

Simultaneous Video Stabilization and Moving Object Detection in Turbulence

Omar Oreifej, *Student Member, IEEE*, Xin Li, *Member, IEEE*, and Mubarak Shah, *Fellow, IEEE*

Abstract—Turbulence mitigation refers to the stabilization of videos with nonuniform deformations due to the influence of optical turbulence. Typical approaches for turbulence mitigation follow averaging or dewarping techniques. Although these methods can reduce the turbulence, they distort the independently moving objects, which can often be of great interest. In this paper, we address the novel problem of simultaneous turbulence mitigation and moving object detection. We propose a novel three-term low-rank matrix decomposition approach in which we decompose the turbulence sequence into three components: the background, the turbulence, and the object. We simplify this extremely difficult problem into a minimization of nuclear norm, Frobenius norm, and ℓ_1 norm. Our method is based on two observations: First, the turbulence causes dense and Gaussian noise and therefore can be captured by Frobenius norm, while the moving objects are sparse and thus can be captured by ℓ_1 norm. Second, since the object's motion is linear and intrinsically different from the Gaussian-like turbulence, a Gaussian-based turbulence model can be employed to enforce an additional constraint on the search space of the minimization. We demonstrate the robustness of our approach on challenging sequences which are significantly distorted with atmospheric turbulence and include extremely tiny moving objects.

Index Terms—Three-term decomposition, turbulence mitigation, rank optimization, moving object detection, particle advection, restoring force

1 INTRODUCTION

THE refraction index of the air varies based on several atmospheric characteristics, including the air's temperature, humidity, pressure, carbon dioxide level, and dust density. Such conditions are typically not homogeneous; for instance, a nonuniform temperature distribution might be observed above a surface receiving sunlight. Therefore, light rays traveling through the air with such nonuniform changes in its relative refraction index will go through a complex series of refraction and reflection, causing extreme spatially and temporally varying deformations to the captured images [1], [2], [3], [4], [5].

On the other hand, if the objects of interest are additionally moving in the scene, their motion will be mixed up with the turbulence deformation in the captured images, rendering the problem of detecting the moving objects extremely difficult. In this paper, we are interested in the dual problem of turbulence mitigation (stabilizing the sequence) and moving object detection under the turbulent medium. To the best of our knowledge, such a problem has never been explored before. Relevant previous approaches have either focused on detecting moving objects or dewarping a deformed sequence, but not on both tasks concurrently. Note that other than image deformation, atmospheric turbulence may cause blur if the camera exposure time is not sufficiently short. In this paper, however, we focus only on image

deformation because of the inherent confusion between the motion of the object and the motion caused by the turbulence.

Given a sequence of frames $\{I_1, \dots, I_T\}$ acquired from a stationary camera residing in a turbulent medium while observing relatively tiny moving objects, we decompose the sequence into background, turbulence, and object components. More precisely, consider the frames matrix $F = [\text{vec}\{I_1\} \cdots \text{vec}\{I_T\}]$ for $I_k \in \mathbb{R}^{W \times H}$ ($k = 1, 2, \dots, T$), where $W \times H$ denotes the frame resolution (width by height) and $\text{vec} : \mathbb{R}^{W \times H} \rightarrow \mathbb{R}^M$ is the operator which stacks the image pixels as a column vector. We formulate our decomposition of F as

$$\min_{A,O,E} \text{Rank}(A) \text{ s.t. } F = A + O + E, \quad (1)$$

$$\|O\|_0 \leq s, \|E\|_F \leq \sigma,$$

where F , A , O , and E are the matrices of frames, background, object, and error (turbulence), respectively. Here, the $\|\cdot\|_0$ -norm counts the number of nonzero entries, $\|\cdot\|_F$ -norm is the Frobenius norm which is equal to the square root of the sum of squared elements in the matrix, s represents an upper bound of the total number of moving objects' pixels across all images, and σ is a constant which reflects our knowledge of the maximum total variance due to corrupted pixels across all images.

Our decomposition is based on the intrinsic properties of each of the components.

1. *The background.* The scene in the background is presumably static; thus, the corresponding component in the frames of the sequence has linearly correlated elements. Therefore, the background component is expected to be the part of the matrix which is of low rank. Minimizing the rank of the low-rank component of the frames matrix F emphasizes the structure of the linear subspace containing

• The authors are with the University of Central Florida, 4000 Central Florida Blvd, Orlando, FL 32816.

E-mail: {oreifej, shah}@eecs.ucf.edu, xli@math.ucf.edu.

Manuscript received 25 Aug. 2011; revised 15 Jan. 2012; accepted 5 Apr. 2012; published online 19 Apr. 2012.

Recommended for acceptance by M. Brown.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2011-08-0584.

Digital Object Identifier no. 10.1109/TPAMI.2012.97.

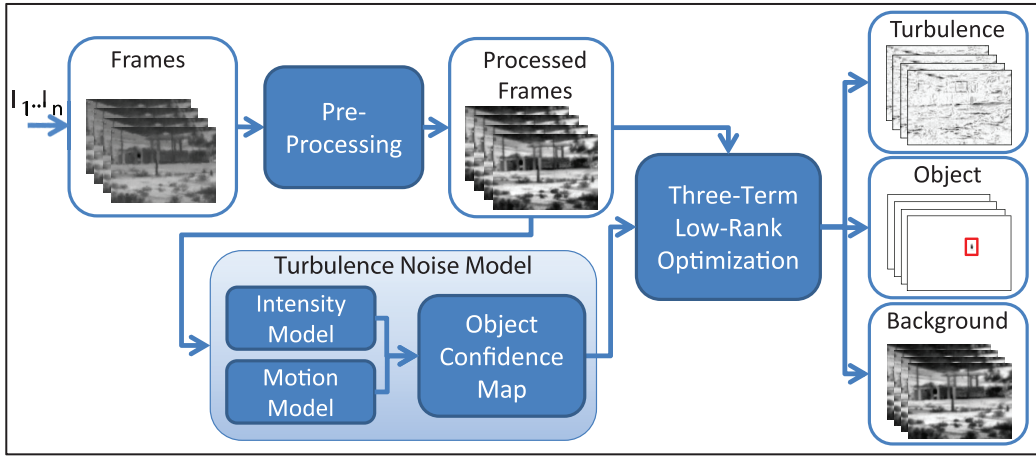


Fig. 1. The various steps of the proposed algorithm.

the column space of the background, which reveals the background.

2. *The turbulence.* Previous work dealing with turbulence, such as [6], [7], [8], [9], [10], demonstrated that the fluctuations of fluids (for instance air and water) attain Gaussian-like characteristics such as being unimodal, symmetric, and locally repetitive; therefore, the projected deformations in the captured sequence often approach a Gaussian distribution (we discuss this in more detail in Section 3.2). For this reason, the turbulence component can be captured by minimizing its Frobenius norm. The Frobenius norm of a matrix is the same as the euclidean norm of the vector obtained from the matrix by stacking its columns. Therefore, as in the well-known vector case, constraining the error in the euclidean norm is equivalent to controlling the sample variance of the error. Furthermore, theoretically, the estimate obtained by the Frobenius norm has several desirable statistical properties [11].
3. *The moving objects.* We assume that the moving objects are sparse in the sequence. This means that the number of pixels occupied by the moving objects is small (or can be considered as outliers) compared to the total number of pixels in the frames. This is a reasonable assumption for most realistic surveillance videos. For this reason, the moving objects are best captured by restricting the number of nonzero entries (denoted by the ℓ_0 norm of the matrix), which is desirable for finding outliers.

In practice, parts of the turbulence could also appear as sparse errors in the object matrix O . Therefore, an additional constraint needs to be enforced on the moving objects. We employ a simple turbulence model to compute an object confidence map which is used to encourage the sparse solutions to be located on regions exhibiting linear motion that is dissimilar from the fluctuations of the turbulence. Under the new constraint, the optimization problem (1) must be reformulated as

$$\begin{aligned} \min_{A, O, E} \text{Rank}(A) \quad \text{s.t.} \quad F = A + O + E, \\ \|\Pi(O)\|_0 \leq s, \|E\|_F \leq \sigma, \end{aligned} \quad (2)$$

where $\Pi: \mathbb{R}^{M \times T} \rightarrow \mathbb{R}^{M \times T}$ is the object confidence map, which is a linear operator that weights the entries of O according to their confidence of corresponding to a moving object such that the most probable elements are unchanged and the least are set to zero.

Fig. 1 shows a diagram of the proposed approach. We first apply a preprocessing step to improve the contrast of the sequence and reduce the spurious and random noise. Consequently, we obtain an object confidence map using a turbulence model which utilizes both the intensity and the motion cues. Finally, we decompose the sequence into its components using three-term rank minimization.

This paper makes three main contributions: First, we propose a new variant of matrix decomposition based on low-rank optimization and employ it to solve the novel problem of simultaneous moving object detection and turbulence mitigation in videos distorted by atmospheric turbulence. Second, we propose a turbulence model based on both intensity and motion cues, where the motion distribution is derived from the Lagrangian particle advection framework [12]. The turbulence model is used to enforce an additional constraint on the decomposition, which encourages the sparse solutions to be located in areas with non-Gaussian motion. Finally, we propose an additional force component in the particle advection framework in order to stabilize the particles in the turbulent medium and handle long sequences without discontinuities.

The rest of the paper is organized as follows: In the next section, we discuss the related works. In Section 3, we present our three-term rank minimization, followed by our probabilistic formulation of the object confidence map using particle advection. Section 4 discusses the details of implementation and the technical challenges of the three-term optimization. The experiments and the results are described in Section 5. Finally, Section 6 concludes the paper.

2 RELATED WORK

Rank optimization-based video denoising has recently flourished with several successful works reported, from which we will only discuss the most related articles. Robust PCA was proposed in [13], where a low rank matrix was recovered from a small set of corrupted observations

through convex programming. Similar concepts were later employed in [14] for video denoising, where serious mixed noise was extracted by grouping similar patches in both the spatial and temporal domains, and solving a low-rank matrix completion problem. Additionally, in [15], linear rank optimization was employed to align faces with rigid transformations, and concurrently detect noise and occlusions. In [16], Yu and Schuurmans proposed an efficient solution to subspace clustering problems which involved the optimization of unitarily invariant norms. Another variant of such space-time optimization techniques is the total variation minimization, where, for instance, in [17] Chan et al. posed the problem of video restoration as a minimization of anisotropic total variation given in terms of ℓ_1 -norm, or isotropic variation given in terms of ℓ_2 -norm. Consequently, the Lagrange multiplier method was used to solve the optimization function.

On the other hand, moving object detection is a widely investigated problem. When the scene is static, moving objects can be easily detected using frame differencing. A better approach would be to use the mean, the median, or the running average as the background [18]. The so-called eigenbackground [19] can also be obtained using PCA. However, when the scene is constantly changing because of noise, light changes, or camera shake, the intensities of image pixels can be considered as independent random variables, which can be represented using a statistical model such as a Gaussian, a mixture of Gaussians, or a kernel density estimator. The model can then be used to compute the probability of each pixel belonging to either the background or the foreground. Examples of such approaches include [20], [21], [22]. Additionally, the correlation between spatially proximal pixels could also be employed to improve the background modeling using a joint domain (location) and range (intensity) representation of image pixels such as in [23].

Approaches for turbulence mitigation focused mainly on registration-based techniques. In [6], [5], and [24], both the turbulence deformation parameters and a super-resolution image were recovered using area-based B-Spline registration. Moreover, in [25], Tian and Narasimhan proposed recovering the large nonrigid turbulence distortions through a “pull-back” operation that utilizes several images with known deformations. In model-based tracking [26], the characteristics of the turbulence caused by water waves were employed to estimate the water basis using PCA. More recently, in [7], turbulence caused by water was overcome by iteratively registering the sequence to its mean followed by RPCA to extract the sparse errors. Averaging-based techniques are also popular for video denoising and turbulence mitigation, including pixel-wise mean/median, nonlocal means (NLM) [27], [28], Fourier-based averaging [29], and speckle imaging [1], [3], [30]. Another category of methods for turbulence mitigation is the lucky region approach [4], [31], [32], where the least distorted patches of the video are selected based on several quality statistics, then those selected patches are fused together to compose the recovered video.

Clearly, previous work in moving object detection in dynamic scenes mostly focused on detecting the objects and

did not consider recovering the background. Inversely, previous work in turbulence mitigation did not consider the possible interest in detecting moving objects in the scene. In this paper, we pose the two problems of moving object detection and turbulence mitigation as one application for our proposed three-term low-rank decomposition. We demonstrate how to decompose a turbulent video into separate background, foreground, and turbulence components. It is important to note that our method is not directly comparable to background subtraction or turbulence mitigation approaches, though we do provide competitive results on each task separately.

3 PROPOSED APPROACH

We decompose the matrix which contains the frames of the turbulence video into its components: the background, the turbulence, and the objects. The decomposition is performed by solving the rank optimization in (2), which enforces relevant constraints on each component. In the next section, we describe the details of the decomposition approach.

3.1 Three-Term Decomposition

When solving (2), it is more convenient to consider the Lagrange form of the problem.

$$\begin{aligned} \min_{A,O,E} \text{Rank}(A) + \tau \|\Pi(O)\|_0 + \lambda \|E\|_F^2 \\ \text{s.t. } F = A + O + E, \end{aligned} \quad (3)$$

where τ and λ are weighting parameters. The optimization of (3) is not directly tractable since the matrix rank and the ℓ_0 -norm are nonconvex and extremely difficult to optimize. However, it was recently shown in [13] that when recovering low-rank matrices from sparse errors, if the rank of the matrix A to be recovered is not too high and the number of nonzero entries in O is not too large, then minimizing the nuclear norm of A (sum of singular values $\sum_i \sigma_i(A)$) and the ℓ_1 -norm of O can recover the exact matrices. Therefore, the nuclear norm and the ℓ_1 -norm are the natural convex surrogates for the rank function and the ℓ_0 -norm, respectively. Applying this relaxation, our new optimization becomes

$$\min_{A,O,E} \|A\|_* + \tau \|\Pi(O)\|_1 + \lambda \|E\|_F^2 \quad \text{s.t. } F = A + O + E, \quad (4)$$

where $\|A\|_*$ denotes the nuclear norm of matrix A . We adopt the Augmented Lagrange Multiplier method (ALM) [33] to solve the optimization problem (4). Define the augmented Lagrange function for the problem as

$$\begin{aligned} L(A, O, E, Y) = \|A\|_* + \tau \|\Pi(O)\|_1 + \lambda \|E\|_F^2 + \\ \langle Y, F - A - O - E \rangle + \frac{\beta}{2} \|F - A - O - E\|_F^2, \end{aligned} \quad (5)$$

where $Y \in \mathbb{R}^{M \times T}$ is a Lagrange multiplier matrix, β is a positive scalar, and $\langle \cdot, \cdot \rangle$ denotes the matrix inner product ($\text{trace}(A^T B)$). Minimizing the function in (5) can be used to solve the constrained optimization problem in (4). We use the ALM algorithm to iteratively estimate both the Lagrange multiplier and the optimal solution by iteratively minimizing the augmented Lagrangian function:

$$(A_{k+1}, O_{k+1}, E_{k+1}) = \arg \min_{A, O, E} L(A, O, E, Y_k), \quad (6)$$

$$Y_{k+1} = Y_k + \beta_k (F_{k+1} - A_{k+1} - O_{k+1} - E_{k+1}).$$

When β_k is a monotonically increasing positive sequence, the iterations converge to the optimal solution of problem (4) [34]. However, solving (6) directly is difficult; therefore, the solution is approximated using an alternating strategy minimizing the augmented Lagrange function with respect to each component separately:

$$A_{k+1} = \arg \min_A L(A, O_k, E_k, Y_k),$$

$$O_{k+1} = \arg \min_O L(A_{k+1}, O, E_k, Y_k), \quad (7)$$

$$E_{k+1} = \arg \min_E L(A_{k+1}, O_{k+1}, E, Y_k).$$

Following the idea of the singular value thresholding algorithm [35], we derive the solutions for the update steps in (7) for each of the nuclear, Frobenius, and ℓ_1 norms. Please refer to the supplementary appendix, which can be found in the Computer Society Digital Library at <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2012.97>, for the complete derivations. Consequently, a closed form solution for each of the minimization problems is found:

$$UWV^T = \text{svd}(F - O_k - E_k + \beta_k^{-1}Y_k),$$

$$A_{k+1} = US_{1/\beta_k}(W)V^T,$$

$$O_{k+1} = S_{\tau/\beta_k \Pi}(F - A_{k+1} - E_k + \beta_k^{-1}Y_k), \quad (8)$$

$$E_{k+1} = \left(1 + \frac{2\lambda}{\beta_k}\right)^{-1} (\beta_k^{-1}Y_k + F - A_{k+1} - O_{k+1}),$$

where $\text{svd}(M)$ denotes a full singular value decomposition of matrix M and $S_\alpha(\cdot)$ is the soft-thresholding operator defined for a scalar x as

$$S_\alpha(x) = \text{sign}(x) \cdot \max\{|x| - \alpha, 0\}, \quad (9)$$

and, for two matrices $A = (a_{ij})$ and $B = (b_{ij})$ of the same size, $S_A(B)$ applies the soft-thresholding entry-wise outputting a matrix with entries $S_{a_{ij}}(b_{ij})$.

The steps of our decomposition are summarized in Algorithm 1. In the next section, we describe our method to obtain the moving object confidence map Π , which is employed as a prior in the rank minimization problem.

3.2 Turbulence Model

We employ a turbulence model to enforce an additional constraint on the rank minimization such that moving objects are encouraged to be detected in locations with non-Gaussian deformations. Exact modeling of the turbulence is in fact ill-posed as it follows a nonuniform distribution which varies significantly in time, besides having an additional complexity introduced during the imaging process, thus rendering the problem of modeling turbulence extremely difficult. Although the refraction index of the turbulent medium is often randomly changing, it is also statistically stationary [3], [30], [36]; thus, the deformations caused by turbulence are generally repetitive and locally centered [6], [7], [8], [9], [37]; this encourages the use of Gaussian-based models as approximate distributions that are general enough to avoid overfitting, but rather capture significant portion of the turbulent characteristics.

Authorized licensed use limited to: CLEMSON UNIVERSITY. Downloaded on April 16, 2024 at 23:22:27 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1: Simultaneous Turbulence Mitigation and Moving Object Detection

Input : Distorted image set stacked as
 $F \in \mathbb{R}^{M \times T} = [\text{vec}\{I_1\} \dots \text{vec}\{I_T\}]$
 Initial Object Confidence Map
 $\Pi : \mathbb{R}^{M \times T} \rightarrow \mathbb{R}^{M \times T}$

Output: Solution to equation (1)
 Background $A \in \mathbb{R}^{M \times T}$
 Turbulence $E \in \mathbb{R}^{M \times T}$
 Moving Objects $O \in \mathbb{R}^{M \times T}$

while not converged do

\Minimize the Lagrange function in equation (5)
 $UWV^T = \text{svd}(F - O_k - E_k + \beta_k^{-1}Y_k),$
 $A_{k+1} = US_{1/\beta_k}(W)V^T,$
 $O_{k+1} = S_{\tau/\beta_k \Pi}(F - A_{k+1} - E_k + \beta_k^{-1}Y_k),$
 $E_{k+1} = (1 + \frac{2\lambda}{\beta_k})^{-1} (\beta_k^{-1}Y_k + F - A_{k+1} - O_{k+1}),$
 $Y_{k+1} = Y_k + \beta_k (F - A_{k+1} - O_{k+1} - E_{k+1}),$
 $\beta_{k+1} = \rho \beta_k.$

end

We use a Gaussian function to model the intensity distribution of a pixel going through turbulence. This is similar to [20], which employs a mixture of Gaussians; however, we found that a single Gaussian worked better since more complicated models often require a period of training which is not available in our sequences. Therefore, the intensity of a pixel at location \mathbf{x} is modeled using a Gaussian distribution:

$$I(\mathbf{x}) \sim \mathcal{N}(\mu_I, \sigma_I), \quad (10)$$

where μ_I and σ_I are the mean and the standard deviation at \mathbf{x} , respectively. On the other hand, the deformation caused by turbulence can be captured in the motion domain besides the intensity. Therefore, we combine the intensity and the motion features to obtain a better model of turbulence. In order to capture the ensemble motion in the scene, we use the concept of a “particle” in a Lagrangian particle trajectory acquisition approach. We assume that a grid of particles is overlaid onto a scene where each particle corresponds to a single pixel (the granularity is controllable). The basic idea is to quantify the scene’s motion in terms of the motion of the particles which are driven by dense optical flow. A so-called particle advection [12], [38], [39] procedure is applied to produce the particle trajectories. Given a video clip $\in \mathbb{R}^{W \times H \times T}$, we denote the corresponding optical flow by (U_w^t, V_h^t) , where $w \in [1, W]$, $h \in [1, H]$, and $t \in [1, T - 1]$. The position vector (x_w^t, y_h^t) of the particle at grid point (w, h) at time t is estimated by solving the following differential equations:

$$\frac{dx_w^t}{dt} = U_w^t, \quad (11)$$

$$\frac{dy_h^t}{dt} = V_h^t.$$

We use Euler’s method to solve them, similarly to [38]. By performing advection for the particles at all grid points with respect to each frame of the clip, we obtain the clip’s

particle trajectory set, denoted by $\{(x_w^t, y_h^t) | w \in [1, W], h \in [1, H], t \in [1, T]\}$.

We employ the spatial locations of the particle trajectories (i.e., (x_w^t, y_h^t)) to model the turbulence motion in the scene. The locations visited by a particle moving due to the fluctuations of the turbulence have a unimodal and symmetric distribution which approaches a Gaussian [6], [30], [10], [9]. This is dissimilar from the linear motion of the particles driven by moving objects. Therefore, we associate each particle with a Gaussian with mean μ_M and covariance matrix Σ_M :

$$\mathbf{x} \sim \mathcal{N}(\mu_M, \Sigma_M). \quad (12)$$

By augmenting the intensity model in (10) with the motion model in (12), the total confidence of corresponding to the turbulence versus the moving objects for a particle at location \mathbf{x} is expressed as a linear opinion pooling of the motion and the intensity cues:

$$C(\mathbf{x}) = wP(I(\mathbf{x}) | \mu_I, \sigma_I) + (1 - w)P(\mathbf{x} | \mu_M, \Sigma_M). \quad (13)$$

The parameters of our model $\{w, \mu_I, \sigma_I, \mu_M, \Sigma_M\}$ can be learned by optimization using training sequences or set to constant values selected empirically. In the context of our three-term decomposition, the obtained confidence provides a rough prior knowledge of the moving objects' locations, which can be incorporated into the matrix optimization problem in (4). Interestingly, this prior employs motion information; therefore, it is complementary to the intensity-based rank optimization and can significantly improve the result.

At frame t , we evaluate all the particles' locations against their corresponding turbulence models and obtain the turbulence confidence map $C_t \in \mathbb{R}^{W \times H}$. While C_t corresponds to the confidence of a particle belonging to turbulence, the desired Π in (4) corresponds to the confidence of belonging to the moving objects; therefore, we define the object confidence map Π as the complement of the stacked turbulence confidence maps:

$$\Pi = 1 - [\text{vec}\{C_1\} \cdots \text{vec}\{C_T\}]. \quad (14)$$

3.3 Restoring Force

The particles carrying the object's motion typically drift far from their original locations, leaving several gaps in the sequence. In the presence of turbulence, the drifting also occurs as a result of the turbulent motion. Therefore, the particles need to be reinitialized every certain number of frames, which, however, creates discontinuities. This is a typical hurdle in the Lagrangian framework of fluid dynamics [12], [38], [39], which constitutes a major impediment for the application of particle flow to turbulence videos. In order to handle the drifting and the discontinuity problems associated with the particle flow, we use a new force component in the advection equation:

$$\begin{aligned} \frac{dx_w^t}{dt} &= U_w^t + G(x, x_o), \\ \frac{dy_h^t}{dt} &= V_h^t + G(y, y_o). \end{aligned} \quad (15)$$

We refer to the new force as "Restoring Force"—a reference to a local restoration force acting in the direction

of the original location of each particle. We use a simple linear function to represent the restoring force

$$G(x, x_o) = \frac{x - x_o}{s}, \quad (16)$$

where s is a scaling factor which trades off the detection sensitivity and the speed of recovery for the particles. In other words, if s is set to a high value, the effect of the restoring force will be negligible and therefore the particles will require a relatively longer time to return to their original positions. In this case, the sensitivity of moving object detection will be higher, but more prone to false positives. If s is low, the particles will be more attached to their original location, thus less affected by turbulence, but will have lower detection sensitivity. In our experiments, we set s to $0.5 \times W = 125$, which we found to be adequate for all sequences.

Using the restoring force allows continuous processing of the sequence without the need to reinitialize the particles. For instance, if an object moves to one side of the frame then comes back, we can still capture its motion when it returns. Additionally, the restoring force maintains the particles' motion within a certain range and provides robustness against random noise, thus reducing the number of false object detections. Fig. 2 shows the overlaid particles on selected frames and the corresponding object particles with and without restoring force. It is clear that the restoring force stabilizes the particles and delivers better moving object confidence.

4 ALGORITHM AND IMPLEMENTATION DETAILS

4.1 Preprocessing

The noise caused by turbulence often has several random and spurious components which are difficult to model. Therefore, we employ temporal averaging to mitigate such components. We use a small averaging window of eight frames to avoid distorting the objects' motion. In order to improve the contrast of the sequences, we apply adaptive histogram equalization, which is similar to the ordinary histogram equalization; however, it operates on small patches of size 8×8 . Deviation around the optimal values of such parameters only causes smooth and graceful degradation in the results.

4.2 Determining the Optimization Parameters

The parameters τ and λ from (4) correspond to the total number of moving objects' pixels across all images and the total variance due to corrupted pixels across all images, respectively. In other words, a higher τ leads to an increased significance in minimizing the O component, thus obtaining sparser moving objects. A higher λ leads to an increased significance in minimizing the noise in the E component, thus obtaining less noise in E and a more turbulent background. On the other hand, decreasing τ and λ leads to placing more emphasis on minimizing the rank, thus obtaining a more static background in A , large turbulence in E , and less sparse moving objects in O .

Several theoretical considerations were previously studied to derive optimal values for similar parameters in [13]. However, such analysis does not apply to all practical

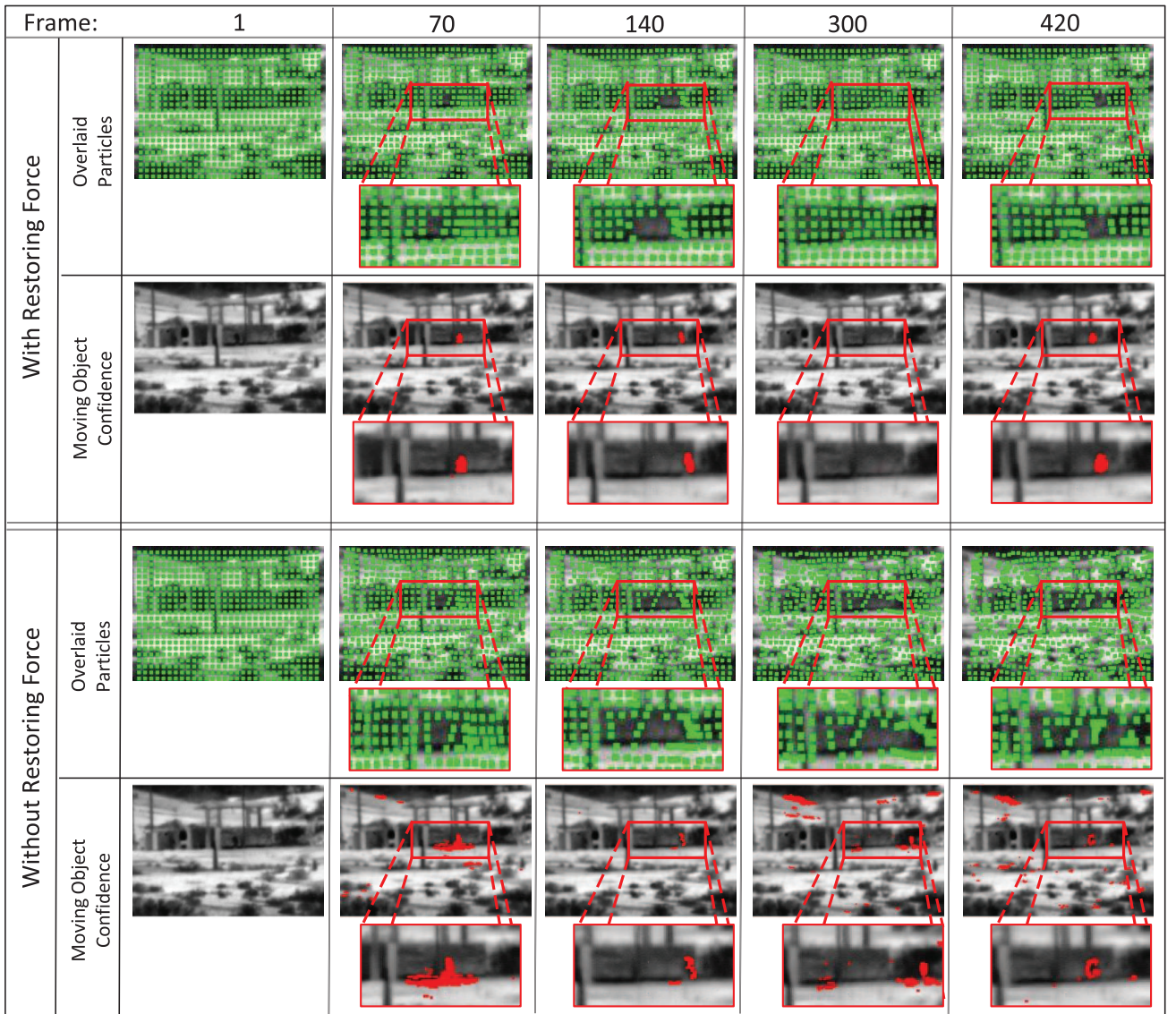


Fig. 2. Overlaid particles (the end points of the particle trajectories) and the corresponding moving object confidence (14) for sample frames. Rows one and three show the overlaid particles (the granularity is reduced for better visualization). Shown in red, in rows two and four, are the particles with high confidence of belonging to a moving object rather than turbulence (a fixed threshold is used). After frame 1, the object starts moving to the right, gets occluded at frame 300, then moves back to the left at frame 420. In the top two rows, the restoring force is employed; therefore, the particles gain two new properties: First, they become intact and attached to their original position, which make them robust against drifting due to the turbulence. Second, the particles automatically return to their original positions after the moving object disappears, which can be seen on frame 300. In the bottom two rows, the restoring force is not employed and such properties are not available; therefore, the particles continue to float and drift along the sequence, resulting in a poor object confidence performance.

scenarios. In the context of our three-term decomposition, the matrix to be decomposed is not an exact composition of the expected components. In addition, the components do not correspond exactly to their expected model. For instance, the background is expected to be of low rank; however, the exact desired rank is debatable as a background of rank 1 is often not desirable since it will be a repetition of a static image, which is not realistic. Similarly, the desired sparsity of error varies significantly among different applications. Therefore, we argue that, in practical scenarios, such parameters are problem-dependant and highly heuristic. Therefore, we empirically set τ to 0.1 and λ to 2.0, which, as we will show, worked as a good compromise among the constraints of the optimization in all sequences.

4.3 Discussion of the Three-Term Model

Our method relies on a special three-term decomposition of matrices, which we formulated as the optimization problem (4). A similar optimization model has been proposed and studied very recently in [40], [41], and [33]. In [40], sufficient conditions are obtained in order to find an optimal solution to (4), recovering the low-rank and sparse components of F . In this paper, however, we are interested in the computational methods for finding the desired decompositions. This leads us to scheme (7) via the Alternating Directions Method of Multipliers (ADMM), which was first introduced in the mid-1970s by Gabay and Mercier [42] and Glowinski and Marroco [43], and is the current method of choice for large-scale nonsmooth convex optimization, as in [41], [44], [45], and [46].

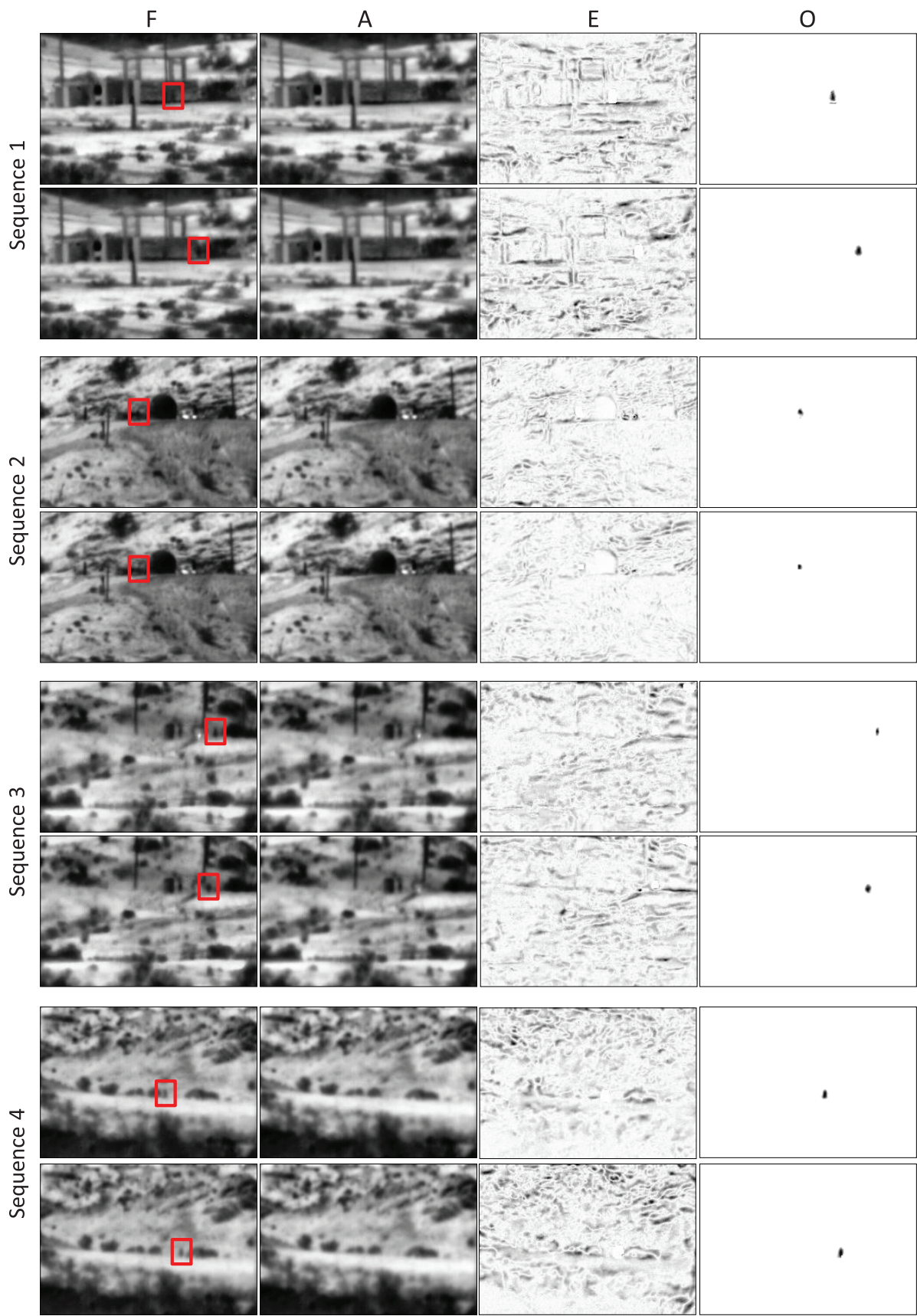


Fig. 3. Three-term decomposition results for two example frames from four testing sequences. Column F shows the original sequence (after preprocessing) which was decomposed into background (column A), turbulence (column E, absolute value of E is shown), and moving object (column O). Please refer to our website for the complete videos as the results of correcting the deformations are difficult to observe in a single frame.

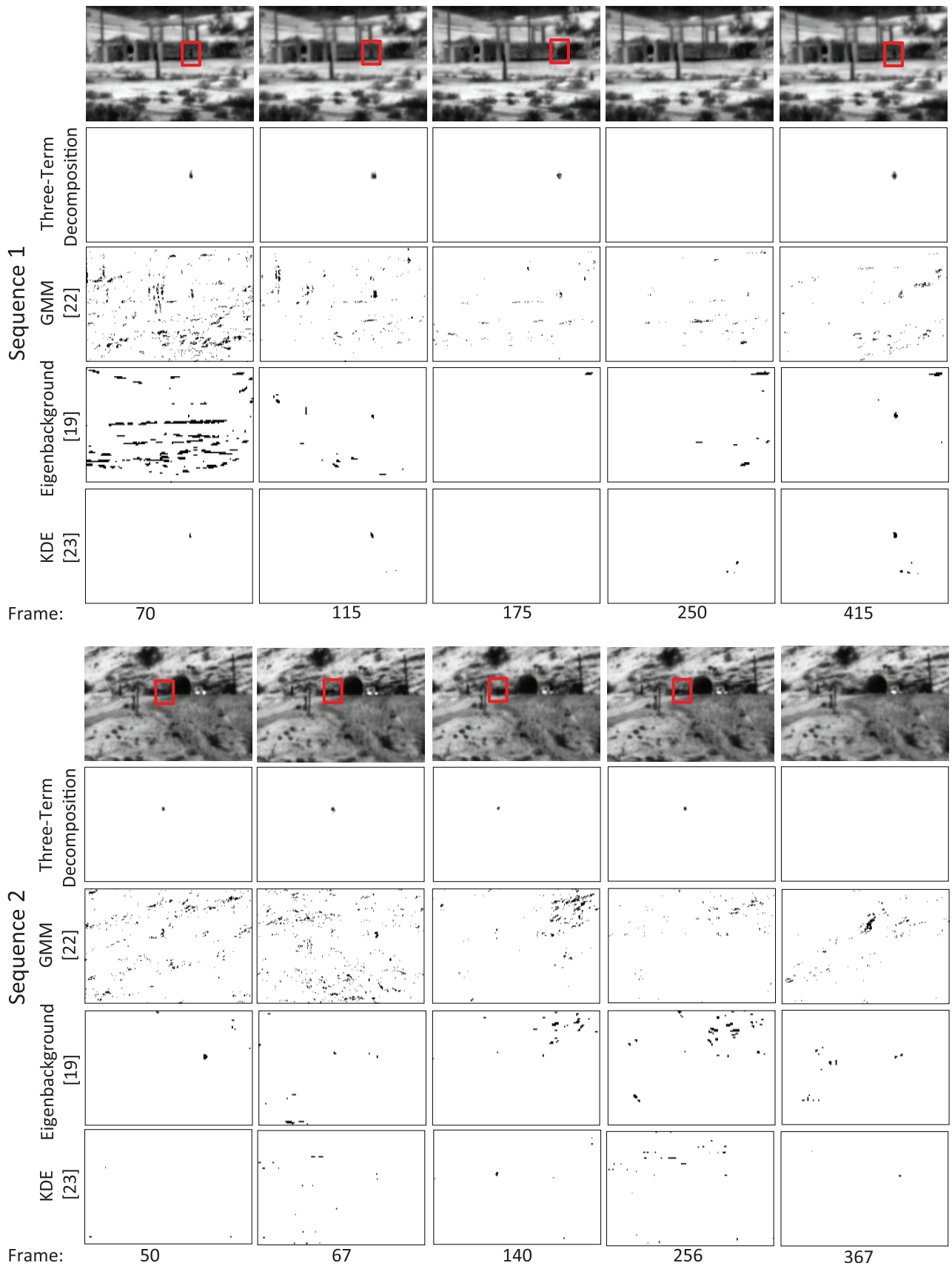


Fig. 4. Our moving object detection results on sample frames from sequences 1 and 2 compared to [22], [19], and [23]. For every sequence, the first row shows the frames, the second row shows the result from our method (taken from matrix O after the decomposition), the third, fourth, and fifth rows show the background subtraction result obtained using [22], [19], and [23], respectively. Please zoom in to see the details.

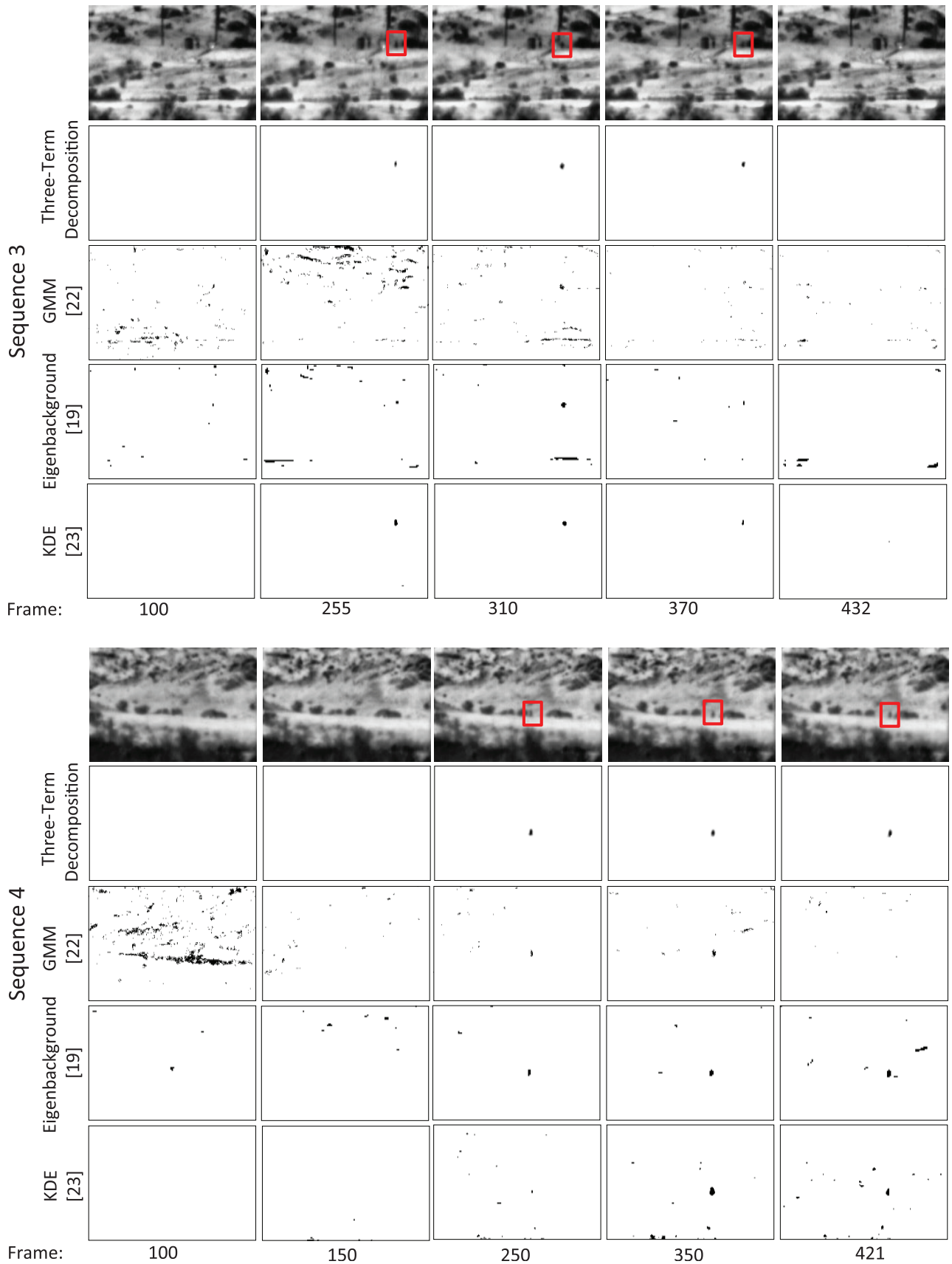


Fig. 5. Our moving object detection results on sample frames from sequences 3 and 4 compared to [22], [19], and [23]. For every sequence, the first row shows the frames, the second row shows the result from our method (taken from matrix O after the decomposition), the third, fourth, and fifth rows show the background subtraction result obtained using [22], [19], and [23], respectively. Please zoom in to see the details.

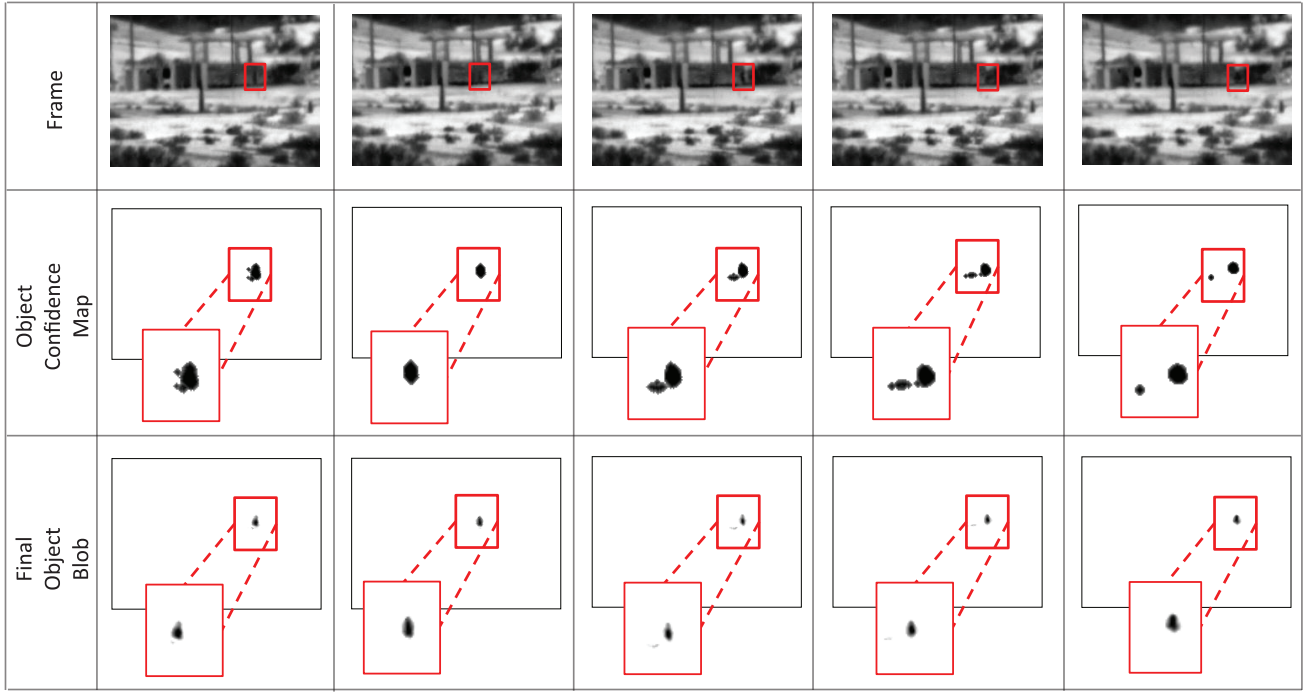


Fig. 6. Example frames illustrating the contribution of the turbulence model and the low-rank optimization in the total moving object detection performance, separately. The first row shows the frames, the second row shows the object confidence map obtained from the turbulence model (confidence values are mapped to $[0 - 255]$, the highest confidence is black), and the third row shows the final object blob after the three-term low-rank decomposition. Clearly, the turbulence model provides a rough estimation of the object location, while the low-rank optimization refines the result to obtain an accurate detection.

A formulation similar to (7) can be obtained by replacing E with $F - A - O$ in the objective function in (4) in order to solve $\min_{A,O} \|A\|_* + \tau \|O\|_1 + \lambda \|F - A - O\|_F^2$. This was indeed done by Lin et al. in [33], in which they proposed two methods to solve this optimization: the Augmented Lagrange Multiplier and the Inexact Augmented Lagrange Multiplier (IALM). The IALM is an alternating direction approach which converges almost as fast as the exact ALM, but with significantly fewer partial SVDs. This was further explored in [41], where an additional three-term alternation method was discussed. As noted in [41], the convergence of the scheme with more than two terms alternation is still an open problem in general, although numerical experiments strongly suggest that the alternation scheme is globally optimal under relatively mild conditions. However, for special three-term alternations with very strict and potentially nonpractical assumptions, there do exist proofs, as given in [45] and [44].

The convergence of scheme (7) to the optimal solution of (4) clearly requires further study; however, we are able to prove the feasibility of the accumulation points produced by the iterations of the algorithm.

Theorem 1. *The sequences $\{A_k\}$, $\{O_k\}$, and $\{E_k\}$ generated by Algorithm 1 are bounded, and any accumulation point (A^*, O^*, E^*) of (A_k, O_k, E_k) is a feasible solution: $F = A^* + O^* + E^*$.*

Refer to the supplementary appendix, which is available online, for an outline of the proof of Theorem 1. Our theorem indicates that the iterations of our algorithm are guaranteed to yield a decomposition of low-rank, sparse, and turbulence components, which is sufficient in our problem. Proving that

the solution that we arrive at is globally optimal is, however, still an open problem which we leave for future work. However, our experimental evaluations agree with [41] and suggest the optimality of the alternation scheme.

5 EXPERIMENTS

We experimented extensively on the proposed ideas using four infrared sequences significantly distorted by atmospheric turbulence and also containing a moving human. The sequences and the code are available on our website: <http://www.cs.ucf.edu/~oreifej>. Each frame is 250×180 with 450 frames per sequence. The moving object occupies around 40 pixels in the frame and moves arbitrarily in the FOV. Typically, the object is static at the beginning of the sequence; therefore, we use the first 50 frames to compute the parameters of the intensity and the motion Gaussian models at every pixel. Our three-term decomposition converges quickly in about 25-35 iterations, which takes approximately three minutes on a conventional laptop. Fig. 3 shows our

TABLE 1
Comparison of the Average PSNR in dB
for the Original Sequences,
After Applying Registration, and Three-Term Decomposition

	Original	Registration [7]	3-Term Decomp.
Sequence 1	26.55	32.20	31.62
Sequence 2	27.49	34.05	34.11
Sequence 3	27.71	31.79	32.25
Sequence 4	27.82	32.72	33.13

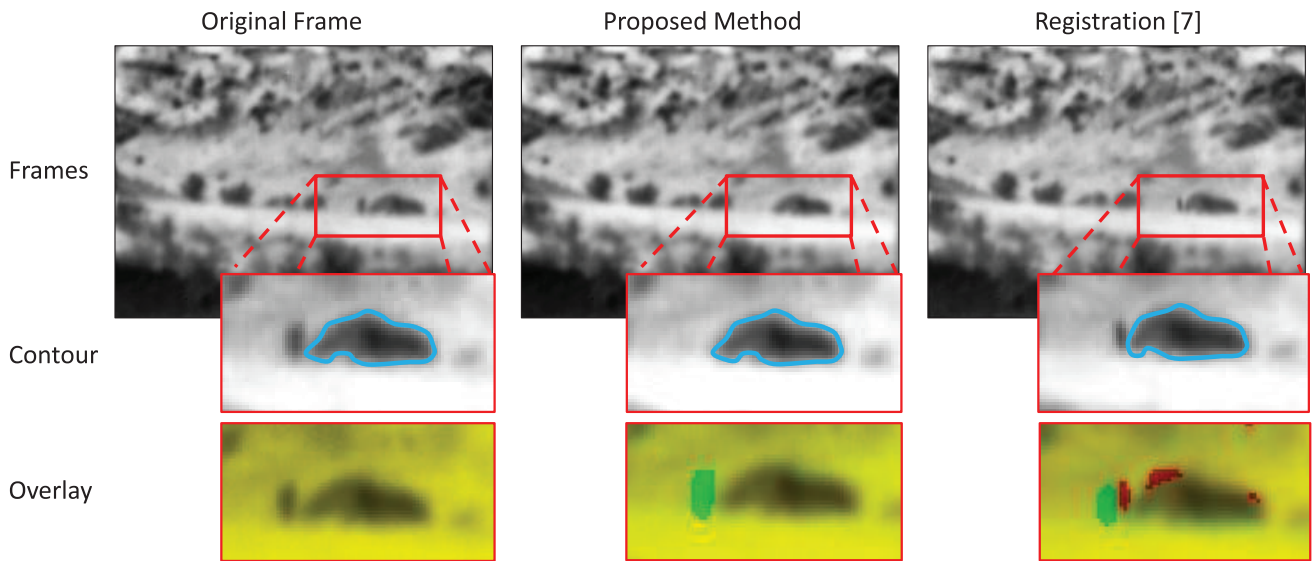


Fig. 7. Our turbulence mitigation results compared to nonrigid registration [7] for an example frame. The first row shows, from left to right: the original frame, the recovered background using our proposed method, and the recovered background using [7]. The second row shows a zoomed-in version of the frames with overlaid contours of a vehicle near the moving object. Note how the object's motion near the vehicle caused a deformation in the contour of the vehicle. The third row demonstrates another visualization of the artifacts experienced by registration due to the moving object. We inserted each of the frames into the R color channel with the original frame in the G channel, and set the B channel to zero. It is obvious that, using our proposed method, the object is eliminated from the background (appears as a green blob) and the vehicle's shape is maintained, while using registration methods such as [7] and [6] do not handle moving objects and accordingly result in deformations in the region surrounding the object (appears as several green and red blobs). The figure is best observed in color. Please zoom in to see the details and refer to our website for the complete videos.

decomposition results. Our algorithm is able to decompose all the turbulence sequences generating a clear background and detecting the moving objects. The results are better observed in the videos available on our website.

Since this problem is novel, there are no directly comparable approaches. Therefore, we consider comparing each of the moving object detection and turbulence mitigation tasks separately. We compare our moving object detection results with the background subtraction method described in [22], where the background is modeled using a mixture of Gaussians. Moreover, we compare our result with an eigen-background model similar to [19], where the background basis for every 3×3 patch is found using PCA, then the patch is marked as foreground if it is not well represented by the PCA basis (i.e., its reconstruction error is above a threshold). We also compare our results with [23], which employs a nonparametric Kernel Density Estimation method (KDE) over the joint domain (location) and range (intensity) representation of image pixels. Figs. 4 and 5 illustrate the results of the comparison on sample frames from our sequences. It is clear that even state-of-the-art methods suffer in turbulence sequences; therefore, our method outperforms such methods significantly. In Fig. 6, we demonstrate the contribution of each of the turbulence model and the sparse optimization in the detection performance. As can be observed from the figure, our method leverages multiple constraints (sparsity, motion model, and intensity model) which complement each other to finally determine the object regions, thus significantly reducing miss-detection rates.

To evaluate the results quantitatively, we used a region-based measure where we applied connected components to the binary mask resulting from the background subtraction in order to obtain contiguous detection regions. Consequently,

a detection region is considered correct if at least 50 percent of it is overlapping with the groundtruth; otherwise it is considered a miss-detection. The ROC curve in Fig. 8 summarizes the obtained results for all sequences.

In addition, we compare our turbulence mitigation results with the robust registration algorithm presented in [7]. The robust registration is an iterative process which recovers the original dewarped sequence by registering the frames to their mean, and then updating the frames and the mean at every iteration. The registration is performed using

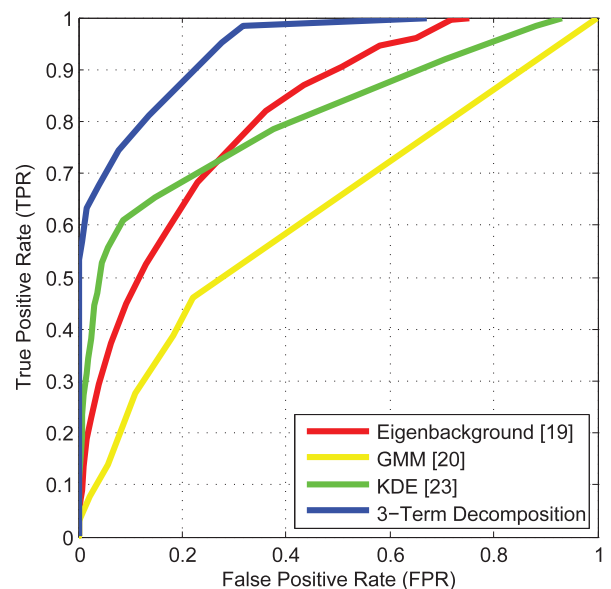


Fig. 8. Performance of our method compared to background subtraction methods.

nonrigid alignment by a means of control points overlaid on the frames. To evaluate the performance, we measured the Peak Signal-to-Noise Ratio (PSNR) between the first frame of the sequence and the rest of the frames, and reported the average results for all the frames in Table 1. It is clear that both our method and registration can significantly stabilize the sequences and improve the PSNR. However, the moving object is not explicitly handled in the registration; therefore, it impedes the process by causing the control points to incorrectly shift in the direction of the object's motion, resulting in several artifacts in the surrounding area, as demonstrated in Fig. 7. In contrast, our three-term decomposition handles such difficulties by separating the moving object; therefore, we recover a clear background without artifacts and with significantly reduced turbulence.

6 CONCLUSION AND FUTURE WORK

We presented a novel method for concurrent turbulence mitigation and moving object detection. Our method leverages the low-rank, the Gaussian, and the sparse properties of the sequence, decomposing it into the background, the turbulence, and the moving objects, respectively.

In the future, we will investigate further applications of our three-term decomposition. In particular, we have recently shown in [39] how RPCA can be employed in action recognition to decompose the trajectories of an action in a moving camera scenario into low-rank components corresponding to camera motion and rigid object motion, and a sparse component corresponding to the articulated object motion. In an extension to that, we will investigate the use of the three-term decomposition to achieve finer decompositions of action sequences, and extract further components including the component corresponding to the Gaussian noise. We are also investigating the potential extension of the three-term-decomposition into a generic n -component decomposition, and its possible applications in clustering and segmentation problems.

ACKNOWLEDGMENTS

This work was supported by the Night Vision and Electronic Sensors Directorate (NVESD) under contract number W15P7T-08-D-P417. The authors thank Teresa Pace from NVESD for introducing them to the problem, providing the dataset, and conducting useful discussions. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NVESD.

REFERENCES

- [1] M. Roggemann and B. Welsh, *Imaging through Turbulence*. CRC Press, 1996.
- [2] D. Li, R.M. Mersereau, and S. Simske, "Atmospheric Turbulence-Degraded Image Restoration Using Principal Components Analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 3, pp. 340-344, July 2007.
- [3] J. Goodman, *Statistical Optics*. Wiley-Interscience, 2000.
- [4] G.C.M. Aubailly, M. Vorontsov, and M. Valley, "Automated Video Enhancement from a Stream of Atmospherically-Distorted Images: The Lucky-Region Fusion Approach," *Proc. SPIE*, vol. 7463, pp. 74630C-74630C, 2009.
- [5] X. Zhu and P. Milanfar, "Image Reconstruction from Videos Distorted by Atmospheric Turbulence," *Proc. SPIE*, vol. 7543, pp. 75430S-75430S, 2010.
- [6] 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*, 2008.
- [7] O. Oreifej, G. Shu, T. Pace, and M. Shah, "A Two-Stage Reconstruction Approach for Seeing through Water," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2011.
- [8] Y. Yitzhaky, I. Dror, and N. Kopeika, "Restoration of Atmospherically Blurred Images According to Weather Predicted Atmospheric Modulation Transfer Function (MTF)," *Optical Eng.*, vol. 36, pp. 3064-3072, 1997.
- [9] M. Chang, C. Font, C. Gilbreath, E. Oh, E. Distefano, S. Restaino, C. Wilcox, and F. Santiago, "Comparing Horizontal Path C2N Measurements over 0.6 km in the Tropical Littoral Environment and in the Desert," *Proc. SPIE*, vol. 6551, p. 65510I, 2007.
- [10] A. Efros, V. Isler, J. Shi, and M. Visontai, "Seeing through Water," *Proc. Neural Information Processing Systems*, 2004.
- [11] A. Bjorck, *Numerical Methods for Least Squares Problems*. SIAM, 1996.
- [12] S. Ali and M. Shah, "A Lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2007.
- [13] E.J. Candes, X. Li, Y. Ma, and J. Wright, "Robust Principal Component Analysis?" *Proc. IEEE Sensor Array and Multichannel Signal Workshop*, 2009.
- [14] H. Ji, C. Liu, and Y.X. Zuowei Shen, "Robust Video Denoising Using Low Rank Matrix Completion," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2010.
- [15] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "RASL: Robust Alignment by Sparse and Low-Rank Decomposition for Linearly Correlated Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2233-2246, Nov. 2012.
- [16] Y. Yu and D. Schuurmans, "Rank/Norm Regularization with Closed-Form Solutions: Application to Subspace Clustering," *Proc. Conf. Uncertainty in Artificial Intelligence*, 2011.
- [17] S.H. Chan, R. Khoshabeh, K.B. Gibson, P.E. Gill, and T.Q. Nguyen, "An Augmented Lagrangian Method for Total Variation Video Restoration," *IEEE Trans. Image Processing*, vol. 20, no. 11, pp. 3097-3111, Nov. 2010.
- [18] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting Moving Objects, Ghosts, and Shadows in Video Streams," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1337-1342, Oct. 2003.
- [19] N.M. Oliver, B. Rosario, and A.P. Pentland, "A Bayesian Computer Vision System for Modeling Human Interactions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 831-843, Aug. 2000.
- [20] C. Stauffer and W. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1999.
- [21] A. Elgammal, D. Harwood, and L. Davis, "Non-Parametric Model for Background Subtraction," *Proc. Sixth IEEE Int'l Conf. Computer Vision*, 1999.
- [22] Z. Zivkovic, "Improved Adaptive Gaussian Mixture Model for Background Subtraction," *Proc. 17th Int'l Conf. Pattern Recognition*, 2004.
- [23] Y. Sheikh and M. Shah, "Bayesian Modeling of Dynamic Scenes for Object Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 11, pp. 1778-1792, Nov. 2005.
- [24] X. Zhu and P. Milanfar, "Stabilizing and Deblurring Atmospheric Turbulence," *Proc. IEEE Int'l Conf. Computational Photography*, 2011.
- [25] Y. Tian and S. Narasimhan, "A Globally Optimal Data-Driven Approach for Image Distortion Estimation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2010.
- [26] Y. Tian and S. Narsimhan, "Seeing through Water: Image Restoration Using Model-Based Tracking," *Proc. 12th IEEE Int'l Conf. Computer Vision*, 2009.
- [27] A. Buades, B. Coll, and J. Morel, "A Non-Local Algorithm for Image Denoising," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2005.
- [28] C. Liu and W. Freeman, "A High-Quality Video Denoising Algorithm Based on Reliable Motion Estimation," *Proc. 11th European Conf. Computer Vision*, 2010.

- [29] Z. Wen, D. Fraser, A. Lambert, and H. Li, "Reconstruction of Underwater Image by Bispectrum," *Proc. IEEE Int'l Conf. Image Processing*, 2007.
- [30] J. Goodman, *Introduction to Fourier Optics*. Roberts and Company Publishers, 2004.
- [31] A. Donate, G. Dahme, and E. Ribeiro, "Classification of Textures Distorted by Water Waves," *Proc. 18th Int'l Conf. Pattern Recognition*, 2006.
- [32] N. Joshi and M. Cohen, "Lucky Imaging for Multi-Image Denoising, Sharpening, and Haze Removal," *Proc. IEEE Int'l Conf. Computational Photography*, 2010.
- [33] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices," technical report, Univ. of Illinois at Urbana-Champaign, 2009.
- [34] D.P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 2004.
- [35] J.-F. Cai, E.J. Candes, and Z. Shen, "A Singular Value Thresholding Algorithm for Matrix Completion," *SIAM J. Optimization*, vol. 20, no. 4, pp. 1956-1982, 2010.
- [36] H. Branover and Y. Unge, *Progress in Turbulence Research*. AIAA, 1994.
- [37] M. Isichenko and A. Gruzinov, "Isotopological Relaxation, Coherent Structures, and Gaussian Turbulence in Two-Dimensional (2-D) Magnetohydrodynamics (Mhd)," *Physics of Plasmas*, vol. 1, 1994.
- [38] S. Wu, B. Moore, and M. Shah, "Chaotic Invariants of Lagrangian Particle Trajectories for Anomaly Detection in Crowded Scenes," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2010.
- [39] S. Wu, O. Oreifej, and M. Shah, "Action Recognition in Videos Acquired by a Moving Camera Using Motion Decomposition Of Lagrangian Particle Trajectories," *Proc. 12th IEEE Int'l Conf. Computer Vision*, 2011.
- [40] A. Agarwal, S. Neghaban, and M. Wainwright, "Noisy Matrix Decomposition: Optimal Rates in High Dimensions," technical report, <http://arxiv.org/abs/1102.4807>, Feb. 2011.
- [41] M. Tao and X. Yuan, "Recovering Low-Rank and Sparse Components of Matrices from Incomplete and Noisy Observations," *SIAM J. Optimization*, vol. 21, pp. 57-81, 2011.
- [42] D. Gabay and B. Mercier, "A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximation," *Computers and Math. with Applications*, vol. 2, pp. 17-40, 1976.
- [43] R. Glowinski and A. Marroco, "Sur l'Approximation, par Elements Finis d'Ordre Un, et la Resolution, par Penalisation-Dualite, d'une Classe de Problemes de Dirichlet Non Lineaires," *Revue Francaise D'Automatique, Informatique et Recherche Operationnelle*, 1975.
- [44] D. Goldfarb and S. Ma, "Fast Multiple Splitting Algorithms for Convex Optimization," preprint, 2009. available online <http://www.arxiv.org/abs/0912.4570>.
- [45] B. He and X. Yuan, "Linearized Alternating Direction Method with Gaussian Back Substitutions for Separable Convex Programming," http://www.optimization-online.org/DB_HTML/2011/10/3192.html, 2011.
- [46] Y. Zhang, "An Alternating Direction Algorithm for Nonnegative Matrix Factorization," technical report, Rice Univ., 2010.



He is a student member of the IEEE.



Omar Oreifej received the BS degree in computer engineering from the University of Jordan, Amman, in 2006, and the MS degree in computer engineering from the University of Central Florida, Orlando, in 2009. Currently, he is working toward the PhD degree in the Computer Vision Laboratory, University of Central Florida. His research interests include video denoising, low rank optimization, sparse representation, object tracking, and object detection.

Xin Li received the BS and MS degrees from Zhejiang University in 1983 and 1986, respectively, and the PhD degree in mathematics from the University of South Florida in 1989. He is working as a professor of mathematics at the University of Central Florida. His research interests include approximation theory and its applications in mathematical image and video analysis. He is a member of the IEEE.



Mubarak Shah is an agere chair professor of computer science and the founding director of the Computer Visions Lab at the University of Central Florida (UCF). He is a coauthor of three books (*Motion-Based Recognition* (1997), *Video Registration* (2003), and *Automated Multi-Camera Surveillance: Algorithms and Practice* (2008)), all by Springer. He has published extensively on topics related to visual surveillance, tracking, human activity and action recognition, object detection and categorization, shape from shading, georegistration, visual crowd analysis, etc. In 2006, he was awarded the Pegasus Professor award, the highest award at UCF, given to a faculty member who has made a significant impact on the university. He is an ACM Distinguished Speaker. He was an IEEE Distinguished Visitor speaker for 1997-2000, and received the IEEE Outstanding Engineering Educator Award in 1997. He received the Harris Corporation's Engineering Achievement Award in 1999, the TOKTEN awards from UNDP in 1995, 1997, and 2000; the SANA award in 2007, an honorable mention for the ICCV 2005 Where Am I? Challenge Problem, and was nominated for the best paper award at the ACM Multimedia Conference in 2005 and 2010. At UCF, he received Scholarship of Teaching and Learning (SoTL) award in 2011; College of Engineering and Computer Science Advisory Board award for faculty excellence in 2011; Teaching Incentive Program awards in 1995 and 2003, Research Incentive Award in 2003 and 2009, Millionaires' Club awards in 2005, 2006, 2009, 2010 and 2011; University Distinguished Researcher award in 2007 and 2012. He is an editor of an international book series on video computing; editor in chief of *Machine Vision and Applications* journal, and an associate editor of the *ACM Computing Surveys* journal. He was an associate editor of the *IEEE Transactions on Pattern Analysis and Machine Intelligence*, and a guest editor of the special issue of *International Journal of Computer Vision* on video computing. He was the program cochair of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008. He is a fellow of the IEEE, IAPR, AAAS, and SPIE.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.