SAR Analysis for Real Video Super-Resolution Improvement

SHI LI\textsuperscript{1,3}, XIAODIAO CHEN\textsuperscript{2}, STEVEN ZHIYING ZHOU\textsuperscript{3}, WANBIN PAN\textsuperscript{1}, AND YIGANG WANG\textsuperscript{1}\textsuperscript{*}

\textsuperscript{1}School of Media and Design, Hangzhou Dianzi University, Hangzhou 310018, China
\textsuperscript{2}School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China
\textsuperscript{3}Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583

Corresponding author: Yigang Wang (yigang.wang@hdu.edu.cn)

This work was supported in part by the Zhejiang Provincial Science and Technology Program in China under Grant 2018C01030, and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LY19F020041.

\textbf{ABSTRACT} This paper presents a simultaneous autoregressive (SAR) analysis method to describe the unknown signal-to-noise ratio (SNR) and the texture feature of low-quality real video frames when the ground-truth images are not available. The real video images degraded by the factors, such as electronic noise, oversaturated pixels, motion blur, and compression artifacts, often result in poor motion registration estimation, which makes the performance of the existing video super-resolution (VSR) algorithms lower than expected. It is hard to estimate the SNR of the low-quality real frames without any prior knowledge. To solve this problem, we made a connection between SAR hyperparameters and the SNR of real images. The relationship expression of them was given in this paper. Using the proposed method, well-registered low-quality real video frames can be selected to decrease the root mean squared error (RMSE) of motion estimation of video frames for VSR reconstruction improvement. The anomalous low-quality frame images whose SAR hyperparameters values are inconsistent with others will be considered for removal. Synthetic experiments were designed to illustrate how the SAR hyperparameters values vary with the variation of synthetic parameters. In order to better illustrate the effectiveness of the proposed method, real low-quality videos captured under different conditions were tested under VSR reconstruction experiments. The VSR reconstruction results show that the results obtained using SAR prior analysis have sharper edges and fewer ringing artifacts than the original results. It indicates that the proposed method is helpful to obtain better results of motion registration estimation for low-quality real video images.

\textbf{INDEX TERMS} Simultaneous autoregressive, video super-resolution, motion registration, Bayesian methods, signal-to-noise ratio.

I. INTRODUCTION

Video super-resolution (VSR) processing is often required by medical imaging (such as Endoscopic imaging, Ultrasound Imaging), surveillance imaging, mobile phone imaging, and other real video applications. VSR methods utilize the non-redundant information contained in the low-resolution (LR) images to reconstruct the high-resolution (HR) image \cite{1,2}. Before HR image reconstruction, the LR images usually need to be aligned/registerd in a unified coordinate system. This registration operation needs to estimate the motion between the target frame (reference frame) and the neighboring frames (supporting frames). It is very hard to accurately estimate motion information only by LR observation, especially in the case of high noise disturbance \cite{15,16,17}. It makes unavoidable estimation errors which will be propagated and amplified during HR image reconstruction and produce artifacts around image structures.

To prevent error-propagation, Bayesian framework was proposed to model the unknown HR image, the acquisition process, the motion parameters and the unknown model parameters in a stochastic sense \cite{1,2}. Many works \cite{1,2,8}, including our previous work \cite{8}, show that Bayesian-based super-resolution (SR) algorithms are fully-automated and do not require parameter tuning. However, when the input image sequence is seriously degraded, such as some surveillance videos, registration estimation problems still exist.

VSR algorithms based on deep learning also rely heavily on the accuracy of motion estimation \cite{9,10,11,12,13,14,15}. One of the difficulties of the learning-based algorithm is to obtain the...
ground-truth images and the necessary training data. The simulated data makes their actual performance on real images often overestimated [18]. The reason is that real images are more difficult to estimate motion than simulated data. And the low-quality real images are also more difficult to estimate motion than high-quality real images.

Sufficiently high-quality motion registration estimation is still a vital issue. For low-quality real video frames, it is very hard to obtain a ground-truth image as a reference image. The quality of the supporting images also affects the accuracy of the registration estimation. How to select a suitable reference frame and a series of acceptable supporting frames among the low-quality LR images is a challenging problem. To solve this problem, this paper presents a SAR analysis method that can evaluate the quality of each LR video frame for frame selection to improve motion registration estimation.

The rest of this paper is organized as follows: Section II introduces the SAR prior model and the SNR descriptor. Section III presents the image description performance of SAR hyperparameters and the SAR analysis method, including frame-interval detection and suitable frames selection. The experimental results consist of synthetic experiments and real experiments are demonstrated in Section IV. Finally, the conclusions are drawn in Section V.

II. SAR IN VARIATIONAL BAYESIAN
The quality of HR image estimation and the accuracy of other unknown estimation depend on the accuracy of the image model, such as the TV prior model [1], SAR prior model [2], [4], and other prior models [3]–[7]. In this paper, we use SAR prior to describe the SNR and texture feature of frame images. In the following sections, we briefly describe the individual distributions used to model the unknowns under the Bayesian framework. And the SNR descriptor is presented in section II-D.

A. SAR MODEL
SAR statistical model [2], [4] is a quadratic model (\(\ell^2\) norm) that imposes image smoothness especially the border structures of the images and attenuates noise. SAR model just autoregressive model itself that makes it simpler and more suitable to describe the characteristics of noise video frames and reflect the differences between video frames [2], [4]. In this paper, we adopt the matrix-vector notation such that and reflect the differences between video frames [2], [4].

Suitable to describe the characteristics of noise video frames is assumed to be zero-mean white Gaussian noise.

B. BAYESIAN FRAMEWORK
According to the Bayesian inference, the distribution of the HR image \(x\) is known as a multivariate Gaussian distribution

\[
q(x) = N(x|\langle x \rangle, \Lambda_x)
\]

The mean value of \(x\) is expressed as

\[
\langle x \rangle = \Lambda_x \left[ \sum_k \langle \beta_k \rangle B_k (s_k) y_k \right]
\]

and the inverse covariance is

\[
\Lambda_x^{-1} = \sum_k \langle \beta_k \rangle B_k (s_k)^T B_k (s_k) + \langle \alpha_{im} \rangle \mathcal{L}^T \mathcal{L}
\]

The distributions of the hyperparameters \(q(\alpha_{im})\) and \(q(\beta_k)\) are found as Gamma distributions given by

\[
q(\alpha_{im}) \propto \alpha_{im}^{N/2-1+\alpha_{im}^0} \exp \left[ -\alpha_{im} b_{im}^0 + E_{\alpha_{im}} ||| \mathcal{L} x |||^2 \right]/2
\]

and

\[
q(\beta_k) \propto \beta_k^{N/2-1+\beta_k^0} \exp \left[ -\beta_k b_{\beta_k}^0 + E_{\beta_k} ||| y_k - B_k (s_k) x |||^2 \right]/2
\]

where \(\alpha_{im}^0\) and \(\beta_k^0\) are the parameters of Gamma distribution, the ‘0’ (zero) indicates their initial estimation values. The details can be found in [2].

C. HYPERPARAMETERS
To analysis the characteristics of video frames, the SAR model is independently applied to each frame based on the variational Bayesian framework. It means every single frame will be processed by (3)(4)(5)(6) alone. Since a single image does not need motion registration and downsampling matrix is a constant matrix, the matrix \(B_k (s_k)\) can be simplified as blurring matrix \(H_k\). Hence, (3) is rewritten as

\[
\hat{f} = \langle x_{yk} \rangle = \Lambda_{yk} \left[ \langle \beta_k \rangle H_k^T y_k \right]
\]

and inverse covariance (4) is rewritten by

\[
\Lambda_{yk}^{-1} = \langle \beta_k \rangle H_k^T H_k + \langle \alpha_k \rangle \mathcal{L}^T \mathcal{L}
\]

The means of the distributions in (5) and (6) can be calculated by

\[
\langle \alpha_k \rangle = \frac{PN}{E_x[||| \mathcal{L} f |||^2] + 2b_{\alpha}}
\]

\[
\langle \beta_k \rangle = \frac{PN}{E_x[||| y_k - H_k f |||^2] + 2b_{\beta}}
\]
where \( b_\alpha = \| \Lambda y_k [L^T L] \|^2 \), \( b_\beta = \| \Lambda y_k [H_k^T H_k] \|^2 \). The initial estimation values of (9) and (10) are

\[
\langle a_k^0 \rangle = \frac{PN}{E_x[||y_k - H_k f||^2]} \quad \langle b_k^0 \rangle = \frac{PN}{E_x[||L f||^2 + 2b_\alpha]}
\]  

(11)

(12)

The Calculation of \( a_k \) and \( b_k \) is summarized in Algorithm 1.

**Algorithm 1** Hyperparameters Calculation

1: Calculate the initial \( a_k^0 \) and \( b_k^0 \) of the kth frame using (11) and (12)

2: while convergence criterion is not met do

3: Compute the kth frame distribution using (7) and (8)

4: Compute the distributions of the hyperparameters \( a_k, b_k \) using (9) and (10)

Based on empirical study, the convergence criterion is set as

\[
\frac{\| f^n - f^{n-1} \|^2}{\| f^{n-1} \|^2} < 10^{-6}
\]

**D. SNR DESCRIPTOR**

According to (9) and (10), we can obtain that

\[
\frac{\langle a_k \rangle}{\langle b_k \rangle} = \frac{E_x[||y_k - H_k f||^2]}{E_x[||L f||^2 + 2b_\alpha]}
\]  

(13)

To further illustrate the meaning of (13), the formula derivation is following:

Using (8) in (7) to obtain

\[
\hat{f} = (H_k^T H_k + \langle a_k \rangle \langle b_k \rangle L^T L)^{-1} H_k^T y_k
\]  

(14)

Define \( F \) is the Fourier transform matrix, identity matrix \( I = F^{-1} F \), the Fourier transform of (14) can be expressed as

\[
F \hat{f} = F (H_k^T H_k + \langle a_k \rangle \langle b_k \rangle L^T L)^{-1} (F^{-1} F) H_k^T (F^{-1} F) y_k
\]

\[
= \left( F H_k^T H_k F^{-1} + \langle a_k \rangle \langle b_k \rangle F L^T L F^{-1} \right)^{-1} F H_k^T F^{-1} F y_k
\]

\[
+ \langle a_k \rangle \langle b_k \rangle F L^T L F^{-1} \left( F H_k^T F^{-1} \right) F y_k
\]

(15)

Since \( H_k \) is real circulant matrix and Fourier transform is an orthonormal basis, \( F H_k F^{-1} = \text{diag}(\gamma) \) and \( F L^T L \) is diagonal, where \( \gamma \) is the Fourier transform of \( H_k \). Letting \( \gamma = F L^T L F^{-1} \). Actually, \( \gamma \) is the power spectral density of Laplacian operator. So \( \gamma \) can be considered as a constant matrix when the size of the matrix is confirmed. (15) can be rewritten as

\[
F \hat{f} = \left( \gamma^* \gamma + \langle a_k \rangle \langle b_k \rangle \gamma \right)^{-1} \gamma^* \gamma (F y_k)
\]  

(16)

The part of \( \left( \gamma^* \gamma + \langle a_k \rangle \langle b_k \rangle \gamma \right)^{-1} \gamma^* \gamma \) is exactly like the Wiener filter \( W \),

\[
W = \left( \gamma^* \gamma + \frac{S_n}{S_y} \gamma \right)^{-1} \gamma^* \gamma
\]

(17)

where \( S_n \) and \( S_y \) represent the power spectral density of the noise \( n \) and LR image \( y_k \), respectively.

We know that Wiener filter \( W \) is the optimal linear filter which minimizes the expected error \( E_x[||y_k - H_k f||^2] \) to obtain an optimal estimated \( \hat{f} \). Algorithm 1 process also produces an estimated image \( \hat{f} \). Comparing the terms in (16) and (17), we can confirm that the following equation is true

\[
\frac{S_n}{S_y} \sim \frac{\langle a_k \rangle}{\langle b_k \rangle} \gamma
\]

(18)

As mentioned above, \( \gamma \) is a constant matrix, (18) can be simplified to

\[
\frac{S_n}{S_y} \propto \frac{\langle a_k \rangle}{\langle b_k \rangle}
\]

(19)

Therefore, the SNR of the kth frame whose ground-truth image is unavailable can be represented by

\[
\frac{S_y}{S_n} \propto \frac{\langle b_k \rangle}{\langle a_k \rangle}
\]

(20)

**III. SAR ANALYSIS**

**A. IMAGE DESCRIPTION**

Based on the content correlation between video frames, the change in adjacent frame image should be tiny, and the \( \alpha_k \) curve of video frames should be smooth without muting. When the value of \( \alpha_k \) changes dramatically and frequently, it indicates that the content of the frame images has been greatly degraded by worse noise, obviously motion or blur, which is inherently unfavorable to frame image registration.

In this subsection, we present the image description performance of SAR hyperparameters on the real video image sequence. If not specially noted, all the real sequences with unknown motion information used in this paper are from publicly available datasets provided by UCSC [19], and all the SAR hyperparameters are obtained by Algorithm 1.

1) **TEXTURE DESCRIPTION**

The single value of the hyperparameter \( \alpha_k \) may not be able to accurately describe the content of the images, but the difference in the \( \alpha_k \) values between the adjacent frames can reflect the change of the adjacent frames.

Equation (9) indicates that the frame image with more details will obtain a smaller \( \alpha_k \) value. Fig.1 describes the hyperparameters values of the 26 frames in the sequence named Disk. The 4 frames, which are designated by red color in Fig. 1, are shown in Fig.2. The result proves that when the difference in \( \alpha_k \) values between two adjacent frame images is large, the corresponding image content usually changes greatly, while the image with smaller \( \alpha_k \) value has more details.
S. Li et al.: SAR Analysis for Real VSR Improvement

FIGURE 1. The $\alpha_k$ and $\beta_k$ values of the 26 frame images in the Disk sequence. The 4 frames, represented by the red circles, are also shown in Fig. 2.

FIGURE 2. Some frame images in Disk sequence. (a) is the 1st frame, $\alpha_1$ is 204.479; (b) is the 2nd frame, $\alpha_2$ is 111.717; (c) is the 20th frame, $\alpha_{20}$ is 93.7198; (d) is the 21st frame, $\alpha_{21}$ is 217.004; The content of the adjacent frames is significantly different.

FIGURE 3. The $\alpha_k$ and $\beta_k$ values of the 40 frames in the Surveillance sequence. The 4 frames marked in full red are also shown in Fig. 4.

FIGURE 4. Some frame images in Surveillance sequence. (a) is the 33rd frame, $\alpha_{33}$ is 75.2573; (b) is the 34th frame, $\alpha_{34}$ is 64.3823; (c) is the 35th frame, $\alpha_{35}$ is 72.766; (d) is the 36th frame, $\alpha_{36}$ is 62.996; there are obviously movement between them.

FIGURE 5. The $\alpha_k$ and $\beta_k$ values of the 80 frames in the Emily sequence. The frames from 61 to 69 are shown in Fig. 6.

FIGURE 6. Some frame images in Emily sequence. (a) is the 33rd frame, $\alpha_{33}$ is 75.2573; (b) is the 34th frame, $\alpha_{34}$ is 64.3823; (c) is the 35th frame, $\alpha_{35}$ is 72.766; (d) is the 36th frame, $\alpha_{36}$ is 62.996; there are obviously movement between them.

2) MOTION DESCRIPTION
The motion between adjacent frame images also influences $\alpha_k$ values. The SAR hyperparameters values corresponding to the 40 frames in the Surveillance sequence are plotted in Fig. 3. Fig. 4 shows 4 adjacent frames marked by red circles filled with red in the graph of Fig. 3. The rectangles marked with the same color in Fig. 4 have the same position and size. By comparing the patterns of the same position on the four adjacent frames, we can see that the four adjacent frames have a displacement offset from each other. And the $\alpha_k$ values corresponding to the 4 frames in the graph of Fig. 4 show sharp fluctuations. This indicates that this video has experienced severe vibrations during this time.

In Fig. 3 case, 6 pairs of $\alpha_k$ have the same values which are designated by red circles filled with green in Fig. 3. (For example, $\alpha_4 = \alpha_5 = 66.0437$, $\alpha_{17} = \alpha_{18} = 65.0155$, and so on.). The adjacent frames having the same $\alpha_k$ value cannot be distinguished from each other.

The hyperparameters values of the 80 frames in the Emily sequence are drawn in Fig. 5. It shows that the average of $\alpha_k$ values after frame 64 is larger than the average of $\alpha_k$ values before frame 64. The corresponding frame images are shown in Fig. 6. And it can be found that the girl in the sequence keeps still until frame 64, after which she moved and caused local movement.

The experimental results in this subsection show that when the $\alpha_k$ values of adjacent frames change greatly, some visible changes will appear in the frame images. The experiment details will be presented in section IV.
B. SAR Prior Analysis

Hyperparameter $\beta_k$ is a description of $E_k[||y_k - H_k f||^2]$ as shown in (10). Its value reveals the level at which $y_k$ is disturbed by noise and other degradation factors. In the frame-interval where the interference, such as noise, tends to be stable, we consider that the curve formed by the $\beta_k$ values corresponding to the $y_k$ frames in the interval should be smooth and moderate.

To achieve a better image registration from the low-quality video, we need to discriminate LR images according to the $\beta_k$ value of each frame image to keep the interference level of each frame image participating in the registration as consistent as possible. In the following subsections, we will introduce a method to obtain the suitable frame-interval and the reference image.

1) Frame-Interval Detection

The curve of $\beta_k$ is defined as $B(k)$, where $k = 1, 2, \ldots, L$. The difference is given by $\Delta B(k) = B(k + 1) - B(k)$. The threshold is expressed as $T = \mu(\text{Max} - \text{Min})/L$, where $L$ is the length of the sequence, $\text{Max}$ and $\text{Min}$ are the maximum and minimum value of $B(k)$, respectively, and $\mu$ is a coefficient, usually equal to 1.

We define a sign description function

$$\text{Sign}(k) = \begin{cases} 1, & \text{if } T < \Delta B(k) \\ 0, & \text{if } -T < \Delta B(k) < T \\ -1, & \text{if } \Delta B(k) < -T \end{cases}$$  \hfill (21)

The point function is determined by the following calculation

$$\text{Point}(k) = \begin{cases} 1, & \text{if } \text{Sign}(k) \times \text{Sign}(k + 1) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$  \hfill (22)

The interval boundary points are the points with value $\text{Point}(k) = 1$. Every two boundary points form an interval. All the points in the same interval have the same value $\text{Point}(k) = 0$.

2) Suitable Frames Selection

The number of frames in each interval is signed as $n(t)$, where symbol $t$ is the ordinal number of the interval. $\bar{n}$ is the average frame number of each interval. Let $\emptyset = \{t|n(t) > \bar{n}\}$. All the intervals satisfy the condition $n(t) > \bar{n}$, will be calculated the expectation of $\Delta B(k)$ represented by

$$E_t[|\Delta B(k)|] = E[|\Delta B(k)|], \quad t \in \emptyset \text{ and } k \in t^{th} \text{ interval}$$  \hfill (23)

The expected interval $\hat{t}$ is the one generates the minimum $E_t[|\Delta B(k)|]$ which means the values of $\beta_k$ in the interval $\hat{t}$ tend to be stable. In other words, the frames in the interval $\hat{t}$ are the wanted supporting frames which help to improve the registration accuracy.

$$\hat{t} = \arg\min_t (E_t[|\Delta B(k)|]), \quad t \in \emptyset$$  \hfill (24)

The selection of reference frame from the interval $\hat{t}$ is another important factor affecting the registration accuracy of low-quality LR images. To ensure the registration accuracy, we select the frame with the best SNR performance as the reference frame in the interval $\hat{t}$ according to the SNR analysis mentioned in section II-D. The expression is as follows:

$$\hat{k} = \arg\max_k \left(\langle\beta_k\rangle / \langle\alpha_k\rangle\right), \quad k \in \hat{t} \text{ interval}$$  \hfill (25)

The frame-selection flow diagram is shown in Fig.7.
IV. EXPERIMENTS
Due to the ground-truth images are unavailable, it is difficult to evaluate the accuracy of motion estimation between low-quality real images. As known that VSR results rely heavily on the accuracy of motion estimation. In this section, we used the improvements of the HR reconstruction results to demonstrate the effectiveness of our proposed method.

We used the following methods for comparison: 1) Bicubic interpolation, 2) Variational Bayesian SR [1], [20] (denoted by VBSR), and 3) sub-pixel motion compensation SR method [14] (denoted by SPMC), which is a deep-learning based SR algorithm. We also experimented with Shift-and-Add method [21] contained in the MDSP SR software [22] (denoted by S&A). It should be noted that SPMC did not provide training code, we directly used their training results here, and the performance of SPMC in the experiments did not represent its best performance.

The corresponding improved algorithms equipped with proposed method are called iBicubic, iS&A, and iSPMC, respectively. VBSR, S&A, and SPMC all are state-of-the-art SR algorithms. Since VBSR has been evaluated as one of the best algorithms which are fully automatic, we will focus on the performance of the iVBSR in the four improved algorithms.

A. SYNTHETIC EXPERIMENTS
In the following subsections, two synthetic experiments were demonstrated to qualitatively and quantitatively evaluate the performance of the proposed method,

1) HYPERPARAMETER ILLUSTRATION
In this subsection, we illustrate how the SAR hyperparameters values vary with the variation of synthetic parameters, such as SNR level and blur kernel size. We generated two groups of synthetic images from “Lena” image. The images in group 1 are degraded by additive white Gaussian noise at SNR levels of 10 dB, 15 dB, 25 dB, 35 dB, . . ., 75 dB. The images in group 2 are uniformly blurred. The kernel sizes are (1, 2, 3, 4, 5, 6, 7, 8, 9, 10). Some example synthetic images are shown in Fig. 8.

The SAR hyperparameters values of group 1 images are plotted in Fig.9. As expected, the estimated SNR(β/α) value increases as the simulated SNR level increases. The growth rate of SNR(β/α) is particularly obvious, when the SNR level is below 30dB. According to (9), the more high-frequency information is contained, the smaller αk value. The curve of the αk value, in Fig.9, well describes the high-frequency information of each frame image. When the image is severely degraded (SNR level is below 20 dB), the estimated αk value drops significantly. When the image simulated SNR level is higher than 30 dB, the estimated αk value tends to be steady.

The estimated results of group 2 images are plotted in Fig.10. It should be emphasized that deblurring for each frame is beyond the scope of this paper. The blurring matrix \( H_k \) of each frame, during the estimation procedure, is fixed as an identity matrix. It means each estimated image \( \hat{f} \) is treated as a blurred image. As the simulated blur-kernel size increases, the corresponding image is degraded worse, the estimated SNR(β/α) value decreases, and the estimated αk value increases. According to (10), the performance of the estimated βk value in Fig.10 and Fig.9 reveals that the image with less high-frequency information will obtain a larger estimated βk value.

FIGURE 8. Example synthetic images. (a) SNR = 10dB (b) SNR = 20dB (c) SNR = 30dB (d) kernel size = 1 (e) kernel size = 3 (f) kernel size = 5.

FIGURE 9. SAR hyperparameters values of the frame images in Group 1.

FIGURE 10. SAR hyperparameters values of the frame images in Group 2.
The curve of SNR ($\beta/\alpha$) in Fig. 10 implies that when the error of $H_k$ increases, the growth rate of $\beta_k$ is lower than $\alpha_k$, which indicate that $\beta_k$ is more stable than $\alpha_k$. The results of the experiment indicate that the $\alpha_k$ parameter can describe the high-frequency information of the image. However, high-frequency information may be image details or noise. A single $\alpha_k$ value is not sufficient to accurately describe the content of the image.

2) FRAME SELECTION EXPERIMENTS
The purpose of this synthetic experiment is to illustrate the importance of the reference-image-selection in VSR. In this section, we generated 10 synthetic LR images from the original HR image shown on the left in Fig. 11 through warping, blurring, and downsampling by a factor of 2. The warping consists of both translation and rotation, where the translations are chosen as
\[
\begin{pmatrix}
0.0 & 0.0 \\
0.0 & 0.5 \\
0.5 & 1.0 \\
1.0 & 0.0 \\
1.0 & 0.5 \\
0.0 & 1.0
\end{pmatrix},
\]
and the rotation angles are ($0^\circ$, $\pm1^\circ$, $\pm2^\circ$, $\pm3^\circ$, $\pm5^\circ$), respectively. As the blur, we applied a $3 \times 3$ uniform PSF on the first image. The LR images obtained after the warping, blurring, and downsampling operations are further degraded by additive white Gaussian noise at SNR levels of 5 dB, 15 dB, 25 dB, 35 dB, and 45 dB. Example LR images corresponding to the 5 dB SNR case are shown on the right in Fig. 11. Note that this resolution chart image is utilized for better illustration of the performance in resolution enhancement.

In each sequence, the first image with uniform blur is chosen as the reference image, and the standard Lucas-Kanade method [23] is utilized to iteratively estimate the motion parameters of the synthetic images. The estimated motion parameters of the synthetic images at each SNR level are shown in Tables 1-3. From Tables 1-3, it can be seen that the accuracy of motion estimation is significantly affected by noise and blur. The estimation errors in 5 dB SNR case and $\pm5^\circ$ rotation Angle case are particularly serious.

In order to reveal the influence of the reference-frame-selection on the SR results, we made 3 combinations of the synthetic images. Group 1 are the synthetic images with rotation angles ($0^\circ$, $\pm1^\circ$, $\pm2^\circ$, $\pm3^\circ$). Group 2 are the synthetic images with rotation angles ($0^\circ$, $\pm2^\circ$, $\pm3^\circ$, $\pm5^\circ$). Group 3 are the synthetic images with rotation angles ($0^\circ$, $\pm3^\circ$, $\pm5^\circ$). All of them include SNR levels of 5 dB, 15 dB, 25 dB, 35 dB, and 45 dB. And the 5 dB synthetic image with uniform blur that has the worst quality is confirmed as a reference image for all sequences. The synthetic images with $\pm5^\circ$ rotation angles play as disturbance images in this experiment. It should be emphasized that all sequences in the experiment including an ideal reference frame image which is only degraded by noise. Theoretically, the SR reconstruction results should be that Group 1 is the best, and Group 3 is the worst.

The peak signal-to-noise ratio (PSNR) values provided by the algorithms are shown in Table 4. The PSNR values of the Bicubic are obtained from the interpolation of the ideal reference frame at each SNR level. For the reason of poor motion estimation, S&A and SPMC methods result in worse reconstructions than Bicubic interpolation. VBSR achieves good reconstruction results in Group 1 and 2. But in Group 3 case, VBSR also obtained worse reconstructions than Bicubic interpolation for the reason of poor motion estimation (it can be seen in Tables 1-3).

In the “iVBSR” column of Table 4, the PSNR values are the results whose reference frames are the ideal reference frames. As expected, it shows that the SR reconstruction results in Group 1 are better than Group 2 and 3, and Group 3 is the worst. The PSNR values in the “Real” column are provided

| Table 1. The estimated rotation parameters of the synthetic images at each SNR level. The first row is the standard value. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| $\theta$ | $0^\circ$ | $1^\circ$ | $-1^\circ$ | $2^\circ$ | $-2^\circ$ | $3^\circ$ | $-3^\circ$ | $5^\circ$ | $-5^\circ$ |
| 5dB | 0.045 | 0.403 | -0.442 | 0.711 | -2.328 | 2.484 | -3.142 | 0.076 | -0.089 |
| 15dB | 0.016 | 1.025 | -1.269 | 1.959 | -2.024 | 2.636 | -2.952 | 0.104 | -0.074 |
| 25dB | -0.141 | 1.097 | -1.303 | 2.330 | -2.457 | 2.724 | -3.283 | 0.514 | -0.572 |
| 35dB | -0.170 | 0.970 | -1.323 | 2.065 | -2.441 | 2.704 | -3.295 | 0.520 | -0.299 |
| 45dB | -0.188 | 0.967 | -1.301 | 2.059 | -2.442 | 2.652 | -3.316 | 0.458 | 1.116 |

| Table 2. The estimated displacement x of the synthetic images at each SNR level. The first row is the standard value. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| $x$ | $0.0$ | $0.0$ | $0.5$ | $1.0$ | $0.0$ | $0.5$ | $1.0$ | $0.0$ | $0.5$ | $1.0$ |
| 5dB | 0.025 | 0.692 | 0.758 | 1.087 | 1.292 | 1.759 | 0.959 | 0.771 | 1.441 |
| 15dB | 0.065 | 0.394 | 0.697 | 0.974 | -0.341 | 0.775 | 0.810 | 1.171 | 2.494 |
| 25dB | 0.174 | 0.242 | 0.688 | 1.196 | 0.072 | 0.679 | 1.076 | 1.689 | 4.230 |
| 35dB | 0.168 | 0.171 | 0.696 | 1.198 | 0.073 | 0.679 | 1.079 | 1.695 | 4.119 |
| 45dB | 0.177 | 0.183 | 0.689 | 1.213 | 0.077 | 0.747 | 1.083 | 1.700 | 4.323 |

| Table 3. The estimated displacement y of the synthetic images at each SNR level. The first row is the standard value. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| $y$ | $0.0$ | $0.0$ | $0.5$ | $1.0$ | $0.5$ | $0.0$ | $1.0$ |
| 5dB | 0.585 | 0.737 | 0.769 | 0.737 | 2.155 | 1.918 | 0.507 | 1.506 | 0.742 |
| 15dB | 0.221 | 0.684 | 0.344 | -0.341 | 1.009 | 0.966 | 0.588 | 2.665 | 1.008 |
| 25dB | 0.096 | 0.715 | 0.282 | 0.196 | 1.228 | 0.968 | 0.637 | 4.467 | 1.189 |
| 35dB | 0.095 | 0.729 | 0.256 | 0.168 | 1.221 | 0.964 | 0.624 | 4.459 | 0.835 |
| 45dB | 0.105 | 0.745 | 0.282 | 0.174 | 1.233 | 0.957 | 0.638 | 4.258 | 1.116 |
The PSNR values in “Ideal” column and “Real” column of Table 4 show that different reference frame selection in the image sequence will result in different HR reconstruction results. PSNR evaluation is a full-reference quality measurement for pixel-wise comparisons. If the coordinate of the reference image is not the default one, the estimated HR need to be warped back to fit the coordinate of the original image to obtain meaningful PSNR values. The PSNR values in the “Revised” column are the calibration of the values in the “Real” column. Compared to the values in the “Ideal” column and “Real” column, it can be found that our method fails to select the ideal reference image. However, the PSNR values in the “Revised” column still indicate that iVBSR method provides better performance at all noise levels than all original methods not equipped with the proposed method. This means that if the first frame image in the image sequence is badly degraded, our method is able to find a better reference image in the image sequence, although not necessarily the best image.

The hyperparameters values of the synthetic images at each SNR level are shown in Table 5. The image sequences are sorted by rotation Angle. The hyperparameters values of the synthetic images at each SNR level. The image sequences are sorted by rotation Angle.

The results of the HR reconstruction in Group 1 are shown in Fig. 12 for the 45dB case, and example HR reconstruction in Group 3 are shown in Fig. 13 for the 25dB case. In Fig. 12 case, VBSR and iVBSR provide better than other frames in the same sequence. For this reason, the proposed method missed the ideal default selection.

The reference-frame-selection strategy based on the SNR performance can effectively improve the motion estimation of the sequence frames. The take the case of 25dB in Group 1 for example. The eventual motion estimation errors of VBSR and iVBSR are shown in Table 6. In the VBSR method, the frame with $-1^\circ$ rotation is selected as the reference image because its $\beta/\alpha$ value is the largest in the sequence (refer to the values in Table 5). Hence, the standard value of the warping parameters in Table 6 is different. The comparison results in Table 6 show that the motion estimation errors in iVBSR are smaller than VBSR’s if ignore the motion estimation errors of their 1st frame which is degraded badly. This is also proved by the RMSE values of the estimation errors in Table 6.
FIGURE 12. Example estimated HR images from different SR methods in the Group 1 case when SNR = 45 dB and motion information is inaccurate. Results of (a) Bicubic interpolation (PSNR = 16.252dB), (b) VBSR (PSNR = 22.269dB), (c) S&A (PSNR = 8.406dB), (d) SPMC (PSNR = 11.439dB), (e) Ideal result of iVBSR (PSNR = 47.265dB), (f) iVBSR (PSNR = 15.688dB), and (g) Revised result of (f) (PSNR = 27.772dB).

FIGURE 13. Example estimated HR images from different SR methods in the Group 3 case when SNR = 25 dB and motion information is inaccurate. Results of (a) Bicubic interpolation (PSNR = 16.220dB), (b) VBSR (PSNR = 10.347dB), (c) S&A (PSNR = 8.216dB), (d) SPMC (PSNR = 11.441dB), (e) Ideal result of iVBSR (PSNR = 18.619dB), (f) iVBSR (PSNR = 10.312dB), and (g) Revised result of (f) (PSNR = 19.482dB).

visually enhanced restorations and much sharper edges compared to other methods. Compared with the original VBSR, iVBSR which uses reference-frame-selection can significantly reduce the ringing artifacts. In Fig.13, the performance of VBSR is worse than Bicubic interpolation, but the performance of iVBSR is still better. The HR constructions with the default ideal reference frame (Fig.13 (e)) and the selected reference frame (Fig.13 (f)) are very similar. After the coordination calibration, the PSNR value of the latter (Fig.13 (g)) is higher than the former (Fig.13 (e)).

B. REAL EXPERIMENTS

In order to evaluate the performance of the proposed method more comprehensively, DUF [10] and RBPN [24], whose deep-learning network can allow them to avoid explicit motion estimation and compensation, are added as new comparative algorithms in this subsection. And all the input sequences are the real videos provided by UCSC [19]. It should be noted that we just directly ran the code, provided by the authors of DUF and RBPN, to test the UCSC datasets. The performance of DUF and RBPN in the experiments did not represent their best performance.

1) SIMPLE TEXTURE VSR

Text sequence shows simple text pattern with global translational motion. It consists of 30 LR images. The $\alpha_k$ and $\beta_k$ values of the 30 frame images are drawn in Fig.14. The 3rd, 4th, 10th frame which are designated by red color in Fig. 14,
S. Li et al.: SAR Analysis for Real VSR Improvement

**FIGURE 14.** The $\alpha_k$ and $\beta_k$ values of the 30 frame images in the Text sequence. The 3 frames marked in red are shown in Fig.15. The frame-interval $[23, 30]$ marked in green is the selected stable frame-interval.

**FIGURE 15.** Sample images in Text sequence. (a) is the 3rd frame, $\alpha_3$ is 381.739; (b) is the 4th frame, $\alpha_4$ is 209.353; (c) is the 10th frame, $\alpha_{10}$ is 342.064.

**FIGURE 16.** The SNR($\beta/\alpha$) values of 30 frame images in the Text sequence. It can be found that the frame with smaller $\alpha_k$ value (such as frame 4) is better than other frames. The curve of $\beta_k$ in Fig. 14 shows that the values are drastically changing until into the frame-interval $[23, 30]$ which can be detected by the proposed method. The SNR($\beta/\alpha$) values of 30 frame images in the Text sequence are illustrated in Fig.16. As can be seen that the $\beta/\alpha$ value of frame 28 is the largest in the interval $[23, 30]$. And it is selected as the reference frame by the proposed method.

The reconstructed HR images obtained by Bicubic interpolation and SR algorithms are shown in Fig. 17. All SR algorithms based on explicit motion estimation have better reconstruction effect than Bicubic interpolation. Fig. 17 (e) and Fig. 17 (f) obtained by SPMC and iSPMC, respectively, have obviously artifacts. But Fig. 17 (f) obtained by iSPMC is a little better. The edges of Fig. 17 (d) and Fig. 17 (h) are sharper than Fig. 17 (c) and Fig. 17 (g), respectively. It reveals that the improved SR algorithms with the proposed method have better motion registration than the original SR algorithms. Fig. 17 (i) and Fig. 17 (j) show that DUF and RBPN are not very good at reconstructing low-quality real video frames. The specific evaluation results are shown in Table 7.

2) RICHER DETAILS VSR

In this subsection, the reconstruction experiment of Book sequence is presented to better illustrate the visual
TABLE 7. BRISQUE scores for the resulted images.

|   | Bicubic | iBicubic | S&A | ISPMC | VBSR | iVBSR | DUF | iDUF | SBPN |
|---|---------|----------|-----|-------|------|-------|-----|------|------|
| Text | 66.251 | 22.496 | 20.831 | 16.922 | 33.045 | 13.421 | 12.204 | 5.428 | 49.434 | 30.055 |
| Book | 45.159 | 48.984 | 42.387 | 40.966 | 47.617 | 45.072 | 75.211 | 40.473 | 55.479 | 51.892 |
| Tree class | 58.616 | 54.153 | 51.798 | 36.612 | 47.419 | 46.531 | 42.953 | 36.991 | 65.345 | 66.034 |
| Alpaca | 67.187 | 65.098 | 51.681 | 37.565 | 71.781 | 69.727 | 37.001 | 39.305 | 72.255 | 78.347 |
| Disk | 42.462 | 27.267 | 21.086 | 12.48 | 70.614 | 22.712 | 69.918 | 0.2975 | 47.998 | 37.858 |
| Adyeron | 42.912 | 40.947 | 96.941 | 21.458 | 52.556 | 51.163 | 99.481 | 38.323 | 68.213 | 61.837 |
| Face | 61.697 | 61.117 | 43.499 | 42.935 | 53.527 | 53.593 | 43.594 | 22.306 | 46.825 | 50.001 |
| Emily | 72.391 | 69.858 | 51.939 | 50.992 | 77.062 | 78.884 | 58.228 | 49.136 | 88.165 | 97.358 |

FIGURE 18. The $\alpha_k$ and $\beta_k$ values of the Book sequence 30 frame images. The frame-interval [6], [11] marked in green is detected as stable frame-interval by the proposed method.

FIGURE 19. The $SNR(\beta/\alpha)$ values of the Book sequence 30 frame images.

FIGURE 20. Super-resolution results of Book sequence (2x resolution increase) by (a) Bicubic interpolation, (b) iBicubic interpolation, (c) S&A, (d) iS&A, (e) SPML, (f) iSPML, (g) VBSR, (h) iVBSR, (i) DUF, (j) RBPN.

improvement introduced by the proposed method. The Book sequence has richer details than the Text sequence, and it is closer to normal video sequence. It also contains 30 LR images. The SAR hyperparameters values of Book sequence are plotted in Fig. 18. The frame-interval [6], [11] marked as green color is detected as stable frame-interval by the proposed method. The $SNR(\beta/\alpha)$ values are drawn in Fig. 19. Frame 9 with the largest $\beta/\alpha$ value in the interval [6], [11] is selected as the reference frame.

The HR construction results are shown in Fig. 20. Fig. 20 (g) obtained by VBSR is fettered by the registration problem and its edges are wreathed in noise and ringing artifacts. The most difference between VBSR and S&A is the times of motion estimation. VBSR iteratively estimate the motion parameters along with the iteration of HR image reconstruction. S&A method makes only one motion estimation and reconstructs HR image very quickly. Therefore, S&A is less interfered by the registration problem. SPML based on motion compensation is less affected by the registration estimation too. In this case, the quality of Book sequence is obvious better than that of Text sequence. Both of DUF and RBPN obtain better results. However, Fig. 20 (h) reconstructed by iVBSR has sharper edges and fewer artifacts.
than others. The specific evaluation results can be found in Table 7. It also shows that the results of iS&A and iSPMC are better than S&A and SPMC, respectively. The experiment results indicate that the proposed method contributes to the improvement of motion registration estimation.

3) COMPLICATE MOTION ANALYSIS

The Surveillance sequence in Fig. 4 has significant jitter, as captured by a handheld device. These frames did not only follow the global translational motion model, but also include a little pitching motion. By analyzing the $\beta_k$ values plotted in Fig. 3, we can obtain two obvious stable frame-intervals, [7], [13] and [17], [22], each of which can contribute a good SR reconstruction result. The proposed method considers the frame-interval [7], [13], which are designated with green circles in Fig. 21, as the better one because of its smaller $E[|\Delta B(k)|]$ value, and selects frame 7 as the reference frame, since its $\beta/\alpha$ value is the largest in the interval [7], [13] (see Fig. 22).

![FIGURE 21. The $\alpha_k$ and $\beta_k$ values of the 40 frame images in the surveillance sequence. The frame-interval [7], [13] marked in green is detected as stable frame-interval by the proposed method.](image)

![FIGURE 22. The SNR($\beta/\alpha$) values of the 40 frame images in the surveillance sequence.](image)

We try to combine the two mentioned frame-intervals, and obtain a result like Fig. 23 (g) with heavy ringing artifacts. It is noted that the registration model in VBSR and iVBSR defaults to 4 degree-of-freedom (DOF) motion registration. Obviously, 4 DOF registration model cannot accurately estimate the 6 DOF motion. Hence, we conclude that there may be different pitch motions between the two...
frame-intervals, which increases the motion DOF of the combined frame-interval from 4 to 6 or more.

The results in Fig. 23 show that iVBSR with the proposed method provides acceptable visually enhanced restoration with sharper edges and fewer ringing artifacts than the original VBSR. It indicated that our method can obtain local frame-intervals with stable motion in low-quality real video to decompose the motion complexity (such as reducing the complex 6 DOF motion to simple 4 DOF motion, which only includes translation and rotation). The results of iS&A and iSPMC are also better than S&A and SPMC, respectively. The results of DUF and RBPN have obvious ringing artifacts. The evaluation results can be found in Table 7.

![SAR values in Alpaca sequence](image)

**FIGURE 24.** The $\alpha_k$ and $\beta_k$ values of the Alpaca sequence 55 frame images.

![SNR($\beta/\alpha$) values in Alpaca sequence](image)

**FIGURE 25.** The SNR($\beta/\alpha$) values of the Alpaca sequence 55 frame images.

4) FRAME INTERVALS COMBINATION
This subsection presents Alpaca sequence VSR reconstruction results. The Alpaca sequence with heavy noise approximately follows the global translational motion model. The hyperparameters values of Alpaca are plotted in Fig. 24, which reveals the complicate content change in the sequence. The proposed method detects several stable frame-intervals, such as [15], [20], [22], [29]. It can be considered that the frame-interval [15], [29] is segmented into two frame-intervals by frame 21. We assume that the motion DOF in the two intervals are the same, then the two frame-intervals can be combined into one large frame-interval to keep more non-redundant image information. Under that assumption, frame 20 is selected as the reference frame of the merged frame-interval according to the $\beta/\alpha$ value analysis in Fig. 25.

![Super-resolution results of Alpaca sequence](image)

**FIGURE 26.** Super-resolution results of Alpaca sequence (2x resolution increase) by (a) Bicubic interpolation, (b) iBicubic interpolation, (c) S&A, (d) iS&A, (e) SPMC, (f) iSPMC, (g) VBSR, (h) iVBSR, (i) DUF, (j) RBPN.

The HR construction results of the merged frame-interval are shown in Fig. 26. The SR result of Alpaca sequence obtained by VBSR (see Fig. 26 (g)) is still full of ringing.
FIGURE 27. Other super-resolution results of UCSC datasets (2x resolution increase) by (a) Bicubic interpolation, (b) iBicubic interpolation, (c) S&A, (d) iS&A, (e) SPMC, (f) iSPMC, (g) VBSR, (h) iVBSR, (i) DUF, (j) RBPN.
FIGURE 27. (Continued.) Other super-resolution results of UCSC datasets (2x resolution increase) by (a) Bicubic interpolation, (b) iBicubic interpolation, (c) S&A, (d) iS&A, (e) SPMC, (f) iSPMC, (g) VBSR, (h) iVBSR, (i) DUF, (j) RBPN.
artifacts. The results of iS&A and iVBSR are all acceptable and better than S&A and VBSR respectively. Due to the heavy noise, learning based algorithms all produce obvious artifacts around image structures in this case. This experiment shows that when two frame-intervals are closely adjacent, they can be combined to obtain a stable large interval. In that situation, the proposed reference-selection-method still works well and mitigates the registration problem to obtain a better SR result.

C. OTHER REAL EXPERIMENTS

More HR reconstruction results are illustrated in Fig.27. It is obvious that the results of the improvement methods are sharper and clearer than the results of the original methods. Although the motion estimation improvement in the real sequences cannot be accurately estimated due to the absence of ground-truth images, the improved HR results with less ringing artifacts can indirectly indicate that the proposed method can improve the registration accuracy.

D. BRISQUE RESULTS

In order to objectively evaluate the aforementioned resulted images whose original HR image is not available, Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [25, 26] is selected to quantitatively measure the quality of the restored HR images. For no-reference quality assessment, BRISQUE shows strong correlations with human visual perception [18]. It scores from 0 to 100 (0 represents the best quality, 100 the worst).

The quality scores are reported in Table 7. In most cases, the performance of the improved method is better than the original SR methods. In the case of sequence filled with simple patterns, such as Text sequence and Disk sequence, the improvement of results is particularly noticeable.

No criteria are perfect, and neither is BRISQUE. For example, the quality scores of Alpaca in Table 7 show the VBSR result is the best one. In fact, the reconstructed image by VBSR (see Fig.26 (g)) is filled with ringing artifacts. But the vast majority of the quality scores can still match the image quality.

V. CONCLUSION

It is a challenging problem to estimate SNR information based only on the low-quality real video image itself. The frame with worse SNR value will often impair the accuracy of motion registration. And the estimation errors will greatly worsen the VSR results. In this paper, a SAR analysis method is proposed to describe the SNR of frame images. Based on the analysis, the frames in the original sequence will be filtered to improve the registration accuracy which will make the HR reconstruction of the original sequence better.

Based on the results of visual evaluation and BRISQUE evaluation, we can draw a conclusion that the proposed method significantly improves the accuracy of motion registration estimation for low-quality real video images. The state-of-the-art SR algorithms using proposed method can get better experimental results when dealing with low-quality real sequences. Our method is used as a preprocessing method for the SR algorithms. A simple frame-selection process gives them a greater advantage in terms of robustness and stability.

REFERENCES

[1] S. D. Babacan, R. Molina, and A. K. Katsaggelos, “Variational Bayesian super resolution,” IEEE Trans. Image Process., vol. 20, no. 4, pp. 984–999, Apr. 2011.

[2] S. Villena, M. Vega, S. D. Babacan, R. Molina, and A. K. Katsaggelos, “Bayesian combination of sparse and non-sparse priors in image super resolution,” Digit. Signal Process., vol. 23, no. 2, pp. 530–541, Mar. 2013.

[3] R. Jin, “Super-resolving barcode images with an edge-preserving variational Bayesian framework,” J. Electron. Imag., vol. 25, no. 3, 2016, Art. no. 033016.

[4] V. F. Campana, K. F. Cocco, E. O. T. Salles, and P. M. Ciarelli, “Modification in the SAR super-resolution model using the fractal descriptor LMME in the term regularizer,” IEEE Access, vol. 6, pp. 39046–39062, 2018.

[5] M. Sonogashira, T. Funatomi, M. Iyama, and M. Minoh, “Variational Bayesian approach to multiimage frame restoration,” IEEE Trans. Image Process., vol. 26, no. 5, pp. 2163–2178, May 2017.

[6] S. Zhao, H. Liang, and M. Sarem, “A generalized detail-preserving super-resolution method,” Signal Process., vol. 120, pp. 156–173, Mar. 2016.

[7] C. Chen, H. Liang, S. Zhao, Z. Lu, S. Fang, and X. Pei, “Integrating the missing information estimation into multi-frame super-resolution,” Circuits, Syst., Signal Process., vol. 35, no. 4, pp. 1213–1238, Apr. 2016.

[8] J. Chen, Y. Wang, and S. Li, “Fast Bayesian super resolution using Jacobi preconditioner and multithread,” in Proc. Int. Conf. Image Graph. Process. (ICIGP), Hong Kong, 2018, pp. 121–125.

[9] J. Caballero, C. Ledig, A. Aitken, A. Acosta, J. Totz, Z. Wang, and W. Shi, “Real-time video super-resolution with spatio-temporal networks and motion compensation,” in Proc. IEEE Comput. Conf. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA, Jul. 2017, pp. 2848–2857.

[10] Y. Jo, S. W. Oh, J. Kang, and S. J. Kim, “Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Salt Lake City, UT, USA, Jun. 2018, pp. 3224–3232.

[11] A. Kappeler, S. Yoo, Q. Dai, and A. K. Katsaggelos, “Video super-resolution with convolutional neural networks,” IEEE Trans. Comput. Imag., vol. 2, no. 2, pp. 109–122, Jun. 2016.

[12] D. Liu, Z. Wang, Y. Fan, X. Liu, Z. Wang, S. Chang, and T. Huang, “Robust video super-resolution with learned temporal dynamics,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Venice, Italy, Oct. 2017, pp. 2526–2534.

[13] W. Yang, J. Feng, G. Xie, J. Liu, Z. Guo, and S. Yan, “Video super-resolution based on spatial-temporal recurrent residual networks,” Comput. Vis. Image Understand., vol. 168, pp. 79–92, Mar. 2018.

[14] X. Tao, H. Gao, R. Liao, J. Wang, and J. Jia, “Detail-revealing deep video super-resolution,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Venice, Italy, Oct. 2017, pp. 4472–4480.

[15] H. Seibel, S. Goldenstein, and A. Rocha, “Eyes on the target: Super-resolution and license-plate recognition in low-quality surveillance videos,” IEEE Access, vol. 5, pp. 20020–20035, 2017.

[16] S. Huang, J. Sun, Y. Yang, Y. Fang, and P. Lin, “Multi-frame super-resolution reconstruction based on gradient vector flow hybrid field,” IEEE Access, vol. 5, pp. 21669–21683, 2017.

[17] I. E. Mourabit, M. E. Rhabi, A. Hakim, A. Laghrib, and E. Moreau, “A new denoising model for multi-frame super-resolution image reconstruction,” Signal Process., vol. 132, pp. 51–65, Mar. 2017.

[18] T. Köhler, M. Bätz, F. Naderi, A. Kaup, A. Maier, and C. Riess, “Toward bridging the simulated-to-real gap: Benchmarking super-resolution on real data,” IEEE Trans. Pattern Anal., Mach. Intell., to be published. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8716546.

[19] P. Milanfar, MDSR Super-Resolution and Demoisering Datasets. Accessed: Aug. 18, 2019. [Online]. Available: https://users.ucr.edu/~milanfar/software/st-datasets.html

[20] S. Villena. Super Resolution Software. Accessed: Aug. 18, 2019. [Online]. Available: http://decsai.ugr.es/pi/superresolution/software.html

[21] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, “Fast and robust multiframe super-resolution,” IEEE Trans. Image Process., vol. 13, no. 10, pp. 1327–1344, Oct. 2004.
[22] P. Milanfar. MDSP Resolution Enhancement Software. Accessed: Aug. 18, 2019. [Online]. Available: https://users.soe.ucsc.edu/~milanfar/software/superresolution.html

[23] B. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” in Proc. Imag. Understand. Workshop, 1981, pp. 121–130.

[24] M. Haris, G. Shakhnarovich, and N. Ukiti, “Recurrent back-projection network for video super-resolution,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Long Beach, CA, USA, Jun. 2019, pp. 3897–3906.

[25] A. Mittal, A. K. Moorthy, and A. C. Bovik. BRISQUE Software Release. Accessed: Aug. 18, 2019. [Online]. Available: http://live.ece.utexas.edu/research/quality/BRISQUE_release.zip

[26] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” IEEE Trans. Image Process., vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

SHI LI graduated in optical and information engineering from Zhejiang University, China, in 2003. He received the Ph.D. degree in optical engineering from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, in 2010. He has been a Lecturer with Hangzhou Dianzi University, since 2010. He is currently an Academic Visitor with the Department of Electrical and Computer Engineering, National University of Singapore. His current research interests include image restoration, super-resolution, and real-time image processing.

XIAODIAO CHEN received the master’s and Ph.D. degrees from the Department of Computer Science, Tsinghua University, in 2003 and 2006, respectively. He was a Professor with the School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China. His current research interests include computer-aided geometric design and computer graphics.

STEVEN ZHIYING ZHOU received the B.Eng. and M.Eng. degrees from Southeast University, Nanjing, China, in 1998 and 2001, respectively, and the Ph.D. degree from the National University of Singapore (NUS), Singapore. He is currently a Professor with the Department of Electrical and Computer Engineering, and the Director of the Interactive Multimedia Laboratory, NUS. His research interests include augmented and virtual reality, computer vision, and multimodal human–computer interaction.

WANBIN PAN received the master’s degree in computer science from Hangzhou Dianzi University, Hangzhou, China, in 2008, and the Ph.D. degree in computer science from Zhejiang University, China, in 2015. He is currently an Academic Visitor with the National University of Singapore. He is also an Associate Professor with the School of Media and Design, Hangzhou Dianzi University. His research interests include 3D CAD model retrieval and adaptation.

YIGANG WANG received the master’s and Ph.D. degrees in applied mathematics from Zhejiang University, China, in 1995 and 1998, respectively. He has been a Professor with Hangzhou Dianzi University, since 2004. He mainly works on the following topics, such as image processing, computer graphics, photorealistic rendering and realistic real-time rendering, and virtual reality. He is currently developing research in the areas of computer vision and augmented reality.