentally that the use of a temporal median filter provides an acceptable solution in terms of both stable scene evaluation quality and computational efficiency [49, 50].

The length in time of the filter temporal window, N , is determined by the correlation interval of turbulence effect over time; that is, the longer the time correlation of the turbulence effect is, the larger the size of the temporal sliding window becomes. Our experiments have shown that for correlation intervals of atmospheric turbulence of order of seconds, temporal window size should be of the order of 100 frames for frame rate of 25 frames per second.

Temporal pixelwise median filtering for estimating the stable scene as a reference image is illustrated in Figure 3, where part (a) presents a sample frame taken from a turbulent distorted sequence acquired with a camera acquiring images in size of 4 times common intermediate format (4CIF—704 × 576 pixels) in a frame rate of 25 frames per second (the sequence can be found at [26]). Figure 3(b) depicts the estimation of the stable scene calculated by temporal median over 117 frames. One can notice that the geometrical distortions in Figure 3(a), in particular around the dune's rim on the left-hand side of the image, are removed from the stabilized estimation in Figure 3(b).

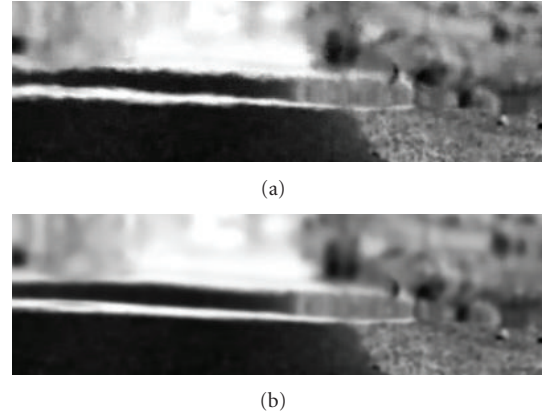


FIGURE 3: Temporal median rank filter as an estimation of the stable scene: (a) is a frame extracted from a turbulent degraded real-life video sequence, while (b) is the stable scene estimation using pixelwise temporal median.

In principle, the median filtering in a moving time window presents high computational complexity. Utilizing a fast recursive method for median filtering [48, 49] enables a real-time implementation at common video rates.

3.3. Motion vector analysis for real-motion discrimination

In order to avoid distortion of real motion due to the turbulence compensation process, real motion should be detected in the observed scene. To this end, a real-time two-stage decision mechanism is suggested in [44, 45, 49]. This method forms, for each pixel in each incoming frame, a real-motion separation mask ($RMSM_{(\vec{p})}$), where \vec{p} is the space-time coordinate vector, $\vec{p} = [x, y, t]$. At the first step, a straightforward fast algorithm is utilized for extracting areas, such as background, that are most easily classified as stable. In most cases, the majority of the image pixels are extracted at this stage. Those parts are not further processed. Only the pixels, which were not tagged as stable at the first phase, are dealt with at the second phase. The second stage uses a more sophisticated though more time-consuming algorithm.

3.3.1. Stage I

At the first stage, the gray-level difference between the current value of each pixel of the incoming frame and its

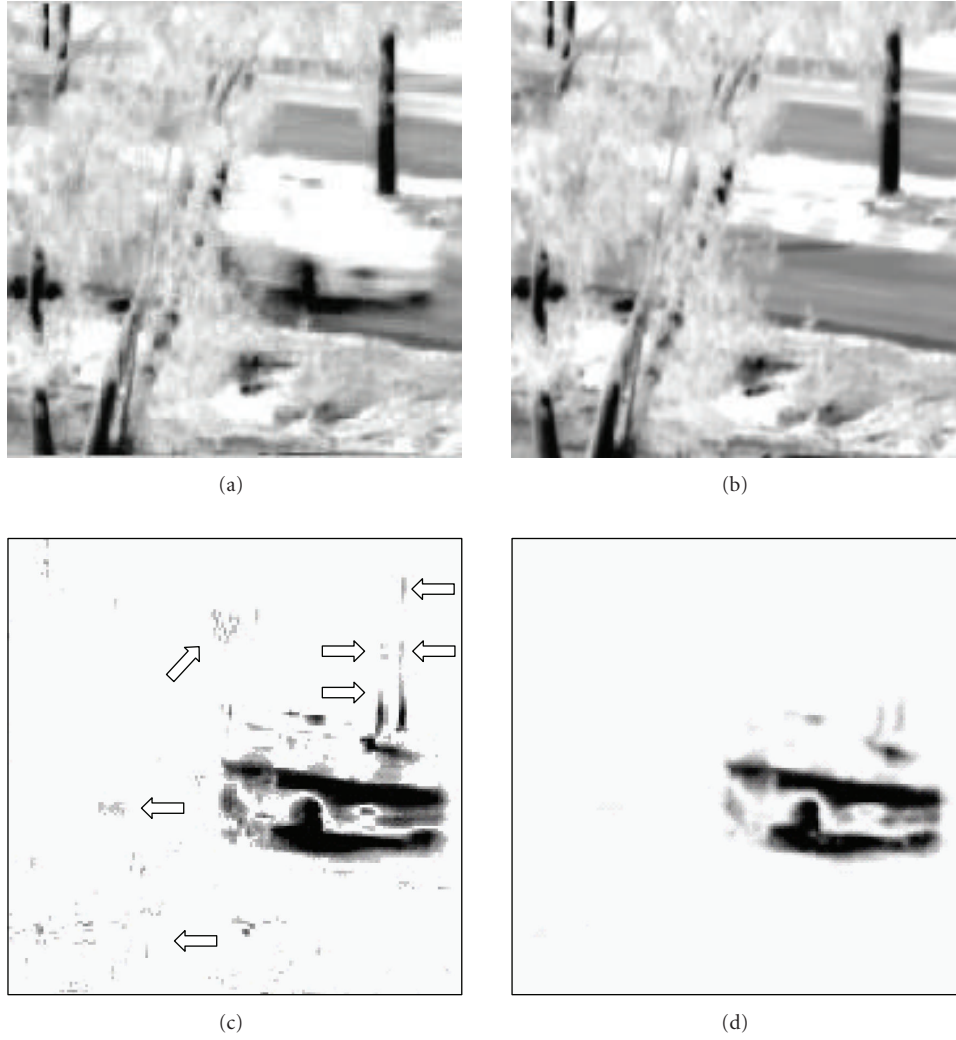


FIGURE 4: Motion extraction and discrimination: (a) is a frame that has been extracted from a real-life turbulent degraded thermal sequence; (b) depicts the stable scene estimation computed over 117 frames; (c) is the result of real-motion extraction of phase I, while (d) is the result of real motion extracted after phase II.

temporal median is calculated as “real-motion measure.” This is referred to as distance-from-median (DFM) measure:

$$\text{DFM}_{\vec{p}} = I_{\vec{p}} - \bar{I}_{\vec{p}}, \quad (1)$$

where t is an index of the current processed frame, and $\bar{I}_{\vec{p}}$ is its median over the temporal window (Ω) centered at \vec{p} :

$$\bar{I}_{\vec{p},\Omega} = \text{MED}\{I_{\vec{p},\Omega}\}. \quad (2)$$

If the distance, $\text{DFM}_{(\vec{p})}$, is below a given predefined threshold, the pixel is considered to be of a stationary object. The threshold is determined by exploiting the observer’s limitation of distinguishing between close gray-levels. In real-time applications, the threshold is an adjustable parameter of the algorithm that can be adjusted by the observer in course of the scene visual analysis. Background areas, which do not belong to a moving object nor are located near edges, will be resolved in this way. All other pixels that are not resolved at this stage are processed at the next one.

Figure 4(a) presents a frame extracted from a real-life turbulent degraded video sequence with moving objects (see [26]). Figure 4(b) is the reference frame computed by applying elementwise temporal median filtering over 117 frames, as described in Section 3.2. Figure 4(c) represents darker pixels, which were tagged as real-motion at the first stage. As one can see, while this first stage detects most of the background pixels as such, it produces some “false alarms” (marked with arrows). Figure 4(d) represents, in darker tones, pixels that contain real motion. As one can see, the real-motion detection errors are eliminated at the second processing stage, which is described in the following section.

3.3.2. Stage II

The second stage improves real-motion detecting accuracy at the expense of higher computational complexity; however, it handles a substantially smaller number of pixels. This stage

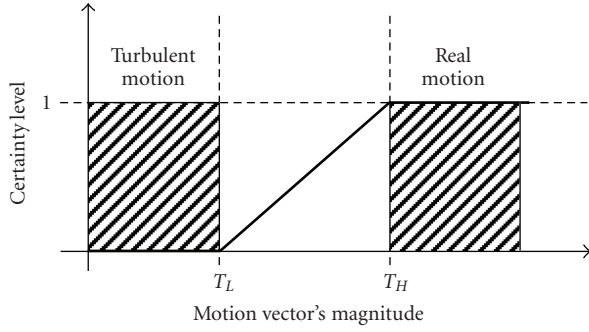


FIGURE 5: Magnitude-driven mask (MDM) certainty level as a function of the motion vector's magnitude.

uses, for motion-driven image segmentation, techniques of optical flow [41, 42, 51–59] and its statistical analysis.

In its simplest form, the optical flow method assumes that it is sufficient to find only two parameters of the translation vector for every pixel. The motion vector $\{\Delta \hat{x}, \Delta \hat{y}\} = \{x - \hat{x}, y - \hat{y}\}$, for every pixel, is the vector difference between the pixel's location in the original image $I_{(x,y)}$ and its location in the reference image $\hat{I}_{(\hat{x},\hat{y})}$. For the subsequent processing stages, the translation vector is presented in polar coordinates as $\{M_{(\vec{p})}, \theta_{(\vec{p})}\}$ through its magnitude $\{M_{(\vec{p})}\}$ and angle $\{\theta_{(\vec{p})}\}$, which are subjected to cluster analysis for discriminating real movement against that caused by atmospheric turbulence.

Real-motion discrimination through motion field magnitude distribution

For cluster analysis of the motion vector magnitude distribution function for all pixels (x, y) in a particular frame, each pixel in the frame is assigned with a certainty grade, the magnitude-driven mask ($MDM_{(\vec{p})}$). The $MDM_{(\vec{p})}$ measure ranges between 0 and 1 and characterizes the magnitude-based likelihood that particular pixel belongs to objects in a real motion. Figure 5 presents the certainty as a function of the motion vector's magnitudes. It is natural to assume that minor movements are caused by turbulence, and larger movements correspond to real motion. The intermediate levels comprise motion vectors' magnitudes upon which concise decision cannot be made. The magnitudes' thresholds T_L and T_H are application-dependent parameters and can be set by the user. Based on the analysis of our visual database, in our experiments with real-life videos, T_L and T_H were set to 2 and 4 pixels, respectively.

Real-motion discrimination through motion field's angle distribution

A pixel's motion discrimination through angle distribution is achieved by means of statistical analysis of the angle component of the motion field. For the neighborhood of each pixel, the variance of angles is computed. As turbulent motion has

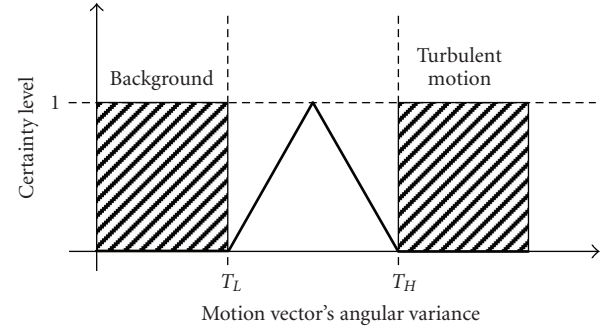


FIGURE 6: Angle-driven mask (ADM) certainty level as a function of the motion vector's local spatial standard deviation.

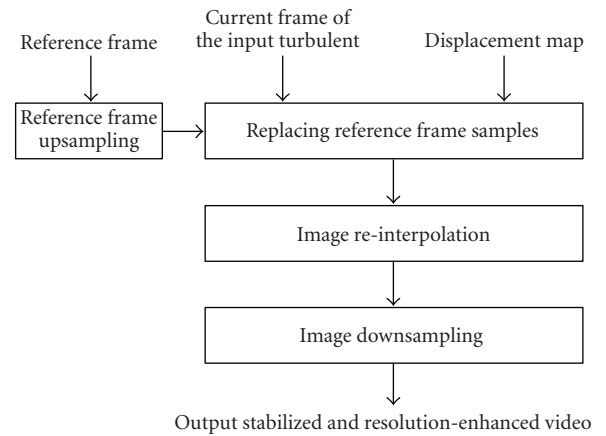


FIGURE 7: Flow diagram of the process of generation of stabilized frames with super-resolution.

fine-scale chaotic structure, motion field vectors in a small spatial neighborhood distorted by turbulence have considerably large angular variance. Real motion, on the other hand, has strong regularity in its direction and therefore the variance of its angles over a local neighborhood will be relatively small.

The neighborhood size, in which the pixel's angular standard deviation is computed, should be large enough to secure a good statistical estimation of angle variances, and as small as possible to reliably localize small moving objects. In our experiments with the dedicated real-life database [26, 27], it was found that neighborhood's sizes of 11×11 and 15×15 present a reasonable compromise.

As a result of variance analysis, each pixel is assigned with an angle-driven mask ($ADM_{(\vec{p})}$), which presents an angle distribution-based likelihood that this pixel belongs to an object in a real motion. This is illustrated in Figure 6. Real moving objects have bounded angular variances, T_L and T_H . Both turbulent and background areas should be regarded as stable. This means that pixels with angular variance smaller than T_L or higher than T_H are regarded as stationary. Those values are set by the observer. In our experiments with real-life video, they were set to $(\pi/6)^2$ and $(\pi/3)^2$, respectively.

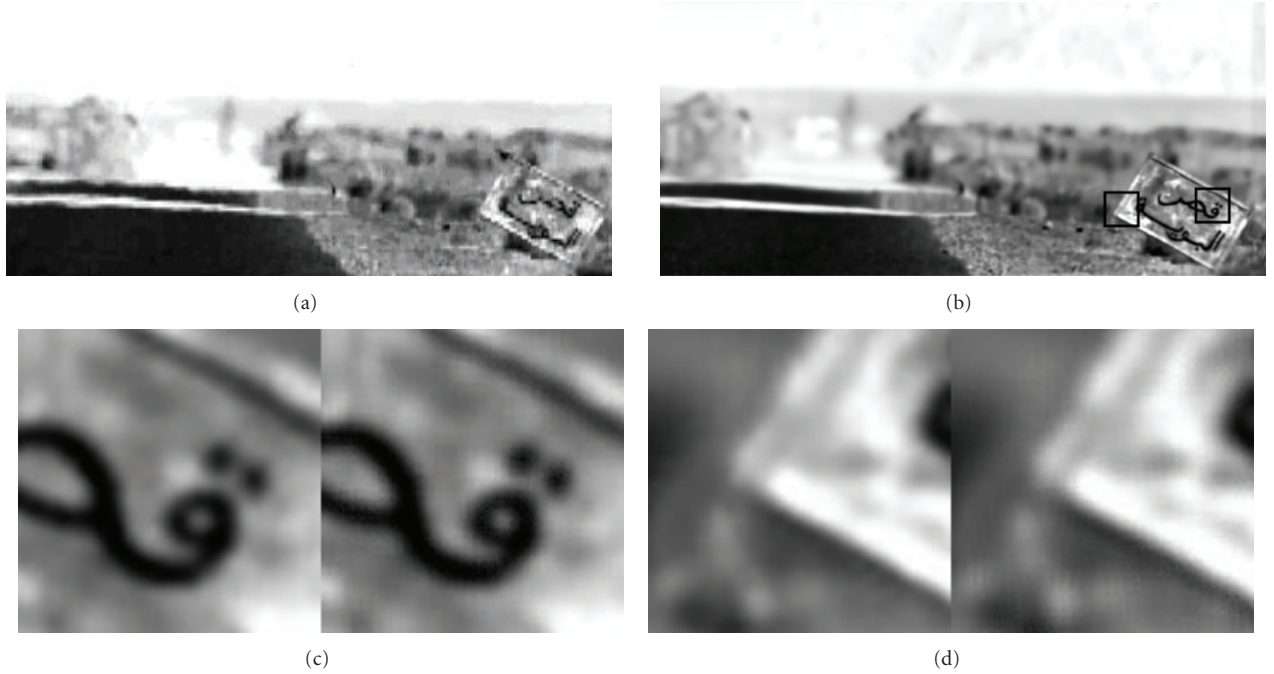


FIGURE 8: Super-resolution through turbulent motion—visual-range sequence. (a) shows a raw video frame; (b) shows a super-resolved frame generated from a visual-range turbulent degraded real-life video; (c)-(d) are the magnified fragments marked on (b)—the left-hand side shows the fragment with simple interpolation of the initial resolution and the right-hand side shows the fragment with super-resolution.

Real-motion separation mask

Having both $MDM_{(\vec{p})}$ and $ADM_{(\vec{p})}$, a combined real-motion separation mask ($RMSM_{(\vec{p})}$) is formed as follows:

$$RMSM_{(\vec{p})} = \begin{cases} ADM_{(\vec{p})}, & \left| ADM_{(\vec{p})} - \frac{1}{2} \right| > \left| MDM_{(\vec{p})} - \frac{1}{2} \right|, \\ MDM_{(\vec{p})}, & \text{otherwise.} \end{cases} \quad (3)$$

Equation (3) implies that the ADM measure is more accurate than the MDM when the term $|ADM_{(\vec{p})} - 1/2|$ has a higher value than $|MDM_{(\vec{p})} - 1/2|$. In this case, the ADM measure will be used; otherwise the MDM value will be applied. Figure 4(d) presents the $RMSM_{(\vec{p})}$, where real moving objects are represented in darker pixels.

3.4. Generation of super-resolved stabilized output frames

In turbulence-corrupted videos, consequent frames of a stable scene differ only due to small atmospheric-turbulence-induced movements between images. As a result, the image sampling grid defined by the video camera sensor may be considered to be chaotically moving over a stationary image scene. This phenomenon allows for the generation of images with larger number of samples than those provided by the camera if image frames are combined with appropriate resampling [2, 60–63].

Generally, such a super-resolution process consists of two main stages [2, 64–68]. The first is determination, with sub-

pixel accuracy, of pixel movements. The second is combination of data observed in several frames in order to generate a single combined image with higher spatial resolution. A flow diagram of this stage of processing is shown in Figure 7.

For each current frame of the turbulent video, inputs of the process are a corresponding reference frame, obtained as a temporal median over a time window centered on the current frame, and the current frame displacement map. The latter serves for placing pixels of the current frame, according to their positions determined by the displacement map, into the reference frame, which is correspondingly upsampled to match the subpixel accuracy of the displacement map. For upsampling, different image interpolation methods can be used. Among them, discrete sinc-interpolation is the most appropriate as the one with the least interpolation error and may also be computed efficiently [69]. As a result, output stabilized and enhanced in its resolution frame is accumulated. In this accumulation process, it may happen that several pixels of different frames are to be placed in the same location in the output-enhanced frame. In order to make the best use of all of them, these pixels must be averaged. For this averaging, the median of those pixels is computed in order to avoid the influence of outliers that may appear due to possible errors in the displacement map.

After all available input frames are used in this way, the enhanced and upsampled output frame contains, in positions where there were substitutions from input frames, accumulated pixels of the input frames and, in positions where there were no substitutions, interpolated pixels of the reference frame. Substituted pixels introduce to the output frame high frequencies outside the baseband defined by the original

sampling rate of the input frames. Those frequencies were lost in the input frames due to the sampling aliasing effects. Interpolated pixels that were not substituted do not contain frequencies outside the baseband. In order to finalize the processing and take full advantage of the super-resolution provided by the substituted pixels, the following iterative reinterpolation algorithm was used. This algorithm assumes that all substituted pixels accumulated, as described above, are stored in an auxiliary replacement map containing pixel values and coordinates. At each iteration of the algorithm, the discrete Fourier transform (DFT) spectrum of the image obtained at the previous iteration is computed and then zeroed in all of its components outside the selected enhanced bandwidth, say, double of the original one. After this, inverse DFT is performed on the modified spectrum, and corresponding pixels in the resulting image are replaced with pixels from the replacement map, thus producing an image for the next iteration. In this process, the energy of the zeroed outside spectrum components can be used as an indicator when the iterations can be stopped.

Once iterations are stopped, the output-stabilized and resolution-enhanced image obtained in the previous step is subsampled to the sampling rate determined by selected enhanced bandwidth and then subjected to additional processing aimed at camera aperture correction and, if necessary, denoising.

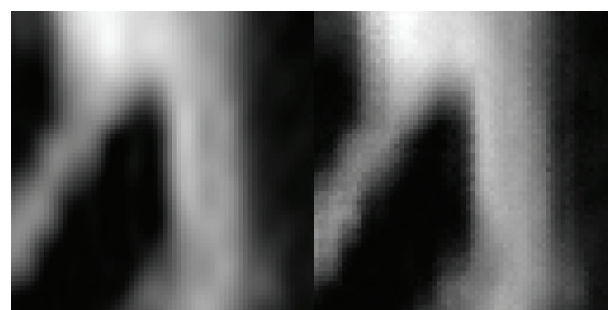
Figure 8 illustrates the feasibility of the method. Figure 8(a) is a frame extracted from turbulent degraded real-life sequence, while Figure 8(b) is its super-resolved stable one. Figures 8(c) and 8(d) are magnified fragments from Figure 8(b). The fragments are marked with black boxes on Figure 8(a). In both Figures 8(c) and 8(d), the original fragments are shown on the left-hand side, while the super-resolved fragments are shown on the right-hand side.

Atmospheric turbulence also affects thermal-range videos. Figure 9 demonstrates application of the method to intermediate infrared wavelengths ($3\text{--}8\ \mu\text{m}$), turbulent video sequence. Figure 9(a) shows an example of a super-resolved frame generated from the thermal sequence (whose stable reference corresponding frame is presented in Figure 4(b)). The marked fragments of Figure 9(a) are presented in Figures 9(b) and 9(c), in which fragments with initial resolution are given on the left-hand side, while the super-resolved fragments, extracted from Figure 9(a), are given on the right-hand side.

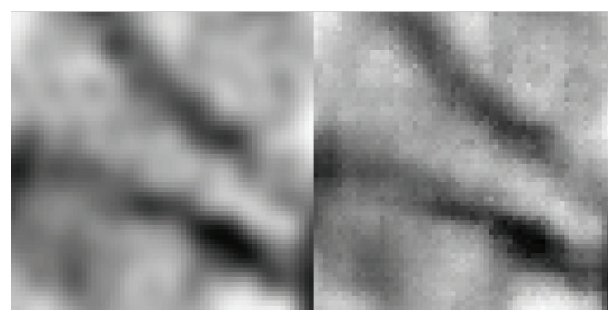
In the evaluation of the results obtained for real-life video, one should take into account that substantial resolution enhancement can be expected only if the camera fill-factor is small enough. The camera fill-factor determines the degree of lowpass filtering introduced by the optics of the camera. Due to this low-pass filtering, image high frequencies in the baseband and aliasing high-frequency components that come into the baseband due to image sampling are suppressed. Those aliasing components can be recovered and returned back to their true frequencies outside the baseband in the described super-resolution process, but only if they have not been lost due to the camera low-pass filtering. The larger the fill-factor is, the heavier the unrecoverable resolution losses will be.



(a)



(b)



(c)

FIGURE 9: Super-resolution through turbulent motion. (a) presents a super-resolved frame generated from a thermal-range turbulent degraded real-life video; (b)-(c) are the magnified fragments marked on (a)—the left-hand side shows the fragment with simple interpolation of the initial resolution and the right-hand side shows the fragment with super-resolution.

For quantitative evaluation of the image resolution enhancement achieved by the proposed super-resolution technique, we use a degradation measure method described in [70]. The method compares the variations between neighboring pixels of the image before and after lowpass filtering. High variation between the original and blurred images means that the original image was sharp, whereas a slight

TABLE 1: Quantitative evaluation of the super-resolved images. The degradation grade is ranging from 0 to 1, which are, respectively, the lowest and the highest degradations.

	Original	Super-resolved
Visual-range video (see Figure 8)		
Entire original image (see Figure 8(a)) versus entire super-resolved image (see Figure 8(b))	0.627	0.3987
Fragment (b)	0.8287	0.5926
Fragment (c)	0.8474	0.5493
Thermal-range video (see Figure 9)		
Entire original image (see Figure 4(b)) versus entire super-resolved image (see Figure 9(a))	0.7323	0.5452
Fragment (b)	0.8213	0.6802
Fragment (c)	0.6630	0.5058

variation between the original and blurred images means that the original image was already blurred. The comparison result presented in a certain normalized scale as in the image degradation measure ranged from 0 to 1 is shown in [70] to very well correlate with subjective evaluation of image sharpness degradation with 0 corresponding to the lowest sharpness degradation and 1 to the highest degradation. The described method might be biased at the presence of substantial noise. To eliminate this, in this example, both visual- and thermal-range sequences were acquired in lighting and sensor conditions to minimize the noise level. Table 1 shows the results of the comparison, using this measure, between images presented in Figures 8 and 9 and their individual fragments before and after applying the described super-resolution process. It is clearly seen from the table that the super-resolved images present better quality in terms of this quantitative quality measure.

3.5. Generation of output frames with preservation of real motion

The algorithm of generating the stabilized output frame $F_{(\tilde{p})}$ is defined by

$$F_{(\tilde{p})} = \bar{I}_{(\tilde{p})} \bullet \left\{ 1 - \left[\text{RMSM}_{(\tilde{p})} \bullet (1 - \text{DFM}_{(\tilde{p})}) \right] \right\} + I_{(\tilde{p})} \bullet \left[\text{RMSM}_{(\tilde{p})} \bullet (1 - \text{DFM}_{(\tilde{p})}) \right], \quad (4)$$

where “ \bullet ” denotes elementwise matrix multiplication, $\bar{I}_{(\tilde{p})}$ is the estimation of the stable scene as described in Section 3.2 or the super-resolved stable scene as described in Section 3.4, $I_{(\tilde{p})}$ is the current processed frame (t), DFM is the “distance-from-median” mask described in Section 3.3.1, and RMSM is the real-motion separation mask detailed in Section 3.3.2.

Figure 10 illustrates results of the described turbulence compensation process. Figure 10(a) is a frame extracted from a real-life turbulent degraded image (see [26]), and Figure 10(b) shows the stabilized frame. As one can notice, the motion of the flying bird located near the upper-left corner of the plaque on the right-hand side of the

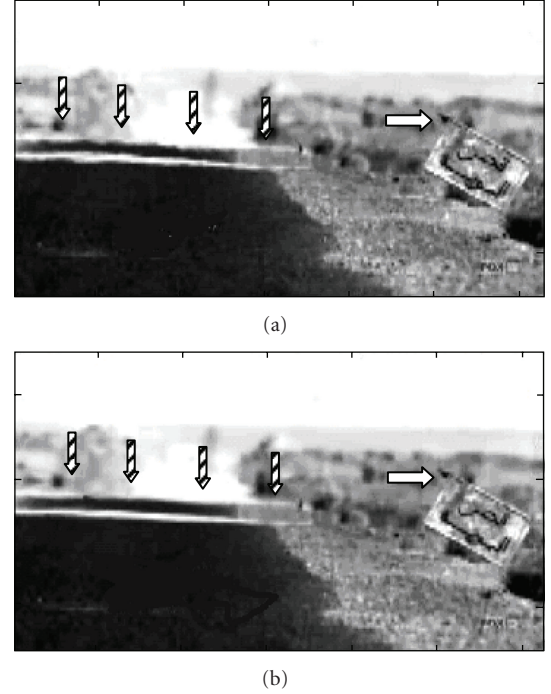


FIGURE 10: Turbulence compensation: (a) is a frame extracted from a turbulent degraded sequence, while (b) is the corresponding turbulent compensated frame.

frame (marked with a white arrow) is retained, while the turbulence-induced distortion of the still rim situated on the frame's left-hand side (marked with striped arrows) is removed.

4. THERMAL-RANGE IMAGE FUSION FOR DENOISING AND RESOLUTION ENHANCEMENT

As detailed in Section 2, the first stage of the fusion algorithm consists of *intrachannel-interframe* processing. The visual-range channel processing was described in Section 3. The thermal channel processing for sensor noise filtering and resolution enhancement by means of 3D (spatial-temporal) local adaptive filtering is depicted in this section.

4.1. Channel characterization and filtering principle

Thermal sensors suffer from substantial additive noise and low image resolution. The thermal sensor noise can be described in terms of the spatial (x, y) and temporal (t) axes using 3D noise models [71, 72]. Resolution degradation is associated with the finite aperture of the sensor sensitive cells.

Video frames usually exhibit high spatial and temporal redundancy that can be exploited for substantial noise suppression and resolution enhancement. In [48, 73], a sliding window transform domain two-dimensional (2D) filtering for still image restoration is described. In this paper, an extension of this method to three-dimensional (3D) spatial/temporal denoising is suggested for thermal image sequence processing [13].

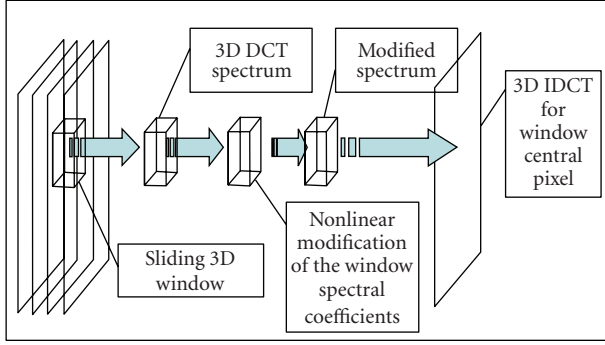


FIGURE 11: Sliding cube 3D transform domain filtering.

A block diagram of the filtering is shown in Figure 11. For each position of the window, the DFT or the discrete cosine transform (DCT) of the signal volume within the spatial/temporal window is recursively computed from that of the previous position of the window. Recursive computation substantially reduces the filter's computational complexity [73, 74]. The signal's spectral coefficients are then subjected to soft or hard thresholding according to

$$\begin{aligned}\hat{\beta} &= \frac{\beta \max \{0, (|\beta|^2 - \text{Thr})\}}{\lambda |\beta|^2}, \\ \hat{\beta} &= \frac{\beta [\text{sign}(|\beta|^2 - \text{Thr}) + 1]}{2\lambda},\end{aligned}\quad (5)$$

where β and $\hat{\beta}$ represent input and modified transform coefficients, correspondingly, and λ represents the set of coefficients of the frequency response of the camera (spatial and temporal indices are omitted for the sake of brevity). The division of image spectra by frequency response of the camera is the implementation of camera aperture correction by means of pseudoinverse filtering [48].

The window spectra which are modified in this way are then used to generate the current image sample of the output, by means of the inverse transform of the modified spectrum. Note that, in this process, the inverse transform need not be computed for all pixels within the window, but only for the central sample, since only the central sample has to be determined in order to form the output signal.

4.2. Tests and results

For the purpose of testing, two sets of artificial movies were generated, having various levels of additive Gaussian noise. The first artificial test movie contains bars with different spatial frequencies and contrasts, and the second is of a fragment of a text. Figure 12 shows results of applying a 3D filtering for image denoising. The images in Figures 12(a) and 12(b) correspond to the original frames. Figures 12(c) and 12(d) show the corresponding frames originating from a sequence possessing temporal and spatial random additive noise. Figures 12(e) and 12(f) show corresponding frames obtained using 3D filtering. Numerical results on noise suppression capability of the filtering obtained for the test images, in terms of

residual filtering error, are provided in Table 2. These images and the table data clearly demonstrate the high noise suppression capability of the filtering stage. Full videos can be found in [27].

The results of 3D filtering of real-life video sequences are illustrated in Figure 13. Figures 13(a) and 13(c) are frames taken from real-life thermal sequences; Figures 13(b) and 13(d) are the corresponding frames from the filtered sequences. As one can see, while noise is substantially suppressed, object edges in the scene are not only well preserved but even sharpened—thanks to aperture correction implemented in the filtering in addition to noise suppression.

5. INTERCHANNEL INTRAFRAME FUSION

5.1. Fusion principles

In accordance with the linear theory of data fusion for image restoration [75], the interchannel fusion process is implemented as a linear combination of thermal- and visual-range channel frames:

$$I_p^{\text{CompositeOutput}} = I_p^{\text{Thermal}} \cdot w_p^{\text{Thermal}} + I_p^{\text{Visual}} \cdot w_p^{\text{Visual}}, \quad (6)$$

where I_p^{Thermal} and I_p^{Visual} are pixel intensities in thermal and visual channels, correspondingly, and w_p^{Thermal} and w_p^{Visual} are the related channel weights.

Several methods for assigning weight coefficients for data acquired from dissimilar sensors' modalities are known [16–19, 25]. Those methods suggest applying a single metric for each channel. This means that the weights are extracted using only one feature of the acquired images. As the aim of the fusion process in the visual observation systems is presenting a superior output (in human observation terms), typically the visual output quality of observation systems is defined by several criteria, such as edge preservation, noise presence, and how active are different areas of the scene. This implies that a composite assignment of weight coefficients, based on those criteria, has to be formulated. To this end, we compose both w_p^{Thermal} and w_p^{Visual} of three sets of weights as

$$w_p = \frac{1}{3} (w_p^{\text{VI}} + w_p^{\text{Noise}} + w_p^{\text{Motion}}). \quad (7)$$

The first set of weights w_p^{VI} is associated with user-defined “visual importance” (“VI”) in the thermal and visual channels. The second set of weights w_p^{Noise} suggests using noise estimation techniques in the fusion process for noise reduction in the fused output. Many observation system applications are intended to evaluate activity of a scene, for example, a car entering a driveway or people in motion. Therefore, the third set of weights w_p^{Motion} is designed to represent the activity level of the scene. Methods for computation of w_p^{VI} , w_p^{Noise} , and w_p^{Motion} are described in Sections 5.2.1, 5.2.2, and 5.2.3, respectively.

TABLE 2: Standard deviation of residual filtering error (RMSE, in gray levels for the image gray-level range of 0–255) for 4-bar test image sequences.

3D window size	Output RMSE (standard deviation of input noise, 30 gray levels of 255)	Output RMSE (standard deviation of input noise, 20 gray levels of 255)
$3 \times 3 \times 3$	13.9	8.86
$5 \times 5 \times 5$	12	7.48

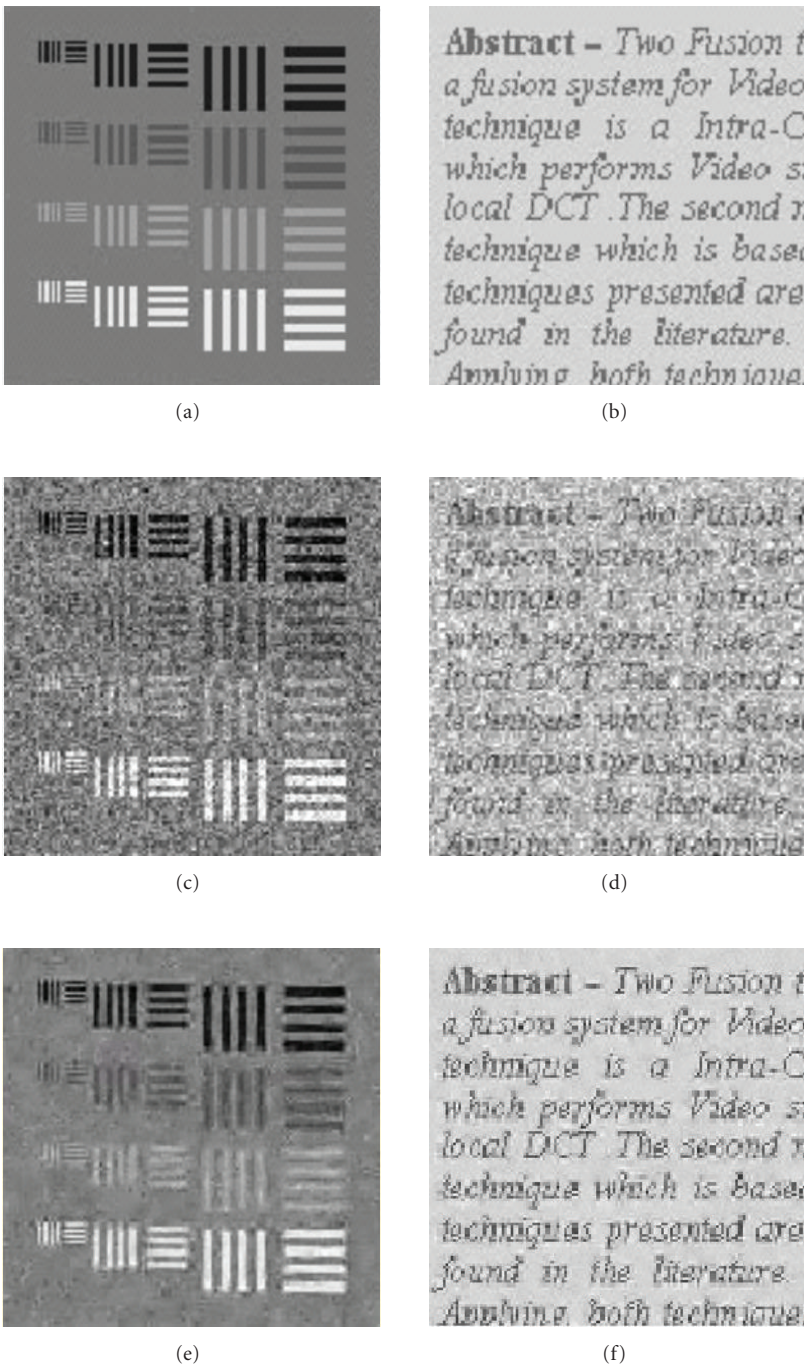


FIGURE 12: Local adaptive 3D sliding window DCT domain filtering for denoising video sequences. Figures (a) and (b) show noise-free test image frames. Figures (c) and (d) are corresponding images with additive white Gaussian noise. Figures (e) and (f) are corresponding filtered frames.

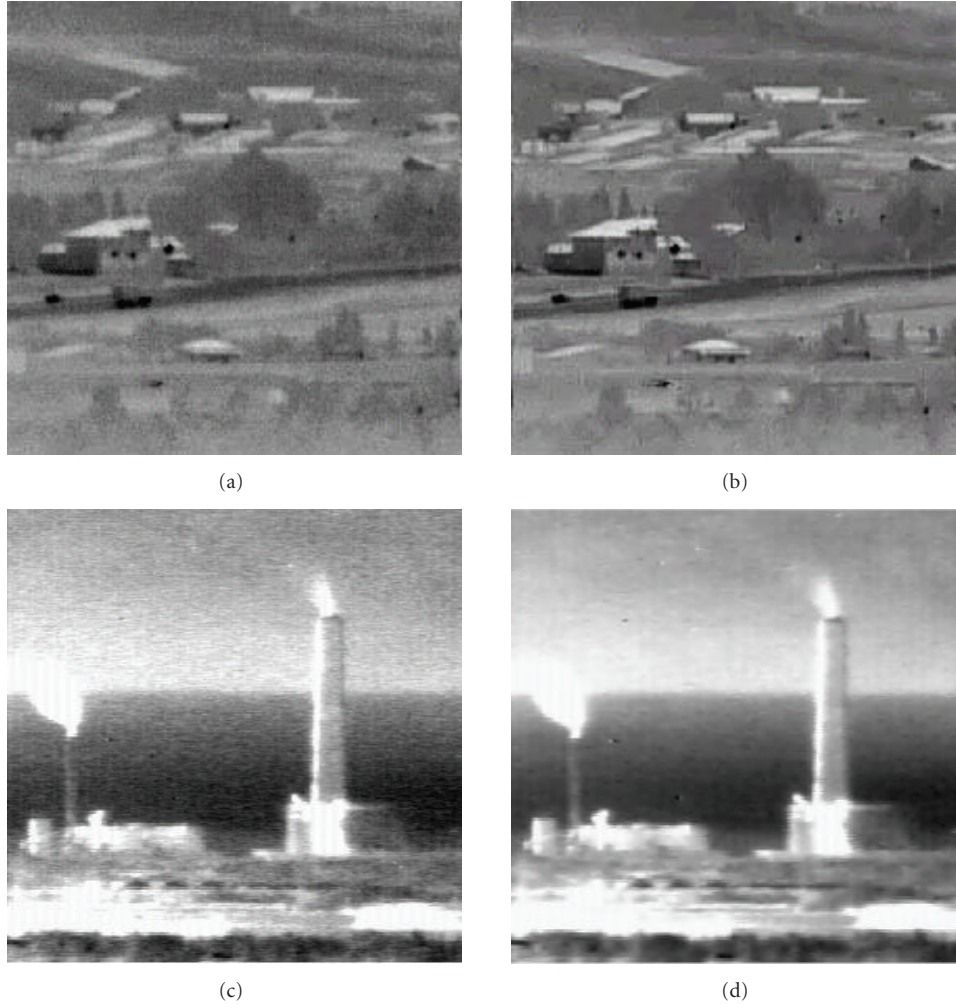


FIGURE 13: Sliding $5 \times 5 \times 5$ 3D window denoising and aperture correction of thermal real-life video sequence. (a) and (c) are frames taken from real-life thermal sequences; (b) and (d) are the corresponding frames from the filtered sequences.

5.2. Weights specification

5.2.1. Visual importance weights

(1) Visual channel

Weighing fused images with local weights determined by *visual importance* of sequences was suggested in [13, 25]. The local spatial/time variances were suggested as the visual-range weights. However, local-variance weighing has some limitations associated with it. First, neighborhoods with only moderate changes in the visual images are assigned with zero weights and are omitted from the fused output even if they may be important visually. Other limitations are due to frequent changes of the same sample's neighborhood variance in sequential frames. This may cause flickering in the output fused sequence and make the observation task more difficult. This is most common in background areas and in areas which are highly affected by noise. As the presence of noise manifests itself in higher local variances, using this criterion will boost noise presence in the output fused image.

The flickering effect can be significantly reduced by using temporal smoothing of the weights. The noise boost presented by the visual-channel VI-weights is dealt with in Section 5.2.2. In order to cope with the omission of visual data, we propose to compute visual VI-weights as follows:

$$w_p^{\text{Visual}} = g_1^{\text{Visual}} + g_2^{\text{Visual}} \cdot \sigma_p^V, \quad (8)$$

where w_p^{Visual} are the computed weights in location, (\vec{p}) and σ_p^V are local intensity standard deviations computed in a spatial running window centered in \vec{p} , and g_1^{Visual} and g_2^{Visual} are user-defined scalars that secure nonzero contribution of the channel in uniform areas, where the local standard deviation is small.

Scalars g_1^{Visual} and g_2^{Visual} are set by the user and are application-dependent. For instance, if the user would like to emphasize edges and changes in higher frequencies, he would choose large g_2^{Visual} with relation to g_1^{Visual} . However, this might result in flickering output and omission of visual information of uniform areas from the composite output.

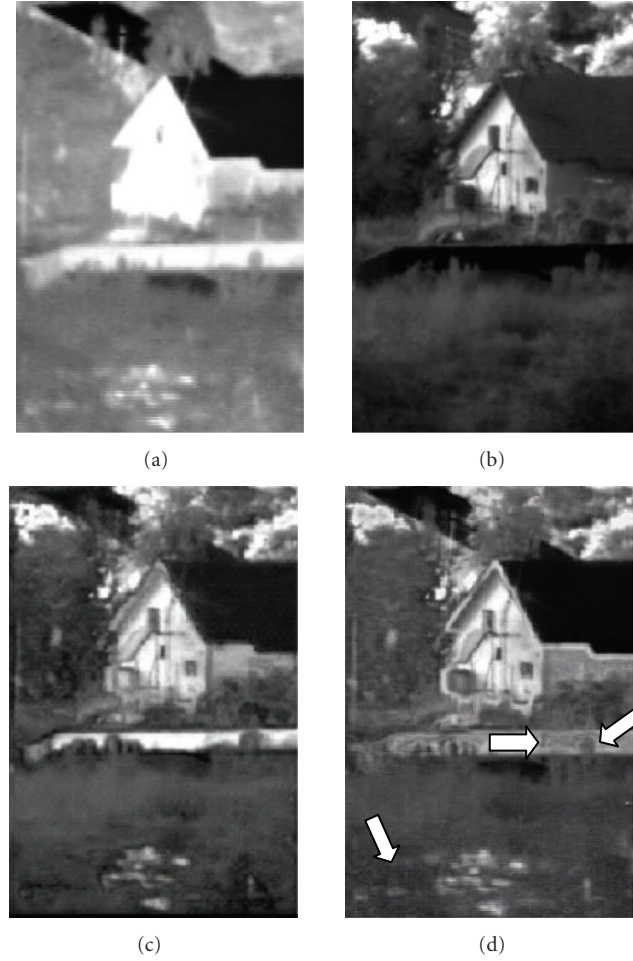


FIGURE 14: Fusing visual- and thermal-range channel images using two described methods for computing the VI-weights. Figure (c) is the fused image using variance and distance from the local average as weights for the visual- and thermal-range channels, respectively. Figure (d) presents the same input images (a) and (b) fused using VI-weights as defined by (8) and (10) for the visual and thermal channels.

Based on the test videos used, g_1^{Visual} and g_2^{Visual} were selected to be 1 and 10, respectively.

(2) Thermal channel

The thermal channel VI-weights are specified under the assumption that *importance* of pixels in the thermal image is determined by their contrast with respect to their background and they are defined as [13, 25]

$$w_p^{\text{IR}} = |I_p^{\text{IR}} - \bar{I}_p^{\text{IR}}|, \quad (9)$$

where I_p^{IR} is the input frame from the thermal channel and \bar{I}_p^{IR} is its local average estimates.

As images are usually highly inhomogeneous, the weight for each pixel should be controlled by its spatial neighborhood. The selection of the size of the neighborhood is application-driven. In our implementation, it is user-selected and is defined as twice the size of the details of objects of interest. Different techniques can be used for estimating the average over the pixel neighborhood, such as local-mean and

median [76]. Both methods have shown good results in experiments without undesired artifacts.

As for background or smooth areas, a similarity can be drawn between the visual and thermal weights. In both weighing mechanisms, those areas are assigned to have weights equal to zero and are omitted from the output image. Therefore, it is suggested to use the user-defined scalars, g_1^{IR} and g_2^{IR} , in the same manner. This brings (9) into the following format:

$$w_p^{\text{IR}} = g_1^{\text{IR}} + g_2^{\text{IR}} \cdot |I_p^{\text{IR}} - \bar{I}_p^{\text{IR}}|. \quad (10)$$

The considerations for setting the values of g_1^{Visual} and g_2^{Visual} are similar to the ones, described under Section 5.2.1(1), used to set g_1^{Visual} and g_2^{Visual} .

We illustrate the described VI-controlled interchannel image fusion in Figure 14. Figure 14(c) shows a fused image of Figure 14(a) (thermal channel) and Figure 14(b) (visual-range channel), using pixel's neighborhood variance σ_p^V for the computation of the visual weighing matrix and difference from local mean $|I_{k,l}^{\text{IR}} - \bar{I}_{k,l}^{\text{IR}}|@$ as the thermal one, while

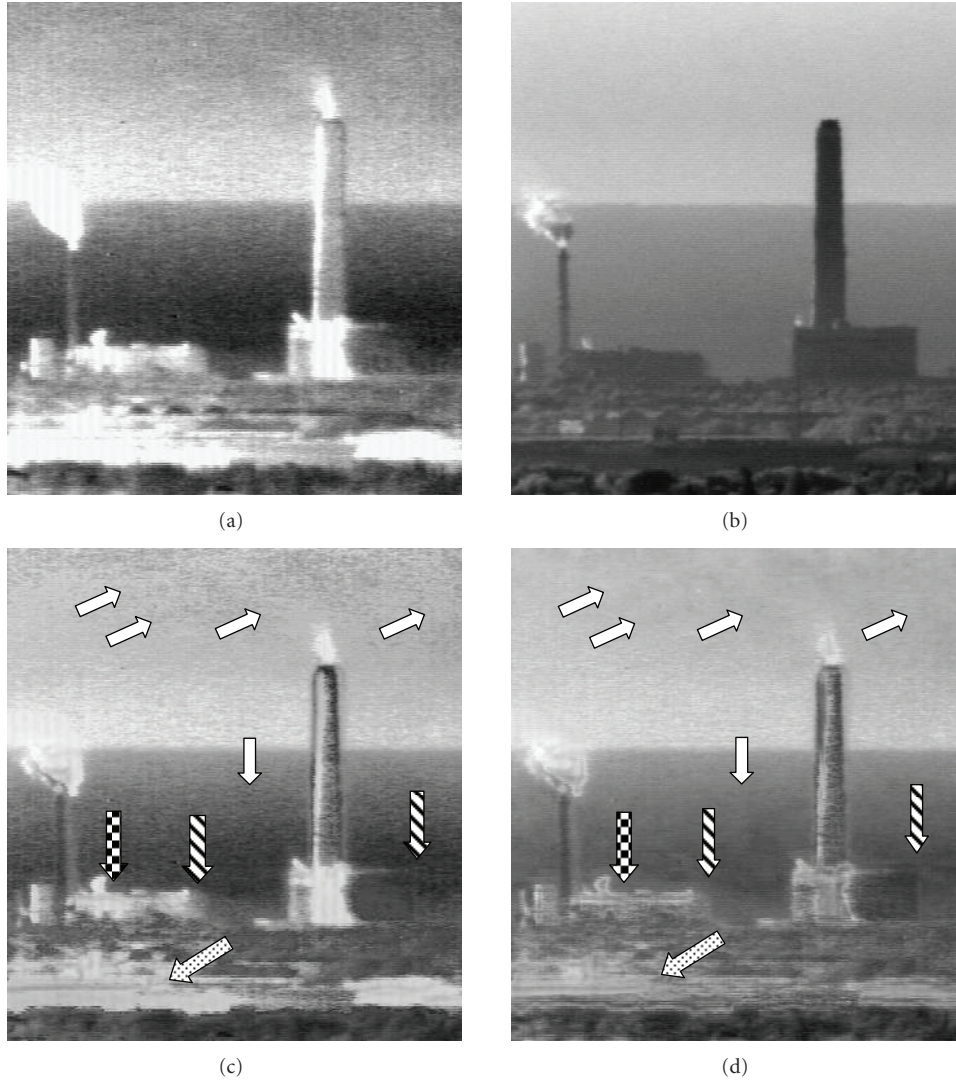


FIGURE 15: Fusion applying noise-defined weights. Figure (c) is the fused output of Figures (a) and (b) using VI-weights. Figure (d) represents the same input images fused using VI-weights and noise-defined weights.

Figure 14(d) shows the same input frames fused applying g_1 and g_2 on each channel.

The brick wall in the image is built from bricks with poles of cement holding them together. The location of the poles might become crucial for military and civil-engineering applications. While it is quite difficult to see the poles in Figure 14(c), they are clearly noticeable in Figure 14(d). This is also true for the hot spots that appear in the field in the lower-left part of the image. Those spots are seen in more detail in Figure 14(d).

5.2.2. Noise-defined weights

We assume that sensor noise acting in each channel can be modeled as additive white signal-independent Gaussian noise [8, 77, 78]. It follows from the linear theory of data fusion for image restoration [79] that noise-defined weights

assigned to each sample of the input channels should be proportional to the signal-to-noise ratio (SNR):

$$w_p^{V, \text{Noise}} = \frac{\sigma_p^V / N_p^V}{\sigma_p^V / N_p^V + \sigma_p^{\text{IR}} / N_p^{\text{IR}}}, \quad (11)$$

$$w_p^{\text{IR}, \text{Noise}} = \frac{\sigma_p^{\text{IR}} / N_p^{\text{IR}}}{\sigma_p^V / N_p^V + \sigma_p^{\text{IR}} / N_p^{\text{IR}}},$$

where σ_p^V and σ_p^{IR} are the image local standard deviations in visual- and thermal-range channels, and N_p^V and N_p^{IR} are the corresponding channel noise standard deviations for the sample neighborhood centered at position \vec{p} .

Two methods for evaluating the noise level of every pixel over its neighborhood may be considered: (i) estimation of the additive noise variance through local autocorrelation function in a running window; (ii) estimation of the additive

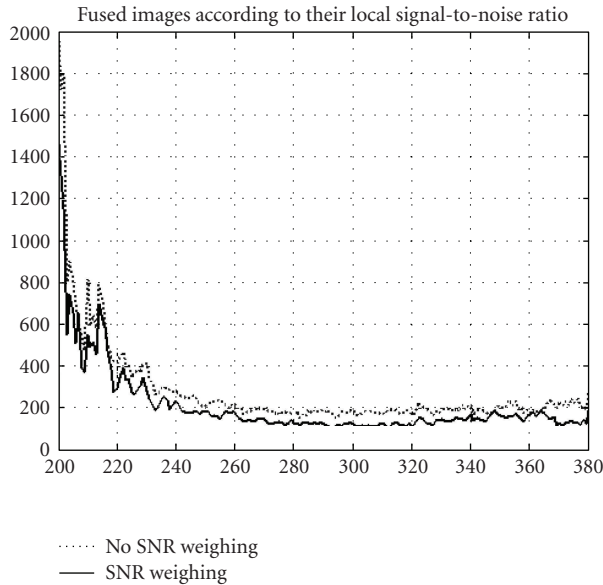


FIGURE 16: Rowwise mean power spectra of image fused with (solid) and without (dotted) SNR weighing.

noise variance through evaluation of noise floor in image local spectra in a running window [76, 79].

The estimation of the noise level yields a quantity measure for each sample. The lower the pixel's noise level estimate is, the heavier the weight assigned to it will be:

$$w_p^{\text{Noise}} = \frac{1}{N_p}. \quad (12)$$

Figure 15 illustrates weighing fused images according to their local SNR estimates. Figure 15(c) presents the output when fusing Figures 15(a) and 15(b), applying only VI-weights. Figure 15(d) shows the same two input frames fused while applying VI-weights along with noise-defined weights. The evaluation of the additive noise variance was performed through analysis of image local correlation function. Local SNRs were evaluated in a moving window of 11×11 pixels.

In evaluating images of Figure 15, observation professionals have pointed out what follows.

- (1) Background noise reduction (see areas pointed by blank arrows \Rightarrow): on video sequences, this type of noise tends to flicker and annoy the user observing the video for several hours.
- (2) Edges preservation (see areas indicated by striped arrows \Rightarrow): one can easily notice how the building edges are better presented in Figure 15(d).
- (3) Details are better presented. The target of interest might not be the power plant itself, but its surrounding. Observing Figure 15(d) reveals more details and allows the observer to make better decisions (dotted arrows \Rightarrow). Additionally, more details can be extracted from the buildings themselves. The chessboard arrows \Rightarrow point to the building floors which are spotted in Figure 15(d) and not in Figure 15(c).

Quantitative assessment of the noise levels in Figures 15(c) and 15(d) is presented in Figure 16 that shows the row-wise average power spectra of Figures 15(c) and 15(d) which were fused with (solid) and without (dotted) noise-defined weights. One can see from this figure that noise floor in the fused image generated with noise-defined weights is substantially lower.

5.2.3. Motion-defined weights

Observation system applications frequently require evaluation of activity of a scene in time. This section suggests a fusion mechanism, which assigns moving objects in the scene with heavier weights. To accomplish that, a quantitative real-motion certainty-level measurement denoting the confidence level of whether this sample is a part of a real moving object, as described in Section 3.3, is used to assign input samples with a weight proportional to their motion level.

Figure 17 presents a typical road scene where a car (marked with striped arrows) is followed by a bus or a truck (marked with blank arrows). The car happens to be very hot, and therefore it exhibits itself as a bright spot in the thermal channel (see Figure 17(a)). The truck is bigger and cooler than the car, and it manifests itself in the visual channel. Both the car and the truck are assigned with higher motion weights in the corresponding channels. The motion-vector-defined weight matrices of the thermal and visual images are shown in Figures 17(a) and 17(b), respectively, where heavier weights are shown in darker pixels.

Figure 18(a) shows an image that was fused using noise-defined and VI-weights, as described in Sections 5.2.1 and 5.2.2, with no motion taken into consideration. It might be difficult to track the vehicles in these images. Modification of the fusion scheme to include motion-defined weights resulted in the output fused image presented in Figure 18(b) in which both car and truck can be spotted much easier than in Figure 18(a) (see marked arrow).

6. CONCLUSIONS

A new multichannel video fusion algorithm, for long-distance terrestrial observation systems, has been proposed. It utilizes spatial and temporal intrachannel-interframe and intrachannel fusion. In intrachannel-interframe fusion, new methods are suggested for

- (1) compensation for visual-range atmospheric turbulence distortions,
- (2) achieving super-resolution in turbulence-compensated videos,
- (3) image denoising and resolution enhancement in thermal videos.

The former two methods are based on local (elastic) image registration and resampling. The third method implements real-time 3D spatial-temporal sliding window filtering in the DCT domain.

The final interchannel fusion is achieved through a technique based on the local weighted average method with weights controlled by the pixel's local neighborhood visual

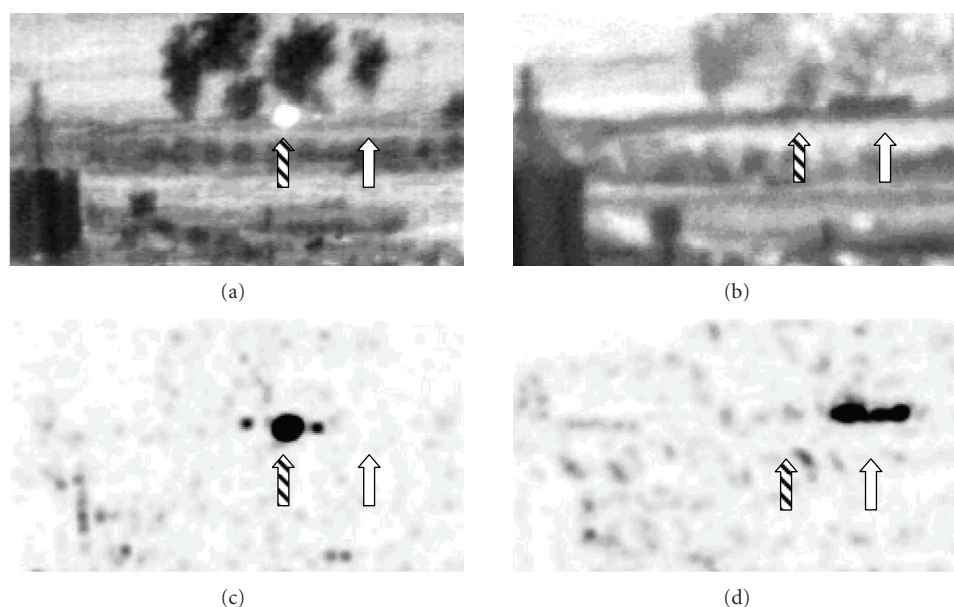


FIGURE 17: Motion weights extracted from real-life sequences. (a) is a sample frame from the thermal channel; (b) is the corresponding frame from the visual-range one. (c)-(d) are the matching motion-defined weights.

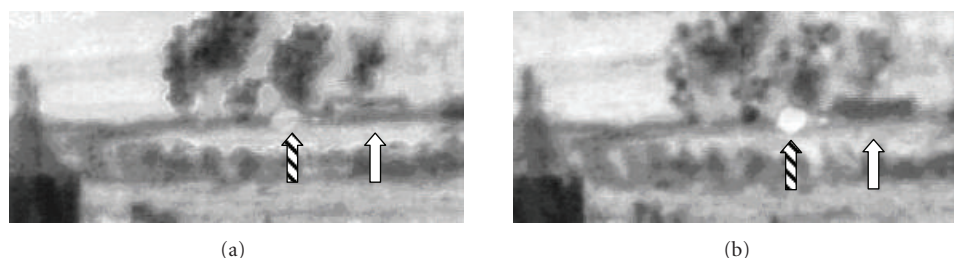


FIGURE 18: Corresponding frames fused using (a) only noise-defined and VI-weights with no motion and using (b) noise-defined and VI-weights along with motion-defined weights.

importance, local SNR level, and local motion activity. While each of the described methods can stand on its own and has shown good results, the full visual- and thermal-range image fusion system presented here makes use of them all simultaneously to yield a better system in terms of visual quality. Experiments with synthetic test sequences, as well as with real-life image sequences, have shown that the output of this system is a substantial improvement over the sensor inputs.

ACKNOWLEDGMENTS

The authors appreciate the contribution of Alex Shtainman and Shai Gepshtein, Faculty of Engineering, Tel-Aviv University (Tel-Aviv, Israel), to this research. They also thank Frederique Crete, Laboratoire des Images et des Signaux (Grenoble, France), for her useful suggestions regarding quantitative evaluation methods. Additionally, they would like to thank Haggai Kirshner, Faculty of Electrical Engineering, The Technion-Israeli Institute of Technology (Haifa, Israel), and Chad Goerzen, Faculty of Engineering, Tel-Aviv University, for their useful suggestions in the writing process. The

video database was acquired with the kind help of Elbit Systems Electro-Optics—ELOP Ltd., Israel. The research was partially funded by the Israeli Ministries of Transportation, Science Culture, and Sport.

REFERENCES

- [1] M. C. Roggermann and B. Welsh, *Imaging Through Turbulence*, CRC Press, Boca Raton, Fla, USA, 1996.
- [2] L. P. Yaroslavsky, B. Fishbain, G. Shabat, and I. Ideses, "Super-resolution in turbulent videos: making profit from damage," *Optics Letters*, vol. 32, no. 20, pp. 3038–3040, 2007.
- [3] C. J. Carrano, "Anisoplanatic performance of horizontal-path speckle imaging," in *Advanced Wavefront Control: Methods, Devices, and Applications*, vol. 5162 of *Proceedings of SPIE*, pp. 14–27, San Diego, Calif, USA, August 2003.
- [4] C. J. Carrano, "Speckle imaging over horizontal paths," in *High-Resolution Wavefront Control: Methods, Devices, and Applications IV*, vol. 4825 of *Proceedings of SPIE*, pp. 109–120, Seattle, Wash, USA, July 2002.
- [5] C. J. Carrano and J. M. Brase, "Adapting high-resolution speckle imaging to moving targets and platforms," in *Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems*

- and Applications, vol. 5409 of *Proceedings of SPIE*, pp. 96–105, Orlando, Fla, USA, April 2004.
- [6] C. J. Carrano and J. M. Brase, “Horizontal and slant path surveillance with speckle imaging,” in *Proceedings of the Technical Conference of Airforce Maui Optical Station (AMOS '02)*, p. 499, Maui, Hawaii, USA, September 2002.
 - [7] T. W. Lawrence, J. P. Fitch, D. M. Goodman, N. A. Massie, R. J. Sherwood, and E. M. Johansson, “Extended-image reconstruction through horizontal path turbulence using bispectral speckle interferometry,” *Optical Engineering*, vol. 31, no. 3, pp. 627–636, 1992.
 - [8] R. D. Hudson Jr., *Infrared System Engineering*, Wiley-Interscience, New York, NY, USA, 1969.
 - [9] Z. Zhang and R. S. Blum, “Image fusion for a digital camera application,” in *Proceedings of the 32nd Asilomar Conference on Signals, Systems & Computers*, vol. 1, pp. 603–607, Pacific Grove, Calif, USA, November 1998.
 - [10] A. M. Achim, C. N. Canagarajah, and D. R. Bull, “Complex wavelet domain image fusion based on fractional lower order moments,” in *Proceedings of the 8th International Conference on Information Fusion (FUSION '05)*, vol. 1, pp. 515–521, Philadelphia, Pa, USA, July 2005.
 - [11] H. Ghassemian, “Multi-sensor image fusion using multirate filter banks,” in *Proceedings of IEEE International Conference on Image Processing (ICIP '01)*, vol. 1, pp. 846–849, Thessaloniki, Greece, October 2001.
 - [12] V. Petrovich and C. Xydeas, “Computationally Efficient Pixel-level Image Fusion,” in *Proceedings of the International Conference on Data Fusion (EuroFusion '99)*, pp. 177–184, Stratford-upon-Avon, UK, October 1999.
 - [13] L. P. Yaroslavsky, B. Fishbain, A. Shteinman, and S. Gepshtein, “Processing and fusion of thermal and video sequences for terrestrial long range observation systems,” in *Proceedings of the 7th International Conference on Information Fusion (FUSION '04)*, vol. 2, pp. 848–855, Stockholm, Sweden, June–July 2004.
 - [14] Q.-S. Sun, S.-G. Zeng, Y. Liu, P.-A. Heng, and D.-S. Xia, “A new method of feature fusion and its application in image recognition,” *Pattern Recognition*, vol. 38, no. 12, pp. 2437–2448, 2005.
 - [15] K. Steinnocher, “Adaptive fusion of multisource raster data applying filter techniques,” in *Proceedings of the International Archives of Photogrammetry and Remote Sensing*, vol. 32, part 7-4-3W6, pp. 108–115, Valladolid, Spain, June 1999.
 - [16] N. Nandhakumar and J. K. Aggarwal, “Integrated analysis of thermal and visual images for scene interpretation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 469–481, 1988.
 - [17] D. L. Hall, M. D. McNeese, E. Rothhoff, and T. Shaw, “Improving the fusion process using semantic level whole-brain analysis,” in *Proceedings of the MSS National Symposium on Sensors and Data Fusion*, Monterey, Calif, USA, May 2005.
 - [18] J. J. Lewis, R. J. O’Callaghan, S. G. Nikolov, D. R. Bull, and C. N. Canagarajah, “Region-based image fusion using complex wavelets,” in *Proceedings of the 7th International Conference on Information Fusion (FUSION '04)*, vol. 1, pp. 555–562, Stockholm, Sweden, June–July 2004.
 - [19] T. A. Wilson, S. K. Rogers, and M. Kabrisky, “Perceptual-based image fusion for hyperspectral data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 1007–1017, 1997.
 - [20] P. J. Burt and R. J. Kolczynski, “Enhanced image capture through fusion,” in *Proceedings of the 4th IEEE International Conference on Computer Vision*, pp. 173–182, Berlin, Germany, May 1993.
 - [21] S. G. Nikolov, D. R. Bull, C. N. Canagarajah, M. Halliwell, and P. N. T. Wells, “Image fusion using a 3-D wavelet transform,” in *Proceedings of the 7th International Conference on Image Processing and Its Applications*, vol. 1, pp. 235–239, Manchester, UK, July 1999.
 - [22] P. Hill, N. Canagarajah, and D. Bull, “Image fusion using complex wavelets,” in *Proceedings of the 13th British Machine Vision Conference (BMVC '02)*, pp. 487–496, Cardiff, UK, 2002.
 - [23] I. Koren, A. Laine, and F. Taylor, “Image fusion using steerable dyadic wavelet transform,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP '95)*, vol. 3, pp. 232–235, Washington, DC, USA, October 1995.
 - [24] H. Li, B. S. Manjunath, and S. K. Mitra, “Multisensor image fusion using the wavelet transform,” *Graphical Models & Image Processing*, vol. 57, no. 3, pp. 235–245, 1995.
 - [25] E. Lallier and M. Farooq, “A real-time pixel-level based image fusion via adaptive weight averaging,” in *Proceedings of the 3rd International Conference on Information Fusion (Fusion '00)*, vol. 2, pp. WeC3-3–WeC3-10, Paris, France, July 2000.
 - [26] <http://www.eng.tau.ac.il/~barak/RealTimeTurbulenceCompensation>.
 - [27] <http://www.eng.tau.ac.il/~yaro/Shtainman/shtainman.htm>.
 - [28] B. M. Welsh and C. S. Gardner, “Performance analysis of adaptive optics systems using slope sensors,” *Journal of the Optical Society of America A*, vol. 6, no. 12, pp. 1913–1923, 1989.
 - [29] B. L. Ellerbroek, “First-order performance evaluation of adaptive-optics systems for atmospheric-turbulence compensation in extended-field-of-view astronomical telescopes,” *Journal of the Optical Society of America A*, vol. 11, no. 2, pp. 783–805, 1994.
 - [30] A. Tokovinin, M. Le Louarn, and M. Sarazin, “Isoplanatism in a multiconjugate adaptive optics system,” *Journal of the Optical Society of America A*, vol. 17, no. 10, pp. 1819–1827, 2000.
 - [31] M. Lloyd-Hart and N. M. Milton, “Fundamental limits on isoplanatic correction with multiconjugate adaptive optics,” *Journal of the Optical Society of America A*, vol. 20, no. 10, pp. 1949–1957, 2003.
 - [32] B. Le Roux, J.-M. Conan, C. Kulcsár, H.-F. Raynaud, L. M. Mugnier, and T. Fusco, “Optimal control law for classical and multiconjugate adaptive optics,” *Journal of the Optical Society of America A*, vol. 21, no. 7, pp. 1261–1276, 2004.
 - [33] T. Fusco, J.-M. Conan, L. M. Mugnier, V. Michau, and G. Rousset, “Characterization of adaptive optics point spread function for anisoplanatic imaging. Application to stellar field deconvolution,” *Astronomy and Astrophysics Supplement Series*, vol. 142, no. 1, pp. 149–156, 2000.
 - [34] W. M. Farmer, *The Atmospheric Filter*, vol. 2, JCD Publishing, Bellingham, Wash, USA, 2001.
 - [35] D. Sadot and N. S. Kopeika, “Imaging through the atmosphere: practical instrumentation-based theory and verification of aerosol modulation transfer function,” *Journal of the Optical Society of America A*, vol. 10, no. 1, pp. 172–179, 1993.
 - [36] B. Cohen, V. Avrin, M. Belitsky, and I. Dinstein, “Generation of a restored image from a video sequence recorded under turbulence effects,” *Optical Engineering*, vol. 36, no. 12, pp. 3312–3317, 1997.
 - [37] B. Ro. Frieden, “Turbulent image reconstruction using object power spectrum information,” *Optics Communications*, vol. 109, no. 3-4, pp. 227–230, 1994.
 - [38] H. van-der-Elst and J. J. D. van-Schalkwyk, “Modelling and restoring images distorted by atmospheric turbulence,” in *Proceedings of the IEEE South African Symposium on Communications and Signal Processing (COMSIG '94)*, pp. 162–167, Stellenbosch, South Africa, October 1994.

- [39] C. J. Carrano, "Progress in horizontal and slant-path imaging using speckle imaging," in *Optical Engineering at the Lawrence Livermore National Laboratory*, vol. 5001 of *Proceedings of SPIE*, pp. 56–64, San Jose, Calif, USA, January 2003.
- [40] D. Fraser, G. Thorpe, and A. Lambert, "Atmospheric turbulence visualization with wide-area motion-blur restoration," *Journal of the Optical Society of America A*, vol. 16, no. 7, pp. 1751–1758, 1999.
- [41] D. Clyde, I. Scott-Fleming, D. Fraser, and A. Lambert, "Application of optical flow techniques in the restoration of non-uniformly warped images," in *Proceedings of the Digital Image Computing: Techniques and Applications (DICTA '02)*, pp. 195–200, Melbourne, Australia, January 2002.
- [42] I. Scott-Fleming, K. Hege, D. Clyde, D. Fraser, and A. Lambert, "Gradient based optical flow techniques for tracking image motion due to atmospheric turbulence," in *Proceedings of the Signal Recovery and Synthesis Symposium, Optical Society of America*, pp. 68–70, Albuquerque, NM, USA, November 2001.
- [43] D. H. Frakes, J. W. Monaco, and M. J. T. Smith, "Suppression of atmospheric turbulence in video using an adaptive control grid interpolation approach," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '01)*, vol. 3, pp. 1881–1884, Salt Lake City, Utah, USA, May 2001.
- [44] B. Fishbain, L. P. Yaroslavsky, I. A. Ideses, O. Ben-Zvi, and A. Shtern, "Real time stabilization of long range observation system turbulent video," in *Proceedings of the International Society for Optical Engineering*, vol. 6496 of *Proceedings of SPIE*, p. 11 pages, San Jose, Calif, USA, January 2007.
- [45] B. Fishbain, I. A. Idess, Sh. Gepstein, and L. P. Yaroslavsky, "Turbulent video enhancement: Image stabilization and super-resolution," in *Proceedings of the 11th Meeting on Optical Engineering and Science*, Tel Aviv, Israel, March 2007.
- [46] L. P. Yaroslavsky, B. Fishbain, I. A. Ideses, D. Slasky, and Z. Hadas, "Simple methods for real-time stabilization of turbulent video," in *Proceedings of the ICO Topical Meeting on Optoinformatics/Information Photonics (ITMO '06)*, pp. 138–140, St Petersburg, Russia, October 2006.
- [47] S. Gepshtein, A. Shtainman, B. Fishbain, and L. P. Yaroslavsky, "Restoration of atmospheric turbulent video containing real motion using elastic image registration," in *Proceedings of the European Signal Processing Conference (EUSIPCO '04)*, pp. 477–480, John Platt, Vienna, Austria, September 2004.
- [48] L. P. Yaroslavsky, *Digital Holography and Digital Image Processing*, Kluwer Academic, Boston, Mass, USA, 2003.
- [49] B. Fishbain, L. P. Yaroslavsky, and I. A. Ideses, "Real-time stabilization of long range observation system turbulent video," *Journal of Real-Time Image Processing*, vol. 2, no. 1, pp. 11–22, 2007.
- [50] S.-C. S. Cheung and C. Kamath, "Robust techniques for background subtraction in urban traffic video," in *Visual Communications and Image Processing 2004*, vol. 5308 of *Proceedings of SPIE*, pp. 881–892, San Jose, Calif, USA, January 2004.
- [51] T. Amiaz and N. Kiryati, "Piecewise-smooth dense optical flow via level sets," *International Journal of Computer Vision*, vol. 68, no. 2, pp. 111–124, 2006.
- [52] R. Ben-Ari and N. Sochen, "A general framework for regularization in PDE based computation of optical flow via embedded maps and minimal surfaces," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, New-York, NY, USA, June 2006.
- [53] A. Bruhn, J. Weickert, T. Kohlberger, and C. Schnörr, *Scale Space and PDE Methods in Computer Vision*, Lecture Notes in Computer Science, Springer, Berlin, Germany, 2005.
- [54] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on theory for wrapping," in *Proceedings of the 8th European Conference on Computer Vision (ECCV '04)*, vol. 4, pp. 25–36, Prague, Czech Republic, May 2004.
- [55] L. J. Barron, D. J. Fleet, and S. S. Beachemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, 1994.
- [56] B. K. P. Horn and B. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1–3, pp. 185–203, 1981.
- [57] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81)*, pp. 674–679, Vancouver, BC, Canada, August 1981.
- [58] S. Periaswamy and H. Farid, "Elastic registration in the presence of intensity variations," *IEEE Transactions on Medical Imaging*, vol. 22, no. 7, pp. 865–874, 2003.
- [59] S. Periaswamy and H. Farid, "Differential elastic image registration," Tech. Rep. TR2001-413, Dartmouth College, Computer Science, Hanover, NH, USA, 2001.
- [60] M. I. Charnotskii, "Imaging in turbulence beyond diffraction limits," in *Adaptive Optical Systems and Applications*, vol. 2534 of *Proceedings of SPIE*, pp. 289–297, San Diego, Calif, USA, July 1995.
- [61] A. Lambert, D. Fraser, M. R. S. Jahromi, and B. R. Hunt, "Super-resolution in image restoration of wide area images viewed through atmospheric turbulence," in *Image Reconstruction from Incomplete Data II*, vol. 4792 of *Proceedings of SPIE*, pp. 35–43, Seattle, Wash, USA, July 2002.
- [62] M. I. Charnotskii, V. A. Myakinin, and V. U. Zavorotnyy, "Observation of superresolution in nonisoplanatic imaging through turbulence," *Journal of the Optical Society of America A*, vol. 7, no. 8, pp. 1345–1350, 1990.
- [63] A. Lambert and D. Fraser, "Superresolution in imagery arising from observation through anisoplanatic distortion," in *Image Reconstruction from Incomplete Data III*, vol. 5562 of *Proceedings of SPIE*, pp. 65–75, Denver, Colo, USA, August 2004.
- [64] S. Srinivasan and R. Chappella, "Image sequence stabilization, mosaicking and superresolution," in *Video and Image Processing Handbook*, A. C. Bovik, Ed., chapter 3.13, pp. 259–268, Academic Press, Dallas, Tex, USA, 2000.
- [65] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models & Image Processing*, vol. 53, no. 3, pp. 231–239, 1991.
- [66] N. Goldberg, A. Feuer, and G. C. Goodwin, "Super-resolution reconstruction using spatio-temporal filtering," *Journal of Visual Communication and Image Representation*, vol. 14, no. 4, pp. 508–525, 2003.
- [67] S. Farsiu, M. Elad, and P. Milanfar, "Multiframe demosaicing and super-resolution of color images," *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 141–159, 2006.
- [68] B. Fishbain, L. P. Yaroslavsky, and I. A. Ideses, "Real-time turbulent video perfecting by image stabilization and super-resolution," in *Proceedings of the 7th IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP '07)*, Palma de Mallorca, Spain, August 2007.
- [69] L. P. Yaroslavsky, "Fast discrete sinc-interpolation: a gold standard for image resampling," in *Advances in Signal Transforms: Theory and Applications*, J. Astola and L. P. Yaroslavsky, Eds., EURASIP Book Series on Signal Processing and Communications, Hindawi, New York, NY, USA, 2007.

- [70] F. Crete, T. Dolmiere, P. Ladref, and M. Nicolas, "The blur effect: perception and estimation with a new no-reference perceptual blur metric," in *Human Vision and Electronic Imaging XII*, vol. 6492 of *Proceedings of SPIE*, pp. 11 pages, San Jose, Calif, USA, January 2007.
- [71] G. C. Holst, *Testing and Evaluation of Infrared Imaging Systems*, SPIE Press, Bellingham, Wash, USA, 2nd edition, 1998.
- [72] G. C. Holst, *CCD Arrays, Cameras and Displays*, SPIE Press, Bellingham, Wash, USA, 2nd edition, 2001.
- [73] L. P. Yaroslavsky, "Local adaptive filtering in transform domain for image restoration, enhancement and target location," in *Proceedings of the 6th International Workshop on Digital Image Processing and Computer Graphics: Applications in Humanities and Natural Sciences*, vol. 3346 of *Proceedings of SPIE*, pp. 2–17, Vienna, Austria, October 1997.
- [74] L. P. Yaroslavsky, "Space variant and adaptive transform domain image and video restoration methods," in *Advances in Signal Transforms: Theory and Applications*, J. Astola and L. P. Yaroslavsky, Eds., Hindawi, New York, NY, USA, 2007.
- [75] L. P. Yaroslavsky and H. J. Caulfield, "Deconvolution of multiple images of the same object," *Applied Optics*, vol. 33, no. 11, pp. 2157–2162, 1994.
- [76] L. P. Yaroslavsky, "Advanced image processing Lab—a tutorial," in *Proceedings of the European Signal Processing Conference (EUSIPCO '00)*, Tampere, Finland, September 2000.
- [77] M. C. Dudzik, Ed., *The Infrared & EO Systems Handbook, Electro-Optical Systems Design, Analysis, and Testing*, vol 4, vol. 2 of *Atmospheric Propagation of Radiation*, SPIE Optical Engineering Press, Bellingham, Wash, USA, 1993.
- [78] R. L. Lagendijk, P. M. B. van-Roosmalen, and J. Biemond, "Video enhancement and restoration," in *Handbook of Image and Video Processing*, Al. Bovik, Ed., pp. 227–241, Academic Press, Canada, 2000.
- [79] L. P. Yaroslavsky and M. Eden, *Fundamentals of Digital Optics*, Birkhauser, Boston, Mass, USA, 1996.