Spatiotemporal singular value decomposition for denoising in photoacoustic imaging with a low-energy excitation light source

MENGJIE SHI, TOM VERCAUTEREN, and WENFENG XIA* 

School of Biomedical Engineering & Imaging Sciences, King’s College London, St Thomas’ Hospital, London, SE1 7EH, UK

*wenfeng.xia@kcl.ac.uk

Abstract: Photoacoustic (PA) imaging is an emerging hybrid imaging modality that combines rich optical spectroscopic contrast and high ultrasonic resolution, and thus holds tremendous promise for a wide range of pre-clinical and clinical applications. Compact and affordable light sources such as light-emitting diodes (LEDs) and laser diodes (LDs) are promising alternatives to bulky and expensive solid-state laser systems that are commonly used as PA light sources. These could accelerate the clinical translation of PA technology. However, PA signals generated with these light sources are readily degraded by noise due to the low optical fluence, leading to decreased signal-to-noise ratio (SNR) in PA images. In this work, a spatiotemporal singular value decomposition (SVD) based PA denoising method was investigated for these light sources that usually have low fluence and high repetition rates. The proposed method leverages both spatial and temporal correlations between radiofrequency (RF) data frames. Validation was performed on simulations and in vivo PA data acquired from human fingers (2D) and forearm (3D) using a LED-based system. Spatiotemporal SVD greatly enhanced the PA signals of blood vessels corrupted by noise while preserving a high temporal resolution to slow motions, improving the SNR of in vivo PA images by 90.3%, 56.0%, and 187.4% compared to single frame-based wavelet denoising, averaging across 200 frames, and single frame without denoising, respectively. With a fast processing time of SVD (~50 µs per frame), the proposed method is well suited to PA imaging systems with low-energy excitation light sources for real-time in vivo applications.

Published by Optica Publishing Group under the terms of the Creative Commons Attribution 4.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

1. Introduction

Photoacoustic imaging is a promising hybrid biomedical imaging technique capable of spatially resolving spectroscopic contrast of tissue at high ultrasonic resolution and depths. It can provide structural, molecular, and functional information of biological tissues, and has shown great potential in various pre-clinical and clinical applications [1]. However, PA signals could be corrupted by electrical noise due to thermal effects, external hardware, and environmental interferences [2]. System-related noise (parasitic noise) is usually associated with solid-state laser systems with high-voltage Q-switching, while thermal noise is randomly distributed in the background.

Various algorithms have been proposed for PA signal denoising on single RF time series and image frame, including frequency filtering, wavelet transform, and signal decomposition. Frequency filtering using bandpass filters is straightforward to implement but is less efficient when noise and signal spectra significantly overlap [3]. Wavelet-based denoising can be effective depending on the appropriate selection of the mother wavelet, the number of decompositions, and the thresholding criteria [4–6]. Decomposition techniques have been used for PA denoising
by exploring the differences of the spatial and/or temporal characteristics of the signal and noise [7–11].

Inherent characteristics of PA signals of tissue can be exploited for PA denoising via different decomposition models. Empirical mode decomposition (EMD) could decompose PA signals into several intrinsic mode functions (IMFs), allowing for noise removal by filtering out noisy IMFs. But the denoising performance is limited by its dependency on selecting IMFs of noise [7–9]. SVD could also be used to remove noise and extract signals of tissue for PA imaging. Hill et al. [10] proposed an SVD-based PA denoising approach for reducing laser-induced noise in the signal domain. Haq et al. [11] investigated K-means singular value decomposition (K-SVD) to compute a sparse representation of PA images. The K-SVD based denoising method involved an iterative process for selecting singular value components (SVCs) corresponding to noise, requiring high computation cost and sources. SVD-based denoising techniques are promising for outperforming frequency filtering and wavelet-based methods but require proper subspace rank selection for the optimality of noise rejection.

Recently, laser diodes (LDs) and light-emitting diodes (LEDs) have shown promise as an alternative to high-power lasers owing to their compact size, safety, and low cost [12], and could be useful in several clinical applications such as the detection of joint inflammation and guidance of minimally invasive procedures [12–15]. Despite much lower optical fluence leading to poor SNRs, LDs and LEDs excitation sources have higher pulse repetition frequencies (several kHz) compared to conventional laser sources (10-100 Hz) that are commonly used for PA tomography, which provides an opportunity for denoising by leveraging the spatial and temporal features of the signals. Anas et al. [16] proposed a recurrent network and a convolutional neural network to exploit the spatial and temporal dependencies respectively. Although they reported a considerable improvement over signal averaging using in vivo data, the performance could be limited by the use of averaged signals as the ground truths for training that could be affected by motion artefacts. Wang et al. [17] proposed a spatiotemporal image reconstruction method for dynamic photoacoustic computed tomography and reported improved image quality and computational cost against conventional frame-by-frame reconstruction methods.

Spatiotemporal SVD has been used for ultrafast and super-resolution ultrasound (US) imaging for extracting blood clutters from tissue background with good sensitivity [18,19]. In US imaging, in contrast to tissue backscatter signals, slow blood flow signals have lower spatiotemporal coherence and power, so they could be separated in the singular value domain. A recent study by Al Mukaddim et al. [20] used spatiotemporal SVD in PA imaging to extract the dynamic cardiac signals during a cardiac cycle.

In this work, for the first time to our knowledge, we propose a spatiotemporal SVD for denoising in PA imaging dedicated to systems with low-fluence light sources. Different from [20], in this study, the background noise is considered dynamic (randomly distributed and having a low spatial and temporal coherence) while tissue signals are considered quasi-static (high spatial and temporal coherence). We validated our method on simulated data, and in vivo data acquired from a human volunteer using a LED-based system in comparison with averaging across multiple imaging frames, and wavelet denoising based on a single frame.

2. Materials and methods

2.1. Randomised spatiotemporal singular value decomposition (rsSVD)

Conventional implementation of the SVD-based clutter filtering for US imaging [18] requires the formulation of a spatiotemporal Casorati matrix S of dimension \((n_t \times n_x \times n_z)\) based on an US dataset with \(n_t\) frames of 2D data of size \((n_x, n_z)\). SVD is then performed on the matrix S as:

\[
S = U \Delta V^\ast
\]  

(1)
Where $U$ and $V$ correspond to the spatial and temporal singular vectors of $S$, respectively. Here, $\Delta$ is a diagonal matrix with diagonal entries listed in descending order. Signals with higher energy and higher spatiotemporal coherence will be captured mostly by the largest singular vectors, which can be filtered out by multiplying a filtering matrix $I^F$. This produces the filtered signals as:

$$S^F = U\Delta I^F V^*$$  (2)

Implementing the full SVD of the matrix $S$ is associated with high computational complexity $O(n_x \times n_z \times n_t^2)$, thereby leading to long computation times. Randomised SVD (rSVD) was proposed for accelerating the SVD process by low-rank matrix approximations [21,22] and could be implemented on a GPU with PyTorch. rSVD is suitable for low energy-based PA denoising since tissue signals with higher spatial and temporal coherence reside in the first few largest singular values. Therefore, instead of performing a full rank SVD, a truncated SVD (e.g., rSVD) is practically efficient. Song et al. [23] employed rSVD for accelerating spatiotemporal SVD-based clutter filtering with US data without substantially degrading the performance. In this work, we implemented rSVD for maximising the acceleration of spatiotemporal SVD for PA denoising. We called it randomised spatiotemporal singular value decomposition (rsSVD) for brevity and described the critical steps of the implementation as follows.

A 3D RF matrix $(n_x, n_z, n_t)$ was formed using stacks of RF data obtained for each pulse of LED excitation with a dimension of $(n_x, n_z)$ where $n_x$ and $n_z$ denotes the number of transducer elements and the number of time steps, respectively. $n_t$ corresponds to the number of frames in use. The 3D RF matrix was then reshaped to the Casorati matrix $S$, with a dimension of $(n_x \times n_z, n_t)$. Instead of calculating a full SVD of the matrix $S$, rsSVD generated a matrix $Q$ with rank $k$ as an approximation of the matrix $S$. We observed that the quasi-static tissue signals primarily corresponded to the largest $k$ rank singular values ($k \ll t$). To calculate $Q$, the matrix $S$ was first multiplied by a random matrix $R$.

$$S' = SR$$  (3)

Each entry of $R$ was generated from a Gaussian distribution $N(0, 1)$. The matrix $Q$ was obtained by performing the pivoted QR decomposition on the matrix $S'$.

$$Q = qr(S')$$  (4)

Power iteration [21] was used to accelerate the convergence of the singular values to the full SVD results. Thus, the rank $k$ matrix $S$ was approximated as

$$S \approx QQ^*S$$  (5)

where $Q^*$ represents complex conjugate transpose. The denoised PA data were obtained by the signal matrix $P$ as:

$$P = QQ^*S$$  (6)

2.2. Rank estimation based on singular vectors

Thresholding strategies for SVD-based clutter suppression methods have been investigated in US imaging [18,24–26]. Estimators based on the characteristics of temporal and spatial singular vectors proposed in [26] were proven efficient for clutter filtering with US images. For the estimator based on the temporal singular vectors, Fig. 1(a) shows the double-side power spectral density (PSD) of all the temporal vectors for an in vivo RF acquisition on human fingers (see Sec. 2.4 below for details on the data acquisition). For each temporal vector, the bandwidth containing 99% of the energy was computed at each side. Compared to noise signals, tissue signals had a narrower bandwidth which was associated with the temporal vectors of high correlation order.
A non-parametric estimator was therefore defined by the inflection point where the 99% bandwidth was expanded towards the entire frequency band. In contrast, the estimator based on the spatial singular vectors leveraged the spatial features within the tissue and noise subspaces. The intensity of the \( k \)th column of the spatial singular vectors \( U \) in Eq. (1) denoted by \( |u_k| \) was observed to be highly correlated only within the tissue subspace. The correlation matrix of \( |u_k| (k = 1 \ldots n_t) \) could intrinsically reveal the differences in spatial statistics, indicating the boundary between the tissue and noise subspaces. The correlation matrix was referred to as the spatial similarity matrix and the calculation was given in Eq. (7) where \( u_n \) and \( \sigma_n \) represent the mean value and the standard deviation of \( |u_n| \), respectively. Figure 1(b) demonstrates the spatial similarity matrix acquired from the in vivo measurements of human fingers. The first correlation square (red dashed square) formed by the spatial vectors of high orders revealed the tissue subspace. The noise was spatially randomly distributed so the corresponding spatial vectors were expected to be less corrected or uncorrelated (blue dashed square). Therefore, the rank estimator based on the spatial vectors was defined according to the boundary of the correlation squares.

\[
C(n,m) = \frac{1}{n_x \times n_z} \sum_{p} n_x n_z (|u_n(p)| - \bar{|u_n|}) \times (|u_n(p)| - \bar{|u_n|}) \sigma_n \times \sigma_m, \quad (m,n) \in [[1, n_t]^2
\]

Fig. 1. Rank estimators based on temporal singular vectors (a) and spatial singular vectors (b) for an in vivo radiofrequency (RF) acquisition using human fingers. (a) Power spectral density (PSD) and 99% bandwidth (BW) of temporal singular vectors. (b) Correlation matrix of spatial singular vectors. The inflection point of the BW curve in (a) was denoted by a black arrow. The red dashed square in (b) from 1 to 10 represents the tissue subspace while the blue dashed square in (b) represents the noise subspace. Noted that the blue dashed square has weak correlation values. Estimations were shown on singular vectors ranging from 1 to 75.

### 2.3. Numerical simulations

The spatiotemporal SVD based denoiser was first evaluated on simulated PA images. To generate the dataset, initial pressure distributions were created by simulating the optical fluence distributions on artificial vasculature using a Monte Carlo simulation program called MCML [28]. A square region of 40.0 mm (X) \( \times \) 40.0 mm (Z) with a grid size of 0.1 mm was considered. The vascular pattern was acquired from the DRIVE database [29]. The optical properties of the vessels were specified by referring to those for oxygenated whole blood (150 g hemoglobin/L) [30]. The optical scattering coefficient, the optical absorption coefficient, and the anisotropy of scattering at the wavelength of 850 nm were 0.5 mm\(^{-1}\), 5.9 mm\(^{-1}\), and 0.9 respectively. The background was assigned as a standard homogenous tissue with a uniform refractive index of...
1.4, optical scattering coefficient of 10 mm\(^{-1}\), optical absorption coefficient of 1.5 mm\(^{-1}\), and anisotropy of 0.9. The PA excitation light was simulated as a homogeneous 38.4 mm photo beam delivering light at the tissue surface. The fluence distribution map was generated with 100 000 photons in 5 mins.

A linear array US transducer with 128 sensor elements over a length of 38.4 mm was simulated in \(k\)-Wave [30]. Each element had a central frequency of 7 MHz and a \(\neg 6\) dB bandwidth of 80.9% for simulating the US sensors employed in the experimental setup (See section 2.4 below for details). During the forward simulations, PA waves were generated and propagated based on the initial pressure distribution map and then received by the sensor elements. To simulate RF acquisitions under a steady condition, the vessels spanning from 10.0 mm to 20.0 mm in the Z direction were assumed to be stationary over 100 frames. Furthermore, considering the slow movements during handheld imaging and mechanical scanning, the total 100 acquisitions were divided into 4 groups where the vessels region was sequentially moved along the Z direction with 0 mm, 1 mm, 2 mm, and \(\neg 1\) mm, respectively. Gaussian random noise was added to the Casorati matrix of the simulated RF data. Three typical noise levels at \(\neg 5\) dB, \(\neg 10\) dB, and \(\neg 15\) dB in SNR were considered according to the noise distributions measured \textit{in vivo}.

2.4. PA imaging of the human fingers and forearm \textit{in vivo}

Figure 2 shows the spatiotemporal SVD based PA denoising framework with \textit{in vivo} data. Human volunteer experiments were approved by the King’s College London Research Ethics Committee (study reference: HR-18/19-8881). An LED-based PA/US imaging system (AcousticX, CYBERDYNE INC, Tsukuba, Japan) was employed for collecting \textit{in vivo} data for validation [15]. US detection used an integrated PA/US probe consisting of a 7 MHz linear array transducer with 128 elements and two 850 nm LED arrays affixed on both sides. The maximum fluence of the illumination area by the LED arrays was 0.11 mJ/cm\(^2\) [15].

![Fig. 2. Schematic illustration of spatiotemporal SVD based PA denoising. A finger of a healthy human volunteer is imaged (for around 22 s) in water with a LED-based photoacoustic (PA)/ultrasound (US) imaging system. Radiofrequency (RF) data are acquired for offline processing. A Casorati matrix is then formulated based on the RF data, and randomised singular value decomposition (rSVD) is performed on the Casorati matrix. The corresponding filtered RF data is then obtained by unrolling the filtered Casorati matrix along the time dimension and reconstructed using a Fast Flourier Transform (FFT)-based algorithm. The reconstruction PA images cover a region of 38.4 mm \(\times\) 10 mm.](image-url)
To validate the proposed method for \textit{in vivo} applications, especially superficial vasculature, fingers of a healthy volunteer were imaged with the imaging probe handheld underwater. Imaging was performed for around 22 s with 1536 frames of raw PA data (maximum number of PA frames could be saved at one time) captured for offline processing. Another validation experiment involved \textit{in vivo} 3-D scanning of superficial vessels of the human forearm. The imaging probe was affixed to a motorised linear translation stage and moved in the Y-direction at a constant speed. To maintain a reliable spatial and temporal resolution during scanning, the probe was translated for around 40 mm in 13 s along the Y-direction with 1299 frames of raw PA data acquired, which were then fed into a Fourier domain algorithm for image reconstruction [31]. The proposed denoising method was further compared with discrete wavelet transform (DWT). DWT was implemented on a single frame by employing ‘db4’ as the mother wavelet and soft thresholding via universal thresholding [32].

2.5. \textbf{Quantitative analysis}

Reference metrics including peak-to-noise ratio (PSNR), edge preservation index (EPI) [33,34], and structural similarity index measure (SSIM) [35] were chosen for analysing the denoising performance on simulated data. With a denoised PA image $X$ and its corresponding noise-free image $Y$ of size $M \times N$, the PSNR is given by:

\begin{equation}
\text{PSNR} = 10 \log_{10} \left( \frac{\max(X)}{\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} [X(i,j) - Y(i,j)]^2} \right)
\end{equation}

The simplified SSIM is computed as:

\begin{equation}
\text{SSIM} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\mu_x^2 + \mu_y^2 + C_1(\sigma_x^2 + \sigma_y^2 + C_2)}
\end{equation}

where $\mu_x, \mu_y, \sigma_x, \sigma_y, \sigma_{xy}$ represents the means, standard deviations, and cross-variance for the denoised image $X$ and noise-free image $Y$ respectively. In Eq. (9), $C_1 = 0.01L^2$ and $C_2 = 0.03L^2$ where $L$ is 255 based on the dynamic range of the input images of data type uint8. The EPI is a commonly used index for edge quality analysis especially during image denoising [34]. Equation (10) defines the computation of the EPI.

\begin{equation}
\text{EPI} = \frac{\Gamma(\Delta s - \overline{\Delta s}, \Delta s - \overline{\Delta s})}{\sqrt{\Gamma(\Delta s - \overline{\Delta s}, \Delta s - \overline{\Delta s})\Gamma(\overline{\Delta s} - \overline{\Delta s}, \overline{\Delta s} - \overline{\Delta s})}}
\end{equation}

\begin{equation}
\Gamma(s_1, s_2) = \sum_{(i,j) \in \text{ROIs}} s_1(i,j)s_2(i,j)
\end{equation}

Where $\overline{\Delta s}$ and $\overline{\Delta s}$ represent mean values of the high pass Laplacian filtered regions of interest (ROIs) $\Delta s$ and $\overline{\Delta s}$ of the denoised image $X$ and noise-free image $Y$, respectively. Here, the EPI was calculated using the whole image instead of ROIs.

To quantitatively evaluate the denoising performance in terms of image quality and sensitivity to slow motions using \textit{in vivo} data, axial resolution, and non-reference metrics including signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) were measured.

For \textit{in vivo} imaging with human fingers, strong PA signals were observed from a double layer structure that may correspond to a digital artery considering the apparent pulsations in real-time display [15]. To further quantify the motion-induced blurring effect with frame averaging and...
compare the sensitivity of spatiotemporal SVD to subtle movements, the axial resolution was calculated at vessel edges. An edge-spread function was obtained by drawing an axial profile across a digital artery. The full width at half maximum (FWHM) value was then measured as the axial resolution of the system.

Signals that may have corresponded to blood vessels and background noise were extracted from specified ROIs of equal pixels based on reconstructed PA images. SNR was defined as the ratio of the mean amplitude of signal regions and the standard deviation of background noise in decibels. CNR was given by the difference of mean amplitude in signal and noise regions divided by the standard deviations of the noise regions. They can be expressed as:

\[
\text{SNR} = 10 \log_{10} \left( \frac{\frac{1}{n} \sum_{i=1}^{n} \mu_i}{\frac{1}{m} \sum_{j=1}^{m} \sigma_j} \right) \tag{12}
\]

\[
\text{CNR} = \frac{\left| \frac{1}{n} \sum_{i=1}^{n} \mu_i - \frac{1}{m} \sum_{j=1}^{m} \mu_j \right|}{\frac{1}{m} \sum_{j=1}^{m} \sigma_j} \tag{13}
\]

where \( \mu_i \), \( \mu_j \), and \( \sigma_j \) denote the mean signal amplitude, mean noise amplitude, and standard deviation of background noise for the \( i \)th and \( j \)th ROI of signals and noise, respectively.

The rank estimators based on temporal and spatial singular vectors were implemented for both numerical simulations and in vivo experiments. To further investigate the impact of the rank selection on image quality, signal averaging and spatiotemporal SVD were performed with different numbers of SVCs using simulated noisy data. For each number of SVCs ranging from 1 to 7, the PSNR, EPI, and SSIM were calculated over 100 image frames. Besides, since signal averaging and spatiotemporal SVD were implemented on a batch of RF frames, the impact of different batch sizes was explored. A sliding window was incorporated for selecting RF frames during the batch processing. For numerical simulations, averaging and spatiotemporal SVD were performed over a total number of 100 RF frames and compared with a batch size of 5, 10, 25, 50, 75, and 100, respectively. For in vivo finger imaging, 300 RF frames with an acquisition time of \( \sim 5 \) s were processed with averaging, DWT, and spatiotemporal SVD. For averaging and spatiotemporal SVD, the axial resolution, SNR, and CNR were compared with a batch size of 50, 100, 150, 200, 250, and 300, respectively.

3. Results

3.1. Numerical experiments

Figure 3(a) shows the denoising performance of averaging and spatiotemporal SVD using the simulated vascular data without slow movements. Two representative noise levels with the SNRs of \(-10\) dB and \(-15\) dB were chosen. Averaging over 100 frames reduced most of the random noise in the background and enhanced the vessel signals. Visually speaking, spatiotemporal SVD achieved similar improvements in noise reduction and signal preservation with the first SVC. The denoising performance decreased with more SVCs used for the inversion. Considering the noisier data, spatiotemporal SVD had comparable performance to averaging with the largest SVC.

Figure 3(b) further demonstrates the impact of the number of SVCs (i)-(iii) and RF frames (iv)-(vi) on denoising. The performances of spatiotemporal SVD and averaging were quantified by PSNR, EPI, and SSIM with noise-free images as reference. Significant decreases in the measured metrics were observed with the largest 7 SVCs. For all three noise levels, with the largest SVC, spatiotemporal SVD achieved higher PSNR, EPI, and SSIM than averaging. Figure 3(b) (iv)-(vi) show the quantitative performance against the different numbers of RF frames in use. The statistical difference at small numbers of frames (e.g., 5 and 10) was barely observed for spatiotemporal SVD and averaging. However, the performances were improved by spatiotemporal
Fig. 3. Numerical experiments comparing spatiotemporal SVD (STSVD) and frame averaging on simulated vascular data without slow movements. **a.** Noisy and denoised photoacoustic (PA) images by STSVD and corresponding averaged and noise-free images. Two typical noise levels at -10 dB and -15 dB respectively were compared. **b.** Quantitative analysis of STSVD and frame averaging in terms of peak-to-noise ratio (PSNR), edge preservation index (EPI), and structural similarity index measure (SSIM). (i)-(iii) reported the quantitative results of STSVD against different numbers of singular value components (SVCs) for signal reconstruction and frame averaging (dotted lines) using simulated vascular data at three noisy levels; 75 frames were used both for STSVD and averaging. (iv)-(vi) showed the quantitative results of STSVD and frame averaging against different numbers of frames using simulated vascular data at the middle noisy level (-10 dB in SNR). Data of STSVD in (i)-(vi) and averaging in (iv)-(vi) represent mean values with shades denoting standard deviations. **c.** Temporal (i) and spatial criteria (ii) for rank selection for simulated vascular data without motions. All PA images share the same scale bar of 5 mm.
SVD with more frames in use. For example, PSNR, EPI, and SSIM of spatiotemporal SVD were 1.2-, 1.1-, and 1.0-fold higher than those of frame averaging with 75 frames. Noted that both for spatiotemporal SVD and averaging, the enhancements in PSNR, EPI, and SSIM became less significant as the number of frames was larger than 50.

The proposed rank estimators based on spatial and temporal vectors were applied to the simulated data. Figure 3(c) (i) and (ii) show the acquired PSD, 99% BW, and spatial similarity matrix. For the simulated data without motions, the tissue subspace was identified by the first transition point during the evolution of the 99% bandwidth and the first correlation square. Those corresponded well to the experimental results with different numbers of SVCs in Fig. 3(b) (i)-(iii).

Figure 4 further compares the denoising performance of averaging and spatiotemporal SVD in terms of signal sensitivity by using simulated data with slow motions. Two frames (corresponding to moment t1 and t2) were chosen from the total of 100 frames for comparison. The time interval was carefully selected to include a certain number of synthetic motions. As shown in Fig. 4(a), frame averaging effectively suppressed background noise, but rendered motion artefacts indicated by the blurring vessel edges. In contrast, spatiotemporal SVD resolved the sharp edges of the blood vessels and removed noise in the background. Figure 4(b) explicitly demonstrates its statistical performance regarding the optimisation of SVCs and frames. The optimal performance of spatiotemporal SVD shown in Fig. 4(a) was achieved with the first 4 SVCs and 75 RF frames. However, for frame averaging, it is worth noting that the improvement of denoising performance was not consistent with the increase of the processed number of frames due to the motions. In Fig. 4(b) (iv)-(vi), the PSNR, EPI, and SSIM (blue dashed lines) suddenly decreased as the number of frames increased from 25 to 50. The non-parametric rank estimators were also evaluated with the simulated data containing slow motions (Fig. 4(c)). The bandwidth and spatial similarity matrix both gave an optimal threshold around the first 4 SVCs. The estimated threshold also appeared consistent with the results acquired by individually testing possible numbers of SVCs [Fig. 4(b) (i)-(iii)].

3.2. In vivo experiments with human fingers

Figure 5 compares representative results of denoised PA images through frame averaging, DWT, and spatiotemporal SVD using in vivo data from human fingers. Differing from averaging and spatiotemporal SVD, the wavelet signal denoiser was employed on a frame-by-frame basis. Quantitative results in terms of axial resolution at vessel edges, SNR, and CNR with averaging and spatiotemporal SVD were reported against the different number of frames in use. A total number of 300 frames with an acquisition time of ∼5 s was processed. Spatiotemporal SVD employed 200 frames (acquisition time: ∼3.5 s) in the Casorati matrix considering the optimality of noise suppression and temporal resolution. Comparisons were performed between noisy PA images, spatiotemporal SVD denoised PA images at three different time points and frame averaged results during the intervals.

Before denoising, PA signals from blood vessels were buried by the randomly distributed background noise of high intensity, leading to poor visibility of vascular structures, and limited tissue penetration depth. The wavelet-based denoiser reduced the noise but also generated substantial signal loss with a mean SNR of 13.8 dB and CNR of 0.3 over 300 frames. Spatiotemporal SVD greatly suppressed the background noise and maintained high sensitivity to fine tissue structures. e.g., the improved contrast of a digital artery denoted by white arrows in Fig. 5(e) with a two-layer structure. Averaging over the first 200 frames achieved an improved SNR of blood vessels, but the movement of the imaging probe and the finger blurred the edges [white arrows in Fig. 5(c)]. Furthermore, movements could also be observed between t2 and t3 where additional vascular signals became visible [yellow arrows in Fig. 5(e)]. As expected, the reconstructed image using the averaged RF data of this interval was largely contaminated by motion artefacts as well as background noise. On the contrary, spatiotemporal SVD effectively
Fig. 4. Numerical experiments comparing spatiotemporal SVD (STSVD) and frame averaging on simulated vascular data with slow movements. 

a. Noisy and denoised photoacoustic (PA) images at t1 and t2 by STSVD and corresponding averaged and noise-free images; noise level: $-10 \text{ dB in SNR}$. 

b. Quantitative analysis of STSVD and frame averaging in terms of peak-to-noise ratio (PSNR), edge preservation index (EPI), and structural similarity index measure (SSIM). (i)-(iii) reported the quantitative results of STSVD against different numbers of singular value components (SVCs) for signal reconstruction and frame averaging (dotted lines) using simulated vascular data at three noisy levels; 75 frames were used both for STSVD and averaging. (iv)-(vi) showed the quantitative results of STSVD and frame averaging against different numbers of frames using simulated vascular data at the middle noisy level ($-10 \text{ dB in SNR}$). Data of STSVD in (i)-(vi) and averaging in (iv)-(vi) represent mean values with shades denoting standard deviations. 

c. Temporal (i) and spatial criteria (ii) for rank selection for simulated vascular data with motions. All PA images share the same scale bar of 5 mm.
Fig. 5. *In vivo* comparison using human finger data. (a) Three noisy photoacoustic (PA) frames, (b) corresponding ultrasound (US) frames, (c) frame averaging PA denoising, (d) discrete wavelet transform (DWT) PA denoising, and (e) spatiotemporal SVD (STSVD) PA denoising (see Visualization 1). Comparison of real-time PA image frames acquired from a human finger before and after spatiotemporal SVD denoising and corresponding US image frames. (f) Axial profiles drawn along the white lines marked in the denoised PA image by spatiotemporal SVD and the corresponding averaged image. (g)-(i) Quantitative performance of frame averaging and STSVD in terms of axial resolution at the vessel edges, signal-to-noise ratio (SNR), and contrast-to-noise ratio (CNR). Data of averaging and STSVD in (g)-(i) represent average values with shades denoting standard deviations. SNR and CNR values of DWT in (h)-(i) represent average values with error bars denoting standard deviations. PA and US images share the same scale bar of 5 mm.
reduced the background noise and had appreciable improvements in restoring fine structures for \textit{in vivo} data containing slow motions. The axial resolution was 0.235 mm ± 0.07 for spatiotemporal SVD with 200 frames in decomposition, which was much smaller than that of averaging (0.81 mm) across the same frames. It is noticed that in Fig. 5(h) and Fig. 5(i), SNR and CNR with frame averaging decreased with the number of averaged frames when it is larger than 200 due to the accumulated stationary noise and slow motions.

3.3. \textit{In vivo 3D scanning of human forearm}

Figure 6 shows the maximum intensity projections (MIPs) of 3D PA images of the human forearm within 5 to 25 mm in depth with and without denoising. The same number of frames (25 frames) were used for frame averaging and spatiotemporal SVD while the wavelet-based method was performed on a single frame basis. Superficial vasculature in the human forearm was greatly enhanced by averaging and spatiotemporal SVD. On the contrary, the wavelet denoiser failed to mitigate the noise and degraded the spatial resolution of vessels. Spatiotemporal SVD largely retained the microvasculature distribution in 3D. However, severe distortions including discontinuity and edge jags of blood vessels were shown by using signal averaging. It is worth noting that the enhancement by spatiotemporal SVD was unstable especially over the first 1 cm in the Y-direction. Fluctuations in background noise levels at different scanning positions led to artefacts manifested as vertical lines on the MIP images. These artefacts were largely removed by averaging and spatiotemporal SVD, but remained on the images processed by DWT.

![Fig. 6.](image)

\textbf{Fig. 6.} Comparison of spatiotemporal SVD, discrete wavelet transform (DWT), and averaging for photoacoustic (PA) denoising with 3D scanning of the human forearm \textit{in vivo}. (a) The experimental set-up with the right forearm of a healthy volunteer imaged in water. (b, c) Maximum intensity projections (MIPs) of the reconstructed 3D PA volumes onto XY-plane (b) and ZY-plane (c) with the depth Z ranging from 5 mm to 25 mm.

4. \textbf{Discussion and conclusions}

This study presented a denoising approach based on spatiotemporal SVD for low-fluence based PA imaging. Compared to signal averaging and wavelet transform, spatiotemporal SVD based denoiser explores both the spatial and temporal statistics of RF data. The proposed method was validated using simulated vascular data, \textit{in vivo} 2D PA image data acquired from human fingers, and 3D scanning data of human forearm, outperforming the frame averaging and DWT in terms of suppression of background noise, sensitivity to slow motions, and vascular signal specificity. The averaged axial resolution was reduced from 0.65 to 0.24 mm by 63.1\%, and averaged SNR and CNR values of the denoised images were improved by 56.0\% and 154.5\% with STSVD respectively compared to those with frame averaging. We demonstrated a processing time of about 10 ms for a matrix of size 600 × 128 × 200 (~50 µs per frame) using a GPU (Tesla T4, PyTorch v1.11.0), indicating the feasibility of real-time processing.
Low-cost light sources such as LEDs and LDs are promising for clinical translation. However, the imaging performance of low-fluence based PA systems is hindered by low output energy of these light sources, especially for deep tissue. Averaging over tens to hundreds of image frames using these light sources with high pulse repetition frequencies is a common way to improve the SNR but decreases the frame rate, leading to a reduced temporal resolution that is critical for real-time studies of physiological responses such as pulsatile and prone to movement artefacts.

Since the temporal resolution of low-fluence based PA imaging systems is sufficiently high, signals from tissue present high spatial and temporal correlation during high-speed acquisitions while thermal noise can reach its full randomness and appears to be spatiotemporally uncorrelated. Therefore, spatiotemporal fluctuations of signals and noise could be efficiently characterised by singular vectors. Besides, compared to conventional solid-state laser-based PA imaging systems, systems employing low-fluence light sources are unlikely to be affected by high-intensity parasitic noise. Thus, denoising was performed by preserving the low-order singular values contributed by signals from critical targets including deep depth with high spatiotemporal correlations. Validation results on simulated data showed that noise reduction with minimal degradation of critical signals could be achieved by constructing the signal subspace using the largest $k$ singular values ($k \ll t$). Besides, the rank estimators based on the characteristics of the singular components were applied to both simulated and in vivo data, showing high efficiency for automatically resolving the boundary of tissue and noise subspaces. The number of frames for spatiotemporal SVD was determined by considering the temporal resolution and denoising performance. For systems with lower repetition frequencies, real-time denoising with spatiotemporal SVD could be implemented with a sliding window across consecutive image frames.

Spatiotemporal SVD was validated with 2D images of the human fingers and 3D scans of the human forearm in vivo. Variations of microvasculature between different image frames were challenging to capture by using signal averaging. Spatiotemporal SVD retained a better sensitivity to the variations including the signal intensity that were reflected by the dynamic ranges of SNR and CNR. In addition, 3D scanning involved high volumes of intensive spatial and temporal variations, requiring the denoising process being specific and sensitive to the tissue signals. It is worth noting that the 3D blood vasculature was well restored by spatiotemporal SVD with a relatively small number of frames in use. However, outliers that were projected as a set of parallel lines could be observed, especially for initial frames of scanning, which could be attributed to unstable scanning conditions that interrupted the consistency of signal or noise statistical distributions.

Spatiotemporal SVD outperformed conventional single frame-based denoiser such as DWT as it leveraged coherence information from multiple frames. State-of-the-art non-learning based denoisers such as Block Matching 3D (BM3D) have shown promising results for processing nature images [36] and medical images [37,38] but less effective for realistic PA data in the image domain [39]. More recently, a new voxel denoising method based on average statistics of signals and noise was proposed for PA imaging to quantify drugs [40]. Learning-based denoisers based on supervised models [41,42] require large noisy input with clean targets for training, which is challenging and expensive in clinical practice. Self-supervised models demonstrate competitive performance to learning-based methods without requiring clean data as references [43,44]. However the denoising performance could be improved with extra information on the noise distribution. In future studies, spatiotemporal SVD could be used to generate denoised in vivo PA data as ground truths for training deep-neural networks, which could be advantageous compared to previous studies using averaged frames as ground truths [16].

We conclude that spatiotemporal SVD is well suited for denoising in PA imaging with low-fluence light sources.

**Funding.** Engineering and Physical Sciences Research Council (NS/A000027/1, NS/A000049/1); Wellcome Trust
Disclosures. T.V. is co-founder and shareholder of Hypervision Surgical Ltd, London, UK. He is also a shareholder of Mauna Kea Technologies, Paris, France.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. P. Beard, “Biomedical photoacoustic imaging,” Interface Focus. 1(4), 602–631 (2011).
2. S. Troumas, A. Rosenenthal, C. Lutzweiler, D. Razansky, and V. Ntziachristos, “Spatiotemporal denoising framework for multispectral photoacoustic imaging based on sparse signal representation,” Med. Phys. 41(11), 113301 (2014).
3. C. Zhang, K. I. Maslov, J. Yao, and L. V. Wang, “In vivo photoacoustic microscopy with 7.6-µm axial resolution using a commercial 125-MHz ultrasonic transducer,” J. Biomed. Opt. 17(11), 1 (2012).
4. S. H. Holan and J. A. Viator, “Automated wavelet denoising of photoacoustic signals for circulating melanoma cell detection and burn image reconstruction,” Phys. Med. Biol. 53(12), N227 (2008).
5. S. A. Ermilov, T. Khamapirad, A. Conjusteau, M. H. Leonard, R. Lacewell, K. Mehta, T. Miller, and A. A. Oraevsky, “Laser optoacoustic imaging system for detection of breast cancer,” J. Biomed. Opt. 14(2), 024007 (2009).
6. G. Güney, N. Uluç, A. Demirkiran, E. Aytaç-Kipergil, M. B. Unlu, and O. Birgul, “Comparison of noise reduction methods in photoacoustic microscopy,” Comput. Biol. Med. 109, 333–341 (2019).
7. M. Sun, N. Feng, Y. Shen, X. Shen, and J. Li, “Photoacoustic signals denoising based on empirical mode decomposition and energy-window method,” Adv. Adapt. Data Anal. 04(01n02), 1250004 (2012).
8. C. Guo, J. Wang, Y. Qin, H. Zhan, J. Yuan, Q. Cheng, and X. Wang, “Photoacoustic imaging optimization with raw signal deconvolution and empirical mode decomposition,” in Photons Plus Ultrasound: Imaging and Sensing 2018 (SPIE, 2018), 10494, pp. 461–470.
9. M. Zhou, H. Xia, H. Zhong, J. Zhang, and F. Gao, “A Noise Reduction Method for Photoacoustic Imaging In Vivo Based on EMD and Conditional Mutual Information,” IEEE Photonics J. 11(1), 1–10 (2019).
10. E. R. Hill, W. Xia, M. J. Clarkson, and A. E. Desjardins, “Identification and removal of laser-induced noise in photoacoustic imaging using singular value decomposition,” Biomed. Opt. Express 8(1), 68 (2017).
11. I. U. Haq, R. Nagaoka, S. Siregar, and Y. Sajo, “Sparse-representation-based denoising of photoacoustic images,” Biomed. Phys. Eng. Express 3(4), 045014 (2017).
12. M. Kuniyil Ajith Singh and W. Xia, “Portable and affordable light source-based photoacoustic tomography,” Sensors 20(21), 6173 (2020).
13. Y. Zhu, T. Feng, Q. Cheng, X. Wang, S. Du, N. Sato, M. Kuniyil Ajith Singh, and J. Yuan, “Towards clinical translation of LED-based photoacoustic imaging: a review,” Sensors 20(9), 2484 (2020).
14. J. Jo, G. Xu, Y. Zhu, M. Burton, J. Sarazin, E. Schiopu, G. Gandikota, and X. Wang, “Detecting joint inflammation by an LED-based photoacoustic imaging system: a feasibility study,” J. Biomed. Opt. 23(11), 1 (2018).
15. W. Xia, M. Kuniyil Ajith Singh, E. Maneas, N. Sato, Y. Shigeta, T. Agano, S. Ourselin, S. J. West, and A. E. Desjardins, “Handheld real-time LED-based photoacoustic and ultrasound imaging system for accurate visualization of clinical metal needles and superficial vasculature to guide minimally invasive procedures,” Sensors 18(5), 1394 (2018).
16. E. M. A. Anas, H. K. Zhang, J. Kang, and E. Doctor, “Enabling fast and high quality LED photoacoustic imaging: a recurrent neural networks based approach,” Biomed. Opt. Express 9(8), 3852 (2018).
17. K. Wang, J. Xia, C. Li, L. V. Wang, and M. A. Anastasio, “Fast spatiotemporal image reconstruction based on low-rank matrix estimation for dynamic photoacoustic computed tomography,” J. Biomed. Opt. 19(05), 1 (2014).
18. C. Demene, T. Delfieux, M. Perriot, B.-F. Osmanski, V. Biran, J.-L. Gennisson, L.-A. Sieu, A. Bergel, S. Franqui, J.-M. Correas, I. Cohen, O. Baud, and M. Tanter, “Spatiotemporal clutter filtering of ultrafast ultrasound data highly increases doppler and ultrasound sensitivity,” IEEE Trans. Med. Imaging 34(11), 2271–2285 (2015).
19. P. Song, A. Manduca, J. D. Trzasko, and S. Chen, “Ultrasound small vessel imaging with block-wise adaptive local clutter filtering,” IEEE Trans. Med. Imaging 36(1), 251–262 (2017).
20. R. Al Mukaddim, A. M. Weichmann, C. C. Mitchell, and T. Varghese, “Enhancement of in vivo cardiac photoacoustic signal specificity using spatiotemporal singular value decomposition,” J. Biomed. Opt. 26(04), 046001 (2021).
21. N. Halko, P.-G. Martinsson, and J. A. Tropp, “Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions,” arXiv:0909.4061 [math] (2010).
22. P.-G. Martinsson and S. Voronin, “A randomized blocked algorithm for efficiently computing rank-revealing factorizations of matrices,” arXiv:1503.07157 [math] (2015).
23. P. Song, J. D. Trzasko, A. Manduca, B. Qiang, R. Kadirvel, D. F. Kallmes, and S. Chen, “Accelerated signal value-based ultrasound blood flow clutter filtering with randomized singular value decomposition and randomized spatial downsampling,” IEEE Trans. Ultrason., Ferroelect., Freq. Contr. 64(4), 706–716 (2017).
24. L. A. F. Ledoux, P. J. Brands, and A. P. G. Hoeks, “Reduction of the clutter component in doppler ultrasound signals based on singular value decomposition: a simulation study,” Ultrason Imaging 19(1), 1–18 (1997).
25. A. C. H. Yu and L. Lovstakken, “Eigen-based clutter filter design for ultrasound color flow imaging: a review,” IEEE Trans. Ultrason., Ferroelect., Freq. Contr. 57(5), 1096–1111 (2010).
26. J. Baranger, B. Arnal, F. Perren, O. Baud, M. Tanter, and C. Demené, “Adaptive Spatiotemporal SVD clutter filtering for ultrafast doppler imaging using similarity of spatial singular vectors,” IEEE Trans. Med. Imaging 37(7), 1574–1586 (2018).

27. S. Hovda, H. Rue, and B. Olstad, “New echocardiographic imaging method based on the bandwidth of the ultrasound Doppler signal with applications in blood/tissue segmentation in the left ventricle,” Computer Methods and Programs in Biomedicine 92(3), 279–288 (2008).

28. L. Wang, S. L. Jacques, and L. Zheng, “MCML—Monte Carlo modeling of light transport in multi-layered tissues,” Computer Methods and Programs in Biomedicine 47(2), 131–146 (1995).

29. J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, “Ridge-based vessel segmentation in color images of the retina,” IEEE Trans. Med. Imaging 23(4), 501–509 (2004).

30. B. E. Treeby and B. T. Cox, “k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields,” J. Biomed. Opt. 15(2), 021314 (2010).

31. M. Jaeger, S. Schüpbach, A. Gertsch, M. Kitz, and M. Frenz, “Fourier reconstruction in optoacoustic imaging using truncated regularized inverse k-space interpolation,” Inverse Problems 23(6), S51–S63 (2007).

32. D. L. Donoho and I. M. Johnstone, “Ideal Spatial Adaptation by Wavelet Shrinkage,” Biometrika 81(3), 425–455 (1994).

33. F. Sattar, L. Floreby, G. Salomonsson, and B. Lovstrom, “Image enhancement based on a nonlinear multiscale method,” IEEE Trans. on Image Process. 6(6), 888–895 (1997).

34. J. Joseph, S. Jayaraman, R. Periyasamy, and V. R. Simi, “An edge preservation index for evaluating nonlinear spatial restoration in MR images,” Curr. Med. Imaging Rev. 13(1), 58–65 (2017).

35. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” IEEE Trans. on Image Process. 13(4), 600–612 (2004).

36. K. Dahov, A. Foi, V. Katkownik, and K. Egiazarian, “Image denoising by sparse 3-D transform-domain collaborative filtering,” IEEE Trans. on Image Process. 16(8), 2080–2095 (2007).

37. L. Fang, S. Li, Q. Nie, J. A. Izatt, C. A. Toth, and S. Farsiu, “Sparsity based denoising of spectral domain optical coherence tomography images,” Biomed. Opt. Express 3(5), 927–942 (2012).

38. C. A. N. Santos, D. L. N. Martins, and N. D. A. Mascarenhas, “Ultrasound image de-speckling using stochastic distance-based BM3D,” IEEE Trans. on Image Process. 26(6), 2632–2643 (2017).

39. D. He, J. Zhou, X. Shang, J. Luo, and S.-L. Chen, “De-noising of photoacoustic microscopy images by deep learning,” arXiv:2201.04302 [cs, eess] (2022).

40. A. Khadria, C. D. Paavola, K. Maslov, F. A. Valenzuela, A. E. Sperry, A. L. Cox, R. Cao, J. Shi, P. L. Brown-Augsburger, E. Lozano, R. L. Blankenship, L. V. Wang, “Photoacoustic imaging reveals mechanisms of rapid-acting insulin formulations dynamics at the injection site,” Molecular Metabolism 62, 101522 (2022).

41. K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising,” IEEE Trans. on Image Process. 26(7), 3142–3155 (2017).

42. M. Weigert, U. Schmidt, T. Boothe, A. Müller, A. Dibrov, A. Jain, B. Wilhelm, D. Schmidt, C. Broaddus, S. Culley, M. Rocha-Martins, F. Segovia-Miranda, C. Norden, R. Henriques, M. Zerial, M. Solimena, J. Rink, P. Tomancak, L. Royer, F. Jug, and E. W. Myers, “Content-aware image restoration: pushing the limits of fluorescence microscopy,” Nat. Methods 15(12), 1090–1097 (2018).

43. A. Krull, T.-O. Buchholz, and F. Jug, “Noise2Void - learning denoising from single noisy images,” in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (IEEE, 2019), pp. 2124–2132.

44. J. Batson and L. Royer, “Noise2Self: Blind Denoising by Self-Supervision,” in Proceedings of the 36th International Conference on Machine Learning (PMLR, 2019), pp. 524–533.