Accurate Heart Rate Measuring from Face Video Using Efficient Pixel Selection and Tracking

Mikiya Koike^{a,*}, Satoru Fujita^a

^aGraduate School of Computer and Information Sciences, Hosei University, Japan

*Corresponding Author: 21t0007@cis.k.hosei.ac.jp

Abstract

As the coronavirus (COVID-19) spreads around the world, we are increasingly cognizant of our health on a daily basis. This paper focuses on heart rate monitoring, utilizing remote monitoring methodology as a vital indicator of health status. Remote photoplethysmography (rPPG), is a well-known technique in human remote monitoring, to calculate the heart rate from face videos. Since rPPG analyzes small changes in: color and motion, physical factors (e.g., breathing and adjusting posture), and environmental factors (e.g., illumination and shade), it is difficult to measure heart rate with precision. To resolve these challenges, this paper proposes a system that effectively combines the following methods: 1) Lucas-Kanade method to dynamically track each skin pixel, 2) selection of proper pixels that are not affected by the environmental fluctuations in light and shade, 3) the delineation of the heart rate signal from noisy to precise data to improve accuracy, and 4) Fast Fourier Transform (FFT) to estimate the main frequency of the signal to determine the heart rate. The results of the experiment showed that the mean absolute error (MAE) of the heart rate was 3.4 bpm for 72 face videos.

Keywords: heart rate measuring, rPPG, feature point tracking, remote monitoring.

1. Introduction

Since a new variant of coronavirus (COVID-19) was found in 2019, public newspapers have reported many cases of contact infection. To mitigate the spread of COVID-19, people are working remotely from home rather than in the marketplace. This situation requires people to be vigilant about their overall health and well-being. Additionally, the use of healthcare devices that require skin contact, further

pose a risk of contact infection.

As a result of the COVID-19 pandemic, non-contact healthcare devices are in demand. We will focus on the remote monitoring of heart rate as one of the vital indicators of health status. Non-contact, remote monitoring systems, such as Remote photoplethysmography (rPPG), are divided into two categories; those that monitor (1) changes in facial color and (2) movement of the face. Regarding facial color, physical factors (e.g., breathing and adjustments to posture) prevent monitoring from correctly tracking identical pixels on the face. In addition, environmental factors (e.g., illumination and shade) make it difficult to detect skin color changes caused by fluctuation in heart rate. Regarding facial movements, the system tracks small nodding movements of the head caused by heart rate; however, other motions of the body prevent the monitoring system from extracting the heartbeat motion. Due to these reasons, the remote monitoring data is less accurate as a result of external noises.

In this paper, we will address the concept of detecting facial color changes to estimate heart rate with a combination of several technologies. The proposed remote monitoring system tracks independent pixels on the face correctly and selects proper feature points to observe the periodical signal.

This paper is organized as follows. In section 2, we review previous research on heart rate measurement. In section 3, we describe our proposed system. In section 4 and 5, the proposed system is evaluated, and we discuss the differences from previous research. In section 6, we address the conclusion of this paper.

2. Heart rate measurement

Electrocardiogram (ECG) developed by Willem

Fig. 1. Overall processes in the proposed system

Einthoven measures the electric potential difference between two points on the human body. It can precisely measure the heart rate by counting periodic spikes of the potential difference. Thus, plethysmography detects the change in volume of the blood vessels. As a derivation of plethysmography, photoplethysmography (PPG) uses photos to observe blood flow with transmitted or reflected light. The transmitted approach uses a light source that shines on a thin skin area and measures the amount of transmitted light. It is necessary to place a part of the human body (e.g., earlobe or finger) between the light source and receiver. The reflected approach also uses a light source, but measures the amount of reflected light. In recent years, the last approach has become widely used because the Apple Watch is equipped with a built-in sensor using the reflected light.

All technologies mentioned above require skin contact for heart rate measurement. Remote photoplethysmography (rPPG), on the other hand, does not require skin contact but remotely observes human skin tone. Juan et al. measured the average skin tone using G-channel of RGB in the ROI (forehead and cheeks) because hemoglobin mainly absorbs green and yellow (1). Yang et al. measured skin tone in CIELab converted from RGB (2). Christian et al. used independent component analysis (ICA) as a blind source separation, which extracts an independent component from three color channels (RGB) (3). Gerard de Haan et al. calculated the chrominance signal for heart rate, called CHROM method, as a linear combination of RGB values under the assumption of standardizing skin colors (4). From another viewpoint, Wu et al. proposed Eulerian video magnification (EVM) that can extract and reveal the head movement by the human heart pulse visually (5). Balakrishnan et al. detected several feature points on the surface of the head, measured the vertical movements of the points, then applied principal component analysis (PCA) to these movements and finally extracted the component with the strongest periodicity in the movements (6). Li et al. measured not only vertical movement but also horizontal

movement and used ICA (7). These methods (6, 7) that extract head movements have an advantage to determine heart rate even when the face is hidden by something, like a mask. The disadvantage of noise tolerance due to body movements, such as breathing and adjusting posture, prevent correct observation of head movement.

3. Noise tolerant heart rate measurement

This paper measures heart rate accurately from facial color by tracking the feature points of the face and selecting noiseless points. Overall processes in the proposed system are shown in Fig. 1. The system consists of tracking feature points, discriminating useful feature points, delineation of heart rate signal via color from noisy to precise data points to improve accuracy and frequency analysis.

3.1 Feature point tracking

Many systems proposed in the previous papers specified an ROI on the face, within which the averages of RGB values are calculated. If the size of ROI is larger, non-skin areas such as hair and beard might be involved and then would make the averaged color values incorrect. On the other hand, if the size of ROI is smaller, the total number of pixels would become smaller, and the estimated time series of the signal would be distorted. For these reasons, previous systems had a difficulty to set the proper size of the ROI. In this paper, we do not set ROI explicitly; we calculate the sequences of RGB values for all pixels, select noiseless points, and average the selected RGB values afterwards.

The following are procedures of the calculation. After the first frame of the video is extracted, the face is detected using haar-like features. The extracted area is shrunk to 90% of the center in height and 50% of the center in width because the extracted face area often includes not only the face but also the background. Since there is a lot of movement around the eyes and mouth, which makes tracking difficult, we separately select two regions on the face with less movement.

The forehead, a region of 20% from the top edge or hairline of the face region, and the mid-facial region with a height of 20% from the top 55% position in Fig. 2. Selection of these two regions make it possible to deal with the case where the forehead region is hidden by hair.

The next step detects skin pixels within the designated regions, shown in Fig. 3. We specify the range of skin color and extract pixels suitable for the range, referring to the paper (8). First, we convert the RGB values to HSV values. The skin tone ranges in HSV are shown in Eq. (1).

$$0 \leq H \leq 17, 15 \leq S \leq 170, 0 \leq V \leq 255 \quad (1)$$

Next, we convert the RGB values to YCbCr values. The skin tone ranges in YCbCr are shown in Eq. (2)

$$\begin{aligned} 0 \leq Y \leq 255, 85 \leq Cb \leq 135, \\ 135 \leq Cr \leq 180 \end{aligned} \quad (2)$$

Pixels within the ranges in Equations (1) and (2) are extracted as candidates, and N pixels are randomly selected among these pixels. The system tracks the selected N points using the Lucas-Kanade method and calculates the sequences of RGB values for the points.

3.2 Normalizing RGB signals using small windows

The angle between the surface of the face and the light source subtly differs due to the shape of the surface, and the intensity of illuminated light also differs depending on the parts of the face. These factors may cause large changes in the sequence of RGB values even when the face slightly moves; therefore, we need to select proper feature points with small changes in RGB values within small windows for the sequence.

For the i -th feature point from frame t to $t + 1$, the normalized amounts of change in RGB values are defined in Eq. (3).

Fig. 2. Two parts of face region

Fig. 3. Selecting feature points

$$\begin{pmatrix} dR_i^t \\ dG_i^t \\ dB_i^t \end{pmatrix} = \begin{pmatrix} \frac{R_i^{t+1} - R_i^t}{R_i^{t+1} + R_i^t} \\ \frac{G_i^{t+1} - G_i^t}{G_i^{t+1} + G_i^t} \\ \frac{B_i^{t+1} - B_i^t}{B_i^{t+1} + B_i^t} \end{pmatrix} \quad (3)$$

R_i^t stands for R value for i -th point at frame t , and dR_i^t for the variation of R from t to $t + 1$. Notations for G and B are the same as those for R. According to the paper (9), dR_i^t , dG_i^t and dB_i^t fall within the range of $[-0.02, 0.02]$ if the point is correctly tracked with no illumination changes.

However, since observation in actual environment is very noisy, it may not get enough points within the range. Therefore, we designed another way to select SN -best feature points for the small windows as follows. We use overlapped windows where the number of frames in the window is l and the number of overlapped frames is $\frac{l}{2}$. The

frame sequence of j -th window w_j starts from $\frac{l_j}{2}$ and ends at $\frac{l_j}{2} + l - 1$ in Eq. (4).

$$w_j = \left[\frac{l_j}{2}, \dots, \frac{l_j}{2} + l - 1 \right] \quad (4)$$

The maximum variance of color value for i -th feature point in the j -th frame is defined as m_{ij} .

$$m_{ij} = \max_{k \in w_j} |dR_i^k| + \max_{k \in w_j} |dG_i^k| + \max_{k \in w_j} |dB_i^k| \quad (5)$$

The index sequence for the ascending sort of m_{ij} for j -th window is defined as l_j . In other word, $l_j(0)$ is the index of the point with the smallest m_{ij} . We define the sequences of averages for top- SN values of color in window w_j as SR_{w_j} , SG_{w_j} and SB_{w_j} . As a notation for sequence, we use $\text{Seq}[x]$ in the following equation.

$$\begin{aligned}
SR_{w_j} &= \text{Seq}_{k \in w_j} \left[\text{average}_{0 \leq p \leq SN} R_{I_j(p)}^k \right] \\
SG_{w_j} &= \text{Seq}_{k \in w_j} \left[\text{average}_{0 \leq p \leq SN} G_{I_j(p)}^k \right] \\
SB_{w_j} &= \text{Seq}_{k \in w_j} \left[\text{average}_{0 \leq p \leq SN} B_{I_j(p)}^k \right]
\end{aligned} \tag{6}$$

These sequences need to be normalized to reduce the discrepancy among windows. We calculate $\overline{SR_{w_j}}$, $\overline{SG_{w_j}}$ and $\overline{SB_{w_j}}$ for the average values in a window and finally define $\overline{R_{w_j}}$, $\overline{G_{w_j}}$ and $\overline{B_{w_j}}$ as follows.

$$\begin{aligned}
\overline{R_{w_j}} &= \frac{SR_{w_j}}{\overline{SR_{w_j}}} - 1 \\
\overline{G_{w_j}} &= \frac{SG_{w_j}}{\overline{SG_{w_j}}} - 1 \\
\overline{B_{w_j}} &= \frac{SB_{w_j}}{\overline{SB_{w_j}}} - 1
\end{aligned} \tag{7}$$

3.3 CHROM method

We calculate the chrominance signal from the sequences of RGB values to amplify the color of blood flow. The chrominance signal is defined in Equations. (8), (9) and (10).

$$X_{w_j} = 3\overline{R_{w_j}} - 2\overline{G_{w_j}} \tag{8}$$

$$Y_{w_j} = 1.5\overline{R_{w_j}} + \overline{G_{w_j}} - 1.5\overline{B_{w_j}}$$

$$P_{w_j} = BPF(X_{w_j}) - s BPF(Y_{w_j}) \tag{9}$$

$$s = \frac{\sigma(X_{w_j})}{\sigma(Y_{w_j})} \tag{10}$$

In these equations, $BPF(X_{w_j})$ and $BPF(Y_{w_j})$ are filtered sequences of X_{w_j} and Y_{w_j} respectively, with a Butterworth bandpass filter (BPF) of [0.8, 4.0] Hz ([48, 240] bpm), P_{w_j} is an amplified signal for window w_j , and s is a ratio of deviations of X_{w_j} and Y_{w_j} . Finally, P_{chrom} is calculated with the sum of P_{w_j} multiplied by Hanning window.

3.4 Selection in relation to environment

The CHROM method amplifies the color of blood flow to measure heart rates accurately under white lights, while simple G-channel performs better in other environments, like

under neutral white light bulb. Therefore, the final heart rate signal P should be selected in relation to their environments shown in Eq. (11).

$$P = \begin{cases} P_{chrom}, & \text{if white light or sunlight} \\ P_G, & \text{otherwise} \end{cases} \tag{11}$$

P_G stands for the sequence of $BPF(\overline{G_{w_j}})$ multiplied by Hanning window.

The FFT is applied to the heart rate signal P with zero padding extension to calculate the main frequency f_p . The estimated heart rate, HR_{estm} , is the multiplication of f_p and 60 in Eq. (12).

$$HR_{estm} = f_p \times 60 \tag{12}$$

4. Experiment

4.1 Experimental setting

For this experiment, we took videos of 10 men and 13 women by the camera of the iPhone SE 2nd generation. The pixel resolution of the video was 3840×2160 (4k), and a frame rate was 30 fps. The lengths of the videos were 30 to 45 seconds. The light source was either sunlight or indoor light. The distances between the camera and subject were 0.3 to 0.5 meters. The subjects attached Citizen electronic blood pressure monitor (CH-602B) and sat in front of the camera without moving.

We implemented the system using OpenCV (Python). We extracted 2000 feature points ($N=2000$) from image and set $SN=200$ as the parameter in section 3.

4.2 Experiment

We evaluated the accuracy of the proposed system, compared with the combination of the other techniques proposed in the previous research. We tested two types of tracking methods: region tracking and feature point tracking. When using region tracking, we set the fixed ROI with rectangle-shape on the forehead, tracked the relative area against whole face region and calculated the average values in the ROI. When using feature point tracking, we tracked several randomly selected feature points on the face using the Lucas-Kanade method to calculate the average values for the points. We also tested three types of signal extraction methods: the use of G-channel, extraction with ICA, and CHROM method. We further evaluated the system based on

Fig. 4. Plots for estimate HR and theoretical HR

the motion detection proposed by Guha Balakrishnan et al. (6).

As a performance measure, mean absolute error (MAE) is calculated for each of them in Eq. (13).

$$MAE = \frac{\sum |HR_{Ref}(n) - HR_{estm}(n)|}{VN} \quad (13)$$

$HR_{Ref}(n)$ indicates the theoretical value measured with the electronic blood pressure monitor and $HR_{estm}(n)$ indicates the estimated value by respective methods. In Eq. (13), n is the index of each video, and VN is the number of videos used in the experiment.

4.3 Results of experiment

The performance results of the experiment with 72 facial videos are shown in Table 1. The accuracies in region tracking were lower than those in feature point tracking because forehead areas of many subjects were hidden by hair, and the size of ROI could not be averaged. Our system determined the heart rate with MAE of 3.4 bpm. CHROM method was the best in the existing methods with MAE of 6.6 bpm, but our system exceeded the CHROM by using n-best feature points for the small windows, even when the feature point tracking was applied to the CHROM. The performance of motion-based method was worse than that of color-based method.

From the viewpoint of processing time, feature point tracking was faster than region tracking because the cost for face detection and skin-area extraction was higher, while the cost for pixel tracking, using Lucas-Kanade method, was lower.

Tables 2 and 3 show MAE of our method in gender and lighting categories. Figure 4 shows the correspondences between theoretical and estimated HRs. In Fig. 4, plots were marked on the line if estimated value was the same as the

Table 1. MAE for applied methods.

	G	ICA	CHROM	Head(6)	Proposed
Region tracking	25.4	26.6	16.1		
Pixel tracking	7.0	9.4	6.6	14.8	3.4

Table 2. MAE for proposed system in gender category.

	Proposed
Men	3.3
Women	3.5
Total	3.4

Table 3. MAE for proposed system in lighting category.

	Proposed
White light bulb	2.3
Neutral white light bulb	6.2
Sunlight	3.5
Total	3.4

theoretical value. Plots in Sub figures (2) and (3) of Fig. 4 were color-coded in different colors for gender and light sources.

Although there was no explicit difference in the accuracy of the proposed system by gender, there was a difference in success rates of tracking because the average number of feature points that could be tracked was 1587 points for men, while 1094 points for women. We presume the surface of the women's skin was covered with cosmetics and smoother in texture.

We used G-channel in the environment under the neutral white light and CHROM method in the other environments according to Eq. (11), but the accuracy under the neutral white light bulb source was the lowest of all in Table 3. The accuracy became worse than this result when we applied CHROM method to the neutral white light source. Therefore, the selection of G-channel was reasonable, and

then we need more examination how to construct better signal from combination of RGB values.

5. Discussion

Most systems for heart rate measurement require human, skin contact. This poses a risk of contact infection when infectious diseases, like coronavirus, are rampant in society. Therefore, we focus on the remote monitoring of heart rate, utilizing Remote photoplethysmography (rPPG). Juan proposed a method of measurement, using the average of the skin tone of the forehead region, analyzed by the G-channel of RGB, called ROI. This method worked well when the surface of the face is uncovered but there are noisy objects, like hair and beard, that distort remote monitoring HR data in actual situations. To solve this problem, our system uses a pixel tracking method, where pixels are selected from a face region and tracked with Lucas-Kanade method, and then noiseless points are carefully extracted.

De Haan proposed a method to amplify the color of blood flow, called CHROM method. This method worked efficiently under the lighting of white light bulbs, but erroneous under the lighting of neutral white light bulbs. Our system combined advantages of CHROM and G-channel methods for signal extraction, selecting CHROM method under white light and sunlight, and G-channel otherwise.

The combination of feature point tracking and signal extraction yields greater accuracy of heart rate measuring than conventional remote monitoring systems.

Our experiments described in section 4 were conducted using high-resolution camera with 4K pixels, but we also tried to apply our method to low-resolution images, like video stream in an online communication tool whose resolution was 1280×720 . It successfully worked for men in most cases, but did not always work for women. We presumed this failure came from the difficulty for tracking feature points. Therefore, we need to extend tracking method to deal with such low-resolution images.

6. Conclusion

In this paper, we proposed a system that is tolerant to physical and environmental factors with a method to effectively combine feature point tracking and signal extraction. The proposed system was able to estimate the heart rate with accuracy of 3.4 bpm evaluated with MAE. The system overcomes the performances of the existing systems. However, there was a slight difference in the

estimation accuracy due to environmental difference in terms of lighting. To build a more stable system, it is necessary to divide the measuring conditions into smaller data points, and to carefully construct models and methods suitable for each condition. In addition, we should propose a highly convenient system that automatically switches the methods according to the environments.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP20K11777.

References

- (1) Juan, L., Cuevas, A. and Sanchez, R. E., "Simplified Heart Rate Measurement from Video.", *Proceeding of the 3rd International Workshop on Interactive and Spatial Computing*, 23-28, 2018.
- (2) Yang, Y., Liu, C., Yu, H., Shao, D., Tsow, F. and Tao, N., "Motion robust remote photoplethysmography in CIELab color space.", *Journal of Biomedical Optics*, 2016;21(11):117001.
- (3) Wiede, C., Richter, J., Apitzsch, A., KhairAldin, F. and Hirtz, G., "Remote Heart Rate Determination in RGB Data.", In *Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods – ICPRAM*, 240-246, 2016.
- (4) De Haan, G. and Jeanne, V., "Robust Pulse Rate from Chrominance-Based rPPG.", *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, 2878-2886, 2013.
- (5) Wu, H., Rubinstein, M., Shih, E., Guttag, J., Durand, F., and Freeman, W. T., "Eulerian video magnification for revealing subtle changes in the world.", *ACM Transactions on Graphics (Proceedings SIGGRAPH 2012)* 31, 4 (2012), 21434-21445.
- (6) Balakrishnan, G., Durand, F. and Guttag, J., "Detecting Pulse from Head Motions in video.", *IEEE Computer Vision and Pattern Recognition conference*, 2013.
- (7) Shan, L. and Yu, M., "Video-based heart rate measurement using head motion tracking and ICA.", *2013 6th International Congress on Image and Signal Processing (CISP)*, vol. 1, pp. 160-164, 2013.
- (8) Djamila, D., Cheref, M. and Larabi, S., "Zero-sum game theory model for segmenting skin regions.", *Image and Vision Computing* 99, 2020.
- (9) Wang, W., Stuijk, S. and De Haan, G., "Exploiting Spatial Redundancy of Image Sensor for Motion Robust rPPG.", *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, 415-425, 2015.