On the benefits of alternative color spaces for noncontact heart rate measurements using standard red-green-blue cameras

Gill R. Tsouri
Zheng Li

On the benefits of alternative color spaces for noncontact heart rate measurements using standard red-green-blue cameras

Gill R. Tsouri* and Zheng Li
Rochester Institute of Technology, Department of Electrical and Microelectronic Engineering, Communications Laboratory, 79 Lomb Memorial Drive, Building 9, Room 3050, Rochester, New York 14623-5603, United States

Abstract. Existing video plethysmography methods use standard red-green-blue (sRGB) video recordings of the facial region to estimate heart pulse rate without making contact with the person being monitored. Methods achieving high estimation accuracy require considerable signal-processing power and result in significant processing latency. High processing power and latency are limiting factors when real-time pulse rate estimation is required or when the sensing platform has no access to high processing power. We investigate the use of alternative color spaces derived from sRGB video recordings as a fast light-weight alternative to pulse rate estimation. We consider seven color spaces and compare their performance with state-of-the-art algorithms that use independent component analysis. The comparison is performed over a dataset of 41 video recordings from subjects of varying skin tone and age. Results indicate that the hue channel provides better estimation accuracy using extremely low computation power and with practically no latency. © *2015 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.JBO.20.4.048002]

1 Introduction

Video photoplethysmography (VPG) is a relatively new technique of monitoring human vital signs requiring no physical contact with the subject being monitored. Instead of a sensor placed on the skin, VPG utilizes a digital camera, signal-processing power, and ambient light. Reflected light off the face changes with the pulsating heart as it delivers blood to and from the face. In a typical implementation, the face would be detected in each frame, followed by averaging of the color pixels per frame within the detected area in each channel. The result is a signal per channel across the video frames. The signals are then processed to extract periodic patterns to assess cardiac activity. The potential applications include automated, continuous and ubiquitous health monitoring in hospitals and residential environments. Since the equipment needed is not specialized, VPG can be implemented using software add-ons to existing electronic devices with a digital camera such as in personal computers, laptops, and smartphones. The most common use of VPG is in providing a noncontact alternative to heart pulse rate estimation using a finger probe oximeter or an earlobe photoplethysmography (PPG) sensor. A recent study showed VPG applications on a mobile phone camera have the potential to not only estimate heart rate but also breathing rates and oxygen saturation (SpO_2).[1] This provides the possibility to perform physiological studies related to emotional states outside of a laboratory environment.[2]

Past work reported in the literature focused primarily on the standard red-green-blue (sRGB) color space, see Refs. 1–13 for examples. This color space is readily available since most electronic devices represent images with sRGB channels. It is widely recognized that the green channel provides the most prominent response to the pulsating heart due to its high absorption in blood.[3,4] Directly processing the green color to obtain a PPG-like signal provides real-time estimation. However, it does not provide an accurate estimation. Increased accuracy is achieved using methods that consider all colors combined. The prominent approach relies on independent component analysis (ICA)[5,6,10,13] and an extension of it called constrained ICA (cICA).[7] ICA is a blind source separation technique applicable when the observed signals are a linear mixture of independent sources. When applying ICA to RGB, the underlying assumption is that one of the independent sources is the pulsating heart. ICA was successfully used to improve the accuracy of estimation[5–7,10,13] to a level comparable with commercially available finger-probe oximeters in a laboratory setting.[5–7,10,13]

The complexity of ICA-based methods requires processing time and considerable processing power. Such processing power is unavailable in compact devices with a low form factor. Compact devices can be wirelessly connected to a remote server with a higher computation power, but this would require transmission of raw data for remote processing resulting in estimation latency, high power consumption, and a reliable wireless communication link for sending data. Another caveat associated with ICA methods arises from their block processing of samples typically corresponding to a 30- to 60-s long recording. Block processing results in substantial latency in pulse estimation thereby limiting the utility of the technology in applications where an immediate response is required.

Processing a single color channel is preferable due to its low latency and low complexity implementation provided that an equally accurate estimation is achieved. Since single sRGB channels do not provide accurate estimation, we investigate the use of other color spaces derived from sRGB as an alternative to high latency and computationally intensive algorithms.

Work on VPG using alternative color spaces has been fairly limited to date. In one work,[11] the CIE XYZ color space was used. CIE stands for Commission International de I'Elcairage (International Commission on Illumination), and X, Y, and Z are the three channels in the CIE XYZ color space also known as the CIE 1931 XYZ. Results showed that estimating pulse rate using color spaces other than sRGB is feasible. However, estimation accuracy was not compared to other color spaces and methods. In another study,[12] a few videos were processed to show that ICA using hue saturation lightness (HSL) and hue saturation value (HSV) color spaces provides a more robust estimation in a dynamic environment. In recent work,[13] a camera with a high pixel dynamic range (16 bits/pixel) capable of capturing five bands of color frequencies was used. The work investigated applying ICA to different combinations of red, green, blue, cyan, and orange. Reported results show that using orange or cyan to replace red in the RGB signals fed to ICA improves estimation accuracy. Implementing this approach requires the use of a specialized five-band camera instead of the readily available sRGB cameras.

In this contribution, we investigate heart pulse rate estimation accuracy of single channel processing in alternative color spaces. To preserve all the benefits of using widely available webcams, we investigate color spaces obtainable by direct transformation of detected sRGB colors, i.e., by applying the transformation per frame during real-time image capture. Our premise is that some color transformations might combine the information from the sRGB color space in a way that would amplify the pulsating heart's effect on the resulting channels. An alternative single channel to sRGB green would help in providing accurate real-time pulse estimation using low computation power and virtually no latency. This is because color transformations can be applied real-time without resorting to block processing as is done in ICA-based methods.

Our investigation includes an experimental study encompassing 41 subjects. Channels from seven color spaces are extracted and used to estimate pulse rate. The estimation accuracy per channel is assessed for all channels and compared with ICA and cICA estimation accuracy. Results demonstrate the highest accuracy for the hue channel of HSL, HSV, and hue saturation intensity (HSI). The U channel of CIE YUV and Y channel of CIE XYZ provide good results as well. Hue, U, and Y provide an estimation accuracy comparable to ICA-based methods.

2 Alternative Color Spaces

We investigate seven color spaces: sRGB, HSL, HSV, HIS, XYZ, CIE XYZ, and CIE YUV. In what follows, we briefly describe each color space and its potential for pulse rate estimation.

Red green blue (RGB) is an additive color system. It is commonly used in computer systems, television, and video. Three chromaticities R, G, and B represent channels red, green, and blue, respectively, for every pixel. The combination of the three colors results in visible colors to the human eye. RGB color spaces are easy to implement but are device dependent.

The videos in our study were taken using a digital camera that uses the sRGB color space.

HSL/HSV/HSI color spaces are cylindrical-coordinate color systems. The transformation from the sRGB color space to these color spaces requires manipulation of the three sRGB channels. The hue (H) channel is the same in all three color systems. Saturation (S) is different and is denoted differentially as S_HSL, S_HSV, and S_HSI. Lightness (L), value (V), and intensity (I) are the remaining channels in these color spaces. To obtain these color spaces, we define maximum and minimum components denoted as M and m for each pixel. Then the value chroma (C) is calculated by subtracting M by m as shown below[14]

$$M = \max(R, G, B), \tag{1}$$

$$m = \min(R, G, B), \tag{2}$$

$$C = M - m. \tag{3}$$

H is then calculated using Eqs. (4) and (5) provided below[14]

$$H' = \begin{cases} 0, & \text{if } C = 0 \\ \frac{G-B}{C} \bmod 6, & \text{if } M = R \\ \frac{B-R}{C} + 2, & \text{if } M = G \\ \frac{R-G}{C} + 4, & \text{if } M = B \end{cases}, \tag{4}$$

$$H = 60 \ \deg \times H'. \tag{5}$$

Hue is defined as the pure spectrum colors measured in degrees between 0 and 360 deg and follows a nonlinear transformation of sRGB. Pure colors like red, green, and blue are located at 0, (360 deg), 120, and 240 deg, respectively.[15] All other visible colors are located in between. The color spaces HSL, HSV, and HSL represent different shades of colors more intuitively than the sRGB color space. The variation in hue represents only changes in color.[15] This may be useful to highlight differences between the absorptions of red, green, and blue ambient light in the blood, since time-variation in the relative reflections will be translated to rotation in H. Variations in the sRGB color space depend not only on the color of the object but also on the intensity of the reflected light from the surface. Hue, on the other hand, does not depend on lightness. This means that hue is more tolerant to changes in ambient light. Given these properties, we speculate that hue could provide an accurate estimation of pulse rate.

The sRGB color space can be transformed to the XYZ color space via multiplication with the M matrix.[15,16–19] The transformation is shown in Eq. (6) where r, g, and b are normalized R, G, and B, i.e., the red, green, and blue channels, respectively, divided by their sum

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124564 & 0.3575761 & 0.1804375 \\ 0.2126729 & 0.7151522 & 0.0721750 \\ 0.0193339 & 0.1191920 & 0.9503041 \end{bmatrix} \begin{bmatrix} r \\ g \\ b \end{bmatrix}. \tag{6}$$

The XYZ color space is normalized and becomes the CIE XYZ color space by following:[17]

$$x = \frac{X}{X + Y + Z}, \tag{7}$$

$$y = \frac{Y}{X + Y + Z}, \tag{8}$$

$$z = 1 - x - y. \tag{9}$$

X, Y, and Z in the CIE XYZ color space are the reformulated tristimulus values converted from the RGB color space. Lower case x, y, and z are the chromaticity coordinates. The values for the XYZ tristimulus do not directly correspond to red, green, and blue. Y in color space XYZ is also called the luminance factor because the Y tristimulus value matches the curve that indicates the response of the human eye to the total power of a light source.[19] Y as the luminance measures the luminous intensity of light traveling in a given direction per unit area. This can be useful when the change in signal power is meaningful, e.g., when color change of the skin is of importance.

The CIE XYZ color space can be converted into a CIE YUV[20] color space. Equations (10)–(12) perform the transformation:

$$y = y, \tag{10}$$

$$U = \frac{2x}{6y - x + 1.5}, \tag{11}$$

$$V = \frac{3y}{6y - x + 1.5}. \tag{12}$$

CIE YUV decreases the nonuniformity considerably compared to CIE XYZ. Y for CIE YUV is the same as for CIE XYZ. U and V are linear transformations from X and Y from CIE XYZ.

3 Comparative Study

The study was approved by the Internal Review Committee for Protecting Human Subjects at the Rochester Institute of Technology and included participants between the ages of 18 and 45 years old from various nationalities. The recorded videos were 60 s long taken using a commercially available RGB Logitech camera set to 15 frames/s. A 320 × 240 pixel resolution was set for all videos. The videos were saved in WMV format and then converted to AVI format. A reference pulse rate measurement was taken per video using an food and drug administration-approved and commercially available finger probe oximeter (Onyx II 9550 Military Model Finger Pulse Oximeter).

The laboratory where the videos were taken was illuminated by fluorescent light fixtures covering the entire ceiling. No significant light entered the lab from the outside. The participants were asked to maintain a relaxed sitting position during the recording but were not constrained in any way, i.e., their heads were not supported and they were free to move slightly. Each participant was seated ∼65 cm in front of a screen with the webcam positioned on top of it. The light coming from the screen illuminated the face with insignificant intensity compared to the ceiling light fixtures.

All recorded videos were included in the performance analysis and processed using MATLAB. In each video, a region of interest (RoI) of the subject's face was manually selected and remained unchanged during the recording. The average value of all the pixels in a frame within the RoI was calculated for each color channel resulting in three channels of red, green, and blue. An infinite impulse response Butterworth bandpass filter of order 9 with a band-pass of [0.75, 4] Hz was applied to each channel corresponding to a pulse range of [45, 240] beats per minute (bpm). Computation of all tested channels was followed by applying the aforementioned color transformations to the filtered RGB channels. The process for obtaining each channel in each alternative color space is depicted in Fig. 1.

For each channel in all color spaces, the first 150 frames were removed to eliminate the effects of focusing and adjustments of the camera. The MATLAB Periodogram spectral estimation function was applied with a Hamming window to each channel. Pulse rate estimation was then performed by finding the peak frequency within the range of 0.75 to 4 Hz corresponding to a pulse rate range of from 45 to 240 bpm.

Two previously reported ICA-based methods were applied to the sRGB channels to form a basis for comparison of estimation accuracy. The first method followed the reported algorithm[6] where ICA was applied to the RGB channels followed by spectral estimation for each of the three resulting independent components (ICs) and selection of the peak frequency exhibiting the highest peak within 0.75 to 4 Hz across the three ICs. The JADE algorithm was used for finding the ICs. The second followed the reported cICA algorithm[7] where a reference harmonic signal is used to extract a single IC followed by spectral estimation and peak frequency selection within the same range. Note that the reported ICA[6] and cICA[7] algorithms originally had mismatching preprocessing steps on the sRGB channels prior to applying ICA/cICA. We avoided implementing mismatching preprocessing steps to maintain consistency across all methods being evaluated.
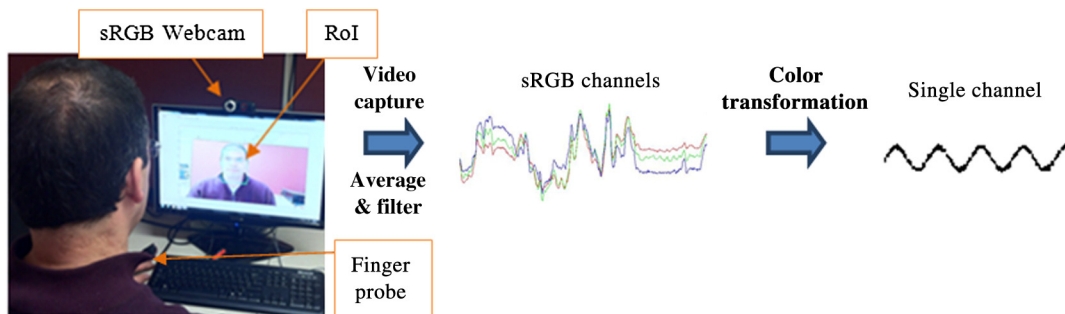

Fig. 1 Experimental setup used in the comparative study.

All pulse rate estimations from alternative color spaces and two ICA-based methods were then compared with the pulse rate measured by the finger probe oximeter. Bland–Altman[21] plots were used to assess the performance over the entire data set. The mean absolute error and its standard deviation were calculated over the entire data set as well.

4 Results and Discussion

We found that some videos resulted in very poor estimation accuracy for all tested methods. Inspection of the videos revealed that in most of them, the subjects were significantly moving their head, laughing, or looking down. In other cases, only part of the ceiling lights were turned on causing the subject's face to be illuminated poorly. The data extracted from these videos were not omitted from the results presented below.

The group of subjects in our study was diverse and included participants with various skin tones. We found that skin tone did not have a notable impact on pulse rate estimation accuracy. Similar observations were made in Ref. 12 where the effect of skin tone was investigated.

Videos obtained from different subjects at different times could result in a different scaling of the signals, e.g., due to different background or lighting conditions. In addition, ICA could result in ICs with varying scale. Recall that pulse rate estimation is obtained per channel separately by evaluating the spectral component with the maximum strength within the frequency range corresponding to the expected range of the pulse. Since pulse rate estimation is based on the frequency of the spectral peak and not on the absolute strength of the spectral peak, scaling of the signals would not affect the estimation of pulse rate and the corresponding error.

We start by analyzing the mean absolute error and its standard deviation across the entire dataset as summarized in Table 1. As expected, the green channel offers the best performance of the three sRGB channels and ICA/cICA exhibit a

better performance compared to the green channel. The following alternative color channels exhibit accuracy comparable to ICA and cICA: hue of HSV/HSL/HSI, Y of CIE XYZ, and U of CIE YUV. Hue performs best with a mean (deviation) of 4.31 (7.04) compared to 5.15 (8.72) for ICA.

Having identified hue as the best single-channel alternative to ICA-based methods, we now use Bland–Altman plots[21] to assess the agreement between our VPG estimations and the finger probe oximeter for hue along with green, ICA, and cICA for comparison. Results are presented in Fig. 2. All channels exhibit a wide spread of data points indicating the study-captured data across the expected range of the pulse.

Observing Fig. 2 while disregarding the outlier estimations corresponding to poor recording quality, it is evident that the hue channel provides consistently higher accuracy than the green channel and a consistently comparable accuracy to ICA and cICA.

As explained earlier, hue represents overall change in color across red, green, and blue. By transforming the sRGB channels to hue, the changes in reflection off the skin due to the pulsating heart are combined and enhanced providing a stronger indication of the pulsating heart. More specifically, note that the definition of hue in Eqs. (4) and (5) is based on the difference between the RGB channels. The RGB channels exhibit varying sensitivity to the pulsating heart where the green channel is most affected. However, all channels are equally sensitive to other variations in reflected light due to slight motion of the subject and changes in ambient light. It follows that taking the difference amplifies the varying reflection off the skin due to the pulsating heart. This amplifying effect could account for the increased estimation accuracy compared to that obtained using the green channel alone.

To gain more insight into the response of the hue channel to the pulsating heart, we recorded a 1-min video of a subject while also recording a PPG signal using an earlobe sensor (Binar Heart-Sensor HRS-07UE). Figure 3 presents the spectrum of the green, hue, and PPG signals. Scaling of the power spectral density was done by normalizing each power spectral density by its maximum value resulting in all peak frequencies having a value of 1. While we see that both green and hue are in agreement with PPG with regard to peak frequency and the corresponding pulse rate, it is also evident that the power spectral density of hue is closer to that of PPG and is more descriptive of the pulse rate.

Obtaining the hue channel following Eqs. (1) to (5) requires very simple mathematical operations and can be performed separately per sRGB frame. It follows that simple sRGB-based sensing platforms with limited computation power are capable of extracting a real-time PPG-like signal. In contradistinction to hue, sensing platforms using ICA and cICA would require significantly more computation power and would suffer from estimation latency.

The second best performing channel is luminance factor Y of the CIE XYZ color space. Transformation into XYZ coordinates may be viewed as measuring the amount of reflected light over several spectral bands. Similarly, traditional PPG data are obtained from the amount of reflected (or transmitted) light over a very narrow spectral band (typically red or infrared). This similarity in terms of the physical data of interest could explain why Y performs well compared to the finger probe PPG oximeter. Deriving the Y channel requires more processing than deriving the hue channel, but it is still much lower than that for ICA methods.

Table 1 Mean and standard deviation of absolute error for all color channels and independent component analysis (ICA) methods.

Tested channel	Mean (standard deviation)	Tested channel	Mean (standard deviation)
Red	15.70 (15.18)	X of XYZ	8.82 (12.27)
Green	**7.41 (12.18)**	Y of XYZ	7.87 (11.90)
Blue	19.80 (19.74)	Z of XYZ	16.93 (17.70)
Chroma	15.36 (14.13)	X of CIE XYZ	13.87 (13.74)
Hue	**4.31 (7.04)**	**Y of CIE XYZ**	**5.33 (7.48)**
L of HSL	17.12 (16.67)	Z of CIE XYZ	15.17 (18.47)
V of HSV	15.70 (15.18)	**U of CIE YUV**	**6.07 (9.85)**
I of HSI	13.02 (15.73)	V of CIE YUV	7.01 (10.08)
S of HSL	15.36 (14.13)	*ICA-based methods*	
S of HSV	19.89 (25.91)	ICA on RGB	5.15 (8.72)
S of HSI	14.47 (16.20)	cICA on RGB	6.85 (13.66)
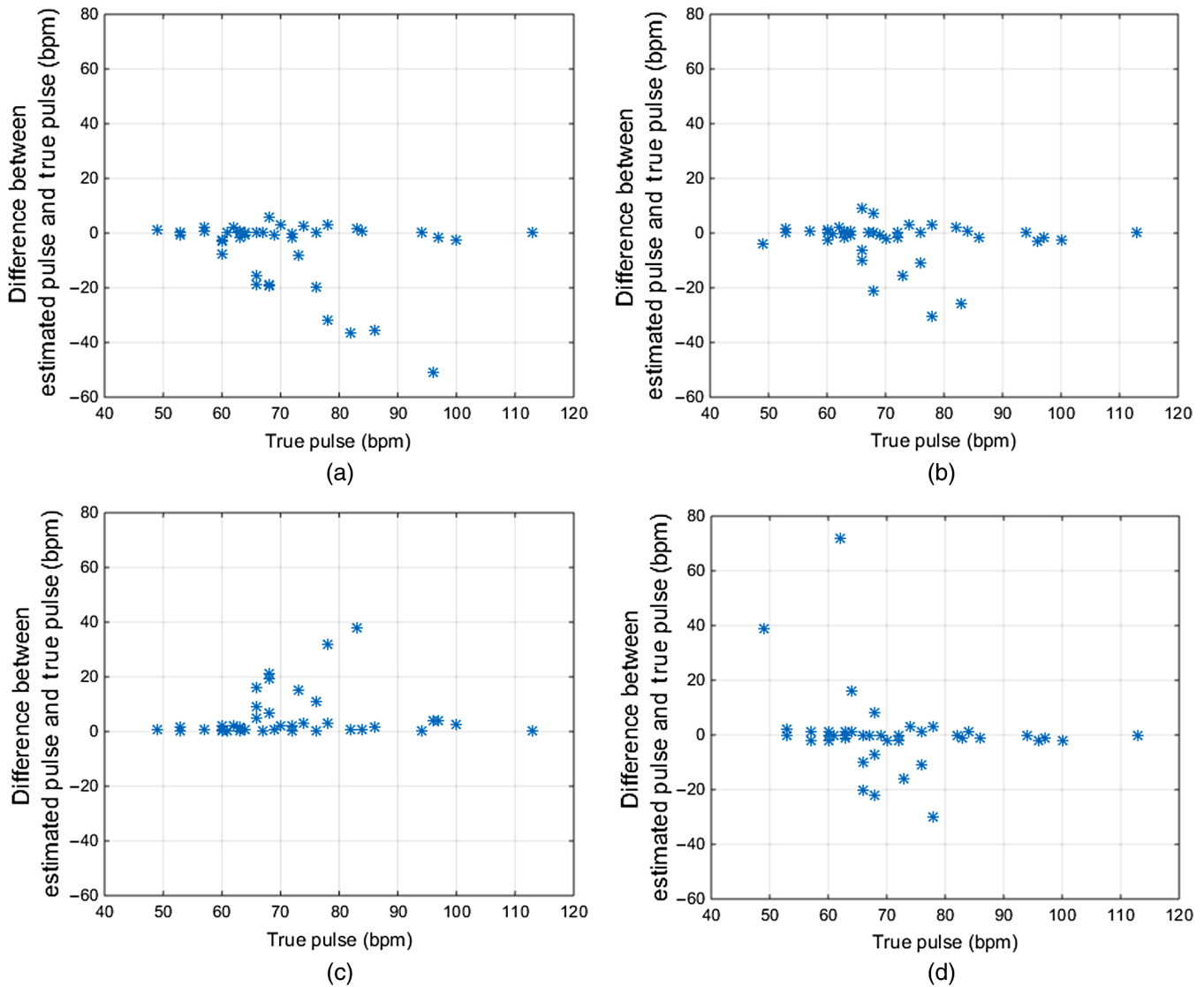
Fig. 2 Bland–Altman plots: (a) green in standard red green blue (sRGB), (b) hue in hue saturation value (HSV)/hue saturation lightness (HSL)/hue saturation intensity (HSI), (c) independent component analysis (ICA) algorithm, and (d) constrained independent component analysis (cICA) algorithm.
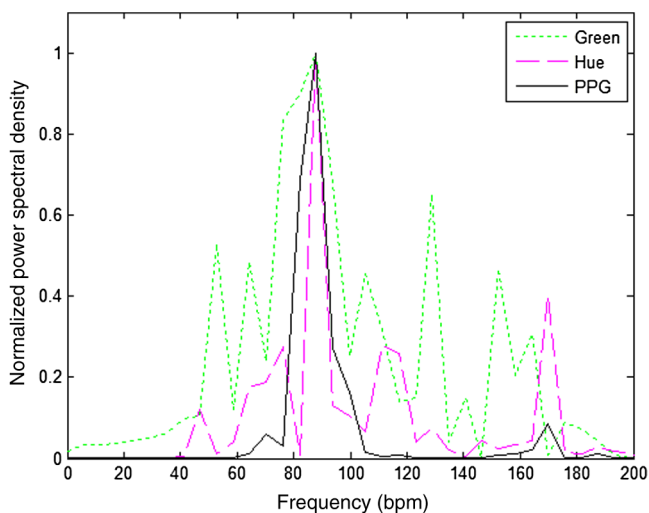
Fig. 3 Demonstration of channels' spectra extracted from the facial region of a single subject.

5 Conclusion

Channels of alternative color spaces derived from the sRGB color space were proposed and investigated as a light-weight alternative to ICA-based methods. A comparative study was performed over a dataset of 41 video recordings, where the channels of seven color spaces were used to estimate the pulse rate compared to a finger probe oximeter. Estimation accuracy was compared to the performance of two ICA-based methods. Results indicate that the hue channel of HSV/HSL/HSI, Y of CIE XYZ, and U of CIE YUV provide an estimation accuracy comparable to ICA methods, where the hue channel offers the best performance.

References

1. C. G. Scully et al., "Physiological parameter monitoring from optical recordings with a mobile phone," *IEEE Trans. Biomed. Eng.* **59**(2), 303–306 (2012).
2. D. Lakens, "Using a smartphone to measure heart rate changes during relived happiness and anger," *IEEE Trans. Affective Comput.* **4**(2), 238–241 (2013).

3. C. Takano and Y. Ohta, "Heart rate measurement based on a time-lapse image," *Med. Eng. Phys.* **29**(8), 853–857 (2007).
4. W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Express* **16**(26), 21434–21445 (2008).
5. M. Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Opt. Express* **18**(10), 10762–10774 (2010).
6. M. Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in non-contact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomed. Eng.* **58**(1), 7–11 (2011).
7. G. R. Tsouri et al., "Constrained-ICA approach to non-obtrusive pulse rate measurements," *J. Biomed. Opt.* **17**(7), 077011 (2012).
8. H. Wu et al., "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graph.* **31**(4), 65 (2012).
9. M. Lewandowska et al., "Measuring pulse rate with a webcam—a non-contact method for evaluating cardiac activity," in *Proc. Federated Conf. Computer Science and Information System*, pp. 405–410 (2011).
10. L. Shan and M. Yu, "Video-based heart rate measurement using head motion tracking and ICA," *IEEE Int. Congr. Image Signal Process.* **1**, 160–164 (2013).
11. I. Nishidate et al., "Noncontact plethysmographic imaging based on diffuse reflectance spectroscopy using a digital RGB camera," *Proc. SPIE* **8798**, 87908D (2013).
12. N. G. Roald, "Estimation of vital signs from ambient-light non-contact photoplethysmography," http://www.diva-portal.org/smash/record.jsf?pid=diva2:622000 (2013).
13. A. Hanbury and J. Serra, "A 3D-polar coordinate colour representation suitable for image analysis," *Comput. Vision Image Underst.* **1**, 804–811 (2002).
14. F. M. Christine, *Advanced Color Image Processing and Analysis*, Springer, New York (2013).
15. "Adobe technical guides," Adobe Systems Incorporated, http://dba.med.sc.edu/price/irf/Adobe_tg/models/main.html (2000).
16. L. W. MacDonald and M. R. Luo, *Colour Imaging Vision and Technology*, John Wiley & Sons Ltd., Hoboken, New Jersey (1999).
17. H. J. Trussell, E. Saber, and M. Vrhel, "Color image processing basics and special issue overview," *IEEE Signal Process. Mag.* **22**(1), 14–22 (2005).
18. A. Ford and A. Roberts, "Colour space conversions," http://www.poynton.com/PDFs/coloureq.pdf (1998).
19. "RGB/XYZ matrices," Bruce Lindbloom.com, http://www.brucelindbloom.com/index.html?Eqn_RGB_XYZ_Matrix.html (4 April 2015).
20. J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet* **327**(8476), 307–310 (1986).
21. D. McDuff, S. Gontarek, and R. W. Picard, "Improvements in remote cardio-pulmonary measurement using a five band digital camera," *IEEE Trans. Biomed. Eng.* **61**(10), 2593–2601 (2014).

Gill R. Tsouri received his BSc, MSc, and PhD degrees in electrical and computer engineering from Ben-Gurion University, Israel, in 2000, 2004, and 2008, respectively. He joined the Rochester Institute of Technology, New York, USA, in 2008. His current research interests include body area networks, biomedical signal processing, and wireless physical layer security.

Zheng Li received his BSc and MSc degrees in electrical engineering from Rochester Institute of Technology, New York, USA, in 2014. She is now with Bendix Commercial Vehicle Systems, where she is working on image-processing applications.