Superresolution in turbulent videos: making profit from damage

L. Yaroslavsky,* B. Fishbain, G. Shabat, and I. Ideses

Department of Interdisciplinary Studies, Faculty of Engineering, Tel Aviv University, Tel Aviv 69978, Israel
**Corresponding author: yaro@eng.tau.ac.il*

Received July 10, 2007; revised September 11, 2007; accepted September 12, 2007;
posted September 17, 2007 (Doc. ID 85126); published October 12, 2007

It is shown that one can make use of local instabilities in turbulent video frames to enhance image resolution beyond the limit defined by the image sampling rate. We outline the processing algorithm, present its experimental verification on simulated and real-life videos, and discuss its potentials and limitations.

OCIS codes: 100.2000, 100.3020, 100.6640, 110.7050.

In long distance observation systems, images and video are frequently damaged by atmospheric turbulence, which causes spatially and temporally chaotic fluctuations in the index of refraction of the atmosphere [1] and results in chaotic spatial, and temporal geometrical distortions in the neighborhood of all pixels. This geometrical instability of image frames heavily worsens the quality of videos and hampers their visual analysis. To make visual analysis possible, first of all it is required to stabilize images of stable scenes while preserving the real motion of moving objects that might be present in the scene. Methods of generating stabilized videos from turbulent ones were reported in [2–5] with emphasis on the real-time implementation and reliable discrimination of atmospheric turbulence-induced pixel displacements from those caused by moving objects present in the scene. Along with this, in [5,6] the idea of making a profit from atmospheric turbulence-induced image geometrical spatial–temporal degradations to compensate for image sampling artifacts and generate stabilized images of the stable scene with higher resolution than that defined by the camera sampling grid was advanced. The idea was supported by the demonstration of proof-of-concept results of perfecting real-life videos. A similar idea was also advanced in [7].

In this Letter, we elaborate on this idea, describe a practical method for producing good quality higher-resolution videos from low-resolution turbulent video streams that implement this idea, and illustrate its performance experimentally using simulated and real-time atmosphere distorted videos. The core of the method of image stabilization with superresolution is elastic registration, with subpixel accuracy, of available video frames of stable scenes followed by resampling of the frames according to the registration results.

The first step in the processing of each current frame is updating a reference frame used for elastic registration. For generating and updating the reference frame, a temporal pixelwise median over a block of frames preceding and following the current frame is used as described in [2–6].

The next step is computing, with subpixel accuracy, a map of pixel displacements in the current frame with respect to the reference frame. It is carried out by means of elastic registration of the current frame and the reference frame. For elastic image registration, several options exist: (i) optical flow methods [8,9]; (ii) correlation methods [10], and (iii) using motion vector data available in MPEG encoded videos [11]. In the reported experiments, optical flow methods were used.

The pixel displacement map is then analyzed and segmented to separate pixels of the real moving objects from those that belong to the stable scene and are displaced solely due to the atmospheric turbulence. Toward this end, the absolute difference from the reference as well as statistical analysis of displacement magnitudes and angles is used [3–6].

The displacement map found for pixels of the stable scene serves, at the next processing step, to place pixels for each 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 [10]. 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 position in the output enhanced frame. To make best use of all of them, these pixels must be averaged. For this averaging, we suggest computing the median of those pixels to avoid the influence of outliers that may appear due to possible anomalous errors in the displacement map.

After all the available input frames of the stable scene are used in this way, the enhanced and upsampled output frame contains accumulated pixels of input frames in positions where there were substitutions from input frames and interpolated pixels of the reference frame in positions where there were no substitutions. Substituted pixels introduce to the out-

0146-9592/07/203038-3/$15.00

put 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 replaced do not contain frequencies outside the baseband. To finalize the processing and take full advantage of the superresolution provided by the replaced pixels, we suggest the following iterative reinterpolation algorithm that implements discrete sinc interpolation of nonuniformly sampled data. The algorithm assumes that all accumulated replaced pixels are stored in an auxiliary replacement map that contains pixels values and coordinates. At each iteration of the algorithm, discrete Fourier transform (DFT) [or discrete cosine transform (DCT), which is less vulnerable to boundary effects [10]] a spectrum of the image obtained at the previous iteration is computed and then zeroed in all its components outside the selected enhanced bandwidth, say, double the original baseband one. After this, inverse DFT (DCT) is performed on such a modified spectrum, and the corresponding pixels in the resulting image are replaced by 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 as to when the iterations can be stopped.

Stabilized and resolution-enhanced images, obtained in this way can be finally downsampled to the sampling rate determined by the desired enhanced bandwidth and then subjected to additional processing aimed at camera aperture correction and, when necessary, denoising.

The potential efficiency of the entire processing is illustrated in Fig. 1 with the results of computer simulations. As one can see from Fig. 1, even from as small a number of frames as 15, a substantial resolution enhancement is possible especially when a sufficiently large number of reinterpolation iterations are carried out.

The method was also tested on real-life turbulent degraded visual-light range and thermal videos. Visual-light range videos (wavelengths of $0.3-1\ \mu$m) were acquired with 4CIF (Common Intermediate Format) 704×576 pixels cameras. Thermal videos were acquired with QCIF near infrared ($2-5\ \mu$m) cameras. The typical distance from the camera to the observed scene ranged from 5 to 15 km. The typical

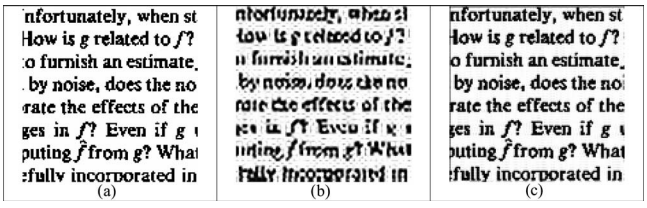

Fig. 1. Illustrative simulation results of the resolution enhancement of turbulent video frames: (a) initial high resolution frame, (b) example of a low-resolution frame obtained by a camera with a fill factor 0.05, (c) resolution-enhanced frame obtained by the described fusion process with 50 iterations of the reinterpolation from 15 low resolution frames distorted by simulated random local displacements with a standard deviation of 0.5 interpixel distance.

environmental temperature, in which the sequences were acquired was above 30°C. In our data, standard deviation of the motion field magnitude, in interpixel distance, was 0.3–0.4 for the thermal range and 0.7–1.2 for the visual range.

Figure 2 illustrates the efficiency of the method with real-life turbulence degraded video. Figure 2(a) and 2(b) shows a 256×256 pixel fragment of a stabilized real-life frame obtained as a temporal median over 117 frames, and the same fragment obtained after replacing its pixels, as it is described here, by pixels of those 117 frames according to their displacement maps. As a visually improved resolution can be appreciated only on a high quality display, image (c) presents the difference between these two images that clearly shows edges enhanced in image (b) compared with image (a). Figure 2(d) presents the final result of the processing after reinterpolation and aperture correction is implemented with the assumption that the camera sensor has a fill factor close to 1.

For the quantitative evaluation of the resolution improvement we used a method for the numerical evaluation of image visual quality degradation due to blur suggested in [12]. The method is based on the discrimination between different levels of blur perceptible on the same picture. The image degradation measure ranges from 0 to 1, which are, respectively, the lowest and the highest degradations. Degradation factors for images shown in Fig. 2 are 0.43 for the stabilized image, before resolution enhancement, 0.4 for the resolution-enhanced image, and 0.28 for the final reinterpolated and aperture corrected image.

In the evaluation of the results obtained for real-life video, one should take into account that substantial resolution enhancement in the described superresolution fusion process can be expected only if the camera fill factor is small enough. The camera fill factor determines the degree of low-pass filtering introduced by the camera. Due to this low-pass filtering, image high frequencies in the base band and aliasing high frequency components that come into the base band due to image sampling are suppressed. Those aliasing components can be recovered and returned back to their true frequencies outside the base band in the described superresolution process, but only if they have not been lost due to the camera low-pass filtering. The larger the fill factor, the heavier the unrecoverable resolution losses. In the described simulation experiment, the camera fill factor is 0.05, whereas in reality fill factors of monochrome cameras are usually close to 1.

In conclusion, one can state that the presented results show that the described technique allows, in addition to compensating for the atmospheric turbulence in video sequences, to improve image resolution thanks to proper elastic local registration and resampling of degraded video frames. The degree of the achievable resolution improvement depends on the number of frames that contain a stable scene, the spread of turbulence-induced local image displacements, and on the camera fill factor. Certain resolution enhancement can be achieved even for conven-

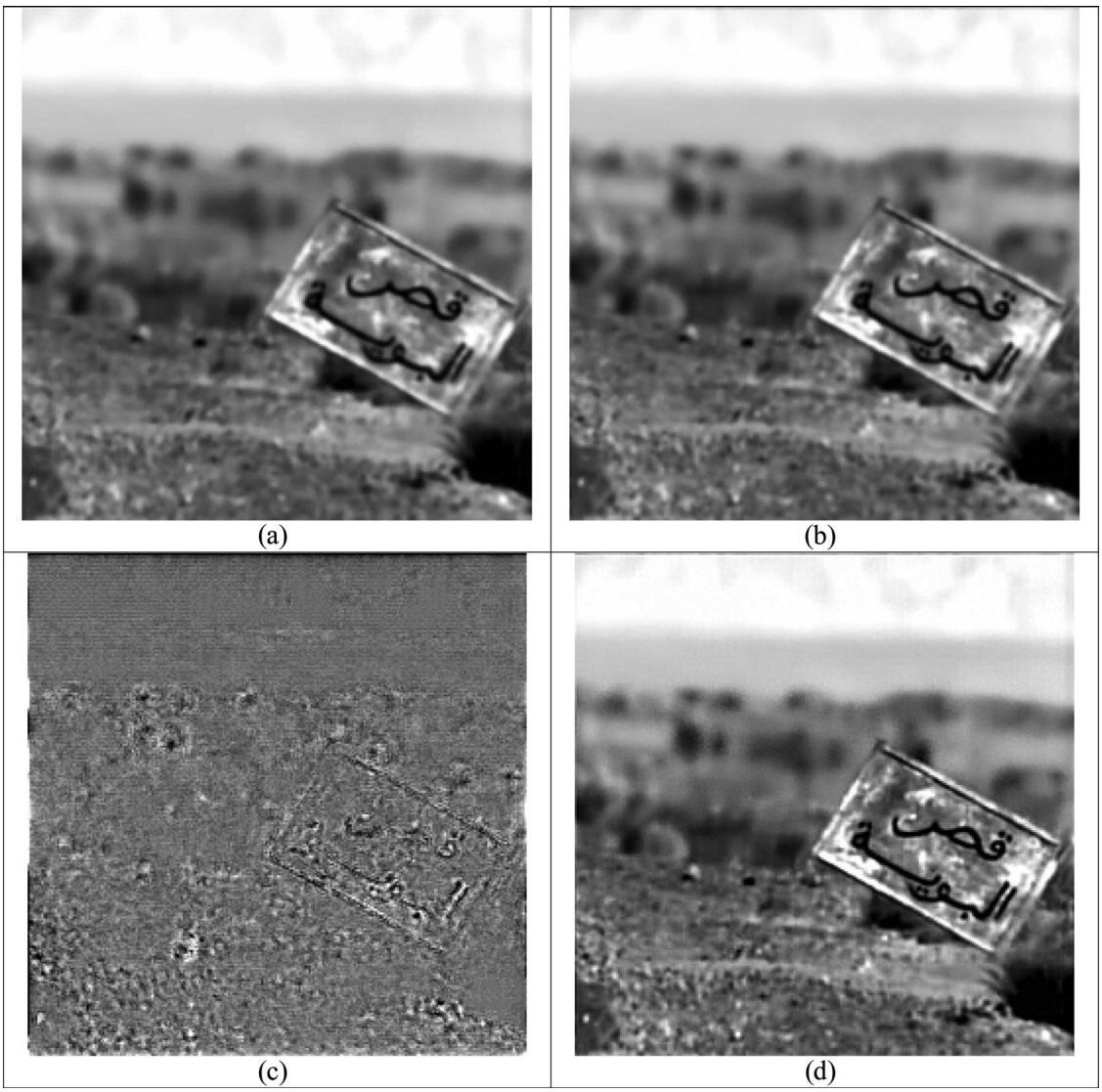

Fig. 2. Illustration of resolution enhancement of real-life video frames (a) a 256×256 pixel fragment of the stabilized frame; (b) same fragment after resolution enhancement; (c) difference between images (a) and (b) that show edges enhanced in image (b); (d) image (b) after reinterpolation and aperture correction.

tional cameras with a large fill factor. For cameras with a small fill factor, such as, for instance, color cameras, the potential resolution enhancement might be, as simulation experiments show, very substantial.

References

1. M. C. Roggermann and B. Welsh, *Imaging Through Turbulence* (CRC Press, 1996).
2. S. Gepshtein, A. Shtainman, B. Fishbain, and L. Yaroslavsky, in *The 2004 European Signal Processing Conference (EUSIPCO-2004)*, J. Platt, ed. (2004), pp. 477–480.
3. L. Yaroslavsky, B. Fishbain, I. Ideses, D. Slasky, and Z. Hadas, in *Proceedings of ICO Topical Meeting on Optoinformatics/Information Photonics* (ITMO, 2006), pp. 138–140.
4. B. Fishbain, L. P. Yaroslavsky, I. A. Ideses, A. Shtern, and O. Ben-Zvi, Proc. SPIE **6496**, 64960C (2007).
5. B. Fishbain, I. A. Ideses, Sh. Gepstein, and L. P. Yaroslavsky, in Proceedings of The 11th Meeting on Optical Engineering and Science in Israel, Tel-Aviv, March 26–27, 2007.
6. B. Fishbain, L. P. Yaroslavsky, and I. A. Ideses, in *Proceedings of the 7th IASTED International Conference on Visualization, Imaging, and Image Processing* (VIIP, 2007), pp. 213–218.
7. Z. Zalevsky, Sh. Rozental, and M. Meller, Opt. Lett. **32**, 1072 (2007).
8. T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, in *Proceedings of the 8th European Conference on Computer Vision*, T. Pajdla and J. Matas, eds., Vol. 4 of Springer Lecture Notes in Computer Science (Springer-Verlag, 2004), pp. 25–36.
9. L. J. Barron, D. J. Fleet, and S. S. Beachemin, Inf. Sci. (N.Y.) **12**, 43 (1994).
10. L. Yaroslavsky, *Digital Holography and Digital Image Processing* (Kluwer Scientific, 2004).
11. I. Ideses, L. Yaroslavsky, B. Fishbain, and R. Vistuch, Proc. SPIE **6490**, 64901C (2007).
12. F. Crete, T. Dolmiere, P. Ladret, and M. Nicolas, Proc. SPIE **6492**, 64920I (2007).