IEEE GlobalSIP
Fourth IEEE Global Conference on Signal and Information Processing
December 7-9, 2016, Greater Washington D.C., USA
http://2016.ieeeglobalsip.org/

IMPROVED REMOTE ESTIMATION OF HEART RATE IN FACE VIDEOS

Alain Malacarne, Mattia Bonomi, Cecilia Pasquini, Giulia Boato

Multimedia Signal Processing and Understanding Lab, DISI, University of Trento

UNIVERSITY OF TRENTO - Italy

mm lab

ABSTRACT

Measuring the Heart Rate (HR) plays an important role in the description of human physiological and psychological state, due to its relationship with cognitive/emotional factors such as attention effort, stress or arousal. For this reason, **remote methodologies for HR measurements have recently been investigated to find a reliable and cost-effective methodology.** Our work aims at the following:

- Development of a **novel technique** for remote HR estimation
- Comparison of the proposed method with the state of the art on a **common dataset**

INTRODUCTION

Traditional methodologies for HR estimation involved the use of physical sensors, although recent studies focused on the use of common inexpensive cameras.

Three main issues affect the SoA approaches:

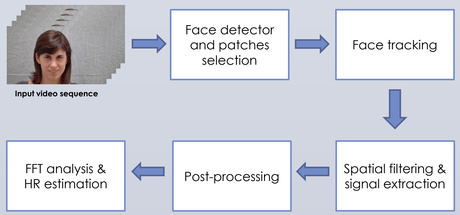
- **Rigid motion**, due to head movements
- **Illumination noise**, caused by light
- **Non-rigid motion**, due to facial expressions

For this reason, we propose an improved technique for remote HR estimation, which deals with all the common issues of the SoA methods.

The main contributions of this work are:

1. Selection of **relevant video patches**, tracked in two different modalities and combined to obtain a time signal
2. Improved **post-processing procedure** for a more accurate HR estimation
3. Creation of a new **validation dataset** and comparison with peculiar SoA techniques

PROPOSED APPROACH



1. FACE DETECTOR AND PATCHES SELECTION

Viola-Jones Face detector on first frame of the sequence

- FACE patch contains **pulse information**
- BG1/BG2 patches to gather **illumination noise**
- MZ patch for **position changes** with respect to the light source

2. FACE TRACKING

Two different tracking modalities:

Shifting mode

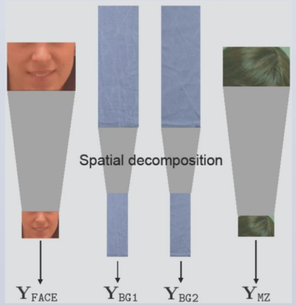
- Tracks only **translation** movements throughout the video sequence

Warping mode

- Also accounts **rotations** and recovers the original head position of first frame

3. SPATIAL FILTERING AND SIGNAL EXTRACTION

- Conversion from RGB to **YCbCR** color space
- **Spatial decomposition** of luminance component at each frame
- Y mean value for every time instant to get patch signal



4. POST-PROCESSING

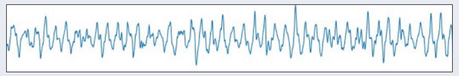
4.1 SIGNAL COMPOSITION

- Combination of the patch signals to **remove illumination and position noise** from pulse information

$$Y = Y_{FACE} - Y_{MZ} - \frac{Y_{BG1} + Y_{BG2}}{2}$$

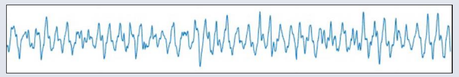
4.2 DETRENDING FILTER

- Filter out **slow and non-stationary** signal components



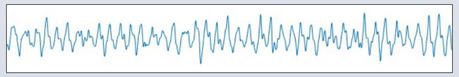
4.3 NON-RIGID MOTION ELIMINATION

- Removes **noise** due to changes in **facial expression**
- Divides the signal in equal length segments and removes with highest variance



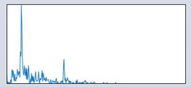
4.4 AVERAGE FILTERING

- Attenuates **small noisy oscillations** in the signal



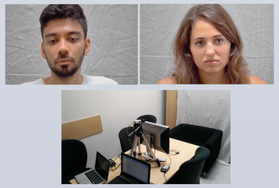
5. FFT ANALYSIS and HR ESTIMATION

- Frequency range corresponding to **40-100 [bpm]**
- **Estimated HR** as frequency of the highest peak



VALIDATION DATASET CREATION

- **29 still subjects** recorded for 30 seconds
- **Point gray** camera
- **Vilistus** device to gather ground truth
- **No engagement stimuli involved**

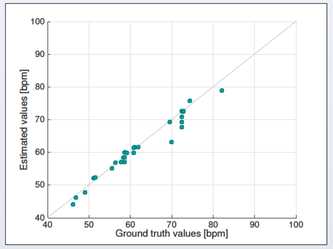


EXPERIMENTAL RESULTS

- **Three proposed variations** of the pipeline
- **Mean HR value** estimated for each sequence

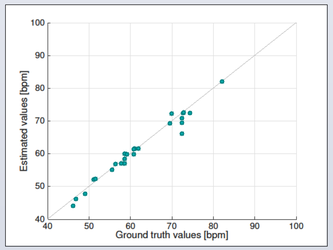
PROPOSED 1

- **Shifting tracking mode** applied to **all** the patches



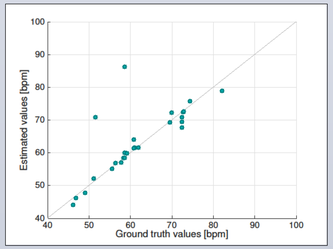
PROPOSED 2

- **Warping tracking mode** applied to **all** the patches



PROPOSED 3

- **Warping tracking mode** applied to only the **FACE** patch



COMPARISON WITH THE STATE OF THE ART

- Total of **7 compared techniques** (4 existing, 3 proposed)

METHOD	M_e	SD_e	RMSE	r
Head Motion[2]	9.71	10.4	14.10	-0.08
Eulerian Magnification[3]	-4.32	12.10	12.69	0.43
I.C.A[4]	3.64	11.19	10.77	0.39
Background Rectification[5]	**0.10**	3.76	3.69	0.92
Proposed 1	1.4	6.68	6.71	0.77
Proposed 2	-0.60	**1.64**	**1.71**	**0.98**
Proposed 3	-0.77	1.98	2.02	**0.98**

- **Proposed 2** and **Proposed 3** outperform the SoA in terms of linear correlation, while **Proposed 2** also exhibits the least values for SD_e and RMSE
- Both techniques employ the **warping tracking** mode, which is accounts more information, although with a heavier computational load

CONCLUSIONS

We have proposed a novel pipeline for estimating the heart rate of a human subject in face videos and results on a new self-collected validation dataset show the benefits of our approach with respect to existing techniques, in particular when the warping tracking mode is employed on the facial patch.

Future work will be dedicated to improve the proposed pipeline, aiming at a reduction of the computational complexity and further improving the estimation accuracy.

REFERENCES

[1] G. Balakrishnan, F. Durand, J. Guttag, "Detecting pulse from head motions in video", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013.

[2] H. Wu, M. Rubinstein, E. Shih, J.V. Guttag, F. Durand, W.T. Freeman, "Eulerian video magnification for revealing subtle changes in the world", ACM Transactions on Graphics, 2012.

[3] M. Poh, D.J. McDuff, R.W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam", IEEE Transactions on Biomedical Engineering, 2011.

[4] X. Li, J. Chen, G. Zhao, M. Pietikainen, "Remote heart rate measurement from face videos under realistic situations", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014.