Moving Object Detection in Turbulence Degraded Video

Nafiz ARICA¹, Tufan CALISKAN²

Abstract: Atmospheric turbulence causes blurring and geometrical distortions in images acquired from a long distance. It makes it difficult to detect moving objects due to both the irregular movements and deformations of the pixels. In this study, we propose a fast method to detect moving objects in turbulence-degraded image sequences. It combines an efficient registration and background subtraction techniques. Since we model the image degradation as local linear deformations, it is estimated by the motion patterns calculated by optical flow. We utilize feature based optical flow and incremental reference frame generation in registration stage. After warping the frames using the registration result GMM based background subtraction technique detects moving objects in stabilized frames. The experiments performed on common image sequences show that the proposed method detects moving objects faster than the available methods, without distorting the objects.

Keywords: Moving object detection, Turbulence degraded video, Background subtraction.

1. Introduction

Atmospheric turbulence occurs by the irregular air motions such as hot air and winds that randomly vary in magnitude and direction. It causes unsteady deformations in an image sequence such as geometric distortion, space-time varying blur and motion blur even if both object and camera are static. The elimination of atmospheric turbulence is emerging as an important need in applications of video surveillance. Especially, the problem becomes extremely difficult if the object of interest moves in the scene. As the turbulence causes deformations and motions in pixel level, the object's motion may become a part of these turbulence effects. Therefore, it's needed to perform turbulence elimination as a preprocessing step for moving object detection.

Several studies were proposed for the turbulence elimination problem such as multi-frame reconstruction [1], [2], [3], Lucky-imaging [4] and Fourier based approaches [5]. Those studies assume that the scene and image sensor are both static (no moving objects). Motions are only caused by turbulence so that the moving objects either blurred or labeled as an outlier and removed from the image.

On the other hand, background subtraction techniques assume that most part of the scenes are static thus foreground can be extracted via background estimation which is obtained from some statistical computation such as mean, and median of all pixels in observed frames [6]. However, most of these methods have low success ratio for modeling widely changing background, which is common in turbulence-degraded images [7]. Therefore, performing any background subtraction technique to images effected by atmospheric turbulence without any preprocessing does not work for moving object detection.

Few studies were presented for moving object detection in turbulence-degraded videos. To the best of our knowledge, the first method in this area utilizes Low-Rank Matrix Decomposition (LRMD) approach and decomposes the image sequence into three parts: the background, the turbulence and the moving object [8]. The authors simplify the problem into a minimization of nuclear norm, Frobenius norm and L1 norm. The distorted image sequence is first preprocessed for reducing the spurious and random noise by temporal averaging. Then turbulence noise, modeled by both the intensity and motion cues is used to obtain an object confidence map (OCM). OCM is a probability density function that gives cues about a moving pixel which belongs to foreground or turbulence. Three term low rank optimization algorithm gets OCM and distorted image sequence as input and produces background, turbulence and moving objects. The main advantage of this method is to prevent moving objects from geometric deformation while removing atmospheric turbulence. However, LRMD method uses a single Gaussian function to model turbulence so that it doesn't work under severe turbulence which is common in long range surveillance. Besides, all frames are needed in minimization problem. Therefore, LRMD method is not proper for online applications.

Recently, another method was proposed to combine non-rigid image registration and Gaussian Mixture Modeling (GMM) [9]. In the first step, a B-spline based registration algorithm is used to estimate the motion field in each observed frame according to the reference frame obtained by temporal averaging of the image sequence. The deformation vectors are generated based on the displacement of the equally spaced dense control points. In the second step, GMM based background subtraction technique is used to detect moving objects from stabilized images. The process time for registration is excessively long and it uses all the image sequence to obtain a reference image so that it is not proper for online applications as well. Besides, some deformations in moving object may appear because of incorrectly shifted control points.

In this paper, we propose a fast method, which combines an effective feature-based image registration technique, cumulative temporal averaging and GMM based background subtraction technique. Unlike the other studies, the degradation caused by atmospheric turbulence is modeled as local linear deformations.

---

¹ Computer Engineering Department, Bahcesehir University, Istanbul, Turkey
² Turkish Naval Academy, Istanbul, Turkey
* Corresponding Author: Email: nafiz.arica@eng.bahcesehir.edu.tr
# This paper has been presented at the International Conference on Advanced Technology & Sciences (ICAT'15) held in Antalya (Turkey), August 4-7, 2015.
instead of nonlinear estimated by the motion patterns of feature based optical flow. We generate the initial reference image by temporal averaging of the predefined number of frames. It is then updated by the subsequent frames of video sequence in a cumulative manner. After warping the frames based on the registration result, GMM based background subtraction technique is applied to extract background and foreground separately. Two advantages of proposed method are as follows. First, it is faster than the available studies in the literature because the reference frame generation doesn’t need all the frames in sequence, and an efficient feature based optic flow is employed in registration stage. Second, it detects the moving objects without deforming them. This is due to utilizing feature based control points in warping process instead of dense samples. The paper is organized as follows: Section 2 presents the proposed framework in detail. The experimental evaluation in public datasets is given in section 3. Then it concludes with remarks in section 4.

2. Proposed Method

The proposed method consists of 3 major parts; reference image generation, image registration and background subtraction. The block diagram of our approach is shown in Figure 1.

A. Incremental Reference Image Generation

Most of the multi-frame image reconstruction approaches for turbulence mitigation utilize the reference image for image registration [1]. Reference image can be estimated simply by averaging all the observed frames in the image sequence. In this study, we use cumulative temporal averaging technique.

\[
\text{Re} f_{i+1} = \frac{F_{i+1} + i \cdot \text{Re} f_i}{i+1}
\]

(1)

where \( F, \text{Ref} \) and \( i \) are the target image, reference image and index of an image respectively. We use first M frames to obtain initial reference image. Then, it is incrementally updated as the new frames in sequence are processed.

B. Image Registration using Feature based Optical Flow

Optical flow is originally developed for motion estimation and video compression. In our case the degradation caused by atmospheric turbulence is modeled as the motion patterns in optical flow [12]. Therefore, unlike the other studies, the displacement vectors between two frames are estimated using feature based optical flow. The optical flow estimation is an alternative and effective way to find local motion between reference and observed frame. In optical flow estimation, it is assumed that the gray value of a pixel does not change due to the displacement;

\[
I(x, y, t) = I(x + u, y + v, t + 1)
\]

(2)

where \( I \) denotes an image and \( w = (u, v, 1) \) is the displacement vector between two images at time \( t \) and \( t + 1 \). However, optical flow estimation runs into problems at gradient vanishing points (mostly at object borders), which is called aperture problem. Thus, to avoid from aperture problem, it is useful to introduce another assumption, called smoothness assumption. It considers the interactions between neighboring pixels in estimating the pixel displacements. Hence, the possible shape deformations caused by the B-spline based method will be prevented. In our study, we implement an algorithm similar to [10]. The fundamental difference lies in the feature extraction and descriptor matching stages, which are described in the following subsections. An optical flow pattern between a target frame and the reference is shown in Figure 2.

1) FAST Interest Points: In the first step, we extract Features from Accelerated Segment Test (FAST) features [11] in order to identify the interest points in the images. FAST interest points have well-defined positions and high local information content. In addition they are robustly detected and repeatable between different images. The most promising advantage of FAST corner detector is its computational efficiency. Hence it is very convenient for real-time or near real-time video processing applications. Unlike the original FAST algorithm, we utilize 9 contiguous pixels out of the 16 to speed up the feature extraction. Additionally, a non-maximal suppression algorithm is used for removing adjacent interest points.

2) Feature Matching: After feature extraction process, feature descriptors with 121 dimensions are obtained using 11x11 neighborhood pixels of any interest point. Finally, sum of absolute difference technique is utilized for descriptor matching for optical flow estimation.

3) Image Warping: Image warping is a transformation, which maps one selected position in observed image plane to position in reference image plane. It’s a critical stage for image registration process. After the local motion vectors in observed frame are found via optical flow, all pixels in observed frame are warped backward using those motion vectors. As a result
geometric distortions are corrected with respect to reference image. We used 2D bilinear interpolation technique to warp images.

4) Moving Object Detection Using GMM Based Background Subtraction Technique:

GMM based background subtraction models multi-modal background distributions. In this study, we employ GMM method in [11] to detect moving objects in stabilized image sequence. In this method, the pixels which correspond to background are determined based on the persistence and the variance of each of the Gaussians of the mixture. If any pixel value does not fit with the background distributions then it is considered as foreground. It adapts lighting changes, repetitive motions and slow-moving objects robustly. Therefore, GMM method is suitable for background subtraction of turbulence-degraded images. In GMM method, online mixture model is proposed which considers the values of pixel over time. Furthermore, it puts emphasis on recent observations in updating Gaussian parameters. The recent history of a pixel is modeled using K (3 to 5) distributions. The probability of observing the current pixel value is defined as:

\[ p(X_t) = \sum_{k=1}^{K} W_{k,t} \eta(X_t, \mu_{k,t}, \Sigma_{k,t}) \]  

(3)

where K, \( W_{k,t} \) and \( \eta \) are number of distributions, an estimate of weight and Gaussian probability density function respectively, of the \( k^{th} \) Gaussian in the mixture at time \( t \). Every new pixel value is compared with the existing K Gaussian distributions until a match is found. If a pixel value within a predefined standard deviations of a distribution, it is labeled as a match. If none of the K distributions match the current pixel value, then the least probable distribution is replaced with a new distribution. The current pixel value is assigned to the new distribution as its mean value and to separate from existing distributions it’s variance is set high. The prior weights of the K distributions at time \( t \) can be adjusted as follows:

\[ W_{k,t} = (1 - \alpha)W_{k,t-1} + \alpha(N_{k,t}) \]  

(4)

where \( \alpha \) and \( N_{k,t} \) are learning rate and match score which is 1 for matched model and 0 for unmatched models respectively. If match score is 0 then mean and variance of unmatched distribution remains same. Otherwise, new observations can be updated as follows:

\[ \mu_t = (1 - \rho)\mu_{t-1} + \rho X_t \]  

(5)

\[ \sigma_t^2 = (1 - \rho)\sigma_{t-1} + \rho(X_t - \mu_t)^T(X_t - \mu_t) \]  

(6)

Background modeling estimates the Gaussians in a mixture that accurately represents the background. Finally, connected component analysis is applied to get foreground object in observed frame.

3. Experimental Results

In performance analysis of our approach, we compare it with two recent methods on common datasets. The first method employs an approach based on Low-Rank Matrix [8]. To our knowledge, it performs the best results in common datasets in evaluation of similar studies. The second method removes air turbulence by B-spline based image reconstruction algorithm and GMM based background subtraction algorithm [9]. We use four infrared videos (A, B, C, D), which are severely distorted by turbulence and contain moving objects. The algorithms are implemented in MATLAB and run on Intel Core i5 CPU (2.5 GHz).

One of the parameters in the method is the initial number of frames, \( M \) in incremental reference frame generation stage. To determine the optimal value for M, we design an experimental set and calculate Peak Signal-to-Noise Ratio (PSNR) between the first frame and all the remaining frames, using various M values starting from 5 to 100. The result is illustrated in Figure 3. It is clearly seen that 50 is the reasonable value for M with respect to PSNR and process time.

**Figure 3.** PSNR of all frames with different values of M

In the experiments, learning rate \( \alpha \) is set as 0.005 in GMM based background subtraction. To compare the recent studies we utilize PSNR as a performance metric as in [8]. We measure the PSNR between the first frame and the rest of the frames, and report the average results for all the frames in Table 1. As shown in the table, our method gives the higher PSNR values in all sequences.

**Table 1.** Comparison of the average PSNR

| Data Set | Original | Matrix Decompos. [8] | B-Spline [9] | Proposed Method |
|----------|----------|----------------------|-------------|-----------------|
| A        | 26.55    | 31.62                | 32.05       | 33.16           |
| B        | 27.49    | 34.11                | 34.74       | 35.36           |
| C        | 27.71    | 32.25                | 32.56       | 34.17           |
| D        | 27.82    | 33.13                | 33.42       | 34.56           |

Figure 4 shows some detection results on random frames in different video data. It can be observed that the foreground pixels are correctly extracted without any spurious pixels. In addition, we can say that the proposed approach prevents geometric deformations in detected objects (see Figure 5). If we compare the detection results with the method in [9], while our method extracts the human without deformation, [9] shears the human shape because of incorrectly shifted control points. Note that the matrix decomposition technique [8] also detects objects without deformation.

As a consequence of cumulative temporal averaging in reference image generation, the proposed method is able to detect moving objects after a processing time of only 2 seconds. Each frame is then processed in 2.76 seconds for the complete registration and moving object detection stages. This is due to the efficient estimation of local motion vectors using optical flow with FAST features. On the other hand, in the matrix decomposition technique [8], all frames are needed to obtain OCM and it takes approximately 17 minutes in the experimental data to obtain a result. Although the decomposition algorithm computes other stages efficiently as in our method, it doesn’t work without calculating OCM. In B-spline based method [9], the process time for registration takes approximately 14 minutes for each frame. Therefore, the experimental results show that the proposed method detects moving objects faster than the available methods, without distorting them.
4. Discussion and Conclusion

In this study, we propose an alternative approach for removing atmospheric turbulence and detecting moving objects from a sequence of degraded images. The proposed method utilizes a speeded up registration technique which employs a feature-based optical flow. It makes our algorithm faster than the other methods. Besides, our method is adaptable to the video surveillance applications easily because of its capability of almost online working. For further studies, our goal is to extend the proposed method for restoring moving objects in degraded image sequence for moving environment in which neither scene nor camera is static. In addition we will try to make our algorithm much faster using GPU or FPGA based implementation, so that it can be used in real-time or near real-time applications.

Figure 5: Deformations in detected objects (a) Detection in [9]. (b) Detection in our method

References

[1] M. Shimizu, S. Yoshimura, M. Tanaka, and M. Okutomi, “Super-Resolution from Image Sequence under Influence of Hot-Air Optical Turbulence,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, June 2008.

[2] X. Zhu and P. Milanfar, “Removing atmospheric turbulence via space-invariant deconvolution,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. PP, no. 99, pp. 1, 2012.

[3] X. Zhu and P. Milanfar, “Image Reconstruction from Videos Distorted by Atmospheric Turbulence,” in SPIE, 2010.

[4] M. Aubailly, M.A. Vorontsov, G.W. Carhat, and M.T. Valley, “Automated Video Enhancement from a Stream of Atmospherically-Distorted Images: The Lucky-Region Fusion Approach,” Proc. SPIE, vol. 7463, 2009.

[5] M. Hirsch, S. Sra, B. Scho¨lkopf, and S. Harmeling, “Efficient Filter Flow for Space-Variant Multiframe Blind Deconvolution,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 607-614, June 2010.

[6] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, “Detecting moving objects, ghosts and shadows in video streams”, IEEE Trans. on Patt. Anal. and Machine Intell., 2003.

[7] C. Stauffer and W. E. L Grimson, “Adaptive background mixture models for real time tracking,” in CVPR, 1999.

[8] O. Oreifej, Xin Li, and M. Shah, “Simultaneous video stabilization and moving object detection in turbulence,” in IEEE Trans. on Patt. Anal. and Machine Intell., 2012.

[9] A. Deshmukh, S. Medasani, andG. Reddy, “Moving Object Detection from Images Distorted by Atmospheric
Turbulence,” in ISSP, 2013.

[10] T. Brox, and J. Malik 2011. Large displacement optical flow: Descriptor matching in variational motion estimation. IEEE Trans. on Pattern Analysis and Machine Intelligence 33, 3 (Mar.), 500–513.

[11] E. Rosten and T. Drummond, “Machine learning for high speed corner detection,” in 9th European Conference on Computer Vision, vol. 1, pp. 430–443. 2006.

[12] T. Caliskan, N. Arica, "Atmospheric Turbulence Mitigation Using Optical Flow", ICPR, pp:883-888, 2014.