IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 9, 2021, accepted July 18, 2021, date of publication August 2, 2021, date of current version August 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3101823

Frequency-Aware SVD Decomposition and Its Application to Color Magnification and Motion Denoising

IBRAHIM KAJO[1], NIDAL KAMEL[2], (Senior Member, IEEE), YASSINE RUICHEK[1], (Senior Member, IEEE), AND ABDULRAHMAN AL-AHDAL[3]

[1]CIAD UMR7533, University of Bourgogne Franche Comté, UTBM, F-90010 Belfort, France
[2]Centre for Intelligent Signal and Imaging Research, Universiti Teknologi PETRONAS, Seri Iskandar, Perak 31750, Malaysia
[3]Electrical Engineering Department, Umm Al-Qura University, Mecca 24381, Saudi Arabia

Corresponding author: Nidal Kamel (nidalkamel2@hotmail.com)

This work was supported in part by the Regional Council of Bourgogne Franche-Comté, France, and in part by the Deanship of Scientific Research at Umm Al-Qura University, Saudi Arabia, under Grant 17-ENG-1-01-0007.

ABSTRACT Videos are full of dynamic changes along both the spatial and temporal dimensions. Large, jerky short-term motions make it difficult to extract significant changes from videos such as subtle color changes and long-term motions occurring in time-lapse sequences. In this paper, we introduce two singular value decomposition (SVD)-based video decomposition schemes to clearly reveal such changes. The first scheme involves enhancing the visual characteristics of small subtle color changes in the presence of a wide variety of motion patterns by magnifying their pixel intensities. The second scheme removes short-term motions that visually distract attention from the underlying content of video sequences such as time-lapse videos, snowing scene, and maritime surveillance. Both schemes involve the decomposition of videos into spatiotemporal slices in which each slice is further decomposed into several singular components. The low-rank components that primarily represent background and color intensity information are then temporally processed to magnify the magnitude of the signal at the subtle color change target frequency. At the same time, an approach similar to that used in denoising time-lapse sequences is applied to temporally filter the singular components representing sparse information, thereby removing jittery short-term motions while preserving long-term motions, which are represented by both low-rank and unfiltered sparse components. We demonstrate promising color magnification and motion denoising results that can be obtained much faster than results estimated using state-of-the-art techniques.

INDEX TERMS Singular value decomposition, Fourier transform, short-term motion, motion denoising, subtle color changes.

I. INTRODUCTION

Computational photography is a quickly growing research field in which dozens of computational imaging techniques have been proposed for the generation of images and videos that could not otherwise be generated using current imaging devices [1]–[3]. The key difference between images and videos is the presence of the time dimension in the latter; videos are primarily sequences of frames that can be played at different. frame-per-second rates. Furthermore, a wide variety of motion patterns and events can coexist in videos. Unfortunately, large-magnitude motions will usually dominate a scene [4], [5], making it difficult to clearly observe many significant details and changes and requiring further processing to reveal these unseen changes. One significant approach to better understanding the dynamics of an object's motion is to manipulate it by magnifying color changes and separating their short- and long-term dynamics. We hypothesize that manipulating the frequencies of motion events in a given video can change the way in which we observe these motion events and reveal/hide certain types of motion information. Our manipulation approach depends on the decomposition of video into several components according to frequency and motion information. This decomposition is performed using singular value decomposition (SVD), through which video content can be divided

The associate editor coordinating the review of this manuscript and approving it for publication was Cesar Vargas-Rosales.

into distinct motion and frequency components. Subsequently, different temporal manipulations such as filtering and magnifying schemes can be used to produce the desired results. Here we propose two manipulation schemes with many potential applications in the field of video processing and manipulation—color video magnification and motion denoising.

In many visual periodicity-based applications such as heart beat estimation [6]–[8], the color intensity of the extracted signal is very small and the related changes cannot be seen by the naked eye. It is therefore necessary to magnify the color magnitude of the periodic signal embedded in the processed video. The proposed color magnification scheme is designed to magnify the SVD components related to the extracted periodic signal to magnify the color changes occurring in the original video. This magnification procedure can significantly reduce the processing burden required under conventional techniques, in which all pixels must be spatiotemporally processed to achieve the desired result. Videos also contain numerous types of motion that can be categorized in terms of their temporal characteristic into short- and long-term and random motions. In certain types of videos, such as time-lapse sequences and dynamic background sequences, random and temporally inconsistent motion must be removed through a procedure known as motion denoising, which was defined and first used in [9]. Temporally filtering the SVD component embedded in undesired motions can significantly reduce their presence. The specific contributions of this paper are as follows:

1) We introduce a technique to magnify subtle color changes that is robust to a wide variety of distorting motions, including rapid, slow, large, and random motion.
2) We introduce an extremely fast and efficient motion denoising technique capable of suppressing short-term jittery motions in time-lapse sequences.

The remainder of the paper is organized as follows. Section II discusses the related work on both color magnification and time-lapse enhancement. Section III presents a brief description of the tensor structure of video representation. Sections IV and V describe the proposed methods for color magnification and motion denoising, respectively. In Section VI, we present and discuss our experimental results and, finally, Section V concludes the study and discusses future research topics.

II. RELATED WORK

Recently, there has been growing interest in extracting invisible changes occurring in video sequences using several computer vision applications [10]–[15]. These efforts have been hindered by the fact that the intensities of the pixels involved in the extracted signals have very small magnitudes, which make the related changes invisible to the naked eye. To solve this problem, numerous change magnification schemes for enhancing the motion and color change intensities of pixels involved in these signals have been proposed. In this section, we review only the proposed techniques related to the task of color magnification, leaving out motion magnification as beyond the scope of this work. The most commonly used approach is the Eulerian framework, which can be used to detect subtle color changes and magnify them accordingly [16]. This approach starts by decomposing the frames of a given video into multiple Gaussian pyramids. These are then temporally filtered to extract the signal of interest, which is later magnified to the desired scale. Finally, a magnified video is generated by reconstructing the magnified pyramids. Recently, Zhang *et al.* [17] proposed an acceleration-based Eulerian color/motion magnification method that can magnify color changes in the presence of large motion in cases in which conventional Eulerian-based methods fail. Their method magnifies the acceleration of color changes by relying on the fact that large motions show linearity with respect to the time scale while small motions do not. More recently, Takeda *et al.* [18], proposed a color magnification technique that makes the acceleration-based approach more robust to rapid, large motions. Using this technique, they developed a jerk-based smoothness filter that removes rapid and large-scale changes while passing the subtle changes that need to be magnified. They further enhanced their change magnification results by proposing a fractional anisotropic filter that passes only meaningful subtle color changes while eliminating the non-meaningful background color fluctuations caused by photographic noise [19]. These techniques, however, are robust against only a limited number of motion types, which can lead to unreliable performance in real-world scenarios. Our color magnification scheme differs from these techniques, in which color changes are separated from motions beforehand, regardless of their type, size, or speed.

As mentioned earlier, certain types of video sequence are accompanied by jerky small-scale and short-term changes that must be removed to achieve a better watching experience. Analysis of video components along the time dimension is the typical approach to segregating temporally consistent and non-consistent objects [20], [21]. Numerous approaches to enhancing the quality of video sequences have been proposed, including editing [22], synthesizing [23], [24], resampling [25], [26], and removing certain events and popping artifacts. Sunkavalli *et al.* [27] introduced an editing approach that separates video components into their reflectance, illumination, and geometry factors. Their technique extracts a spatiotemporal intensity-based time series (profile) for each pixel and then applies matrix factorization to the extracted profiles to decompose them into sunlight, shadow, and skylight areas, enabling different applications such as shadow removal, relighting, and advanced image editing. Under a different decomposition approach proposed by Rubinstein *et al.* [9] time-lapse sequence are separating into short- and long-term motion components by implementing a Bayesian-based computational framework that temporally smooths and reduces the randomness of video content. Despite their promising results, these approaches

are limited to very short sequences and are computationally expensive. Martin-Brualla *et al.* [28] introduced an approach for synthesizing time-lapse sequences of popular landmarks from large photo collections. The appearances of their output time-lapse sequences are stabilized by reducing the lighting effect and minimizing the flicker by applying a temporally median filter to the warped frames followed by temporal regularization. Concurrently, Tan *et al.* [29] introduced a decomposition framework that decomposes time-lapse painting videos into sequences of subsequent "stroke" images. Based on this technique, they proposed a processing pipeline that is robust to long-term occlusion, global color shifts, and different painting-related challenges. Their technique detects and removes pixels that are unstable with respect to the time dimension in terms of moving standard deviation. The removed pixels are replaced by corresponding pixels extracted from keyframes that contain no occlusion. Recently, Tian *et al.* [30] proposed a low rank/sparse decomposition framework that removes the snowflakes from both static background images and foreground objects. The background desnowing is achieved by decomposing the input video into low rank component that represents the snowflake free background, and a sparse component that represents both moving objects and snowflakes. On the other hand, the foreground objects are separated from snowflakes by performing a block wise alignment approach where the blocks that belong to a certain foreground object are arranged in a new tensor for a second round of decomposition. However, such alignment procedure cannot properly handle several scenarios with challenges such as large occlusion, scale variation, and non-rigid motions.

The target of our motion denoising approach is similar to Rubinstein's framework [9] in that it decomposes both time-lapse sequences and normal videos into short- term and long-term motion-based components. Our motion denoising scheme differs from Rubinstein's framework, however, in that it is much faster and can be extended to the processing of more challenged sequences such as long time-lapse painting, snowflakes removal, and maritime surveillance videos.

III. TENSOR STRUCTURE OF VIDEOS

A video can be considered a sequence of frames (images), that is represented spatiotemporally in a three-dimensional coordinate system (i, j, t), also known as a tensor. In general, a video can be represented as a three-dimensional real array, $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{m\times n\times r}$, in which the indices m and n are the spatial dimensions of the video frame and r is the frame number. By using the symbols i, j, and t as variables over the three dimensions of the video tensor, m, n, and r, respectively, a video frame can be represented as $\boldsymbol{\mathcal{A}}_{::t} \in R^{m\times n}$ while the YT and XT spatiotemporal slices can be represented as matrices $\boldsymbol{\mathcal{A}}_{i::} \in R^{r\times n}$ and $\boldsymbol{\mathcal{A}}_{:j:} \in R^{r\times m}$, respectively. The columns in each slice represent the time series of the corresponding pixels contained in the slice. Thus, based on the events captured in the video, the columns of various slices can be identified as discrete functions. Symbols and other notation used in this paper are defined upon their first use and are summarized in Table 1 for reference.

TABLE 1. Mathematical notations and symbols.

Symbol	*Description*
$\boldsymbol{\mathcal{A}}$	Tensor
$\boldsymbol{\mathcal{A}}_{i::}$	Spatiotemporal slice-matrix
$\mathbf{A}$	Classical matrix
$\boldsymbol{a}$	vector
$\boldsymbol{a}[\mathrm{t}]$	Vector as discrete time signal function
a	Matrix entry
$\tilde{\boldsymbol{a}}$	Filtered vector constructed by the target frequency
$\breve{\boldsymbol{a}}$	Filtered vector constructed by the remaining frequency
$\acute{\boldsymbol{a}}$	Vector with magnified magnitude engaged at the target frequency
$\ddot{\boldsymbol{a}}$	Vector with manipulated freqeuncy
q	Number of singular components used
α	Magnitude maginfication factor
$\boldsymbol{\tau}$	Iteration index
Δ	Reduction in width value
ε	Increment in window width at each iteration
$\acute{\boldsymbol{a}}$	Motion denoised vector
γ	Number of synthesized frames
δ	Number of excluded frames
$\boldsymbol{\rho}$	Start index of synthesization at each iteration

Tensor-based processing is considered as a better alternative to extract features from a sequence of multidimensional data in numerous applications such as background subtraction [31], and microseismic events detection [32]. To this end, two obvious ways of processing the observation tensor are reported in the literature: the vectorization procedure where each matrix is converted into a vector and the unfolding procedure, which is also known as matricization. Nonetheless, neither of these processing procedures is capable of taking full advantage of the temporal dimension of the tensor structure. Therefore, we propose to employ our processing scheme which was recently introduced for spatiotemporally decomposing videos via SVD [36].

A. SINGULAR VALUE DECOMPOSITION OF A MATRIX

If A is a real m-by-n matrix, then there exist orthogonal matrices $\mathbf{U} = [\boldsymbol{u}_1|\ldots|\boldsymbol{u}_m] \in R^{m\times m}$ and $\mathbf{V} = [\boldsymbol{v}_1|\ldots|\boldsymbol{v}_n] \in R^{n\times n}$ such that [33]

$$\mathbf{U}^{\mathrm{T}}\mathbf{A}\mathbf{V} = \mathbf{S} = diag\left(s_1, \ldots, s_p\right) \in R^{m\times n}, \quad p = min\left\{m, n\right\}, \tag{1}$$

where $s_1 \geq s_2 \geq \cdots \geq s_n \geq 0$, matrices $\mathbf{U}$ and $\mathbf{V}$ have orthonormal columns known as the left and right singular vectors of $\mathbf{A}$, respectively, and $\mathbf{S}$ is diagonal with real nonnegative diagonal entries. Based on this, the SVD of a spatiotemporal slice $\boldsymbol{\mathcal{A}}_{i::}$ can be expressed as follows:

$$\boldsymbol{\mathcal{A}}_{i::} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}} \tag{2}$$

IV. PROPOSED COLOR MAGNIFICATION

Use Our goal is to magnify the color subtle changes associated with a specific frequency to make them visible to the naked eye. The proposed color magnification technique decomposes a video sequence into m spatiotemporal slices $\boldsymbol{\mathcal{A}}_{i::}$. Subsequently, SVD is applied to the extracted slices using a set of matrices $\mathbf{U}$, each of which should contain the same temporal information as the corresponding original slice $\boldsymbol{\mathcal{A}}_{i::}$. Given this requirement, the structures of the left singular vectors of a matrix $\mathbf{U}$ should be further investigated. The main challenge facing the majority of proposed color magnification techniques is the issue of color/motion separation; in other words, the color intensities of the components of interest should be magnified independently of any motion that might be present in the scene. Solutions proposed to address this challenge include different signal filtering schemes that are temporally applied on each pixel profile. Each proposed filter is also designed to eliminate certain types of motion such as large, rapid, or slow motion. By contrast, our proposed magnification scheme decomposes video content into low-rank (color/background) and sparse (foreground/changes) components. This separation allows the color changes to be magnified separately from the motion present in the processed video.

A. LEFT SINGULAR VECTOR REPRESENTATION

As mentioned above, the matrix U represents the temporal changes embedded in a video sequence. Accordingly, this matrix can be subjected to structural analysis to determine what each left singular vector represent with respect to the original pixel profiles. In [34]–[36], comprehensive investigations of SVD component structures were carried out for different motion pattern cases, including periodic, arbitrary, and intermittent motion. From a signal processing perspective, the projection of a spatiotemporal slice onto the first left singular vector $\boldsymbol{u}_1$ subspace represent the low-rank (intensity) component of this slice, while the projections of the same slice onto the remaining left singular vectors' $\boldsymbol{u}_2, \boldsymbol{u}_3 \ldots \boldsymbol{u}_m$ subspaces represent the changes in the low-rank color component along the time dimension. Based on this conclusion, the magnification of the first left singular vector enables magnification of the color intensity separately from the motion.

B. TEMPORAL ANALYSIS OF LEFT VECTOR

To magnify the target frequency of a subtle color change event, the first left singular vector $\boldsymbol{u}_1$ must be temporally processed. A discrete Fourier transform (DFT) is used to decompose $\boldsymbol{u}_1$ into different frequency components through a process in which the target frequency is separated from the other frequencies and its amplitude is magnified by a factor of α. The vector $\boldsymbol{u}_1$ can be represented as a discrete time signal function within the interval $[-1, 1]$ as follows

$$\boldsymbol{u}_1 = \{\boldsymbol{u}_1[t]\} \quad 0 < t < k \tag{3}$$

The t^{th} number in this sequence is denoted by $\boldsymbol{u}_1[t]$, where t refers to the frame index. For simplicity, $\boldsymbol{u}_1[t]$ is used to refer to the overall vector $\boldsymbol{u}_1$. Subsequently, The DFT of the signal $\boldsymbol{u}_1[t]$ is given as

$$\boldsymbol{U}_1(k) = \sum_{t=0}^{N-1} \boldsymbol{u}_1[t]\, e^{j2\pi tk/N} \tag{4}$$

where $k = 0, 1, \ldots, N-1$. Symbolically, this can be written as

$$\boldsymbol{u}_1[t] \overset{D\mathcal{F}}{\Leftrightarrow} \boldsymbol{U}_1(k) \tag{5}$$

where $\mathcal{F}$ and $D\mathcal{F}^{-1}$ denote, respectively, the operation of computing the DFT of $\boldsymbol{u}_1[t]$ and the inverse DFT. Furthermore, the N entries in $\boldsymbol{u}_1[t]$ and $w = (2\pi/N)k$, where $k = 0, 1, \ldots, N-1$, represent the set of distinguishable frequencies needed to synthesize $\boldsymbol{u}_1[t]$.

The target frequency for magnification/denoising is separated from the remaining component in the frequency domain and an inverse Fourier transform is employed to reproduce the following two discrete signals

$$\tilde{\boldsymbol{u}}_1[t] = D\mathcal{F}^{-1}\{\boldsymbol{U}_1\left(\tilde{k}\right)\} \tag{6}$$

$$\check{\boldsymbol{u}}_1[t] = D\mathcal{F}^{-1}\{\boldsymbol{U}_1\left(\check{k}\right)\} \tag{7}$$

where the Fourier series engaged with the target frequency are represented as $\boldsymbol{U}_1\left(\tilde{k}\right)$ and the remaining series are represented as $\boldsymbol{U}_1\left(\check{k}\right)$. As a result of this process, the target frequency components are magnified relative to the other frequency components as follows

$$\acute{\boldsymbol{u}}_1[t] = \alpha.\tilde{\boldsymbol{u}}_1[t] + \check{\boldsymbol{u}}_1[t] \tag{8}$$

To preserve the orthogonality of $\acute{\boldsymbol{u}}_1[t]$, the vector is normalized as follows

$$\boldsymbol{u}_1^{norm}[t] = \frac{\acute{\boldsymbol{u}}_1[t]}{\left\|\acute{\boldsymbol{u}}_1[t]\right\|} \tag{9}$$

To magnify the intensity of the target frequency components, vector $\boldsymbol{u}_1^{norm}[t]$ is rescaled by factor α_1 to produce a new vector $\boldsymbol{u}_1^{final}[t]$ with lower minimum and higher maximum values. Subsequently, the resulting video with magnified color changes can be reconstructed based on the SVD construction theorem by rewriting the original slices $\boldsymbol{\mathcal{A}}_{i::}$ as the sum of rank-1 matrices as follows

$$\boldsymbol{\mathcal{A}}_{i::}^{magnified} = \boldsymbol{u}_1^{final} s_1 \boldsymbol{v}_1^{\mathrm{T}} + \cdots + \boldsymbol{u}_d s \boldsymbol{v}_d^{\mathrm{T}} \tag{10}$$

where $d < p$ is the number of nonzero singular values.

Figure 1 demonstrates an example of the magnification of a bulb light in which the structures of different frequency-based components of the left singular vector $\boldsymbol{u}_1$ are clearly shown.

V. PROPOSED MOTION DENOISING TECHNIQUE

The goal of motion denoising is to eliminate short-term jittery motions from a given video while preserving long-term scene changes. Conventional temporal filtering techniques such as median- and mean-based approaches process each pixel independently, with pixels from different moving objects processed equally. This procedure leads to poor performance in

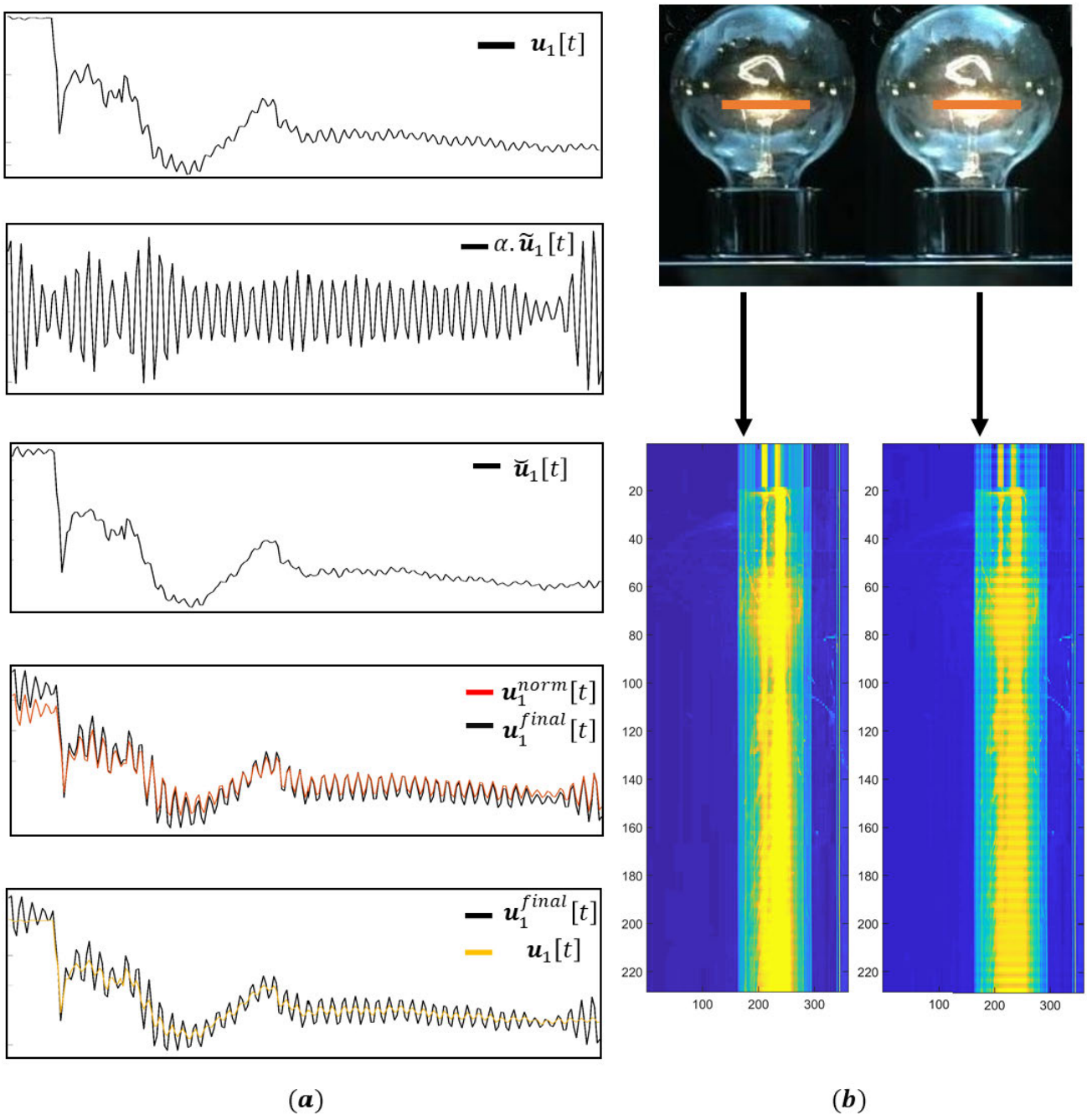

FIGURE 1. (a) from top to bottom: original first singular vector obtained by decomposing the spatiotemporal slice extracted from "bulb" video, filtered target frequency component, remaining frequency components, the final magnified first singular vector compared to both the normalized and the original vectors, respectively. (b). two spatiotemporal slices that represent the temporal changes of the bulb color changes extracted from the original video (left) and the magnified video (right).

cases of extreme dynamic scenarios in which objects move rapidly, producing videos that suffer from blur and discontinuity artifacts. To address this issue, the proposed method processes video content at the spatiotemporal slice level, at which the spatial information of pixel profiles can be taken into account. Furthermore, the proposed technique considers only the moving pixels in the denoising procedure, which in turn ensures that none of the background pixels are negatively affected.

From the discussion in subsection IV.A, the first rank-1 matrix $\mathcal{B}_{i::} = \boldsymbol{u}_1 s_1 \boldsymbol{v}_1^T$ represents the stationary-pixel content over the course of a video (i.e., the background) while the sum $\mathcal{F}_{i::} = \sum_{l=2}^{q} \boldsymbol{u}_l s_l \boldsymbol{v}_l^T$ represents the sparse information denoting the foreground objects. The residual, i.e., the sum $\mathcal{N}_{i::} = \sum_{l=q+1}^{d} \boldsymbol{u}_l s_l \boldsymbol{v}_l^T$, represents *additive white Gaussian noise* such as illumination changes and thermal and quantization noise. Thus, proper selection of low-rank and sparse parts of the spatiotemporal slices $\mathcal{A}_{i::}$, enable its decomposition

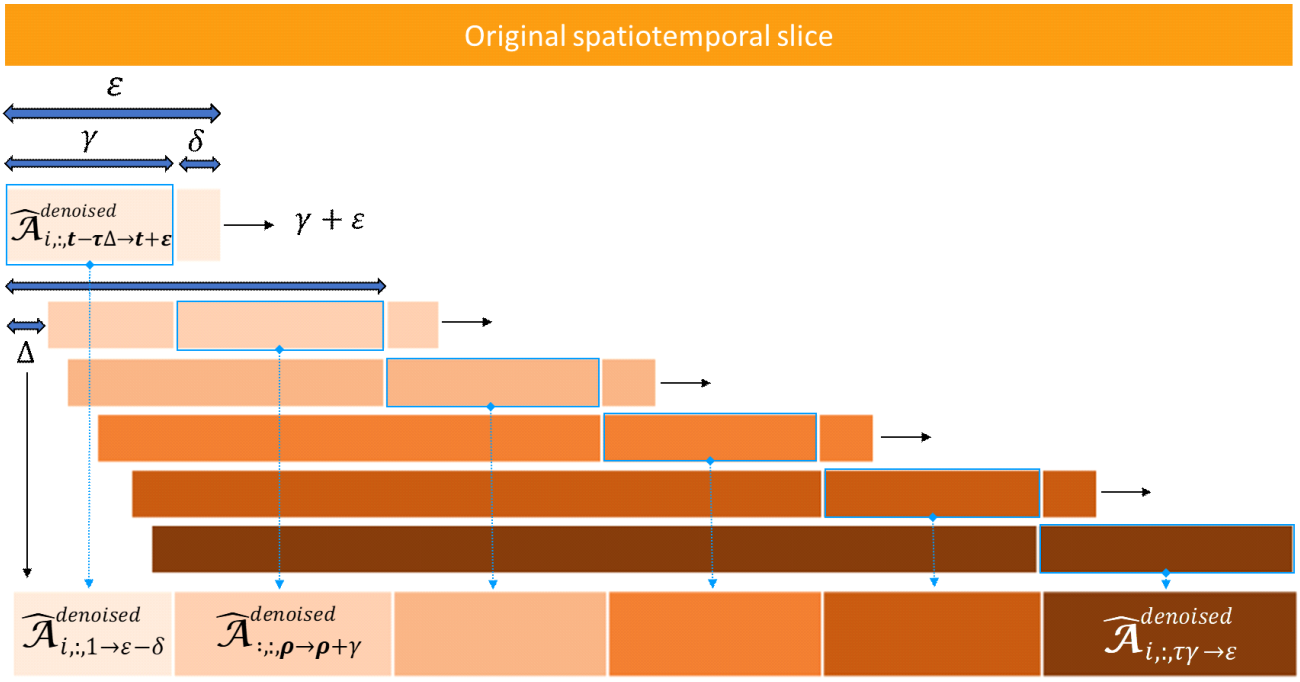

FIGURE 2. Block diagram of proposed motion denoising scheme that shows in detail the sequential steps of incrementally decomposing a spatiotemporal slice and reconstruction of a new slice denoised from jittery short-term motion.

into noise-free background $\mathcal{B}$ and foreground $\mathcal{F}$ tensors as follows

$$\mathcal{A}_{i::} \approx \mathcal{B}_{i::} + \mathcal{F}_{i::} \tag{11}$$

SVD is performed using an incremental window approach in which the temporal size of the window is incrementally increased by $\boldsymbol{\varepsilon}$ at each iteration. To reduce the computational complexity and control the contribution of the history information to the resynthesized video, the first Δ columns of the window are removed at each iteration and the extracted SVD components are used to reconstruct both the foreground and background components

$$\mathcal{A}_{\mathbf{i},:,\boldsymbol{\tau}\Delta\rightarrow(\boldsymbol{\tau}+\mathbf{1})\boldsymbol{\varepsilon}} = \mathcal{B}_{:,:,\boldsymbol{\tau}\Delta\rightarrow(\boldsymbol{\tau}+\mathbf{1})\boldsymbol{\varepsilon}} + \mathcal{F}_{:,:,\boldsymbol{\tau}\Delta\rightarrow(\boldsymbol{\tau}+1)\boldsymbol{\varepsilon}} \tag{12}$$

for $\boldsymbol{\tau} = 0, 1, \ldots, \frac{k}{\boldsymbol{\varepsilon}} - 1$.

A straightforward temporal filtering is then applied to the foreground slice to remove the short-term motions. This is achieved by temporally filtering the left singular vectors that represent the motion information as follows

$$\grave{\boldsymbol{u}}_l = f\left(\boldsymbol{u}_l\right) \quad for\ l = 2, 3, \ldots, q \tag{13}$$

where f is the filtering operator, which here is taken as an ideal temporal bandpass filter with sharp low- and high-cutoff frequencies denoted by f_l and f_h, respectively. Consequently, the foreground components can be reconstructed using the filtered left singular vectors as follows

$$\mathcal{F}^{\boldsymbol{denoized}}_{:,:,\boldsymbol{\tau}.\Delta\rightarrow(\boldsymbol{\tau}+\mathbf{1}).\boldsymbol{\varepsilon}} = \sum\nolimits_{l=2}^{q} \grave{\boldsymbol{u}}_l \mathrm{s}_l \boldsymbol{v}_l^{\mathrm{T}}, \tag{14}$$

and the motion-denoised matrix-slice can be expressed as

$$\mathcal{A}^{\boldsymbol{denoized}}_{:,:,\boldsymbol{\tau}\Delta\rightarrow(\boldsymbol{\tau}+\mathbf{1})\boldsymbol{\varepsilon}} = \mathcal{B}_{:,:,\boldsymbol{\tau}\Delta\rightarrow(\boldsymbol{\tau}+\mathbf{1})\boldsymbol{\varepsilon}} + \mathcal{F}^{\boldsymbol{denoized}}_{:,:,\boldsymbol{\tau}\boldsymbol{\Delta}\rightarrow(\boldsymbol{\tau}+\mathbf{1})\boldsymbol{\varepsilon}} \tag{15}$$

For each iteration, the denoised slices $\mathcal{A}^{denoised}_{:,:,\boldsymbol{\tau}_1\Delta\rightarrow\boldsymbol{\tau}_2\boldsymbol{\varepsilon}}$ are used to resynthesize video frames corresponding to their respective temporal window sizes excluding the frames that have been processed in previous iterations. To produce a smoother resynthesized video, the final δ frames of $\mathcal{A}^{\boldsymbol{denoised}}_{:,:,\boldsymbol{\tau}_1\boldsymbol{\Delta}\rightarrow\boldsymbol{\tau}_2\boldsymbol{\varepsilon}}$ are also excluded from the synthesization process and replaced by their corresponding frames in the next window. Thus, we obtain the resynthesized tensor

$$\mathcal{A}^{\text{resynthesized}} = \sum\nolimits_{\boldsymbol{\tau}=0}^{\frac{k}{\varepsilon}-1} \mathcal{A}^{\boldsymbol{denoised}}_{:,:,\boldsymbol{\rho}\rightarrow\boldsymbol{\rho}+\gamma} \tag{16}$$

where $\gamma = \varepsilon - \delta$ and $\boldsymbol{\rho} = (\tau + 1)\,\gamma - \tau\Delta$. Figure 2 shows the processing steps used by the proposed motion denoising scheme.

VI. EXPERIMENT

This section consists of two main subsections that concern the color magnification experiment and the motion denoising experiment, respectively. In the case of color video magnification, it is worth mentioning that the frequency of interest of each video is predefined based on the selected application (as implemented in all related color magnification frameworks). In the case of motion denoising framework, the implementation parameters are divided into two groups: 1) SVD based parameters and 2) motion denoising based parameters. In order to determine the number of the significant SVD components, q, the concept of the retained energy

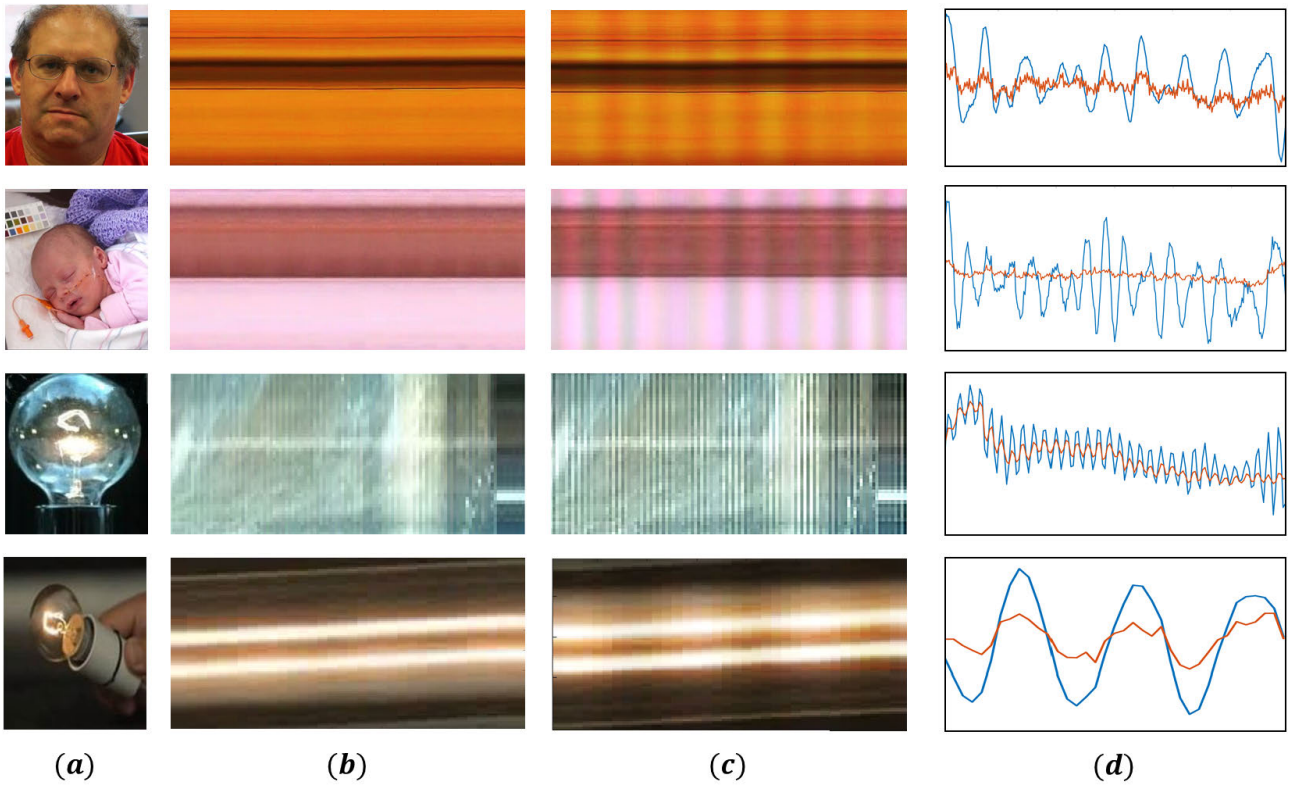

FIGURE 3. (a) examples of original videos. (b) spatiotemporal slices extracted from original videos. (c) spatiotemporal slices extracted from magnified videos. (d) the magnified first singular vector (blue) extracted from the magnified video compared to the first singular vector extracted from the original video (red).

in the principal subspace of the spatiotemporal slice $\mathcal{A}_{i::}$ is employed [37]. The total energy $E = \sum_{l=1}^{d} s_l^2$ of the $\mathcal{A}_{i::}$ is defined as the sum of the squares of its singular values. The retained energy in the q-rank approximation of $\mathcal{A}_{i::}$ is given as the sum of the q largest singular values squared. Accordingly, the energy ratio between the q-rank approximation and the total energy is given as

$$Er = \frac{1}{E} \sum_{l=1}^{q} s_l^2 \times 100 \tag{17}$$

This ratio is utilised in determining the value of q as a trade-off between the minimum number of the SVD components that are used in q-rank approximation and the maximum retained energy. The parameters of the second group, namely the parameters that are related to the size of the temporal window (such as Δ, ε, and δ) are proposed to reduce the computational complexity and downgrade the contribution of history information to the final resynthesized video while the rest of the parameters are related to the cutoff high frequency which differs from one application to another (dynamic sea surface, snow falling). Such parameters are empirically estimated.

A. COLOR MAGNIFICATION EXPERIMENT

To evaluate the performance of the proposed color magnification technique, it was applied to a number of real videos reflecting different motion scenarios. In the first assessment, involving videos containing no motion, the goal was to magnify facial color changes arising from the blood flow corresponding to the heart beat pattern. The first and second videos presented the faces of a stationary man and a new-born baby, respectively. The results shown in Fig. 3 (1st and 2nd row) indicate that the proposed technique could appropriately magnify the color changes in both videos without producing negative artifacts. The second experiment involved the processing of a video in which color changes were accompanied by slow large-scale motion. The video presented a moving hand holding a bulb and featured barely visible changes in color intensity. As shown in Fig. 3 (4th row), the proposed technique was able to successfully magnify the frequency component related to the target frequency in vector $\boldsymbol{u}_1$, which represents the intensity changes independent of the motion changes represented by the remaining singular vectors.

The third experiment involved a video in which the color change-accompanying motion was faster than the motion in the preceding video. Most of the color magnification techniques failed to handle this scenario because of the presence of large-scale motion, which is generally excluded from the magnification process under dedicated jerk-aware filtering. The video showed the shattering of three bulbs by bullets as shown in Fig. 1 and Fig. 3 (3rd row). Although both the proposed and jerk-aware techniques had no difficulties in magnifying the light changes produced by the bulbs

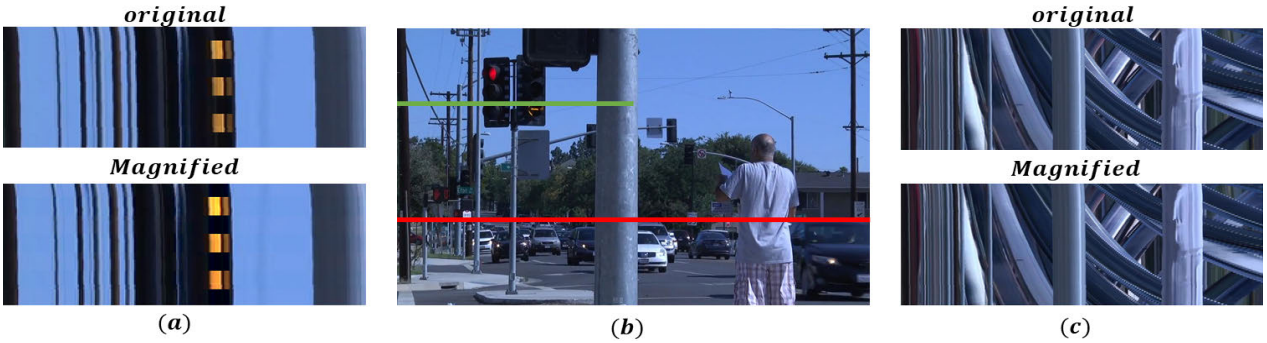

FIGURE 4. (a) the extracted spatiotemporal slices (green line) extracted from both the original and magnified videos. (b) a frame example of the processed video. (c) the extracted spatiotemporal slices (red line) extracted from both the original and magnified videos showing the untouched motion events.

without introducing clipping or messy artifacts, jerk-based techniques are designed to eliminate the impacts of certain types of motion (rapid and large-scale). Furthermore, they are fundamentally based on the assumption that the subtle color changes and motion that need to be eliminated are spatiotemporally independent. Such constraints limit their magnification ability, making the processing of uncontrolled scenarios such as outdoor scenes challenging. Therefore, a fourth assessment involving the magnification of color changes in the presence of several types of random motion was carried out. Fig. 4 shows an example of a video containing a blinking traffic light along with traffic flow, trees, and a pedestrian. As seen in Fig 4. (a), the proposed color magnification technique succeeded in magnifying the color changes in the traffic light while leaving the remaining related to car and pedestrian motion untouched as in the original.

TABLE 2. Significant parameter values for each sequence.

sequence	ε (frames)	Δ (frames)	f_h (hz)	*fps*
Pool	240	0	3	30
Construction1	20	5	2	24
Construction2	25	5	3	10
Construction3	20	10	3	10
Plant1	150	0	2	30
Plant2	50	0	4	30
Road	100	0	1	30
Airport	300	0	2	30
painting	20	10	0.5	30

B. MOTION DENOISING EXPERIMENT

Figures The proposed motion denoising approach is explored in three different applications: time-lapse enhancement, snowflakes removal, and foreground boat detection.

The performance of the proposed approach in addition to the state-of-the-art techniques are evaluated via visual comparisons and quantitative comparisons when possible. For better visualization, video examples are provided in supplementary files.

1) VISUAL RESULTS OF TIME-LAPSE VIDEOS

To evaluate the proposed motion denoising technique, the videos used in Rubinstein *et al.* [9], along with some additional videos, were analyzed. The three new videos—*construction1*, *construction2*, and *construction3*—demonstrate different construction processes in which several construction tools and vehicles are shown under varying lighting conditions. In the *airport* video, the short-term motion patterns are represented by the daily traffic motion of airplanes and vehicles while clouds produce long-term motion patterns. The *painting* video presents a time-lapse sequence in which painting is carried out over time and the scene is corrupted by the jerky motions of the painter's hand and head. It is worth noting that the new input sequences were downloaded from YouTube using the key word "time-lapse." Owing to the unique nature of each time-lapse sequence, the parameters of the proposed technique were set to slightly different values to produce optimal visual results for each sequence. Table 2 lists the selected values of significant parameters for each time-lapse sequence. The number of excluded frames was fixed at $\delta = 5$ and the low-cutoff frequency was fixed at $f_l = 0$.

In some sequences such as *plant1*, *pool*, and *airport* in which the number of frames was low, the value of the increment in the moving window was set equal to the total number of frames to prevent the creation of moving windows. The estimated number q of remaining left singular vectors used was set equal to the length of the temporal moving window and the performance of the proposed technique in eliminating short-term and jerky motion was evaluated by estimating the optical flow of frame pair sequences $(t, t+1)$ in each video. The optical flow results included both horizontal and vertical motion fields that indicated the magnitude and direction of the estimated motion.

A good motion denoising technique produces very small magnitudes of motion in both the horizontal and vertical directions. An optical flow technique called "classic+NL," [34] which is efficient and robust against

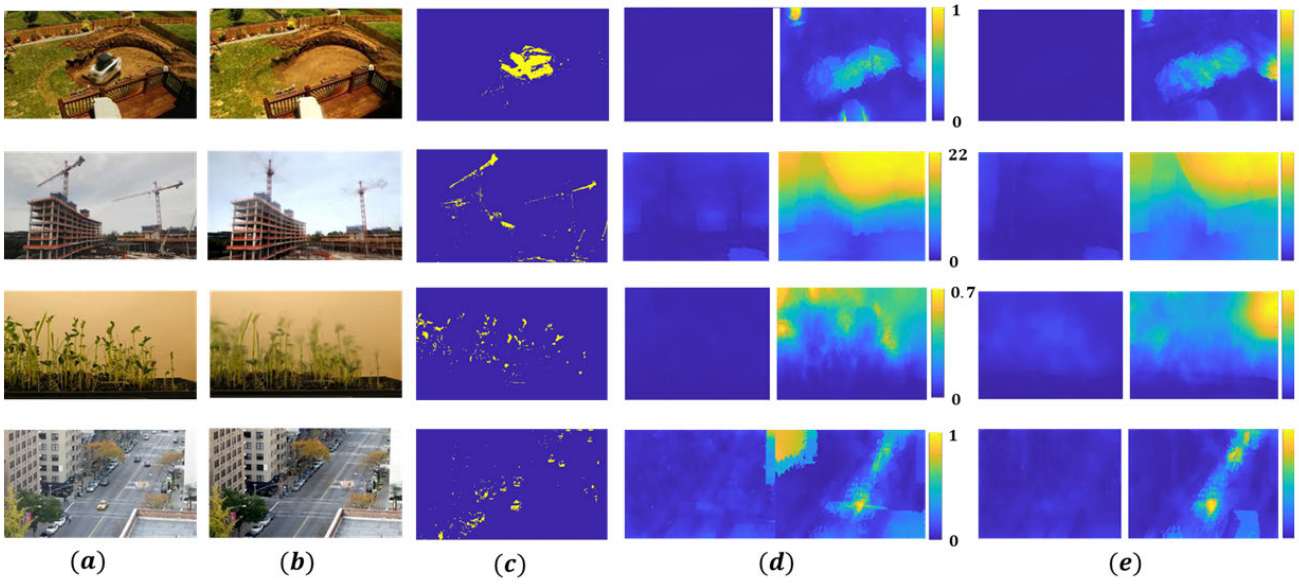

FIGURE 5. (a) frame examples of original videos. (b) frame examples of denoised videos. (c) the difference between the original frame and the denoised frame. (d) the average optical flow results along the horizontal direction (d) and the vertical direction (e) of both the original (right) and denoised videos (left).

several challenges, including motion discontinuities, large motions, and motion rigidity, was used to produce the motion fields. The average estimated values of horizontal and vertical optical flow results for both the original and denoised videos are shown in Fig. 5, in which, for the sake of visual comparison, the average results extracted from the denoised videos are magnified by a factor of five.

The average results extracted from the original video feature large-magnitude motion vectors caused by the short-term motion of cars. By contrast, the results extracted from the denoised video feature small-magnitude motion vectors in both the vertical and horizontal motion components. This indicates that the proposed technique was effective at canceling the short-term motions of several objects, including the crane, plant leaves, and vehicles. Time-lapse sequences such as construction and painting videos that display progressive change are good examples of the efficacy of using a moving window scheme. For instance, the painting sequence experiment revealed how the proposed technique appropriately extracted subsequent stroke images of the painted sketch, while the results obtained for the three construction sequences clearly reveal the long-term changes occurring during the construction progress while significantly reducing jittery changes caused by construction tools as well as heavy lighting changes. Furthermore, the airport video results show only cloud movement with all airplane traffic removed. Figure 6 shows motion denoising results extracted from the newly added time-lapse sequences.

2) VISUAL RESULT OF SNOWFLAKES VIDEOS

Snowflake removal is a significant preprocessing procedure that enhances outdoor video processing tasks such as segmentation, tracking and abnormal event detection. Snowflakes show a repetitive behavior along the time dimension where their frequencies are higher than other foreground objects and lower than the background components. Therefore, snowflakes can be easily removed from a given video by employing our motion denoising approach video where low temporal frequencies that carry the snowflakes patterns are filtered out resulting in a snowflake free background image. In this experiment, we implemented our proposed snowflake removal on several video sequences introduced by Tian *et al.* [30]. The videos show different snow scenarios that range from mid snow falling to heavy snow falling in addition to different sizes of snowflakes. Figure 7 shows a visual comparison between the proposed approach and different existing techniques [39], [40], [30]. Both methods of [39] and [40] show poor performance in the majority of the tested videos where snowflakes residue, object distortion, and blur can be easily seen in their results. On the other hand, the proposed approach and the method of [30] show similar satisfactory visual results where the snowflakes are removed and the visibility in the scene is significantly enhanced. However, one advantage of our proposed approach over the method of [30] is that it does not require the additional processing phase proposed in [30] to remove the snowflakes from foreground objects as can be seen in the 5^{th} row in Fig.7.

3) QUANTITATIVE RESULTS

Two assessments were carried out to numerically measure the efficiency of the proposed motion denoising technique. In the first, the abilities of the proposed technique and the technique in Rubinstein *et al.* to remove short-term motion were compared based on estimates of the average optical flow magnitude of all pairs of subsequent frames in each test sequence. The temporal step size was limited to a one-frame

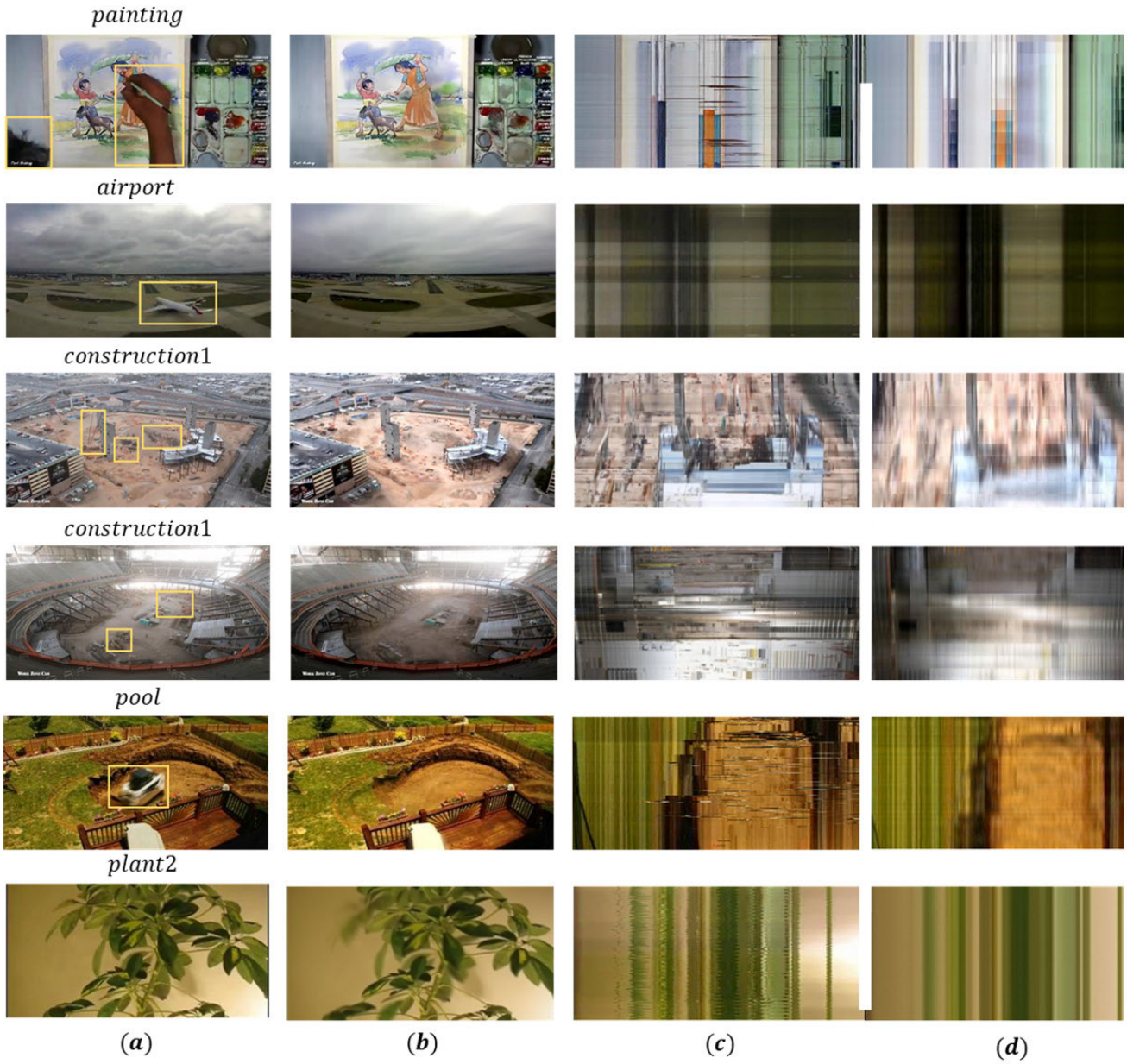

FIGURE 6. (a) examples of original videos. (b) examples of denoised videos. (c) spatiotemporal slices extracted from the original videos. (d) spatiotemporal slices extracted from the denoised videos.

difference to ensure that the estimated optical flows represented short-term motion. In this case, smaller average flow magnitudes corresponded to better motion denoising results. The estimated optical flow magnitude results obtained for five videos provided by Rubinstein *et al.* are listed in Table 3.

Considering the small differences between the results obtained using the respective techniques, it is difficult to conclude which performed better. Both techniques produced visually comparable results and were successful in removing jerky short-term events from the time-lapse sequences. However, because the proposed moving window-based SVD requires only $O(n\varepsilon^2)$ operations to compute the decomposition results for a spatiotemporal slice, its computational load was much lower than that required by the

TABLE 3. Estimated motion errors of proposed technique and rubinstein technique.

sequence	Proposed	Rubinstein et al
Pool	**0.010**	0.017
Street	**0.022**	0.034
Plant1	**0.009**	**0.009**
Plant2	0.042	**0.035**
pond	**0.024**	0.084

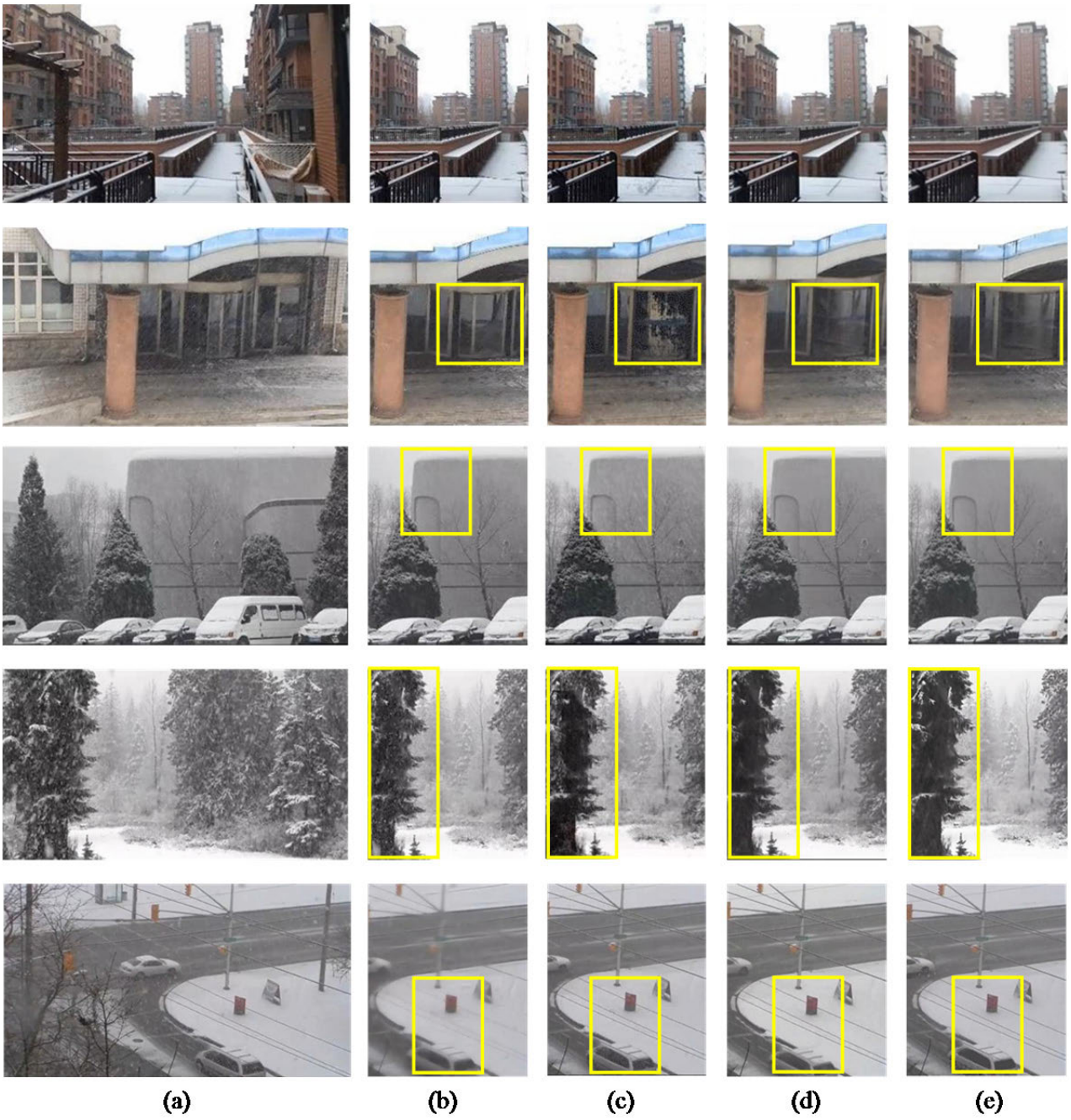

FIGURE 7. Visual examples of snowflake removal results: (a) input video, (b) Kim *et al.* method [39], (c) Sakaaino method [40], (d) Tian et al method [30], and (e) proposed method.

Rubinstein *et al.* technique, which uses a motion compensation algorithm based on a space-time loopy belief propagation algorithm with a computational complexity of $O(k^3)$ for each dimension, where k indicates the size of the search volume. Additional computational burden such as sequential scanning along the row, column, and time spaces, bilinear interpolation and scaling, and iterative processing further increase the processing time required by Rubinstein's method. Finally, their massive 3D grid-based technique introduces spatially related computational difficulties in which the maximum number of frames per sequence cannot exceed. On average, Rubinstein's method requires up to 50 h to denoise three-hundred 300 $\times$ 300 frames, whereas the proposed technique requires only 67.26 s to denoise a video with these dimensions The second assessment followed the approach presented in [41], in which the performance of a given technique can be evaluated by measuring its impact when added to a conventional foreground segmentation. To carry out this evaluation, a group of challenged maritime video sequences in which a significant component of the background (sea surface)

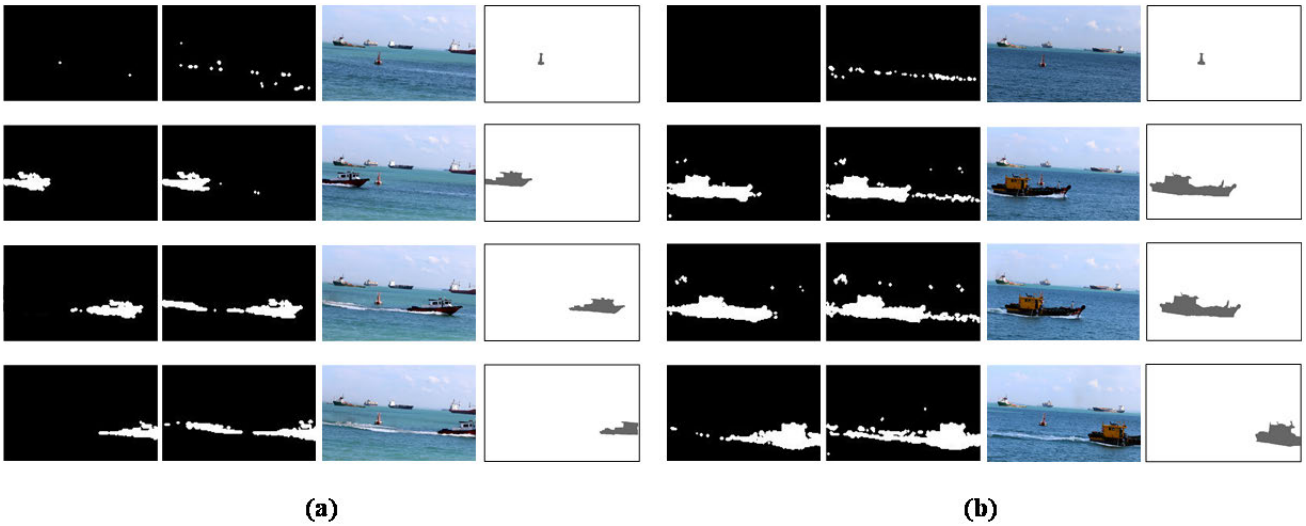

FIGURE 8. Examples of foreground segmentation results of proposed (SVD+MD) method (1st column) and traditional SVD (2nd column) extracted from the original frames (3rd column) in MVI-1470-VIS (a) and MVI-1471-VIS (b) videos. The groundtruth images are shown in 4th column.

TABLE 4. Estimated foreground segmentation results with and without the proposed motion denoizing scheme.

Sequence		*P*	*R*	*F*-measure
video 1452	SVD	0.786	0.756	0.770
	SVD+MD	0.824	0.804	**0.801**
video 1470	SVD	0.583	0.771	0.669
	SVD+MD	0.872	0.720	**0.779**
video 1471	SVD	0.630	0.766	0,636
	SVD+MD	0.961	0.676	**0.791**
video 1481	SVD	0.304	0.698	0.406
	SVD+MD	0.504	0.751	**0.602**
video 1484	SVD	0.461	0.868	0.601
	SVD+MD	0.501	0.852	**0.656**
video 1582	SVD	0.842	0.854	0.847
	SVD+MD	0.901	0.855	**0.864**
video 1613	SVD	0.397	0.321	0.346
	SVD+MD	0.501	0.401	**0.442**
video 1617	SVD	0.344	0.246	0.274
	SVD+MD	0.421	0.452	**0.412**
video 1622	SVD	0.430	0.737	0.538
	SVD+MD	0.507	0.759	**0.621**
video 1624	SVD	0.624	0.634	0.626
	SVD+MD	0.629	0.684	**0.646**

was dynamically moving on a short-term basis while foreground objects (boats) were present on a long-term basis was selected. Theoretically, an efficient motion denoising technique should improve the results of foreground segmentation because the majority of pixels that represent the sea surface are usually detected as foreground objects. To carry out the assessment, the Singapore Maritime (SM) dataset [42], which contains several challenged maritime videos, was used. The provided ground truths of the moving objects were limited to bounding boxes to enable the use of pixel-based ground truth images, as in [43]. The motion denoising mechanism was added to the framework of the best technique reported in [43] to decompose the estimated foreground tensor into long- and short-term moving objects. The extracted short-term moving objects were then integrated with the background tensor to provide better representation of the dynamic nature of the background. Table 4 lists the results obtained by combining, respectively, the original SVD-based foreground segmentation alone and a combination of the foreground segmentation method and the proposed motion denoising mechanism (SVD+MD). Both the numerical and the visual results (see Fig.8) clearly show that the foreground segmentation is significantly enhanced under the addition of the proposed segmentation method, which increases the values of both the precision and recall metrics relative to the original SVD-based segmentation.

VII. CONCLUSION

In this paper, two SVD-based schemes for color magnification and motion denoising were presented. These schemes apply different processing operations to temporally process the decomposed singular components of a given video to achieve a desired processing result. To magnify the intensity of color changes in a specific event, a particular set of left singular vectors that carries sufficient color change information can be selected, filtered, and rescaled accordingly. The results of our color magnification scheme are robust to the presence of several types of accompanying motion. Further temporal filtering of the sparse components by removing

their high-frequency components enables the elimination of short-term events and the preservation of long-term events. The global nature of the proposed scheme in terms of the spatiotemporal slicing process applied to video segments results in a significantly reduced computational complexity relative to conventional techniques, which are mostly pixel-based. This reduction in complexity arises from the fact that processing the singular components (instead of pixel profiles) representing a target event produces a desired result with much less computational burden. For instance, the magnification of the first singular left vector magnifies the intensities of only those pixel profiles containing color changes. While the problem of magnification remains one of the most important issues in the field of video processing, motion magnification is no less important. In future work, we will address the motion magnification problem through the use of right singular vectors to theoretically represent the displacements of pixel intensities caused by moving objects.

REFERENCES

[1] T.-A. Song, F. Yang, S. R. Chowdhury, K. Kim, K. A. Johnson, G. El Fakhri, Q. Li, and J. Dutta, "PET image deblurring and super-resolution with an MR-based joint entropy prior," *IEEE Trans. Comput. Imag.*, vol. 5, no. 4, pp. 530–539, Dec. 2019.

[2] Y. Huang, Y. Quan, Y. Xu, R. Xu, and H. Ji, "Removing reflection from a single image with ghosting effect," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 34–45, 2020.

[3] A. M. Teodoro, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Image restoration and reconstruction using targeted plug-and-play priors," *IEEE Trans. Comput. Imag.*, vol. 5, no. 4, pp. 675–686, Dec. 2019.

[4] P.-M. Jodoin, V. Saligrama, and J. Konrad, "Behavior subtraction," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4244–4255, Sep. 2012.

[5] V. Mahadevan and N. Vasconcelos, "Background subtraction in highly dynamic scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–6.

[6] S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, J. F. Cohn, and N. Sebe, "Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2396–2404.

[7] J. Kranjec, S. Beguš, G. Geršak, and J. Drnovšek, "Non-contact heart rate and heart rate variability measurements: A review," *Biomed. Signal Process. Control*, vol. 13, pp. 102–112, Sep. 2014.

[8] X. Li, J. Chen, G. Zhao, and M. Pietikainen, "Remote heart rate measurement from face videos under realistic situations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 4264–4271.

[9] M. Rubinstein, C. Liu, P. Sand, F. Durand, and W. T. Freeman, "Motion denoising with application to time-lapse photography," in *Proc. CVPR*, Jun. 2011, pp. 313–320.

[10] M. A. Elgharib, M. Hefeeda, F. Durand, and W. T. Freeman, "Video magnification in presence of large motions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4119–4127.

[11] J. F. P. Kooij and J. C. van Gemert, "Depth-aware motion magnification," in *Computer Vision—ECCV*, vol. 9912, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, pp. 467–482.

[12] J. G. Chen, N. Wadhwa, Y.-J. Cha, F. Durand, W. T. Freeman, and O. Buyukozturk, "Structural modal identification through high speed camera video: Motion magnification," in *Topics in Modal Analysis I*, vol. 7. Cham, Switzerland: Springer, 2014, pp. 191–197.

[13] A. Davis, M. Rubinstein, N. Wadhwa, G. J. Mysore, F. Durand, and W. T. Freeman, "The visual microphone," *ACM Trans. Graph.*, vol. 33, no. 4, pp. 1–10, Jul. 2014.

[14] C. Liu, A. Torralba, W. T. Freeman, F. Durand, and E. H. Adelson, "Motion magnification," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 519–526, Jul. 2005.

[15] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman, "Phase-based video motion processing," *ACM Trans. Graph.*, vol. 32, no. 4, p. 80, 2013.

[16] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 1–8, 2012.

[17] Y. Zhang, S. L. Pintea, and J. C. Van Gemert, "Video acceleration magnification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 502–510.

[18] S. Takeda, K. Okami, D. Mikami, M. Isogai, and H. Kimata, "Jerk-aware video acceleration magnification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1769–1777.

[19] S. Takeda, Y. Akagi, K. Okami, M. Isogai, and H. Kimata, "Video magnification in the wild using fractional anisotropy in temporal distribution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1614–1622.

[20] A. Davis and M. Agrawala, "Visual rhythm and beat," *ACM Trans. Graph.*, vol. 37, no. 4, pp. 1–11, Aug. 2018.

[21] N. Bonneel, J. Tompkin, K. Sunkavalli, D. Sun, S. Paris, and H. Pfister, "Blind video temporal consistency," *ACM Trans. Graph.*, vol. 34, no. 6, pp. 1–9, Nov. 2015.

[22] J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan, "Webcam clip art," *ACM Trans. Graph.*, vol. 28, no. 5, pp. 1–10, Dec. 2009.

[23] S. Meyer, A. Djelouah, B. McWilliams, A. Sorkine-Hornung, M. Gross, and C. Schroers, "PhaseNet for video frame interpolation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 498–507.

[24] Z. W. Wang, W. Jiang, K. He, B. Shi, A. Katsaggelos, and O. Cossairt, "Event-driven video frame synthesis," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 4320–4329.

[25] X. Ding, Y. Gaobo, R. Li, L. Zhang, Y. Li, and X. Sun, "Identification of motion-compensated frame rate up-conversion based on residual signals," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 7, pp. 1497–1512, Jul. 2018.

[26] T.-H. Tsai, A.-T. Shi, and K.-T. Huang, "Accurate frame rate up-conversion for advanced visual quality," *IEEE Trans. Broadcast.*, vol. 62, no. 2, pp. 426–435, Jun. 2016.

[27] K. Sunkavalli, W. Matusik, H. Pfister, and S. Rusinkiewicz, "Factored time-lapse video," *ACM Trans. Graph.*, vol. 26, no. 3, p. 101, Jul. 2007.

[28] R. Martin-Brualla, D. Gallup, and S. M. Seitz, "Time-lapse mining from internet photos," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 62:1-62:8, Jul. 2015.

[29] J. Tan, M. Dvorožňák, D. Sýkora, and Y. Gingold, "Decomposing time-lapse paintings into layers," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 61:1–61:10, Jul. 2015.

[30] J. Tian, Z. Han, W. Ren, X. Chen, and Y. Tang, "Snowflake removal for videos via global and local low-rank decomposition," *IEEE Trans. Multimedia*, vol. 20, no. 10, pp. 2659–2669, Oct. 2018.

[31] T. Bouwmans, A. Sobral, S. Javed, S. K. Jung, and E.-H. Zahzah, "Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset," *Comput. Sci. Rev.*, vol. 23, pp. 1–71, Feb. 2017.

[32] N. Iqbal, E. Liu, J. H. McClellan, A. Al-Shuhail, S. I. Kaka, and A. Zerguine, "Detection and denoising of microseismic events using time–frequency representation and tensor decomposition," *IEEE Access*, vol. 6, pp. 22993–23006, 2018.

[33] L. N. Trefethen and D. Bau, III, *Numerical Linear Algebra*, vol. 50. Philadelphia, PA, USA: SIAM, 1997.

[34] N. Kamel, I. Kajo, and Y. Ruichek, "On visual periodicity estimation using singular value decomposition," *J. Math. Imag. Vis.*, vol. 61, no. 8, pp. 1135–1153, Oct. 2019.

[35] I. Kajo, N. Kamel, and Y. Ruichek, "Incremental tensor-based completion method for detection of stationary foreground objects," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 5, pp. 1325–1338, May 2019.

[36] I. Kajo, N. Kamel, Y. Ruichek, and A. Malik, "SVD-based tensor-completion technique for background initialization," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 3114–3126, Jun. 2018.

[37] N. S. Kamel, S. Sayeed, and G. A. Ellis, "Glove-based approach to online signature verification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 1109–1113, Jun. 2008.

[38] D. Sun, S. Roth, and M. J. Black, "Secrets of optical flow estimation and their principles," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 2432–2439.

[39] J.-H. Kim, J.-Y. Sim, and C.-S. Kim, "Video deraining and desnowing using temporal correlation and low-rank matrix completion," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2658–2670, Sep. 2015.

[40] H. Sakaino, "A semitransparency-based optical-flow method with a point trajectory model for particle-like video," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 441–450, Feb. 2012.
[41] C. H. Bahnsen and T. B. Moeslund, "Rain removal in traffic surveillance: Does it matter?" *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 8, pp. 2802–2819, Aug. 2019.
[42] D. K. Prasad, C. K. Prasath, D. Rajan, L. Rachmawati, E. Rajabally, and C. Quek, "Object detection in a maritime environment: Performance evaluation of background subtraction methods," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 5, pp. 1787–1802, May 2019.
[43] I. Kajo, N. Kamel, and Y. Ruichek, "Tensor-based approach for background-foreground separation in maritime sequences," *IEEE Trans. Intell. Transp. Syst.*, early access, Jul. 1, 2020, doi: 10.1109/TITS.2020.3001687.

IBRAHIM KAJO received the M.Sc. and Ph.D. degrees in electrical and electronic engineering from PETRONAS University of Technology (UTP), Malaysia, in 2015 and 2018, respectively. He is currently a Postdoctoral Researcher with the University of Technology of Belfort-Montbéliard (UTBM). His research interests include computer vision, motion detection, visual surveillance, motion magnification, crowd analytics, and artificial intelligence.

NIDAL KAMEL (Senior Member, IEEE) received the Ph.D. degree (Hons.) in telecommunication from the Technical University of Gdansk, Poland, in 1993. Since then he has been involved in research projects related to subspace techniques in signal and image processing, estimation theory, noise reduction, optimal filtering, and pattern recognition. He is currently an Associate Professor with PETRONAS University of Technology.

YASSINE RUICHEK (Senior Member, IEEE) received the Ph.D. degree in control and computer engineering and the Habilitation à Diriger des Recherches (HDR) degree in physic science from the University of Lille, France, in 1997 and 2005, respectively. Since 2007, he has been a Full Professor with the University of Technology of Belfort-Montbéliard (UTBM). His research interests include computer vision, image processing and analysis, pattern recognition, data fusion, and localization, with applications in intelligent transportation systems and video surveillance.

ABDULRAHMAN AL-AHDAL received the M.Sc. degree in analogue and digital IC design from Imperial College London, London, U.K., in 1999, and the Ph.D. degree from the Electrical Engineering Department, Imperial College London. He spent a few years in industry and then joined the Electrical Engineering Department, Umm Al-Qura University, as a Lecturer, in 2005. His research interests include circuits and systems for biomedical applications.