# Contact-Free Heart Rate Measurement Using a Camera

Kual-Zheng Lee, Pang-Chan Hung, and Luo-Wei Tsai
Information & Communications Research Lab.
Industrial Technology Research Institute, HsinChu, Taiwan
kzlee, PC_Huang, lwtsai@itri.org.tw

***Abstract*—In this paper, a contact-free heart rate measurement method using an ambient light camera is proposed. The color features of skin-like pixels are used for time domain frequency analysis, in which a skin color classifier is applied. A novel data adjustment scheme is further integrated to automatically expand sampling time length and increase measurement precision. Five people of different age, six cameras of different type, and four different body parts have been tested in this paper. Experimental results show the effectiveness of the proposed method comparing with a pulse oximeter device. The advantages of the proposed method include: 1) it uses a low-cost ambient light camera; 2) it eases the discomfort of people based on contact-free methodology; 3) it measures multiple persons' heart rates fully automatic; and 4) it can be applied in multiple parts of a human body such as head and neck, arm, and palm regions.**

***Keywords-heart rate, skin color detection, ambient light camera***

## I. INTRODUCTION

The heart rate is one of important physiological signals of a human body, so medical professionals or individuals usually measure the heart rate to judge the physiological state [1, 2]. For example, resting heart rate has been identified as an independent risk factor (comparable with smoking, dyslipidemia or hypertension) for cardiovascular disease [2].

Existing heart rate measurement (HRM) apparatuses are mainly contact-based devices, such as pulse oximeter [3, 4], sphygmomanometer, and electrocardiograph devices. A pulse oximeter uses optical properties to measure blood oxygen concentrations and heart rate values, in which a light emitter with red and infrared LEDs is used that shines through a reasonably translucent site with good blood flow such as fingers, and then signals for HRM are obtained by measuring the light of transmission or reflectance. Pulse oximeters that attach to the fingertips or earlobes are inconvenient for subjects and the spring-loaded clips can cause pain if worn over a long period of time. Sphygmomanometers could not measure heart rate at continuous time points. Electrocardiograph devices are cost expensive and require subjects to wear adhesive gel patches or chest straps that may cause skin irritation and discomfort.

In order to ease the discomfort of subjects and measure multiple persons' heart rates at a time, methods for contact-free HRM have been developed, which can be categorized into two classes:

- *Far-infrared light sources* [5-7]: a thermal infrared camera is used to sense the information contained in the thermal signal emitted from major superficial vessels of a person and then analyzes the frequency to measure the heart rate. In general, thermal cameras are cost expensive and may be difficult to be applied in low-cost products.
- *Ambient light sources* [8-10]: an ambient light camera is used to capture human faces for HRM. Multiple groups of regions on the human face are labeled manually or a whole face region is used to analyze a periodic variation caused when blood flows through the human face, so as to measure the heart rate. It is merely applicable in front faces and needs to use a face detector with a high computation amount. Furthermore, the human face region includes many meaningless regions without heart rate information, for example, eyebrows, eyes, nares, or beards, which may affect the accuracy.

It is known that skin is one of the most important features for contact-free HRM. To our knowledge, there is no method applying skin detection to solve the HRM problem so far. In this paper, a skin color classifier based on neural networks [11] is used to identify skin-like pixels from captured images. The skin-like pixels are then labeled into regions and tracked with a mean-shift tracker [12, 13]. The RGB color values in each object are extracted and weighted as a feature value for time domain frequency transformation [14]. In order to achieve higher precision during HRM, a data adjustment scheme is proposed to automatically expand sampling time length and increase measurement precision. The proposed method is tested by five people of different age, six camera of difference type, and four different body parts. Acceptable results are obtained comparing with a pulse oximeter product. For example, the error of the proposed method is smaller than 3.71% with various skin color thresholds on measuring the head/neck part.

The proposed method achieves contact-free HRM with a low-cost ambient light camera, which eases the discomfort of people based on contact-free methodology. Multiple persons' heart rates can be measured once fully automatic and it can be applied in multiple parts of a human body such as head and neck, arm, and palm regions. Moreover, it's practicable on embedded systems since no complex face detector is required. The proposed method is applicable in fields such as in general health assessment, ill physiological and mental conditions prediction, polygraph testing, intent identification, smart room, human-computer interaction, etc.

978-0-7695-4683-4/12 $26.00 © 2012 IEEE
DOI 10.1109/CRV.2012.27

## II. Related Work

Existing optical-based HRM methods can be categorized into two classes: contact-based and contact-free methods. Contact-based methods are usually implemented by using near-infrared light sources. For example, Blouin [3] proposed a HRM device that emits near-infrared light into skin region and detects the scattered light signal from the region to analyze user's heart rate. Mix and Viala [4] proposed a goggles-like device with infrared sensing module located at earlobe to observe blood flow and then measure the heart rate. To our knowledge, contact-based methods are mature so far and have been implemented as various hardware devices.

Contact-free methods are typically implemented by using far-infrared light or ambient light sources. Chekmenev *et al*. [6] used a passive thermal infrared sensor to capture the heat pattern from superficial arteries, and a blood vessel model was used to describe the pulsatile nature of the blood flow. Pavlidis *et al*. [5] used a thermal infrared camera to capture the images of large superficial blood vessel such as carotid and temple, then analyzed the changes of temperature in vessel areas to measure the heart rate. In [7], Garbey *et al*. proposed a method for estimation of blood flow speed and vessel location from thermal infrared video, which analyzed the areas of arm and wrist to measure the heart rate. The above mentioned methods are mainly dependent on cost expansive thermal infrared cameras, which may be difficult to be applied in low-cost products.

Recently, Verkruysse *et al*. [8] proposed a contact-free HRM method using ambient light sources. They used color movies captured by an ambient camera as test data and manually identified some regions of interest (ROI) on a human face for analyzing. The fast Fourier transform was performed to determine the power of frequencies provided by the RGB colors in a ROI. They showed that pulse measurements from the human face are attainable with normal ambient light as the illumination source.

Based on the theory of [8], Pol *et al*. [9, 10] proposed an automatic contact-free HRM method by using a face detector. The frontal human faces in images were detected and tracked as the ROIs mentioned in [8]. The independent component analysis (ICA) was further applied on the RGB color channels to uncover independent signals. However, the human face region includes many meaningless regions without heart rate information, for example, eyebrows, eyes, nares, or beards, which may affect the accuracy. Moreover, it is merely applicable in front faces and needs to uses a face detector with a high computation amount.

## III. Frequency Transformation

Frequency transformation is a mathematical operation that expresses a mathematical function of time as a function of frequency, known as its frequency spectrum. In this paper, the discrete-time Fourier transform (DFT) [14] is used to analyze color signals, its equation can be defined as:

$$X(b) = \sum_{t=1}^{T-1} f_t \times \mathrm{e}^{\frac{-i2\pi tb}{T}}, \quad b = 0,1,\dots,T-1, \qquad (1)$$

where $T$ is the data count to be transformed, $t$ is a time point, e is the base of natural logarithm, $i$ is an imaginary unit, $f_t$ is the data value at time $t$, $X(b)$ is magnitude of $b^{\text{th}}$ band after the transformation. So a magnitude set corresponding to $T$-1 bands can be obtained through transformation to form a power spectrum. In the HRM problem, the data count $T$ is the number of frames sampled from a camera, which is a major factor that influences the measurement accuracy and the sampling time length for HRM.

Table I shows the relationships between camera frame rate (fps, frame per second), DFT data count $T$, measurement precision (bpm, beat per minute), and sampling time length (sec). Details of calculating measurement precision can be found in next chapter. It reveals that a higher frame rate cannot improve the precision and shorten the sampling time length simultaneously. Since the value of multiplying measurement precision and sampling time length is a constant, i.e., 60, a good precision cannot be obtained in a short sampling time.

According to our experiences, most capture devices can provide 30fps video data and the first heart rate should be reported in 10 seconds for comfortable. The data count $T$=256 is considerable at the beginning of measurement with the precision 7.03bpm and sampling time length 8.53sec. Since HRM is a continuous process, the collected data will be increased with the time length of measurement. Therefore a higher precision can be obtained if the measurement time is long enough because a large data count can be used. In this paper, a data adjustment scheme is proposed to implement this concept. Details can be found in next chapter.

TABLE I. Measurement precision and sampling time at various camera frame rate and DFT data count.

| Frame Rate | DFT Data Count | | | | | |
|---|---|---|---|---|---|---|
| | 128 | 256 | 512 | 1024 | 2048 | 4096 |
| 120 | 56.25 (1.07s) | 28.13 (2.13s) | 14.06 (4.27s) | 7.03 (8.53s) | 3.52 (17.07s) | 1.76 (34.13s) |
| 60 | 28.13 (2.13s) | 14.06 (4.27s) | 7.03 (8.53s) | 3.52 (17.07s) | 1.76 (34.13s) | 0.88 (68.27s) |
| 30 | 14.06 (4.27s) | 7.03 (8.53s) | 3.52 (17.07s) | 1.76 (34.13s) | 0.88 (68.27s) | 0.44 (136.53s) |
| 15 | 7.03 (8.53s) | 3.52 (17.07s) | 1.76 (34.13s) | 0.88 (68.27s) | 0.44 (136.53s) | 0.22 (273.07s) |

## IV. Proposed Method

The flowchart of the proposed method is shown in Fig. 1, which comprising the steps of skin color detection, region labeling, object tracking, feature extraction, frequency transformation, and HRM with data adjustment scheme.

### *A. Skin Color Detection*

Skin detection using color is very effective and thus is useful for algorithms for face detection, hand gesture analysis, and objectionable image filtering to reduce the search space [15]. In this paper, a neural network based skin color classifier is used [11] with 3-25-1 multi-layer perceptron architecture. More than 200 images are collected and manually labeled skin/non-skin pixels for training. Taking color R, G, B values of a pixel as the input, the

classifier outputs the regression value which indicate the likelihood of skin color. By setting the threshold $T_{skin} \in [0, 255]$, each pixel can be classified as skin or not.

Let $I_t$ be the frame data at time $t$ and $\mathbf{p}_t$ be the color value of a pixel $\mathbf{x}_t$ in frame $I_t$, where $\mathbf{p}_t$ comprising 8-bit colors for red, green, and blue channels. The skin color detection process judges a pixel that is similar to a skin color in the pattern information and outputs a binary flag value whether the pixel in the pattern information is a skin-like point, denoted as $\mathbf{x}_t^s$ with a color value $\mathbf{p}_t^s$.

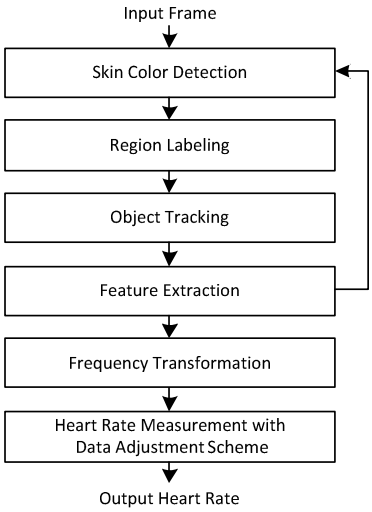



Figure 1. Flowchart of the proposed method.

## B. Feature Extraction of an Object

According to the binary skin flag values, the connected component labeling method [12] is applied to label adjacent skin-like points with the same label into a region. A region with an excessively large or small area can be filtered according to preset thresholds, and a region that falls within the thresholds is regarded as a candidate object. After the labeling process, we have the region count $N_t$ in frame $I_t$ and region information $\mathbf{R}_t^i$ comprising the total count $n_t^i$ and coordinates of skin-like pixels in region, where $i = 1,2, \dots N_t$. Therefore the feature vector $\mathbf{u}_t^i$ of $i$-th region in frame $I_t$ can be defined as:

$$\mathbf{u}_t^i = \frac{1}{n_t^i}\sum(\mathbf{p}_t^s \times \delta), \text{ where } \delta = \begin{cases} 1 & \text{if } \mathbf{x}_t^s \in \mathbf{R}_t^i, \\ 0 & \text{else.} \end{cases} \quad (2)$$

In order to collect the feature values of all regions in continuous frames, the mean-shift object tracker [13] is applied. Let $M_t$ be the number of tracked objects and $\mathbf{O}_t^j$ be the information of $j$-th tracked object comprising a feature vector $\mathbf{v}_t^j = (c_R, c_G, c_B)$, where $j = 1,2, \dots M_t$, and $c_R, c_G$ and $c_B$ are the values of red, green, and blue channels, respectively. Note that $\mathbf{v}_t^j = \mathbf{u}_t^i$ if region $\mathbf{R}_t^i$ belongs to object $\mathbf{O}_t^j$.

The frequency transformation is performed with a single color channel at a time. In the investigated problem, it is time consuming to run the transformation for red, green, and blue channels individually. Moreover combination of the frequency results of different color channels is another issue. In this paper, a weighted sum is used for representing the feature value of $j$-th object, defined as:

$$f_t^j = c_R w_R + c_G w_G + c_B w_B. \quad (3)$$

Therefore the magnitudes of all bands can be further obtained through (1) by using $f_t^j$ as the input data value. In [8], it has been shown that the green channel featuring the strongest plethysmographic signal comparing with red and blue channels. According to our simulation, ($w_R$, $w_G$, $w_B$) = (-0.17, 0.97, -0.18) performs well in most case, which are estimated by a brute-force strategy. Fig. 2 shows an example of RGB traces and the weighted feature values in a period of time. The variant of red channel is reduced through the weighted process.

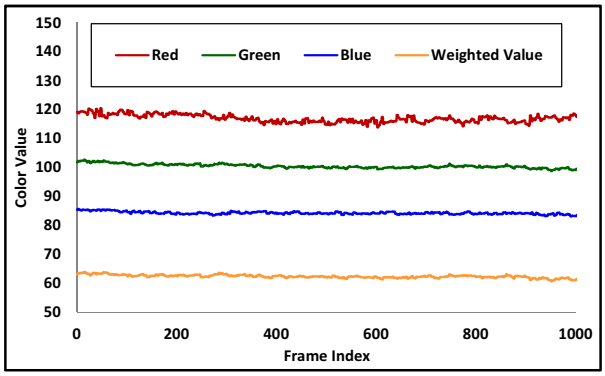



Figure 2. RGB Traces and weighted feature values in a period of time .

## C. Heart Rate Measurement with Data Adjustment Scheme

The heart rate $H(b)$ can be calculated with beat per minute (bpm) unit represented by the band $b$ according to a time interval of adjacent frames in the pattern information. A frame rate of the pattern information is set to be $K$ fps, and the transformation between the band $b$ and the heart rate $H(b)$ bpm can be defined as:

$$H(b) = \frac{60 \times K \times b}{T}. \quad (4)$$

A rational minimum and maximum value for the heart rate are set. For the target $\mathbf{O}_t^j$, a band $b_t^j$ having the largest magnitude in the rational heart rate range is taken and an equation for transformation is used to calculate a heart rate $H_M^j$ of the target $\mathbf{O}_t^j$. Taking a rational heart rate in [40, 240] as an example, the equation for calculating the band $b_t^j$ is as follows:

$$b_t^j = \arg\max_b \mathbf{X}^j(b), \quad \frac{40 \times T}{60 \times K} \le b \le \frac{240 \times T}{60 \times K}. \quad (5)$$

As the descriptions in Chapter III, data count $T$ is a major factor that influences the time length required for HRM. In this paper, a data count adjustment scheme is proposed to dynamically adjust $T$, so as to rapidly obtain a frequency transformation result.

In a preset time period for HRM, the smallest and largest data counts can be obtained according to a frame rate of a video capture device. Several data counts are selected as preset parameters in ascending order in a time period, which can be defined as a set $\mathbf{W}=\{w_1, w_2, \ldots, w_m\}$. The values in $\mathbf{W}$ are arranged in ascending order and a total number of elements is $|\mathbf{W}|$, and an initial value $m = 1$, so that the data count $T = w_m$. Referring to the data adjustment scheme show in Fig. 3, the data count will be expanded if the sampled data count meets the defined criteria. Therefore, a small amount of data is used for the frequency transformation in the early period of video capture, so the heart rate values can be obtained within a very short time. As the sampling time length increases, it can be automatically switch to the mode of higher data count with better precision. In our program, $\mathbf{W}=\{256, 512, 1024\}$ is defined.

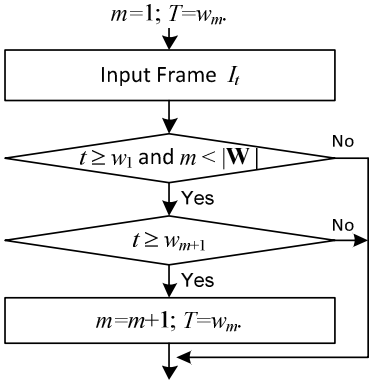



Figure 3. Flowchart of data adjustment scheme.

## V. RESULTS

### A. Experimental Procedure

In order to verify the performance of the proposed method, four USB cameras (Asus UVC2M, Logitech V-UU22, Logitech V-UAV35, and Samsung PWC-3800) and two network cameras (Axis P1311 and Axis 1031W) are used as the capture devices. The settings of USB cameras are: 1) resolution 640×480; and 2) frame rate 30fps. The setting of network cameras are: 1) resolution 640×480; 2) frame rate 30fps; 3) H.264 video codec; and 4) bit rate 1024kbps. Fig. 4 shows the snapshots of various body part used in experiments. A pulse oximeter product Rossmax SA310 (called *SA310* for short) is used to provide ground truth data. All the programs are implemented by the language C/C++ in Visual Studio 2008 on x86 computers.

The heart rate value $H_{\mathrm{M}}^{j}$ is denoted as $H_{\mathrm{M}}$ for short since there is only one target with one region in all test cases. Let $H_{\mathrm{D}}$ be the heart rate value measured by SA310 device, the measurement error $E$ can be defined as:

$$E = \frac{|\mathrm{H_D} - \mathrm{H_M}|}{\mathrm{H_D}}. \quad (6)$$

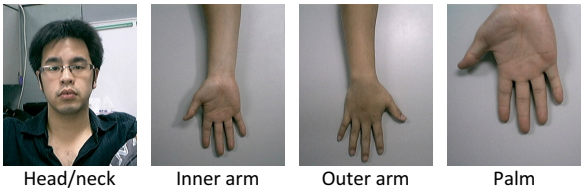



Figure 4. Snapshots of various body part used in experiments.

### B. Various People

In this experiment, 47 videos captured from the head/neck part of five males are used as the test data. Each video comprises more than 5000 frames. The parameters are fixed as skin threshold $T_{\mathrm{skin}}$=80 and data count $T$=1024. In Table II, the heart rate variation of each person can be observed from the standard deviation value. It reveals that an average error 3.22% can be achieved for all persons. The error of M2 case is large 5.48% (±3.29 bpm for 60bpm) because the subject has significant movements during testing. Note that an error value is averaged from the errors of all test frames, instead of the error between average $H_{\mathrm{D}}$ and $H_{\mathrm{M}}$.

TABLE II. EXPERIMENTAL RESULTS OF VARIOUS PEOPLE.

| Case | Age | Video Count | Avg ($H_D$) | Std ($H_D$) | Avg ($H_M$) | Std ($H_M$) | Error |
|---|---|---|---|---|---|---|---|
| M1 | 26 | 17 | 53.04 | 1.07 | 52.49 | 1.32 | 1.63% |
| M2 | 35 | 6 | 60.56 | 1.81 | 57.70 | 2.56 | 5.48% |
| M3 | 31 | 12 | 70.34 | 1.10 | 68.13 | 3.35 | 3.82% |
| M4 | 34 | 6 | 70.46 | 2.46 | 69.11 | 4.04 | 3.16% |
| M5 | 34 | 6 | 68.27 | 2.34 | 66.69 | 4.19 | 4.25% |

Note: Avg is average value and Std is standard deviation.

### C. Various Cameras

The test videos and parameters in this experiment are the same with those in section B. Each test case comprises at least one video of each person. In Table III, it shows that the proposed method performs well with USB cameras (smaller than 2.97%). However, the errors of network cameras are high, especially in the C5 case 5.04% (±3.02 bpm for 60 bpm). Due to the variations of video encoding complexity, encoded data size, and network state, the time interval between two frames of a network camera may be unstable after video decoding. One way to improve the performance might port the measurement kernels into a network camera and run the measurement before video encoding.

TABLE III. EXPERIMENTAL RESULTS OF VARIOUS CAMERA.

| Case | Camera Type | Video Count | Error |
|---|---|---|---|
| C1 | Asus UVC2M | 8 | 2.62% |
| C2 | Logitech V-UU22 | 8 | 2.40% |
| C3 | Logitech V-UAV35 | 8 | 2.42% |
| C4 | Samsung PWC-3800 | 7 | 2.97% |
| C5 | Axis P1311 | 8 | 5.04% |
| C6 | Axis 1031W | 8 | 3.93% |

### D. Various Body Parts

In this experiment, the videos of head/neck, inner arm, outer arm, and palm parts are used, which are captured from five males. The data count $T$ is fixed as 1024. The experimental results can be found in Table IV. It shows that the proposed method can be applied in various body parts and obtain a maximum error of 6.74% in palm case (±4.04 bpm for 60 bpm). Best results with various skin thresholds are obtained in head/neck cases. It slightly indicates the accuracy may be dependent on the distance between a body part to heart while using contact-free strategies. On the other hand, we find that the shadows on the desk may affect the accuracy of skin color detection. A higher skin threshold value may decrease the shadow effects but also reject some useful pixels on skin area. According to our simulation, it is recommended to use a skin threshold $T_{skin}$ smaller than 150.

TABLE IV. EXPERIMENTAL RESULTS OF VARIOUS BODY PARTS.

| Part | Video Count | $T_{skin}$ | | | |
|---|---|---|---|---|---|
| | | 60 | 80 | 100 | 120 |
| Head/neck | 47 | 3.23% | 3.22% | 3.22% | 3.25% |
| Inner arm | 35 | 3.38% | 3.35% | 3.42% | 3.40% |
| Outer arm | 35 | 6.69% | 5.39% | 5.15% | 5.23% |
| Palm | 35 | 6.74% | 6.70% | 6.68% | 6.59% |

### E. Effects of Various Data Count

In this experiment, a video of the person M1 in Section B is selected to demonstrate the effects of various data count. The experimental results with a fixed skin threshold $T_{skin}$= 80 and various data count $T$ can be found in Fig. 5. The measurement is performed between $I_{2610}$ and $I_{5500}$ thus the number of measurements in all cases is 2890. The average errors of $T$=256, 512, 1024, and 2048 are 4.72%, 2.24%, 1.23%, and 2.31%, respectively.

It can be found that the accurate heart rate of M1 is in the range of [50, 54]. In this case, two best results of $T$=256 are 49.22 and 56.25 since its heart rate precision is 7.03 according to Table I. Therefore the major factor is the inaccurate precision while data count is small. However, a higher data count may not perform better because the heart rate is not stationary. In the case of $T$=2048, it collects the data in 68.27 sec to measure a heart rate value, which may not response correct heart rate immediately. According to our experiences, a suitable value $T$=1024 is recommended for most cases if the data adjustment scheme is not implemented.

### F. Performance Comparison

The method presented in [10] (called *Poh's method* for short) is implemented for performance comparison, which uses a face detection algorithm provided by the open source library OpenCV. In Poh's method, the ICA is used to improve its accuracy. However the performance of ICA is not significant (even worse) comparing with non-ICA mode in our simulation. Therefore the reported values of Poh's method are obtained without the ICA process.

In this experiment, 47 head/neck videos are used as test data and the errors are shown in Table V. The errors of both methods are high with a small data count $T$=256 (larger than 6.83%) since its precision is not good. In most cases, data count 1024 and 2048 perform better for both methods. From Table V, it reveals that the performance of proposed method is better than Poh's method in all cases because meaningless regions of a face are filtered by the skin color classifier. The improved accuracy of our method for all cases is in the range of 0.06% and 1.35%.

Note that the errors of Poh's method in Table V are better than the errors reported in their paper. In real-time programs, the frame rate of a camera may be slightly unstable. We have implemented an interpolation process to fix the time interval of two frames into a constant 1/30 second. Such process may also improve the performance of Poh's method.

TABLE V. RESULTS OF PERFORMANCE COMPARISON.

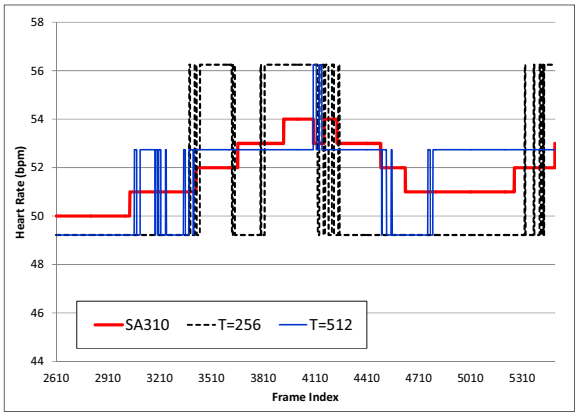
| Data Count | Poh's Method [10] | Propose Method with Various Skin Thresholds (improved Value to [10]) | | | | |
|---|---|---|---|---|---|---|
| | | 40 | 60 | 80 | 100 | 120 |
| 256 | 8.18% | 6.83% (1.35%) | 6.93% (1.25%) | 6.92% (1.26%) | 6.92% (1.26%) | 7.54% (0.64%) |
| 512 | 5.13% | 4.45% (0.68%) | 4.45% (0.68%) | 4.54% (0.59%) | 4.60% (0.53%) | 5.07% (0.06%) |
| 1024 | 3.95% | 3.37% (0.58%) | 3.20% (0.75%) | 3.22% (0.73%) | 3.38% (0.57%) | 3.71% (0.24%) |
| 2048 | 3.87% | 3.32% (0.55%) | 3.25% (0.62%) | 3.23% (0.64%) | 3.33% (0.54%) | 3.67% (0.20%) |
| 4096 | 4.62% | 3.80% (0.82%) | 3.39% (1.23%) | 3.67% (0.95%) | 3.64% (0.98%) | 3.89% (0.73%) |

## VI. CONCLUSIONS AND FUTURE WORK

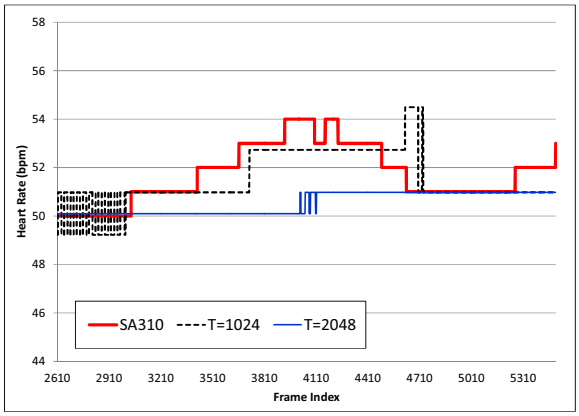
In this paper, a contact-free HRM method is proposed based on skin color detection. Comparing with previous contact-free HRM methods, the proposed method can be applied in multiple parts of a human body to automatically measure the heart rate, such as head /neck, arm, and palm regions. Experimental results show that the error of the proposed method is smaller than 3.71% with a data count $T$=1024 and various skin color thresholds on measuring the head/neck part. The advantages of the proposed method include: 1) it uses a low-cost ambient light camera; 2) it eases the discomfort of people based on contact-free methodology; 3) it measures multiple persons' heart rates fully automatic; and 4) it can be applied in multiple parts of a human body such as head and neck, arm, and palm regions.

In real-world applications, there may exists some *noise* pixels in the scene having similar colors to skin pixels. For solving this problem, a mask can be displayed on the screen to highlight the hot zone for measurement, and only the skin-like pixels in the hot zone are regarded as the feature points of an object. We have implemented this function on an Apple iPad2 platform and it works well.

In order to improve the robustness in real-world applications, two key issues have to be handled in the future: light effects and body movements. For the former issue, color enhancement algorithms such as Retinex may be applicable to decrease the light effects. For the latter issue, more robust object tracking algorithms can be further integrated to collect stable feature values. Besides, the




(a)




(b)

Figure 5. Heart rate measurement with various data count.

proposed method can detect more than one skin regions from a person. The performance may be further improved by combining the measured heart rates from multiple body parts.

## REFERENCES


[1] E. N. Bruce, *Biomedical Signal Processing and Signal Modeling*. New York: John Wiley & Sons, 2001.

[2] S. Cook, M. Togni, M. C. Schaub, P. Wenaweser, and O. M. Hess, "High heart rate: a cardiovascular risk factor?," *European Heart Journal,* vol. 27, pp. 2387-2393, 2006.

[3] D. Blouin, "Optical sensor for sports equipment," US Patent 20070049813, 2007.

[4] J. Mix and R. Viala, "Heart rate monitor," US Patent 20100016741, 2010.

[5] I. Pavlidis, J. Dowdall, N. Sun, C. Puri, J. Fei, and M. Garbey, "Interacting with human physiology," *Computer Vision and Image Understanding,* vol. 108, pp. 150-170, 2007.

[6] S. Y. Chekmenev, A. A. Farag, and E. A. Essock, "Thermal imaging of the superficial temporal artery: an arterial pulse recovery model," in *Proc. Computer Vision and Pattern Recognition*, 2007, pp. 1-6.

[7] M. Garbey, A. Merla, and I. Pavlidis, "Estimation of blood flow speed and vessel location from thermal video," in *Proc. Computer Vision and Pattern Recognition*, 2004, pp. 356-363.

[8] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Optics Express,* vol. 16, pp. 21434-21445, 2008.

[9] 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, pp. 10762-10774, 2010.

[10] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomedical Engineering,* vol. 58, pp. 7-11, 2011.

[11] K. K. Bhoyar and O. G. Kakde, "Skin color detection model using neural networks and its performance evaluation," *Journal of Computer Science,* vol. 6, pp. 955-960, 2010.

[12] L. G. Shapiro and G. C. Stockman, *Computer Vision*. Upper Saddle River: Prentice Hall, 2001.

[13] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, pp. 564-577, 2003.

[14] B. Boashash, *Time-Frequency Signal Analysis and Processing - A Comprehensive Reference*. Oxford: Elsevier Science, 2003.

[15] S. L. Phung, A. Bouzerdoum, and D. Chai, "Skin segmentation using color pixel classification: analysis and comparison," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, pp. 148-154, 2005.