A Method for Removal of Turbulence Disturbance from Video, Enabling Higher Level Applications

Tomer Avidor and Moty Golan

RAFAEL

Abstract. A common method for reconstructing a turbulence scene is through the creation of an artificial reference image. The reference image is usually obtained by averaging video through time. Using optical flow from that reference image to input images would give rise to such applications as: super-resolution, tracking, and so forth. This technique, however, suffers from several drawbacks: the resulting artificial reference image is blurred, so the calculated optical flow is not precise and limits the ability of higher level applications (such as super-resolution, tracking, mosaics). We present a mathematical framework for reconstructing a video scene as would have been seen without turbulence interference, yielding an observable live video output. We then use both frames and optical flow to get the aforementioned applications while dealing with camera motion, and draw guidelines to deal with in-scene motion inherently.

Keywords: turbulence reconstruction, optical flow, super-resolution, surveillance.

1 Introduction

Long range observation systems are greatly affected by the existence of atmospheric turbulence disturbance, rendering them at times unsuitable for vision applications such as surveillance, scene inspection, and so forth. The disturbance is characterized by two separate phenomena: non-uniform geometrical deformation and non-uniform image quality degradation (blur).

A common way of dealing with the above disturbances involves the creation of a reference image which is free of the first phenomena. The reference image is usually estimated by some form of pixel wise temporal filtering such as averaging, as is done in [1] and [2]. The resulting image is equivalent to a Gaussian filtering of the true world image. After the reference image estimation, the turbulence distortion may by estimated using an optical flow from each image to it. In this paper we present a mathematical scheme to estimate the turbulence distortion without the formation of a reference image but rather from the distorted images themselves. We do this by calculating the optical flow between the turbulence distorted images. Using our direct approach results in a sharper

M. Kamel and A. Campilho (Eds.): ICIAR 2009, LNCS 5627, pp. 647–656, 2009.

(a) (b) (c)

Fig. 1. (a)Original distorted image (b)Estimated by averaging images (c)Reconstructed by proposed method

(estimated) turbulence free image and opens the way of dealing with camera and in scene movement in a natural way. A comparison between an original image[1] (of a turbulent scene) and a reconstruction using image averaging and reconstruction is shown in Fig. 1. We further show how to use our scheme to get rid of the second phenomena by using super-resolution on the degraded images. We also show applications on scanning (pan) cameras, including formation of high-resolution panorama, tracking and scene reconstruction involving in-scene movement. This paper is organized as follows: Section 2 describes the mathematical scheme estimating the turbulence optical flow. We further develop the scheme to handle simple camera movement and in scene movement. Section 3 describes further application using the turbulence optical flow field. We describe estimation of a turbulence geometric disturbance free video, for scene inspection, tracking, super-resolution and panorama generation. Section 4 shows results of the afore mentioned applications on real data. We conclude with ideas for further researches.

2 Single Image Reconstruction

2.1 Reconstruction of a Static Scene from a Static Camera

Let I_k, $k \in \{-L..M\}$, be M+L+1 sequential images from video of a turbulent scene, and let J_k, $k \in \{-L..M\}$, be image I_k without the turbulence. Let I_0 be some image from the video. Our wish is to estimate the optical flow from J_0 to I_0. Let Δ_k, $k \in \{-L..M\}$, be the optical flow from I_0 to I_k, and let T_k be the optical flow from J_0 to I_k as described in Fig. 2.

According to the physical model, each pixel vibrates about its true location; by the law of large numbers, averaging the pixel location over time gives us the pixels true location. Formalizing the last statement we get:

$$\sum_{k=-N}^{N} 1/(2N+1) \cdot T_k \xrightarrow[N\to\infty]{} 0. \tag{1}$$

[1] Generously granted by Prof. Leonid Yaroslavsky, TA University, IL.

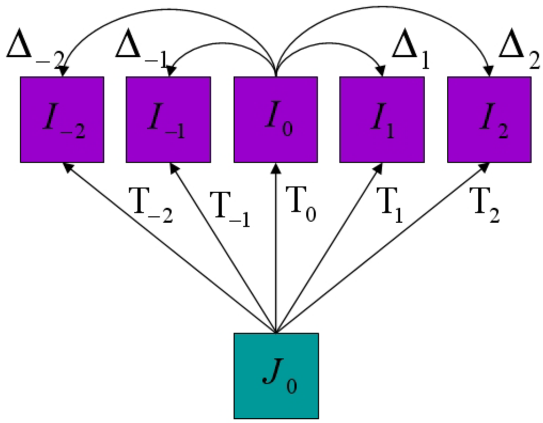


Fig. 2. Proposed reconstruction scheme

We use the approximation

$$1/\left(M+L+1\right)\cdot\sum_{k=-L}^{M}T_k=0, \tag{2}$$

and a simple equality

$$T_0+\Delta_k=T_k. \tag{3}$$

Dividing by M+L+1, summing over all images and using (1), we get

$$1/\left(M+L+1\right)\cdot\sum_{k=-L}^{M}\left(T_0+\Delta_k\right)=0, \tag{4}$$

or

$$T_0=-\sum_{k=-L}^{M}\Delta_k/\left(M+L+1\right). \tag{5}$$

Warping the image I_0 using T_0, we get the reconstructed image J_0, i.e. image I_0 without the turbulence spatial disturbance.

2.2 Reconstruction of a Static Scene from a Moving Camera

The above formulation may be extended to the case of a moving camera. We distinguish between two common movements

1. Camera panning, as in scanning mode.
2. Camera translation.

Camera panning manifests itself as translation on the image plane. However, camera translation doesn't manifest itself as a translation, for the magnitude of a translation is proportional to the inverse of the range, and hence is not global.

However, for long range observations which are most affected by turbulence when their FOV (Field Of View) is narrow, the difference between the object's ranges in the scene is negligible compared to the ranges themselves; so the image translation is approximately global. Furthermore, most outdoor videos contain some global translation between the images due to camera instability, hence could be modeled as a camera translation. Let U_k be the global translation from the reference image I_0 to image I_k. Then substituting T_k for $T_k + U_k$ in (3), we get

$$T_0 = -\sum_{k=-L}^{M} \left(\Delta_k - U_k\right) / \left(M + L + 1\right). \tag{6}$$

2.3 Reconstruction of a Dynamic Scene from a Static Camera

By a dynamic scene we mean a scene in which, aside from static objects, several objects move rigidly. Although the following formulation is general, we assume that optical flow for a moving object can be reliably estimated only if it moves according to some simple motion, such as constant velocity or constant acceleration. The following formulation is for constant velocity of a single moving object, though it may be generalized easily. For all Pixels that move only due to the turbulence, we use the formulation in 2.1, The following applies for the rest of the pixels. Let $\Omega_t \subset I_t$ be the moving object, and let $U\left(x, y, t\right)$ be the displacement of a pixel in image I_t at image coordinate (x, y). For the case of constant velocity, $U\left(x, y, t\right) = u$ for all $(x, y) \in \Omega_t$. In the following, we use the same notations as in Sect. 2.1. Let K_k be the optical flow from image J_0 to image I_k, and T_k be the optical flow attributed to the turbulence, i.e.

$$K_k = k \cdot u + T_k, \tag{7}$$

with

$$1/\left(M + L + 1\right) \cdot \sum_{k=-L}^{M} T_k = 0. \tag{8}$$

Then

$$1/\left(M + L + 1\right) \cdot \sum_{k=-L}^{M} K_k = 1/\left(M + L + 1\right) \cdot \sum_{k=-L}^{M} \left(k \cdot u + T_k\right) \tag{9}$$

$$= u/\left(M + L + 1\right) \cdot \sum_{k=-L}^{M} k + 0. \tag{10}$$

On the other hand

$$1/\left(M + L + 1\right) \cdot \sum_{k=-L}^{M} \left(K_0 + \Delta_k\right) = K_0 + 1/\left(M + L + 1\right) \cdot \sum_{k=-L}^{M} \Delta_k, \tag{11}$$

hence

$$u/(M+L+1) \cdot \sum_{k=-L}^{M} k = K_0 + 1/(M+L+1) \cdot \sum_{k=-L}^{M} \Delta_k. \tag{12}$$

So, the Reconstruction Equation is

$$K_0 = \frac{u/2\,[M \cdot (M+1) - L \cdot (L+1)] - \sum_{k=-L}^{M} \Delta_k}{M+L+1}. \tag{13}$$

3 Applications

Having the ability to reconstruct a turbulent frame, we are now capable of developing a whole set of applications. Here we present some of those.

3.1 Turbulent Video In - Turbulent Free Video Out

Turbulent video might become unobservable to the human eye. Removing the disturbance, even without any in-scene motion analysis, can assist the video inspector. Equation (6) shows the way to remove turbulence disturbance from video, keeping camera movement. To produce such a video we can divide the input video into separate slices, each containing enough images (in our case, 25 were enough) to satisfy the approximation $\sum_k \Delta_k = 0$. This forces us to deal with a real-time issue: To compute the motion fields between any pair of images within a collection of N images. That means N(N-1) times running the optical flow solver, which is typically relatively expensive. Transformation concatenation for non-rigid Motion fields requires interpolation, since between three censequtive images, e.g. Im1, Im2, and Im3, the endpoints of the motion field from Im1 to Im2 ($\Delta_{1,2}$) don't meet the origins of the motion field from Im2 to Im3 ($\Delta_{2,3}$). Notice that we know $\Delta_{2,3}$ on Im2's ordered grid and want to predict its values on $\Delta_{1,2}$ cloud of endpoints, and the transformation between these sets is $\Delta_{1,2}$ itself. So the missing motion field, from $\Delta_{1,2}$ cloud of endpoints to $\Delta_{2,3}$ cloud of endpoints, is $\Delta_{1,2}(\Delta_{2,3})$. Namely,

$$\Delta_{1,3} = \Delta_{1,2} + \Delta_{1,2}(\Delta_{2,3}) \tag{14}$$

In that way, we lower computational cost from N(N-1) to 2(N-1) optical flow solver running time, plus (N-2)(N-1) times of applying motion field, which is significantly faster.

3.2 Panorama and Super-Resolution

Reconstructing images true appearance while maintaining the camera movement opens the way for higher-level algorithms such as panorama and super Resolution. Panorama is a useful tool for viewing and monitoring an area which is several times larger than the camera's Field Of View. Scene understanding grows and relationships between objects become clearer.

Super resolution is another helpful tool. A basic way of using it takes a collection of images' which contain a shift in between, and build a single output image as would have been seen using a higher definition camera with a narrow Point Spread Function, and adding less noise. Classic algorithm's input contains the images and the transformation between all of them and a common coordinate system. We refer to the Iterative Back Projection (IBP) scheme ([3] [4]), whose basic idea is as follows:

On each iteration, let us consider the best approximation to the SR image as if it would be the actual view, and computationally produce the set of images that would have been acquired from it using a camera (that is, for each image follow its transformation from the high-resolution common coordinates, optical lens' Point-Spread-Function and down-sampling). Then subtract this set from the set of input images to get the "error" hidden in the SR approximation, and back-transform this error to the common high resolution coordinates, to fix the SR approximation. In this way, on every iteration the information hidden in the input sequence is used to improve the result image till convergence.

After estimating the turbulence flow field with sub-pixel accuracy from every image to the high resolution image, we need also to get the backward flow fields as well. This is done the same way by simply reversing the arrow direction in Fig. 2, resulting in similar reconstruction equation.

Though some previous works suggested performing Super-Resolution on turbulent images, they are based mainly on creating an artificial blurred reference image, and calculating motion field between it and any input image. In this way, after averaging the information from input images to a common blurred grid, all what remains is to sharpen this image. Our method offers a direct solution, not using a virtual image, but resolving the exact transformations from input images to a non-deformed grid, and then resizing it to become directly from the common higher-resolution coordinates to each frame and vise versa. This set of transformations is what the IBP algorithm requires.

3.3 In-Scene Movement Detector and Tracker

After removing camera movement, each pixel's movement statistic is supposed to be unbiased. Pixels on moving objects would behave differently. First, the mean of their movement will be significantly larger than "regular" pixels, and will show the approximated direction. Second, if for short-time video snip we can assume that the movement's velocity is almost constant, then we can expect any pixel to be found in a well-known radius around its main line of movement. In this way, we can filter outliers that do not belong to moving objects, while catching all objects' traces in the video snip. When noticing these pixels, we showed in equation (13) the mathematic way of reconstructing the images, so scene would be seen static with a separated notable movement. Another alternative is not showing anything until a movement occurs, then alarming, showing up with a live display having the object's location and direction marked, and letting a human eye decide whether the suspected area contains a threat or not.

4 Experiments

4.1 Turbulence Removal from Static Scene and Static Camera

To demonstrate the behavior of our method, we took 30 frames from a gray-scale video stream taken by a static camera and artificially deformed. We used optical flow algorithm described in [5] to calculate the motion fields between input frames. Some tests we performed showed that the pyramid sample factor could be lowered to 0.7 to affect real-time performance, maintaining good accuracy. Samples of calculated motion fields are presented in Fig. 3.

Notice the fluency in these fields as it resembles the hot air waves. The reconstructed frame compared to its deformed version are shown in Fig. 4.

4.2 Turbulence Removal from Static Scene and Moving Camera

Here we took 30 frames taken by another gray-scale camera investigating a view from a straight angle. As explained before, we subtracted the camera's motion in order to reconstruct the images and then shifted it back to preserve the original video's appearance. Examples are presented in Fig. 5.

We point at buildings' edge lines and windows' arcs reconstructed.

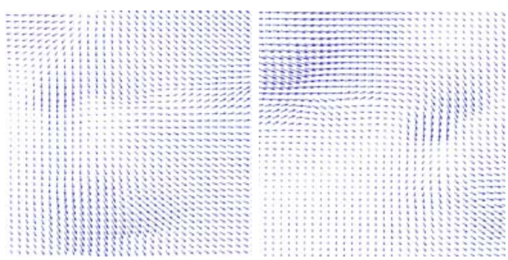
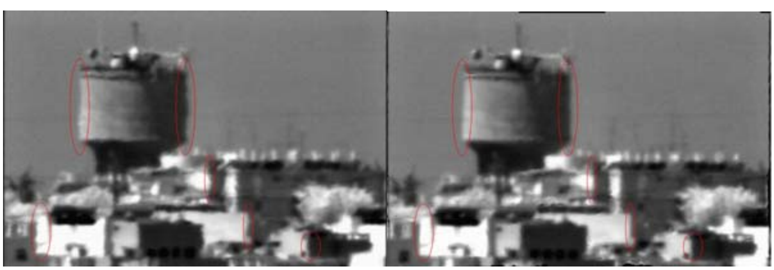


Fig. 3. Turbulence motion fields



(a) (b)

Fig. 4. (a)Deformed image (b)Reconstructed image

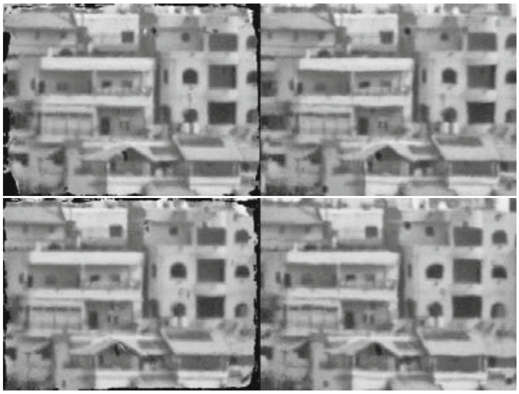
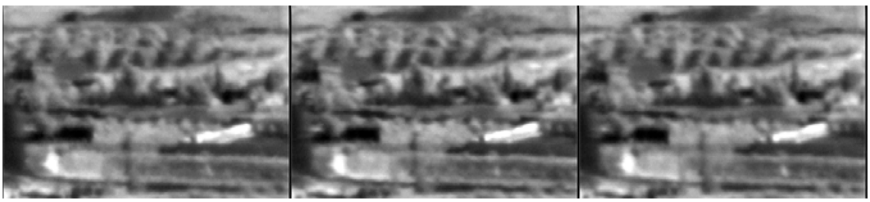
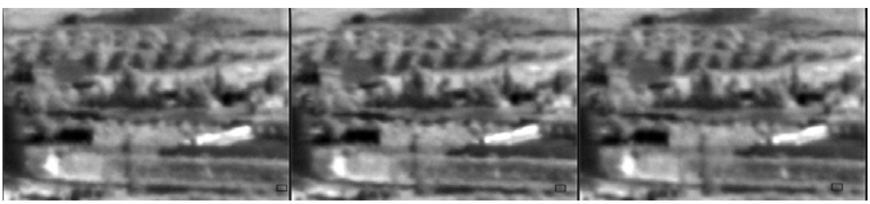


Fig. 5. Aligned images on the left versus their inputs on the right



(a) (b) (c)

Fig. 6. Input images



(a) (b) (c)

Fig. 7. Independent motion detected

4.3 Independent Motion Detection

Independent motion detection was tested on 30 frames from a strongly deformed video stream. After defining the mid-frame as a reference frame, and calculating motion fields from all frames to reference frame, we could see that pixels at right-bottom of the frame had different statistics: their mean move vector was biased, pointing to the left. Gathering them was done using a threshold based on all pixels' movement means, and then a morphologic filter. We show in Fig. 6 three input frames and the algorithm's output on them (fig. 7).

Fig. 8. Original images

Fig. 9. Original image interpolated

Fig. 10. Super resolution result

4.4 Super-Resolution

Taking 30 frames from a translating camera viewing a turbulent scene, we defined again the mid-frame as the reference, and calculated the motion field from it to all other frames and from all frames to it. Then we resized the transformations by a factor of 2, and got the set of transformations the Super Resolution algorithm requires. Added to the camera's PSF we applied the IBP algorithm on the input (results are shown in Figs. 8 – 10).

5 Conclusions and Further Research

In this paper we showed a new mathematical, direct way of extracting the turbulence motion field from a turbulence deformed scene, for static and moving camera without priors, and showed how to deal with independent, simple modeled motion in it. Experiments were performed using that motion field to reconstruct the scene and higher-level algorithms were applied. We showed the success and effects of such a reconstruction. Further researches should include several issues:

1. Better understanding of the phenomena behavior, especially the statistical model. Computation could become more accurate having all motion fields calculated using bundle adjustment, and a spatial constraint added.
2. Local Point Spread Functions' parameters conclusion from motion fields. This could lead to an even better reconstruction, and a new deBlur algorithm should be developed.
3. Solving the case when a more complicated camera movement is present. Though translations seem to match many relevant scenarios, this may add accuracy to the process.
4. Independent motion automatic detection from pixels statistical properties. Experiments showed that only the first steps in this direction were made.

We believe that a basis for a new ability of inspecting and investigating this kind of images was found.

References

1. Shimizu, M., Yoshimura, S., M.T., Okutomi, M.: Restoration of atmospheric turbulent video containing real motion using rank filtering and elastic image registration. In: CVPR (2008)
2. Gepshtein, S., Shtainman, A., B.F., Yaroslavsky, L.: Restoration of atmospheric turbulent video containing real motion using rank filtering and elastic image registration. In: EUSIPCO (2004)
3. Elad, M., Feuer, A.: Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. Image Processing (1997)
4. Irani, M., Peleg, S.: Improving resolution by image registration. Graphical Model and Image Processing (1991)
5. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004)