Self-Learning-Based Signal Decomposition for Multimedia Applications: A Review and Comparative Study

Li-Wei Kang1,2, Chia-Hung Yeh3,*, Duan-Yu Chen4, and Chia-Tsung Lin2

1Graduate School of Engineering Science and Technology National Yunlin University of Science and Technology, Yunlin, Taiwan E-mail: lwkang@yuntech.edu.tw
2Department of Computer Science and Information Engineering National Yunlin University of Science and Technology, Yunlin, Taiwan
3Department of Electrical Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan E-mail: yeh@mail.ee.nsysu.edu.tw
4Department of Electrical Engineering, Yuan Ze University, Taoyuan, Taiwan E-mail: dychen@saturn.yzu.edu.tw

Abstract—Decomposing a signal (e.g., image or video) into multiple semantic components has been an effective research topic for various image/video processing applications, such as image/video denoising, enhancement, and inpainting. In this paper, we present a survey of signal decomposition frameworks based on the uses of sparsity and morphological diversity in signal mixtures and its applications in multimedia. First, we analyze existing MCA (morphological component analysis) based image decomposition frameworks with their applications and explore the potential limitations of these approaches for image denoising. Then, we discuss our recently proposed self-learning based image decomposition framework with its applications to several image/video denoising tasks, including single image rain streak removal, denoising, deblocking, joint super-resolution and deblocking for a highly compressed image/video. By advancing sparse representation and morphological diversity of image signals, the proposed framework first learns an over-complete dictionary from the high frequency part of an input image for reconstruction purposes. An unsupervised or supervised clustering technique is applied to the dictionary atoms for identifying the morphological component corresponding to the noise pattern of interest (e.g., rain streaks, blocking artifacts, or Gaussian noises). Different from prior learning-based approaches, our method does not need to collect training data in advance and no image priors are required. Our experimental results have confirmed the effectiveness and robustness of the proposed framework, which has been shown to outperform state-of-the-art approaches.

I. INTRODUCTION

Decomposing a signal into its building components has been of great interest in many multimedia applications, where the input signal is assumed to be a linear mixture of several source signals [1]. A typical example is to decompose an image into its texture and non-texture parts, which has been recently investigated and extended to the applications of image analysis and synthesis, image enhancement, image restoration, and image content editing [2]–[16]. Let’s consider a fundamental problem of decomposing an image of $N$ pixels into $C$ different content components of $N$ samples for each. To solve this problem, there are as many as $N \times C$ unknowns for the contribution of each component to each pixel in the image. In fact, it is not easy to solve such ill-posed problems with more number of the unknowns than that of the equations. Nevertheless, it has been shown that such decomposition might be possible if sparsity priors of an image can be properly exploited [1]. To morphologically decompose a signal into its building blocks based on sparsity priors has been recently studied extensively and applied to solve many problems in signal and image processing. In this paper, we present a signal decomposition framework based on morphological component analysis (MCA) [1]–[3] for self-learning-based removal of structured noise (e.g., rain streaks removal [7]–[13], deblocking [14]–[16]) and unstructured noise (e.g., Gaussian noise removal [11]) as well as joint super-resolution and deblocking for a highly compressed image/video. We also compare our frameworks with recent state-of-the-art approaches via sparse representation [4], [17].

The rest of this paper is organized as follows. In Sec. II, we briefly review the concepts of MCA-based image decomposition, sparse coding, and dictionary learning techniques. Sec. III presents the proposed self-learning-based signal decomposition framework and its variations to several multimedia denoising applications. In Sec. IV, some experimental comparisons are demonstrated. Finally, Sec. V concludes this paper.

II. MCA-BASED IMAGE DECOMPOSITION

The key idea of MCA is to utilize the morphological diversity of different features contained in the data to be
decomposed and to associate each morphological component to a dictionary of atoms [1]–[3]. Suppose an image \( I \) of \( N \) pixels is a superposition of \( S \) components (called morphological components), denoted by \( I = \sum_{s=1}^{S} I_s \), where \( I_s \) denotes the \( s \)-th component, such as the geometric or textural component of the image \( I \). To decompose \( I \) into \( I_s, s = 1, 2, ..., S \), the MCA algorithms iteratively minimize the following energy function:

\[
E((I_s)^{S}_{s=1}, (\theta_s)^{S}_{s=1}) = \frac{1}{2} \left\| I - \sum_{s=1}^{S} I_s \right\|_2^2 + \tau \sum_{s=1}^{S} E_s(I_s, \theta_s),
\]

where \( \theta_s \) denotes the sparse coefficients corresponding to \( I_s \) with respect to the dictionary \( D_s \), \( \tau \) is a regularization parameter, and \( E_s \) is the energy function defined according to the type of \( D_s \) (global or local dictionary). The MCA algorithms solve (1) by iteratively performing for each component \( I_s \), the following two steps: (i) update of the sparse coefficients: this step performs sparse coding [18]–[20] to solve \( \theta_s \) or \( \{\theta_s^p\}_{p=1}^P \), where \( \theta_s^p \) represents the sparse coefficients of the \( p \)-th patch extracted from \( I_s \), and \( P \) is the total number of extracted patches, to minimize \( E_s(I_s, \theta_s) \) while fixing \( I_s \); and (ii) update of the components: this step updates \( I_s \) or \( \{b_s^p\}_{p=1}^P \) while fixing \( \theta_s \) or \( \{\theta_s^p\}_{p=1}^P \). More details about MCA can be found in [1]–[3].

It should be noted that separating and removing noises considered in this paper (e.g., rain streaks or blocking artifacts) from a single image is not a trivial work as the noises are usually highly mixed with the image to be decomposed, making the decomposition process very challenging. Therefore, existing MCA-based image decomposition technique [1]–[3] introduced in this section is hard to be directly applied. Even if some signal decomposition or denoising results have been achieved by the existing MCA-like approaches [1]–[5], most of them rely on some prior knowledge about the component to be removed (e.g., pre-collected related training examples or the standard deviation of Gaussian noise). To make a signal decomposition framework more realistic, we present a self-learning-based signal decomposition framework as well as its various extensions, which can be adapted to several multimedia applications, introduced in Sec. III.

III. SELF-LEARNING-BASED SIGNAL DECOMPOSITION FRAMEWORK FOR MULTIMEDIA APPLICATIONS

In this section, we present our self-learning based signal decomposition framework for single image denoising including rain streak removal [7]–[13], deblocking [14]–[15], joint deblocking and super-resolution in a highly compressed image [16], and Gaussian noise removal [11], which can be naturally extended to video denoising with temporal information/constraint incorporated. To decompose an input image into two or more morphological components, we propose to first decompose the high-frequency (HF) part of the image, since the image components to be preserved and those to be removed (e.g., noise patterns) are typically mixed in an image. Unlike prior sparse representation or MCA-based image decomposition works which require the collection of training data for observing the image dictionaries, we advocate the learning of the dictionaries directly from the input image itself, so that the image components associated with undesirable patterns can be automatically identified and extracted, while most original image details can be preserved.

A. Self-Learning-Based Single Image Rain Removal

For simplicity, we start to briefly present our MCA-based framework with our first single image rain removal work [7]–[8] as an example, shown in Fig. 1. For an input rain image \( I \), we first apply a filter (e.g., the bilateral filter [21] or guided filter [22]) to roughly decompose \( I \) into the low-frequency (LF) part \( (I_{LF}) \) and HF part \( (I_{HF}) \). Then, our method learns a dictionary \( D_{HF} \) via online dictionary learning [20] based on the training patches extracted from \( I_{HF} \) itself to further decompose \( I_{HF} \), where \( D_{HF} \) can be further divided into two sub-dictionaries, \( D_{HF,G} \) and \( D_{HF,N} \) ( \( D_{HF} = [D_{HF,G}D_{HF,N}] \)), for representing the non-noise (e.g., non-rain) and noise (e.g., rain) components of \( I_{HF} \), respectively. Here, we perform unsupervised clustering (e.g., the K-means algorithm) to partition \( D_{HF} \) into \( D_{HF,G} \) and \( D_{HF,N} \), where the diversities of them have been well verified. As a result, we formulate the denoising (e.g., rain removal) problem for...
image \( I \) as a sparse representation-based signal decomposition problem as:

\[
\min_{\theta_{HF} \in \mathbb{R}^m} \left\| b_{HF}^k - D_{HF} \theta_{HF}^k \right\|^2_2 \quad \text{s.t.} \quad \left\| \theta_{HF}^k \right\|_0 \leq L,
\]

(2)

where \( b_{HF}^k \in \mathbb{R}^m \) represents the \( k \)-th patch extracted from \( I_{HF} \).

\( \theta_{HF}^k \in \mathbb{R}^{n \times m} \) are the sparse coefficients of \( b_{HF}^k \) with respect to \( D_{HF} \in \mathbb{R}^{n \times m} \), \( n \leq m \), and \( L \) denotes the sparsity or maximum number of nonzero coefficients of \( \theta_{HF}^k \). Eq. (2) can be solved via \( l_1 \)-minimization [18], and then each patch \( b_{HF}^k \) can be reconstructed and used to recover either the non-noise or noise component of \( I_{HF} \) depending on the corresponding nonzero coefficients in \( \theta_{HF}^k \), i.e., the used atoms from \( D_{HF,L} \) or \( D_{HF,R} \). Finally, integrating the recovered non-noise component with \( I_F \), the denoised version of \( I \) can be produced. More details and experimental results of this framework can be found in [7]–[8].

In the following subsections, we shall show how to extend this framework [7] to solve advanced single image [10]–[13] or video [9] rain removal, deblurring [14]–[15], joint super-resolution and deblurring/deringing for highly compressed images/videos with severe compression artifacts [16], and Gaussian noise removal [11]. These structured noise removal problems can all be successfully addressed with the proposed framework by properly integrating the dictionary learning and unsupervised/supervised clustering modules in a unified framework. Our experiments have verified that the proposed framework is powerful and versatile.

### B. Extension to Static Video Rain Removal

To extend our signal decomposition framework for single image rain removal [7] to video rain removal [9], we consider a rain video of static scene without significant moving objects. Fig. 2 shows the proposed video-based rain removal framework, where we find that by averaging a number of successive frames in a static scene, the rain streaks can be eliminated in this “average frame,” which can be used to replace the filtered image (LF part) in our single-image-based method [7]. For an input rain video of \( V \) frames \( I_i \), \( i = 1, 2, \ldots, V \), we average the first \( Z \) successive frames \( I_i \), \( i = 1, 2, \ldots, Z \), to generate the common LF part \( I_{AVE} \) for all of the remaining frames \( I_i \), \( i = Z + 1, Z + 2, \ldots, V \), in the video. We then apply our single-image-based method [7] to \( I_{Z+1} \) with LF part being set to \( I_{AVE} \) to obtain the rain and non-rain dictionaries, \( D_{(Z+1)_R,I} \) and \( D_{(Z+1)_G,I} \), to decompose the HF part \( I_{Z+1} = I_{AVE} + I_{HF} \) of \( I_{Z+1} \) into the rain and non-rain components, \( I_{Z+1,R} \) and \( I_{Z+1,G} \), respectively. Then, the rain-removed version of \( I_{Z+1} \) can be obtained via \( I_{Z+1,\text{Rain removed}} = I_{AVE} + I_{(Z+1)_G} \). For removing rain streaks from \( I_i \), \( i = Z + 1, Z + 2, \ldots, V \), in the video, we use the same LF part \( I_{AVE} \) to obtain the HF part \( I_{HF,i} = I_i - I_{AVE} \). We then directly perform MCA decomposition to \( I_{HF,i} \), \( i = Z + 1, Z + 2, \ldots, V \), using the same two dictionaries, \( D_{(Z+1)_R,I} \) and \( D_{(Z+1)_G,I} \), learned from \( I_{HF,Z+1} \) to obtain the rain and non-rain components, \( I_{HF,R,i} \) and \( I_{HF,G,i} \), respectively. Finally, the rain-removed version of \( I_i \), \( i = Z + 1, Z + 2, \ldots, V \), can be obtained via \( I_{i,\text{Rain removed}} = I_{AVE} + I_{HF,G,i} \). More details and experimental results of this framework can be found in [9].

### C. Extension to Context-aware Single Image Rain Removal

To further improve the performance of our single-image-based rain removal framework [7]–[8], we relied on the fact that in most rain images, the rain streaks are present in the entire image and with similar gradients [10]. Therefore, the dictionaries learned from different context categories should share common atoms which indicate the rain patterns. This observation inspired us to present a context-aware framework for single image rain removal. As illustrated in Fig. 3, our framework learns context information in an unsupervised setting, while the rain patterns can be automatically identified and removed from dictionaries learned for each context category. In this extension [10], instead of learning only one dictionary for an image in [7]–[8], we proposed to learn the

Fig. 2. Block diagram of the proposed video-based rain streak removal method [9].
dictionary for each context category (image segment), followed by an automatic identification of atoms which correspond to rain and non-rain patterns. Therefore, one of the major contributions of this framework is to exploit the local characteristics of context categories to further improve the performance of [7]–[8].

Based on the characteristics that the dictionaries learned from different context categories should share common atoms (rain patterns), while for the atoms strongly uncorrelated to each other across different context categories should be associated with non-rain patterns, we collected two sets of atoms which correspond to rain and non-rain atoms with high confidence, so that we could train a support vector machine (SVM) [23] classifier on them. This classification process has been shown to be more confident than the K-means clustering process used in [7]. More details and experimental results of this framework can be found in [10].

D. Extension to Visual Depth Guided Single Image Rain Removal

To further improve the rain removal performance, we also proposed a single-color-image-based rain removal framework [12]–[13] by formulating rain removal as an image decomposition problem. Similar to [7], an input color image is also first decomposed into low-frequency part and high-frequency part by a low-pass filter (the guided image filter [22] was used in [12]–[13]). Different from [7], to separate rain streaks from the high-frequency part, a hybrid feature set, including HoG (histogram of oriented gradients) [24], DoF (depth of field) [25]–[26], and Eigen color [27], is employed to further decompose the high-frequency part.

Fig. 3. Block diagram of the proposed context-aware single-image-based rain streak removal method [10].

Fig. 4. Block diagram of the proposed visual depth guided single image rain removal framework [12], where “α” denotes the α-blending operation and α is a weighting map.
As shown in Fig. 4, in [12], an input rain image $I$ is first roughly decomposed into the LF part $\hat{I}_{LF}$ and the HF part $\hat{I}_{HF}$ using the guided filter [22]. $\hat{I}_{HF}$ can then be further roughly decomposed into the rain component $\hat{I}_{R}$ and the non-rain (geometric) component $\hat{I}_{NR}$ via signal decomposition. In the decomposition of $\hat{I}_{HF}$, the learned dictionary $D_{\hat{I}_{HF}}$ for $\hat{I}_{HF}$ is separated into the two sub-dictionaries, $D_{\hat{I}_{R}}$ and $D_{\hat{I}_{NR}}$, respectively, where $D_{\hat{I}_{HF}} = [D_{\hat{I}_{R}} \mid D_{\hat{I}_{NR}}]$, via the HoG feature-based dictionary atom clustering [7]. Then, to refine $\hat{I}_{R}$ by extracting misclassified non-rain region from $\hat{I}_{R}$, a new weighting map $\text{DoF}_I$ derived from $\text{DoF}_I$ is used to blend $I$ and $\hat{I}_{R}$ to obtain the enhanced version $\hat{I}_{R,\text{final}}$. Then, the regions $\hat{I}_{R,\text{final}}, \hat{I}_{R}$, and $\hat{I}_{NR,\text{final}}$ are integrated to obtain the final rain-removed output image. Compared with [7], the key of [12] is three-fold: (i) enhancement of the LF part using the DoF saliency map of the input image; (ii) recovery of the misclassified non-rain region in the rain component; and (iii) enhancement of the color information of the non-rain region extracted from the HF part. More details and experimental results of this framework can be found in [12].

**E. Extension to Blocking Artifacts Removal for Compressed Image/Video**

To investigate other multimedia applications of our signal decomposition framework, we found that this framework is suitable to be extended to blocking artifacts removal (or deblocking) of compressed image/video [14]–[15]. Inspired by [7], we proposed a self-learning-based post-processing framework for image/video deblinking by properly formulating deblinking as an MCA-based image decomposition problem via sparse representation [14]–[15]. Without the need of any prior knowledge (e.g., the positions where blocking artifacts occur, the algorithm used for compression, or the characteristics of image/video to be processed) about the blocking artifacts to be removed, our method can automatically learn two dictionaries for decomposing an input decoded image into its “blocking component” and “non-blocking component.”

As shown in Fig. 5, an input image with blocking artifacts is first roughly decomposed into the LF part and the HF part using a low-pass filter, where the most basic information will be retained in the LF part while the blocking artifacts and the other edge/texture details will be included in the HF part of the image as illustrated in Figs. 5(b) and 5(c), respectively. Then, we perform our MCA-based image decomposition to further decompose the HF part into the non-blocking component [see Fig. 5(d)] and the blocking component [see Fig. 5(e)]. In the decomposition step, a dictionary learned from the training exemplars extracted from the HF part of the image itself can be divided into two sub-dictionaries [see Figs. 5(f) and 5(g)] by performing our HOG feature-based dictionary atom clustering modified from [7]. Then, we perform sparse coding [20] based on the two sub-dictionaries to achieve MCA-based image decomposition, where the non-blocking component in the HF part can be obtained, followed by integrating with the LF part of the image to obtain the deblinking version of this image as illustrated in Fig. 5(h). More details and experimental results of this framework can be found in [14].
F. Extension to Joint Super-Resolution and Deblocking for a Highly Compressed Image/Video

Furthermore, by considering the fact that low-quality images/videos are usually not only with low-resolution, but also suffer from compression artifacts (e.g., blocking artifacts), we proposed a self-learning-based super-resolution (SR) framework to simultaneously achieve single-image SR and compression artifact removal for a highly-compressed image [16]. In this method [16], we proposed to self-learn image sparse representation for modeling the relationship between low and high-resolution image patches in terms of the learned dictionaries, respectively, for image patches with and without blocking artifacts, where the signal decomposition characteristic in [7] has been involved. As a result, image SR and deblocking can be simultaneously achieved via sparse representation and MCA-based image decomposition [7]. More details and experimental results of this framework can be found in [16].

G. Extension to Gaussian Noise Removal

Other than removing structured noise (e.g., rain streaks [7]−[13] or blocking artifacts [14]−[16]), we have found that our signal decomposition framework is also suitable to be applied to unstructured noise removal (e.g., Gaussian noise [11]). Even Gaussian noise is randomly generated without regular pattern, its characteristic is still very different from other image components. Therefore, our signal decomposition method can still self-learn two sets of dictionary atoms for sparsely representing the Gaussian-noise component and the noise-free component to achieve noise removal of an input noisy image. Unlike existing denoising methods (e.g., [4], [21]), it is not required to know the standard deviation of such noise to be given in advance, which makes our method more practical for real-world applications. More details and experimental results of this framework can be found in [11].

IV. COMPARATIVE STUDY RESULTS

In this section, we present some experimental results of our signal decomposition frameworks for rain streak removal, blocking artifact removal, Gaussian noise removal, and joint super-resolution and deblocking for images/videos. We also compare our results with those of some state-of-the-art approaches [4], [17], [21], [22], [28]. More experimental results and discussions can be found from [7]−[16].

A. Performance Evaluation on Image/Video Rain Removal

To evaluate the performance of image/video rain removal, we collected several natural/synthetic rain images/videos from the Internet (with ground-truth images/videos for a few of them). Figs. 6-7 show some single image and video rain removal results, respectively. Based on Fig. 6, our frameworks [7], [10], [12] significantly outperform existing approaches [4], [21], [22] used for comparisons in single image rain removal. The results demonstrate that although these existing denoising filter-based methods [4], [21], [22] can remove most rain streaks, they both simultaneously remove much image detail as well. On the other hand, the proposed methods [7], [10], [12] successfully remove most rain streaks while preserving most non-rain image details in these test cases. In addition, by advancing the techniques of context-constrained image segmentation, categorization, and sparse coding, our context-aware method [10] can outperform our MCA-based method [7] which may remove parts of non-rain components as well due to the heuristic dictionary

Fig. 6. Single image rain removal results: (a) the original non-rain image; (b) the rain image of (a); the rain-removed versions of (b) via the: (c) bilateral filtering [21]; (d) guided filtering [22]; (e) K-SVD-based denoising [4]; (f) proposed MCA-based [7]; (g) proposed context-aware [10]; and (h) proposed visual depth guided [12] methods.
partition. Furthermore, our visual depth guided method \cite{12} can recover/protect non-rain component while removing rain streaks, which can, to the best of our knowledge, achieve state-of-the-art performance in single image rain removal. Moreover, our video rain removal method \cite{9} also outperforms our single image-based method (lack of temporal information) \cite{7} and the state-of-the-art video-based method (based on very accurate rain detection result and simple interpolation-based method) \cite{28} in static video case.

B. Performance Evaluation on Image/Video Blocking Artifacts Removal

To evaluate the performance of our blocking artifacts removal method \cite{14}, H.264/AVC-decoded video sequences were used. Fig. 8 shows some deblocking results obtained by the H.264/AVC with in-loop filter \cite{29}, a signal adaptive weighted sum (SAWS) technique \cite{30}, and our method \cite{14}. It can be found from Fig. 8 that the proposed method \cite{14} outperforming the H.264/AVC with in-loop filter \cite{29} and SAWS \cite{30} methods can remove blocking artifacts from the original H.264/AVC-decoded versions while preserving acceptable visual quality, mainly benefiting from the proposed self-learning-based image decomposition strategy.

C. Performance Evaluation on Gaussian Noise Removal

To evaluate the performance of our Gaussian noise removal method \cite{11}, we compared our method with $K$-SVD-based denoising \cite{4}, and show the results (with standard deviation value $\sigma$ of the Gaussian noise set to 25) in Fig. 9. The exact value of the parameter $\sigma$ is required for the $K$-SVD method, while it is not required to be known in advance for our method. We assume this parameter is not exactly known in practical and simply set a large standard deviation value $\sigma = 35$ for our experiment. As a result, our method quantitatively and qualitatively outperformed the other without needing to know the standard deviation of Gaussian noise to be removed.

D. Performance Evaluation on Joint Super-Resolution and Deblocking for a Highly Compressed Image

To evaluate the performance of our joint super-resolution (SR) and deblocking method \cite{16}, we compared our method with a baseline SR method (bicubic interpolation \cite{31}) and a state-of-the-art learning-based SR method via sparse coding (ScSR) \cite{17}. Fig. 10 shows the SR results for a highly compressed JPEG image with quality factor (QF) set to 15. It can be observed from Fig. 10 that the visual quality of our SR result outperforms those obtained by the pure SR approaches used for comparisons, where blocking artifacts are significantly magnified while directly enlarging the image.
V. CONCLUSIONS

In this paper, we have presented a brief review and comparative studies for our self-learning-based signal decomposition framework via sparse representation as well as several its extensions and applications in multimedia. Our signal decomposition framework mainly based on the uses of sparsity and morphological diversity in signal mixtures has demonstrated good performances in variety of applications, including image/video rain streaks removal \cite{7}–\cite{13}, blocking artifacts removal \cite{14}–\cite{15}, Gaussian noise removal \cite{11}, and joint super-resolution and deblurring \cite{16}. The major contribution of our framework is three-fold: (i) to the best of our knowledge, our method is among the first to achieve both structured and unstructured noise removal in a self-learning manner while preserving geometrical details in an image, where no temporal or motion information among successive images is required; (ii) our automatic MCA-based signal decomposition framework is adapted to several denoising applications in multimedia; and (iii) the learning of our framework for decomposing noises from an image is fully automatic and self-contained, where no extra training samples are required in the dictionary learning stage. For future works, we will study more real applications of our signal decomposition framework as well as efficient implementation strategy of our method.

REFERENCES


