

# The OBF Database: A Large Face Video Database for Remote Physiological Signal Measurement and Atrial Fibrillation Detection

Xiaobai Li<sup>†</sup>, Iman Alikhani<sup>†</sup>, Jingang Shi<sup>†</sup>, Tapio Seppänen<sup>†</sup>, Juhani Junttila<sup>‡</sup>, Kirsi Majamaa-Voltti<sup>‡</sup>,  
Mikko Tulppo<sup>‡</sup> and Guoying Zhao<sup>§†</sup>

<sup>†</sup>Center for Machine Vision and Signal Analysis (CMVS),  
University of Oulu, Oulu, Finland

<sup>‡</sup>Research Unit of Internal Medicine, Medical Research Center Oulu,  
Oulu University Hospital and University of Oulu, Oulu, Finland

<sup>§</sup>School of Information and Technology, Northwest University, Xi'an, China

Email: {xiaobai.li, iman.alikhani, jingang.shi, tapio.seppanen, juhani.junttila, mikko.tulppo, guoying.zhao}@oulu.fi  
& kirsi.majamaa-voltti@ppshp.fi

**Abstract**—Physiological signals, including heart rate (HR), heart rate variability (HRV), and respiratory frequency (RF) are important indicators of our health, which are usually measured in clinical examinations. Traditional physiological signal measurement often involves contact sensors, which may be inconvenient or cause discomfort in long-term monitoring sessions. Recently, there were studies exploring remote HR measurement from facial videos, and several methods have been proposed. However, previous methods cannot be fairly compared, since they mostly used private, self-collected small datasets as there has been no public benchmark database for the evaluation. Besides, we haven't found any study that validates such methods for clinical applications yet, e.g., diagnosing cardiac arrhythmias/disease, which could be one major goal of this technology.

In this paper, we introduce the Oulu Bio-Face (OBF) database as a benchmark set to fill in the blank. The OBF database includes large number of facial videos with simultaneously recorded reference physiological signals. The data were recorded both from healthy subjects and from patients with atrial fibrillation (AF), which is the most common sustained and widespread cardiac arrhythmia encountered in clinical practice. Accuracy of HR, HRV and RF measured from OBF videos are provided as the baseline results for future evaluation. We also demonstrated that the video-extracted HRV features can achieve promising performance for AF detection, which has never been studied before. From a wider outlook, the remote technology may lead to convenient self-examination in mobile condition for earlier diagnosis of the arrhythmia.

## I. INTRODUCTION

Physiological signals such as heart rate (HR), heart rate variability (HRV), and respiratory frequency (RF) are vital biomarkers of the body and necessary information for diagnostic and monitoring purposes of clinicians. They are constantly regulated by the autonomic activity of the nervous system and thus are indirect sources of information regarding nervous system functionality. Those biomarkers are monitored

in many health-care scenarios in clinical settings, whether it is a treatment procedure or examination and diagnostics. Conventional yet most prevalent means of cardiovascular monitoring often involves invasive or non-invasive sets of sensors, e.g., electrocardiography (ECG) sensors attached to the chest to measure cardiac electrical activity, or a pulse oximeter attached to finger tip or earlobe in order to measure the blood volume pulse (BVP). The wired connection of sensor sets causes discomfort (particularly for elderlies and infants) for those undergo long-term monitoring.

In recent years, there were studies reported that heartbeat could be remotely measured from facial videos [1], [2], [3], [4], [5]. Some studies [2], [4], [5] proposed color-based methods to measure heartbeat information by analyzing the color change of the face videos, while others [3] proposed motion-based methods. No doubt whatsoever the idea of estimating physiological signals remotely from face could be a promising alternative platform for inconvenient clinical and daily monitoring, and it explains the reason of growing number of research papers in this area. However, considering the emerging technology of remote physiological signal measurement, two key questions are yet to be addressed:

- 1) How can we fairly evaluate the proposed methods?,
- 2) Are the methods sufficiently accurate for medical purposes?

Previous studies were not able to comprehensively address these two prominent questions, and one major reason is the lack of a benchmark database. To our best knowledge, as of now we were not able to find any database published for the purpose of validating video-extracted physiological signals to detect abnormal cardiac activity patterns for clinical purpose.

In this paper, we introduce the Oulu BioFace (OBF) database, as a benchmark testing set to fill in the blank. OBF database includes large number of controlled subjects' data as well as patients' suffering from atrial fibrillation (AF). A

Corresponding Author: Guoying Zhao, email: guoying.zhao@oulu.fi

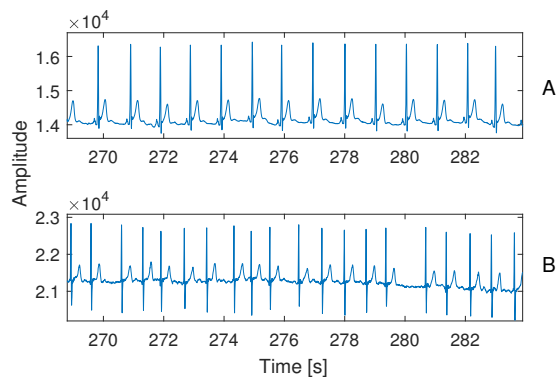


Fig. 1. (A). ECG of a healthy person, and (B). ECG of an AF patient. The irregular pulsation of the heart and morphological change of ECG are descriptive features of AF. If AF could be examined and detected by some convenient means at home, e.g., self-checking through a smart camera recording one's face, it would greatly help people who are in need of continuous monitoring.

sample ECG signal measured from a healthy person as well as an AF patient's signal is shown in Figure 1. Our aim is first to introduce the database, second to carry out experiments on OBF to primarily evaluate the accuracy of remote HR, HRV and RF measurements; and third to demonstrate the possibility of using video-extracted features for AF detection, which has never been studied before.

## II. BACKGROUND

### A. Remote physiological signal measurement from face

We first briefly review the main progress of studies regarding remote physiological signal measurements, derived from facial videos. By physiological signals we mean average HR, HRV and RF, while the literature mostly offers studies exploring average HR, and a few of them considered the measurements of HRV features and RF. Verkruyssen et al. [1] in 2008 first reported that HR can be measured from color facial videos recorded under ambient light. The authors also found that among the three color channels, the green one contains the strongest heart beat related component. Poh et al. in [6] proposed a method to use independent component analysis (ICA) of the three color channels for HR measurement from web-cam-recorded facial videos. Since then many studies [7], [4], [8], [5] proposed improved color-based approaches for average HR measurement from facial videos. The later approaches mainly focused on face region selection, reducing artifacts, caused by motions and illumination variations.

Nevertheless, there are also studies considering motion-based methods for HR measurement from facial videos. In 2013, Balakrishnan et al. [3] first proposed that HR can be measured by tracking the motion trajectories of facial points due to the interaction of gravity and arterial pulsations, and two studies [9], [10] followed their work by improving the trajectory-analyzing methods for HR measurement from facial videos. More detailed background of remote HR measurement studies is referred to one review paper [11].

All the mentioned studies focused on average HR measurement. However, HR counts the number of heartbeats within a specific period of time, which is a valuable yet very coarse and limited description of the cardiac activity. The HRV (also referred as the inter-beat-interval, IBI) on the other side, indicates the variation in time interval between heartbeats, which offers much more detailed information about cardiac and autonomic nervous system activity. But HRV requires more delicate measurement as it is computed using time points of each individual heartbeat. For rigorous applications and complex analysis such as clinical diagnosis, HRV is needed, while HR is too coarse for such purposes.

Compared to remote HR measurement studies, remote HRV analysis were only addressed in a few papers. In 2011 Poh et al. [2] extended their previous work and analyzed HRV and RF from facial videos using a similar yet slightly improved approach. The method was validated on a small self-collected dataset of 12 persons. McDuff et al. [12] used the same approach to compute HRV features derived from face, and applied the face-derived HRV features for stress level classification on a self-collected dataset of ten subjects. In one later work [13], the authors proposed a chrominance-based method with continuous wavelet transform (CWT) to attenuate the HRV measurement noise, which was validated on a self-collected dataset of six subjects. Besides, the feasibility of RF estimation from facial videos was also proved in [2], [14].

To sum up, remote physiological signal measurements is a new area in machine vision research, and growing number of pieces of research is being reported. Considering the prospective future of the topic, two important questions need to be addressed:

- **How one can fairly evaluate the proposed methods?** Previous studies share one common problem about the testing datasets. Most of the proposed methods were only validated on self-collected datasets, containing small sample size. The limitation of using private datasets is that there is no benchmark that methods can be fairly evaluated and compared for this new emerging topic. There were some studies [4], [8], [5], [10] tested on a shared MAHNOB-HCI database [15]. The MAHNOB-HCI was originally designed for emotion analysis, and the videos involve various issues and poor illumination, which might be too complicated for delicate HRV measurements. It is better to start the exploration of remote HRV analysis methods first on a fine and well-controlled dataset (as the OBF database proposed here), and then proceed to more challenging datasets later.
- **Are the derived signals sufficiently qualified for medical purposes?** Although the physiological signals can be measured for other purposes (e.g. emotion analysis), one ultimate and the most demanding purpose is clinical usage (e.g. diagnosing cardiac arrhythmia such as AF). At the moment, the authors could not find any proposed remote measurement method which has been validated for medical diagnosis, and thus, the primary step towards that is to collect a proper dataset for such target.

## B. Atrial fibrillation (AF)

AF is one of the most common sustained widespread cardiac arrhythmia and according to American Heart Society at least 2.7 million Americans are living with AF. In AF electrical impulses do not originate in the sinoatrial (SA) node and cause disorganized and uncoordinated pulse through cardiac muscle. Thus, AF is usually accompanied with rapid and irregular heartbeats. It can be diagnosed using electrical activity of the heart recorded by ECG. ECG signal measured from a healthy person versus a patient diagnosed with AF is shown in Figure 1. The peak population of AF is owned by elders over 65-year-old, and it is likely to happen for younger people with either family history, or with related diseases, like hypertension or diabetes. AF is a major cause of hospitalization today. Constant AF may lead to severe or even fatal consequences (e.g., ischemic stroke). But in case of early diagnosis, AF could be controlled or even cured with medications and other treatments, such as electrical cardioversion. Unfortunately the AF in early stage is less symptomatic for some people, who may miss the opportunity of early intervention. Remote examination of AF in a convenient mean, e.g., with a camera at home, could increase the chance of early intervention and cut a substantial expenses in treatment and hospitalization of suspected AF patients.

## C. Contribution of the current study

In this paper, we introduce the OBF database. OBF database includes large number of face videos and synchronized reference physiological signals recorded from both healthy subjects and AF-diagnosed patients. The OBF database can be used as a benchmark for:

- 1) Evaluating the methods of physiological signal derivation (HR, HRV and RF) from facial videos (of resting-state vs. post-exercise; or of three groups of different skin tones);
- 2) Training computational models to use video-extracted physiological features for AF detection.

We also carry out experiments on OBF database to:

- 1) Evaluate the accuracy of HR, HRV and RF measurement from face, and report the performance as baseline results for future comparison with other proposed methods;
- 2) Use the face-extracted HRV features for detecting AF cases [16].

## III. OULU BIO-FACE (OBF) DATABASE

### A. Equipment and setup

The data collection was first carried out for healthy subjects at the University of Oulu, and then at the Oulu University Hospital for AF patient's recording. The recording environment and setups were strictly controlled so that data achieved at both sites are of the same properties. A multi-modal setup was arranged for synchronized recording of two video channels and three channels of physiological signals from each participant. The recording set includes a computer, a color (RGB) camera, a near infra-red (NIR) camera, a biosignal data acquisition

(BIO) device with three sets of sensors, and two LED lights. Detailed information of all equipment are listed in Table II. The software used for recording physiological signals is Bio Trace+ NX10<sup>1</sup>, which is provided by the device company. We also used a free software VirtualDubMod<sup>2</sup> for the recording of NIR videos.

TABLE I  
STATISTICAL INFORMATION OF THE PARTICIPANTS.

	Healthy (n=100)	Patient (n=6 <sup>3</sup> )
Age (y)	31.6 ± 8.8, [18, 68]	68.1 ± 13.2, [43, 81]
Gender	61% M, 39% F	50% M, 50% F
Ethnic	Caucasian:32%, Asian:37%, Others:31%.	Caucasian: 100%
Weight (Kg)	71 ± 16	88 ± 13
Wear eyeglasses	39%	50%

RGB videos recorded using the color camera are the core data of OBF database. Most commercial oximeters use NIR bandwidth light sources for contact BVP measurements, while the NIR face videos haven't been studied yet for the same purpose in non-contact video-extracted context. Thus we also added an additional NIR video set to explore the feasibility of physiological signal derivation from the NIR facial videos.

The recording environmental setup is illustrated in Figure 2. A chair was placed in front of cameras at one meter distance. Two lights standing on both sides of the cameras and illuminate the faces at 45 degree angle with a distance of 1.5 meters. As OBF is the first database built for evaluating methods for remote measurement of physiological signals, including very challenging HRV features, there is possibility that the video-extracted HRV features are too poor for the AF detection at this stage. We intentionally design the setup to make sure the videos could be of relatively high quality, so that future studies can explore methods on the data as a fundamental starting point. More challenging data (recorded in wild condition) will be considered as the next step.

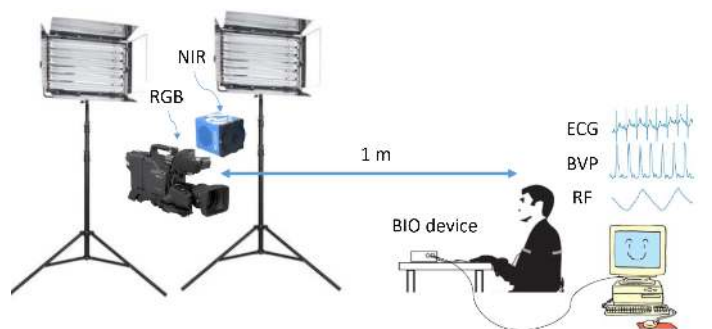


Fig. 2. The setup of the OBF database collection.

<sup>1</sup><https://www.mindmedia.com/downloads/>

<sup>2</sup><http://www.virtualdub.org/>

<sup>3</sup>More patients data are being collected for a target number of 100. Statistic info are based on current 6 patients.

TABLE II  
EQUIPMENT AND SPECIFICATIONS AND SETTINGS.

Device	Specifications	Settings	Output
Computer	HP EliteDesk	Windows 7 OS	N/A
RGB camera	Blackmagic URFA mini	ISO 400, FPS 60, HD 1920 x 1080	RGB video
NIR camera	Customized USB 2.0 Camera box (SN9C201&202)	640 x 480, FPS 30	NIR video
LED lights (2)	Aputure, LightStorm LS 1c	Brightness: 3500 lux, Temperature: 5500 k	N/A
Biosignal Acquisition	NeXus-10 MKII	N/A	N/A
ECG sensor	NX-EXG2B	256Hz	ECG Signal
Respiratory Belt	NX-RFP1B	32Hz	RF Signal
BVP sensor	NX-BVP1C	128Hz	BVP Signal

TABLE III  
DATA COMPOSITION OF THE OBF DATABASE.

Modalities	Data length / person		Total(min)
	Healthy (n=100)	AF (n=6)	
RGB Video, NIR Video ECG, BVP, RF	Resting state: 5min	Prior-treatment: 5min	10600
	Post-exercise: 5min	Post-treatment: 5min	

### B. Participants

Healthy participants were recruited from University of Oulu through internal email calls. The research content and data collection procedure were explained in the emails, and altogether 100 healthy (no heart disease history) volunteers were recruited. AF patients were recruited from Oulu University Hospital by cardiologists and nurses. Most of the patients were reserved on a list to receive cardioversion treatment, while the rest were acute patients recruited from emergency center. The collection of AF patients data is ongoing at the moment. Already six patients participated before the experiment, while more patients are reserved and data of altogether 100 patients of diverse ethnicities will be collected in the following months and published as another database. General statistical information of the participants are summarized in Table I. The healthy group covers a wide range of age, and owns a diverse ethnicity.

**Ethical issues:** The study was performed according to the Declaration of Helsinki, and the local committee of research ethics of the Northern Ostrobothnia Hospital District approved the protocol. All the subjects gave written informed consent. The patients were accompanied and arranged by professional cardiologists and nurses through the whole measurement.

### C. Acquisition procedure

Before the execution of the recording, each participant read and signed the consent agreement. Then the participants were guided to seat in front of the cameras, and bio-sensors were attached to them. The participants were instructed to remain comfortable and seated facing the cameras. There were two sections of recording, each lasts for five minutes. Healthy participants were asked to exercise (climbing the stairs) for about five minutes between the two sessions in order to record higher heart rate for the next session. Thus, in session one we recorded the resting state, and in session two we recorded post-exercise state of elevated HR and RF.

For AF patients, session one was recorded after they were examined by the doctor prior to the cardioversion treatment (with the AF symptom), while session two was recorded three hours later after the cardioversion treatment (with recovered sinus heart rhythm).

Regarding the data synchronization, for each session, the cameras and the bio-sensors were synchronized by passing a visual cue (for both of the cameras) and pressing a trigger button (on the bio-signal data acquisition) simultaneously, both at the start and at the end of the recording.

### D. Data composition

Table III lists the data included in the database. There are data from 100 healthy participants and currently six AF patients (ongoing collection). For each participant, multi-modal data were recorded simultaneously for two sessions of five minutes. Sample data of each modality is shown in Figure 3. For healthy participants, session one is resting state and session two is post-exercise with elevated HR; for AF patients, session one is recorded before the cardioversion treatment (with AF) while session two is after the treatment (with recovered healthy heartbeat rhythm for most cases). The total data sum up to 10600 minutes so far for each of the five modalities.

## IV. METHOD

We propose a framework to measure average HR, HRV features, and RF for AF detection. The main steps of the framework is shown in Figure 4. Our framework was inspired by paper [2], while adjustments were made in several steps through our prior testing and hence, the current framework achieves superior performance than the original method.

### A. Facial ROI tracking

Given an input video of  $n$  frames, we use discriminative response map fitting (DRMF) method [17] to detect the face and locate 66 landmarks on the first frame. Then a customized region of interest (ROI) was defined by the facial contour landmarks. The defined ROI was tracked through the video using the Kanade-Lukas-Tomasi (KLT) algorithm [18] (see Figure 4 A). The raw heartbeat signals (raw RGB signals) are calculated by averaging the pixel values within the ROI of the three color channels of each frame, as  $R_{raw}$ ,  $G_{raw}$  and  $B_{raw}$ , which all are vectors of length  $n$ .

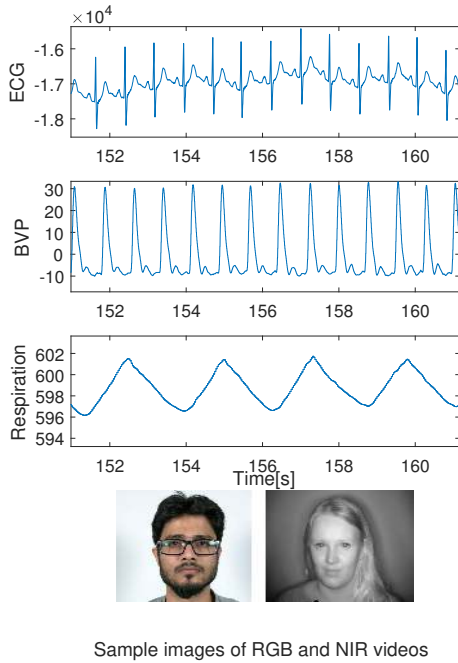


Fig. 3. Samples of raw data of the OBF database.

We use ROI tracking instead of face detection on each frame as [2], since our method is able to avoid including background pixels and acquire smoother raw signal. Besides, in [2] the independent component analysis (ICA) was utilized to decompose the three signals of RGB channels for reducing motion artifacts. Here we use the green channel signal  $G_{raw}$  (see Figure 4 B), because in our prior testing we found the ICA output component sometimes might include unaccountable signal patterns, especially for the cases that contain short but dramatic fluctuations.

### B. Temporal filtering and interpolation

The raw green signal ( $G_{raw}$ ) is then filtered and interpolated. We first use a detrending filter [19] with the smoothing parameter  $\Lambda = 300$  (cutoff frequency of 0.66 Hz) to remove the slow floating trend of the raw signal. After that the signal is normalized by subtracting its mean and divided by its standard deviation. Then the normalized signal is smoothed using a five-point moving average filter.

The aim of these three steps of filtering are to remove irrelevant noise and acquire a smoother signal. Since the reference ECG signal were recorded at 256Hz, the smoothed signal is interpolated with a cubic spline function at a sampling frequency of 256 Hz. In paper [2] a bandpass filter was applied. We did not include any bandpass filter, as we found that it might over-smooth the signal and alter the pulse peak locations. The filtered and interpolated signal is defined as  $G_{smooth}$  (see Figure 4 C) for the next step of IBI signal computing and HRV analysis.

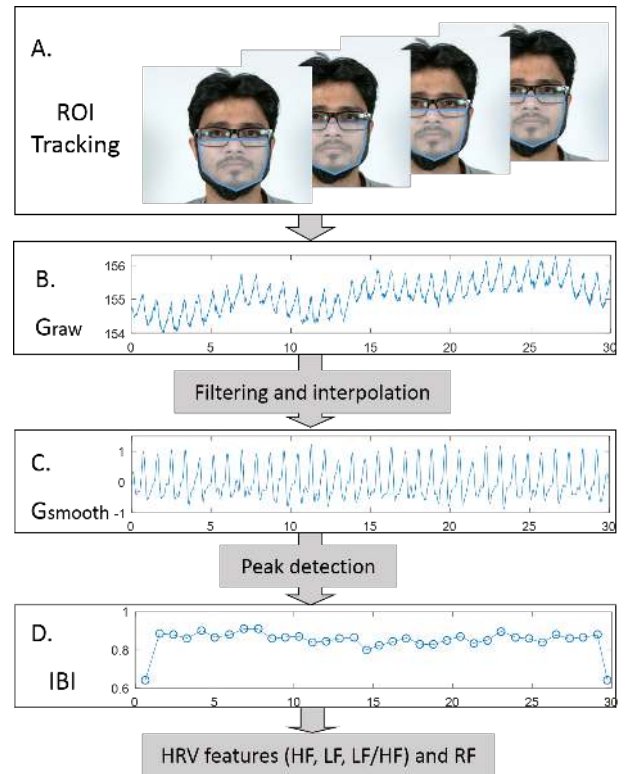


Fig. 4. Framework proposed for HRV measurement from facial videos.

### C. Quantification of physiological parameters

Subsequently, we developed a customized peak detection function for detecting pulse peaks from  $G_{smooth}$  and compute the IBI signal (see Figure 4 D). The average HR (expressed as  $\overline{HR}$ ) of the input video can be calculated from the average IBI (known as  $\overline{IBI}$  and in second unit) as  $\overline{HR} = \frac{60}{\overline{IBI}}$  in beat-per-minute (bpm) unit.

In this study, we perform standard spectral analysis of HRV (a.k.a., the IBI signals), as an important indicator of autonomic nervous system activity [20]. According to the literature, there are three main frequency components in HRV signals, including very low frequency (VLF), low frequency (LF) and high frequency (HF) components. However, the justification of existence and explanation of VLF components are still under question and thus, we consider only LF and HF components in this study. We measured those two components in normalized units (n.u.), which represent the relative value of each component to the total power. These two branches of spectrum indicate the controlled and balanced behavior of the autonomic nervous system.

For the sake of stationarity of the signals, we segmented the five-minute-long signals into segments of 30 seconds. Then we applied Lomb-Scargle periodogram which returns the power spectral density (PSD) estimation and can handle irregularly spaced time series, as in our case the IBI signal is not equidistant. Thus, the LF and HF in n.u. and their ratio, extracted from PSD is reported as well as the extracted RF and  $\overline{HR}$ . Our results are reported with the same style as the



results reported in [2].

However, for the AF detection, we use some of IBI standard features [20], including

- Time-domain: mean IBI, standard deviation of IBI, root mean square of successive differences (RMSSD), square root of the sum of the squares of differences of individual values compared to the mean value, divided by the number of RRI in a period (RMSM) and percentage of samples with more than 50ms difference from the consecutive beat (pNNI50),
- Geometrical-domain: Poincaré plot standard deviations (SD1, SD2)
- Spectral-domain: LF, HF and their ratio in normalized unit (n.u.).

## V. EXPERIMENTS AND RESULTS

There are two parts of experimental results. First, we evaluate the accuracy of our framework on OBF database by comparing physiological signals (HR, HRV features, and RR) measured from facial videos with reference ECG and BVP signals. Second, we use the face-extracted HRV features to perform AF detection.

### A. Physiological signals measured from facial videos

100 healthy participants data from OBF database are used for the evaluation of the proposed framework. All data are cut into 30 seconds long segments, in which each participant yields 20 segments of the data and there are entirely 2000 units of sample data.

Using the proposed framework, we first compute  $\overline{HR}$  of each segment and compare it with the corresponding reference ECG. BlandAltman plot [21] is used to evaluate the  $\overline{HR}$  measured from RGB videos as shown in Figure 5, and statistical results are listed in Table IV. It can be seen that by using the proposed method framework we can measure  $\overline{HR}$  from RGB videos with high accuracy on most of the cases, with the average error of  $-0.03 \pm 2.16$  bpm. We also implemented the original method proposed in [2] which includes the process of ICA and bandpass filtering, but the accuracy (average error =  $6.55 \pm 11.69$  bpm) is significantly lower than our results for the  $\overline{HR}$  measurement. One possible reason might be that the ICA is not very effective for some of the cases, as we mentioned in the method. In the following experiments, the IBI signal computing will completely be based on the heartbeats derived in the current step, since our framework outperformed [2]’s method by acquiring more accurate  $\overline{HR}$ .

HR measured from NIR videos are shown in Figure 6, and statistical results are listed in Table V. The results show that heartbeat information can also be extracted from NIR facial videos, but with poorer performance than those from RGB videos. The average error is  $-1.66 \pm 6.60$  bpm.

For HR measurement from RGB videos, there are a few cases with large errors. One sample of those erroneous cases is illustrated in Figure 7. One possible reason of the large error for this case (and several other similar cases) is the darker skin tone. Although on general the accuracy of HR measurement

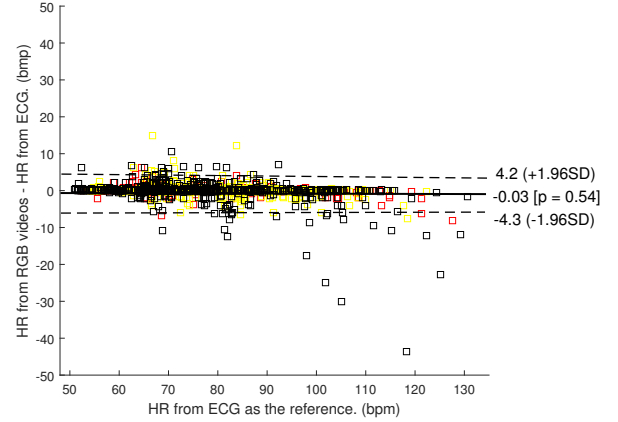


Fig. 5. BlandAltman plot of the  $\overline{HR}$  measured from RGB videos vs reference ECG. Data samples are divided into three groups according to the skin tone: group 1 (red) includes samples of lighter skin tone, group 2 (yellow) includes samples with middle-range skin tone, and group 3 (black) includes samples of darker skin tone. Samples with large error are mostly from group 3.

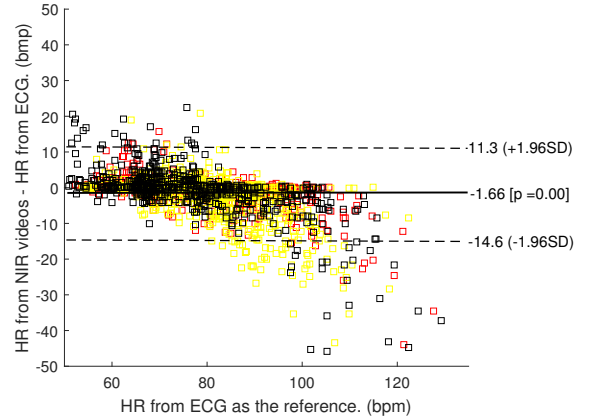


Fig. 6. BlandAltman plot of the  $\overline{HR}$  measured from NIR videos vs reference ECG. Data samples are divided into three groups the same way as Figure 5.

on NIR videos is lower comparing to RGB videos, one note worth being pointed out is that, NIR videos might provide supplementary information for some challenging RGB cases of dark skin tone. As for the sample shown in Figure 7, the heartbeat signal extracted from NIR video is superior than the corresponding RGB video, as NIR image is more robust to skin tone. How to use the complimentary information from RGB and NIR videos for remote physiological signal measurement is worthy of exploration in future, and the OBF database provides a decent platform for this study.

We further computed IBI signals to extract HRV features (LF, HF and LF/HF). Figure 8 shows the video-measured signals and features and compared them with those of the reference ECG. The statistical results of remote HRV measurement from RGB videos are reported in Table IV, and those of NIR videos are reported in Table V. The RF is computed using the peak location of the HF component from the PSD of

TABLE IV  
SUMMARY RESULTS ON OBF RGB VIDEOS USING THE PROPOSED FRAMEWORK.

	HR (bpm)	RF (Hz)	HRV		
			LF (n.u.)	HF(n.u.)	LF/HF
Error Mean	-0.03	0.13	0.17	0.17	1.2
Error SD	2.16	0.17	0.19	0.19	2.2
RMSE	2.16	0.22	0.26	0.26	2.5
R	0.99	0.23	0.55	0.55	0.38

TABLE V  
SUMMARY RESULTS ON OBF NIR VIDEOS USING THE PROPOSED FRAMEWORK.

	HR (bpm)	RF (Hz)	HRV		
			LF (n.u.)	HF(n.u.)	LF/HF
Error Mean	-1.66	0.15	0.24	0.24	1.4
Error SD	6.60	0.17	0.23	0.23	2.3
RMSE	6.80	0.23	0.33	0.33	2.8
R	0.87	0.09	0.25	0.25	0.20

the IBI signal. According to the reported figures, the HR can be measured at a higher accuracy, while basically the HRV features are more challenging with larger error margin and much lower correlation efficiency  $R$  (mostly less than 0.5).

As mentioned in Section II, the primary focus of the current paper is to introduce OBF database as a benchmark database to facilitate investigation of remote physiological signal measurement. The proposed framework only includes basic steps of HRV analysis, which could serve as a foundation for future work. Results in Table IV and V are listed as baseline results for later evaluation on OBF database. More advanced methods will be explored in future for more reliable physiological signal derivation, especially HRV features.

### B. AF detection using video-extracted HRV features.

We also explore whether it is possible to use video-extracted HRV features for detecting AF cases against healthy ones. The five-minute resting-state recording of each healthy participant, and the five-minute prior-treatment recording of each AF patient are utilized for a binary classification test. Data are also divided into segments of 30-second length. Entirely, there are 60 AF segments (of six AF patients) and 1000 healthy control samples (of 100 healthy participants).

For the task of AF detection, we only considered the variation of IBIs, derived from ECG and RGB videos. A set of features explained in IV-C are extracted first from the IBI data of reference ECG, and then from the IBI data of RGB videos. Radial-basis function (RBF) kernel support vector machine (SVM) is used as the classifier and the kernel parameters are optimized using MATLAB heuristic procedure optimization. About the testing protocol, we run 100 rounds of classification tests and summarize the average accuracy. For each round, We randomly select six (out of 100) healthy subjects as the control group so that we have balanced samples of the two categories. The data is divided into two halves, each contains 30 healthy samples from three healthy subjects, and 30 AF samples from three patients. We first train on one half and test on the other half, and then exchange for cross validation.

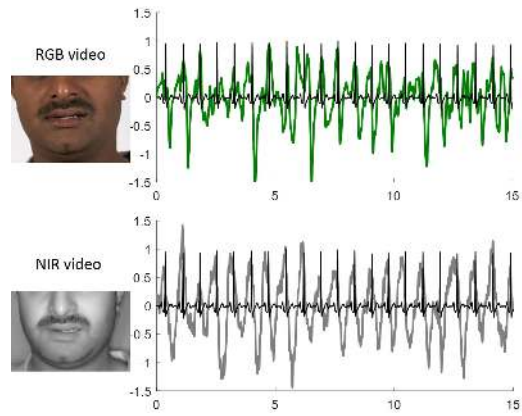


Fig. 7. One error sample. HR signal computed from RGB video is noisy due to the dark skin tone; while HR signal computed from NIR is superior, as NIR video is less prone to skin color. The black curve is the reference ECG.

TABLE VI  
THE PERFORMANCE OF AF DETECTION USING ECG- AND RGB-DERIVED IBI SIGNALS.

	ACC	SP
ECG-derived features	88.8%	89.1%
RGB-video-derived features	78.3%	78.8%

The accuracy ( $ACC = \frac{TP+TN}{TP+TN+FP+FN}$ ) and specificity ( $SP = \frac{TN}{TN+FP}$ ) are used as the validation metrics. The average ACC and SP of the 100 rounds classification are reported in Table VI for both ECG-derived features and video-derived features. We achieved accuracy of about 78.3% using video-extracted HRV features for AF detection. ECG-derived features of course performed better, as ECG is the golden standard employed by doctors for AF diagnosis. Our results using video-extracted features are quite promising, since this is the first work that ever demonstrated using remotely measured HRV features for AF detection. We are gathering more data of AF patients, and more advanced methods will be explored in future to reduce errors of video-extracted features and further improve the AF detection performance.

## VI. CONCLUSION

In this paper we introduce the OBF database, which includes large number of facial videos with corresponding reference physiological signals as a benchmark database for validating methods on remote physiological signals measurement. OBF includes data recorded from 100 healthy subjects and six patients diagnosed with AF, which is the most widespread sustained cardiac arrhythmia in clinical settings. This new database fills the gap between remote physiological signal analysis and demonstration of the new technology for medical applications. Accuracy of HR, HRV and RF measured from OBF videos are provided as the baseline results using a proposed framework method. We also demonstrated that the video-extracted HRV features can achieve promising performance for AF detection, which has never been studied before.

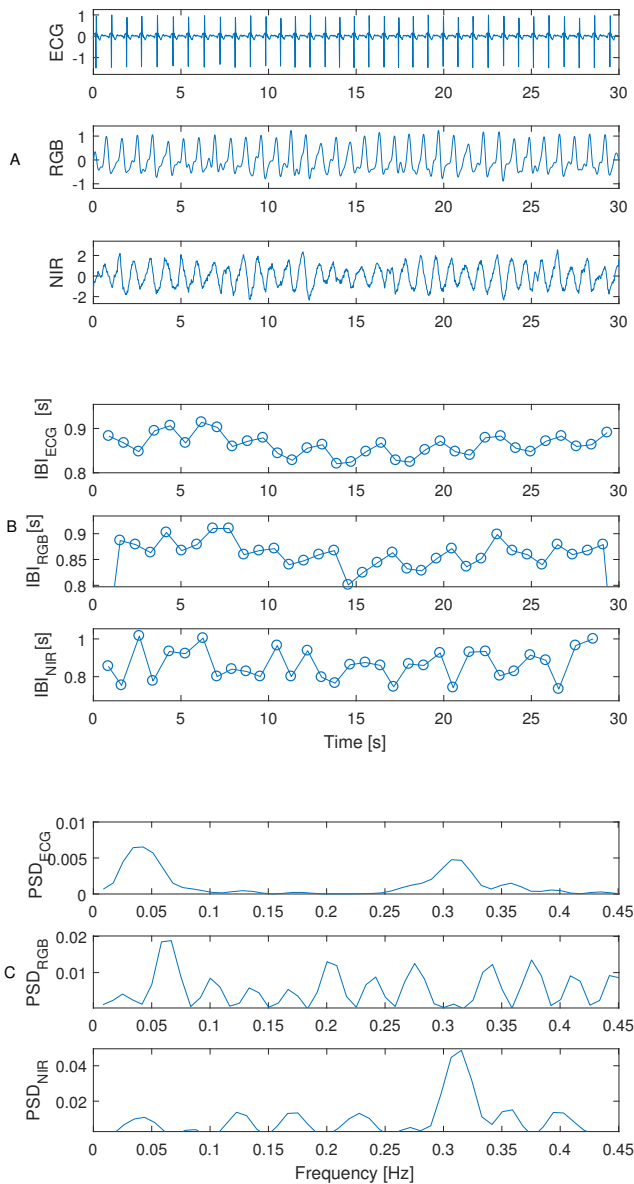


Fig. 8. Data measured from facial videos versus the reference ECG. Three sets of subplots are depicted. Set A are the heartbeat signals recorded from reference ECG and extracted from RGB and NIR face videos. Set B shows the IBI signals derived from corresponding heartbeat signals from the set A. Set C is the PSD of the IBI signals depicted in the set B.

The OBF database can provide sufficient amount of data for future explorations regarding remote physiological signal measurement from face. More data of AF patients is under collection, and will be released later as an extended dataset which includes 100 AF patients with different ethnicities. More advanced methods will be explored in future (e.g., by combining RGB and NIR videos) to achieve more accurate HRV features for AF detection.

## VII. ACKNOWLEDGMENTS

This work was supported by Academy of Finland, Tekes Fidipro program (Grant No. 1849/31/2015) and Tekes project (Grant No. 3116/31/2017), Infotech, Tekniikan Edistämisyätiö Foundation, National Natural Science Foundation of China (No. 61772419).

## REFERENCES

- [1] W. Verkruyse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Optics express*, 2008.
- [2] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in non-contact, multiparameter physiological measurements using a webcam," *IEEE transactions on Biomedical Engineering*, 2011.
- [3] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.
- [4] X. Li, J. Chen, G. Zhao, and M. Pietikainen, "Remote heart rate measurement from face videos under realistic situations," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.
- [5] S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, J. F. Cohn, and N. Sebe, "Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [6] 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*, 2010.
- [7] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rppg," *IEEE Transactions on Biomedical Engineering*, 2013.
- [8] A. Lam and Y. Kuno, "Robust heart rate measurement from video using select random patches," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015.
- [9] L. Shan and M. Yu, "Video-based heart rate measurement using head motion tracking and ica," in *Image and Signal Processing (CISP), 2013 6th International Congress on*. IEEE, 2013.
- [10] M. A. Haque, R. Irani, K. Nasrollahi, and T. B. Moeslund, "Heartbeat rate measurement from facial video," *IEEE Intelligent Systems*, 2016.
- [11] M. Hassan, A. Malik, D. Fofi, N. Saad, B. Karasfi, Y. Ali, and F. Meriaudeau, "Heart rate estimation using facial video: A review," *Biomedical Signal Processing and Control*, 2017.
- [12] D. McDuff, S. Gontarek, and R. Picard, "Remote measurement of cognitive stress via heart rate variability," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, 2014.
- [13] R.-Y. Huang and L.-R. Dung, "Measurement of heart rate variability using off-the-shelf smart phones," *Biomedical engineering online*, 2016.
- [14] F. Zhao, M. Li, Y. Qian, and J. Z. Tsien, "Remote measurements of heart and respiration rates for telemedicine," *PLoS One*, 2013.
- [15] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, 2012.
- [16] P. Przystup, A. Poliński, A. Bujnowski, T. Kocejko, and J. Wtorek, "A body position influence on ecg derived respiration," in *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*. IEEE, 2017.
- [17] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Robust discriminative response map fitting with constrained local models," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.
- [18] C. Tomasi and T. Kanade, *Detection and tracking of point features*. CMU, 1991.
- [19] M. P. Tarvainen, P. O. Ranta-aho, and P. A. Karjalainen, "An advanced detrending method with application to hrv analysis," *IEEE Transactions on Biomedical Engineering*, 2002.
- [20] A. J. Camm, M. Malik, J. Bigger, G. Breithardt, S. Cerutti, R. J. Cohen, P. Coumel, E. L. Fallen, H. L. Kennedy, R. E. Kleiger *et al.*, "Heart rate variability: standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology," *Circulation*, 1996.
- [21] J. M. Bland and D. G. Altman, "Measuring agreement in method comparison studies," *Statistical methods in medical research*, 1999.