Assessing an Automatic Procedure of Extraction of Physiological Parameters from Skin using Video Photoplethysmography

R. Barbieri, *Senior Member, IEEE*, R. Levi, M. Mollura, I. Marsella, L. Ficarelli, M. Negro, L. Cerina, L. Mainardi, *Senior Member, IEEE*.

***Abstract*— Video Photoplethysmography (vPPG) allows for estimation of blood volume pulse (BVP) from the skin by means of a video camera recording at high frequency rate. The estimation procedure presents several drawbacks in its application to real world conditions, such as light changes or movements that often generate artifacts in the extracted BVP waveform. In addition, the process requires a skin segmentation algorithm to distinguish skin pixels from the background. To date, even the most refined skin segmentation algorithms still need a manual definition that could lead to incorrect pixel classification, and consequently to a decrease in the signal-to-noise ratio (SNR). We here propose a fully autonomic procedure able to extract BVP from video recordings of the skin in real world conditions.**

The experimental protocol is designed to record the signals of interest and to evaluate changes in the Autonomic Nervous System modulation of the heart during a baseline condition and a controlled breathing phase. Video recordings are gathered from 4 young healthy subjects (age: 21$\pm$1 years). vPPG signals are processed in order to extract the BVP waveform, and a peak detection algorithm detects pulse wave peaks that are then used to compute specific measures of heart rate variability (HRV).

The procedure is successfully validated by comparing the extracted HRV measures against those extracted using a finger photoplethysmograph (fPPG) using three different skin segmentation algorithms from BVP signals.

***Clinical Relevance*— The proposed procedure paves the way for a monitoring tool able to collect high resolution BVP measurements from the subject's hand in controlled clinical settings.**

I. Introduction

In recent years, video photoplethysmography (vPPG) has been employed in order to extract cardiovascular parameters with no contact with the subject, and validated both on face and on skin regions. In particular, heart rate variability (HRV) indexes have been investigated by applying experimental protocols that could elicit the Autonomic Nervous System (ANS), such as rest to stand [1] or thermal shock [2] protocols. These analyses have provided evidence that vPPG is able to correctly estimate the influence of the ANS on the heart by HRV frequency domain indexes established in the last decades [3].

In order to properly extract the BVP waveform by vPPG technology, previous studies have proposed algorithms able to distinguish skin pixels from the background. One of the most popular methods analyzes face skin pixels after the application of a face detection algorithm [4], such as the Viola-Jones algorithm [5]. Drawbacks of this procedure are related to improper recognition of features from the face, as well as to irregular illumination on the skin, due to real world luminosity conditions.

In our previous study, a semi-automatic algorithm was proposed in order to extract the vPPG signal from video recordings of the hand palm, showing an overall agreement in terms of time domain features of HRV indexes [2]. This methodology required the manual selection of skin pixels and background pixels, which lacks automatism and homogeneity. Here, we present a novel automatic vPPG signal extraction procedure from skin pixels and we test it on 4 healthy subjects by comparing time and frequency domain HRV indexes in baseline resting state and during a controlled breathing phase. In addition, we compare three different classification algorithms able to distinguish skin pixels from background pixels.

HRV indexes are calculated after automatic systolic peak detection [6]. Standard HRV parameters are validated by characterizing the activity of the ANS, in particular pNN50 and SDNN, and very-Low Frequencies (VLF), Low Frequencies (LF) and High Frequencies (HF) [3].

II. State of the art

The Video Photoplethysmography (vPPG) is an innovative technology which aims at estimating the blood pressure pulse from video recordings of the skin. The first scientific publication related to vPPG was published by Verkruysse et al. in 2008 [7], were the authors provided a solution to overcame problems related to the fingerPPG (fPPG) technology, which needs a continuous contact with only small portions of the skin, such as fingers or earlobes. However authors provided results only on a limited cohort of subjects, therefore it was necessary to carry out studies on a larger sample. In the last years, several studies were conducted in order to extract heart rate (HR) from video recording of the skin, both for resting state condition [8] and during acute hypoxic challenges [9].

Other studies investigated HRV indexes, which are recognized as related to the activity of the Autonomous Nervous System (ANS) on the heart [3]. Iozzia et al. evaluated the vPPG signal during a rest-to-stand protocol on 60 subjects and estimated HRV indexes (both in time and frequency domain) [4], demonstrating a strong correlation with those calculated from ECG. However, some shortcomings of this approach were the movement's artifacts of subject's faces and real world condition of luminosity.

R. Barbieri, R. Levi, M. Mollura, I. Marsella, L. Ficarelli, M. Negro, L. Cerina and L. Mainardi, are with the Department of Electronics, Informatics and Bioengineering, Politecnico di Milano, Milano, Italy .

A key stage for the extraction of the vPPG signal is the implementation of a skin detection algorithm able to distinguish skin pixels from the background. There are mainly two kinds of approaches to this issue: pixel-based and ROI-based algorithms. Pixel-based approach relies on the definition of a-priori thresholds that could distinguish skin pixels [10], using different color spaces (RGB,YCbCr,HSV). ROI-based algorithms are based on the definition of region of interest on the skin after the search for features, such as eyes and nose [5]. However, both algorithms present some drawbacks when applied to real world condition of luminosity or on different skin types. In order to perform a better selection of the skin pixels, Oghaz et al. provided a study where several machine learning algorithms were compared together in order to reach the highest value of accuracy and area under the receiver operating curve (ROC-AUC) [11]. The authors showed higher performances than methods described before in terms of skin pixel classification.

III. Methods

To address current limitations, we developed a novel procedure that autonomously extracts vPPG signal from video recordings of the hand, according to the following steps:

1) Automatic selection of skin pixels from pixels related to background or sensors placed on fingers;
2) vPPG signal extraction from skin pixels using three different classification algorithms;
3) Signal processing stage to obtain BVP waveform from raw vPPG signal;
4) Statistical comparison of HRV indexes between BVP signal from fPPG and vPPG.

A. Experimental protocol

Four healthy subjects (age: 21 ± 1) volunteered for this study. Each signed a written informed consent describing the experimental protocol, whose procedure received the approval from the Politecnico di Milano Ethical Committee. Volunteers' skin type range from type II to type III of the Fitzpatrick scale [12].

Each subject was equipped with 4 different sensors:

- ECG sensor with chest-placement configuration;
- respiration sensor located in the middle of the chest;
- fPPG sensor placed on the middle finger of the left hand;
- skin-conductance sensor on the index and ring finger of the left hand.

Left hand is then placed on a hand support in order to limit artifacts due to movements.

After a resting phase of at least 15 minutes for each subject, we recorded 5 minutes of baseline condition (*Rest*). Then, after a cold pressure maneuver and a grip task, we asked the subject to perform deep inspirations and exhalations with a pace of 1 second each for 2 minutes in total (*Resp*). In this analysis we perform a comparison between the *Rest* phase and the *Resp* phase in terms of HRV-indexes.

The subjects' left hand was centered in the middle of the frame and an Imaging Source DFK-23UM021 camera equipped with a 15 mm mixed focal length lenses was used with a frame-rate of 115 frames-per-seconds. The physiological signals were acquired through FlexComp Infiniti by Thought Technologies, Inc. device, sampling ECG, BVP, skin conductance and respiration sensors at 2048 Hz.

B. Automatic Algorithm for Pixel Classification

The algorithm is based on K-Means clustering technique, an unsupervised learning method that divides a set of N samples into K disjoint clusters C, each described by the mean of the samples in the cluster, called *centroid* (μ); given a distance metric, the algorithm groups each observation x_i in the corresponding cluster by minimizing its distance with the *centroid* [13]. We used K-Means algorithm to group pixels of the first frame for similarity, using 4 clusters (background, shadows, sensors equipment, hand) and evaluating inertia distance metric, or within-cluster sum-of-squares criterion:

$$inertia = \sum_{i=0}^{N} min_{\mu_j \in C}(\|x_i - \mu_j\|^2) \tag{1}$$

The one cluster representing the skin was identified by looking at which cluster the central pixel of the image belongs to.

Then, pixels belonging to skin cluster are grouped together and compared with the others to create a binary classification training set to feed the skin detection algorithms.

C. Skin Segmentation Algorithms

Once the hand region was defined, skin segmentation algorithms were trained on the first frame of the video recording using three different supervised learning algorithms in order to validate the model with higher performances in estimating HRV indexes.

The first algorithm is based on a Decision Tree classifier (DT), a non-parametric supervised method minimizing an impurity function H at each split node [14]. In particular we applied the Cross-Entropy function defined as:

$$H(X_m) = -\sum_{k} p_{mk}(1 - p_{mk}) \tag{2}$$

where p_{mk} is the proportion of class k observations in node m.

The second method used entailed Support Vector Machines (SVM), defining an hyperplane that is able to properly separate records looking at the target feature [15]. A quadratic optimization problem with linear constraints needs to be solved in order to evaluate the coefficients w and the intercept b of the hyperplane:

$$\begin{aligned} min_{w,b}\ & \frac{1}{2}\|w\|^2 \\ s.to\ \ & y_i(w'x_i - b) \geq 1 \end{aligned} \tag{3}$$

The third method is based on Logistic Regression (LR), evaluating the posterior probability $P(y = 0|x)$ by using a logistic function:

$$P(y = 0|x) = \frac{1}{1 + e^{w'x}} \tag{4}$$

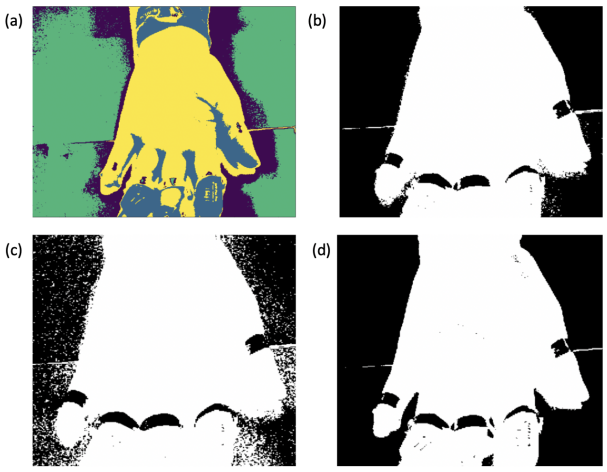

Fig. 1. Skin detection algorithms applied on subject's hand: (a) k-means clustering, (b) Decision Tree, (c) Support Vector Machines, (d) Logistic Regression

where w are the slope regression coefficients [16].

Machine learning algorithms are trained with the binary classification training set defined in the previous section and, for each video frame, skin segmentation is performed as demonstrated in Fig.1 by applying the trained algorithms.

In order to extract the vPPG signal, we calculated the mean values of the R,G,B channels in each skin region (S_k) and for each frame k.

$$M_k(c) = \sum_{i=1}^{H}\sum_{j=1}^{W} P_{ij}(c)\frac{1}{N_k}, \forall (i,j) \in S_k \tag{5}$$

where k is the number of video-frames, c is the color channel (Red, Green, Blue), H is the height and W is the width of the frames, N_k is the number of skin-pixel in each frame k and P is the matrix $HxWx3$ of the frames, resulting in a $3xk$ matrix M, containing the vPPG signal.

D. Video Signal Processing

The steps to compare vPPG with fPPG use a non-linear trend removal and filtering procedure with a pass band (0.3-4Hz) Hamming filter as already presented in [2].

So far, the extracted signals appear very similar to the plethysmographic one with all the segmentation methods and frequently details as the dicrotic notch appears more evident, in line with the shape expected on the original blood pressure waveform, as illustrated in Fig.2. Of note, the proposed steps were able to maintain the signal amplitude, a quite relevant feature for monitoring.

E. Feature Extraction and Statistical Analysis

The Automatic Multiscale-based Peak Detection (AMPD) method [6], was used to extract the systolic peak locations as surrogate of ECG R-peaks both in vPPGs and fPPG.

The comparison between different skin segmentation methods was performed by extracting HRV-related features from both fPPG and vPPG, during *Rest* and *Resp*.

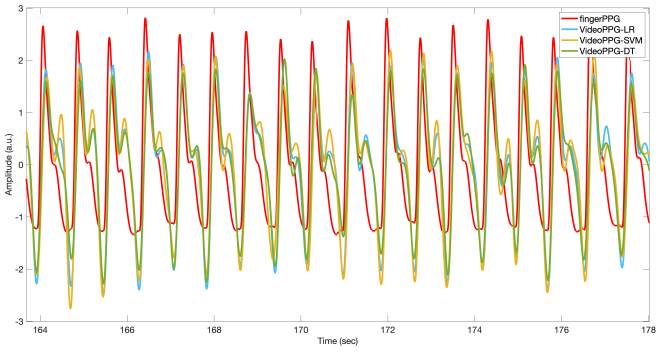

Fig. 2. Comparison of BVP signal extracted with fPPG and vPPG using different classification algorithms : Logistic Regression (blue), Support Vector Machines (yellow) and Decision Tree (green)

For each feature i an error measure was defined as the difference between the *i-th* feature extracted from the reference (fPPG) and the same feature ($x_k(i)$) computed with the *k-th* algorithm and then normalized by the reference itself.

$$e_k(i) = \frac{fPPG(i) - x_k(i)}{fPPG(i)} \tag{6}$$

For statistical analysis, we considered the mean of *PP* interval (μ_{PP}), *pNN50* and *SDNN* parameters, as important time domain measures.

In the frequency domain, we analyzed the Power Spectral Density (PSD) of the *PP* intervals in the 2 phases. The PSD is calculated in three range of frequencies:

- Very Low Frequencies (VLF): from 0.0033 to 0.04Hz
- Low Frequencies (LF): from 0.04 to 0.15Hz
- High Frequencies (HF): from 0.15 to 0.40Hz

IV. Results

The extracted time and frequency domain features were averaged among all subjects. Each feature computed from the different vPPG signals is compared with the one computed on the fPPG by Mann-Whitney U test for each stimulus condition.

The obtained averages and mean absolute deviations (MAD) for the chosen features are shown in table I. Of note, no significant statistical difference between time domain features can be observed. Conversely, the SVM segmentation algorithm shows a significant statistical difference ($p < 0.05$) on the *HF* and *HFn* features only under *Resp* condition.

In accordance with fPPG recordings, all the algorithms show a decrease of the average PP interval, whereas *SDNN* and *pNN50* show an increase for all the subjects during the *Resp* phase with respect to *Rest*. In addition, *HF*, *LF*, *LFn*, *VLF* and *LF/HF*, decrease under *Resp* conditions and only *HFn* shows an increase during the paced breathing phase.

Finally, SSEs, summarized in table II, show a smaller SSE for the Decision Tree classifier against the others.

V. Discussion and Conclusions

We have presented an automatic procedure for skin segmentation that successfully recognizes skin pixels from the

TABLE I

TIME- AND FREQUENCY-DOMAIN FEATURES FOR FINGER-PPG (FPPG) AND VIDEOPPG (V-K) EXTRACTED WITH DECISION TREE (K=DT), SUPPORT VECTOR MACHINE (K=SVM) AND LOGISTIC REGRESSION (K=LR) EXPRESSED AS $Mean(x) \pm MAD(x)$. STATISTICAL SIGNIFICANCE IS DEFINED AS FOLLOWS: $^{*} = p < 0.05$, $^{**} = p < 0.001$ AND $^{***} = p < 0.0001$

Feature	Session	fPPG	v-DT	v-SVM	v-LR
μ_{PP}	Rest	0.755 ±0.050	0.760 ±0.049	0.761 ±0.048	0.758 ±0.050
	Resp	0.707 ±0.039	0.723 ±0.045	0.723 ±0.039	0.716 ±0.050
SDNN	Rest	0.072 ±0.008	0.077 ±0.009	0.079 ±0.007	0.085 ±0.014
	Resp	0.085 ±0.015	0.081 ±0.016	0.089 ±6.535	0.089 ±0.017
*pNN*50	Rest	0.200 ±0.102	0.376 ±0.066	0.403 ±0.049	0.450 ±0.124
	Resp	0.312 ±0.122	0.493 ±0.147	0.563 ±0.154	0.505 ±0.109
HF	Rest	119.031 ±31.832	124.290 ±18.082	114.038 ±14.920	127.293 ±7.025
	Resp	91.921 ±11.042	70.480 ±16.850	58.215 ±12.121*	69.587 ±7.364
HFn	Rest	0.324 ±0.058	0.353 ±0.070	0.325 ±0.057	0.361 ±0.040
	Resp	0.551 ±0.031	0.440 ±0.071	0.383 ±0.066*	0.437 ±0.0458
LF	Rest	190.081 ±35.263	156.829 ±25.744	146.299 ±16.887	158.869 ±29.330
	Resp	44.995 ±15.697	49.563 ±13.429	50.930 ±8.951	58.258 ±5.776
LFn	Rest	0.538 ±0.115	0.412 ±0.067	0.437 ±0.070	0.397 ±0.020
	Resp	0.278 ±0.116	0.300 ±0.061	0.280 ±0.036	0.341 ±0.025
VLF	Rest	111.538 ±24.392	74.543 ±27.260	73.235 ±29.524	81.787 ±16.517
	Resp	14.830 ±6.535	12.618 ±6.210	10.453 ±2.431	10.711 ±5.936
LF/*HF*	Rest	1.800 ±0.546	1.327 ±0.383	1.423 ±0.204	1.171 ±0.2650
	Resp	0.521 ±0.234	0.710 ±0.171	0.899 ±0.204	0.867 ±0.163

first frame of the video recording, allowing the training of a model that can be easily applied to the successive frames, thus reducing the overall computational cost of running an unsupervised method for each frame. HRV measures extracted from the resulting vPPG are not different from the ones extracted with fPPG, suggesting that the automatic procedure for vPPG extraction can be used to quantify HRV automatically and in a contactless way. The comparison between three different machine learning methods for hand segmentation and vPPG extraction shows that a DT classifier performs better when computing HRV measures than the others. Finally, the measured autonomic changes suggest that paced breathing shifts the autonomic activity toward Vagal tone and increases baroreflex sensitivity despite a decrease in the average PP intervals. The observed increase of *HFn* and a decrease in *LFn* and *LF/HF* are in agreement

TABLE II

SUM OF SQUARED ERRORS (e_k) FOR THE PROPOSED ALGORITHMS IN *Rest* AND *Resp* CONDITIONS

Algorithm	Rest	Resp
v-DT	1.533	1.046
v-LR	2.198	1.370
v-SVM	1.658	1.362

with previous findings on contact-derived HRV indexes [17]. Further studies with an increased number of subjects needs to be performed in order to validate this preliminary study.

REFERENCES

[1] G. Valenza, L. Iozzia, L. Cerina et al., "Assessment of Istantaneous cardiovascular dynamics from video plethysmography", in *Proceedings of IEEE-EMBC*, 2016.

[2] R. Barbieri, L. Ficarelli, R. Levi et al., "Identification and Tracking of Physiological Parameters from Skin using Video Photoplethysmography" in *Proceedings of IEEE-EMBC*, 2019.

[3] Heart Rate Variability, "Standards of measurement, physiological interpretation, and clinical use: Task force of the European society of cardiology and the North American society for pacing and electrophysiology " *Circulation*, vol. 93, no. 5, pp. 1043-1065, 1996.

[4] L. Iozzia, L. Cerina, L. Mainardi, "Relationships between heart-rate variability and pulse-rate variability obtained from video-PPG signal using ZCA", *Physiological Measurement*, vol. 37, no. 11, pp. 1934-1944, 2016.

[5] P. Viola, M. Jones, "Robust real-time face detection", *International Journal of Computer Vision*, 57(2), pp. 137-154, 2004.

[6] F. Scholkmann, J. Boss, M. Wolf, "An efficient algorithm for automatic peak detection in noisy periodic and quasi-periodic signals", *Algorithms*, vol. 5, no. 4, pp. 588-603, 2012.

[7] W. Verkruysse, L. O. Svaasand, J. S. Nelson, "Remote plethysmographic imaging using ambient ligth", vol. 16, no. 26, pp. 72-75, 2008.

[8] A.V. Moco, S. Stuijk , G. De Haan, "Motion robust PPG-imaging through color channel mapping". *Biomed Opt Express.*,vol 7, pp.17371754, 2016.

[9] P.S. Addison, D. Jacquel et al., "Video-based heart rate monitoring across a range of skin pigmentations during an acute hypoxic challenge", *Journal of Clinical Monitoring and Computing*, vol.32, pp. 871-880, 2018.

[10] S. Kolkur, D. Kalbande, et al., "Human Skin Detection Using RGB, HSV and YCbCr Color Models", *Proceedings of ICCASP*, vol. 137, pp. 324-332, 2016.

[11] M.M. Oghaz, M.A. Maarof et al., "A Hybrid Color Space for Skin Detection Using Genetic Algorithm Heuristic Search and Principal Component Analysis Technique", *PLOS ONE*, vol. 10, 2015.

[12] T.B. Fitzpatrick, Sun and skin, *Journal de Medecine Esthetique*, pp. 33-34, 1975.

[13] A. David, S. Vassilvitskii, "k-means++: The advantages of careful seeding", *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics*, 2007.

[14] L. Breiman, J. Friedman, R. Olshen, C. Stone, *Classification and Regression Trees*, 1st ed., Taylor Francis, 1984.

[15] J. C. Platt, "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods", *Advances in Large Margin Classifiers*, pp.61-74, 1999.

[16] H. Yu, F. Huang, C. Lin, "Dual coordinate descent methods for logistic regression and maximum entropy models", *Machine Learning*, vol.85, pp.41-75, 2011.

[17] L. Changjun, C. Qinghua et al., "Effects of slow breathing rate on heart rate variability and arterial baroreflex sensitivity in essential hypertension", *Medicine (Baltimore)*, vol. 97, 2018.