

# A Novel Framework for Remote Photoplethysmography Pulse Extraction on Compressed Videos

Changchen Zhao<sup>1,2</sup>, Chun-Liang Lin<sup>\*2</sup>, Weihai Chen<sup>\*1</sup>, Zhengguo Li<sup>3</sup>

<sup>1</sup>School of Automation Science and Electrical Engineering, Beihang University, Beijing, 100191, China

<sup>2</sup>Department of Electrical Engineering, National Chung Hsing University, Taichung, 402, Taiwan

<sup>3</sup>Signal Processing Department, Institute for Infocomm Research, 138632, Singapore

chunlin@dragon.nchu.edu.tw, whchen@buaa.edu.cn

## Abstract

*Remote photoplethysmography (rPPG) has recently attracted much attention due to its non-contact measurement convenience and great potential in health care and computer vision applications. However, almost all the existing rPPG methods are based on uncompressed video data, which greatly limits its application to the scenarios that require long-distance video transmission. This paper proposes a novel framework as a first attempt to address the rPPG pulse extraction in presence of video compression artifacts. Based on the analysis of the impact of various compression methods on rPPG measurements, the problem is cast as single-channel signal separation. The framework consists of three major steps to extract the pulse waveform and heart rate by exploiting frequency structure of the rPPG signal. A benchmark dataset which contains stationary and motion videos has been built. The results show that the proposed algorithm significantly improves the SNR and heart rate precision of state-of-the-art rPPG algorithms on stationary videos and has a positive effect on motion videos at low bitrates.*

## 1. Introduction

Remote photoplethysmography (rPPG), also known as imaging photoplethysmography (iPPG), aims at detecting human cardiac activities by means of a digital camera. The periodical cardiac activities lead to fluctuations in the amount of blood volume in the microvascular beneath the skin tissues thereby causing skin color variations that, though invisible to naked eyes, can be perceived by an optical camera. Compared with the conventional contact PPG measurement like the pulse oximeter, the non-contact rPPG causes no skin irritation and imposes no constraints on body motion. This technique has led to a variety of applications such as heart rate (HR) [12], respiratory rate [25], and SpO<sub>2</sub>

[7] monitoring, arterial stiffness assessment [20], living-skin detection for face anti-spoofing [13, 31], etc. It also has a great potential in health care and computer vision applications such as fitness training [29], home care services, sleep monitoring, driver auxiliary system, etc.

Numerous methods have been proposed during the last few decades for robust signal extraction. The biggest challenge is the suppression of motion artifacts (MAs), *i.e.*, the color variation caused by body motion whose signal power is usually larger than that of the pulse signal. The methods can be classified into three major groups: 1) blind source separation (BSS)-based methods, which separate rPPG signal from multiple sources using independent component analysis (ICA) [19, 21, 16] or principal component analysis (PCA) [11]; 2) data driven methods, which extract pulse signal without physiological considerations by exploiting, for example, spatial subspace rotation [32], spatial redundancy [30], or head motion [3]; and 3) model based methods, which are based on the optical/physiological principles, *e.g.*, the skin reflection model [5, 6, 28], or the camera-based signal acquisition model [33, 10]. For a thorough review of the development of rPPG, please refer to [18, 24, 23, 2].

There exists one question that is crucial but has not gained much attention: the impact of video compression on rPPG measurements (referred to as video compression artifacts, VCAs). Nearly all the existing rPPG methods are designed based on uncompressed video data. One of the main issues is that video data occupy a large amount of storage space, which is unsuitable for sharing databases online. For example, a 1-minute uncompressed video file of size  $640 \times 480$  needs approximately 1.7 GB hard disk space. rPPG researches always require video data of duration in hours. The most important issue is that, without compression, it is impossible to apply rPPG to the cases that need telecommunication, *e.g.*, remote homecare where video data need to be transmitted from remote clients to the health

care center for analysis. The uncompressed video usually has a bitrate over 220 Mb/s, which largely exceeds the transmission capability of current telecommunication technology. Existing rPPG algorithms can only be applied on site rather than remotely. Another example is the mobile apps for heart rate monitoring. The uncompressed video has to be processed on the user’s cell phone rather than uploaded to the cloud, leading to a less accurate result due to insufficient computing resources on the mobile devices. Therefore, to develop algorithms for reliable rPPG extraction based on the compressed videos plays a crucial role in widening its real-world applications.

Only a few researchers have noticed this issue and done some preliminary works. Hanfland and Paul [8] pointed out that rPPG signals are mostly conserved by video compression but the overall quality is altered. McDuff *et al.* [17] studied the effects of two sets of commonly used compression methods (x264 and x265) and a series of compression ratios (constant factor rate). They found that compression considerably degrades the signal to noise ratio (SNR) of extracted rPPG signals given by existing rPPG algorithms but retains rPPG signal undestroyed. These researches only analyzed the impact of compression on rPPG measurements. To the best of our knowledge, no solution has been given for pulse signal extraction on compressed videos, which is exactly the main purpose of this paper.

A thorough analysis of VCAs is first conducted by compressing the raw videos using 4 compression methods to a wide range of bitrates. We formulate this problem as single-channel pulse extraction in presence of VCAs and MAs. A novel framework is proposed which consists of three major steps: G-channel band-pass filtering, singular spectrum analysis (SSA) with reconstructed component (RC) selection, and spectral masking. A benchmark dataset, in which subjects were asked to perform stationary and motion task, was built to evaluate the performance of the proposed algorithm in comparison with several state-of-the-art algorithms in terms of SNR and heart rate precision. Strength and weakness of the proposed algorithm are discussed. The contributions of this paper are summarized as follows:

1. A novel framework that employs single-channel signal processing is proposed as a preliminary attempt to address the problem of rPPG pulse extraction on compressed videos.
2. Analysis of the video compression artifacts is given, which serves as a reason to explain the failure of the existing rPPG methods and an insight to developing new algorithms.
3. Four compression methods (x264, x265, vp8, & vp9), 2 motion types (stationary and motion), and a wide range of target bitrates (from 100 kb/s to 20 Mb/s) have been studied in order to cover a wider range of rPPG applications.

## 2. Analysis of video compression artifacts

The raw videos are compressed using four popular compression methods widely used in video compression applications. The impact of VCAs is analyzed by comparing differences between rPPG measurements from raw and compressed videos. rPPG measurements (traces) denote the three-channelled time series representing the color variations in RGB channels of the facial skin region pixels obtained according to the spatial averaging described in Section 3.1. Three types of VCAs are observed: a) amplitude deterioration; b) high-frequency structured noise; and c) trace discontinuity. We first describe the three VCAs, briefly point out the causes, explain the failure of the existing rPPG algorithms, and finally present the idea of the proposed algorithm.

**Amplitude deterioration** means that the pulse amplitude tends to be attenuated as the bitrate decreases. An example is shown in Figure 1, where the trace in Green channel and its corresponding spectrum are plotted. A clear deterioration of the pulse amplitude is observed as the bitrate decreases. Even the waveform, not only the amplitude, has got a serious erosion at a bitrate of 100 kb/s. The deterioration effect can be seen more obviously in the frequency domain, where a dramatic drop in the power of the components with frequency at 1.2 Hz is observed.

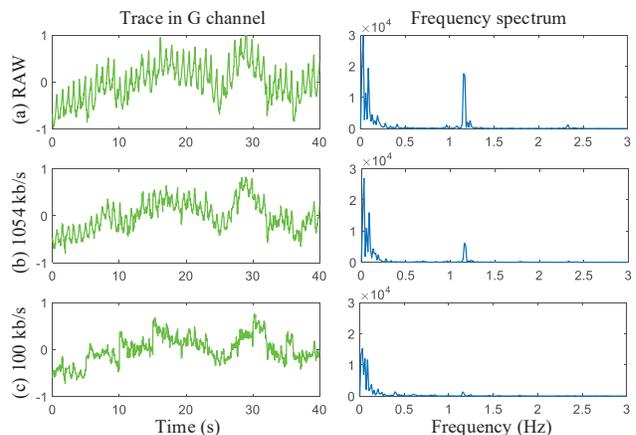


Figure 1. Amplitude deterioration. G-channel trace and its frequency spectrum are plotted in each row. (a) Trace extracted from the uncompressed video; (b) Trace extracted from the compressed video at a bitrate of 1054 kb/s; (c) Trace extracted from the compressed video at a bitrate of 100 kb/s.

**High-frequency structured noise** is another type of VCAs that arises during compression. It mainly appears in x264 compressed videos. An example is shown in Figure 2, where the R-channel trace and its corresponding spectrum are plotted. In comparison with the RAW video trace, 3420 kb/s video trace has an obvious high-frequency structured noise added to the original rPPG signal, *i.e.*, two peaks at

around 7.5 Hz in the spectrum are observed. As the bitrate decreases, the power of the noise first increases to be comparable to the pulse signal and then decreases to disappear when bitrate = 100 kb/s. The noise has fixed frequency across subjects and bitrates but has different powers in three color channels, *i.e.*, significant in R and B but less in G channel.

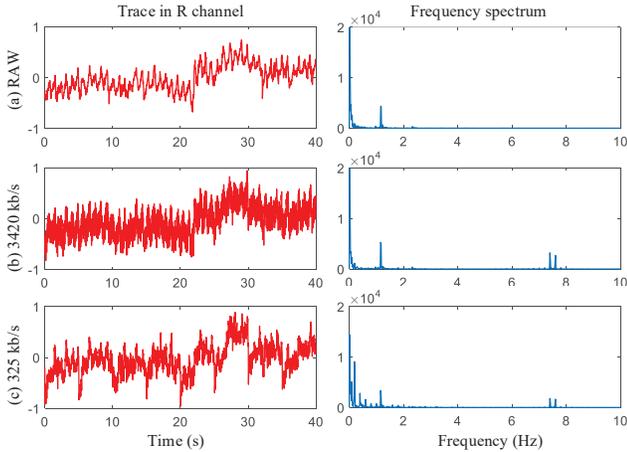


Figure 2. High-frequency structured noise. R-channel trace and its frequency spectrum are plotted in each row. (a) Trace extracted from the uncompressed video; (b) Trace extracted from the compressed video at a bitrate of 3420 kb/s. (c) Trace extracted from the compressed video at a bitrate of 325 kb/s.

**Trace discontinuity.** The traces start to be discontinuous when the bitrate becomes very low, *e.g.*, the trace is added by impulse or step noise. Examples are shown in Figure 3. Trace discontinuity is observed in x264, vp8, and vp9 video in 300-100 kb/s. However, it is not observed in x265 videos at least when bitrate is above 100 kb/s. Table 1 summarizes the presence of VCAs in each of the four compression methods.

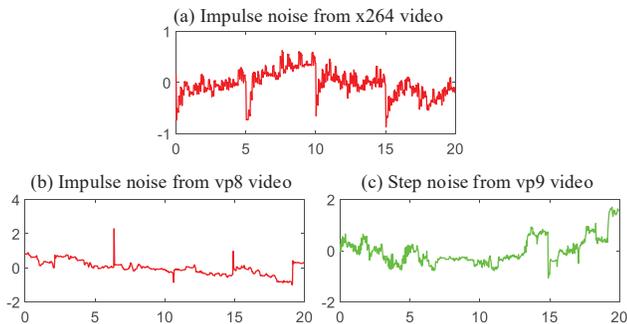


Figure 3. Trace discontinuity. Sample traces containing impulse and step noise extracted from x264, vp8, and vp9 videos.

Various steps in the video compression pipeline may cause such detrimental effects on the rPPG measurements

Table 1. Presence of VCAs in four compression methods

	x264	x265	vp8	vp9
presence of VCAs	a,b,c	a	a,c	a,c

because rPPG signals belong to the imperceptible color variations that video compression is designed to eliminate. Amplitude deterioration has a close relationship with intra-frame compression (reduced spatial redundancy in the region-of-interest (ROI)), chroma subsampling (reduced spatial redundancy in color space), and motion compression (smoothed temporal oscillation). Trace discontinuity is probably due to the intra-frame (I-frame) compression, where the I-frame is encoded independently without taking any previous/subsequent frames into account.

The failure of the existing rPPG methods becomes easy to explain, *i.e.*, no techniques are designed for VCA suppression. For example, PCA is sensitive to the outliers [34], which is the case of trace discontinuity. The skin reflection model has indeed a noise term but it is designed for Gaussian noise rather than the high-frequency structured noise. Moreover, amplitude deterioration is most likely to change the color variation directions in the skin reflection model such that the methods based on this model would be influenced.

Although VCAs have introduced corruption to the rPPG measurements, the pulse signal can still be extracted as long as it is not fully corrupted. We cast this problem as single-channel pulse extraction. For each color channel, the observed signal  $\mathbf{y} \in \mathbb{R}^{l_1}$  of length  $l_1$  can be expressed as:

$$\mathbf{y} = \mathbf{y}_{dc} + f_c(\mathbf{m} + \mathbf{p}) \quad (1)$$

where  $\mathbf{y}_{dc}$ ,  $\mathbf{p}$ ,  $\mathbf{m}$  denote the DC component, the zero-mean pulse signal, and the motion artifact, respectively, and  $f_c$  denotes a function that introduces VCAs. The purpose is to extract  $\mathbf{p}$  from  $\mathbf{y}$ .

### 3. The proposed framework

The overview of the proposed framework is plotted in Figure 4. Windowed G trace signal is decomposed by SSA and reconstructed by an RC selection criterion. Overlap adding concatenates windowed signals to a full length signal, based on which the instantaneous HR is calculated as a reference to spectral masking. The final output is the refined pulse signal and the instantaneous HR.

#### 3.1. ROI tracking and spatial averaging

The ROI is designated as the nose and cheek regions (see Figure 4) as they contain little motion artifacts caused by talking or blinking. The initial ROI is selected manually and tracked by Staple (Sum of Template And Pixel-wise LEarners) [4]. Given the bounding box of a frame, spatial averaging [19] is applied to compute the average value

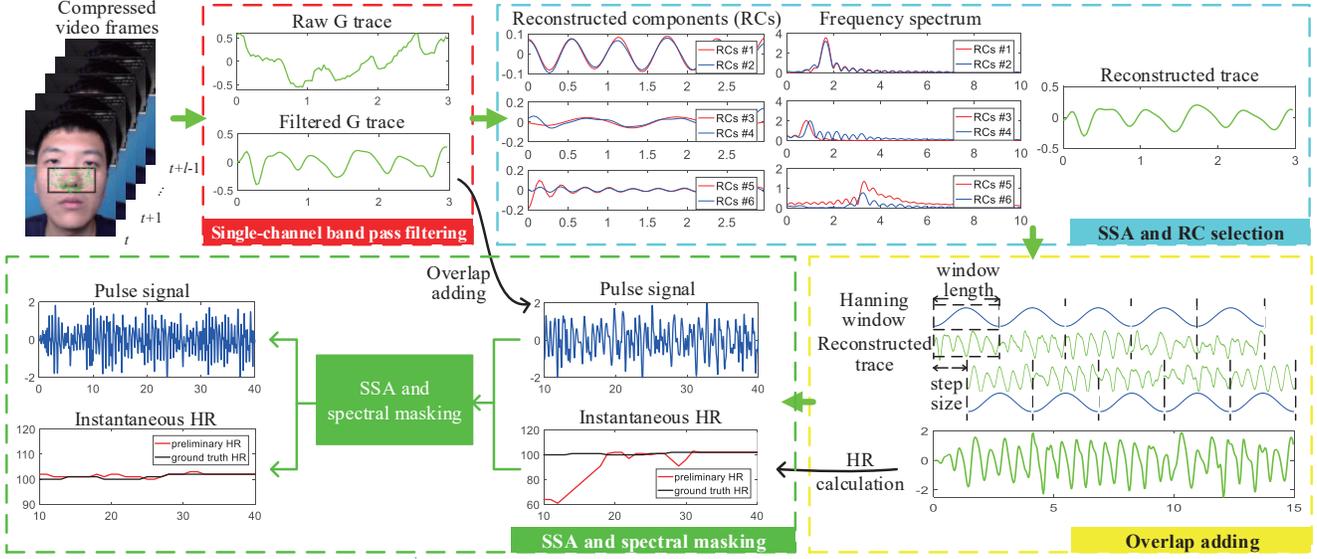


Figure 4. Framework overview.

of pixels in each color channel within the bounding box, resulting in a  $3 \times 1$  column vector. The rPPG traces are obtained by concatenating the column vector of each frame into a matrix.

### 3.2. Single-channel band-pass filtering

The analysis of VCAs indicates that a band-pass filter is necessary for pulse extraction. However, the combination of three color channels as in most existing algorithms proves to be infeasible because R and B channels are highly polluted by VCAs. Note that the VCAs have different strengths in three color channels. The high-frequency structured noise and trace discontinuity mainly appear in R and B channels and become less significant in G. G channel is mainly affected by amplitude deterioration. Hence, we choose the signal in G channel and discard R and B channels for further processing. A band-pass filter is applied to the G channel signal with cutoff frequency [0.8 5] Hz ([48 300] bpm).

### 3.3. SSA decomposition and RC selection

The benefit of single-channel pulse extraction is that the impact of VCAs can be reduced to the minimum. However, a new issue arises because existing rPPG methods cannot be applied since they are based on multi-channel signal processing. Numerous methods have been proposed in the signal processing community for the purpose of single-channel signal separation, *e.g.*, empirical mode decomposition (EMD) [9, 22], non-negative matrix factorization (NMF) [1], singular spectrum analysis (SSA) [27, 26], etc. However, they cannot be applied directly here and have to make some modifications. The proposed framework em-

ployes SSA and a RC selection strategy based on the frequency structure of the rPPG signals.

Given the filtered G-channel signal  $\mathbf{y}$ , SSA first reshapes it into a matrix  $\mathbf{Y}$  by Hankel matrix embedding, and decomposes  $\mathbf{Y}$  by singular value decomposition (SVD),  $\mathbf{Y} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ , where  $\sigma_i$  for  $i = 1, \dots, r$  denote the singular values of  $\mathbf{Y}$  with singular vectors  $\mathbf{u}_i$  and  $\mathbf{v}_i$ . The collection  $(\sigma_i, \mathbf{u}_i, \mathbf{v}_i)$  is called an eigentriple and every eigentriple is transformed back to a time series called reconstructed component (RC) by diagonal averaging. As a result, the original signal  $\mathbf{y}$  can be denoted by the sum of RCs,  $\mathbf{y} = \mathbf{y}_1 + \dots + \mathbf{y}_r$ .

The RCs are sorted according to corresponding singular values in a descending order and the first 10 RCs are retained because they preserve the most information of the trace segment. Among them, we need to select the RCs that are related to the pulse signals. The rPPG signal frequency structure is utilized as the RC selection criterion. A typical frequency structure is plotted in Figure 5, where two peaks that associate, respectively, with HR and  $2 \times \text{HR}$  (first harmonic) are observed.

The frequency spectrum of each candidate RC is computed using fast Fourier transform (FFT). The RC pairs whose dominant frequencies (frequency with maximum power) have the twice relationship are selected, while others are rejected. If there are no satisfying RC pairs, all the candidate RCs are retained in case of information loss.

### 3.4. Overlap adding

This step aims at concatenating the windowed trace to the final output [5]. The procedure is illustrated in Figure 4.

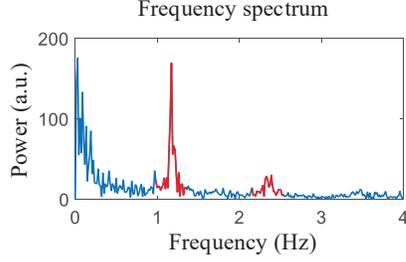


Figure 5. Example of frequency structure of a typical rPPG signal. The heart rate and its first harmonic are highlighted in red.

Every windowed trace is multiplied by a Hanning window of the same length as the windowed trace. The step size is half of the window length. The multiplied windowed trace of each step is added up to get the preliminary output.

### 3.5. Spectral masking

The RC selection step would not identify the pulse signal perfectly due to the presence of large compression and motion artifacts. We propose a spectral masking algorithm to refine the preliminary output. Note that the RC selection criterion is purely data-driven, *i.e.*, selecting RC pairs satisfying the frequency structure. The basic idea of spectral masking is that the output should be more precise if a reference heart rate is involved as *a priori*. We take the instantaneous HR of the preliminary output as the reference, set it as the central frequency of a spectral mask, and select RCs from the filtered G trace that overlap with the mask.

First, to obtain the instantaneous HR of the preliminary output, a sliding window of length  $l_2 = 10$  s and step size 1 s is applied to locate a windowed trace. The windowed trace is transformed to frequency domain using FFT. The dominant frequency is set to be the HR of this time instant. The instantaneous HR is updated every 1 second and the values between successive updating instants are equal to the previous instant. Then, the same procedure in the SSA decomposition step is repeated to obtain the spectrum of RCs. Meanwhile, a reference HR is computed by averaging the instantaneous HR within the same interval located by the sliding window of length  $l_1$ . Let the reference HR be  $f_r$  and the set of dominant frequencies of RCs be  $\mathcal{F} = \{f_i\}_{i=1}^{10}$ , the spectral masking algorithm determines the following index function,

$$\mathbf{1}_F(i) = \begin{cases} 1, & f_r - \frac{\omega_1}{2} \leq f_i \leq f_r + \frac{\omega_1}{2} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where  $\omega_1$  denotes the window length of the spectral mask. The RCs are selected by this indexing function for reconstructing the output signal. The final pulse waveform is then obtained by overlap adding.

## 3.6. Discussion

The design of single-channel pulse extraction is mainly because R and B channels are highly contaminated by VCAs. A combination of all three channels will not fully remove VCAs. However, single-channel processing methods usually suffer from MAs. The RC selection and spectral masking in the proposed framework are especially designed to deal with this problem.

## 4. Assessment details

### 4.1. Compression methods

**x264**: an open source library for encoding video streams into the H.264/AVC compression format. H.264/AVC is a current generation video compression standard that has been widely adopted in Blu-ray Discs, internet source streaming, web software, etc.

**x265**: an open source library for encoding video streams into the H.265/HEVC compression format, and one of the several potential successors to H.264/AVC. It keeps the same video quality with about half the bitrate of H.264/AVC.

**vp8**: an open source and royalty free video compression format owned by Google and created by On2 Technologies, primarily designed for webpage videos.

**vp9**: a successor to vp8 and competes mainly with H.265/HEVC, widely supported by browsers on PC and mobile devices.

### 4.2. Dataset

This research has been approved by the Research Committee of National Chung Hsing University. The informed consent was obtained from each subject. 18 healthy subjects (14 males and 4 females, aged 21-35, averaged HR 79.75 bpm, East Asian skin tone) were recruited. A total number of 55 videos of duration 1 min were recorded using a regular webcam<sup>1</sup> and stored in uncompressed AVI format (640x480 pixels, 24 bit/pixel, 30 fps). Two motion types are considered: stationary and motion. The subjects were asked sitting still in front of the camera (with distance 0.5 m) while recording the first video and then moving left and right (with speed about 1 Hz) while recording the second video. The PPG signal measured by a fingertip pulse oximeter<sup>2</sup> was simultaneously recorded as the ground-truth for reference. The PPG signals were synchronized with the videos by setting a same timer. The videos were then compressed to 10 target bitrates evenly selected in the range [100, 20000] kb/s in log scale. Compression was performed using the latest FFmpeg, an integrated video processing library which includes libx264 for x264, libx265

<sup>1</sup>Model HD pro webcam c920, Logitech.

<sup>2</sup>Model CMS50E, Contec Medical.

for x265, and libvpx for vp8 and vp9. The compressed file size ranges from 157 MB (20000 kb/s) to 846 KB (100 kb/s) per file. The raw video information in the benchmark dataset is summarized in Table 2. Figure 6 plots the averaged compressed file size as a function of bitrate. It can be seen that the file size of the compressed videos has been significantly reduced compared with the uncompressed ones.

status	stationary	motion
# of videos	30	25
total duration (min)	33.75	28.36
frame rate (fps)	30	30
average file size (GB)	1.73	1.73

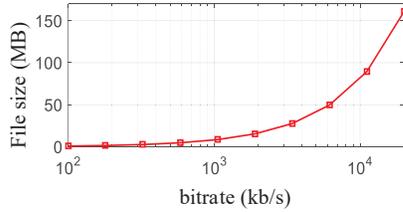


Figure 6. Averaged compressed file size as a function of bitrate.

### 4.3. Evaluation metrics

**Signal to noise ratio (SNR)** is widely adopted by existing rPPG methods [28, 17] to evaluate the quality of the rPPG signal in the extracted pulse waveform. A template window is first defined,

$$U(f) = \begin{cases} 1, & f_c - \frac{\omega_2}{2} \leq f \leq f_c + \frac{\omega_2}{2} \\ 1, & 2f_c - \frac{\omega_2}{2} \leq f \leq 2f_c + \frac{\omega_2}{2} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where  $f_c$  denotes the reference HR computed from the contact PPG signal using the same instantaneous HR calculation algorithm described in Subsection 3.5, and  $\omega_2$  denotes the spectral window length. SNR is defined as the ratio of the power in the template window to that outside the window.

$$\text{SNR} = 10 \log_{10} \left( \frac{\sum_{f=0.8}^5 U(f) S^2(f)}{\sum_{f=0.8}^5 (1 - U(f)) S^2(f)} \right) \quad (4)$$

where  $S(f)$  denotes the spectrum of the extracted pulse waveform.

**Precision** is used to evaluate the accuracy of the heart rate  $HR$  obtained from the extracted pulse waveform compared to the ground-truth heart rate  $HR_r$  from the contact PPG signal. Precision with a given threshold  $T$  is defined as:

$$p_T = \frac{\# \text{ of } \{t | \text{abs}(HR(t) - HR_r(t)) \leq T\}}{\text{total } \# \text{ of } t \text{ in } HR} \quad (5)$$

where  $\text{abs}(\cdot)$  returns the absolute value. The final Precision is the averaged result over four thresholds  $\mathcal{T} = \{0, 1, 2, 3\}$  bpm.

### 4.4. Compared methods

**CHROM** [5]: a chrominance-based method that builds upon the skin reflection model and the standardized skin-tone assumption.

**CorDiff** [6]: a color difference method that utilizes strength difference of pulse signal in color channels to eliminate motion artifacts.

**POS** [28]: a model-based method that projects the rPPG measurements to the ‘plane orthogonal to skin’ (POS) for pulse extraction.

In order to examine the performance of each step in the proposed framework, the following three combinations are considered.

**filter**: consists of only G-channel band-pass filtering and overlap adding.

**filter+SSA**: a combination of G-channel band-pass filtering, SSA RC selection, and overlap adding.

**filter+SSA+refine**: the proposed algorithm that consists of all the steps.

## 5. Results

The results for stationary and motion videos are shown, respectively, in Figures 7 and 8. Each figure contains SNR and Precision results as a function of bitrate with the compared methods for each compression method. The results on the raw videos are also plotted at the last column in each subfigure.

### 5.1. Performances on the stationary case

Figure 7 shows that the extracted signal quality of existing algorithms is getting worse as the bitrate decreases, which is in agreement with previous researches [8, 17], verifying that video compression has a detrimental effect on rPPG measurements and current techniques are ineffective. For example, the signal quality has a sharp drop at 2 Mb/s for x264 and vp9. x265 preserves the best pulsatile information among other compression methods. There is a linear relationship between bitrate and SNR for vp8.

The proposed algorithm (filter+SSA+refine) achieves a significant improvement in SNR and Precision when compared with existing algorithms. For x264, x265, and vp8, the Precision curve does not decrease until 320 kb/s, suggesting that the videos can be compressed to the bitrate as low as 320 kb/s to obtain the performance equivalent to that on the raw video. The Precision is above 60% at the lowest considered bitrate 100 kb/s whereas existing algorithms have Precision below 20% (increased by over 40%). Note that, Precision is the averaged results over four thresholds.

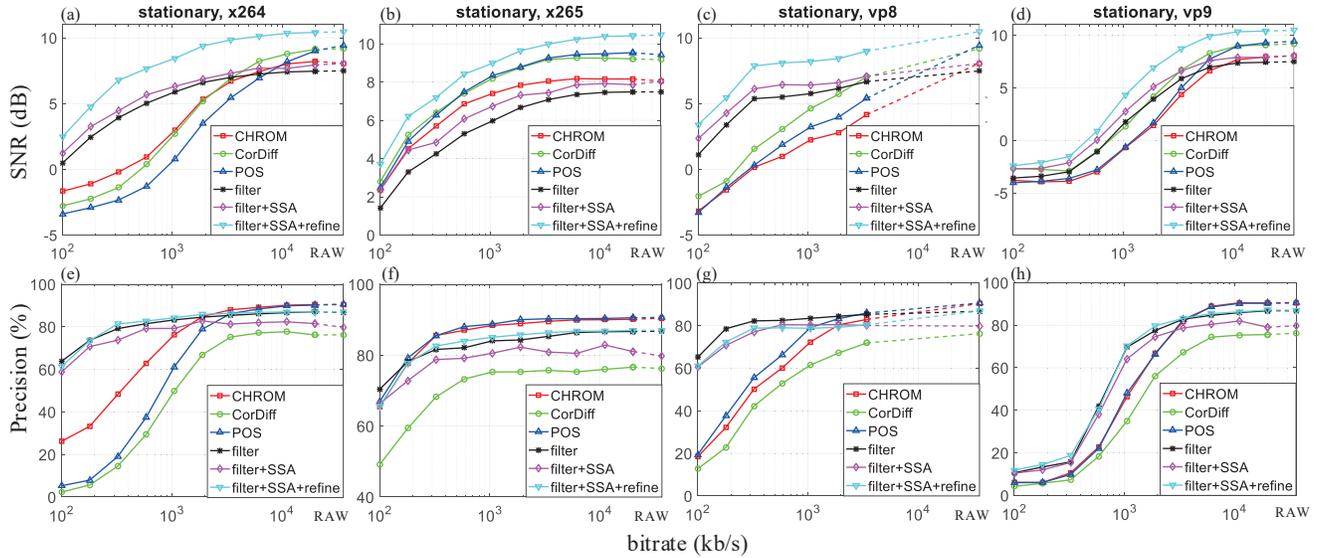


Figure 7. SNR and Precision results as a function of bitrate on stationary videos.

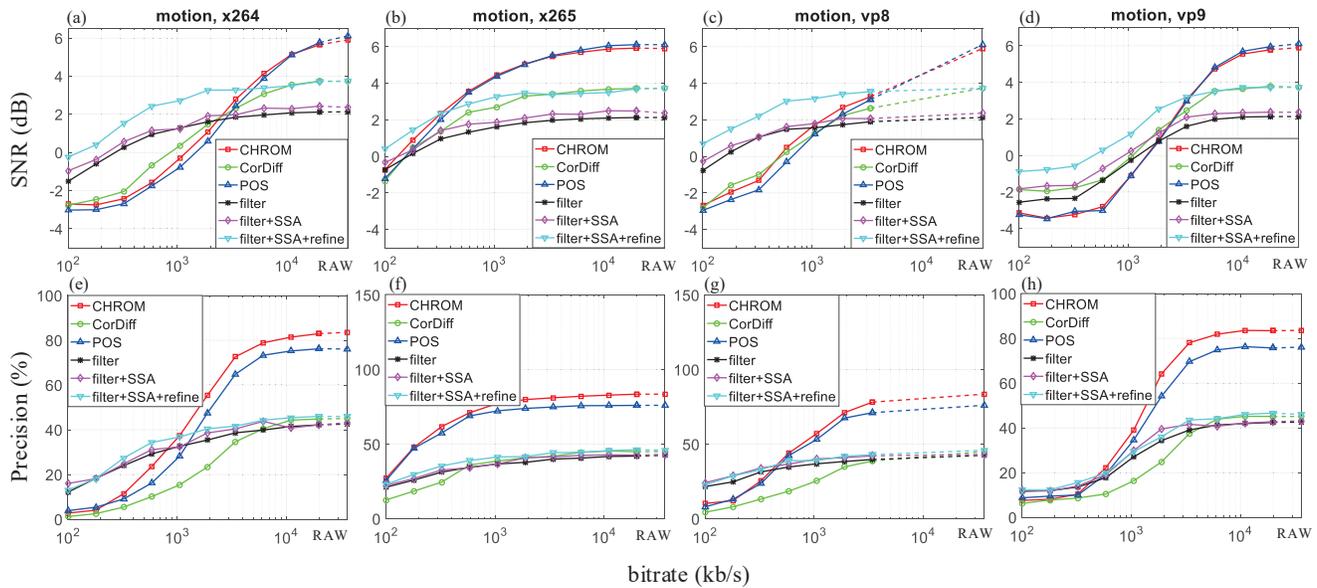


Figure 8. SNR and Precision results as a function of bitrate on motion videos.

The scores should be higher if only  $T = 3$  is considered. The results also demonstrate that all the rPPG algorithms are not effective on vp9 compressed videos, which implies that vp9 poses a serious threat to the physiological signals that there is little pulse signal remaining.

## 5.2. Performances on the motion case

Similar results can be observed from Figure 8 that a decrease in SNR and Precision of existing algorithms along with the decrease of bitrate, and that x265(vp9) pre-

serves the best(worst) signal quality among other compression methods. The proposed algorithm improves the SNR and Precision results at low bitrates. The performances at high bitrates are outperformed by CHROM and POS. This comes at no surprise as CHROM and POS employ multi-channel pulse extraction while the proposed algorithm employs single channel. At high bitrates, the VCAs become less significant and the MAs dominate the problem. The primary concern turns out to be the MA suppression, which is the main purpose of existing algorithms. The performance

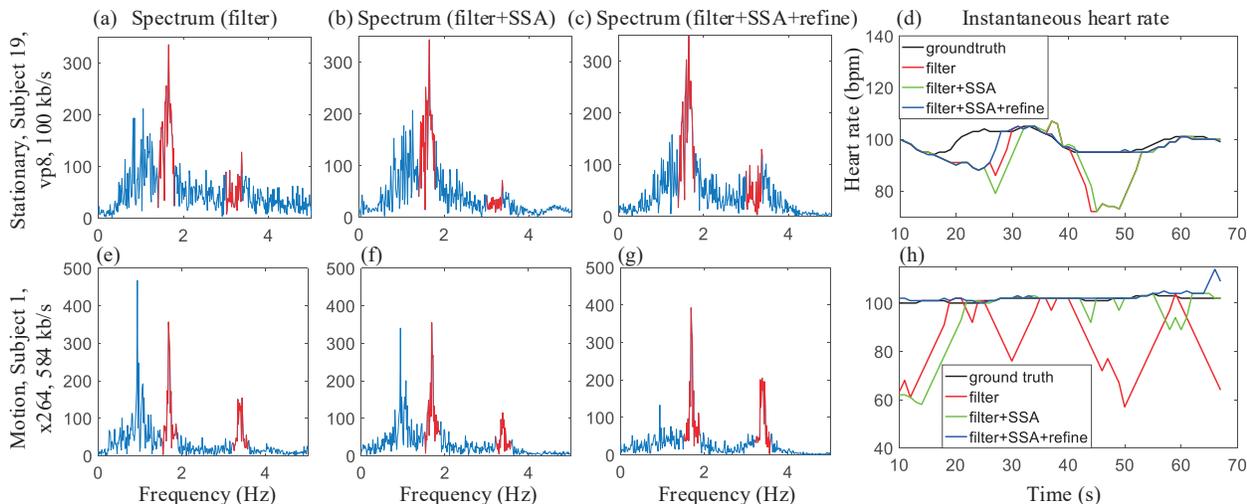


Figure 9. Comparison of three steps in the proposed framework. The first three columns plot the spectrum of output by each step. Red region denotes the HR and its first harmonic of the ground-truth pulse waveform measured by a fingertip pulse oximeter.

inferiority of the proposed method at high bitrates implies a weak MAs suppression ability by just exploiting rPPG frequency structure compared with existing methods. Nevertheless, one can still observe a performance improvement at low bitrates on x264, vp8, and vp9, demonstrating a VCAs suppression strength of the proposed algorithm over existing methods.

### 5.3. Effect of each step

Figure 9 depicts two examples illustrating the effect of each step in the proposed framework. For the stationary case (Figure 9, first row), the first step filters out signals with frequencies outside the considered region (Figure 9(a)). The second step (Figure 9(b)) highlights two peaks related to rPPG signals by removing low-frequency MAs. The last step retains the two peaks and removes more noise (Figure 9(c)). Figure 9(d) shows clearly the Precision increase step by step, in which the result given by the entire algorithm (blue line) is the closest to the ground-truth HR.

For the motion case (Figure 9, second row), there is an obvious peak to the left of the HR peaks (Figure 9(e)), which is associated with the MA. Its power has been reduced after the second step (Figure 9(f)). It almost disappears in the last step (Figure 9(g)) and the HR peaks become more obvious. The instantaneous HR comparison (Figure 9(h)) verifies this effect as the resulting HR curve goes from the motion frequency to the ground-truth HR.

## 6. Conclusion and discussion

The issue of pulse extraction on compressed videos is highlighted in this paper as a key factor for applying rPPG techniques into cases that need long-distance video trans-

mission. Three types of video compression artifacts are observed, *i.e.*, amplitude deterioration, high-frequency structured noise, and trace discontinuity. A novel framework based on single-channel signal separation is proposed as a preliminary attempt to address this problem. Three major steps constitute the proposed framework: 1) single-channel band-pass filtering to reduce the impact of VCAs; 2) SSA decomposition and RC selection to separate pulse signal from MAs; and 3) spectral masking to refine the output.

Extensive experiments with four popular compression methods and two motion types are conducted. The results show that the proposed approach significantly improves the performance of existing algorithms on stationary videos and has a positive effect on motion videos at low bitrates. The extracted pulse signal quality degrades along with the decrease of bitrates. x265 preserves the best signal quality and vp9 is the least effective. For stationary videos by x264, x265, and vp8, the proposed algorithm achieves equivalent result to the raw videos when the bitrate is as low as 320 kb/s. The proposed algorithm achieves more than 3 times improvement over existing algorithms at 100 kb/s.

Compared with existing rPPG algorithms which combine multi-channel signals, the single-channel signal processing usually suffers from motion artifact, as is reflected in the motion case. Future work will focus on algorithmic design for motion robustness. Moreover, the performance can be improved using ROI-based rate control [15, 14].

## Acknowledgement

This research was sponsored by National Natural Science Foundation of China under Grand No. 61620106012, 61573048, 61603020, 61773042.

## References

- [1] R. Aihara, T. Takiguchi, and Y. Ariki. Activity-mapping non-negative matrix factorization for exemplar-based voice conversion. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4899–4903, 2015.
- [2] A. Al-Naji, K. Gibson, S. H. Lee, and J. Chahl. Monitoring of cardiorespiratory signal: Principles of remote measurements and review of methods. *IEEE Access*, 5(99):15776–15790, 2017.
- [3] G. Balakrishnan, F. Durand, and J. Guttag. Detecting pulse from head motions in video. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pages 3430–3437, 2013.
- [4] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. S. Torr. Staple: Complementary learners for real-time tracking. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, June 2016.
- [5] G. de Haan and V. Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013.
- [6] L. Feng, L. M. Po, X. Xu, Y. Li, and R. Ma. Motion-resistant remote imaging photoplethysmography based on the optical properties of skin. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(5):879–891, 2015.
- [7] A. R. Guazzi, M. Villarroel, J. Jorge, J. Daly, M. C. Frise, P. A. Robbins, and L. Tarassenko. Non-contact measurement of oxygen saturation with an rgb camera. *Biomedical Optics Express*, 6(9):3320–3338, 2015.
- [8] S. Hanfland and M. Paul. Video format dependency of ppgi signals. In *Proceedings of the International Conference on Electrical Engineering*, 2016.
- [9] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. T. Chi, and H. H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings Mathematical Physical & Engineering Sciences*, 454(1971):903–995, 1998.
- [10] A. Lam and Y. Kuno. Robust heart rate measurement from video using select random patches. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3640–3648, 2015.
- [11] M. Lewandowska, J. Rumiński, T. Kocejko, and J. Nowak. Measuring pulse rate with a webcam - a non-contact method for evaluating cardiac activity. In *Proceedings of the Federated Conference on Computer Science and Information Systems*, pages 405–410, 2011.
- [12] X. Li, J. Chen, G. Zhao, and M. Pietikainen. Remote heart rate measurement from face videos under realistic situations. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pages 4264–4271, 2014.
- [13] S. Liu, P. C. Yuen, S. Zhang, and G. Zhao. 3d mask face anti-spoofing with remote photoplethysmography. In *Proceedings of the European Conference on Computer Vision*, pages 85–100, 2016.
- [14] Y. Liu, Z. G. Li, and Y. C. Soh. A novel rate control scheme for low delay video communication of h.264/avc standard. *IEEE Transactions on Circuits & Systems for Video Technology*, 17(1):68–78, 2006.
- [15] Y. Liu, Z. G. Li, and Y. C. Soh. Region-of-interest based resource allocation for conversational video communication of h.264/avc. *IEEE Transactions on Circuits & Systems for Video Technology*, 18(1):134–139, 2008.
- [16] K. Mannapperuma, B. D. Holton, P. J. Lesniewski, and J. C. Thomas. Performance limits of ica-based heart rate identification techniques in imaging photoplethysmography. *Physiological Measurement*, 36(1):67–83, 2014.
- [17] D. J. McDuff, E. B. Blackford, and J. R. Estepp. The impact of video compression on remote cardiac pulse measurement using imaging photoplethysmography. In *Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition*, pages 63–70, 2017.
- [18] D. J. McDuff, J. R. Estepp, A. M. Piasecki, and E. B. Blackford. A survey of remote optical photoplethysmographic imaging methods. In *Proceedings of the 37th Annual International Conference of the IEEE*, pages 6398–6404, 2015.
- [19] D. J. McDuff, M. Z. Poh, and R. W. Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express*, 18(10):10762–10774, 2010.
- [20] A. Moço, L. Mondragon, W. Wang, S. Stuijk, and G. de Haan. Camera-based assessment of arterial stiffness and wave reflection parameters from neck micro-motion. *Physiological Measurement*, 38(8):1576–1598, 2017.
- [21] M.-Z. Poh, D. J. McDuff, and R. W. Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering*, 58(1):7–11, 2011.
- [22] G. Rilling and P. Flandrin. One or two frequencies? The empirical mode decomposition answers. *IEEE Transactions on Signal Processing*, 56(1):85–95, 2008.
- [23] A. Sikdar, S. K. Behera, and D. P. Dogra. Computer-vision-guided human pulse rate estimation: A review. *IEEE Reviews in Biomedical Engineering*, 9:91–105, 2016.
- [24] Y. Sun and N. Thakor. Photoplethysmography revisited: From contact to noncontact, from point to imaging. *IEEE Transactions on Biomedical Engineering*, 63(3):463–477, 2016.
- [25] L. Tarassenko, M. Villarroel, A. Guazzi, J. Jorge, D. A. Clifton, and C. Pugh. Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. *Physiological Measurement*, 35(5):807–831, 2014.
- [26] R. Vautard, P. Yiou, and M. Ghil. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D: Nonlinear Phenomena*, 58(1):95–126, 1992.
- [27] S. Wang, H. L. Tang, Y. Hu, S. Sanei, G. M. Saleh, T. Peto, et al. Localizing microaneurysms in fundus images through singular spectrum analysis. *IEEE Transactions on Biomedical Engineering*, 64(5):990–1002, 2017.
- [28] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan. Algorithmic principles of remote ppg. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2016.

- [29] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan. Robust heart rate from fitness videos. *Physiological Measurement*, 38(6):1023C1044, 2017.
- [30] W. Wang, S. Stuijk, and G. De Haan. Exploiting spatial redundancy of image sensor for motion robust rppg. *IEEE Transactions on Biomedical Engineering*, 62(2):415–425, 2015.
- [31] W. Wang, S. Stuijk, and G. de Haan. Living-skin classification via remote-ppg. *IEEE Transactions on Biomedical Engineering*, 64(12):2781–2792, 2017.
- [32] W. Wang, S. Stuijk, and G. D. Haan. A novel algorithm for remote photoplethysmography: Spatial subspace rotation. *IEEE Transactions on Biomedical Engineering*, 63(9):1974–1984, 2016.
- [33] S. Xu, L. Sun, and G. K. Rohde. Robust efficient estimation of heart rate pulse from video. *Biomedical Optics Express*, 5(4):1124–1135, 2014.
- [34] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and L. Zhang. Robust principal component analysis with complex noise. In *Proceedings of the International Conference on Machine Learning*, pages 55–63, 2014.