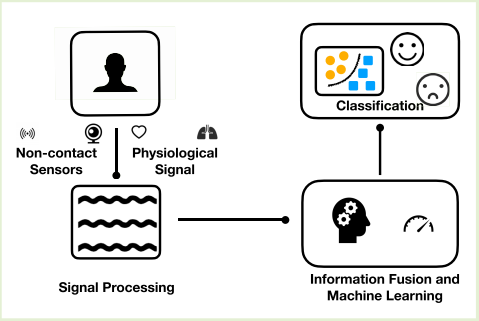
Non-Contact Dual-Modality Emotion Recognition System by CW Radar and RGB Camera

Li Zhang, *Student Member, IEEE*, Chang-Hong Fu, *Member, IEEE*, Hong Hong, *Senior Member, IEEE*, Biao Xue, *Member, IEEE*, Xuemei Gu, Xiaohua Zhu, *Member, IEEE*, and Changzhi Li, *Senior Member, IEEE*

***Abstract*—Emotion recognition has a significant impact on people's health and life quality. Recent studies have shown that emotion recognition can be achieved by analyzing audio-visual emotion channels and physiological signals. However, the main challenges are: 1. the audio-visual techniques can be easily fooled by fake facial expressions; 2. the contact physiological signal monitoring introduces additional stress and is usually unsuitable for long-term monitoring; 3. The non-contact physiological signal monitoring methods are easily affected by complex environmental conditions. In this paper, a non-contact dual-modality emotion recognition system is proposed. First, the respiratory and heartbeat signals are measured by radar and camera simultaneously. Then, a hybrid signal optimization approach is proposed to remove the influence of body motion and light conditions on the physiological signal. It includes a light-intensity-based scheme selecting proper heartbeat signal for different light conditions, and an optical-flow-based algorithm pruning signals with significant radial body motion. Finally, the features extracted from the optimized physiological signals are fused to train an emotion recognition system. As shown in the experimental results, the proposed system could achieve high classification accuracy of 89.6 % for 10-fold cross-validation at sample level, and 71.0 % for cross-validation at subject level. The respiratory and heartbeat signals by non-contact approaches are demonstrated to be reliable in emotion recognition.**



***Index Terms*—Non-contact, dual-modality, radar, video, emotion recognition.**

Manuscript received August 8, 2021; accepted August 11, 2021. Date of publication August 24, 2021; date of current version October 18, 2021. This work was supported in part by China Scholarship Council under Grant 201806845028, Grant 201906845009, and Grant 201906840058; in part by the National Key Research and Development Program of China under Grant 2020YFC2005302; in part by the National Natural Science Foundation of China under Grant 61871224; in part by the Key Research and Development Plan of Jiangsu Province under Grant BE2018729; and in part by the Fundamental Research Funds for the Central Universities under Grant 30917011316 and Grant 30917011319. The associate editor coordinating the review of this article and approving it for publication was Dr. Xiaojin Zhao. *(Corresponding author: Chang-Hong Fu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Nanjing Integrated Traditional Chinese and Western Medicine Hospital Human Research Project under Application No. 201812001.

Li Zhang, Chang-Hong Fu, Biao Xue, and Xiaohua Zhu are with the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: enchfu@njust.edu.cn).

Hong Hong is with the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China, and also with the Pazhou Laboratory, Guangzhou, Guangdong 210094, China.

Xuemei Gu is with Nanjing Brain Hospital Affiliated to Nanjing Medical University, Nanjing 210029, China.

Changzhi Li is with the Department of Electrical and Computer Engineering, Texas Tech University, Lubbock, TX 79409 USA.

Digital Object Identifier 10.1109/JSEN.2021.3107429

I. Introduction

EMOTION is a comprehensive statement that people produce when they meet their own needs. As the advanced function of the human brain, it guarantees the survival and adaptation of the organism, affecting people's learning, memory, and decision-making to varying degrees. Emotion is also a critical factor in personality traits and psychopathology and plays an essential role in daily life [1]. Emotions are the psychological and physiological states that accompany the process of cognition and consciousness and play a significant role in human communication and human-computer interaction. Recognizing other's emotion is crucial in social activities and benefits the relationship. However, the misjudgment of other people's emotion can lead to poor communication [2]. In human-computer interaction, emotion recognition techniques can enhance user experience and create more advanced artificial intelligence [1].

In recent years, emotion recognition has been highly valued by industry and academic [3], [4], including human-computer interaction, education, speech recognition and medical fields. It has carried out the common approaches for recognizing emotions using expressive modalities including facial expressions [5]–[8], speech [9]–[12], body gestures [13]–[16], and

other explicit features. However, these approaches have their limitations that cannot measure inner feeling [17]–[19] and are difficult to remove the influence of subjective factors such as fake expression. For example, people can smile when they are self-deprecating or even feel sad. At the same time, they may keep a poker face when they are happy. Emotions have not only outer behaviors but also complex neurological processes and physiological changes. The physiological response accompanying emotions that is dominated by the nervous and endocrine systems, is spontaneous and not easily controlled or faked [13].

The physiological response such as breathing rate, heartbeat rate, blood pressure, brain wave, etc. are tightly regulated and strong correspond with emotion [2]. Katsigiannis and Ramzan [1] presented DREAMER, a multimodal system using both electroencephalogram (EEG) and electrocardiogram (ECG) signals for emotion recognition. Chanel *et al.* [20] proposed the use of emotions detected from EEG signals for adapting the difficulty levels of a game. Jenke *et al.* [21] leveraged EEG signals to assess the emotional states of people. However, the traditional methods of monitoring those physiological signals are contact sensors like EEG and ECG, which are tightly attached to the skin. They make the subjects uncomfortable and introduce additional measurement error, since the contact will reinforce the feeling of be monitoring, which influence the emotion itself [13], [17].

Straightforwardly, non-contact physiological signal monitoring can overcome these limitations, with no cables, electrodes or straps which may cause distraction [22]. The non-contact approaches rely on different sensors, such as RGB camera [23]–[27], infrared (IR)/near-infrared (NIR) camera [28], [29], frequency-modulated continuous-wave (FMCW) radar [30], continuous-wave (CW) Doppler radar [31]–[35] and Wi-Fi [36]. The video based approaches extract a 1-D signal from the face area of video, which is used to extract the hidden blood volume pulse (BVP) for generating heartbeat signal [23]–[29]. The radio frequency (RF) based approaches analyze the reflected RF signal, which is modulated by the human body, to monitor the heartbeat and breath [30]–[36]. However, different sensors have different limitations in measuring physiological signals. For example, the lighting conditions of the environment have a significant impact on accuracy on RGB camera based approaches. The performance of vital sign monitoring by video degrades in poor lighting conditions [37]. The IR/NIR camera can overcome the problem. But when the background lights stably illuminate scenes, the use of IR/NIR signals makes trouble for heartbeat estimation because of light absorption characteristics of blood [29]. On the other hand, body movement is the most challenging problem for RF based approaches [38].

In this paper, a non-contact hybrid system is built for emotion recognition. First, radar and video camera are employed for dual-modality physiological signal monitoring. The respiratory signal (by radar sensor) and the heartbeat signal (by radar sensor and video sensor) are measured and optimized for feature extraction. Then the features extracted from different sensors are fused to train a classification system that could achieve more reliable emotion recognition. The main contributions are:

1) To our best knowledge, it is the first time to combine the video camera and Doppler radar for non-contact physiological signal monitoring in emotion recognition.
2) A hybrid radar and video signal optimization approach is proposed to achieve a complementary between them:
   a) A light intensity based scheme is proposed to select proper heartbeat signal from radar and video camera for different lighting conditions.
   b) An optical flow based algorithm is proposed to detect the body movement accurately. It can eliminate the influence of body movement on physiological signals from radar.
3) The relationship between emotions and physiological signals in different dimensions (Valence and Arousal) has been explored preliminarily, which makes it possible to include more physiological signals in a non-contact way for emotion recognition in the future.

The rest of this paper is organized into five sections. Section II shows the proposed non-contact dual-modality physiological monitoring system, including the basis of radar based and video based vital sign monitoring, the light intensity based sensor selection of heartbeat signal, and the optical flow based radial body movement elimination. Next, the modeling of emotions, feature extraction, feature normalization, and hybrid sensors fusion algorithms are illustrated in Section III. The experimental results of physiological monitoring and emotion classifications are presented in Section IV. Finally, some conclusions are drawn in Section V.

II. Non-Contact Dual-Modality Physiological Signal Monitoring

As shown in Fig. 1, the proposed system comprises two parts: non-contact dual-modality physiological signal monitoring, and emotion recognition by feature level fusion. The first part will be illustrated in this section. In this part, we combined the digital intermediate frequency (IF) Doppler CW radar with the camera in an iPhone to attain the physiological signals: breathing signal and heartbeat signal. The CW radar was used to catch the chest motion which is modulated by breathing and heartbeat. The RGB camera in iPhone was used to be a video sensor to capture the heartbeat signal.

Furthermore, a hybrid physiological signal optimization approach was raised to overcome the disadvantage of separate video and radar sensors. The light intensity of the face is measured to select the heartbeat signal dynamically from the video and radar channel. At the same time, another region of interest (ROI) (chest) is extracted from the video to generate the corresponding optical flow map of each frame, which helps to eliminate the influence of body movement on physiological signals from radar sensor.

A. The Basis of Radar Sensor Vital Sign Monitoring

A CW Doppler radar with digital-IF architecture from our previous work [32], [39] was utilized to detect micro physio-

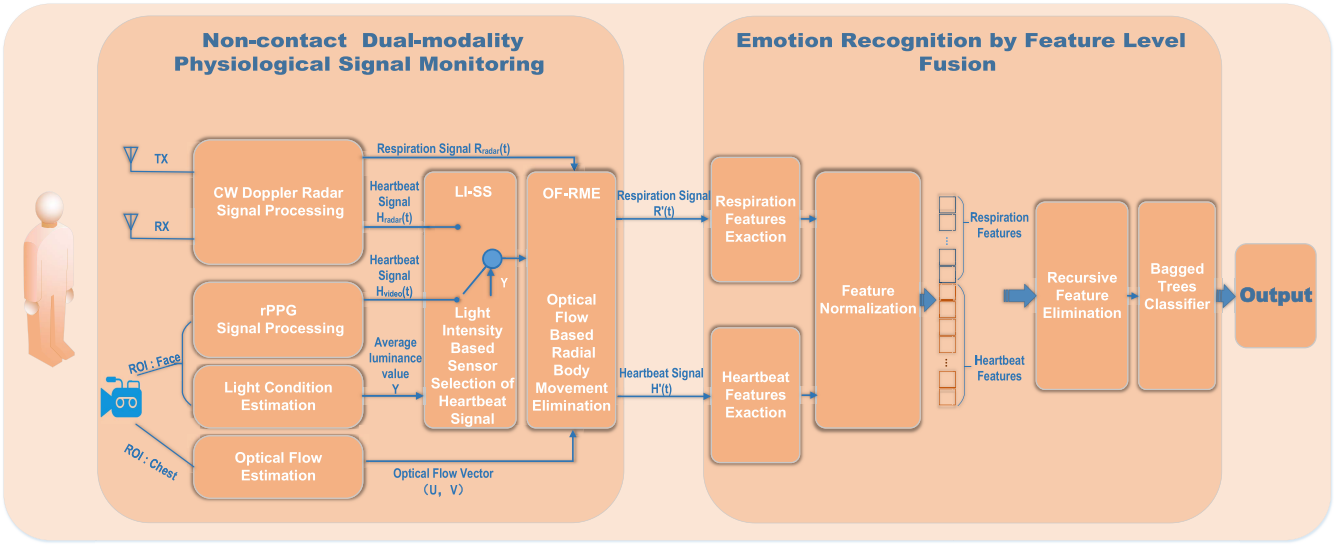


Fig. 1. The block diagram of the proposed non-contact dual-modality emotion recognition system.

TABLE I
THE COMPONENTS OF OUR CW RADAR

Block	Maunfacturer	Description
Crystal Oscillator	Tianma Telecom TX3213B	Providing the 100 MHz reference clock signal
Power Spliter	MiniCiruits AD4PS-1+	Dividing the clock signal into four channals (ADC, FPGA, LO_1 and LO_2)
Local Oscillators (LO_1, LO_2)		Producing 2.475 GHz transmtted signal (LO_1) Producing 2.4 GHz reference signal (LO_2)
LNA	MiniCiruits PSA-5451+	0.05 to 4 GHz 0.7 dBm
ADC	Anaglog Devices AD9430	Digitizing the analog IF signals with a 100 MHz sampling frequency
FPGA	Xilinx C6VLX130T	Produce the baseband quadrature I/Q signals

logical signals due to its high sensitivity. The Doppler radar is composed of a microstrip patch antenna array, a frequency synthesizer, and a digital-IF receiver. The detail is shown in Table I.

The baseband quadrature I/Q signals can be written as

$$B_I(t) = sin\left[\theta + \frac{4\pi x(t)}{\lambda} + \Delta\varphi(t)\right] \quad (1)$$

$$B_Q(t) = cos\left[\theta + \frac{4\pi x(t)}{\lambda} + \Delta\varphi(t)\right] \quad (2)$$

where $\theta = 4\pi d_0/\lambda + \theta_0$ is the constant phase shift dependent on the nominal distance to target d_0, $\Delta\varphi(t) = \varphi(t - 2d_0/c)$ is the total residual phase noise, $x(t)$ denotes the chest wall displacement which consists of heartbeat signal $H_{radar}(t)$ and respiration signal $R_{radar}(t)$.

Based on this scheme, the sensitivity of our system has been significantly improved, which enables a much lower transmit power of -13 dBm compared with 0 dBm in [40], and makes our radar a suitable solution for the long-term physiological signal monitoring. Compared with other sensors, the advantage of radar is that the radar sensor is robust against ambient light, which may vary in different environments of emotion recognition. However, due to the complexity of a real environment, radar-based detection systems suffer from other error sources [41], [42], such as random body movement, especially the radial movement which is towards or far away from the radar, will have a significant impact on the vital sign monitoring by Doppler radar [38]. Therefore, an accurate radial motion detection is desired to eliminate the interference caused by motion in the physiological signal.

B. Video Based Heartbeat Monitoring

The study on heart rate (HR) estimation based on a video camera was first raised by Verkuvesse [27] and further investigated in the last decade [5], [24]–[26], [43], [44]. It is a non-invasive optical technique that measures the micro variation of skin color and usually named as remote image photoplethysmography (rPPG or iPPG). It is assumed that the face skin is illuminated by a normal ambient light source and registered with an RGB video camera. Since the blood volume in vessels varies periodically according to the cardio cycle [45], the reflection of each skin pixel registered by the camera could be modeled as a time-varying function, which is highly correlated with heartbeat cycle.

The video based heartbeat monitoring approach employed in this paper mainly consists of two steps: ROI tracking and signal analysis. ROI is defined within the human face of video and the average value of pixels in ROI is calculated for each frame to form 1-D sinusoid-like time-varying signals in RGB channels. Since the reflection and absorption of hemoglobin

are different for red/green/blue lights, the heartbeat signal in the green channel has a much larger amplitude than others. Simple mode from [43] is adopted in our approach to extract the common heartbeat component from RGB channels with fixed linear weights in the following equation instead of other complicated source separation methods such as the independent component analysis (ICA) [44].

$$H_{video}(t) \approx 2G(t) - R(t) - B(t) \tag{3}$$

It is also shown in our previous work [24], [25] that video sensor was able to achieve accurate results in heart rate monitoring with proper ROI tracking and motion compensation when the subject is in motion, even intensive exercises. However, compared with radar, the performance of iPPG would be affected by poor ambient light [37].

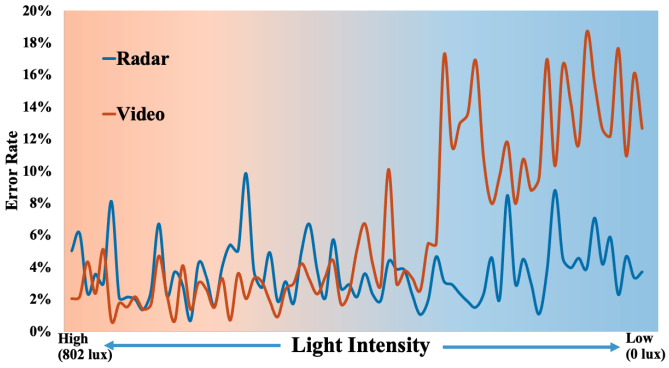


Fig. 2. Comparison of heart rate detection between radar and video sensors under different light intensities. When the light intensity decreases, the error rate of the video sensor increases significantly.

C. Hybrid Video and Radar Sensors Optimization Method

As mentioned above, radar and video sensors have their advantages and disadvantages in physiological signal extraction. The radar sensor is sensitive to body movement. Meanwhile, the video sensor is not. At the same time, a low illumination environment has impacts on video sensor instead of the radar sensor. In this section, a hybrid video and radar sensors optimization mechanism is proposed to utilize both sensors fully. The mechanism consists of two parts: light intensity based sensor selection (LI-SS) and optical flow based radial body movement elimination (OF-RME).

1) Light Intensity Based Sensor Selection (LI-SS) of Heartbeat Signal: For radar sensors, it is relatively difficult to extract the heartbeat signal from the radar signal since the radio frequency is modulated by both the movement of the chest and heart simultaneously. As we know, the spectrum of the heartbeat signal (around 1 Hz) is overlapped with the harmonic components of the respiratory spectrum (around 0.3 Hz). Moreover, the amplitude of the chest movement caused by breathing is much larger than that of the heartbeat. Therefore, the harmonic components of the respiratory spectrum are large in amplitude and cause interference in heartbeat monitoring. On the contrary, the iPPG technique monitors the heart rate by measuring the average oxygen saturation levels in the blood vessels of the face area, which will not be interfered with by the stronger breathing cycle and has better performance than radar in heartbeat monitoring. However, the performance of the video sensor degrades in poor ambient light (around 100 lux) [37].

As shown in Fig. 2, in the bright environment, the error rate of the video sensor is lower than that of the radar sensor. When the light intensity decreases, the radar sensor still has stable performance. On the contrary, the error rate of the video sensor increases significantly.

As a result, a simple but efficient method LI-SS is proposed to ensure the most accurate heartbeat signal is selected for the future processing of emotion classification as

$$H(t) = \begin{cases} H_{video}(t), & Y \geq T_{intensity} \\ H_{radar}(t), & Y < T_{intensity} \end{cases} \tag{4}$$

Both radar based and video based approaches are used to monitor heartbeat signal simultaneously. It is noted that the most relevant light condition that affects the performance of iPPG is the light intensity on the human face. Therefore, instead of measuring the ambient light intensity by lux meter [37], the average luminance value Y of ROI (face) pixels in a video is calculated in the proposed algorithm as the estimation of light intensity. If the average luminance value Y is higher than the threshold $T_{intensity}$, the heartbeat signal extracted by the video sensor is selected for further feature extraction. Otherwise, the signal from the radar is employed. In our system, $T_{intensity}$ is set as 149 empirically through intensive comparative experiments (142 pairs) in different ambient lights.

2) Optical Flow Based Radial Body Movement Elimination (OF-RME): Body movement is one of the most challenging problems in the realization of the non-contact vital signs monitoring using Doppler radar. However, not all types of body movements have serious impacts on the vital signs monitoring by radar. As shown in Fig. 3, we define the motion that towards or far away from the sensor as radial motion and the motion that perpendicular to the radial direction as tangential motion as Fig. 3. According to the principle of radar detection, the radial motion will affect the detection of radar vital signs, while the tangential motion will not [46]. In the real case, any random motion vector could be decomposed into two orthogonal axes: radial and tangential. When radial motion component exists, the corresponding segment of the heartbeat and respiration signal generated by radar sensor will be interfered significantly. The extracted features are not proper for emotion recognition. Therefore, it is desired to detect only the radial motion case of the subject and eliminate the corresponding piece of physiological signals generated by radar from the next steps: feature extraction and emotion recognition. In the proposed system, optical flow is found to be a proper choice to achieve radial motion detection since a camera is used together with radar in the dual-modality system.

Optical flow is a computer vision method commonly used for motion estimation and object tracking [47]. There are two basic hypotheses as follows:

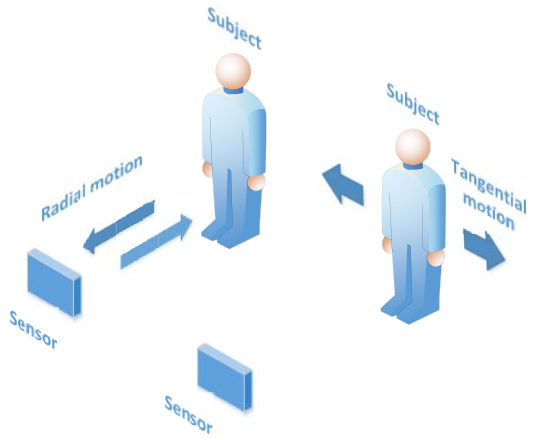


Fig. 3. Schematic diagram of radial motion and tangential motion. On the left is radial motion and on the right is tangential motion.

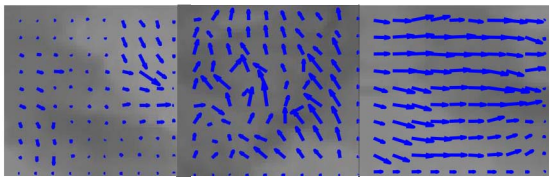


Fig. 4. The optical flow graphs of (a) non-motion (b) radial motion (c) tangential motion. Significant difference of optical flow vectors can found in different states of motion.

1) The pixel intensities of an object do not change between consecutive frames;
2) Neighbouring pixels have similar motion.

According to those assumptions, the brightness pattern can be indicated as

$$I(x(t), y(t), t) = I(x(t + dt), y(t + dt), t + dt) \quad (5)$$

By taking Taylor series approximation of right-hand side, remove common terms and divide by dt, we can obtain the optical flow constraint equation:

$$I_x u + I_y v + I_t = 0 \quad (6)$$

where I_x, I_y and I_t donate spatiotemporal image brightness derivatives. u donates the horizontal and v is the vertical optical flow. We used the Lucas-Kanade (LK) method to solve the optical flow equation for u and v [48].

As shown in Fig. 4, another ROI (chest) could be extracted from the video to generate optical flow graphs of the chest area, which is monitored by radar for respiratory and heartbeat signal. Three examples of optical flow graphs are demonstrated in Fig. 4, corresponding to different human status: static, in radial motion, or tangential one. It could be easily found that when the body keeps still, the optical flow vectors (u, v) are extremely zero-biased, as shown in Fig. 4(a). When the body is in radial motion, the corresponding optical flow vectors are diverse in various directions as shown in Fig. 4(b). On the contrary, the tangential motion of the body leads to consistent vectors in Fig. 4(c).

Based on this, an optical flow based radial body movement elimination (OF-RME) method was proposed to not only detect the body movement in the video but also classify the motion type as radial or tangential. When the radial motion of the chest area is detected, the corresponding segment of the physiological signal from the radar is eliminated. The detail could be found in the following pseudo-code.

Algorithm 1 Optical Flow Based Radial Body Movement Elimination

Require:
ROI (Chest) of the video sequence, $S_v(n)$;
Physiological signals from both two sensors, $Phy(m)$;
Ensure:
The label of the physiological signal, $Label(k)$;
The Physiological signal without radial body movement, $Phy'(k)$;
for each n **do**
$(u_i, v_j) = OpticalFlow(S_v(n), S_v(n+1)), i, j \in S_v(n)$;
$M(n) = \frac{1}{LK}\Sigma_{i=1}^{L}\Sigma_{j=1}^{K}|(u_i, v_j)|$;
$Std(n) = \sqrt{\frac{1}{LK}\Sigma_{i=1}^{L}\Sigma_{j=1}^{K}(|(u_i, v_j)| - M(n))^2}$;
end for
$M(k)$, $Std(k)$, $Phy(k)$=Synchronize($M(n)$, $Std(n)$, $Phy(m)$);
for each k **do**
if $M(k) \geq T_1$ **then**
if $Std(k) \geq T_2$ **then**
$Label(k) \leftarrow$ Radial motion;
$Phy'(k) = 0$;
else
$Label(k) \leftarrow$ Tangential motion;
$Phy'(k) = Phy(k)$;
end if
else
$Label(n) \leftarrow$ Non-motions;
$Phy'(k) = Phy(k)$;
end if
end for
return $Label(k)$, $Phy'(k)$;

As shown in the above pseudo-code, for the chest part of the input video signal, each frame is processed with the optical flow to generate the corresponding matrix of vectors (u,v). The mean $M(n)$ and standard deviation (STD) $Std(n)$ is calculated. Due to the different sampling rates of radar and camera, the physiological signal from radar $Phy(m)$ has a different data rate from $M(n)$ and $Std(n)$. Sliding window and averaging operations are performed to synchronize data from different sensors to the same data rate k, where each k means one data point for all 1-D data. Then, for each data point, $M(k)$ is used to classify the "non-motion" case. $Std(k)$ helps to distinguish the radial motion from tangential one. When $Label(k)$ is decided as "radial motion", the corresponding $Phy(k)$ is set to zero to prevent the influence of radial motion on the calculation of heart rate and respiration rate.

The effect of OF-RME could be demonstrated by the example in Fig. 5. It is noted that the influence of motion on

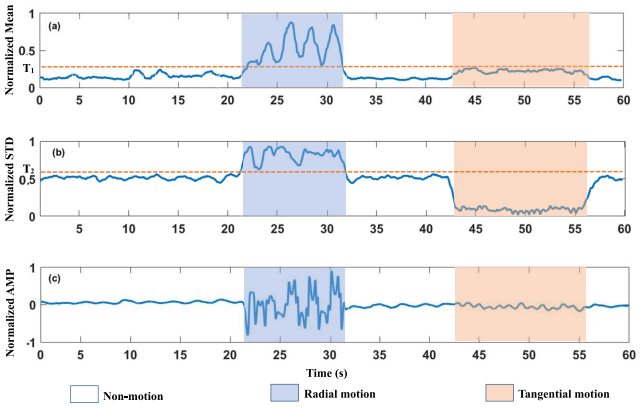


Fig. 5. An example of optical flow based radial body movement elimination. (a) Mean of the optical flow vectors: *M(k)*, (b) STD of optical flow vectors: *Std(k)*, (c) Radar signal. It is capable to distinguish different motion cases due to the significant difference of various motion cases in *M(k)* and *Std(k)*.

the radar signal is difficult to model in amplitude directly. The fluctuation in amplitude may come from tangential motion, which does no influence heart/respiration rate estimation. Therefore, straightforward thresholding on the radar signal could not eliminate the interference due to radial motion properly. On the other hand, in the video domain, it is efficient enough to distinguish different motion cases due to the significant difference of various motion cases in $M(k)$ and $Std(k)$.

III. Emotion Recognition

A supervised machine learning algorithm is employed in the second part of the proposed system to combine the optimized respiration and heartbeat signals from two sensors at the feature level, as shown in Fig. 1.

A. Modeling of Emotions

As we all know, it is difficult for an untrained person to describe his or her emotions accurately. And differentiating emotions in experiments is not a simple task. The researchers try to delimit the boundary of different emotions.

A circumplex model of affecting was proposed by Russell that all affective states arise from cognitive interpretations of core neural sensations that are the product of two independent neurophysiological systems [49]. With this model, psychologists can represent the structure of emotion with two standard dimensions: valence and arousal. Valence represents the evaluation of stimuli, with positive (pleasant) at one end and negative (misery) at the other. The other dimension "arousal" measures the active level of emotion. It could also be called as "activity". These two dimensions make up a coordinate system and 2D emotion model, as shown in Fig. 6. The Happy (positive valence and high arousal), Relaxed (positive valence and low arousal), Sad (negative valence and low arousal) and Afraid (negative valence and high arousal) four basic emotions are located in the four quadrants.

B. Feature Extraction

In previous work, we have used respiration information from the radar sensor to classify Happy, Relaxed, Sad, and Afraid

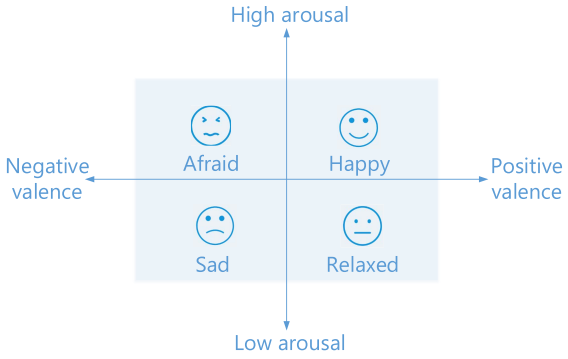


Fig. 6. Russell's circumplex model of affect [49].

emotions [50]. In this paper, we introduce a new feature set by combining the radar information and the video information. Both signals extracted from radar and video are divided into 60-seconds non-overlapping segments. Totally 63 features are extracted from each segment and these features can be divided into time-domain, frequency-domain, nonlinear-domain (including Poincaré plots [51], Approximate Entropy (AE) [52] and Detrended Fluctuation Analysis (DFA) [53]. More details are summarized in Table II.

Among the features, 24 of them are extracted from the respiratory signal, and 39 of them are from the heartbeat signal. In this paper, respiratory features were employed to extract the characteristics. Some of them (F1-F6) reflect some of the characteristics of the waveform in the time domain and the literature [54]. In the frequency domain, we used the power spectral density (F8-F11) to represent the respiratory frequency information. As shown in [54], we studied four features that represent the average energy in each of the first four 0.1 Hz bands of the power spectral density range 0.0-0.4 Hz. Some features are nonlinear features (F12-F24). These features refer to the estimation and characterization of the phase space (or state space) of the cardiovascular system.

There is a large literature on extracting emotion-dependent features from human heartbeats [17]. We extract those features from raw heartbeat signal (F25-F28) and heart rate variability (HRV) (F29-F63) which was measured by the variation in the time interval between heartbeats. It was confirmed that HRV is widely used in the analysis of heart rate [55]. The time-domain (F25 and F29-F38), frequency-domain (F39-F50), and nonlinear features (F51-F63) were extracted as listed in Table II.

C. Feature Normalization

Due to the differences between different people and different environments even the biological rhythm, the physiological signals obtained by various subjects, even the same subject on different days, are different, which we called the "person-dependence" and "day-dependence" [17], [54], [56].

To eliminate the effect of the above problems, a normalization is needed to be performed. In this paper, we calculated the mean of each feature from the neutral emotion signal for a given subject on a given day as the baseline. Then, the features were calibrated by subtracting from each feature

TABLE II
THE LIST OF EMOTIONS FEATURES

No.	Source		Features name	
F1	Respiratory signal		The means of the raw signals	
F2			The standard deviations of the raw signals	
F3			The means of the absolute values of the first difference of the raw signals	
F4			The means of the absolute values of the second differences of the normalized signals	
F5			The means of the absolute values of the second difference of the raw signals	
F6			The means of the absolute values of the second difference of the normalized signals	
F7			Respiratory rate	
F8,F9,F10, F11			Power spectral density from 0.0-0.1 Hz, 0.1-0.2 Hz, 0.2-0.3 Hz and 0.3-0.4 Hz	
F12			Approximate Entropy	
F13			Detrended Fluctuation Analysis	α_1
F14				α_2
F15-F16			Poincaré plots	SD_1(m=1,m=10)
F17-F18				SD_2(m=1,m=10)
F19,**F20**				SD_{12}(m=1,m=10)
F21-F22				S (m=1,m=10)
F23-F24				SDRR(m=1,m=10)
F25	Heartbeat signal	Raw	Heart beat Rate	
F26			Approximate Entropy	
F27			Detrended Fluctuation Analysis	α_1
F28				α_2
F29		HRV	The means of the HRV	
F30			The standard deviations of the HRV	
F31			The means of the absolute values of the first difference of the HRV	
F32			The means of the absolute values of the first differences of the normalized HRV	
F33			The means of the absolute values of the second difference of the HRV	
F34			The means of the absolute values of the second difference of the normalized HRV	
F35			Skewness	
F36			Kurtiosis	
F37			RMSSD	
F38			pNN50	
F39,F40-F41			The power calculated within the VLF, LF and HF bands	
F42,F43,F44			The frequencies having maximum magnitude of,VLF peak, LF peak and HF peak	
F45-F47			The VLF, LF and HF power expressed as percentage of the total power	
F48			The power normalized to the sum of the LF power	
F49			The power normalized to the sum of the HF power	
F50			The LF/HF power ratio	
F51			Approximate Entropy	
F52			Detrended Fluctuation Analysis	α_1
F53				α_2
F54-F55			Poincaré plots	SD_1(m=1,m=10)
F56-F57				SD_2(m=1,m=10)
F58-F59				SD_{12}(m=1,m=10)
F60-F61				S (m=1,m=10)
F62-F63				SDRR(m=1,m=10)

Selected features in **bold**.

its corresponding values calculated at the neutral state for a given person on a given day to extract person-independent and day-independent features [17].

$$Fnormal_j = F_j - \frac{1}{N}\sum_{i=1}^{N} Fneutral_j(i) \tag{7}$$

where $Fnormal_j$ is the feature after normalization, F_j is the j^{th} feature extracted from the signal during watching an arousing video and $Fneutral_j$ is the j^{th} feature extracted from the signal during the neutral state for a given person on a given day.

D. Hybrid Sensor Fusion Algorithm

In recent years, multi-sensor fusion has received significant attention for pattern recognition. The fusion of information from multiple sensors could combine the advantages of the two sensors and compensate for the disadvantages of those sensors to improved accuracies and more specific inferences that could be achieved by the use of a single sensor alone.

In this paper, the fusion of information was carried out on a feature level fusion. In feature layer fusion, before integrating the feature vectors from individual sensors into a single larger feature vector, data alignment and association functions are applied to perform the multi-sensor task. Fusion at the feature level involves the integration of feature sets corresponding to multiple modalities. Since the feature set contains more valuable information about the raw biometric data than the match score or the final decision, integration at this level is expected to provide better recognition results.

In this paper, as shown in Fig. 1, each sensor provides observational data, the radio signal from the CW radar and the video signal from the camera sensor, from which a feature vector is extracted. Then after the optical flow based random body movement detection and light intensity based feature selection, we concatenated those feature sets together into a single feature vector which represents an input to the machine learning algorithms.

In this paper, we used bagged trees learning algorithm, which is an ensemble learning using multiple decision tree learning algorithms as the classifier to differentiate among various emotional states. To provide an unbiased evaluation of the classifier, a stratified ten-fold cross-validation (CV) procedure was adopted. The dataset was divided into ten folds, with one fold used as the test set, and others used as the training set. Each set contains approximately the same percentage of samples of each target class. The process is repeated until all folds have been used as the test set.

IV. Experiment and Discussion

To evaluate the feasibility and performance of the proposed system comprehensively, experiments are conducted on the physiological signal monitoring and emotion recognition part in Fig. 1 separately. Firstly, the participants and the setup of the main experimental scenario are introduced in Section IV-A. Then, the performance of LI-SS and OF-RME is shown in Section IV-B to verify the efficiency of dual-modality non-contact physiological signal monitoring. In Section IV-C, the results of feature fusion based emotion recognition are presented, including classification accuracy, feature selection, and the discussion on the Valence/Arousal dimension. Finally, the overall performance of the system is compared with state-of-the-art algorithms in Section IV-D.

A. Participants and Experiment Setup

The experiments were carried out on eighteen healthy volunteers including seven females and eleven males with ages from 21 to 25. It is noteworthy that not everyone participated in all the experiments, details are shown in TABLE III. Before the experiment, a survey based on questionnaire was made to confirm that none of them had respiration, cardiovascular, mental diseases or related medical history. All of them kept their normal life habit and drank no alcohol one week before and during the period of experiments. Moreover, the volunteers were informed about the protocol of this study.

All experiments were conducted in a small laboratory room ($<15\ m^2$) with only the equipment shown in Fig. 7. As shown in the figure, the volunteers sat on a chair facing the digital IF Doppler radar (2.4GHz, 50μW transmitting power) and an iPhone 6s (Sony Exmor RS IMX315 sensor, 12MP, 1.22μm pixel size, f/2.2 aperture, no optical zoom, 1X digital zoom) with the videos recoded in a H.264 MOV format (1920 $\times$ 1080 pixels, 8-bit depth, 30fps). The radar was set up one meter away with the antenna facing the chest of volunteers. The iPhone was set up 1.3 meter away. The iPhone and the desktop computer to process the radar signals were synchronized to the same internet time to align the time stamps of data measured by different modalities. Besides, a laptop was placed between the radar and camera to play the experimental videos, which was only employed for the emotion recognition experiment in Section IV-C.

TABLE III
SUMMARY OF THE VOLUNTEERS

Volunteer	EX.1	EX.2	EX.3	Gender	Age	Hight(m)	Weight(kg)	BMI(kg/m^2)
V1	√	√	√	female	25	1.63	63	23.7
V2	√	√	√	male	24	1.65	58	21.3
V3	√	√	√	male	25	1.76	70	22.6
V4	√	√	√	female	23	1.62	55	21
V5	√	√	√	male	23	1.80	62	19.1
V6	√	√	√	male	22	1.75	55	18
V7		√	√	female	23	1.66	48	17.4
V8		√	√	male	24	1.80	77	23.8
V9			√	male	23	1.82	75	22.6
V10			√	male	23	1.83	80	23.8
V11			√	female	24	1.70	58	20.0
V12			√	male	24	1.82	82	24.7
V13			√	male	24	1.72	65	22.0
V14			√	male	25	1.72	57	19.3
V15			√	male	23	1.75	66	21.6
V16			√	female	22	1.62	43	16.4
V17			√	female	24	1.60	53	20.7
V18			√	female	22	1.64	52	19.1

EX.1: Evaluation of the LI-SS;
EX.2: Evaluation of the OF-RME;
EX.3: The experiments of emotion recognition.

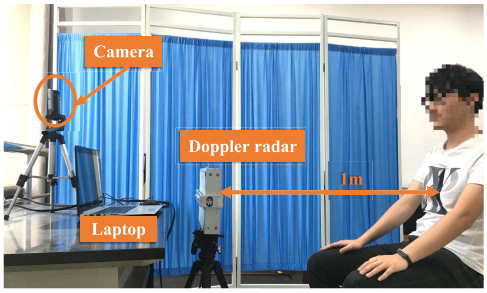


Fig. 7. Experimental Setup.

B. Dual-Modality Physiological Signal Monitoring

Experiments were conducted to demonstrate the advantage of the proposed LI-SS and OF-RME over single sensor (radar or video) in physiological signal monitoring.

1) LI-SS: The proposed LI-SS integrates the results from both sensors by measuring the environment intensity through recorded video, as illustrated in Section II-C-1. Besides the experimental scenario described in Fig. 7, the volunteer wore a heart rate band (Polar H10) to provide the reference of heart rate. During this experiments, the volunteer kept still on the chair. Dual-modality data was recorded by radar and video camera simultaneously and divided into 60-seconds segments. The ambient light conditions were controlled by moving the window curtain and toggling two pairs of overhead lights.

TABLE IV
EXPERIMENT SETUP OF LI-SS

Volunteer	Light Condition Control	Ambient Light Intensity by Luxmeter	Reference of Heart Rate	Total Data Set
v1, v2,	light off, curtain closed	0-19 lux	Polar H10	180 segments
v3, v4,	light on, curtain closed	99-152 lux		
	light off, curtain open	180-620 lux		(60s video & 60s radar
v5, v6,	light on, curtain open	560-802 lux	heart rate band	data per segment)

TABLE V
PERFORMANCE OF HEART RATE MONITORING WITH VARYING AMBIENT LIGHT INTENSITY. (%)

Volunteer	Radar Sensor		Video Sensor		LI-SS	
	$Mean_{ER}$	STD_{ER}	$Mean_{ER}$	STD_{ER}	$Mean_{ER}$	STD_{ER}
V1	3.2	1.7	4.5	3.2	**2.7**	**1.4**
V2	4.3	1.6	4.9	1.7	**3.4**	**0.9**
V3	4.8	2.0	4.7	2.7	**3.5**	**1.9**
V4	5.5	2.0	7	2.7	**5.1**	**2.0**
V5	9.0	3.7	7.7	4.4	**6.1**	**1.9**
V6	8.3	2.9	7.7	2.5	**6.4**	**1.8**
Average	5.9	2.3	6.1	2.9	**4.5**	**1.7**

Ambient light intensity by the luxmeter (lux) : 0-802.
Face light intensity by average pixel value: 15-180 (valid range 0-255).

The ambient lights intensity was measured in terms of Lux by digital Luxmeter (Smart Sensor AS803, ARCO SCIENCE AND TECHNOLOGY LTD, DongGuan, GuangDong, China). As the measured ambient light intensity was very sensitive to the position and angle of Luxmeter and did not exactly equal to the light intensity on human face that directly influencing the performance of iPPG, the measured light intensity was only employed in the ambient light condition control in the experiment. Instead, the average luminance value Y of pixels within face region was calculated to estimate the light intensity of face in the proposed scheme LI-SS, as illustrated in Section II-C-1. More details are shown in Table IV. It is noted that the 180 segments of data were evenly collected from six volunteers with gradually changed light conditions (by moving the window curtain or toggling the lights), although only the light intensity range of four categories were tabulated.

The heart rate of volunteer was monitored by three approaches: single radar, single video, and the proposed LI-SS, respectively. The results were then compared with the reference value from heart rate band. The performance comparison recorded in terms of Mean and STD of Error Rate (ER, in %, see Appendix B) are shown in Table V. It is verified that when the ambient light intensity varied, radar performs slightly better than video sensor on average (slightly smaller error rate and less variation), which is consistent with the demonstrative result on single volunteer in Fig. 2. Among the three approaches, the proposed LI-SS achieves the best performance with the smallest average error rate (Mean) and less fluctuation (STD) with variation of the light intensity, since the performance degradation of video sensor in poor ambient light could be compensated by the other modality radar. It could be concluded that LI-SS can effectively eliminate the influence of weak ambient lights on heartbeat signal monitoring and increase the robustness of the system in environments with complicated light conditions.

TABLE VI
EXPERIMENT SETUP OF OF-RME

Volunteer	Body movement (5-20s)	Reference of Respiration	Total Data Set
v1, v2, v3, v4, v5, v6, v7, v8	radial motion, tangential motion, other random motion	HKH-11C respiration belt	44 (60s a segment)

Ambient light intensity by the luxmeter (lux) : 99-152.

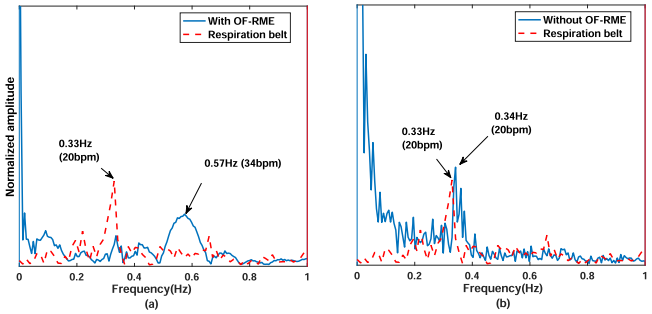


Fig. 8. Spectra of one segment of radar signal (five minutes) (a) without OF-RME (b) with OF-RME compared to the reference value from respiration belt. After removing the slot of signal containing radial motion with OF-RME, the result is adjusted to be same as reference.

2) OF-RME: The proposed OF-RME eliminated the body movement impact from the radar by using the optical flow through recorded video, as described in Section II-C-2. Besides the common experimental setting described in Fig. 7, the volunteer wore a respiration belt (HKH-11C, Electronic Technology Research Institute, Hefei, China, Sampling frequency: 50Hz, Sampling accuracy: 8 bits, Communication baud rate: 9600) as the reference of respiration rate. Dual-modality data was recorded by radar and video camera simultaneously and divided into 60-seconds segments as the last experiment. In this experiment, during each 60s segment, the volunteer made one of the following body movements on the chair: radial motion (tilting the upper body forward and backward), tangential motion (tilting the upper body left and right), random motion (adjusting the posture according to personal habit). More details of the specific setting are summarized in Table VI.

The respiration rate of volunteer was monitored by two approaches: without OF-RME and with OF-RME. One segment of data from volunteer 1 (v1) containing radial motion is used as example, as shown in Fig. 8. The spectra of that segment for both approaches are compared with the reference value from respiration belt. It is noted that the peak frequency could be transformed to the respiration rate in units of beats per minute (bpm) by multiplication of 60. Both peak frequency and respiration rate are labeled in the figure. As demonstrated in Fig. 8(a), the radial motion has significant influence on the peak frequency by single modality radar. However, when OF-RME is employed to remove the slot of signal containing radial motion with optical flow graph from video, the result is adjusted to be same as reference as shown in Fig. 8(b).

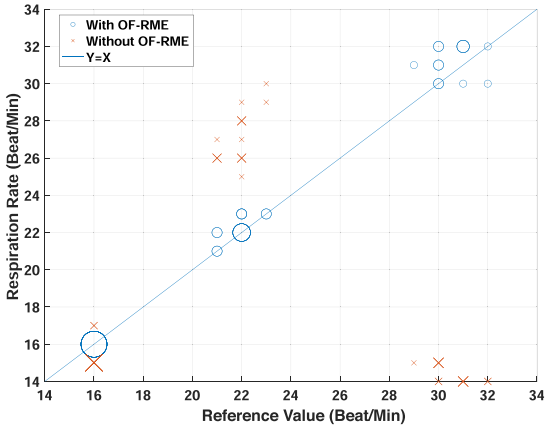


Fig. 9. Correlation between the respiration rate measured by from our system and respiration belt. And the size of the marker represents the number of segments at the same respiration rate. The results of OF-RME are highly correlated with the reference value of respiration belt.

When the respiration rate is influenced by radial motion, the gap between the measured data (without OF-RME) and reference is quite significant as shown in Fig. 8. The Root-Mean-Square-Error (RMSE) for all data segments is 9.7 for the approach without OF-RME and only 0.8 for the approach with OF-RME. It is noted that the average error rate or RMSE is no longer an intuitive metric in this case, since it highly depends on the percentage of segments containing radial motion. Therefore, the correlation between the extracted respiration rate and the reference is plotted in Fig. 9 for the two approaches. It is observed that the radial motion has serious interference on the respiration rate only by radar (without OF-RME). On the other hand, the results of OF-RME are highly correlated with the reference value of respiration belt.

C. Emotion Recognition

Besides the common experimental scenario described in Section IV-A, four emotions, including Happy, Relaxed, Sad, and Afraid, were triggered via videos clips, which were played in the laptop between the radar and camera as shown in Fig. 7. All of the video clips were selected based on similar studies [57]–[59] with two video clips (traditional Chinese crosstalk and sketch) that matched the participants' cultural background. The details are shown in the Table VII. The duration of each video clip was 30 minutes. To record the emotional state at that moment, there was a one break after the playback of each video clip. The volunteers were asked to fill in the emotional self-evaluation form (see Appendix A) designed with reference to the Self-Assessment Manikin (SAM) [60]. Before each emotional video, a short neutral clip (around 15s) was randomly selected from the clips provided in [57] to remove the bias from the emotional state [58].

It is well known that accurate emotion assessment is a challenging task. A human may not describe their own emotions accurately, and the error of emotional self-evaluation is hard to control [58]. And we found some emotions such as "Sad" is challenging to stimulate in a short period. To minimize error introduced by self-evaluation, the video recorded by iPhone

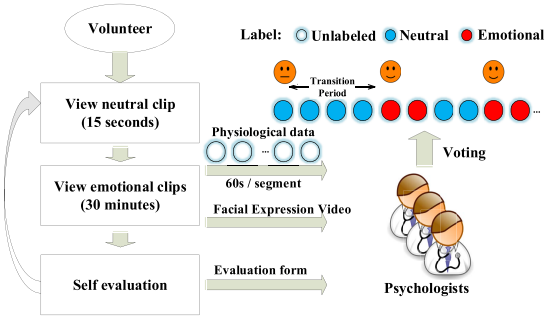


Fig. 10. Flowchart for emotion recognition experiment.

TABLE VII
THE VIDEO CLIPS FOR EACH TARGET EMOTION

Target Emotion	Video Name
Happy	When Harry Met Sally Mr. Bean's Holiday Traditional Chinese Crosstalk Traditional Chinese Sketch
Relaxed	Mozart - Classical Music for Relaxation Candlelight with Chopin
Sad	The Champ Bambi
Afraid	The Shining Slience of the Lamb The Ring, Japanese version 28 Days Later

6s was used not only to extract the heartbeat signal for HR monitoring, to detect the light intensity and optical flow vector for the hybrid optimization method, but also as an essential indicator for professional psychologists to confirm the mood of volunteers. The physiological data by dual-modality sensors and the corresponding facial expression video were divided into 60-seconds segments. It is noted that 60-seconds segment is believed to be a good trade-off between avoiding multiple emotion mixing and maintaining enough time to observe the effects of emotions [58]. Three psychologists were invited to observe the facial expression in each segment of data and compare their observation with the feedback form by volunteers to refine the emotion annotation separately. Then the majority voting method was used to remove segments where the observed emotion state does not match the emotions on the feedback form. The pair of physiological data and emotion label would be used in the training process later as the true label.

The flowchart for the emotional classification experiment is shown in Fig. 10. Two types of labels are made: "Neutral" and "Emotional" (four emotions). A total of 2010 segments of data with emotional labels were collected, which is summarized in Table VIII. As mentioned in Section II, the proposed OF-RME removes the period of data with significant radial motion. When the length of radial motion in one segment (60s) was over 40s, the whole segment was discarded for refinement.

As illustrated in Section II, the features shown in Table II were extracted from the data segments. The features and the corresponding labels were passed into the classifier to build

TABLE VIII
THE EMOTIONS DATA OF THE EXPERIMENT

Volunteer	Happy	Relaxed	Sad	Afraid	Total
V1	6	9	9	7	31
V2	13	7	10	14	44
V3	18	15	28	18	78
V4	33	26	46	13	118
V5	16	15	39	17	87
V6	12	9	0	3	24
V7	0	3	0	28	31
V8	5	15	0	2	22
V9	48	37	65	40	190
V10	41	21	73	31	166
V11	0	35	55	28	118
V12	59	0	20	26	105
V13	63	138	29	33	263
V14	63	0	38	49	150
V15	22	77	22	36	157
V16	86	20	56	0	162
V17	22	144	20	10	196
V18	3	26	35	4	68
Total	509	597	544	359	2010

Ambient light intensity by the luxmeter (lux) : 99-152.

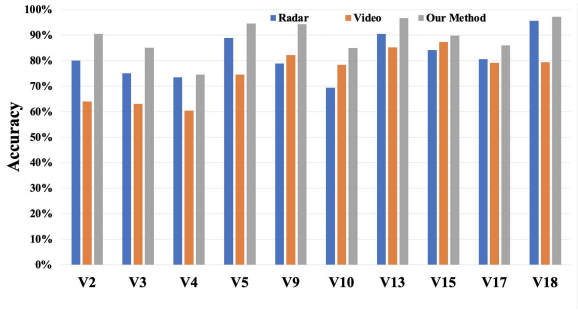


Fig. 11. The classification results comparison of the V2, V3, V4, V5, V9, V10, V13, V15, V17, V18 by our method and single modal data. The accuracy of classification are improved for all volunteers in the proposed dual-modality system.

the mapping model between physiological signals and human emotions. The MATLAB software and the scikit-learn with Python [61] were employed.

1) Accuracy of Emotion Recognition: To evaluate the performance of the proposed system in complex environments, the results are compared with the single sensor system (only Radar or Video). Two types of emotion classification approaches were trained: person-dependent and person-independent. Training and testing were done by using the stratified 10-fold CV. At the same time, p-values are calculated to measure the probability that the observed difference could have occurred just by random chance. The lower the p-value, the greater the statistical significance of the observed difference. Typically, when p-value is smaller than 0.05, the observed difference is believed to be statistically significant.

Firstly, we evaluated our emotion recognition system in a person-dependent way. In the person-dependent way, the classifier was trained and tested on the data of the same volunteer. It is worth noting that only ten subject's data shown in Fig. 11 are used in the person-dependent approach due to

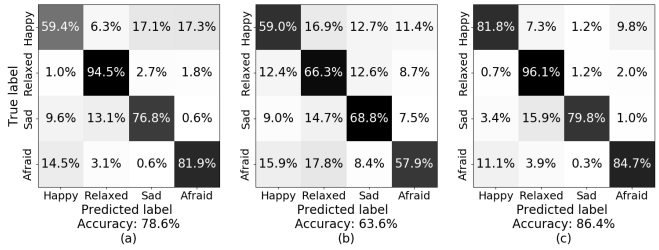


Fig. 12. The emotion classification results of (a) only radar sensor and (b) only video sensor (c) our system. The feature level fusion of two sensors results in better overall performance in emotion classification.

the lack of some emotional types or small datasets for other volunteers. The classification results are shown in Fig. 11. The average accuracy and variance of the radar sensor are 81.6 % and 0.007. And the average accuracy and variance of the video sensor are 75.3 % and 0.009, respectively. Meanwhile, the accuracy and variance of the proposed method are 89.3 % and 0.005. The p-values of the observed difference between the proposed method and single radar is 0.04, while it is 0.0017 between the proposed one and video radar. It is demonstrated that the accuracy of classification are improved for all volunteers in the proposed dual-modality system.

For a person-independent approach, all data from Table III were employed for training all features. The results are shown in Fig. 12. The average accuracy of the radar sensor and video sensor is 78.6 % and 63.6 %, respectively. The performance of the video sensor is relatively low since only heartbeat signals are extracted by the video sensor currently, and the feature set is much less than that of the radar sensor and dual-modality system. The proposed system has the best performance with an accuracy of 86.4 %. It was also found that for both radar and video sensors, there are significant errors for particular emotion pairs, such as Afraid vs. Happy for radar sensor, and Afraid vs. Relaxed for video sensor. The errors are partially reduced in the proposed dual-modality system. It demonstrates that the feature level fusion of two sensors results in better overall performance in emotion classification.

2) Feature Selection and Parameter Adjusting: As mentioned in Table II, 63 features could be extracted from dual-modality physiological signals. Directly using all of those features with limited training data may lead to overfitting. Besides, the contribution of features to the final classification result varies significantly. Thus, feature selection is desired to not only reduce the amount of training data but also improve the classification accuracy. In this work, a recursive feature elimination (RFE) method based on Gini gain, which was calculated using the Gini impurity to help determine which feature gives us the most information, was used with a base model to perform multiple rounds of training. After each round of training, the features of the lowest weight coefficients are eliminated, and the next round of training is performed based on the new feature set until the best results are obtained. The details of RFE could be found in [62].

As shown in Fig. 13, the accuracy of person-independent emotion classification is 86.4 %. It is increased to 88.1 % by

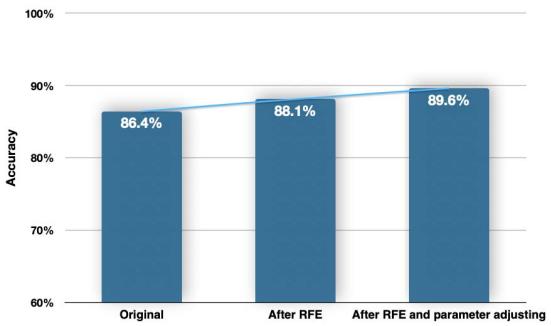


Fig. 13. The result of emotion classification after feature selection and machine learning parameter adjustment. Both the RFE feature selection and the parameter adjustment can improve the classification performance.

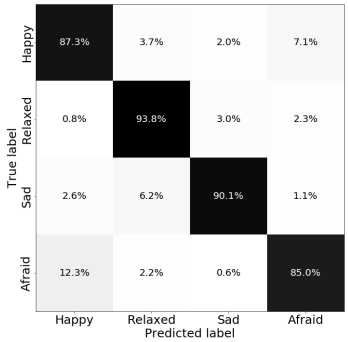


Fig. 14. Confusion matrix of stratified ten fold cross-validation after parameter optimization.

the RFE algorithm (p-value = 0.048). The most significant features are retained including: F1, F2, F7, F13, F15-F18, F20-F25, F27, F29, F31, F39, F42, F45-F51. Both the number of features and the computational complexity are reduced with better classification performance.

Further, by appropriately tuning the parameters of the classifier, the classifier can be better adapted to the entire sample space. For bagged trees classifier, the number of component classifiers and the max quantity of the split node are two critical parameters. Instead of utilizing the default value 30 and 100, these two parameters are further tuned by searching with a Bayesian Optimization Algorithm [63]. As shown in Fig. 13, the final accuracy could be further improved to 89.6 % (p-value = 0.0464) when the number of the component classifiers is adjusted to 298, and the max number of split nodes is set as 415.

3) Emotion Recognition in Valence and Arousal Dimension: The final confusion matrix of classification results is shown in Fig. 14. As demonstrated in the diagonal line, the accuracy of recognition is improved to 87.3 %, 93.8 %, 90.1 %, and 85.0 % for Happy, Relaxed, Sad, and Afraid, respectively. Furthermore, it is interesting to notice that besides the classification accuracy, the number in the anti-diagonal line is usually the second large one, as shown in both Fig. 12 (c) and Fig. 14. For example, in the third row of Fig. 14, 90.1 % of segments with label "Sad" is predicted correctly. Considering

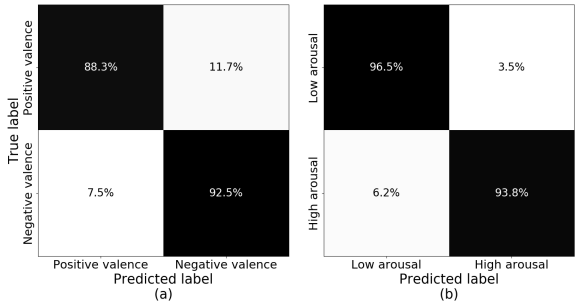


Fig. 15. Confusion matrix of grouped emotions in two dimensions: (a) positive/negative valence (b) high/low arousal.

the remaining misclassified segments, 6.2 % of "Sad" segments are classified as "Relaxed", which is much larger than the error rate of "Sad → Happy" (2.6 %) and "Sad → Afraid" (1.1 %). Similarly, in the fourth row, the error rate of "Afraid → Happy" is much larger than that of "Afraid → Relaxed" and "Afraid → Sad". These emotion pairs (Sad / Relaxed, Afraid / Happy) with a much higher error rate are found to be in the same level of arousal, as shown in Fig. 6.

In Fig. 12(c), the observation is consistent for all emotions, i.e., the error rate in the anti-diagonal line is much larger than the remaining. In Fig. 14, one special case is found between "Happy" and "Relaxed", where the error rate of "Happy → Relaxed" and "Relaxed → Happy" are larger than expectation. The reason may be the overfitting problem introduced by the feature selection and parameter tuning procedure.

To verify the observation, four emotions are grouped with two different dimensions in Fig. 6: arousal and valence. "Happy" and "Afraid" are grouped as high arousal class, while low arousal class contains "Relaxed" and "Sad". On the other hand, "Happy" and "Relaxed" are labeled as positive valence, while "Sad" and "Afraid" are labeled as negative valence. The confusion matrix of different groups is shown in Fig. 15. It is easy to found that although the proposed system has good performance in both valence and arousal classification, the arousal classes are detected with higher accuracy comparing to valence labels. It is due to that the activity of emotions (arousal) is more correlated with the physiological signals: heartbeat and breath. The result is also consistent with [58], where Soleymani classified emotions with EEG signals.

4) Cross Validation at Subject Level: The above results illustrate that the extracted features are appropriate to distinguish different emotions. However, the bias from each subject is a grave concern in safety and efficacy aspects of any application. To demonstrate the robustness of our method and avoid biases effects arise from different subjects, 12 subjects (2/3 of the dataset) are randomly selected as the training set, and the remaining 6 subjects (1/3 of the dataset) are employed as the validation set. The training set is used in the classifier training, feature selection and parameter adjusting. The validation set is used to verify the prediction accuracy. The above experiment is repeated 10 times for the cross validation at subject level. The average accuracy /variance is 75.0 % /0.003,

TABLE IX
THE REFERENCE TO STATE-OF-THE-ART LITERATURE BASED ON PHYSIOLOGICAL SIGNALS

Authors	Signals	Emotion	Overall accuracy	
			Dependent	Independent
Hsu et al. [64]	ECG	Joy,Tension, Sandness,Peacefulness	-	80.51 % (10-fold) 68.1 % (LOSO)
Zhang et al. [65]	EEG	Joy, Fear, Sadness, Relaxtion	62.59 %	58.75 % (10-fold)
Kim et al. [2]	EMG, ECG, SC, Respiration	Joy, Anger, Sad, Pleasure	95.00 %	70 % (LOO)
Zhao et al. [17]	Heartbeat Respiration	Sadness, Anger, Pleasure, Joy	87 %	72.3 % (LOSO)
Our method	Heartbeat Respiration	Happy, Relaxed, Sad, Afraid	89.30%	89.6 % (10-fold) 71 % (CVSL)

SC: skin conductance;
EEG: electroencephalogram, ECG: electrocardiogram, EMG: electromyogram;
Dependent: Person-dependent, Independent: Person-independent;
10-fold: 10-fold cross validation, LOSO: leave one subject out cross validation, LOO: leave one out cross validation;
CVSL : Cross Validation at Subjects Level.

TABLE X
EMOTION SELF-EVALUATION FORM

	Level-2	Level-1	Level0	Level1	Level2
Arousal					
Valence					
Emotion	Happy	Relaxed	Neutral	Sad	Afraid

TABLE XI
EVALUATION METRICS

Metric	Definition
Error Rate	$ErrorRate = \frac{\lvert Rate(i) - Rate_{gt}(i)\rvert}{Rate_{gt}(i)}$
Mean of Error Rate	$Mean_{ER} = \frac{1}{N}\Sigma_{i=1}^{N} \frac{\lvert Rate(i) - Rate_{gt}(i)\rvert}{Rate_{gt}(i)}$
STD of Error Rate	$STD_{ER} = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}(\frac{\lvert Rate(i) - Rate_{gt}(i)\rvert}{Rate_{gt}(i)} - Mean_{ER})^2}$
RMSE	$RMSE = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}(Rate(i) - Rate_{gt}(i))^2}$
Precision	$PPV = \frac{TP}{TP+FP}$
Accuracy	$Acc = \frac{TP+TN}{TP+TN+FP+FN}$

$Rate$ represents respiratory rate (R_r) and heart rate (H_r) respectively, depending on usage.
$Rate_{gt}$ represents the results from the reference.

90.8 % /0.0002 and 71.0 % / 0.0007 in terms of High-Low valence, High-Low arousal, and four emotions recognition, respectively. It shows that the validity of our method when using new data from new subjects. And it is noteworthy that the result is lower than 10-fold CV at sample level (89.6 %). In emotion recognition, individual difference in physiological signaling is still a topic worth investigating.

D. Comparison With Existing Approaches

The performance comparison with existing approaches is shown in Table IX. According to our best knowledge, there are no common database using RF and video sensors in the field of emotion recognition, the use of different signals, stimuli, and emotion classes makes it difficult to compare our system with work directly. First, the research on analyzing the facial expression and other explicit features that could be faked is skipped. Only the emotion recognition methods employing physiological signals are demonstrated in the table. Among these work, Hsu *et al.* [64], Zhang *et al.* [65] and Kim and André [2] analyzed physiological signals such as EEG, EMG, ECG for emotion recognition. However, all of them utilized complex contact sensors, which are uncomfortable to wear and may result in changes of participant's physiology induced by psychological factors and seriously limit the application [66]. Among existing studies, Zhao's work [17] is the most similar one which also employed non-contact signals (from radar sensors) for the classification of four emotions: Sadness, Anger, Pleasure, and Joy. As shown in Table IX, for the person-dependent classification and person-independent classification, the proposed system achieves a better or a similar recognition rate compared with both contact [2], [64], [65] and non-contact [17] methods. Besides, none of the existing works investigate the impact of the complex environment such as body movement and ambient light variation as the proposed one. With the LI-SS and OF-RME and the feature level fusion from two sensors, the proposed non-contact dual-modality system is expected to work well in more complicated cases than the laboratory environment in most prior work. It is noted that due to the different hardware structure, number of sensors, subjects, and target scenario, the comparison in Table IX is only a reference to literature rather than the comparison in statistical significance.

V. CONCLUSION AND FUTURE WORK

This paper presents a technology that is capable of recognizing human emotions through the physiological signals from a hybrid radar and video system. The proposed hybrid video and radar optimization methods (LI-SS and OF-RME) help to utilize the characteristic of each sensor and achieve better non-contact physiological signals monitoring in a complex environment with body motion and ambient light variation. The machine learning system with the feature level fusion of two non-contact sensors achieves accurate emotion recognition. Since the hybrid physiological signal acquisition system with radar and video is non-contact and robust, it may extend the application of emotion recognition to more fields such as education and psychotherapy. It may also be applied to other biological science research and applications, such as health monitoring.

In practical implementation, the traditional manual-selected features cannot cope well with the variability of individual physiological signals. It affects the performance of emotion classification results. More advanced feature selection algorithms are desired [67]. In this paper, it does not take into account the case that the two sensors fail at the same time. Therefore, the future work will mainly focus on the following points:

1) expanding the database and share it online for the worldwide researchers in this area to train a more robust personal-independent system;
2) studying the cross-correlation between dual-modal data and investigating more advanced feature extraction,

sensor fusion and classification algorithms, such as CNN and LSTM [68], [69] to further improve the generalizability of emotion recognition.;

3) investigating the cross-correlation between physiological signals and emotions in the arousal dimension.

APPENDIX A

See Table X.

APPENDIX B

See Table XI.

REFERENCES

[1] S. Katsigiannis and N. Ramzan, "DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 98–107, Jan. 2018.

[2] J. Kim and E. André, "Emotion recognition based on physiological changes in music listening," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 12, pp. 2067–2083, Dec. 2008.

[3] R. Cowie *et al.*, "Emotion recognition in human-computer interaction," *IEEE Signal Process. Mag.*, vol. 18, no. 1, pp. 32–80, Jan. 2001.

[4] R. W. Picard, "Affective computing: From laughter to IEEE," *IEEE Trans. Affect. Comput.*, vol. 1, no. 1, pp. 11–17, Jan. 2010.

[5] B. de Gelder, "Why bodies? Twelve reasons for including bodily expressions in affective neuroscience," *Philos. Trans. Roy. Soc. B, Biol. Sci.*, vol. 364, no. 1535, pp. 3475–3484, Dec. 2009.

[6] P. C. Schmid, M. Schmid Mast, D. Bombari, F. W. Mast, and J. S. Lobmaier, "How mood states affect information processing during facial emotion recognition: An eye tracking study," *Swiss J. Psychol.*, vol. 70, no. 4, pp. 223–231, Dec. 2011.

[7] D. J. Martino, S. A. Strejilevich, G. Fassi, E. Marengo, and A. Igoa, "Theory of mind and facial emotion recognition in euthymic bipolar I and bipolar II disorders," *Psychiatry Res.*, vol. 189, no. 3, pp. 379–384, Oct. 2011.

[8] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin, "Static and dynamic 3D facial expression recognition: A comprehensive survey," *Image Vis. Comput.*, vol. 30, no. 10, pp. 683–697, Oct. 2012.

[9] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognit.*, vol. 44, no. 3, pp. 572–587, 2011.

[10] S. Kwon, "CLSTM: Deep feature-based speech emotion recognition using the hierarchical ConvLSTM network," *Mathematics*, vol. 8, no. 12, p. 2133, Nov. 2020.

[11] S. G. Koolagudi and K. S. Rao, "Emotion recognition from speech: A review," *Int. J. Speech Technol.*, vol. 15, no. 2, pp. 99–117, Jun. 2012.

[12] S. Kwon, "Att-Net: Enhanced emotion recognition system using lightweight self-attention module," *Appl. Soft Comput.*, vol. 102, Apr. 2021, Art. no. 107101.

[13] M. Raja and S. Sigg, "Applicability of RF-based methods for emotion recognition: A survey," in *Proc. IEEE Int. Conf. Pervas. Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2016, pp. 1–6.

[14] N. Dael, M. Mortillaro, and K. R. Scherer, "Emotion expression in body action and posture," *Emotion*, vol. 12, no. 5, p. 1085, 2012.

[15] H. Aviezer, Y. Trope, and A. Todorov, "Body cues, not facial expressions, discriminate between intense positive and negative emotions," *Science*, vol. 338, no. 6111, pp. 1225–1229, 2012.

[16] M. Karg, K. Kühnlenz, and M. Buss, "Recognition of affect based on gait patterns," *IEEE Trans. Syst., Man, Cybern., B (Cybern.)*, vol. 40, no. 4, pp. 1050–1061, Aug. 2010.

[17] M. Zhao, F. Adib, and D. Katabi, "Emotion recognition using wireless signals," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, Oct. 2016, pp. 95–108.

[18] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Trans. Affect. Comput.*, vol. 1, no. 1, pp. 18–37, Jan. 2010.

[19] P. Ekman, W. V. Friesen, and P. Ellsworth, *Emotion in the Human Face: Guidelines for Research and an Integration of Findings*. Amsterdam, The Netherlands: Elsevier, 2013.

[20] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Emotion assessment from physiological signals for adaptation of game difficulty," *IEEE Trans. Syst., Man, Cybern., A, Syst. Humans*, vol. 41, no. 6, pp. 1052–1063, Nov. 2011.

[21] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 327–339, Jul. 2014.

[22] A. D. Droitcoour, "Non-contact measurement of heart and respiration rates with a single-chip microwave Doppler radar," Ph.D. dissertation, Stanford Univ., Stanford, CA, USA, Jun. 2006.

[23] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1479–1491, Jul. 2017.

[24] Q. Zhu, C.-W. Wong, C.-H. Fu, and M. Wu, "Fitness heart rate measurement using face videos," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2000–2004.

[25] K. Xie, C.-H. Fu, H. Liang, H. Hong, and X. Zhu, "Non-contact heart rate monitoring for intensive exercise based on singular spectrum analysis," in *Proc. IEEE Conf. Multimedia Inf. Process. Retr. (MIPR)*, Mar. 2019, pp. 228–233.

[26] Y. Cui, C.-H. Fu, H. Hong, Y. Zhang, and F. Shu, "Non-contact time varying heart rate monitoring in exercise by video camera," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2015, pp. 1–5.

[27] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Exp.*, vol. 16, no. 26, pp. 21434–21445, 2008.

[28] M. V. Gastel, S. Stuijk, and G. D. Haan, "Motion robust remote-PPG in infrared," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 5, pp. 1425–1433, May 2015.

[29] K. Kurihara, D. Sugimura, and T. Hamamoto, "Adaptive fusion of RGB/NIR signals based on face/background cross-spectral analysis for heart rate estimation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 4534–4538.

[30] M. Mercuri, I. R. Lorato, Y.-H. Liu, F. Wieringa, C. Van Hoof, and T. Torfs, "Vital-sign monitoring and spatial tracking of multiple people using a contactless radar-based sensor," *Nature Electron.*, vol. 2, no. 6, pp. 252–262, Jun. 2019.

[31] C. Gu, C. Li, J. Lin, J. Long, J. Huangfu, and L. Ran, "Instrument-based noncontact Doppler radar vital sign detection system using heterodyne digital quadrature demodulation architecture," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 6, pp. 1580–1588, Jun. 2010.

[32] H. Zhao, H. Hong, L. Sun, Y. Li, C. Li, and X. Zhu, "Noncontact physiological dynamics detection using low-power digital-IF Doppler radar," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1780–1788, Jul. 2017.

[33] H. Hong *et al.*, "Microwave sensing and sleep: Noncontact sleep-monitoring technology with microwave biomedical radar," *IEEE Microw. Mag.*, vol. 20, no. 8, pp. 18–29, Aug. 2019.

[34] C. Li, J. Ling, J. Li, and J. Lin, "Accurate Doppler radar noncontact vital sign detection using the RELAX algorithm," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 3, pp. 687–695, Mar. 2010.

[35] C. Li, J. Cummings, J. Lam, E. Graves, and W. Wu, "Radar remote monitoring of vital signs," *IEEE Microw. Mag.*, vol. 10, no. 1, pp. 47–56, Feb. 2009.

[36] C. Chen, Y. Han, Y. Chen, and K. J. R. Liu, "Multi-person breathing rate estimation using time-reversal on WiFi platforms," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2016, pp. 1059–1063.

[37] L. Ren, L. Kong, F. Foroughian, H. Wang, P. Theilmann, and A. E. Fathy, "Comparison study of noncontact vital signs detection using a Doppler stepped-frequency continuous-wave radar and camera-based imaging photoplethysmography," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 9, pp. 3519–3529, Sep. 2017.

[38] C. Gu, G. Wang, Y. Li, T. Inoue, and C. Li, "A hybrid radar-camera sensing system with phase compensation for random body movement cancellation in Doppler vital sign detection," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 12, pp. 4678–4688, Dec. 2013.

[39] L. Sun *et al.*, "Noncontact vital sign detection based on stepwise atomic norm minimization," *IEEE Signal Process. Lett.*, vol. 22, no. 12, pp. 2479–2483, Dec. 2015.

[40] F.-K. Wang *et al.*, "A novel vital-sign sensor based on a self-injection-locked oscillator," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 12, pp. 4112–4120, Dec. 2010.

[41] J. Kranjec, S. Beguš, G. Geršak, and J. Drnovšek, "Non-contact heart rate and heart rate variability measurements: A review," *Biomed. Signal Process. Control*, vol. 13, pp. 102–112, Sep. 2014.

[42] N. Hjortskov, D. Rissén, A. K. Blangsted, N. Fallentin, U. Lundberg, and K. Søgaard, "The effect of mental stress on heart rate variability and blood pressure during computer work," *Eur. J. Appl. Physiol.*, vol. 92, no. 1, pp. 84–89, Jun. 2004.

[43] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2878–2886, Oct. 2013.
[44] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, Jan. 2011.
[45] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas.*, vol. 28, no. 3, pp. R1–R39, Mar. 2007.
[46] J. Tu, T. Hwang, and J. Lin, "Respiration rate measurement under 1-D body motion using single continuous-wave Doppler radar vital sign detection system," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 6, pp. 1937–1946, Jun. 2016.
[47] M. Frigola, J. Amat, and J. Pagès, "Vision based respiratory monitoring system," in *Proc. 10th Medit. Conf. Control Automat. (MED)*, Lisbon, Portugal, 2002, pp. 9–13.
[48] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. Int. Joint Conf. Artif. Intell.*, 1981, pp. 674–679.
[49] J. A. Russell, "A circumplex model of affect," *J. Personality Social Psychol.*, vol. 39, no. 6, p. 1161, Dec. 1980.
[50] Q. Gao *et al.*, "Non-contact emotion recognition via CW Doppler radar," in *Proc. Asia–Pacific Microw. Conf. (APMC)*, Nov. 2018, pp. 1468–1470.
[51] A. Goshvarpour, A. Abbasi, and A. Goshvarpour, "Indices from lagged Poincare plots of heart rate variability: An efficient nonlinear tool for emotion discrimination," *Australas. Phys. Eng. Sci. Med.*, vol. 40, no. 2, pp. 277–287, Jun. 2017.
[52] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Amer. J. Physiol.-Heart Circulatory Physiol.*, vol. 278, no. 6, pp. H2039–H2049, Jun. 2000.
[53] T. Penzel, J. W. Kantelhardt, L. Grote, J. H. Peter, and A. Bunde, "Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 10, pp. 1143–1151, Oct. 2003.
[54] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, Oct. 2003.
[55] M. Malik *et al.*, "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use," *Eur. Heart J.*, vol. 17, no. 3, pp. 354–381, 1996.
[56] J. Fleureau, P. Guillotel, and Q. Huynh-Thu, "Physiological-based affect event detector for entertainment video applications," *IEEE Trans. Affect. Comput.*, vol. 3, no. 3, pp. 379–385, Jul. 2012.
[57] J. J. Gross and R. W. Levenson, "Emotion elicitation using films," *Cogn. Emotion*, vol. 9, no. 1, pp. 87–108, 1995.
[58] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE Trans. Affect. Comput.*, vol. 3, no. 2, pp. 211–223, Apr. 2012.
[59] A. Hanjalic and L.-Q. Xu, "Affective video content representation and modeling," *IEEE Trans. Multimedia*, vol. 7, no. 1, pp. 143–154, Feb. 2005.
[60] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *J. Behav. Therapy Exp. Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.
[61] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.
[62] P. M. Granitto, C. Furlanello, F. Biasioli, and F. Gasperi, "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products," *Chemometric Intell. Lab. Syst.*, vol. 83, no. 2, pp. 83–90, Sep. 2006.
[63] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 2951–2959.
[64] Y.-L. Hsu, J.-S. Wang, W.-C. Chiang, and C.-H. Hung, "Automatic ECG-based emotion recognition in music listening," *IEEE Trans. Affect. Comput.*, vol. 11, no. 1, pp. 85–99, Jan. 2020.
[65] J. Zhang, M. Chen, S. Zhao, S. Hu, Z. Shi, and Y. Cao, "ReliefF-based EEG sensor selection methods for emotion recognition," *Sensors*, vol. 16, no. 10, p. 1558, 2016.
[66] N. G. Frelih, A. Podlesek, J. Babič, and G. Geršak, "Evaluation of psychological effects on human postural stability," *Measurement*, vol. 98, pp. 186–191, Feb. 2017.
[67] S. Kwon, "Optimal feature selection based speech emotion recognition using two-stream deep convolutional neural network," *Int. J. Intell. Syst.*, vol. 36, no. 9, pp. 5116–5135, 2021.
[68] J. Zhao, X. Mao, and L. Chen, "Speech emotion recognition using deep 1D & 2D CNN LSTM networks," *Biomed. Signal Process. Control*, vol. 47, pp. 312–323, Jan. 2019.
[69] S. Kwon, "1D-CNN: Speech emotion recognition system using a stacked network with dilated CNN features," *Comput., Mater. Continua*, vol. 67, no. 3, pp. 4039–4059, 2021.

Li Zhang (Student Member, IEEE) received the B.S. degree from Nanjing University of Science and Technology, Nanjing, China, in 2015, where he is currently pursuing the Ph.D. degree. He was a Visiting Scholar with the Department of Electrical and Computer Engineering, Texas Tech University, Lubbock, TX, USA, in 2019.

Chang-Hong Fu (Member, IEEE) received the B.Eng. (Hons.) and Ph.D. degrees from The Hong Kong Polytechnic University in 2002 and 2008, respectively. He is an Associate Professor with Nanjing University of Science and Technology. He has authored or coauthored over 50 research articles in international journals and conferences.

Hong Hong (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Nanjing University, Nanjing, China, in 2010. He is currently an Associate Professor with Nanjing University of Science and Technology. He has authored or coauthored over 80 research articles in international journals and conferences.

Biao Xue (Member, IEEE) received the M.Eng. and Ph.D. degrees from Nanjing University of Science and Technology, Nanjing, China, in 2016 and 2020, respectively. He is currently an Assistant Professor with Nanjing University of Science and Technology.

Xuemei Gu received the bachelor's degree from Nanjing Medical University, Nanjing, China. She is currently a Nurse Supervisor with the Brain Hospital of Nanjing Medical University. Her research interests include the research on the Alzheimer's disease, cognitive disorders, and neuroimmune diseases.

Xiaohua Zhu (Member, IEEE) received the Ph.D. degree in communication and information system from Nanjing University of Science and Technology, Nanjing, China, in 2002. He is currently a Professor with the School of Electronic and Optical Engineering. He is also the Director of the Radar and High-Speed Digital Signal Processing Laboratory. He has authored or coauthored four books and more than 100 articles.

Changzhi Li (Senior Member, IEEE) received the B.S. degree in electrical engineering from Zhejiang University, China, in 2004, and the Ph.D. degree in electrical engineering from the University of Florida, Gainesville, FL, USA, in 2009. He is a Professor with Texas Tech University. His research interests include microwave/millimeter-wave sensing for healthcare, security, energy efficiency, structural monitoring, and human–machine interface. He was a recipient of the IEEE Microwave Theory and Techniques Society (MTT-S) Outstanding Young Engineer Award, the IEEE Sensors Council Early Career Technical Achievement Award, the ASEE Frederick Emmons Terman Award, the IEEE-HKN Outstanding Young Professional Award, the NSF Faculty Early CAREER Award, and the IEEE MTT-S Graduate Fellowship Award. He is an Associate Editor of the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES and the IEEE JOURNAL OF ELECTROMAGNETICS, RF AND MICROWAVES IN MEDICINE AND BIOLOGY.