Visual Heart Rate Estimation

Radim Špetlík, Jan Čech, Jiří Matas

Center for Machine Perception (CMP)
Czech Technical University in Prague

Overview

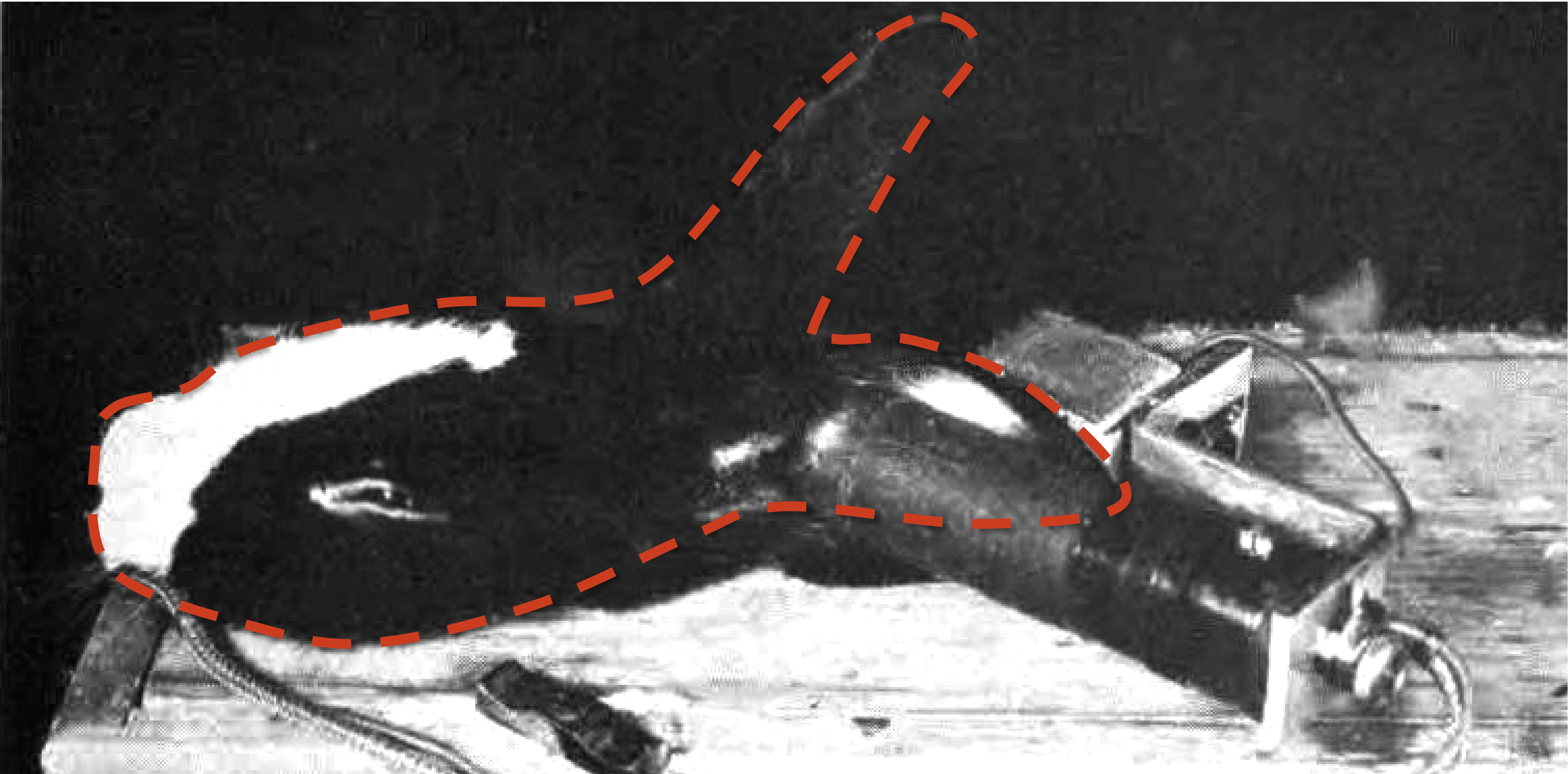
- Non-Contact Reflectance Photoplethysmography: Progress, Myths, and Failures
  - non-contact reflectance PPG (rPPG)
  - rPPG experiments
- Robust Visual Heart Rate Estimation
  - HR-CNN
    - Extractor
    - Estimator
  - Comparative study

Non-Contact Reflectance Photoplethysmography: Progress, Myths, and Failures

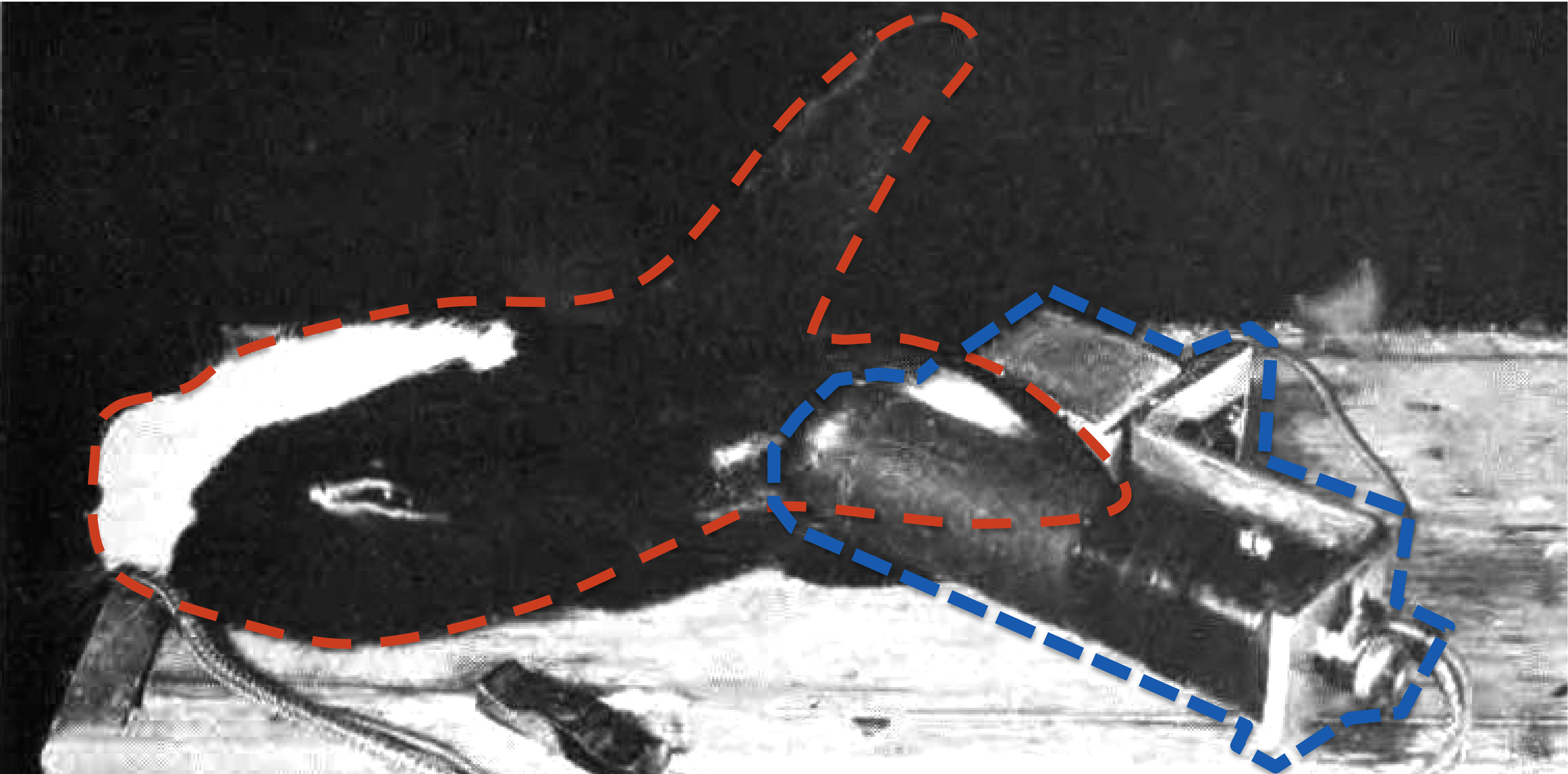
Radim Špetlík, Jan Čech, Jiří Matas

Do you see a rabbit’s head?

Do you see a rabbit's head?

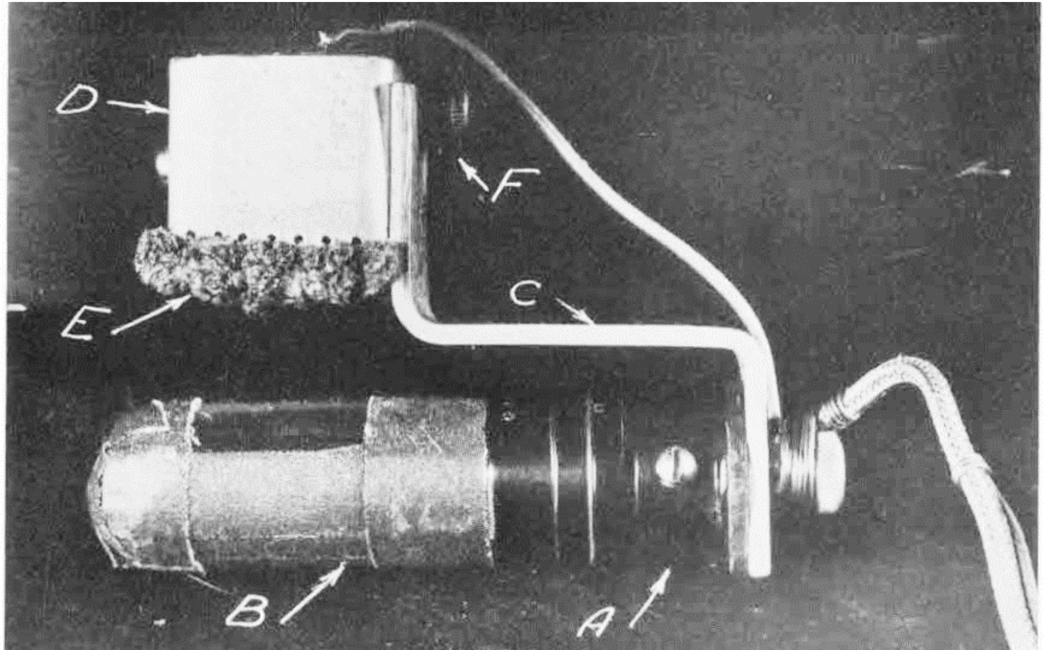


Do you see a rabbit's head?



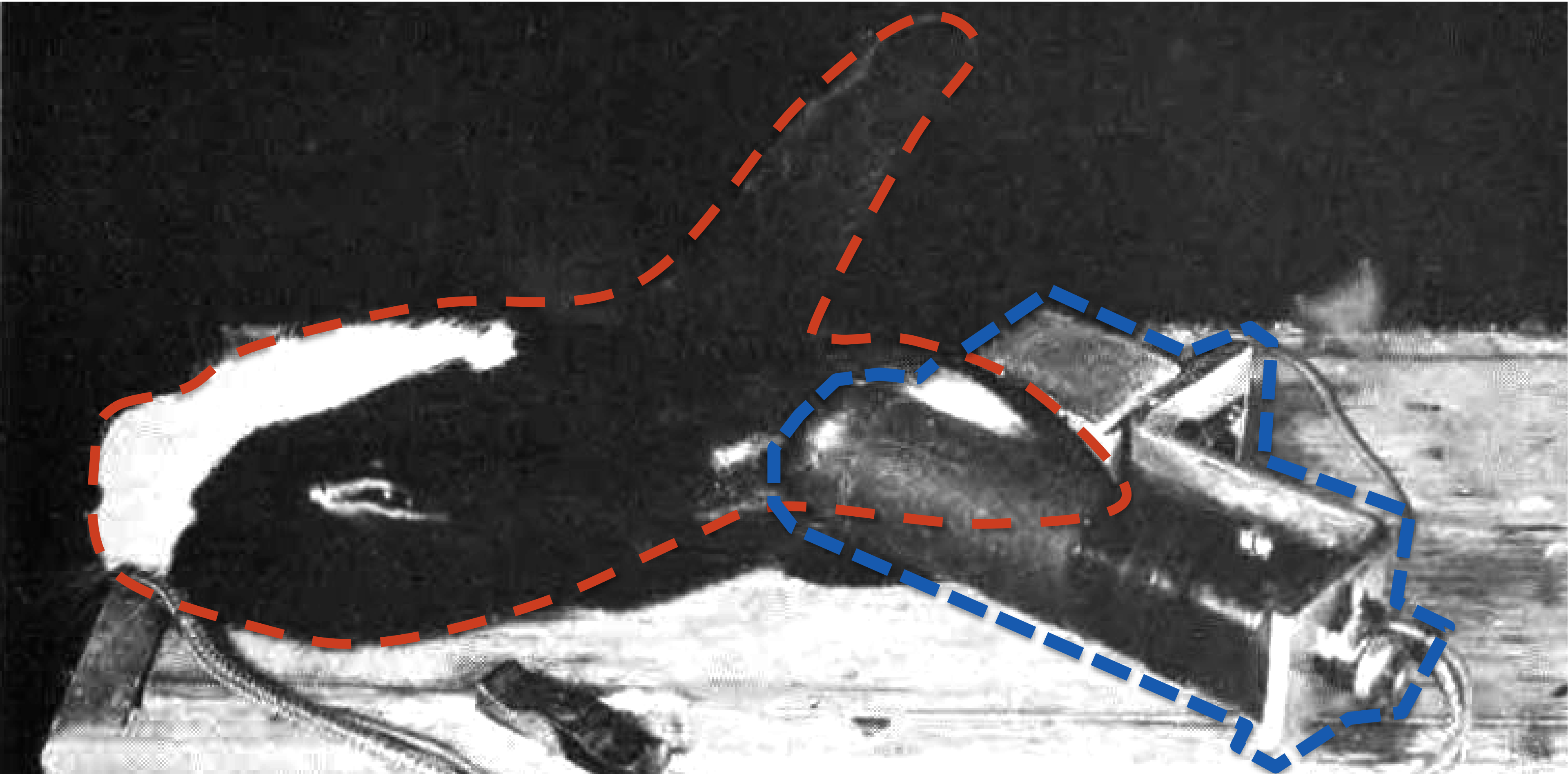
Photoplethysmography

- introduced in **1936** by **H. Molitor and M. Kniazuk**
- detects blood volume changes in the microvascular bed of tissue

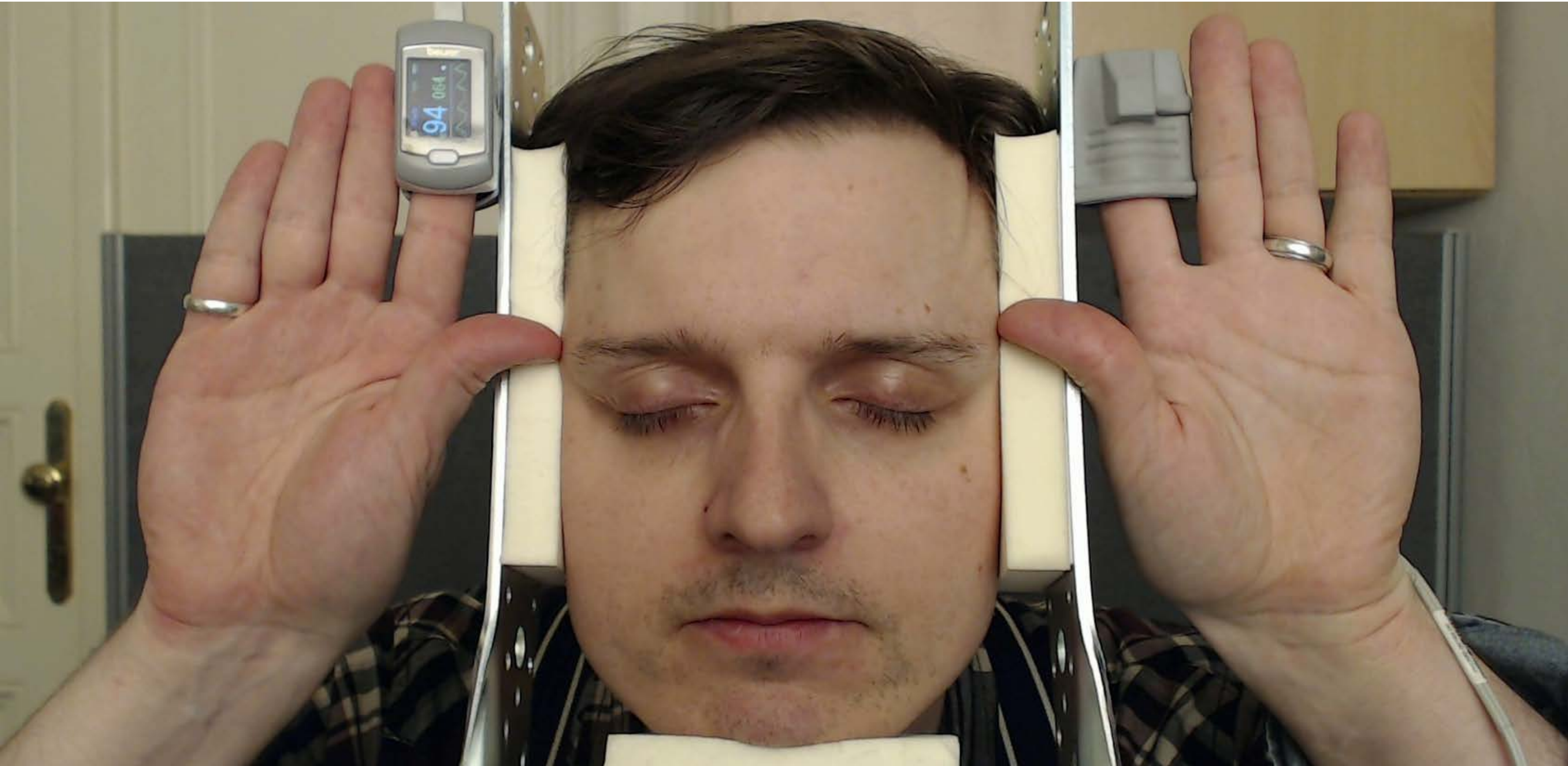


B. photocell
D. light-housing
E. sponge rubber

Do you see a rabbit’s head?



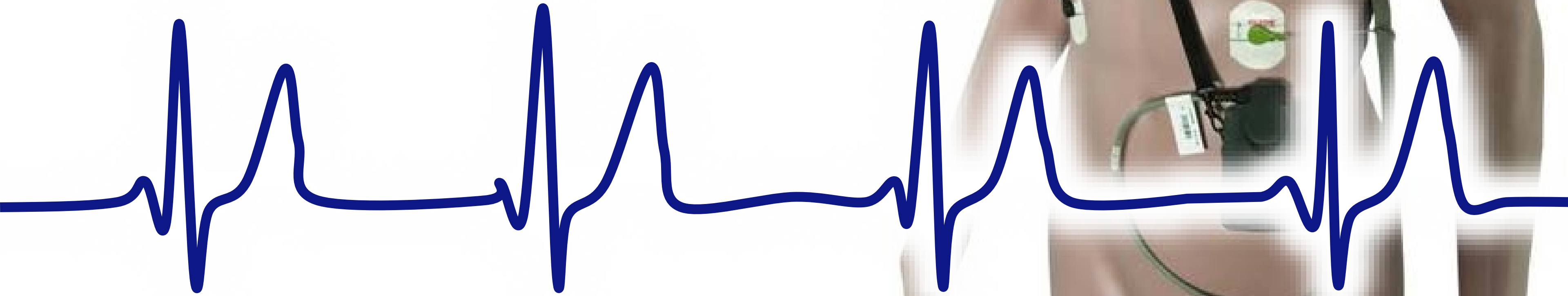
Photoplethysmography



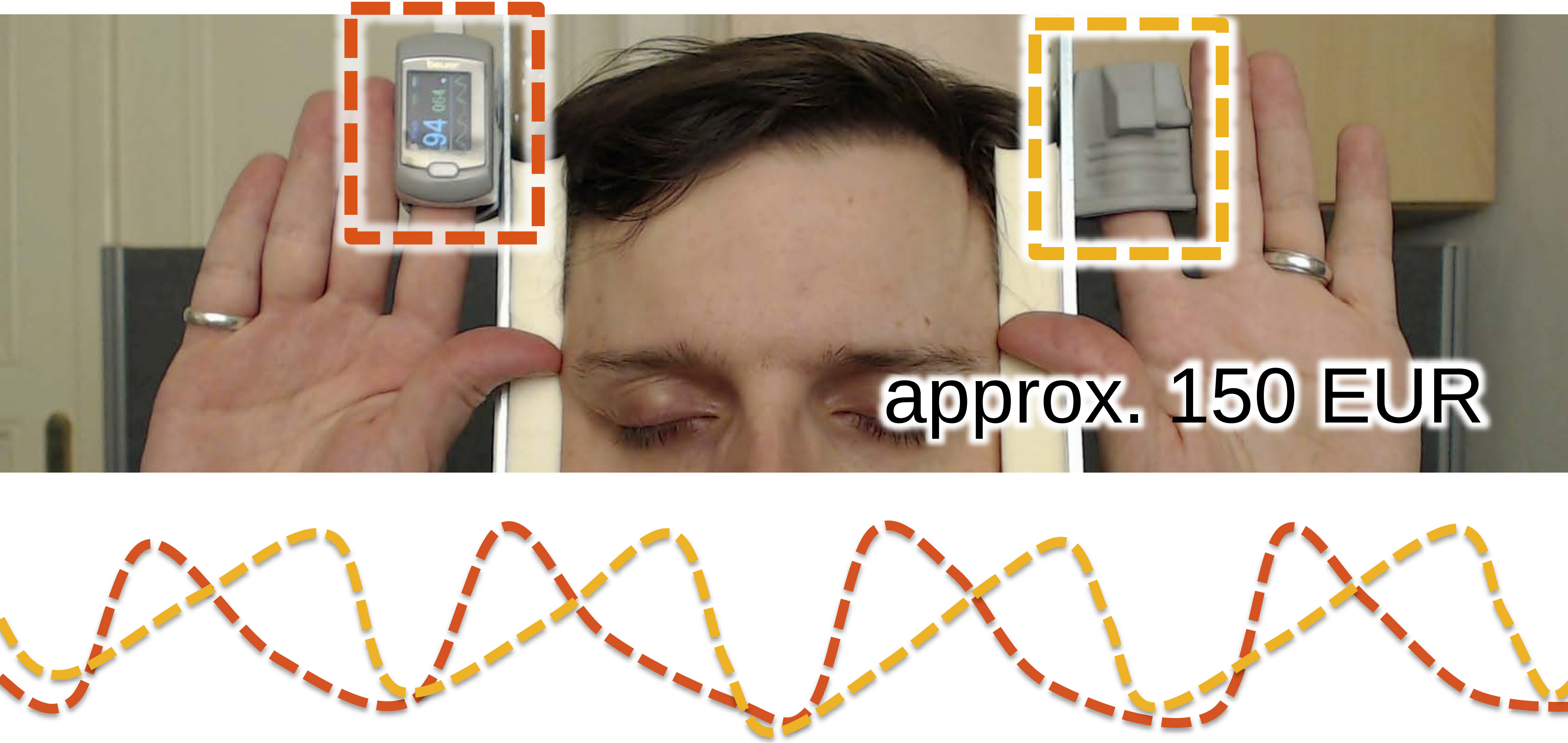
ElectroCardioGraphy

approx. 500 EUR

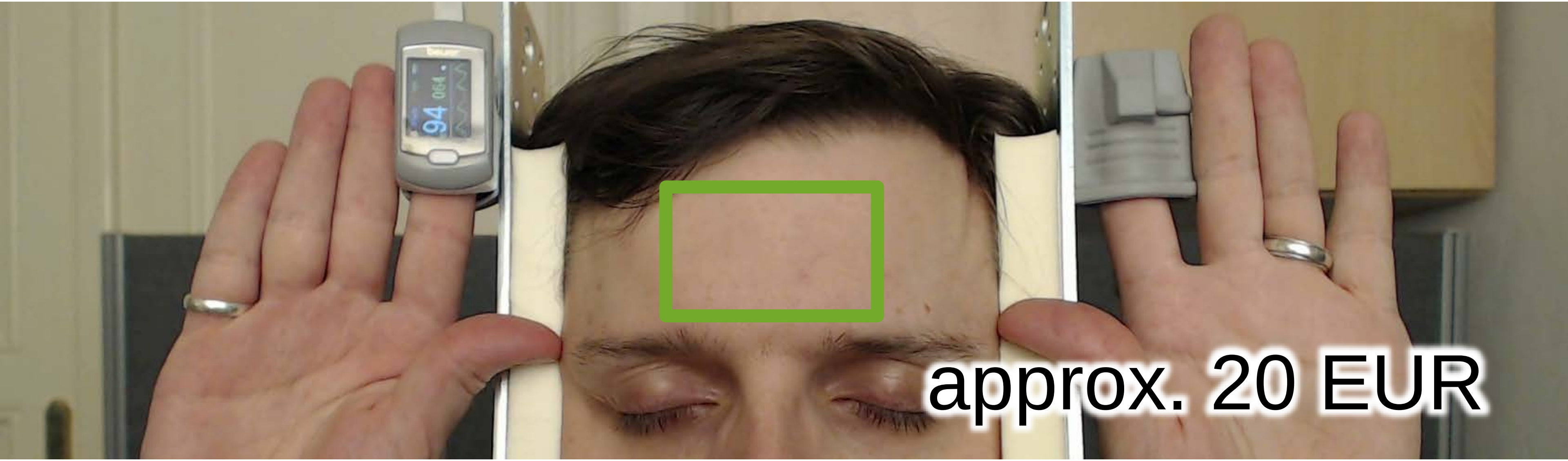
“very contact”

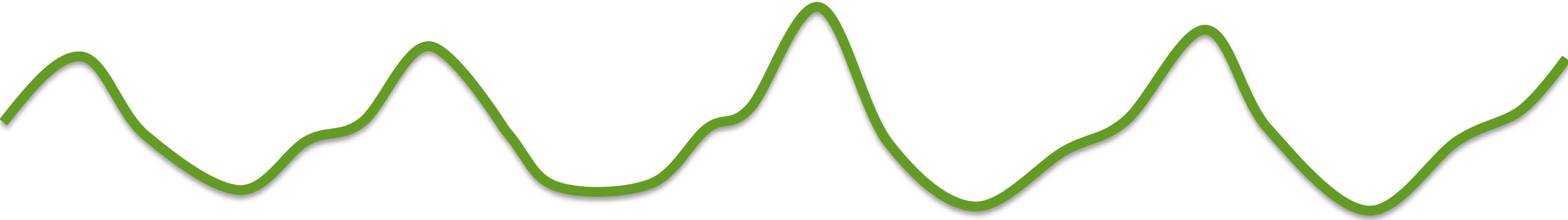


Transmittance Photoplethysmography

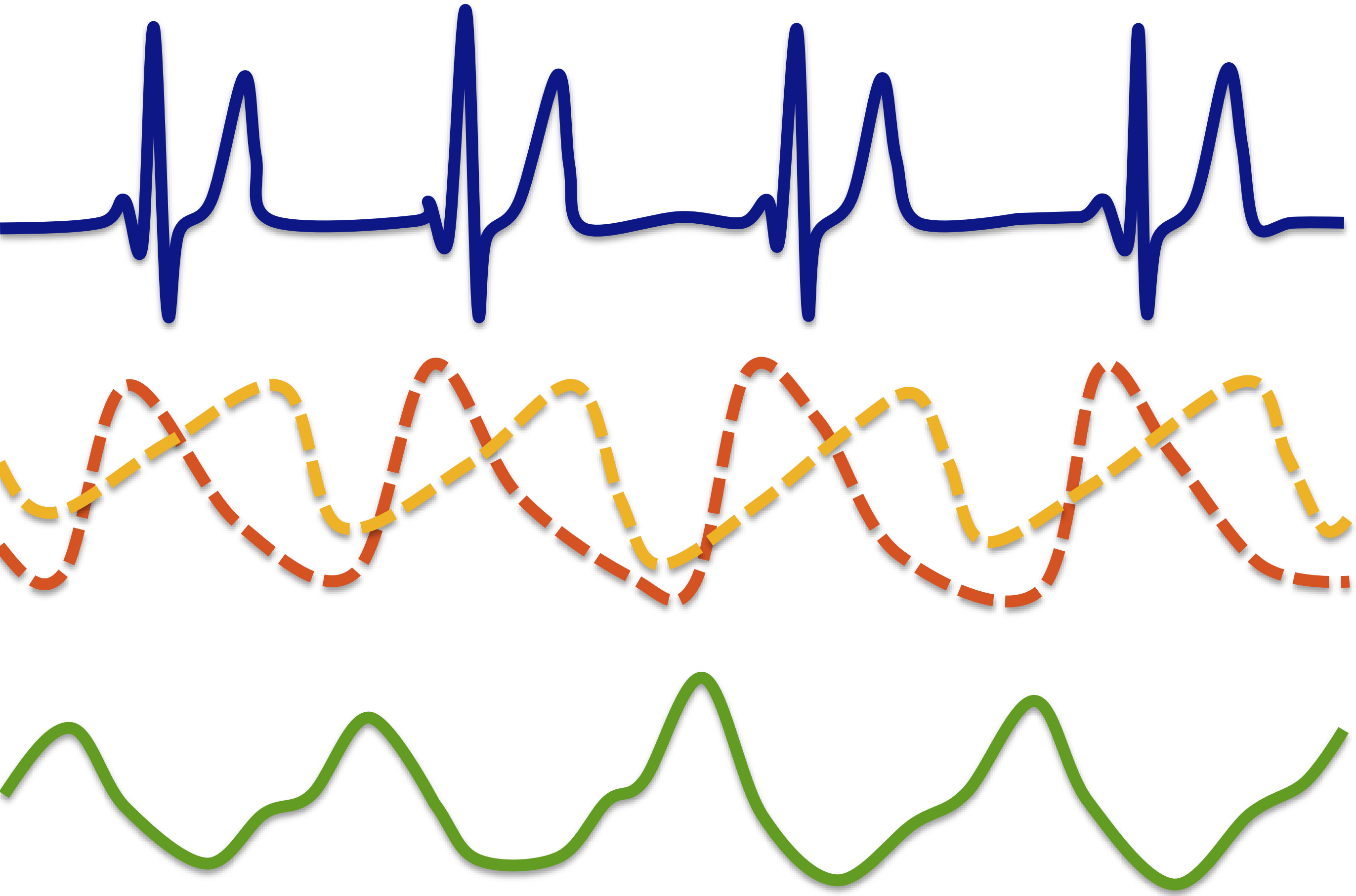


Reflectance Photoplethysmography





Heart Rate Measurements



ECG
500 EUR

tPPG
150 EUR

rPPG
20 EUR

Motivation

- **price**
- patients too fragile for contact measurements (i.e. newborns, patients with burns, etc.)
- improving comfort in fitness related HR measurements

[2013] G. Balakrishnan et al., Detecting Pulse from Head Motions in

www.forbes.com/sites/paullamkin/2017/03/03/fitbits-dominance-diminishes-but-wearable-tech-market-bigger-than-ever/

rPPG Heart Rate Measurement

very low signal-to-noise ratio

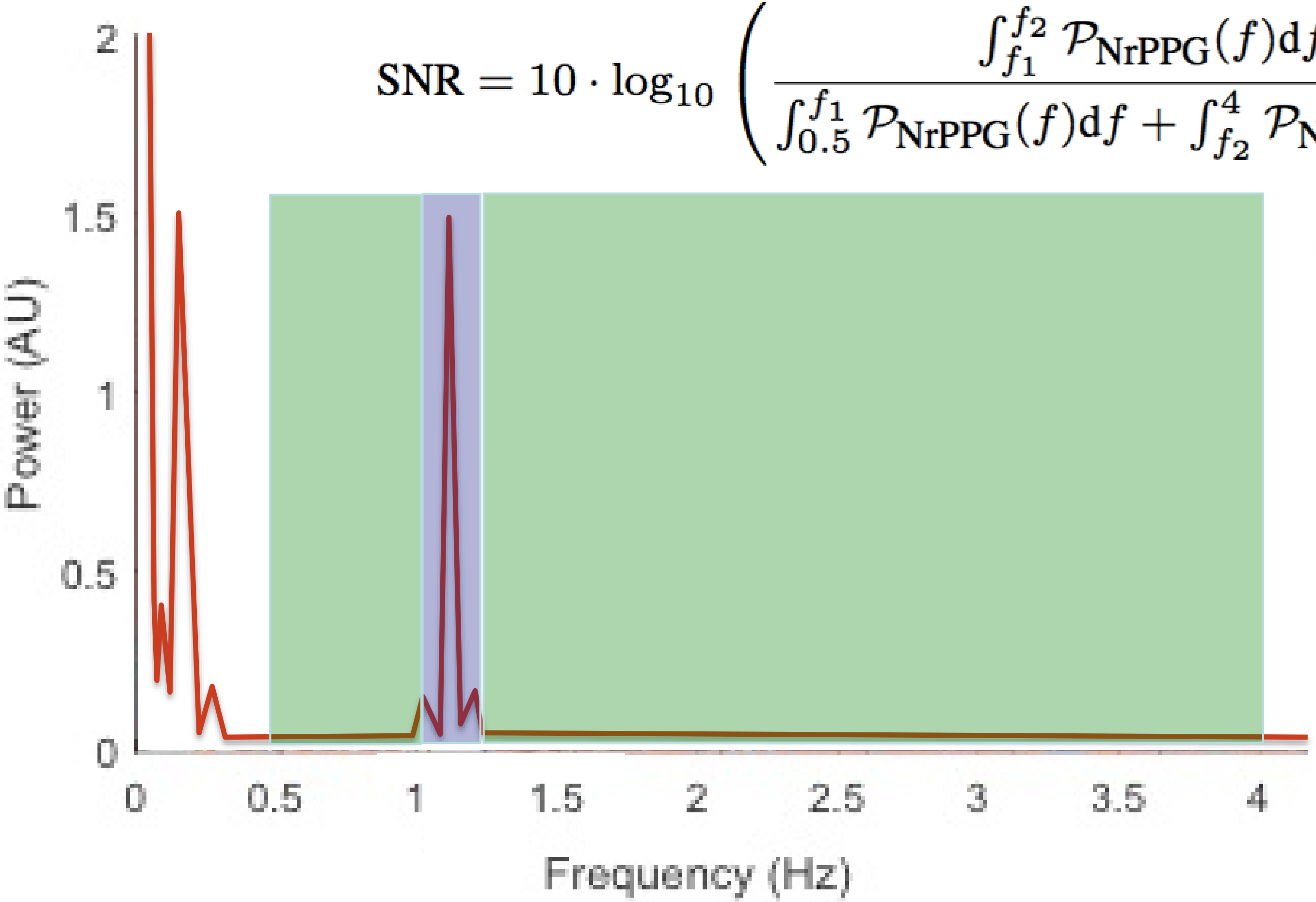
=> it is hard to perform

SNR improved by:

- careful data acquisition
- precise pixel-to-pixel face registration

rPPG – Experimental setup

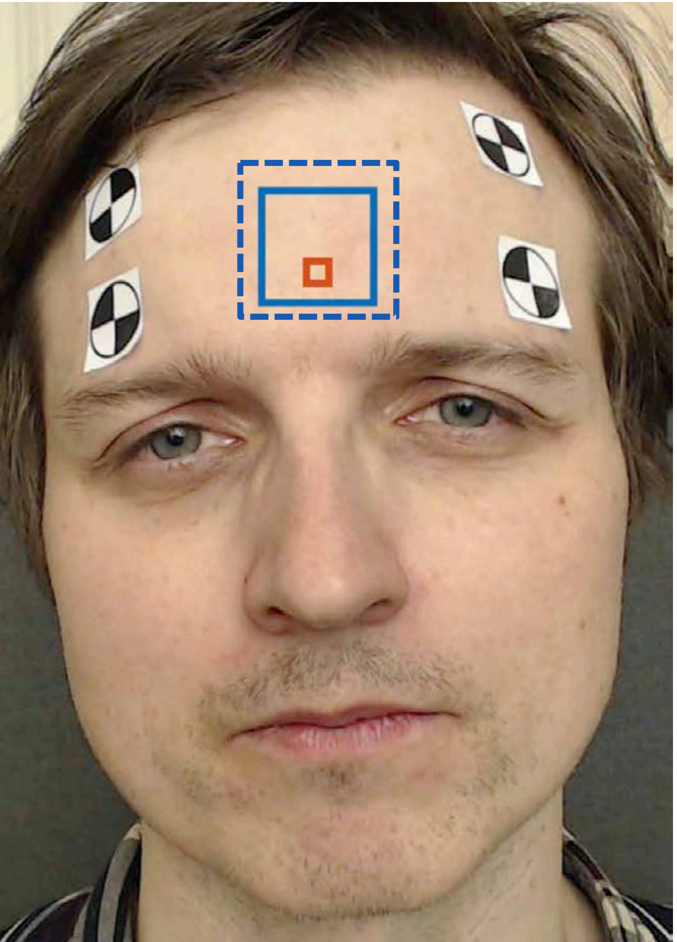
$$\mathrm{SNR} = 10 \cdot \log_{10}\left(\frac{\int_{f_1}^{f_2} \mathcal{P}_{\mathrm{NrPPG}}(f)\mathrm{d}f}{\int_{0.5}^{f_1} \mathcal{P}_{\mathrm{NrPPG}}(f)\mathrm{d}f + \int_{f_2}^{4} \mathcal{P}_{\mathrm{NrPPG}}(f)\mathrm{d}f}\right)$$



f_1 = trueHR – 0.15

f_2 = trueHR + 0.15

rPPG – Experimental setup

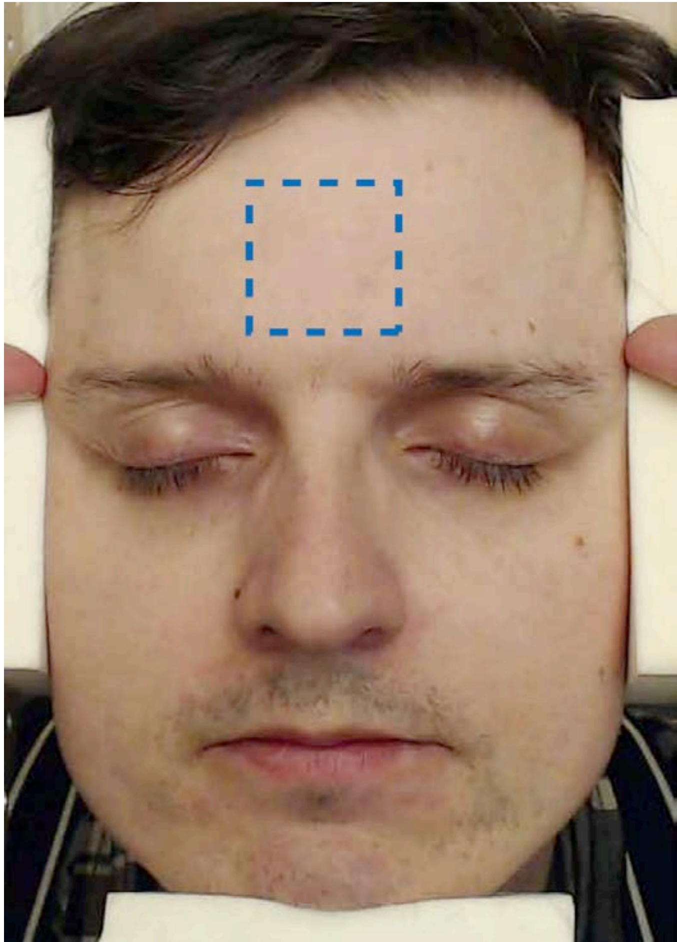


5 volunteers

100 x 100 px
region of interest

spatial
averaging in a
green channel

signal-to-noise
ratio of heart rates



rPPG – Data Acquisition

1920x1080 px

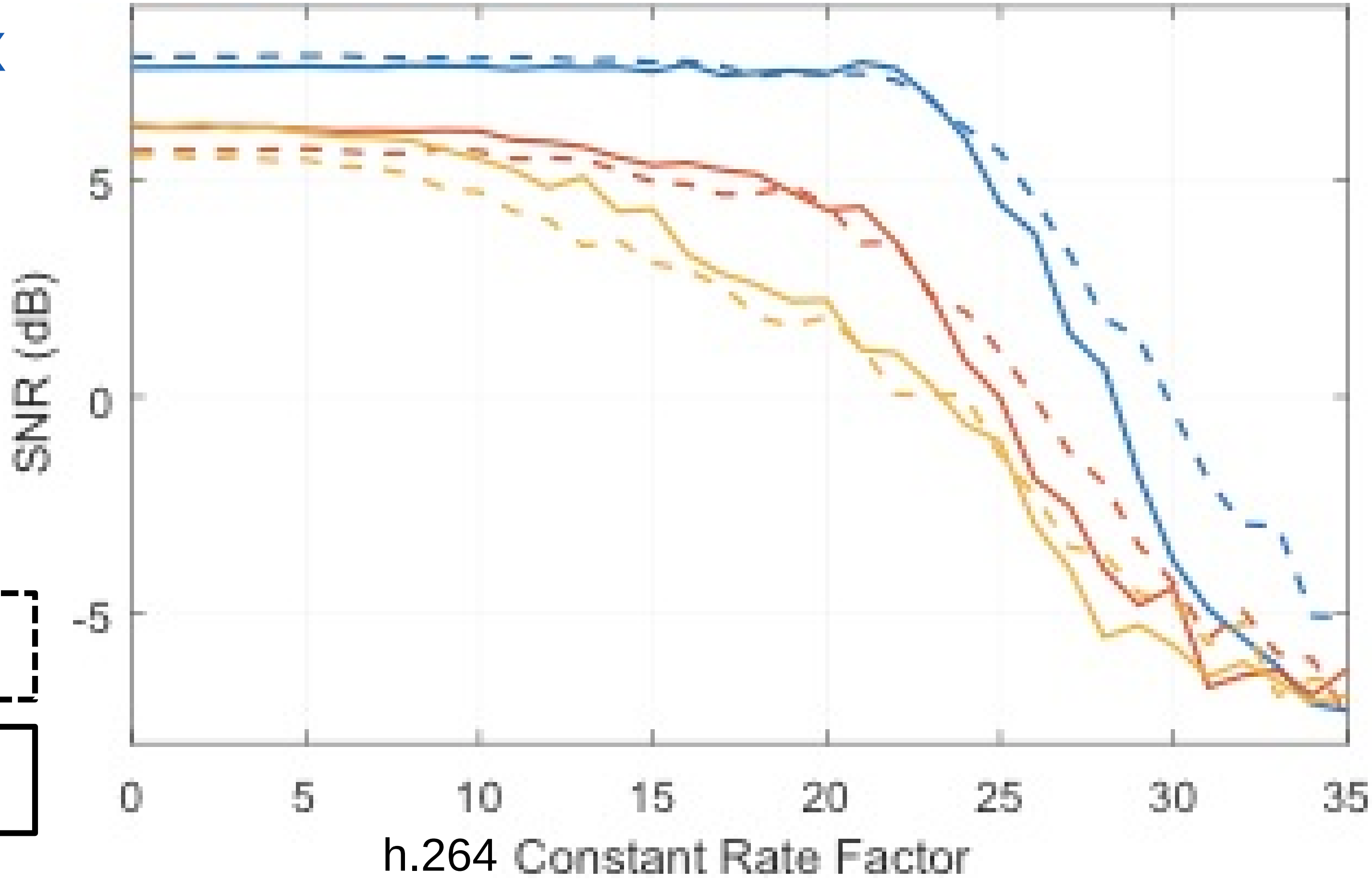
878x494 px

434x234 px

stabilization:

software

physical



rPPG – Data Acquisition

… videos **“were recorded in 24-bit RGB (with 8 bits per channel)”**.

[2014] Hsu et al. Learning-based heart rate detection from remote photoplethysmography features

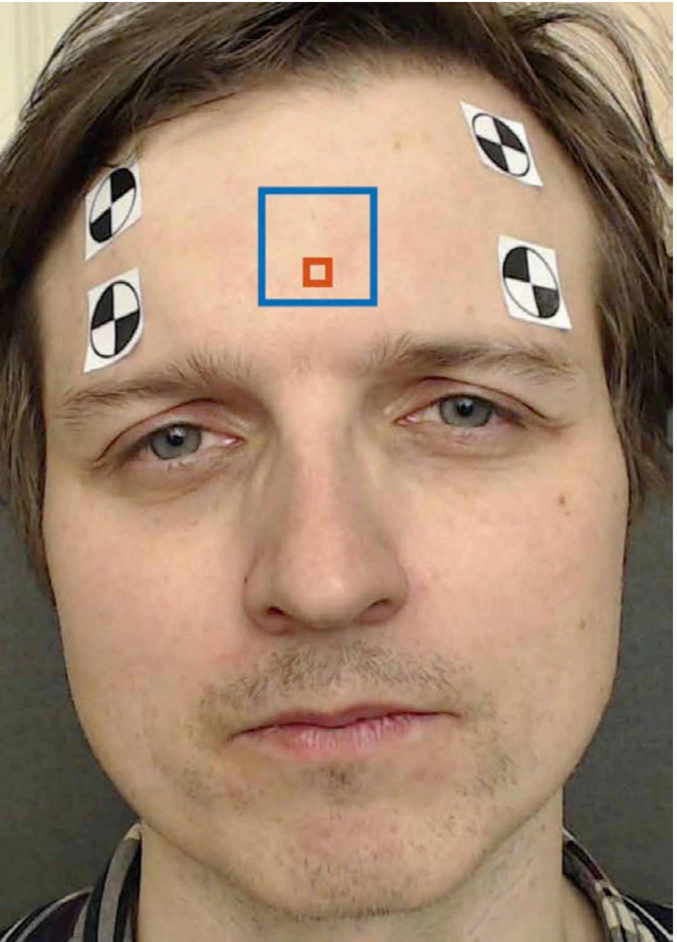
[2015] Yu, Dynamic heart rate estimation using principal component analysis

- **MPEG-4 AVC/H.264**
- bitrate **up to 24 Mbps**

- **3 x 8 bits**

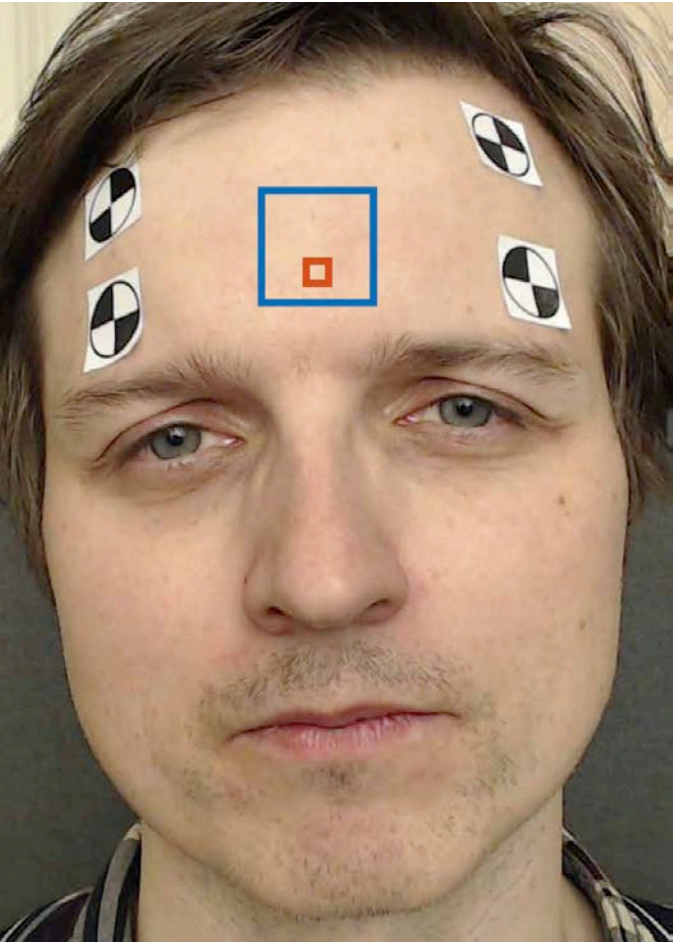
  0.64 bits

rPPG – Precise Face Registration



- four stickers in the first frame set as interest points
- tracked with MATLAB implementation of Lukas-Kanade tracker
- homography found between first (reference) frame and all consecutive
- found homographies used to register the forehead over frames
- spatial average over ROI

rPPG – Precise Face Registration



subject ID	1	2	3	4	5
		15×15 px ROI			
not registered	1.70	-6.17	-2.74	1.88	-7.08
registered	5.47	-6.03	2.6	2.99	-6.42
		75×75 px ROI			
not registered	8.88	-5.42	6.77	6.80	-6.59
registered	9.15	-5.34	7.39	7.52	-5.68

TABLE I

SIGNAL-TO-NOISE RATIO IN DECIBELS OF A NRPPG SIGNAL FOR 5 SUBJECTS. THE SIGNAL IS COMPUTED BY SPATIAL AVERAGING OVER THE GREEN CHANNEL OF REGIONS SHOWN IN FIG. 2. RESULTS BEFORE AND AFTER REGISTRATION OF THE REGIONS.

rPPG – Conclusion

Two experiments with 5 participants showed that SNR of non-contact rPPG is:

A. improved by precise pixel-to-pixel registration
B. deteriorated by the reduction of resolution

In the paper an extensive review of rPPG literature reveals key factors limiting performance and reproducibility as:

A. incomplete description of the datasets
B. heterogeneous methodology
C. absence of publicly available datasets
D. vague terminology

Robust Visual Heart Rate Estimation

Radim Špetlík, Vojtěch Franc, Jan Čech, Jiří Matas

Overview

Given a video with a human’s face

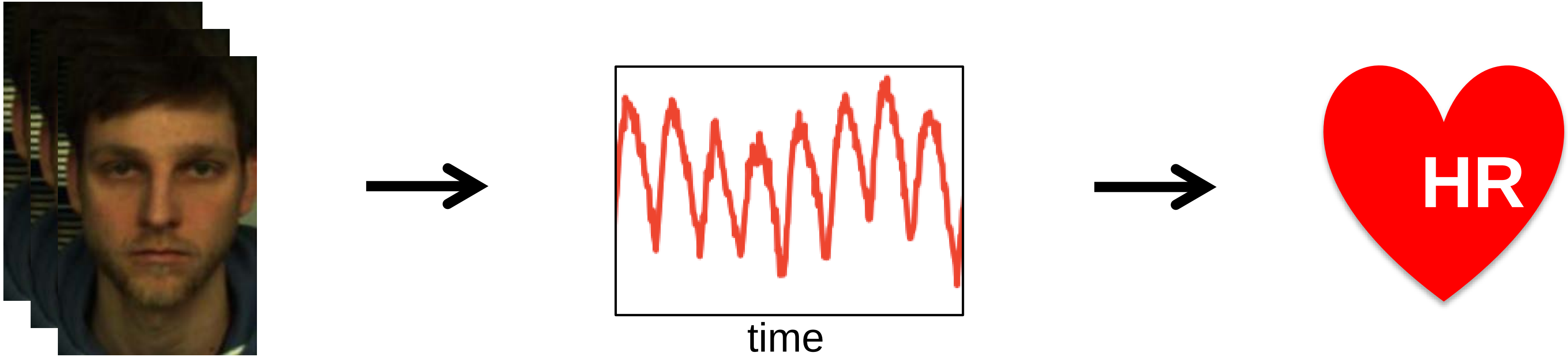
Estimate a heart rate (HR) of the person in the video

= 70 beats per minute

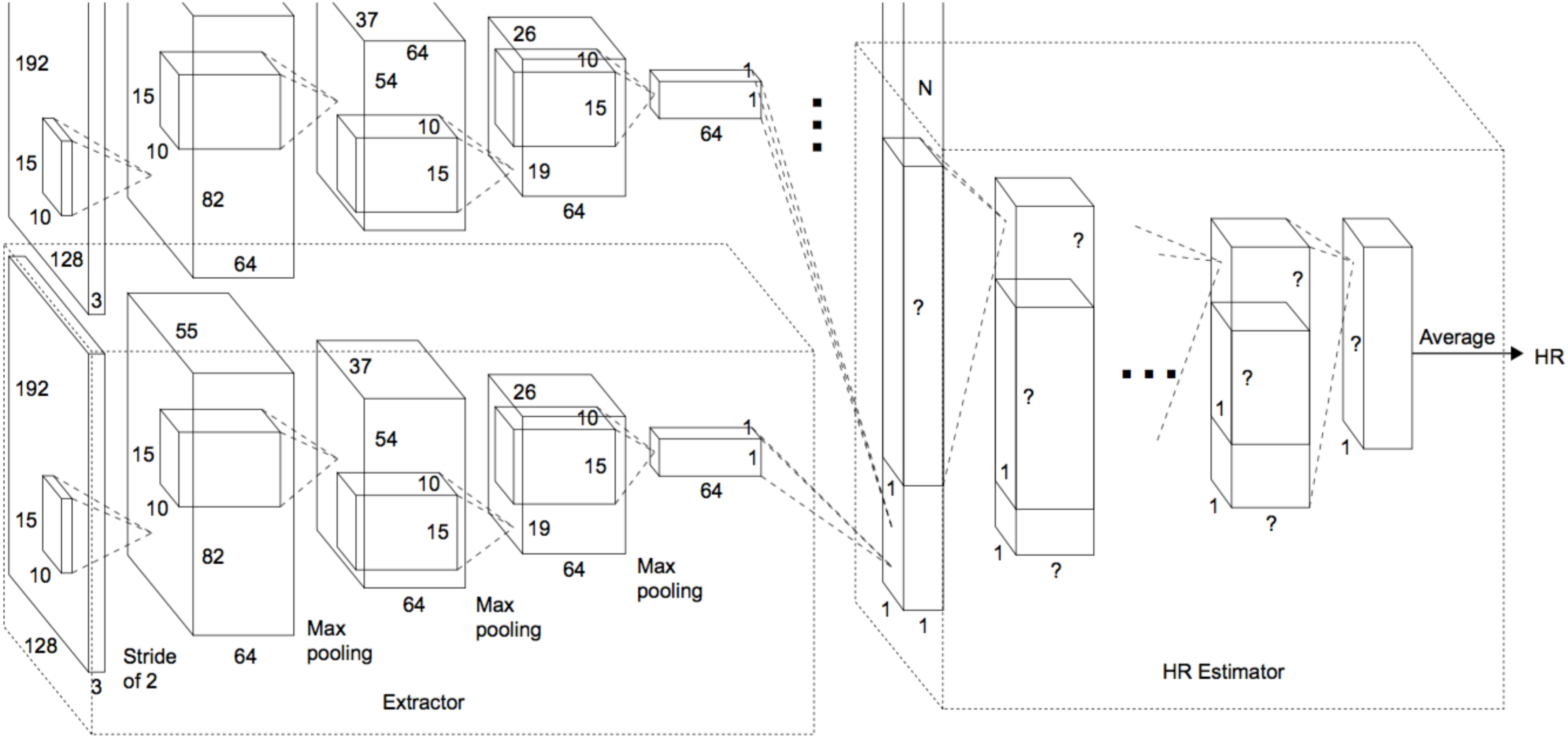
Overview

Our approach – two-step Convolutional Neural Network

1. step – extract relevant signal from face, frame by frame
2. step – given signal from the first step, estimate the heart rate



Method - HR-CNN



Method - HR-CNN

CNN:

- standard chain of convolution, MaxPool and activation blocks
- learned by alternating optimization

Extractor: maximization of SNR of relevant signal

HR Estimator: minimization of MAE

HR-CNN - Extractor

- training set $\mathcal{T} = \{(\boldsymbol{x}_j^1, \dots, \boldsymbol{x}_j^N, f_j^*) \in \mathcal{X}^N \times \mathcal{F} \mid j = 1, \dots, l\}$ with l sequences of N facial RGB image frames $\boldsymbol{x} \in \mathcal{X}$ and their labels $f^* \in \mathcal{F}$,

- given frequency f, power spectral density of signal (PSD) is

$$\mathrm{PSD}(f, \mathbf{X}; \mathbf{\Phi}) = \left(\sum_{n=0}^{N-1} h(\mathbf{x}^n; \mathbf{\Phi}) \cdot \cos\left(2\pi f \frac{n}{f_s}\right)\right)^2 + \left(\sum_{n=0}^{N-1} h(\mathbf{x}^n; \mathbf{\Phi}) \cdot \sin\left(2\pi f \frac{n}{f_s}\right)\right)^2$$

where $h(\boldsymbol{x}^n; \mathbf{\Phi})$ is output of Extractor CNN for n-th image and f_s is sampling frq, $\mathbf{\Phi}$ is concatenation of CNN parameters

- the SNR computes as

$$\mathrm{SNR}(f^*, \mathbf{X}; \mathbf{\Phi}) = 10 \cdot \log_{10}\left(\sum_{f \in \mathcal{F}+} \mathrm{PSD}(f, \mathbf{X}; \mathbf{\Phi}) \Big/ \sum_{f \in \mathcal{F} \setminus \mathcal{F}+} \mathrm{PSD}(f, \mathbf{X}; \mathbf{\Phi})\right)$$

HR-CNN - Extractor

- the SNR computes as

$$\mathrm{SNR}(f^*, \mathbf{X}; \mathbf{\Phi}) = 10 \cdot \log_{10} \left(\sum_{f \in \mathcal{F}+} \mathrm{PSD}(f, \mathbf{X}; \mathbf{\Phi}) \Big/ \sum_{f \in \mathcal{F} \setminus \mathcal{F}+} \mathrm{PSD}(f, \mathbf{X}; \mathbf{\Phi}) \right)$$

where f^* is true HR, $\mathcal{F}^+ = (f^* - \Delta, f^* + \Delta)$, and tolerance interval Δ accounts for true HR uncertainty

- CNN parameter $\mathbf{\Phi}$ is found by minimizing the loss function

$$\ell(\mathcal{T}; \mathbf{\Phi}) = -\frac{1}{l} \sum_{j=1}^{l} \mathrm{SNR}(f_j^*, \mathbf{X}_j; \mathbf{\Phi}).$$

HR-CNN - Estimator

- *Estimator* takes 1D signal – output of *Extractor* – and produces HR

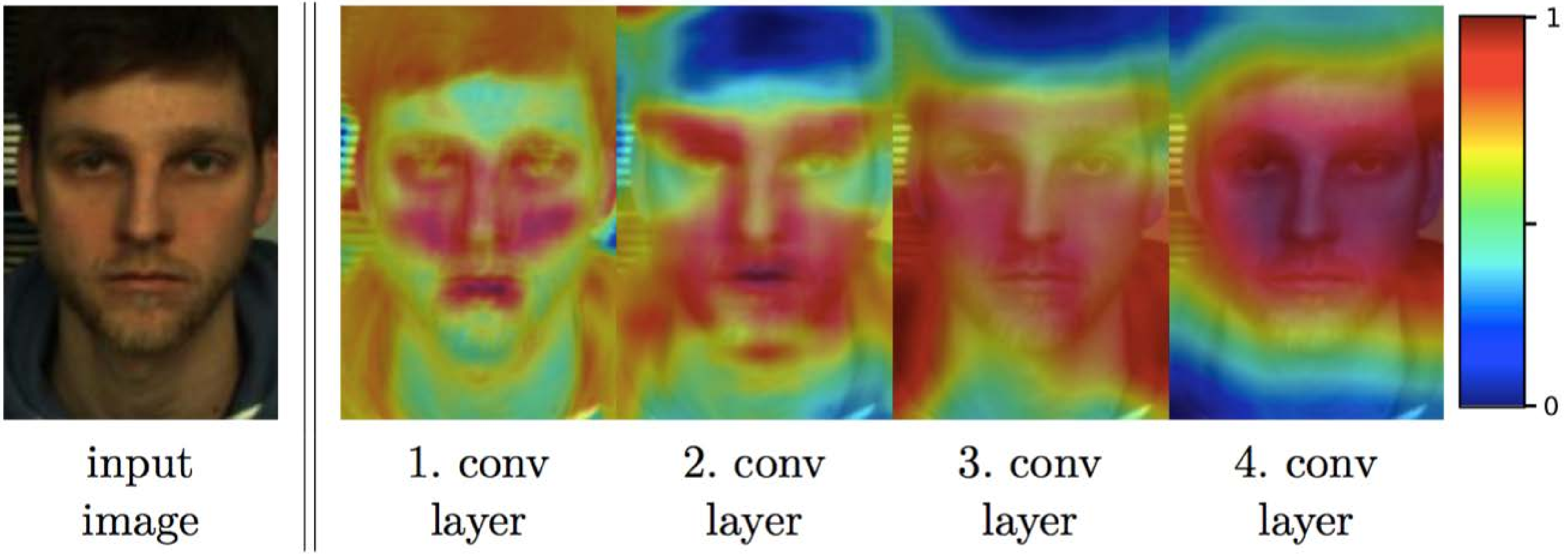
- Trained by minimization of mean absolute error of predicted and true HR f_j^*

$$\ell(\mathcal{T};\boldsymbol{\theta}) = \frac{1}{l}\sum_{j=1}^{l}\left|g\left(\left[h(\mathbf{x}^1;\boldsymbol{\Phi}),\cdots,h(\mathbf{x}^N;\boldsymbol{\Phi})\right];\boldsymbol{\theta}\right) - f_j^*\right|$$

where $g([h(\boldsymbol{x}^1;\boldsymbol{\Phi}),\ldots,\mathrm{h}(\mathbf{x}^{\mathrm{N}};\boldsymbol{\Phi})];\boldsymbol{\theta})$ is output of the CNN for a sequence of N outputs of *Extractor* and $\boldsymbol{\theta}$ is a concatenation of all parameters of *Estimator* CNN.

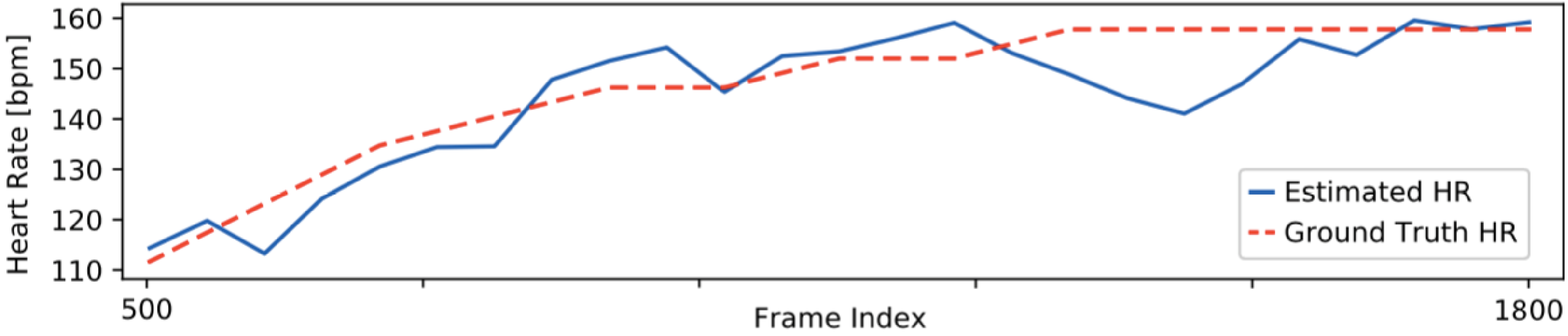
Experiments - Introspection

Extractor - GRAD-CAM



Experiments - Introspection

HR Estimator

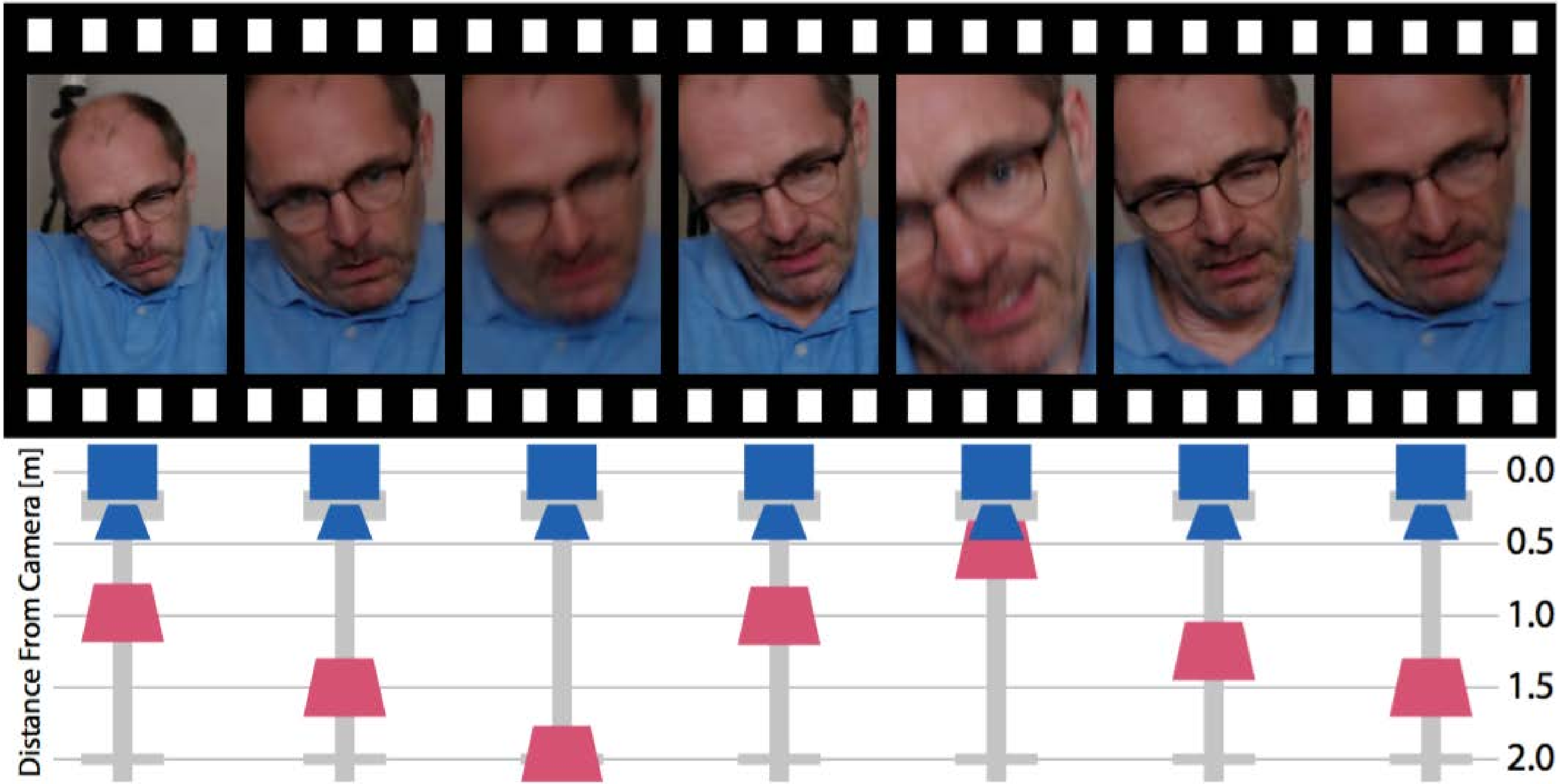


sequence with a significant HR change

Experiments - Database

- a new challenging database where subjects perform fitness activities
- 4 activities, 3 lighting setups, 2 RGB and one THERMAL camera
- in total 204 one minute RGB RAW videos, ECG ground truth

Experiments - Database



Why new database?

Experiments – Comparative Study

- comparison with 3 published methods and baseline on 4 datasets

		COHFACE	ECG-Fitness	MAHNOB	PURE	PURE MPEG-4 Visual
Pearson's corr. coeff.	baseline	—	—	—	—	—
	2SR	−0.32	0.06	0.06	0.98 ②	0.43
	CHROM	0.26 ②	0.33 ②	0.21	0.99 ①	0.55 ②
	LiCVPR	−0.44	−0.58	0.45 ②	−0.38	−0.42
	HR-CNN	0.29 ①	0.82 ①	0.51 ①	0.98	0.70 ①
MAE	baseline	8.98	17.35 ②	9.19	9.29	9.29
	2SR	20.98	43.66	17.37	2.44	5.78 ①
	CHROM	7.80 ①	21.37	13.49	2.07 ②	6.29 ②
	LiCVPR	19.98	31.90	7.41 ②	28.22	28.39
	HR-CNN	8.10 ②	9.46 ①	7.26 ①	1.84 ①	8.72

HR-CNN – Conclusion

A novel method for HR estimation HR-CNN was shown performing SOTA on:

A. three non-challenging public datasets
B. newly collected challenging ECG-Fitness dataset

Thank you very much for your time…