Mathematical Biosciences and Engineering

MBE, 19(5): 5153–5168.
DOI: 10.3934/mbe.2022241
Received: 21 January 2022
Revised: 08 March 2022
Accepted: 09 March 2022
Published: 21 March 2022

http://www.aimspress.com/journal/MBE

Research article

Shielding facial physiological information in video

Kun Zheng, Junjie Shen, Guangmin Sun, Hui Li* and Yu Li*

Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

* **Correspondence:** Email: lihui@bjut.edu.cn, yuli@bjut.edu.cn; Tel: +8618810899182, +8613811136728.

Abstract: With the recent development of non-contact physiological signal detection methods based on videos, it is possible to obtain the physiological parameters through the ordinary video only, such as heart rate and its variability of an individual. Therefore, personal physiological information may be leaked unknowingly with the spread of videos, which may cause privacy or security problems. In this paper a new method is proposed, which can shield physiological information in the video without reducing the video quality significantly. Firstly, the principle of the most widely used physiological signal detection algorithm: remote photoplethysmography (rPPG) was analyzed. Then the region of interest (ROI) of face contain physiological information with high signal to noise ratio was selected. Two physiological information forgery operation: single-channel periodic noise addition with blur filtering and brightness fine-tuning are conducted on the ROIs. Finally, the processed ROI images are merged into video frames to obtain the processed video. Experiments were performed on the VIPL-HR video dataset. The interference efficiencies of the proposed method on two mainly used rPPG methods: Independent Component Analysis (ICA) and Chrominance-based Method (CHROM) are 82.9 % and 84.6 % respectively, which demonstrated the effectiveness of the proposed method.

Keywords: face video; non-contact physiological signal detection; physiological information shielding; forgery

1. Introduction

Face video contains abundant basic personal information, such as identity information, emotional information and physiological information, as shown in Figure 1. With the spread of video,

the personal privacy information is easy to leak.

In order to protect identity information, a large number of face forgery methods have been proposed. There are two main fields in forging face videos, namely, forgery based on graphics and forgery based on learning. In the first field, forgery based on graphics, Garrido utilized 3D method to replace the actor's face in the target video [1], while retaining the original expression. In the second field, Shan developed Fawkes [2], which can change the feature space representation of an image by perturbation, so as to protect our personal identity and other information effectively. Nirkin proposed an independent face replacement and reconstruction method based on GAN technology [3]. Antipov used conditional-GAN technology to change people's ages [4]. Huang generated different facial views by means of GAN while global structure and local details were preserved [5]. In the aspect of face video forgery detection, people often ignore the real physiological characteristics of human beings in the forged video, and it cannot be consistent with real people in general. And some researchers begin to study the methods based on physiological signal characteristics. Ciftci et al. extracted three regions from the face to measure the pulse wave signal [6], and converted the signal into consistent and coherent features. Fernandes et al. utilized heart rate biological signals to distinguish forged videos [7]. Most of the previous face video forgery methods concentrates on the forgery of identity information, emotion, and age. However, according to our knowledge, the forgery of physiological parameters is still an untouched research field, leaving the personal physiological privacy unprotected.

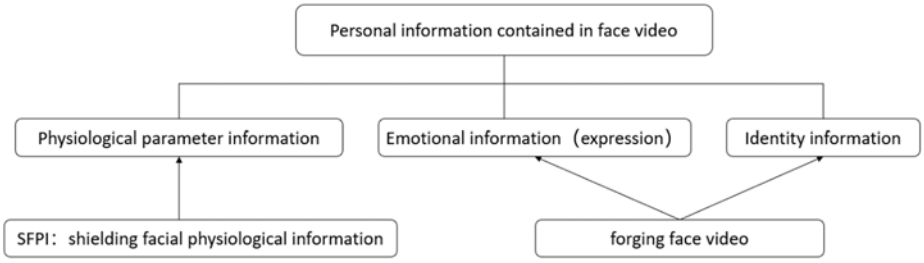

Figure 1. Personal information contained in face video.

With the rapid development of remote health monitoring [8,9] and telemedicine technology, the advantages of remote photoplethysmography (rPPG) has been proved in the detection of physiological parameters from video. In 2010, Poh proposed blind source separation (BSS) to reduce the errors caused by motions [10]. It has been improved that the cleaner PPG signals can be obtained by adding several time filters before and after ICA [11]. De Haan analyzed the limitations of blind source separation to solve motion problems, and proposed a chrominance-based method [12]. Although rPPG technology could achieve remote heart rate measurement quickly, the personal physiological information may be leaked, analyzed and utilized with the spread of video. At present, independent component analysis (ICA) and chrominance-based method (CHROM) are mainly utilized to measure physiological parameters at home and abroad. The physiological parameters can be obtained accurately by analyzing the common face video through the two methods, such as heart rate and heart rate variability. Prakash proposed a bounded Kalman filter for motion estimation and feature tracking to reduce the influence of motion artifacts [13], which a mean error of ± 3 bpm was attained. Yang proposed a framework based on patch fusion for face motion [14], which can estimate

the accurate heart rate from face video. The RMSE of this method is only 2.91 bpm on motion videos. Qiu combined spatiotemporal filtering with convolutional neural network to realize remote measurement of heart rate in practical situations [15]. 69 % of the HR changes are predicted correctly. Niu proposed an end-to-end RhythmNet for remote estimation of HR from the face [16], with RMSE below 8.14 bpm. Zheng proposed a symmetrical substitution method in the case of partial facial missing [17], with RMSE below 7.64 bpm. Based on the traditional methods, combined with skin segmentation and face detection, the measurement error of heart rate can be controlled within 10 bpm [17]. Face detection based on deep learning [18] has better effect and can provide support for more accurate heart rate detection. In good conditions, the measurement error can be even controlled within 0.4 bpm [12]. Therefore, it is a challenging task to protect the facial physiological information without largely alternating the video quality.

This paper provides a method for shielding facial physiological information (SFPI) without changing the visual effect significantly, whose key steps is shown in Figure 2. On the basis of fully considering the principle and detection process of rPPG, the non-contact physiological signal detection is interfered by using single channel periodic blur filtering and single channel periodic noise addition combined with brightness fine-tuning method. It can resist wildly used physiological signal detection methods efficiently such as ICA and CHROM, and protect personal physiological privacy.

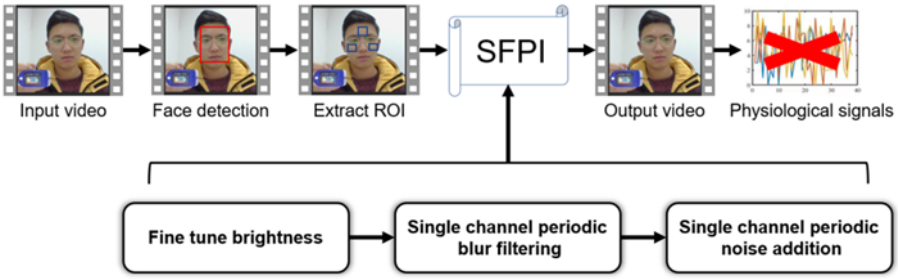

Figure 2. The method for shielding facial physiological information.

2. Methods

The flow chart of the method proposed in this paper is shown in Figure 3. This algorithm can shield the facial physiological information and doesn't change the video effect.

2.1. Video frame processing

Firstly, face detection is performed on the selected video frame by frame. Secondly, the feature face key points detection based on the Gradient Boosting Decision Tree [19] is used to realize the 68 face feature points marking, which can automatically seek the landmark feature positions of eyes, nose, mouth and face contour on the basis of the detected face.

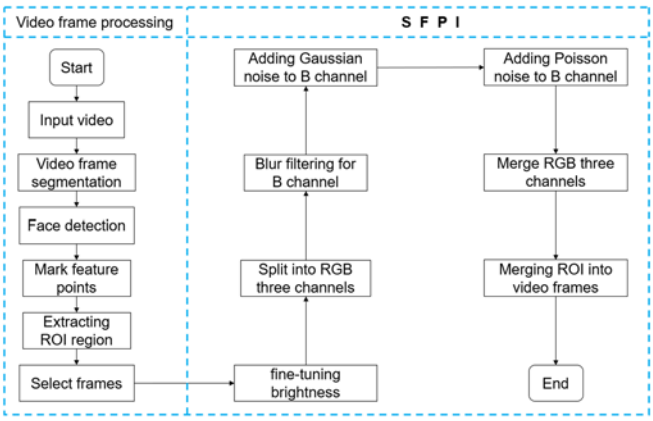

Figure 3. The flow chart for shielding facial physiological information.

2.2. ROI determination

The selection of reliable ROI is the key to extract physiological parameters based on rPPG method, and blood volume pulse (BVP) signal of pulse wave related to physiological signal needs to be extracted by physiological parameter detection from ROI. Therefore, after video capture, face detection and feature point marking, it is necessary to select ROI of each face frame for processing. In the process of extracting physiological parameters based on rPPG method, due to the existence of non-rigid motion of human face, such as blinking, expression changes and speech, noise is inevitably added to the face. Therefore, it is necessary to select the face region which does not contains these places as many as possible to reduce the interference of noise on useful signals. It is shown that the smaller ROI determined from the left and right cheek and forehead regions has a higher signal-to-noise ratio [13]. Therefore, the left cheek, right cheek and forehead were selected as regions of interest, and the final position was determined by referring to 68 feature points. Three ROIs are selected for subsequent processing.

2.3. Brightness fine-tuning

Adjusting the brightness or contrast of local areas of video images before outputting video will change the original pixel value of the skin area and have a significant impact on the measurement of physiological indexes. Therefore, brightness fine-tuning is applied by the proposed method for shielding facial physiological information.

To select the optimum interframe space (IFS) period, we conducted an experiment to adjust the brightness or add noise to the same video with different IFS periods. Figure 4 shows the interference rate of physiological parameter detection caused by different IFS. If the difference between the measured heart rate and the original video heart rate is greater than 10, it has played an effective interference. It is found that the effective rate of interference to physiological parameter detection is 78.2 % when the video is processed once every 12 frames, the highest interference efficiency is obtained. Therefore, in this study 12 is selected as the IFS period for processing, which not only obtain the best interference effect, but also greatly improve the efficiency of video processing.

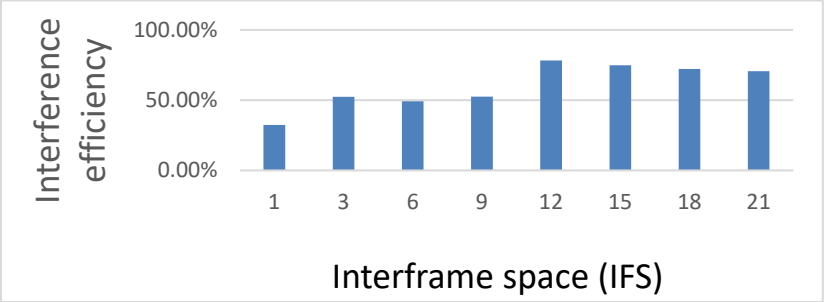

Figure 4. Interference rate of different IFS on physiological parameter detection.

In the process of extracting physiological parameters based on rPPG method, it is necessary to separate the RGB channels of the three ROI regions in each frame, and calculate the spatial average values of the three channels to r_i, g_i, b_i respectively. Then the whole video sequence is converted into three one-dimensional signals. The gray mean value will be changed by slight modification of the image brightness, the original signal will be affected and the detection of physiological parameters will be interfered. Therefore, the contrast and brightness of the image is changed pixel by pixel firstly, the formula is as follows:

$$G(x, y) = \alpha f(x, y) + \beta \tag{1}$$

where $f(x, y)$ represents the value of channel c of the pixels in x rows and y columns of the source image, $g(x, y)$ represents the value of channel c of pixels in x rows and y columns of the target image, α is the value of contrast, β is the value of the brightness modification. The brightness of the image can be modified by adding β to all channels of each pixel. In this paper, the value of β is 1. When β = 1, this step has the least impact on the video visual effect. The mathematical expression of periodic brightness adjustment for video is as follows:

$$Brightness\ (Frame_{12n}) = Brightness + \beta \tag{2}$$

where $\beta = 1$ is the value of the brightness modification. $n = 0, 1, 2, 3, 4, 5\ldots\ldots$ $Frame_{12n}$ represents the first image selected every 12 frames, *Brightness* represents the brightness of the source image, $Brightness\ (Frame_{12n})$ represents the brightness of the first frame selected every 12 frames after fine-tuning. $Bri\ (Frame_{12n})$ is the first frame of every 12 frames after brightness adjustment. After this processing, the brightness of the image frame can be changed periodically.

Natural Image Quality Evaluator (NIQE) [22] is utilized to evaluate the video quality after brightness change. As shown in Figure 5, the blue dot represents the NIQE value of the original video, and the red dot represents the NIQE value of the video after brightness fine-tuning. The coincidence degree of the two parts is high, therefore the objective quality of the video processed has not been changed greatly in this step.

2.4. Single channel periodic blur filtering

The blur filtering of image frame can change some image features and their gray mean value, which can also affect the original signal. However, if the image is blur filtered directly, it will have a serious impact on the visual effect of the image. In this paper, firstly, we segment the image channel.

Secondly, the segmented B-channel is blur filtered. Finally, the RGB channels are merged. It has no change in the visual effect, and it can effectively interfere with the accuracy of physiological parameter detection.

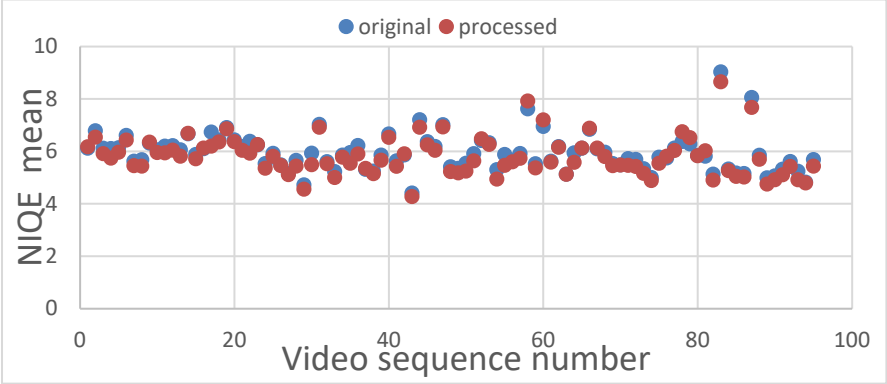

Figure 5 . NIQE mean. The blue dot represents the NIQE value of the original video, and the red dot represents the NIQE value of the video after brightness fine-tuning.

2.5. Single channel periodic noise addition

If the noise is added to the image directly, every frame after processing will produce obvious noise stains. After a lot of experiments, it is found that if the noise is added to the single channel of the image, the visual effect of the processed image has no obvious change. Among them, if the R channel or G channel is added with noise in a single channel, the video modification traces after processing are more obvious, while the video effect of B channel is almost unchanged, and it can effectively interfere with the physiological parameter detection system.

The result of the experiment is illustrated that Gaussian noise with expected value of 0 and variance of 0.01 and Poisson noise with expected value of 1 have the most obvious disturbance on the method of extracting physiological parameters based on rPPG, and have little influence on the video effect. Therefore, Gaussian noise and Poisson noise are added to the selected video frame.

To sum up, the single channel noise adding method is as follows:

Firstly, the image is segmented to prepare for single channel processing. Secondly, the B-channel of the first frame of every 12 frames is extracted and Gaussian noise and Poisson noise are added. Finally, the RGB channels are merged.

The mathematical expression of single channel periodic noise adding for video is as follows:

$$Noise\ (Frame_{12n}) = Noise_{gaussian} + Noise_{poisson} \tag{3}$$

where n = 0, 1, 2, 3, 4, 5…… $Noise_{gaussian}$ represents the added Gaussian noise, $Noise_{poisson}$ represents the added Poisson noise, $Noise\ (Frame_{12n})$ represents the first frame of video after adding Gaussian noise and Poisson noise every 12 frames.

NIQE is used to evaluate the video quality after adding noise to a single channel. As shown in Figure 6, the blue dot represents the NIQE value of the original video, and the red dot represents the NIQE value of the video processed in this step. 22.3 % of the dots have a large gap between the original video and the processed video. It is shown that the objective quality of video after adding

noise to a single channel will be affected. However, conducting single channel filtering before can reduce the loss of video quality caused through this step.

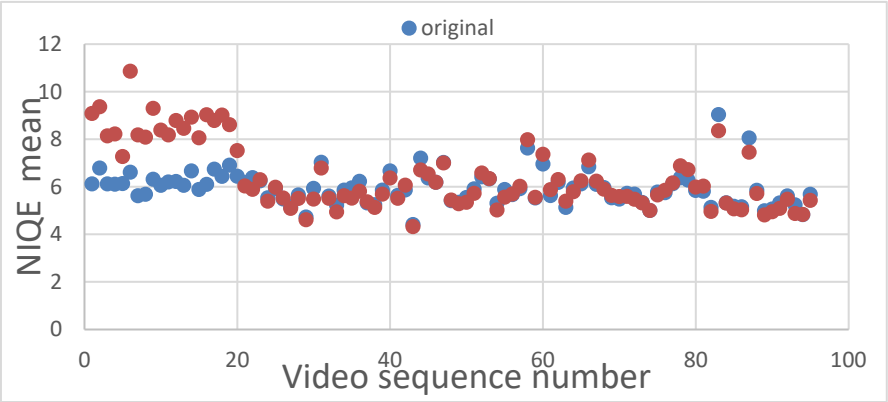

Figure 6 . NIQE mean. The blue dot represents the NIQE value of the original video, and the red dot represents the NIQE value of the video processed in this step.

2.6. ROI images merging

The processed ROI images are merged into video frames, and finally all the image frames are combined into face video. To sum up, the complete mathematical expression of this method is as follows:

$$SFPI(Frame_{12n}) = Bri\ (Frame_{12n}) + Blur\ (Frame_{12n}) + Noise\ (Frame_{12n}) \quad (4)$$

where n = 0, 1, 2, 3, 4, 5…… Then the expression of the first frame of every 12 frames after brightness adjustment is $Bri\ (Frame_{12n})$. $Blur\ (Frame_{12n})$ represents the image frame after blur filtering the first frame every 12 frames. $Noise\ (Frame_{12n})$ represents the first frame of video after adding Gaussian noise and Poisson noise every 12 frames. $SFPI\ (Frame_{12n})$ represents the first frame of video processed by the method in this paper every 12 frames.

3. Experiment results

3.1. Effect on shielding heart rate information

In order to verify the efficiency of the proposed method, the still video data set of VIPL-HR [16] dataset is used. The effective rates of interference against ICA and CHROM are 82.9 % and 84.6 % respectively, the detailed results are shown in Figure 7. Using Cohface [20] data set to verify the proposed method, the effective rates of interference against ICA and CHROM are 90.3 % and 89.6 % respectively, the specific results are shown in Figure 8. As can be illustrated from Figures 7 and 8, the method of fine-tuning brightness used only can cause weak interference effect. The interference effect of blur filtering used only or adding noise are not obvious enough. However, integrating these different video processing methods will cause effective interference to the detection of physiological parameters.

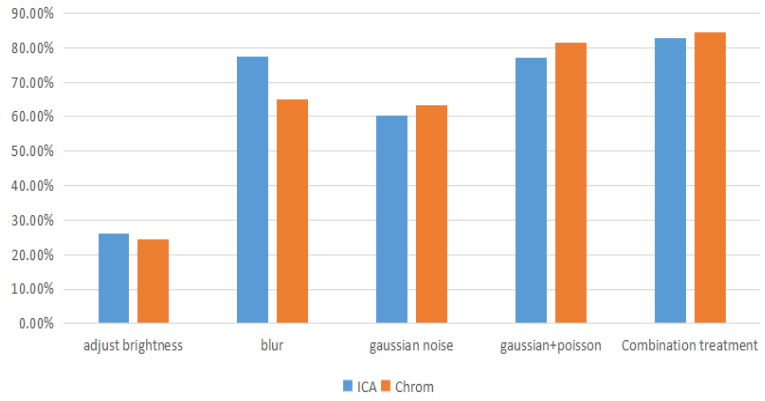

Figure 7. The interference efficiency of ICA and CHROM using VIPL-HR dataset.

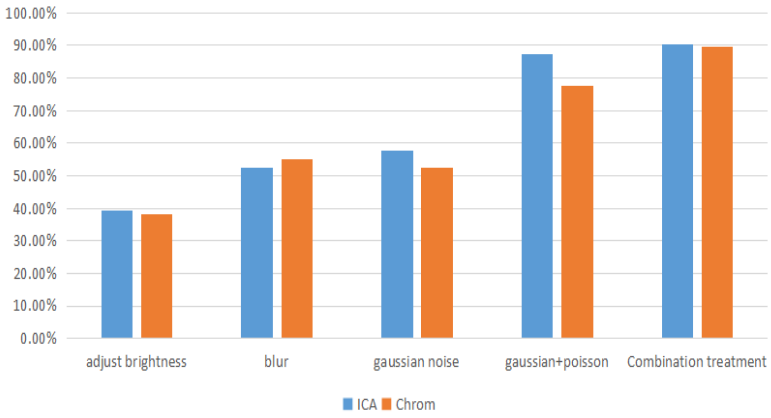

Figure 8. The interference efficiency of ICA and CHROM using Cohface database.

To evaluate the interference effect of this method on heart rate detection, we use Root Mean Square Error (RMSE). A series of analyses are carried out on the heart rate measured by the dataset video before and after processing. The measured heart rate RMSE is shown in Figure 9. The RMSE results of ICA and CHROM are 60.92 and 52.73 respectively. From the RMSE results, it is shown that the heart rate measured by the face video after processing by this method deviates greatly from the results before processing. The interference effect of this method on heart rate detection is significant.

Furthermore, Mean Absolute Error (MAE) is utilized to verify the interference effect of this method on the detection of heart rate by ICA and CHROM directly. MAE represents the average of the absolute error between the heart rate measured in the original video and the video processed by this method. For ICA method, the MAE is 47.7. For CHROM method, the MAE is 41.9. In terms of heart rate detection, the general allowable error is ± 12, and the MAE measured by this method far more than 12. By calculating MAE, it is illustrated intuitively that the heart rate detection is obviously disturbed.

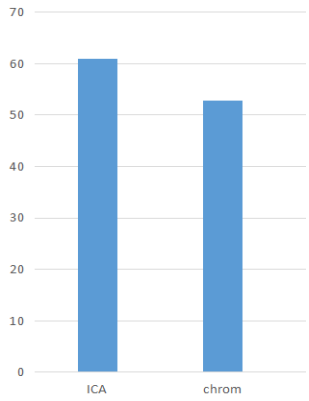

Figure 9. The heart rate RMSE measured by the VIPL-HR dataset video before and after processing.

To prove the superiority of the combination of the three processing methods in this paper. The MAE of heartrate extraction from the video data set processed by three methods are measured, which are SFPI without brightness fine tuning, SFPI without blur filtering and SFPI without noise addition respectively. The experimental results are shown in Figure 10. The MAE of video processed by SFPI is higher than the above three methods, therefore our method has the most obvious interference effect on the detection of physiological parameters.

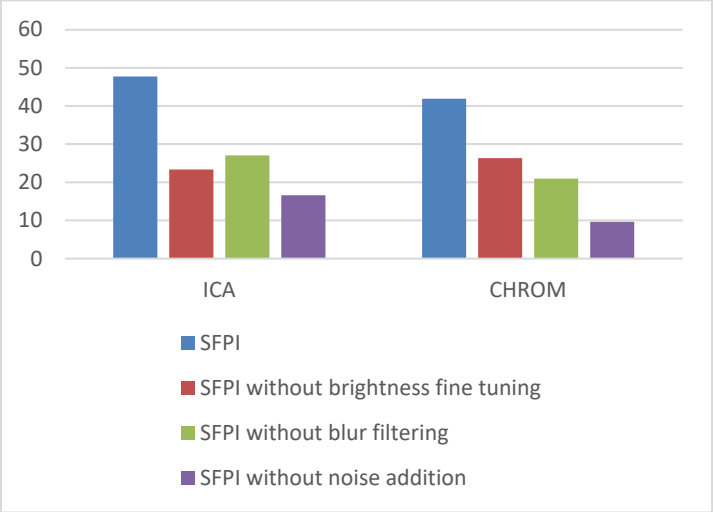

Figure 10. The heart rate MAE measured from the VIPL-HR database video.

The MAE of heartrate extraction from the video data set processed by SFPI with different types of noise is measured as well. The comparative experimental results are shown in Figure 11. It is illustrated that the MAE of the salt noise is higher than the MAE of the Gaussian noise and Poisson noise selected in this paper. But the salt noise has a significant impact on visual effect, as shown in Figure 12(c), while the noise selected in this paper has little impact on visual effect, as shown in Figure 12(b). Therefore, considering both interference effect and visual effect, Gaussian noise and Poisson noise are selected.

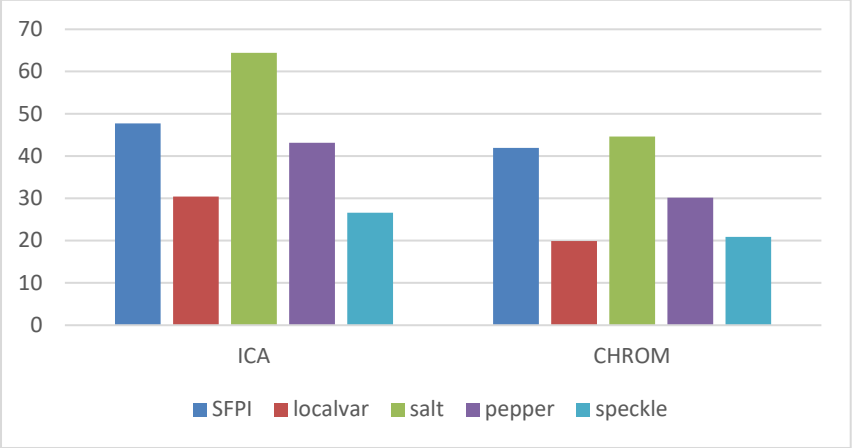

Figure 11. The heart rate MAE measured from the VIPL-HR dataset video processed by SFPI with different types of noise.

Figure 12. The visual effect of videos. (a) the visual effect of the original video. (b) the visual effect of the video processed by SFPI. (c) the visual effect of the video processed by SFPI with salt noise.

To verify the effect of the proposed method, the expected Bland Altman graph is plotted. As shown in Figure 13, the blue horizontal solid line in the middle represents the average value of the difference. The higher the consistency of the two groups of data, the closer the average value of the difference is to 0. The greater the difference between the two figures and the average value, the lower the correlation between the two groups of data. In this experiment, it is proved that the difference between the heart rate value estimated from the video processed by this method and the real heart rate value is great. The obvious difference between Figure 13(a) and (b), as well as between (c) and (d), shows that this method can shield heart rate detection effectively.

We further investigate the sample distribution of the interference effect of our method. The number of samples whose measurement error is less than a specific threshold is shown in Figure 14. As shown in this figure, the number of samples with ICA and CHROM original measurement errors which are less than n (n = 5, 10, 15) is more than the number of samples from the processed video. The number of samples whose measurement error is less than each specific threshold significantly decreased. This method increases the error of heart rate measurement greatly. After the face video is processed by this method, the heart rate measured from the processed video is inaccurate, and only a small part of the heart rate error is less than a specific threshold.

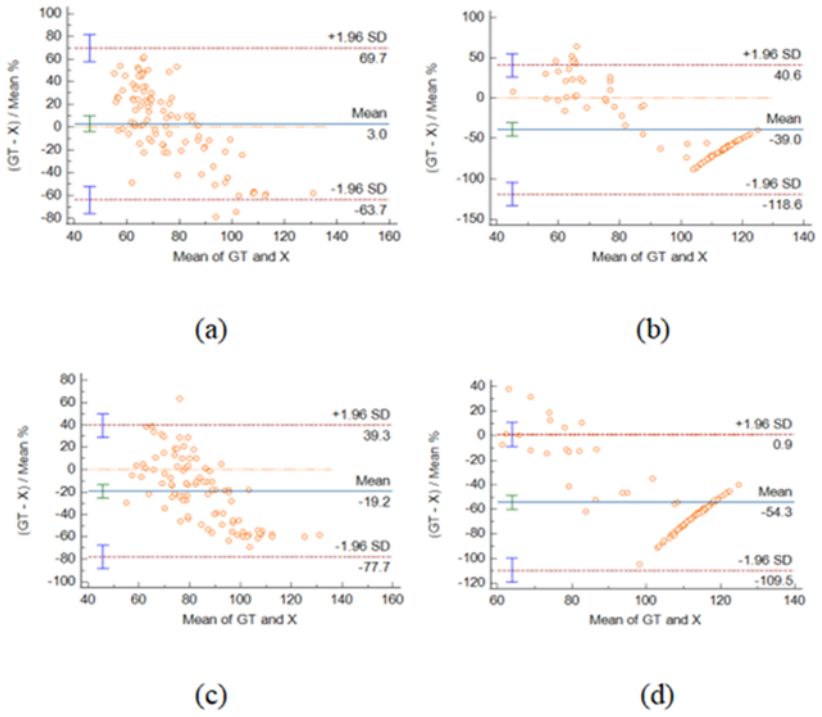

Figure 13. The Bland-Altman diagram of measured and real values. (a) is the Bland-Altman diagram of ICA detecting the heart rate of unprocessed face video, (b) is the Bland-Altman diagram of ICA detecting the heart rate of processed face video, (c) is the Bland-Altman diagram of CHROM detecting the heart rate of unprocessed face video, and (d) is the Bland-Altman diagram of CHROM detecting the heart rate of processed face video.

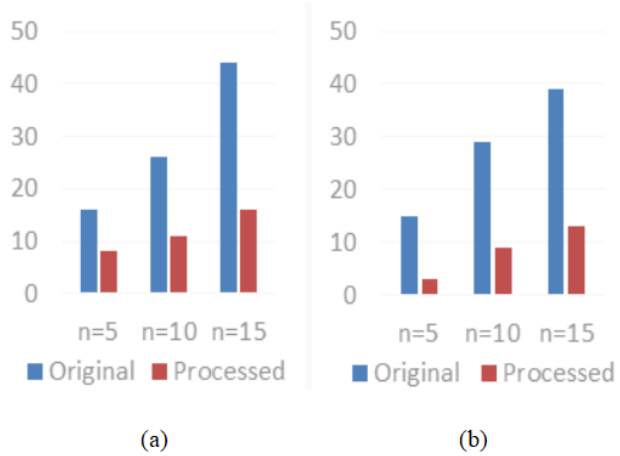

Figure 14. Number of samples with (a) ICA and (b) CHROM measurement errors which are less than a specific threshold.

3.2. Video quality assessment

3.2.1 Objective quality assessment

The goal of this method is to shield facial physiological information, but the premise is not to change the visual effect and objective quality of video. Therefore, Structural Similarity (SSIM) [21] and Natural Image Quality Evaluator (NIQE) [22] are introduced to evaluate the quality of the video processed objectively through this method.

1) SSIM

SSIM is an index to measure the similarity of two images. The mean is used as the estimate of brightness, the standard deviation as the estimate of contrast, and the covariance as the measure of structural similarity. The range of structural similarity is -1 to 1. The higher the similarity of the two images, the closer the SSIM value is to 1. When two images are identical, the value of SSIM is equal to 1.

2) NIQE

In terms of image quality evaluation, some common standards are Peak Signal to Noise Ratio (PSNR) and SSIM, but for super score or other low-level visual tasks image evaluation, these indicators do not meet our human senses, so NIQE came into being. NIQE index is an objective evaluation index, which extracts the features in the natural landscape to test the image. These features are fitted into a multivariate Gaussian model. This model is a measure of the difference in the multivariate distribution of an image to be tested. This distribution is constructed by these features extracted from a series of normal natural images.

NIQE is used to evaluate and compare the video quality before and after processing. The image quality of the video in the dataset is evaluated. The results are shown in Figure 15. The blue dot represents the NIQE value of the original video, and the red dot represents the NIQE value of the video processed by the method in this paper. The NIQE dot of the unprocessed dataset video coincides with the processed. No significant difference is observed between them, which shows that the objective quality of the processed video has not changed greatly.

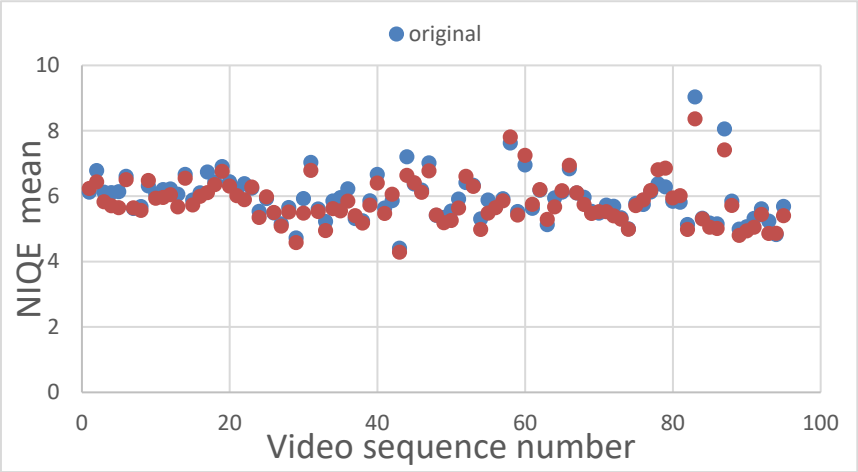

Figure 15. NIQE mean. The blue dot represents the NIQE value of the original video, and the red dot represents the NIQE value of the video processed by SFPI.

In order to conduct the objective quality evaluation and comparison of the video processed, SSIM is also utilized, which is an index to measure the similarity of two images. The video in the data set is processed through this method, and SSIM analysis is achieved on the processed video frame by frame. Then the SSIM values measured frame by frame for each group of videos is averaged to obtain the SSIM values of this group of videos. The SSIM measurement results of the whole video dataset are shown in Figure 16. It is illustrated that the SSIM value of each group of videos before and after processing is greater than 0.9. Therefore the video processed by this method has high similarity with the original video. It is proved that the method in this paper has no great impact on the objective quality of video.

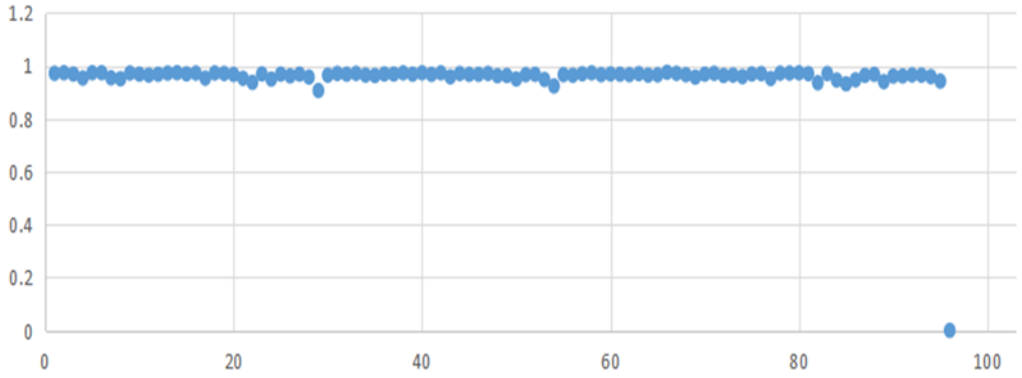

Figure 16. The SSIM measurement results between the original video and the processed video using VIPL-HR dataset.

3.2.2. Visual effect evaluation

We invited 15 colleagues in the laboratory to help us evaluate the subjective quality of video. We selected ten groups of videos as the test set, of which nine groups of videos are composed of the original video and the video processed by this method, and one group of videos is composed of the original video and the video after direct noise processing as the control group. The testers choose a group of videos from ten groups of videos that they can distinguish visual differences. In the end, fifteen testers could only identify the control group from ten groups of videos. In this experiment, 100 % of the testers cannot see the difference between the video processed by this method and the original video. The experimental results are shown that the visual effect of the video processed by this method has not changed obviously.

3.3. Supplementary experiment

In order to further verify the interference effect of this method on physiological parameter detection, PFF [23] data set is used to verify this method. The effective rates of interference against ICA and CHROM are 86.3 % and 82.6 % respectively. The PURE [24] data set is used to verify the proposed method. The effective rates of interference against ICA and CHROM are 92.2 % and 85.3 % respectively.

4. Discussion

We found that when videos are processed periodically, the original signal will have periodic abrupt peaks, which can make the heart rate detection system output numerically controllable heart rate values. When IFS period T is within 8 to 19, the heart rate detection result of heart rate detection system based on rPPG meets the following formula:

$$PR = (F/T) * 60 \tag{5}$$

where PR is the heart rate detection result, F is the frame rate of the input video, and T is the IFS period. However, this operation cannot stably output controllable values, and further research is still needed.

Most face videos processed by this method can shield physiological information and interfere with the heart rate detection results without affecting the visual effect significantly and reducing the video quality. However, for few individual videos, the effect of shielding heart rate by this method is not stable. Due to the single channel processing of the image, the correlation analysis in the process of heart rate detection is affected. More digital image processing methods can be tried in future research and combine with this method to improve the efficiency of shielding physiological information. Considerably more work, hopefully, will be done in this area.

5. Conclusions

A method of shielding facial physiological information without changing the effect of videos is proposed in this paper. Combined with the principle of rPPG, the problem that personal physiological information is leaked, analyzed and utilized with the spread of video has been solved preliminarily. The method can shield the physiological information without reducing the video quality significantly, resist the traditional physiological signal detection methods such as ICA and CHROM, and protect the personal physiological privacy information. The main contributions of this paper are as follows: firstly, the experiments show that the effect of interfering with heart rate detection is the most obvious while the video is processed every 12 frames. Secondly, a B-channel periodic blur filtering and noise adding method is proposed to resist the traditional physiological signal detection methods such as ICA and CHROM. And they have little impact on the objective quality of video. Thirdly, blur filtering is selected for the processing of B-channel, and the interference effect of Gaussian and Poisson mixed noise on heart rate detection is the most obvious.

Acknowledgments

This work was supported in part by Scientific Research Plan of Beijing Education Commission (SZ202110005002), CN.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. P. Garrido, L. Valgaerts, O. Rehmsen, T. Thormaehlen, P. Perez, C. Theobalt, Automatic face reenactment, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2014), 4217–4224. https://doi.org/10.1109/CVPR.2014.537
2. S. Shan, E. Wenger, J. Zhang, H. Li, H. Zheng, B. Y. Zhao, Fawkes: Protecting privacy against unauthorized deep learning models, in *Proceedings of 29th USENIX Security Symposium*, (2020), 1589–1604. https://dblp.org/rec/conf/uss/ShanWZLZZ20
3. Y. Nirkin, Y. Keller, T. Hassner, Fsgan: Subject agnostic face swapping and reenactment, in *Proceedings of the IEEE/CVF international conference on computer vision*, (2019), 7184–7193. https://doi.org/10.1109/ICCV.2019.00728
4. G. Antipov, M. Baccouche, J. L. Dugelay, Face aging with conditional generative adversarial networks, in *Proceedings of IEEE international conference on image processing*, (2017), 2089–2093. https://doi.org/10.1109/ICIP.2017.8296650
5. R. Huang, S. Zhang, T. Li, R. He, Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis, in *Proceedings of the IEEE international conference on computer vision*, (2017), 2439–2448. https://doi.org/10.1109/ICCV.2017.267
6. U. A. Ciftci, I. Demir, L. Yin, Fakecatcher: Detection of synthetic portrait videos using biological signals, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020. https://doi.org/10.1109/TPAMI.2020.3009287
7. S. Fernandes, S. Raj, E. Ortiz, I. Vintila, M. Salter, G. Urosevic, et al., Predicting heart rate variations of deepfake videos using neural ode, in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, (2019), 1721–1729. https://doi.org/10.1109/ICCVW.2019.00213
8. H. Ghayvat, M. Awais, S. Pandya, H. Ren, S. Akbarzadeh, S. C. Mukhopadhyay, et al., Smart aging system: uncovering the hidden wellness parameter for well-being monitoring and anomaly detection, *Sensors*, **19** (2019), 766. https://doi.org/10.3390/s19040766
9. C. I. Patel, D. Labana, S. Pandya, K. Modi, H. Ghayvat, M. Awais, Histogram of oriented gradient-based fusion of features for human action recognition in action video sequences, *Sensors*, **20** (2020), 7299. https://doi.org/10.3390/s20247299
10. M. Z. Poh, D. J. McDuff, R. W. Picard, Non-contact, automated cardiac pulse measurements using video imaging and blind source separation, *Opt. Express*, **18** (2010), 10762–10774. https://doi.org/10.1364/OE.18.010762
11. M. Z. Poh, D. J. McDuff, R. W. Picard, Advancements in noncontact, multiparameter physiological measurements using a webcam, *IEEE Trans. Biomed. Eng.*, **58** (2011), 7–11. https://doi.org/10.1109/TBME.2010.2086456
12. D. H. Gerard, V. Jeanne, Robust pulse rate from chrominance-based rPPG, *IEEE Trans. Biomed. Eng.*, **60** (2013), 2878–2886. https://doi.org/10.1109/TBME.2013.2266196
13. S. K. A. Prakash, C. Tucker, Bounded Kalman filter method for motion-robust, non-contact heart rate estimation, *Biomed. Opt. Express*, **9** (2018), 873–897. https://doi.org/10.1364/BOE.9.000873
14. Z. Yang, X. Yang, J. Jin, X. Wu, Motion-resistant heart rate measurement from face videos using patch-based fusion, *Signal Image Video Process.*, **3** (2019), 423–430. https://doi.org/10.1007/s11760-018-01409-w

15. Y. Qiu, Y. Liu, J. Arteaga-Falconi, H. Dong, A. E. Saddik, EVM-CNN: Real-time contactless heart rate estimation from facial video, *IEEE Trans. Multimedia*, **21** (2018), 1778–1787. https://doi.org/10.1109/TMM.2018.2883866
16. X. Niu, S. Shan, H. Han, X. Chen, Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation, *IEEE Trans. Image Process.*, **29** (2020), 2409–2423. https://doi.org/10.1109/TIP.2019.2947204
17. K. Zheng, K. Ci, J. Cui, J. Kong, J. Zhou, Non-contact heart rate detection when face information is missing during online learning, *Sensors*, **20** (2020), 7021. https://doi.org/10.3390/s20247021
18. D. Garg, P. Goel, S. Pandya, A. Ganatra, K. Kotecha, A Deep learning approach for face detection using YOLO, in *Proceedings of the IEEE Punecon*, (2018), 1–4, https://doi.org/10.1109/PUNECON.2018.8745376
19. V. Kazemi, J. Sullivan, One millisecond face alignment with an ensemble of regression trees, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2014), 1867–1874. http://doi.org/10.1109/CVPR.2014.241
20. G. Heusch, A. Anjos, S. Marcel, A reproducible study on remote heart rate measurement, preprint, arXiv:1709.00962.
21. Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.*, **13** (2004), 600–612. https://doi.org/10.1109/TIP.2003.819861
22. A. Mittal, R. Soundararajan, A. C. Bovik, Making a completely blind image quality analyzer, *IEEE Signal Process. Lett.*, **20** (2012), 209–212. https://doi.org/ 10.1109/LSP.2012.2227726
23. G. S. Hsu, A. Ambikapathi, M. S. Chen, Deep learning with time-frequency representation for pulse estimation from facial videos, in *Proceedings of IEEE International Joint Conference on Biometrics*, (2017), 383–389. https://doi.org/10.1109/BTAS.2017.8272721
24. R. Stricker, S. Müller, H. M. Gross, Non-contact video-based pulse rate measurement on a mobile service robot, in *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication*, (2014), 1056–1062. https://doi.org/10.1109/ROMAN.2014.6926392

AIMS Press