Signal Processing: Image Communication 63 (2018) 19–29

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

Stabilization of atmospheric turbulence-distorted video containing moving objects using the monogenic signal

Chao Zhang [a], Fugen Zhou [a], Bindang Xue [a,*], Wenfang Xue [b]

[a] Image processing center, Beihang University, Beijing, 100191, China
[b] Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China

ARTICLE INFO

Keywords:
Atmospheric turbulence
Video stabilization
Monogenic signal

ABSTRACT

The motion of moving objects in turbulence-distorted videos is affected by both the atmospheric turbulence fluctuations and the objects' own movements. Therefore, simultaneous video stabilization and preservation of the moving objects is a challenging task. Using the monogenic signal, we propose a fast two-stage approach to mitigate these erratic motions in videos and preserve the moving objects. In the first stage, each frame of a video is represented by the monogenic signal, which is used to model the turbulence-induced random brightness scintillation and local wiggles through local monogenic amplitude and phase. Then, a low-pass filter is employed to attenuate the high spatial frequency of local monogenic amplitude and phase variations to remove turbulence-induced distortions and obtain stable background frames. In the second stage, a two-step mask generating scheme is proposed to preserve the moving objects. Firstly, the coarse masks of the moving objects are obtained using the difference images between distorted frames and stable background frames. The coarse masks are then refined through analyzing the difference images of the local monogenic amplitude and phase between distorted frames and stable background frames. Finally, the stable video frames containing moving objects are reconstructed from the refined masks and the stable background frames. Experimental results show that the proposed approach is efficient and provides stabilized video and preservation of moving objects simultaneously in atmosphere turbulent conditions.

1. Introduction

When imaging in atmosphere turbulent conditions, the turbulence-induced wiggles will be superimposed over the moving objects' motions. In order to stabilize the turbulence induced wiggles efficiently and retain the real objects' motion, it is necessary to propose a method for simultaneous video stabilization and moving object preservation. There are two challenges in this task. One is the speed of video stabilization; the other is how to eliminate the turbulence distortion without harming the moving objects.

The existing methods for turbulence-degraded video processing generally fall into two categories.

The first category of methods is mainly used for obtaining a single high-quality image from a distorted video under the static scene, such as lucky imaging [1], speckle imaging [2,3], image restoration methods [4,5], Sobolev gradient flow method [6], grid smoothing method [7], non-rigid registration method [8], and registration-fusion hybrid method [9]. However, these methods cannot be used for the stabilization of turbulence-distorted video containing moving objects.

The other category of methods is used for detecting moving objects in the turbulence-distorted video without video stabilization. To detect moving objects, Gaussian Mixture Model (GMM) [10] is a popular method of background modeling. This method defines the variation of pixel intensity over time as GMM, which is effective for the slow background changes in the flat region. However, in the turbulent conditions, the random brightness scintillation and the variation of edge regions will contain much high-frequency information, and not conform to Gauss distribution, because the temporal probability distribution of the pixels in the turbulence-distorted video is space-variant [11]. GMM based background subtraction method [12] was proposed to detect moving objects. To mitigate turbulent distortion and use GMM to model the background effectively, B-spline based non-grid registration technique was first used to estimate the motion field between each frame and temporal averaging frame. However, this method is limited by the high computational cost introduced by B-spline based registration. Besides, using the temporal averaging frame as the reference frame for registration will generate an inaccurate motion field. In atmosphere turbulent conditions, Work [13] proposed a moving object detection and

---

* Corresponding author.
E-mail address: xuebd@buaa.edu.cn (B. Xue).

https://doi.org/10.1016/j.image.2018.01.006
Received 15 December 2017; Received in revised form 24 January 2018; Accepted 24 January 2018

tracking method. Work [14] investigated the effects of the distortions on the classification of moving objects. Work [15] quantified the effects of image restoration on the acquisition of moving objects. All of these works [13–15] involve a moving object detection step, in which these methods first adopt temporal median filter [16] to estimate background model, and then use background subtraction and an adaptive threshold to segment the moving objects from the background. Finally, the morphological opening and closing are used to eliminate the false object regions. However, especially in the long-distance video, the size of moving object region is generally small. When the size of the false object region is larger than the small moving object region, morphological operations will also eliminate the correct moving object mask and result in missing detection.

All the above-mentioned methods can do only one thing, either reconstructing a single high-quality image from the turbulence-distorted video or detecting moving objects without stabilizing the distorted video containing moving objects. In order to simultaneously stabilize video and preserve moving objects, the low-rank matrix decomposition method [17] was proposed. This method decomposes the turbulence sequence into three components: the background, the turbulence and the objects. However, matrix decomposition is a time consuming method for high resolution image. The optical flow based method [18] first computes the motion vector maps using optical flow technique and then stable background frames are estimated by motion vector maps, finally moving objects are preserved by the segmentation between the motion vector maps and each frame. However, the optical flow technique will take up a large amount of computational burden. In addition, the used optical flow technique in this method is based on the brightness consistency constraint, which cannot be met due to the wave-front angle-of-arrival (AOA) fluctuations of the optical wave propagating through atmospheric turbulence. Therefore, preserved objects are incomplete. The phase based method [19] reconstructs a video through temporal low-pass filtering of the local phase and amplitude within a complex-valued pyramid. However, this method is sensitive to the cutoff frequency of the low-pass filters. The small cutoff frequency is harmful to moving objects. On the contrary, the large cutoff frequency will not mitigate turbulence effectively. Therefore, simultaneous stabilization of atmospheric turbulence-degraded videos and preservation of moving objects remain an open problem.

In this paper, we focus on simultaneously turbulence-distorted video stabilization and moving objects preservation. From the physical point of view, inhomogeneities in the atmosphere result in refractive index variations according to various meteorological parameters. Variations in the refraction index induce multiple refractions of the optical wave propagating in the atmosphere, namely optical wave-front angle-of-arrival (AOA) fluctuation, which is the phase related parameter of the optical wave propagating through atmospheric turbulence. It results in the appearance of high-frequency random brightness scintillation and local wiggles in the recorded long-distance video.

The monogenic signal includes monogenic phase and monogenic amplitude [20], they are insensitive to image contrast and can be effectively used to represent degraded image structure and motion information [21,22]. Therefore, in this paper, we employ monogenic signal to model turbulence effects and propose a fast and efficient method to stabilize turbulence-degraded video containing moving objects. Our method includes two stages, namely turbulence mitigation and moving objects preservation.

In the first stage, each frame of a video is represented by the monogenic signal, which is used to model the turbulence induced random brightness scintillation and local wiggles through the local monogenic amplitude and phase. Therefore, a low-pass filter is adopted to temporally remove the high spatial frequency of the local monogenic phase and monogenic amplitude to obtain stable background frames.

In the second stage, we use background subtraction to generate coarse masks of moving objects. Subsequently, a mask refinement technique is proposed to remove unwanted regions and generate accurate masks for moving objects through temporal analysis of the monogenic signal. Finally, we combine the stable background frames and the accurate masks to reconstruct a stable video containing moving objects.

The paper is organized as follows: Section 2 describes the monogenic signal based video stabilization and moving objects preservation framework in detail. Experiments are conducted in Section 3 to test the proposed approach. Finally, we conclude and discuss the direction of future research in Section 4.

2. Proposed method

The proposed video stabilization framework contains two stages (see the diagram in Fig. 1): obtaining the stable background frames and reconstructing the stabilized video preserving moving objects. The first stage consists of three steps: (1) the monogenic signal representation of each frame of the original video; (2) the temporal filtering of the local monogenic amplitude and phase; (3) the reconstruction of the corrected background frames. The second stage includes three steps: (1) the generation of coarse masks for moving objects in the video; (2) mask refinement through temporal analysis of the monogenic signal; (3) the reconstruction of the corrected frame preserving moving objects.

2.1. *Generating stable background frames using the monogenic signal*

2.1.1. *Monogenic signal representation*

The monogenic signal is an important extension of the analytic signal for multidimensional data [20], which is built based on the Riesz transform. The Riesz transform is a multidimensional extension of the Hilbert transform. The Riesz operator's frequency response is $-j\boldsymbol{\omega}/\|\boldsymbol{\omega}\|$. For an image $I(\mathbf{z})$ with $\mathbf{z} = [x, y]$, the Riesz transform is expressed as:

$$I_R(\mathbf{z}) = \begin{pmatrix} I_x(\mathbf{z}) \\ I_y(\mathbf{z}) \end{pmatrix} = \begin{pmatrix} h_x * I(\mathbf{z}) \\ h_y * I(\mathbf{z}) \end{pmatrix} \tag{1}$$

where filters h_x and h_y are characterized by the 2D frequency responses $H_x(\boldsymbol{\omega}) = -j\omega_x/\|\boldsymbol{\omega}\|$ and $H_y(\boldsymbol{\omega}) = -j\omega_y/\|\boldsymbol{\omega}\|$ with $\boldsymbol{\omega} = [\omega_x, \omega_y]$.

The monogenic signal of image $I(\mathbf{z})$ is defined as the combination of $I(\mathbf{z})$ and its Riesz transform.

$$I_M(\mathbf{z}) = (I(\mathbf{z}), I_x(\mathbf{z}), I_y(\mathbf{z})) \tag{2}$$

where $I(\mathbf{z})$ is called the real part of the monogenic signal, $I_x(\mathbf{z})$ and $I_y(\mathbf{z})$ are two imaginary parts.

Based on the real and imaginary parts, the original image $I(\mathbf{z})$ can be decomposed into local amplitude $A(\mathbf{z})$ and local phase $\varphi(\mathbf{z})$, which are defined as

$$A(\mathbf{z}) = \sqrt{I(\mathbf{z})^2 + I_x(\mathbf{z})^2 + I_y(\mathbf{z})^2} \tag{3}$$

$$\varphi(\mathbf{z}) = -sign(I_x(\mathbf{z}))\,\mathrm{atan2}\left(\sqrt{I_x(\mathbf{z})^2 + I_y(\mathbf{z})^2}/I(\mathbf{z})\right) \quad \varphi \in [0, 2\pi] \tag{4}$$

where $A(\mathbf{z})$ describes the local energetic information and $\varphi(\mathbf{z})$ describes the local structural and motion information.

2.1.2. *Temporal low-pass filtering of the local amplitude and local phase*

The impacts of atmospheric turbulence on imaging appear as pixel intensity scintillation and pixel temporal oscillation. According to the phase-based image processing [19,23] techniques, the degraded effects of atmospheric turbulence can be transformed to analyze the local amplitude and local phase of the monogenic signal.

Assume we are given an image sequence without turbulence effects $I(\mathbf{z}, t)$, where t is the frame number. The image model used in phase-based processing [20] is

$$I(\mathbf{z},t) = A(\mathbf{z},t)cos\,[\varphi(\mathbf{z}, t)] \tag{5}$$

where $A(\mathbf{z}, t)$ is the local amplitude and $\varphi(\mathbf{z}, t)$ is the local phase. If the video is not influenced by turbulence effects, $A(\mathbf{z}, t)$ and $\varphi(\mathbf{z}, t)$ will be