

# An Improved Approach to the Performance of Remote Photoplethysmography

Yi Sheng<sup>1</sup>, Wu Zeng<sup>1,\*</sup>, Qiuyu Hu<sup>1</sup>, Weihua Ou<sup>2</sup>, Yuxuan Xie<sup>3</sup> and Jie Li<sup>1</sup>

<sup>1</sup>School of Electrical and Electronic Engineering, Wuhan Polytechnic University, Wuhan, 430000, China

<sup>2</sup>School of Electronic and Information Engineering, Guizhou Normal University, Guizhou, 550000, China

<sup>3</sup>Gina Cody School of Engineering and Computer Science, Concordia University, W. Montreal, Quebec, H3G1M8, Canada

\*Corresponding Author: Wu Zeng. Email: zengwu@whpu.edu.cn

Received: 30 January 2022; Accepted: 27 April 2022

**Abstract:** Heart rate is an important metric for determining physical and mental health. In recent years, remote photoplethysmography (rPPG) has been widely used in characterizing physiological signals in human subjects. Currently, research on non-contact detection of heart rate mainly focuses on the capture and separation of spectral signals from video imagery. However, this method is very sensitive to the movement of the test subject and light intensity variation, and this results in motion artifacts which presents challenges in extracting accurate physiological signals such as heart rate. In this paper, an improved method for rPPG signal preprocessing is proposed. Based on the well known red green blue (RGB) color space, we segmented skin tone in different color spaces and extracted rPPG signals, after which we use a skin segmentation training model based on the luminance component, the blue-difference chroma components, and red-difference chroma components (YCbCr), as well as hue saturation intensity (HSI) color models. In the experimental verification section, we compare the robustness of the signal on different color spaces. In summary, we are experimentally verifying a better image pre-processing method based on real-time rPPG, which results in more precise measurements through the comparative analysis of skin segmentation and signal quality.

**Keywords:** Remote photoplethysmography; skin segmentation; heart rate

## 1 Introduction

As modern communication technologies continue to improve, advanced medical technologies are rapidly developing. In the era of artificial intelligence (AI), computer technology is enabling many disciplines such as pathology and diagnostic medicine, and is significantly improving the efficiency of the diagnostic and treatment phases of medical care. Telemedicine. Which is an important medical method and healthcare service, has also received increasing attention from researchers. Zhang et al. proposed a new audio watermarking algorithm to ensure the safe transmission and storage of medical data, and have made many contributions towards technologies which protect medical audio data in telemedicine [1,2].



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Of course, improving the standard of medical care and disease prevention and control is ultimately for health. Hypertension, hyperlipidemia, and hyperglycemia are now common in for middle-aged and elderly people. These cardiovascular-related geriatric diseases can be controlled if the patient is continuous monitored and actively given treatment. Heart rate is an important indicator of health, and if heart rate could be monitored in real-time during daily activities, it would be of great help to prevent heart disease. There are two methods for heart rate detection. The first method, called Electrocardiogram (ECG), obtains the bioelectric measurement on the body surface, and changes of biological potential are captured and analyzed which are then used to form the ECG heart rate signal. The second method called photoplethysmography (PPG), is a non-contact optical imaging method which detects the variation in the skin's hemoglobin concentration that occurs as a result of the rhythmic heart activity. As light passes through the skin, the change of blood component concentration corresponding to the change of light absorption is collected to form PPG signals.

There has always been a need for contactless detection of human physiological signals in medical, transport, and military fields. Since the novel Coronavirus outbreak in 2019, a significant increase in the demand for contactless video pulse wave extraction was realized. Unlike previous PPG technologies, which require dedicated sensor devices, remote photoplethysmography (rPPG) can obtain human blood volume pulse (BVP) signals through the cameras of smart phones and various monitoring devices. rPPG estimates heart rate parameters by extracting the color changes of human skin caused by the beating of the heart from video information captured by the device. The speed of heart rate is directly related to human health in these two aspects.

The rPPG tracing important progress as follows, from single channel extract heartbeat signal light green [3], for the first time, verify the technical feasibility. Heart rates that are too fast or too slow may indicate that a person has heart diseases, indicating that further medical assessment is advisable. For example, Kusuma et al. used multi-Kernel-principal components analysis (PCA) and mixed deep learning to study the classification of heart disease, which helped improve the efficiency of the treatment of heart disease [4], rPPG technology is also widely used, making many contributions not only to heart rate measurement but also to blood pressure measurement [5,6]. RPPG provides advantaged for individuals interested in self-health monitoring during daily activities. Considering that a fixed camera is used to obtain the facial video in a stable environment, rPPG technology has obvious disadvantages due to its non-contact measurement characteristics. When the camera collects video signals, it is prone to various environmental interference, such as motion interference and light interference. In recent years, many scholars have made efforts to improve signal stability using blind source separation technology [7], With this approach three separate signal channels are extracted, and compare to select the appropriate signal component as the final extract heartbeat signal source of information. Subsequently a series of different test subjects are used to overcome motion artifact, the non-rigid movement, the influence of mirror reflection, and the change of light [8–11]. The blind source separation algorithm helps to improve the quality of rPPG, but it often fails to recover rPPG signals from serious motion artifacts. There are also machine learning methods that are applied to this, for example using the extracted heart rate features into support vector machines (SVM) for heart rate prediction [12]. In recent years, deep learning has also been applied to extracting heart rate signals [13,14], Chenglin Yao et al. applied this technique to face-spoofing detection [15]. However, most studies measuring vital signs did not take the comprehensive analysis of the color channel combination space but rather choose a color space for the analysis based on the combination of two or more-color spaces. Thus, choosing a suitable color space to extract the rPPG signal is particularly important.

The signal quality collected by rPPG is much worse than that obtained from PPG. To compensate for this deficiency, improving the accuracy of the image pre-processing algorithm has become the focus

of research in recent years. The most critical step of image pre-processing is the selection of the region of interest (ROI), which involves face recognition technology. This element of the analysis was proposed in the early 1990s, however at that time the performance of early face detection systems was limited. In 1996 [16] Sobottka et al. proposed a skin segmentation model based on hue saturation intensity (HSI) color space for face detection technology. In 2003 [17] Kovac et al. proposed a skin segmentation model based on red green blue (RGB) color space for face detection technology. The RGB color space used by this technique is the default color space for most available image formats. However, the high correlation between channels, significant perceptual heterogeneity, and the mixing of chromaticity and luminance data make RGB unsuitable for color analysis and color recognition needed for face recognition algorithms. Dahmani, D et al. proposed a skin segmentation model based on luminance component blue-difference chroma components and red-difference chroma components (YCbCr) and HSI color space for face detection technology in 2020 [18]. Therefore, the skin segmentation model based on YCbCr and HSI color space, and the skin segmentation model based on YCbCr color space, is used for comparative analysis. We chose HSI and YCbCr color spaces because skin color has better clustering distribution in these color spaces.

In 2006, the deep learning framework proposed by Hinton et al. promoted the rapid development of this field [19]. Systematically summarized, the current mainstream three neural network frameworks [20], and image pre-processing algorithm also benefited from continuous improvement on this basis. Skin segmentation model based on neural networks was proposed by Vasanthi et al. in 2021.

To improve the performance of rPPG, with the more widely used RGB color space, this paper first introduces two important color spaces, HSI color space and YCbCr color space. In the pre-processing part of extracting the rPPG signal, a skin segmentation method is used to eliminate the interference caused by non-face regions, and the heart rate detection results in different color spaces are studied. By analyzing the detection results, it is observed that the detection algorithms using these two-color spaces is better than that of the traditional RGB color space. At the same time, the clustering effect of skin color in the YCbCr color space is better, and the two-color difference components representing the brightness and color information are also separated, so the noise effect caused by motion can be better eliminated in the YCbCr color space.

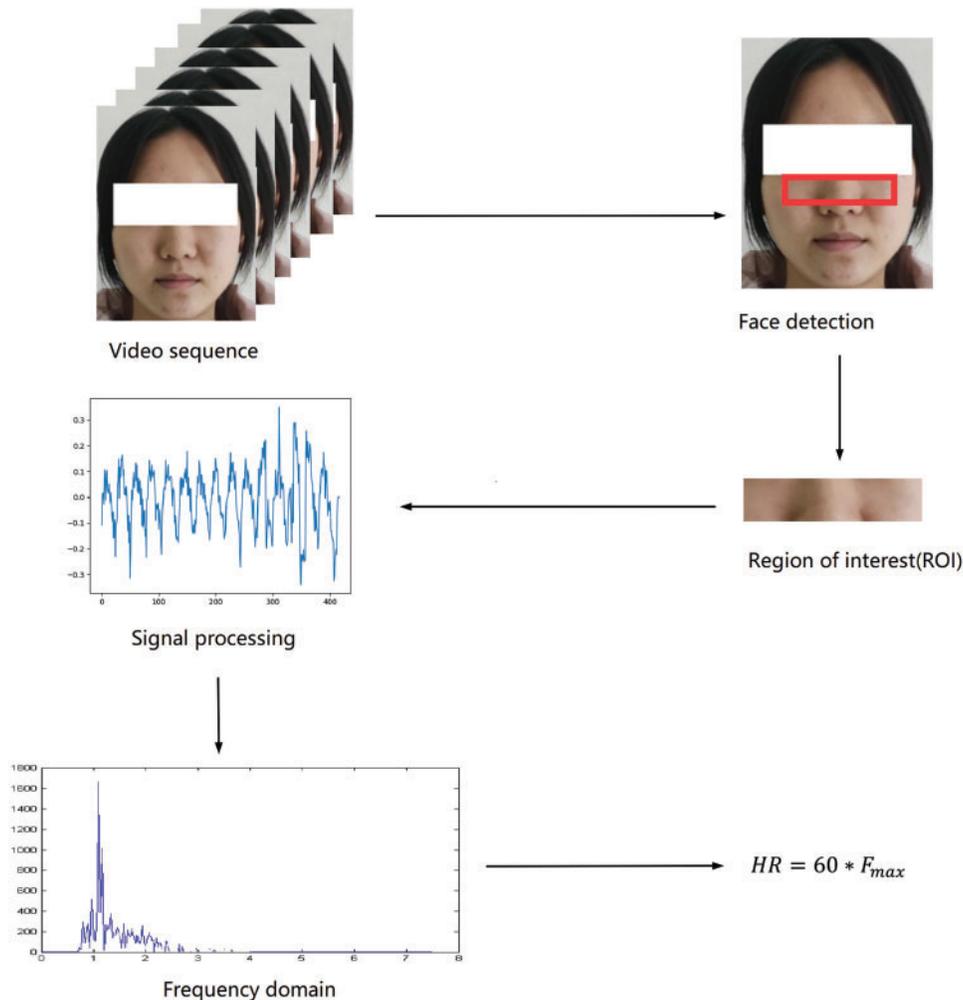
## **2 Principles and Methods**

### ***2.1 Data Collection and Processing***

The human face contains abundant capillaries. Compared with other body parts, the blood volume changes more obviously, which is more conducive to obtaining natural light reflected and scattered by skin, blood vessels, and other tissues. Through a series of techniques, the light signal was collected and processed. The processing generated the digital signal required to obtain the vital signal that is the most relevant data for the human body. For the experiment, a day with sunny weather was chosen and an ordinary mobile phone camera was used to collect facial videos of multiple test subjects under natural indoor lighting conditions. The duration of the video is 1 min and it is saved as a local file. In the first step, the collected videos were initially screened and the videos that are not clear or meet the 1-minute duration requirement are eliminated, and additional video data is collected as a replacement. The video was obtained using the rear camera of the Redmi K30 Pro mobile phone at 30 frames per second (fps), and a resolution of  $1920 \times 1080$ .

## 2.2 The Structure Arrangement

The structure arrangement is shown in Fig. 1. First we used video data obtained from the facial region, then the tracking algorithm was used to locate the selected ROI. In the time domain, the mean value of single or multiple color channels is calculated from the ROI to obtain original raw signal. The correlation algorithm of signal processing is applied to the original signal to obtain the heart rate value. Then the peak detection algorithm, which uses fast Fourier transform (FFT), is used to determine the corresponding peak value of the amplitude spectrum, which is the estimated heart rate value.



**Figure 1:** Basic structure arrangement

In this paper, the pretreatment part is improved, we use the method of skin segmentation. In terms of signal processing, we adopt a chromaticity-based approach to extract signals in HSI and YCbCr color space according to skin reflection's relevant optical and physiological characteristics.

## 2.3 Face Recognition Algorithm and Skin Segmentation Algorithm

The first step of the approach is face detection. In this step we detect the bounding box of the face in the first frame and using the dlib's face detector.

The color space used to extract rPPG signals is mainly based on RGB channel values, but the measurement of physiological signs based on rPPG will be seriously affected due to the interference caused by test subject movement or lighting conditions. Therefore, many scholars began to study in other color spaces, such as HSI [21] and YCbCr [22], and also analyzed rPPG in CIE Lab color space, such as Yuting Yang [23]. These have achieved good results.

To select a color space most suitable for extracting rPPG signals, we compare the signal-to-noise ratio (SNR) of extracting rPPG signals on three channels (RGB, HSI, and YCbCr). The HSI color space uses three parameters, H, S, and I to describe the color characteristics, where Hue (H) represents hue, which is used to define the wavelength of the color; Saturation (S) represents saturation, which is used to indicate the depth of color; Intensity (I) indicates brightness, indicating the brightness of the color. Because hue and saturation (H and S) are independent of brightness in the HSI color space, that is, brightness has little influence on hue and saturation. Thus the heart rate detection results are less sensitive to the influence of light intensity changes with HSI. The formula for converting RGB color space to HSI is as follows (1).

$$H = \cos^{-1} \left\{ \frac{\frac{1}{2} [(R - G) + (R - B)]}{\left[ \frac{(R - G)^2 + (R - B)(G - B)}{3} \right]^{\frac{1}{2}}} \right\} \quad (1)$$

$$S = 1 - \frac{3}{(R + G + B)} [\min(R, G, B)]$$

$$I = \frac{1}{3} (R + G + B)$$

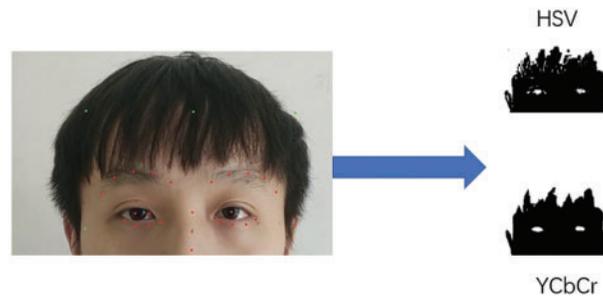
The YCbCr color space uses the three parameters Y, Cb, and Cr to describe the color characteristics, where Y represents the brightness, and Cb and Cr represent the blue component and the red component, respectively. The formula for converting RGB color space to YCbCr is as follows (2). In YCbCr color space, the brightness information and the two chromatic components representing the color information are also phase separated, which is usually used for continuous image processing in film or digital photography systems. In YCbCr color space, image brightness information and chromaticity information are separated from each other. Chromaticity has clustering property and obeys certain distribution in this space, skin color has good clustering property in this space.

$$Y = 0.299 \times R + 0.587 \times G + 0.114 \times B$$

$$Cb = (B - Y) \times 0.564 + 128$$

$$Cr = (R - Y) \times 0.713 + 128 \quad (2)$$

Considering that the interference of background noise has an impact on signal extraction, and the rPPG signal command is reflected in the skin area, the non-skin area should be removed to obtain a clean and accurate pulse signal. Therefore, the skin segmentation method is adopted as a means of pre-treatment as shown in Fig. 2.



**Figure 2:** Skin segmentation

#### **2.4 rPPG Signal Extraction**

After image pre-processing of the collected video frames, it is necessary to separate their signals. To compare and analyze the influence of face recognition and target tracking algorithms and skin segmentation algorithms on rPPG performance, the rPPG method based on chroma is uniformly adopted in the experiment. It should be noted that the research focus of this paper is on image pre-processing rather than signal separation. In the experiment, signal separation adopts the chrominance-based (CHROM) pulse extraction method [8].

It is assumed that when the intensity of diffuse light is stronger than that of specular light, the superposition of strong diffuse light and relatively weak specular light will cause small changes in skin chroma and saturation. Therefore, the illumination change caused by motion (mainly the change of the specular reflection light) has little effect on the chroma and saturation in the HSI color space, and the pulse wave signal can be extracted from the two components of the chroma and the saturation.

In YCbCr color space, since chromaticity information is contained in Cb and Cr components, pulse wave signals can be extracted from these two components to calculate the signal sequences of blue component Cb and red component Cr, and then the FFT of the signal sequences of the two components can be calculated.

To compare and analyze the face tracking and segmentation methods in rPPG, the rPPG method based on CHROM was used to extract rPPG from the pre-processed frames. Then, in the second step, the tracker tracks the regions of interest of the faces in the subsequent frames for comparison. The region of interest was divided into skin mask regions by the thresholding-based skin segmentation method [20,21]. It should be noted that this paper makes a comparative analysis of the basic image processing algorithm of rPPG, rather than signal processing of the extracted rPPG signal. The third step is signal processing, and the rPPG signal is extracted from the average color signal of the skin region obtained from image processing. The chromaticity based method is suitable for real human-computer interaction scenes with frequent angle changes or frequent illumination changes between the camera, skin, and light sources. By separating brightness and chromaticity, chromaticity-based rPPG measurements help to separate rPPG signals from diffuse reflections with noise. In signal processing, the chromaticity signal is used for rPPG extraction, followed by detrending to remove high-frequency components, and then noise reduction and finite impulse response (FIR) bandpass filtering is carried out to obtain a 42–240 beat per minute (bpm) frequency band. Finally, in the fourth step, the signal is converted into a frequency-domain signal by the discrete FFT to estimate the heart rate.

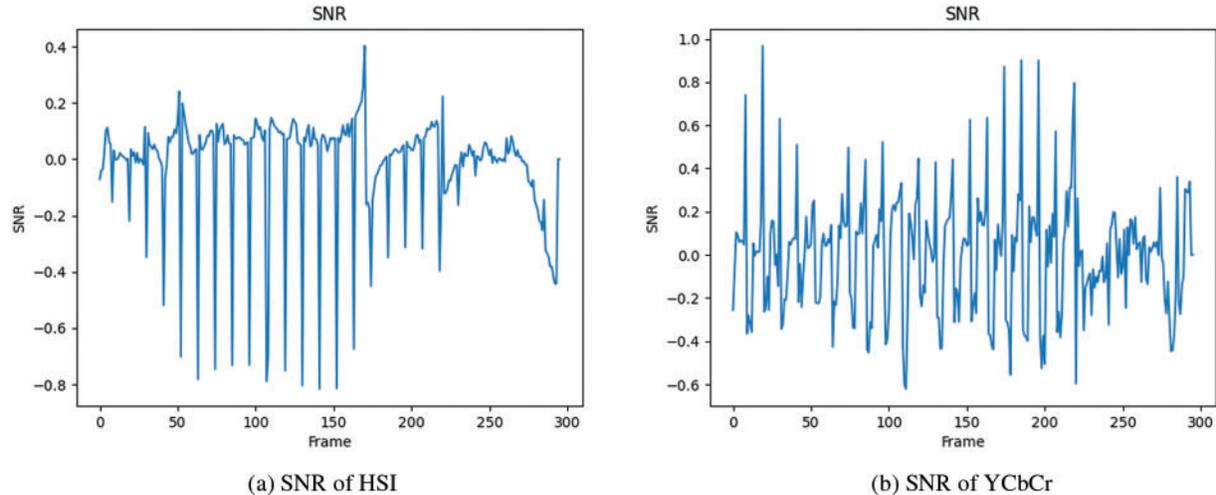
### 2.5 Evaluation Indicators

Finally, the average SNR and SNR box graph are used to compare the performance of the combined methods. The SNR of the signal described in [8] is calculated as follows (3). In the formula,  $U_i(f)$  represents the template window function with signal components, and  $\hat{S}(f)$  is the power spectral density function of rPPG signal. The time filtering bandwidth for the SNR calculation is 42–240 bpm. As the final index of processing speed measurement, the frame number is calculated according to rPPG measurement methods of different color space methods.

$$SNR = 10 \log_{10} \left( \frac{\sum_{42}^{240} (U_i(f) \hat{S}(f))^2}{\sum_{42}^{240} ((1 - U_i(f)) \hat{S}(f))^2} \right) \quad (3)$$

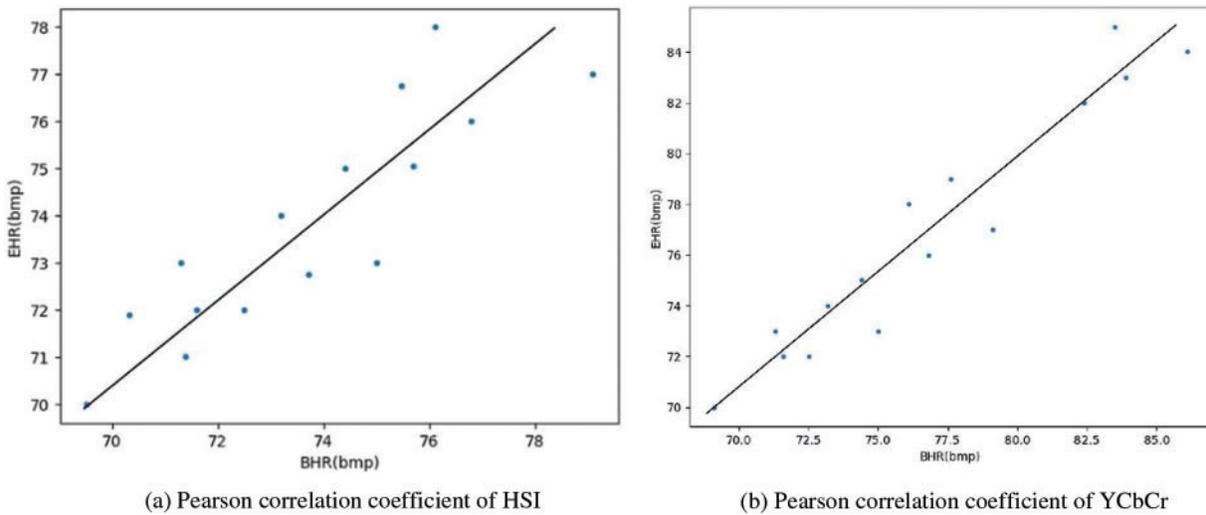
### 3 Result Analysis

Firstly, we analyze the skin segmentation method that defines the skin color space in the YCbCr color model and the skin segmentation method that defines the skin color space in YCbCr and HSI color model. In the case of HSI and YCbCr, noise is generated by the skin mask, and may be sensitive to changes in motion and light. On the other hand, because skin masks are more stable than HSI and YCbCr models, skin segmentation using the YCbCr model showed higher pearson correlation coefficient (PCC) and SNR. Therefore, the skin segmentation method using the YCbCr model has advantages in the correlation between reference data and rPPG signal and the quality of rPPG signal. As shown in Fig. 3.



**Figure 3:** SNR of HSI (a) and YCbCr (b)

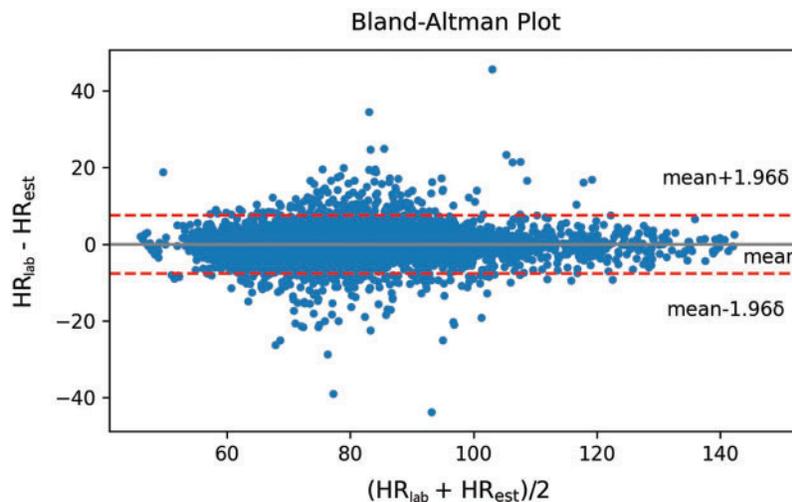
Pearson correlation coefficient is the linear correlation between real heart rate and estimated heart rate through a scatter plot. The greater the absolute value of the correlation coefficient, the stronger the correlation. The result is shown in Fig. 4.



**Figure 4:** Pearson correlation of HSI (a) and YCbCr (b)

The VIPL-HR [24,25] dataset is a data set designed for remote heart rate measurement in recent years. It covers several typical rPPG application scenarios and contains a large amount of data, so it is an ideal choice for training network models. The dataset contains 2378 visible and 752 near-infrared videos in different scenarios. The dataset was derived from 107 test subjects, comprised of 79 men and 28 women, ranging in ages from 22 to 41.

Bland-Altman plots were used to visualize the results, showing a good match between the estimated and labeled heart rate values Fig. 5.



**Figure 5:** Bland-Altman analysis

Tab. 1 shows the comparison of our measurement results in different color spaces. It can be found that the measurement results of HSI and YCbCr are better than those of the traditional RGB color space, which also verifies the knowledge we expounded on before. Tab. 2 shows that 8 groups of data are selected from the 15 groups tested. The 15 sets of video data cover different scenarios. Among

them, object 1 and object 2 are the video data taken when the head is still, object 3 and object 4 are the video data taken when the head moves slightly, and object 5 and object 6 are the video data taken in the environment with light changes. Object 7 and object 8 are video data taken with head movement and different lighting at the same time. On the whole, our scheme is helpful to the accuracy of measurement results.

**Table 1:** Contrast of different color spaces

Color space	Reference	Estimate	Error percentage
RGB	75	81.35	8.46
HSI	75	79.47	5.96
YCbCr	75	78.91	5.21

**Table 2:** Error percentage of different methods

Object no.	CHROM	Green	Our method
1	5.47	7.12	4.39
2	4.78	6.38	4.51
3	6.14	6.97	5.21
4	5.23	5.68	4.89
5	7.48	8.73	7.53
6	3.92	5.29	4.37
7	4.67	6.77	4.18
8	5.79	6.84	5.34

#### 4 Conclusion

In this paper, the rPPG method based on chromaticity is compared and analyzed with different color space models. The experiment shows that in the pre-processing part, the skin segmentation method can effectively filter the skin area, reduce the background noise, and reduce the interference caused by the head movement. By comparing the heart rate detection results of the YCbCr and HSI color spaces, the analysis shows that the detection results of these two-color spaces are better than those based on traditional RGB color space, and the detection results in the YCbCr color space were better than those in the HSI color space. So, this is very helpful for us to improve the accuracy of rPPG in measuring heart rate.

Although the improved method proposed in this paper has achieved good results, low precision data can be obtained when the results are measured in the experimental process. Meanwhile, the measurement environment will also affect the robustness of data results, so there is still a lot of room for improvement in this scheme. In the part of skin segmentation, it is noted that wearing masks may lead to misidentification of facial skin areas and affect the accuracy of extraction signals, which is also part of the follow-up study.

**Acknowledgement:** The authors gratefully acknowledge the laboratory of electronic and information engineering, Multimedia Computing Laboratory, Guizhou Normal University, Guizhou Province, China.

**Funding Statement:** This work was financially supported by the National Nature Science Foundation of China (Grant Number: 61962010).

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] X. Zhang, X. Sun, X. Sun, W. Sun and S. K. Jha, "Robust reversible audio watermarking scheme for telemedicine and privacy protection," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 3035–3050, 2022.
- [2] X. Zhang, W. Zhang, W. Sun, X. Sun and S. K. Jha, "A robust 3-d medical watermarking based on wavelet transform for data protection," *Computer Systems Science and Engineering*, vol. 41, no. 3, pp. 1043–1056, 2022.
- [3] W. Verkruyse, L. O. Svaasand and J. S. Nelson, "Remote plethysmo-graphic imaging using ambient light," *Optics Express*, vol. 16, no. 26, pp. 21434–21445, 2008.
- [4] S. Kusuma and D. Jothi, "Heart disease classification using multiple K-PCA and hybrid deep learning approach," *Computer Systems Science and Engineering*, vol. 41, no. 3, pp. 1273–1289, 2022.
- [5] C. Yen, S. Chang, L. Jia-Xian and Y. Huang, "A deep learning-based continuous blood pressure measurement by dual photoplethysmography signals," *Computers, Materials & Continua*, vol. 70, no. 2, pp. 2937–2952, 2022.
- [6] F. Schruppf, P. Frenzel, C. Aust, G. Osterhoff and M. Fuchs, "Assessment of non-invasive blood pressure prediction from PPG and rPPG signals using deep learning," *Sensors*, vol. 21, no. 18, pp. 6022, 2021.
- [7] M. Z. Poh, D. J. McDuff and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Optics Express*, vol. 18, no. 10, pp. 10762–10774, 2010.
- [8] G. de Haan and A. van Leest, "Improved motion robustness of remote-PPG by using the blood volume pulse signature," *Physiological Measurement*, vol. 35, no. 9, pp. 1913–1926, 2014.
- [9] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878–2886, 2013.
- [10] W. Wang, A. C. den Brinker, S. Stuijk and G. de Haan, "Algorithmic principles of remote PPG," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1479–1491, 2017.
- [11] W. Wang, S. Stuijk and G. D. Haan, "A novel algorithm for remote photoplethysmography: Spatial subspace rotation," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 9, pp. 1974–1984, 2015.
- [12] W. Zeng, Y. Sheng, Q. Hu, Z. Huo, Y. Zhang *et al.*, "Heart rate detection using SVM based on video imagery," *Intelligent Automation & Soft Computing*, vol. 32, no. 1, pp. 377–387, 2022.
- [13] X. Niu, Z. Yu, H. Han, X. Li and G. Zhao, "Video-based remote physiological measurement via cross-verified feature disentangling," in *European Conf. on Computer Vision*, Cham, Switzerland: Springer, 2020.
- [14] Z. Yu, W. Peng, X. Li and X. Hong, "Remote heart rate measurement from highly compressed facial videos: An end-to-end deep learning solution with video enhancement," in *Proc. ICCV*, Seoul, Korea, pp. 621–632, 2019.
- [15] C. Yao, S. Wang, J. Zhang, W. He, H. Du *et al.*, "rPPG-Based spoofing detection for face mask attack using efficientnet on weighted spatial-temporal representation," in *Proc. 2021 IEEE Int. Conf. on Image Processing (ICIP)*, Anchorage, AK, USA, pp. 3872–3876, 2021.
- [16] K. Sobottka, and P. Ioannis, "Segmentation and tracking of faces in color images," in *Proc. of the Second Int. Conf. on Automatic Face and Gesture Recognition.*, Killington, VT, USA, pp. 236–241, 1996.
- [17] J. Kovac, P. Peer and F. Solina, "Human skin color clustering for face detection," *IEEE*, vol. 2, pp. 144–148, 2003.

- [18] D. Dahmani, M. Cheref and S. Larabi, "Zero-sum game theory model for segmenting skin regions," *Image and Vision Computing*, vol. 99, pp. 103925, 2020.
- [19] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [20] A. Krizhevsky, I. Sutskever and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
- [21] K. Lee, K. Jin, Y. Kim, J. H. Lee and E. C. Lee, "A comparative analysis on the impact of face tracker and skin segmentation onto improving the performance of real-time remote photoplethysmography," in *Int. Conf. on Intelligent Human Computer Interaction*, Cham, Switzerland: Springer, pp. 27–37, 2020.
- [22] D. Cho, J. Kim, K. J. Lee and S. Kim, "Reduction of motion artifacts from remote photoplethysmography using adaptive noise cancellation and modified HSI model," *IEEE Access*, vol. 9, pp. 122655–122667, 2021.
- [23] Y. Yang, C. Liu, H. Yu, D. Shao, F. Tsow *et al.*, "Motion robust remote photoplethysmography in CIELab color space," *Journal of Biomedical Optics*, vol. 21, no. 11, pp. 117001, 2016.
- [24] X. Niu, H. Han, S. Shan and X. Chen, "VIPL-HR: A multi-modal database for pulse estimation from less-constrained face video," in *Asian Conf. on Computer Vision*, Cham, Switzerland: Springer, pp. 562–576, 2018.
- [25] X. Niu, S. Shan, H. Han and X. Chen, "Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation," *IEEE Transactions on Image Processing*, vol. 29, pp. 2409–2423, 2019.