Effects of Video Encoding on Camera-Based Heart Rate Estimation

Michal Rapczynski, Philipp Werner, and Ayoub Al-Hamadi

***Abstract*—*Objective:* Public databases are important for evaluating and comparing different methods and algorithms for camera-based heart rate estimation. Because uncompressed video requires huge file sizes, a need for compression algorithms exists to store and share video data. Due to the optimization of modern video codecs for human perception, video compression can influence heart rate estimation negatively by reducing or eliminating small color changes of the skin (PPG) that are needed for camera based heart rate estimation. In this paper, we contribute a comprehensive analysis to answer the question of how to compress video without compromising PPG information. *Methods:* To analyze the influence of video compression, we compare the effect of several encoding parameters: two modern encoders (H264, H265), compression rate, resolution changes using different scaling algorithms, color subsampling, and file size on two publicly available datasets. *Results:* We show that increasing the compression rate decreases the accuracy of heart rate estimation, but that resolution can be reduced (up to a cutoff point) and color subsampling can be applied for reducing file size without a big impact on heart rate estimation. *Conclusions:* From the results, we derive and propose guidelines for the recording and encoding of video data for camera-based heart rate estimation. *Significance:* The paper can sensitize the research community toward the problems of video encoding, and the proposed recommended practices can help with conducting future experiments and creating valuable datasets that can be shared publicly. Such datasets would improve comparability and reproducibility in the research field.**

***Index Terms*—Heart rate, remote photoplethysmography, remote PPG, camera, non-contact, video encoding, video codecs.**

I. Introduction

HEART rate detection is of utmost importance in todays modern medicine. The heart rate and its variability are used and measured for a broad array of health issues, e.g. during operations, routine checkups and risk-assessments [1]. Other useful areas of applications measuring heart rate lie, for example, in the field of competitive sports [2], where certain vital parameters should be kept in a desirable range.

The most accurate and widely used method for heart rate measurement is the electrocardiography (ECG). It measures the electrical activity of the heart muscles at certain locations of the patient's body and requires the attachment of up to ten electrodes. A medically trained person is required for attaching the electrodes correctly. Moreover the used pads and gel at the electrodes can cause skin irritation and significantly impede the freedom of movement of the patient.

A pulse oximetry sensor can be used alternatively to obtain a heart rate signal with less effort. The sensor is usually attached to a finger and measures the heart rate using the light absorption changes caused by the changing blood volume in the skin during a pulse period. Due to the commonly used spring-clip sensors a measurement over longer time periods can be uncomfortable or painful and hinders the normal use of one hand.

The measurement principle of the oximetry sensor (photoplethysmography, PPG) has been adopted for remote measurement through cameras. Several authors have proposed approaches to detect heart beats by analyzing the slight color differences caused by the absorption rate shifts due to the blood volume changes in the skin, including using common off-the-shelf cameras. This method offers easy to setup heart rate measurement with no obstruction of movement of the subject and medical staff and new applications for tele-medicine, competitive sports or human machine interfaces.

Many publications use self-generated or "their own" private data [2]–[11] when evaluating and presenting new heart rate estimation algorithms and methods. These datasets are usually not publicly available for other scientists, which impedes the development of new approaches and comparisons with existing methods. Therefore, publicly available databases for heart rate estimation are needed to advance the field faster, but the creation of a comprehensive dataset represents a big effort in time and money.

The archiving and transfer of video data can also represent a considerable effort in regard to the huge file sizes of uncompressed video data. A one minute uncompressed 1080p video with a color-depth of 8bit and 25 FPS (frames per second) would have a size of 8.7 GB. This is the reason video compression is necessary for sharing video data on a bigger scope.

Manuscript received November 28, 2018; revised February 25, 2019; accepted March 1, 2019. Date of publication March 13, 2019; date of current version November 20, 2019. This work was supported by the Federal Ministry of Education and Research of Germany (BMBF) (Vitalkam2, 03ZZ0465; HyperSense/AutoStress 03ZZ0471A; HuBa 03ZZ0470) within the Zwanzig20 Alliance 3Dsensation. *(Corresponding author: Michal Rapczynski.)*

M. Rapczynski is with the Institute for Information Technology and Communications, Neuro-Information Technology Otto-von-Guericke, University Magdeburg, 39106 Magdeburg, Germany (e-mail: michal.rapczynski@ovgu.de).

P. Werner and A. Al-Hamadi are with the Institute for Information Technology and Communications, Neuro-Information Technology Otto-von-Guericke, University Magdeburg.

Digital Object Identifier 10.1109/TBME.2019.2904326

Modern standard video compression algorithms like H.264 and H.265 are psycho-visually optimized and compress the video data in a way that quality and detail reduction is, as far as possible, invisible to human perception. This often includes color subsampling, reduced image quality during fast movements, and removing and filtering of small color changes. These optimization steps help to reduce the video information to manageable sizes so that video streaming and archiving is today possible with a high perceived image quality. The information reduction applied in these algorithms could have a strong impact on the PPG signal. This problem is often neglected in the current literature. Most papers lack details regarding the used codecs, encoding parameters, and video container formats. This impedes the comparison and reproducibility of published results.

Only a few papers explicitly address the issues of compression and how to reduce file sizes without compromising the PPG signal. McDuff *et al.* [12] tested the effect of video compression on PPG signals with 25 participants engaged in two 5min tasks. They tested the x264 and x265 codecs with different constant rate factor (CRF) values (explained in Sec. II-A2) and compared the peak signal-to-noise ratio (SNR), bit rate and mean estimation error. They concluded that videos "with a bit rate of 10 Mb/s still retained a BVP [blood volume pulse] with reasonable SNR and the pulse rate estimation error was 2.17 BPM" [12, p. 69] and suggested that the x265 algorithm may be more effective than x264 on videos with greater motion.

While McDuff *et al.* [12] proposed a minimal compressed bit rate, which is dependent on the image size and content, they did not recommend which parameters to choose for the video encoding to guarantee a good PPG signal quality.

Blackford *et al.* [13] tested the effect of reduced frame rate and resolution on heart rate estimation with 25 subjects. They varied the frame rate from 120 to 60 and 30 FPS and reduced the image resolution from 658×492 to 329×246 using bilinear and zero-order downsampling. They concluded that "there is little observable difference in mean absolute error or error distributions resulting from reduced frame rates or image resolution".

The paper of Sun *et al.* [14] also confirms the effect of the chosen frame rate by stating that, "Statistical results presented no significant difference among the various sample rates, which was in keeping with the independent relationship between the variations of [pulse rate variability] measurements and sample frequency [20-200 FPS]."

A good overview regarding the problems in the current literature, especially about the lack of information on video compression and recoding setting was done by Špetlík *et al.* [15]. This paper concluded that a higher compression rate and bilinear resolution downscaling reduces the SNR and reducing the video size increases the detrimental effect of compression on the SNR. The authors do not recommend the use of H.264 compression for camera based heart rate estimation.

In the experimental part they calculated and analyzed the signal-to-noise ratio for different compression rates and resolutions. The results disagree with those of others: They neither show a gradual SNR decrease as observed by McDuff *et al.* [12] nor that changing the resolution has little observable effect as reported by Blackford *et al.* [13]. This can be possibly explained by the used video data. Overall, only 10 videos of 5 subjects each with 60 sec. runtime where recorded and used for the analysis. The setup contained no or minimal movement, which probably made the video data easier to encode without pulse information losses due to the interframe compression (see Section II-A1) used by the encoder. The compression effects were only analyzed using the SNR as a benchmark. The authors did neither report the actual effect on the heart rate estimation, nor did they analyze which SNR is sufficient for achieving any specific heart rate estimation error.

We contribute in this paper a comprehensive and reproducible analysis to answer the question of how to compress video without compromising the PPG signal. We analyze the effects of different video encoding parameters on the heart rate estimation and develop encoding recommendations by systematically evaluating two publicly available datasets with overall 161 videos and 50 participants. Using different parameters for the constant rate factor, resolution, color subsampling and two modern video codecs, we analyzed 13,084 videos with an overall size of 1.2 TB. Four different PPG signal extraction methods and two ROIs from the literature were implemented to assess the influence of the encoding parameters on the heart rate estimation. Sec. II describes the methods for the video encoding and heart rate estimation. Sec. III contains the experimental results and Sec. IV a discussion of the results with recommendations for the encoding of video data for heart rate estimation.

II. Methods

A. Video Encoding

All videos for this paper have been generated using FFMPEG [16]. The encoding methods compared in this paper are confined to a few important options, which have a big impact on the video data, quality and the evaluation used for the heart rate estimation.

Many more options are available and possible in video encoding, but this paper is meant as a guide for the encoding and archiving options of future datasets for video based heart rate estimation and evaluates mostly the essential options which have to be chosen for the video encoding process.

1) Codecs x264/x265: We compared the influence of two codecs on the target parameters. First the *H.264* standard or *Advanced Video Coding* (AVC), which is widely used today in video streaming (e.g. *YouTube*, *iTunes Store*), HDTV broadcasts or Blu-rays. Secondly the newer, more advanced *H.265* standard or *High Efficiency Video Coding* (HEVC). This codec offers "approximately a 50% bit-rate savings for equivalent perceptual quality relative to the performance of prior standards [...]." [17, p. 1667]. We used the free x264 and x265 implementations of these codecs incorporated in FFMPEG.

Both codecs are generally trying to find redundant parts in different areas of the video, in single frames (intraframe) and in previous or succeeding frames (interframe). While the *intraframe* compression should have little effect on most heart rate estimation algorithms, due to the fact that the RGB values are usually averaged for every frame, the *interframe* compres-

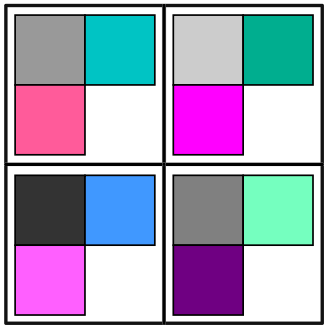

Fig. 1. Schematic representation (2 × 2 pixel) of YUV444 with full chromatic information saved for every pixel.

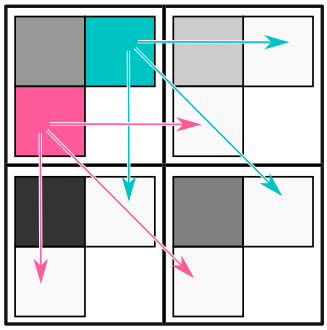

Fig. 2. Schematic representation (2 × 2 pixel) of YUV420 using chroma subsampling with chromatic information shared between pixels.

sion could have a detrimental effect on the PPG signal quality by copying the same color information into different frames, which would result in resembling a multi-frame smoothing filter.

2) Constant Rate Factor: The *constant rate factor* (CRF) is the default rate control mode for x264 and x265 and is used to set the overall perceived video quality. The value ranges from 0 (lossless) to 51 (highest compression).

This mode encodes the video to achieve a constant perceived visual quality. The compression rate can vary throughout the video to optimize the encoding, like reducing the bitrate in high motion frames to reduce file size. This is possible due to the fact that the human eye does perceive less detail in moving objects and is used in modern video encoding methods.

3) Pixel Format: Most of the cameras and screens used today have an RGB sensor or display. This means the data will be recorded and displayed in RGB. However, video data is typically encoded in the YUV color space.

Two *FFMPEG* video color pixel formats are used in this paper, YUV444p and YUV420p. The YUV color encoding system separates the color information in an image into the luminance part Y and the chrominance parts U and V, sometimes also known as C_b, C_r. We used only progressive pixel formats (p) which contain all pixel information for every frame. While the YUV444 format (see Fig. 1) saves the values of all three channels for every pixel, the YUV420 format implements chroma subsampling which results in a reduced resolution in the chrominance channels. In every 2 × 2 pixel block four Y values and only one U and one V value are saved for the whole block (see Fig. 2).

Fig. 3. Example of the FaceMid ROI.

Due to the human visions higher acuity for achromatic than chromatic color components, a reduction of color information and file size is possible without visual degradation of the image for the human perception. For this reason *YUV420* is the default pixel format for most of the modern video streaming and storage that is used today.

Important to note is that the color transformation from RGB to YUV is not reversible for all colors. The around 16.7 million RGB (8bit) colors are mapped to around 11 million YUV (8bit) colors when using the ITU recommendations in rec.601 [18] or rec.709 [19] defined colorspaces. Besides these out-of-gamut colors, rounding errors in the quantization can happen as well during the encoding (RGB → YUV) and the decoding (YUV → RGB) color transformation.

4) Resolution: To reduce the file sizes we tested the impact of changing the image resolution on the heart rate estimation. The videos were downsampled during the encoding process using three different algorithms from the *FFMPEG* scaler to calculate the new pixel values: *bicubic* (default), averaging *area* and *nearest neighbor*.

5) Video Decoding: The videos were decoded and processed using *OpenCVs* (Version 2.4.2) *VideoCapture* class in *C++* which is based on *FFMPEG*.

B. Heart Rate Estimaton

Two different regions of interest (ROI) and four PPG signal extraction methods for RGB color data were used to test the impact of different compression levels on the heart rate estimation accuracy.

1) Region of Interest: We used the Haar-Cascade classifier from *OpenCV* 2.4, to find the face in every image. In the next step the *DLib* facial landmark detector is used to calculate the pixel coordinates (u, v) of 68 points. Both steps were implemented in C++. The landmark points are stabilized over several frames corresponding to 1/10 of the video frame rate by calculating the mean for each u and v coordinate. Based on face detection and landmarks, we extracted two ROIs called FaceMid and Skin. While the FaceMid ROI is used in many approaches the Skin ROI has shown to generate the best results [20].

The **FaceMid** ROI was proposed in [3] and is a widely used ROI in the field of heart rate estimation [4], [5], [21], [22]. The Region uses the full height of the bounding box enclosing the facial landmarks, but trims the sides and utilizes only the middle 60% of the region (see Fig. 3). This is supposed to

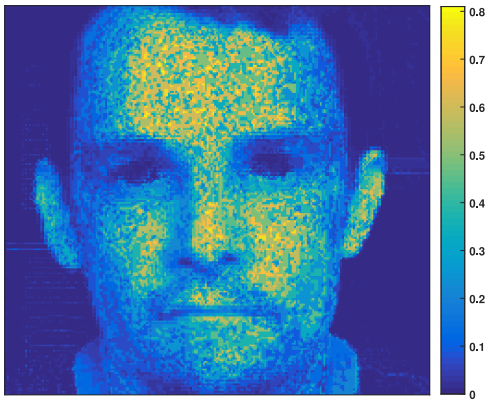

Fig. 4. Example of the Skin ROIs skin probability p.

improve the signal-noise-ratio (**SNR**) of the extracted PPG signal by removing non skin pixels at the borders [20].

The other ROI used is the **Skin** ROI approach proposed by Rapczynski *et al.* [23]. This is based on a lookup table approach from Jones and Rehg [24], using the implementation presented by Saxen and Al-Hamadi [25], which provides the skin probability p for each color pixel c (see Fig. 4). We used the skin probability p_i for each pixel i as a weighting factor for the color value c_i when calculating the pixel mean for the PPG signal, instead of a binary masking to skin/non-skin pixel.

2) Signal Extraction: DeHaan and Jeanne [26] developed a chrominance based approach (**DeHaan**) to eliminate the effect of specular reflections produced by movement. Defining two orthogonal chrominance signals, the algorithm tries to separate the motion-induced noise, which should influence both signals, from the blood volume change induced pulse signal, which should only influence one chrominance signal.

Adaptive Green-Red-Difference (**aGRD**) was presented by Feng *et al.* [27] and is based on the Green-Red-Difference (GRD) method from Hülsbusch *et al.* [28]. The approach removes diffuse and scattered light in the green and red signals to calculate a cPPG signal, with the blood pulse as highest amplitude.

Wang *et al.* [29] presented a new signal extraction approach based on an inverse of the Fourier transformation (**IFFT**). The method decomposes the RGB signals in the frequency domain and extracts the pulse signal from every calculated frequency-band to suppress distortions.

Normalized green (**normG**) was used by Stricker *et al.* [30] and Rapczynski *et al.* [23] as a signal extraction method. The green channel is normalized by the sum of all channels to compensate for different or changing spatial and temporal light intensity levels in the video.

$$PPG = \frac{g}{r + g + b}$$

3) Algorithm: We use a graph-based heart rate estimation algorithm presented by Rapczynski *et al.* [23].

The PPG signal is first filtered using an adaptive bandpass. In the initialization case, a wide (30–240 BPM) bandpass is used on a 30s long signal window. A shorter signal window (10 s) of the PPG signal is filtered in the following time steps, if a previous heart rate value could be estimated, with a much smaller passband ($\pm$15 BPM) centering around the last heart rate estimation value to increase the SNR and enable a faster reaction to a changing heart rate. The signal peaks are then isolated and represent the possible blood pulse peaks. The algorithm then analyses the Inter-Beat-Intervals (IBI) of all detected peaks in the filtered PPG signal window by creating a graph connecting all peaks with an IBI of 0.25 s to 2 s (corresponds to 30–240 BPM) between them.

In the initialization case several possible peak sequences are created by pairing peaks with similar IBI values at the start and the end of the estimation window and connecting them through the peaks in the graph while minimizing the shared mean error of the sequence. The sequence with the smallest mean squared error from their shared mean is selected and used to calculate the estimated heart rate from the mean of the IBIs of the selected peaks. If a heart rate value from the last time step is available the graph algorithm estimates the optimal continuation of the sequence by adding new peaks to the end of the sequence from the last time step. The heart rate estimation is calculated once per second.

4) IEC Error Calculation: The ground truth heart rate was calculated by using a QRS heartbeat detection method described by Schmidt *et al.* [31] followed by a manual check for missed or falsely classified heartbeats. For every heart rate estimation the same window was analyzed in the video (PPG) and ground truth (GT) data. The mean of the extracted ground truth inter-beat-intervals from the window is calculated and converted to the ground truth heart rate BPM value. The error for each heart rate estimation step is then calculated as $E = HR_{GT} - HR_{PPG}$.

The error calculation described in the IEC standard 60601-2-27 for medical ECG devices is used as validation benchmark for the heart rate estimation. Using the calculation above, an estimation is considered valid, if the absolute error between the estimated and the ground truth heart rate is less then 10% of the ground truth or 5 BPM, whichever is higher. The percentage of measurements of a dataset which meet this IEC standard is further referred to as the **IEC accuracy** (in %).

III. Experiments

We tested the impact of different encoding parameters on two datasets. The used data and the influence of the chosen CRF value, color subsampling, and video resolution are described in the following section. All videos were encoded with the other parameters at their default values and using the preset *ultrafast*.

A. Datasets

We used two different datasets to test the influence of the chosen encoding techniques, the MMSE-HR and the PURE databases. Both datasets are publicly available, contain video and synchronously recorded physiological data. Furthermore, both datasets are composed of separate image files without any interframe compression effects, thus all video encoding parameters and processing steps can be controlled for.

The MMSE-HR dataset has a bigger image resolution (>1 Megapixel) which can be used to test the influence of downsampling. It is saved in the *JPEG* format with 2 × 2 pixel color subsampling and lossy compression.

The PURE dataset consists of separate *PNG* files, with all color information saved lossless, but a small image resolution.

1) PURE: The PURE dataset was introduced by Strickler *et al.* in 2014 as a benchmark database to "compare the different [face segmentation] approaches and to examine the artifacts introduced by head motion in more detail" [30, p. 1059]. The PURE dataset contains 10 subjects (8 male, 2 female) performing each six controlled head motions. The physiological signals were captured using a finger pulse oximeter (Pulox CMS50E). Pulse rate wave and SpO2 readings were recorded with a sampling rate of 60 Hz.

The videos were recorded using an *evo274CVGE* camera in color with a resolution of 640 × 400 pixels and 30 fps. Every frame was saved as a separate *png* image file.

The setup was lighted by daylight through a large window frontal to the face. The illumination conditions changed slightly over time depending on the cloud coverage.

2) MMSE HR: The MMSE-HR is a part of the MMSE dataset presented by Zhang *et al.* [32] in 2016. The dataset was created to further the research on multi-modal emotion analysis. The MMSE-HR subset was specifically created to challenge heart rate estimation algorithms. The subset contains 102 videos of different length from 40 different subjects (17 male, 23 female; 18–66 years old) from diverse ethnic backgrounds. During an interview, the participants were exposed to different stimuli to elicit emotional reactions. For samples shorter then the used time window no estimation was calculated. The results were marked as *NAN* and ignored in the error calculation.

The physiological data was collected using a *Biopac MP150* system, which captured the blood pressure and heart rate at 1000 Hz. Other physiological signals were captured but are not part of the MMSE-HR subset. The video data was captured using a *Di3D dynamic imaging system* in color and 1040 × 1392 pixel resolution at 25 fps. "Two symmetric lights" [32, p. 3440] were used to illuminate the scene. Every frame was saved as a separate *jpg* image file with the *quality* setting at 100% and 2 × 2 color subsampling.

B. Differences Between Source and Video

To check the results of the video encoding process we compared the color information from the encoded and then decoded video to the original image. Comparisons were done for different CRF values. Videos with a CRF = 0 should result, according to the FFMPEG documentation [33], in a lossless encoded video. Due to the fact that the MMSE-HR database is saved in an already lossy format, the compression differences can only be compared up to the saved image quality and not the "original" recording data.

Fig. 5 depicts the mean squared error of the pixel value RGB differences between the video and source images of all frames of the first 10s in every video in the PURE and MMSE dataset.

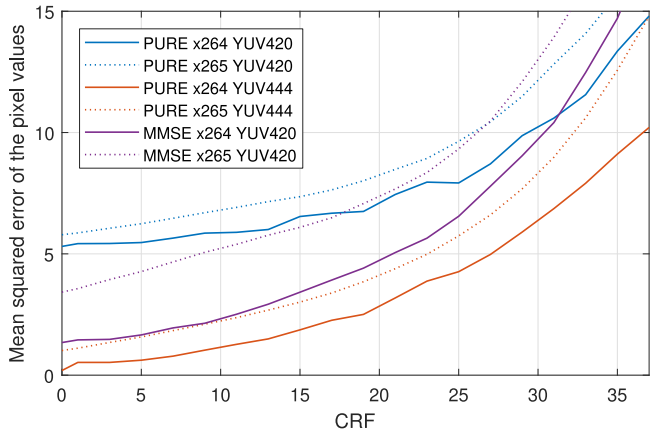

Fig. 5. Mean squared error of the differences of the pixel RGB values from the frames of the first 10 s in every video and original images in the datasets for different CRF values.

Both encoders (x264 and x265) were compared ob both datasets and different pixel color formats on the PURE database.

The high error values for the YUV420 format using the PURE dataset can be attributed to the color subsampling between the *png* images and the videos, which are much smaller using full color information with the YUV444 format. The MMSE-HR database is already color subsampled in its original JPG format, so we used only the YUV420 format.

Both codecs show bigger errors at increasing CRF values, with lower errors in the PURE dataset than the MMSE in the lower CRF values, using their corresponding native pixel color format. The x264 codec has a lower error than x265 at all CRF values. Small color changes can be detected even at CRF = 0, which contradicts the stated losslessness of the video encoding. This is due to quantization errors during the conversion from RGB to YUV and back.

The error rates using x264 with YUV420 show similar "steps" in both datasets. In the PURE dataset at a CRF = 25 the error even decreases using a higher value. The exact cause of this effect could not be estimated due to the complexity of the used codec, but it can be assumed that some dynamic quantization table, block-matching or similar internal values are calculated differently at lower CRF values and result in jumps in the compression errors.

To exclude encoding or decoding errors during the calculations or loading of the images or videos a further test was performed. A video was encoded and compared with the original images using the lossless HuffYUV codec resulting in an MSE error of 0.

C. Impact of the ROI

Two regions of interest methods are compared to test the impact of the image compression on the region of interest. The skin based method is based on a color lookup table and therefore possibly subject to additional deteriorating effects of high compression artifacts.

Figs. 6 and 7 show the mean IEC accuracy over all four signal extraction methods (see Sec. II-B2) for different CRF

TABLE I
MEAN AND STANDARD DEVIATIONS OF THE ABSOLUTE ERRORS OF THE HEART RATE ESTIMATIONS ON THE PURE AND MMSE DATASETS USING DIFFERENT ROIS AND CODECS IN IN RELATION TO THE CRF VALUE

			CRF	0	1	3	5	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35	37
Mean error	MMSE	Face	x264	0,10	-0,15	-0,13	-0,02	0,13	0,22	-0,18	0,04	-0,84	-2,10	0,30	-0,76	-2,37	-3,99	-3,76	-3,84	-3,58	-3,43	-2,43	-2,40
			x265	-0,26	-0,12	0,02	0,20	0,74	0,00	-0,64	-1,82	-2,73	-3,35	-3,62	-3,45	-3,75	-3,18	-3,58	-3,95	-2,63	-3,73	-2,27	-3,36
		Skin	x264	-0,12	-0,40	-0,33	-0,10	-0,38	-0,02	0,00	-0,11	-0,64	-1,85	-0,20	-0,84	-1,78	-3,18	-3,42	-3,31	-4,28	-3,32	-2,56	-2,87
			x265	-0,30	-0,22	0,06	0,04	0,75	0,43	-0,96	-1,97	-2,72	-3,50	-3,24	-3,06	-4,50	-3,42	-3,47	-3,69	-3,38	-2,97	-3,10	-2,94
	PURE	Face	x264	2,99	3,43	2,48	2,95	3,43	2,79	3,42	3,35	5,38	6,93	3,54	5,50	6,57	6,78	5,00	6,27	6,66	6,65	7,45	8,09
			x265	4,33	3,35	3,56	4,18	3,72	3,14	2,13	2,36	1,55	1,89	2,98	3,10	2,17	2,97	3,26	3,13	5,67	4,71	2,71	5,51
		Skin	x264	1,58	2,63	2,07	2,35	2,46	2,97	3,57	3,43	3,45	6,39	4,08	4,64	5,80	6,36	5,56	6,21	6,08	6,68	7,29	8,55
			x265	3,80	3,77	4,32	5,18	4,16	3,55	2,71	2,26	3,25	3,10	4,36	2,75	2,86	3,40	4,04	2,85	4,93	3,96	2,78	4,63
Std. dev.	MMSE	Face	x264	9,42	9,77	10,51	10,04	10,26	10,99	11,23	11,16	15,07	15,80	11,47	14,39	17,34	19,13	18,46	20,01	21,64	21,46	21,34	22,32
			x265	9,39	9,71	10,30	11,04	11,49	14,17	16,44	19,32	20,48	21,06	21,45	22,33	22,70	22,75	23,04	23,09	23,02	22,88	23,85	23,19
		Skin	x264	9,11	9,77	9,08	9,04	10,27	10,44	10,87	11,32	14,25	15,67	11,03	13,68	15,70	17,29	18,13	20,06	21,13	21,07	21,35	22,28
			x265	8,98	9,25	9,81	10,15	11,14	12,83	16,03	19,25	19,84	21,21	21,33	22,06	22,69	22,21	22,72	23,19	23,10	22,90	23,61	23,98
	PURE	Face	x264	15,54	15,10	13,03	13,63	15,22	14,91	16,57	17,16	18,68	21,10	20,95	19,18	21,98	23,23	23,57	24,22	24,05	25,74	25,26	25,14
			x265	16,63	16,65	18,03	21,44	23,00	24,22	25,17	25,47	26,31	26,21	26,37	26,11	26,79	27,06	26,26	26,05	25,18	25,43	26,35	25,58
		Skin	x264	10,12	12,66	11,74	12,18	12,66	13,57	15,75	17,37	15,74	21,94	21,14	17,43	20,02	23,26	22,61	23,13	24,39	24,77	25,01	25,17
			x265	16,13	15,54	17,08	19,89	21,41	23,08	23,57	24,45	24,67	24,23	24,58	25,48	25,33	24,97	24,92	25,79	25,33	25,56	26,13	25,71

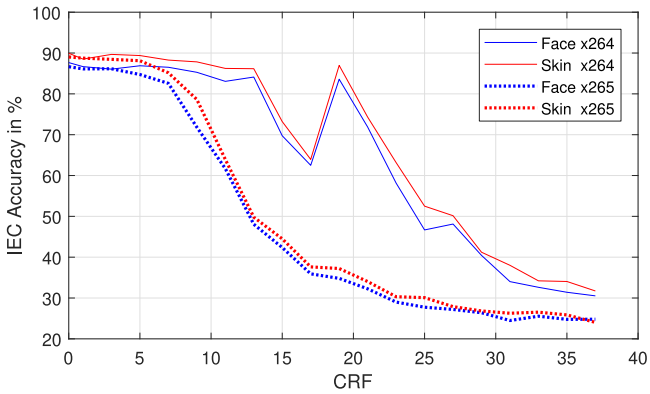

Fig. 6. Mean IEC accuracy for different CRF values for the Skin and FaceMid ROIs using x264 and x265 codecs on the MMSE dataset.

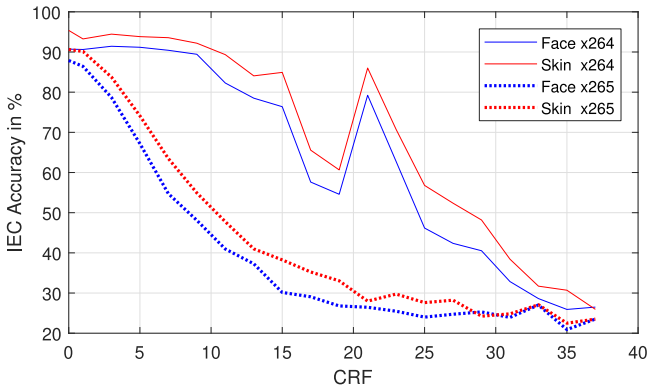

Fig. 7. Mean IEC accuracy for different CRF values for the Skin and FaceMid ROIs using x264 and x265 codecs on the PURE dataset.

values. The mean and standard deviation of the error is shown in Table I at the end of the chapter. Qualitatively, the results of both ROIs change similarly with the CRF value, but the *Skin* ROI was superior in almost all cases (except CRF = 37) over the *FaceMid* ROI. Therefore only the skin ROI is used further in the analysis.

We tested the subject wise error differences between the skin and faceMid ROIs. The skin ROI has better results over all subjects in the PURE dataset. In the MMSE dataset only two subjects had significantly worse results($>5\%$ mean IEC) using the skin ROI. One subject was a Caucasian female (F008) the other an African American female (F009). In comparison with the other subjects in the database no subject specific physiognomic cause like skin color, glasses, or other differences could be found that explained the results. Both women made a lot of moth movements (open/close/smiling/talking) in relation to other subjects and their teeth were classified as skin by the ROI algorithm, which would introduce non skin pixels to the ROI and artifacts into the PPG signal and could, therefore, lower the SNR.

The x264 codec shows a dip in both datasets around a CRF of 15–21. The accuracy falls very quickly to rise again sharply to a higher level than before the dip. The reason for this effect is unknown, it could be a result of the internal video quality scaling (see Sec. III-B). The same effect can also be seen in the pulse rate estimation error in [12, Table 1] at the stationary task and the random motion task using the x264 codec and a CRF between 6 to 9, where the estimation error increases at lower CRF values and then again decreases at higher values.

D. Impact of CRF

Several different CRF levels were tested between the values of 0–37 to account for a wide range of possible compression levels. The calculations were done on both datasets using the x264 and x265 codecs. All four different signals extraction methods (see Sec. II-B2) were used and the mean *IEC accuracy* was calculated for every CRF level.

Fig. 8 and Fig. 9 show the IEC accuracy and the space savings (in relation to uncompressed video) obtained by varying the CRF value for the x264 and x265 codecs. The means and standard deviations of the error are shown in Table I at the end of the chapter.

For both codecs, the highest accuracies are achieved with the least amount of compression with a CRF = 0. Comparing the results for similar CRF values shows a much higher compression rate for the x265 codec. The x265 codec has also a faster decreasing IEC accuracy in relation to an increase of the CRF

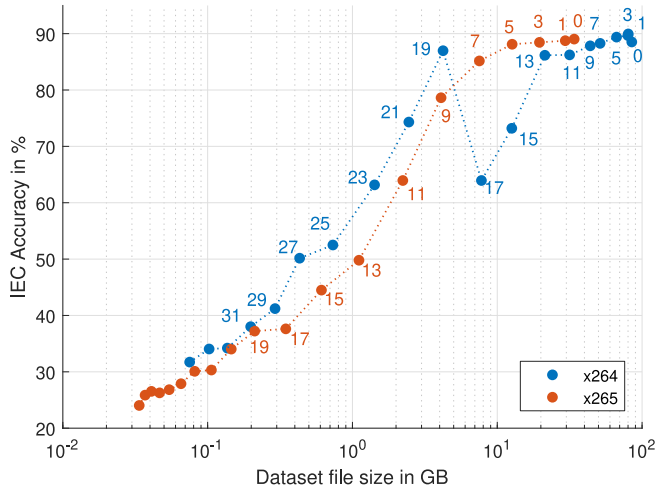

Fig. 8. Mean IEC accuracy for different CRF values (numbers) and the space savings in relation to uncompressed video for the x264 and x265 codecs (YUV420) on the MMSE dataset.

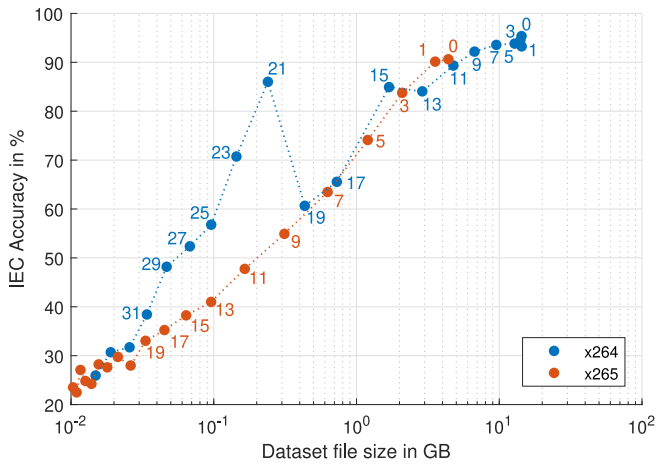

Fig. 9. Mean IEC accuracy for different CRF values (numbers) and the space savings in relation to uncompressed video for the x264 and x265 codecs (YUV420) on the PURE dataset.

value and a lower peak accuracy than x264. The x264 codec shows a dip in both datasets (see explanation in III-C) around 96–99% reduced file size and between a CRF of 15–21. The accuracy falls very quickly to rise again sharply to a higher level than before the dip. The x264 codec achieved overall the better IEC accuracies at the cost of bigger files.

E. Differences Between the Signal Extraction Methods

The Figs. 10, 11, 12 and 13 show the IEC accuracy for the different signal extraction methods.

The accuracy of the different methods was comparable for very low CRF values. No extraction method was clearly dominant over both datasets, codecs and CRF values. A different signal extraction method has the best result In one combination of codec and dataset.

The GRD approach performed generally better on higher CRFs in relation to the other methods for low CRF values. The differences were most noticeable for the stronger compressed

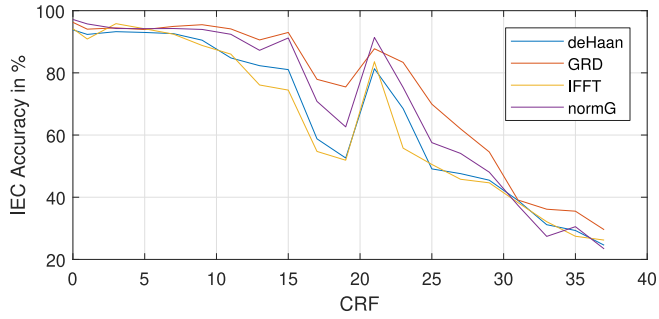

Fig. 10. IEC accuracy for different CRF values using different signal extractions and the x264 codec on the PURE dataset.

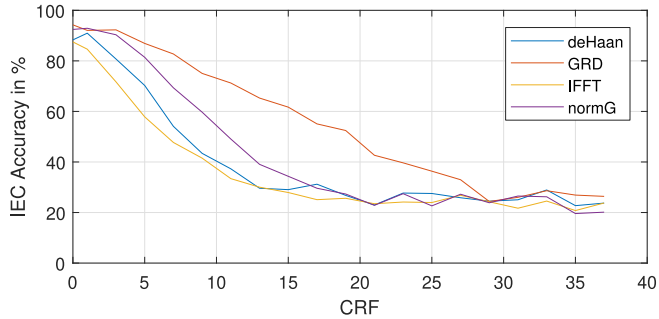

Fig. 11. IEC accuracy for different CRF values using different signal extractions and the x265 codec on the PURE dataset. PURE dataset.

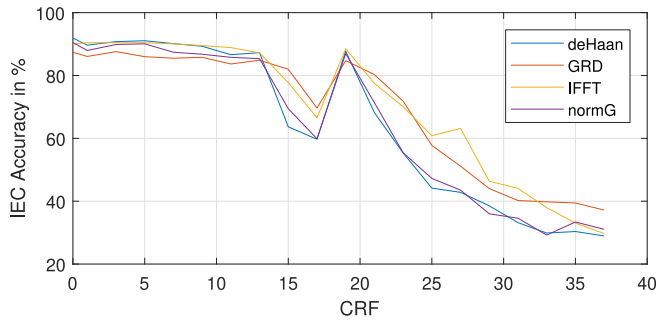

Fig. 12. IEC accuracy for different CRF values using different signal extractions and the x264 codec on the MMSE dataset. PURE dataset.

x265 codec. The IFFT approaches results worsen in relation to the other methods using the x265 codec, but only on the PURE dataset.

F. Impact of Color Subsampling

The impact of color subsampling on the heart rate estimation accuracy was tested. Only the *PURE* dataset was used for this because the *JPG* images of the *MMSE* dataset are already color subsampled. The videos were encoded in the default YUV420 and the YUV444 pixel format (see Sec. II-A3).

Figs. 14 and 15 show the mean IEC heart rate estimation accuracy, over all four signals extraction methods (see Sec. II-B2), in relation to the saved file size. The means and standard deviations of the error are shown in Table II at the end of the chapter. Both codecs, x264 and x265, were tested using the YUV420 (default) and YUV444 pixel format. Using both codecs the YUV420 pixel

TABLE II
MEAN AND STANDARD DEVIATIONS OF THE ABSOLUTE ERRORS OF THE HEART RATE ESTIMATIONS ON THE PURE DATASET USING DIFFERENT PIXEL FORMATS IN RELATION TO THE CRF VALUE

		CRF	0	1	3	5	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35	37
Mean Error	YUV420	x264	1,58	2,63	2,07	2,35	2,46	2,97	3,57	3,43	3,45	6,39	4,08	4,64	5,80	6,36	5,56	6,21	6,08	6,68	7,29	8,55
		x265	3,80	3,77	4,32	5,18	4,16	3,55	2,71	2,26	3,25	3,10	4,36	2,75	2,86	3,40	4,04	2,85	4,93	3,96	2,78	4,63
	YUV444	x264	2,48	2,49	2,74	2,29	3,36	2,30	4,29	3,62	5,23	5,02	7,57	8,26	8,62	8,24	7,21	7,14	6,98	6,65	7,39	7,37
		x265	3,50	3,19	5,02	5,22	5,52	4,00	3,04	2,77	3,27	3,98	3,98	4,10	4,31	3,67	5,22	5,50	5,50	4,83	5,54	6,65
Std. dev.	YUV420	x264	10,12	12,66	11,74	12,18	12,66	13,57	15,75	17,37	15,74	21,94	21,14	17,43	20,02	23,26	22,61	23,13	24,39	24,77	25,01	25,17
		x265	16,13	15,54	17,08	19,89	21,41	23,08	23,57	24,45	24,67	24,23	24,58	25,48	25,33	24,97	24,92	25,79	25,33	25,56	26,13	25,71
	YUV444	x264	12,94	13,03	13,50	13,25	15,70	11,82	16,84	15,18	17,64	17,01	21,79	23,29	23,59	24,45	24,64	25,14	25,71	25,63	25,26	24,08
		x265	16,38	16,41	18,50	20,78	22,27	24,01	23,69	24,87	25,18	24,62	24,40	26,05	25,45	25,85	25,66	26,29	25,87	26,13	25,06	25,90

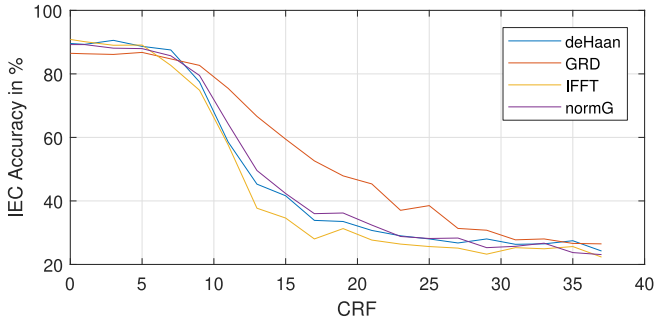

Fig. 13. IEC accuracy for different CRF values using different signal extractions and the x265 codec on the MMSE dataset. PURE dataset.

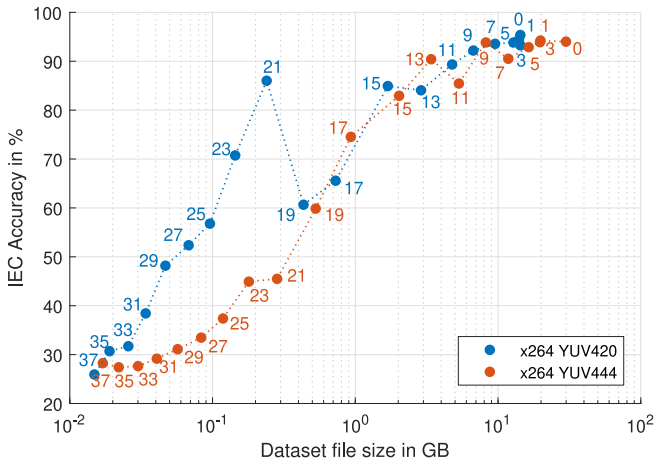

Fig. 14. Mean IEC accuracy for different CRF values (numbers) and space savings, in relation to uncompressed video, using different pixel formats and the x264 codec on the PURE dataset.

format outperformed YUV444, besides two cases with the x264 codec with a CRF of 9 and 13. The differences in IEC accuracy between the pixel formats are 1.4% using x264 and 2.8% using x265. While the file sizes are similar using the x265 codec, the YUV444 files where around double the size of the YUV420 format using x264. This fits the doubling of the mean pixel color format from 12 to 24 bit.

G. Impact of Resolution

To test the impact of video resolution on the estimation accuracy, fifteen different resolution steps were calculated using three scaling methods implemented in FFMPEG. Due to the

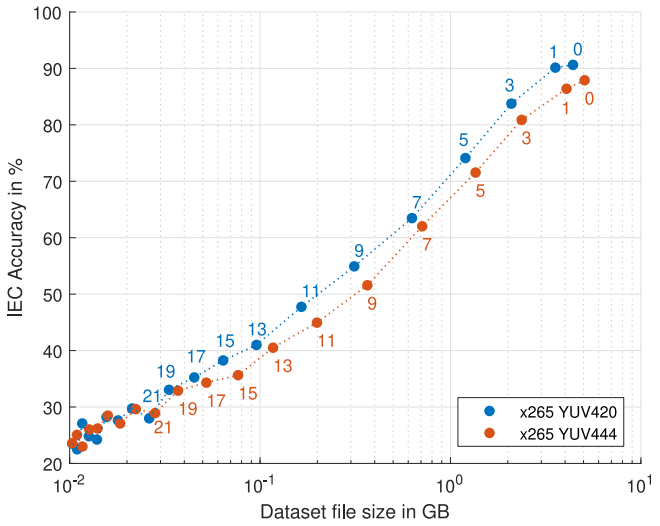

Fig. 15. Mean IEC accuracy for different CRF values (numbers) and space savings, in relation to uncompressed video, using different pixel formats and the x265 codec on the PURE dataset.

already small original resolution of the PURE dataset (640 × 400) only the MMSE dataset was used for these calculations.

The video resolution was linearly decreased from the original 1040 × 1392 pixel in steps of 1/16 of the original pixel resolution to a minimum of 130 × 174 pixel. Odd pixel dimensions were increased by one due to the need for even dimensions using the video codecs. All resolution scaling videos were created from the original *jpg* images using the x265 codec and a CRF = 0. The used scaling algorithms are *nearest neighbor*, *area* and *bicubic* (FFMPEG default). While the *area* and *bicubic* algorithms calculate their new pixel value from the information of multiple pixels, the *nearest neighbor* approach sets the target color from the color value of the spacial nearest pixel in the original image discarding the remaining color information.

Fig. 16 represents the mean IEC accuracy over all four signal extraction methods for different video resolutions. The means and standard deviations of the error are shown in Table III at the end of the chapter. It shows a stable accuracy trend up to around 100.000 pixels (∼316 pixels squared) in the facial bounding box. The *area* and *bicubic* scaling algorithm have a noticeable decline in accuracy from this point on, while the *nearest neighbor* algorithms accuracy stays over 85% with the exception of the smallest tested resolution.

TABLE III
MEAN AND STANDARD DEVIATIONS OF THE ABSOLUTE ERRORS OF THE HEART RATE ESTIMATIONS ON THE MMSE DATASET FOR DIFFERENT RESOLUTIONS USING VARIED SCALING ALGORITHMS IN RELATION TO THE MEAN OF FACE BOUNDING BOX PIXELS

Number of Pixels		5.607	12.711	22.438	35.232	50.456	68.892	89.638	113.698	139.958	169.875	201.549	237.015	274.592	315.720	358.524
Mean Error	Neighbor	5,05	3,84	3,35	3,41	2,83	3,41	3,19	2,78	2,55	3,00	2,57	2,95	3,23	2,76	-0,30
	Area	10,27	8,33	6,61	6,67	5,72	5,11	3,17	4,14	3,57	2,75	3,05	2,90	3,11	2,80	-0,30
	Bi-cubic	9,87	8,48	6,60	5,90	5,09	4,94	3,81	3,42	2,92	2,81	2,88	2,96	3,16	2,85	-0,30
Std. dev.	Neighbor	14,37	12,46	11,52	11,66	10,18	12,43	11,50	10,64	10,01	11,50	10,13	11,14	11,67	10,99	8,98
	Area	18,25	16,39	15,33	15,84	15,65	14,80	10,97	13,14	12,68	10,48	11,15	10,74	11,13	10,63	8,98
	Bi-cubic	17,50	16,81	15,11	14,79	14,16	14,40	12,05	11,71	10,53	10,80	11,07	10,83	11,34	10,78	8,98

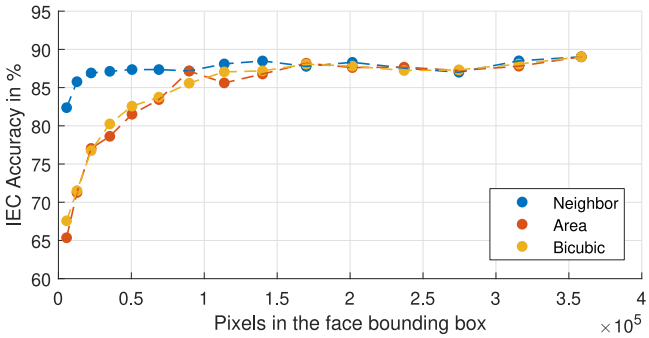

Fig. 16. Mean IEC accuracy over the mean face bounding box size for different resolutions using three scaling algorithms on the MMSE dataset (x265, CRF = 0).

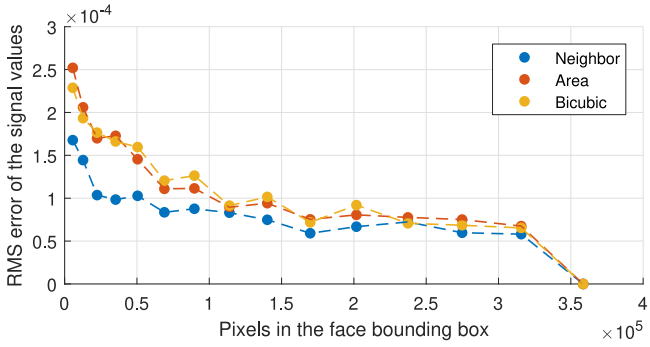

Fig. 17. RMS error of the PPG signals (*normG*) from different video sizes in relation to the original size for using three scaling algorithms on the MMSE dataset (x265, CRF = 0).

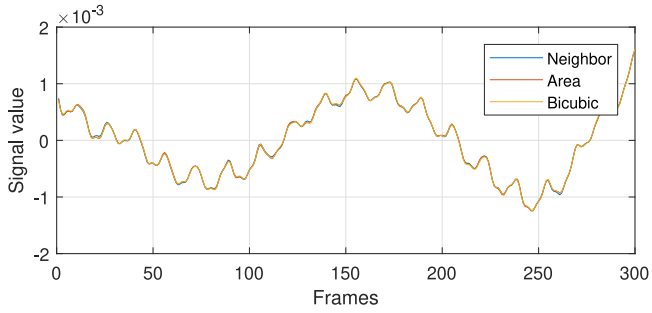

Fig. 18. Example PPG signal (*normG*) from the MMSE dataset (videoID: F005/T10) of the first 300 frames with a resolution of 976 × 1306 pixel using different scaling algorithms.

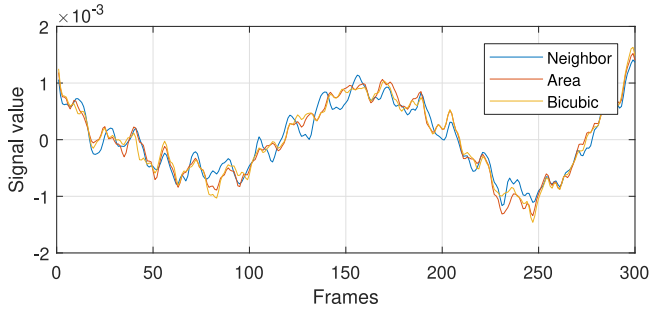

Fig. 19. Example PPG signal (*normG*) from the MMSE dataset (videoID: F005/T10) of the first 300 frames with a resolution of 130 × 174 pixel using different scaling algorithms.

The differences in the PPG signal quality using the different scaling algorithms can be seen in Fig. 17. It shows the root mean square error of all downsampled *normG* PPG signals in the MMSE dataset to the PPG signals from the original 1040 × 1392 pixel source. The RMS errors have a similar trend to the IEC accuracy seen in Fig. 16. The error increases slowly with the reduction of the resolution and are comparable for all three methods up to around 100.000 face pixels. The *nearest neighbor* approach shows a distinctly smaller error in relation to the original signals and rises slower than the *area* and *bicubic* scaling algorithms, while the RMS error of the other two approaches rises steadily from 100.000 pixels, corresponding with the decrease of the IEC accuracy.

Figs. 18 and 19 show an example PPG signal from videos with different resolutions and scaling algorithms. While the PPG signals are almost identical with only a small amount of size reduction (see Fig. 18), the signals in the smallest calculated resolution differ considerably from each other. The *nearest neighbor* algorithm has a much clearer peak prominence than the other two methods. Especially at the slopes of the signal (see Fig. 19 Frames 100–175) the peak height of the *area* and *bicubic* scaling algorithms is distinctly lower.

Below a bounding box size of around 20.000 face pixels (∼141 pixels squared) the error increases considerably for all algorithms in a similar manner.

IV. DISCUSSION

A. Choosing the Right CRF Value

Of all the tested parameters the CRF is the most important. Depending on the used codec already very small values can have an incredibly detrimental effect on the extractable PPG signal. The default values of 23 for x264 and 28 for x265 reduce the PPG signals quality by such a great amount that the videos would be practically useless for heart rate estimation (see Sec. III-D).

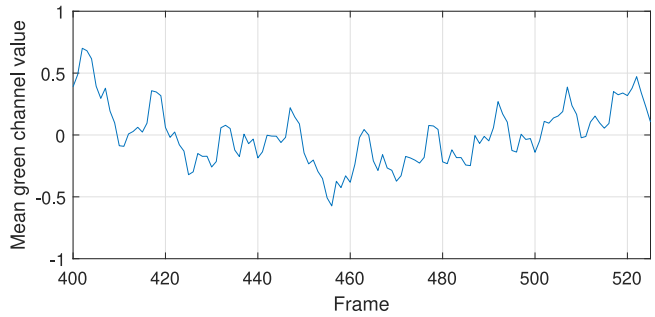

Fig. 20. Linear detrended green channel sample (MMSE, videoID: F005/T10, x265, CRF = 0).

While the optimal value depends on the quality, resolution and content of the encoded dataset, a $CRF = 0$ is the safest option in regards to PPG quality and the value should only be increased if file size is an issue and other space saving options are already exploited.

B. Use of Subsampling for File Size Reduction

Two forms of information reduction through subsampling have been tested. Color subsampling and decreasing the image resolution. Both forms of information reduction show that the accuracy of the heart rate estimation can be held constant on reduced color and pixel information. Subsampling methods which determine the new pixel color through one source pixel without any filtering or averaging (*YUV420* and *neighbor*) show stable IEC results during information reduction.

1) Color Subsampling: While using the x264 codec, only small differences in the IEC accuracy could be detected while applying color subsampling to the data (see Fig. 14). The additional color information resulted in doubled file sizes for the YUV444 format (see Fig. 14). Using the x265 codec, the size differences for the color-subsampled videos were much smaller (+14% at $CRF = 0$) (see Fig. 15). In both cases the *YUV420* pixel format achieved better results than *YUV444* and x264 slightly better than x265. The better performance of the YUV420 format, with less color information, could be explained with better optimization in the encoding process (default pixel format), or a high PPG information content in the Y channel. Another explanation could be that some kind of color subsampling was already carried out by the cameras, in which case the color information was later upsampled during the conversion into the *PNG* format. These effects should be further examined in the future.

2) Resolution: The data in Figs. 16-17 shows that the amount of face pixels can be reduced without a big negative impact on the IEC accuracy. An interesting effect can be seen at small image sizes. If reducing the image to less than 10 k of facial bounding box pixels only the *nearest neighbor* algorithm continues with stable results (see Fig. 16), while the results for the *area* and *bilinear* methods begin to decrease.

A possible explanation for this observed effect is, that the heartbeat in the PPG signal has a lower amplitude than the quantization steps of the video. Fig. 20 shows a detrended green signal sample from the MMSE dataset. The heartbeat peaks have a height of less than 0.5 and are therefore smaller than the color quantization steps of the video. The PPG signal is hence only detectable due to the mean color value of enough skin pixels.

But in the case of video encoding the results of the averaging by reducing the video resolution are not saved as a float but an integer pixel value (0–255) in the new video. In this case the mean of a subset of the original pixel values achieves a better result, than the mean of an interpolated and quantized subset of pixel values, which loses this information in the process. An example would be the array [2 3 3 3 2 3 2 3 3] with a mean of 2.66. By reducing the "resolution" by 1/3 and take a random subset of every three values (*nearest neighbor*) the expected result would be [3 2 3] (not necessarily in that order) with also a mean of 2.66. But if the new values are averaged and rounded (*area*, *bilinear*) the new result would be [3 3 3] with a different mean than the starting array.

Therefore, any local filtering or averaging should be avoided to preserve this subpixel color information from rounding during the re-quantization. This also applies to image transformations, rotations, and similar operations. If for example, some steps are necessary to calculate the ROI, the ROI should be mapped back onto the image instead of using the transformed image. Therefore, only one global averaging over all ROI pixels per frame from the original image should be calculated during the heart rate estimation process, especially when using small ROIs.

C. Conclusions

Video compression is a very important issue for camera-based heart rate estimation. Every experiment or dataset can be invalidated by using the wrong compression method or default parameter, wiping out many hours of work and a lot of money. File size reduction through video compression is possible and – if used correctly–facilitates manageable file sizes for sharing of data and the comparison of different algorithms to advance this field of study. We showed that some of the options for reducing file size seem to preserve the PPG information better than others.

From the results and discussions in this paper, we derived some **guidelines** to increase the quality of video data for camera based heart rate estimation including the recording setup, video and encoding parameters:

- Hardware
  - When possible use industrial cameras, with a low signal to noise ratio, to control all aspects of the recording process.
  - Take care when using consumer products. Automatic compression algorithms can be included in certain hardware (webcams, camcorders, etc.) and could lead to information loss.
  - Use lighting and appropriate shutter speeds to achieve a high dynamic range of color and brightness.
- Recording
  - Set the frame rate between 20–30 Hz. Higher FPS are not adding much information (see [13], [14]) but increase file sizes.
  - Use the highest color-depth possible, to optimize the detection of small color changes.

  - Record using uncompressed RGB *avi* format, the HuffYUV codec or png images during the session.
- Encoding
  - Encode using x264 with a CRF = 0 (saves >80% compared to uncompressed data). The default CRF should not be used.
  - Use chroma subsampling (YUV420) to save 50% file size (default using x264).
  - The resolution *can* be reduced (keeping >50.000 face pixels) using *nearest-neighbor* downsampling to avoiding loss of subpixel color information while saving additional file space with small estimation accuracy losses.

D. Future Work

More datasets with a higher variance of image content are needed to validate the stated hypotheses of this paper to draw a more general conclusion about the influence of video compression on camera based heart rate estimation. This can be seen in the slightly noisy IEC error changes for different CRF values which should smooth out with enough data.

An analysis of the heart rate accuracy dip and spike at higher CRF values, seen in Figs. 8 and 9 using the x264 codec could lead to much smaller files with preserved PPG information if the effect could be predicted and reliably reproduced.

The effect of other video parameters beyond the scope of this paper can be tested. A possible example would be to set the CRF_{max} parameter equal to the CRF value to prevent the reduction of quality during movement.

Dedicated PPG codecs could be developed in the long term, which could be specialized to preserve the PPG information by reducing the video bitrate only in non-essential areas of the image by e.g. using facial or skin detection.

References

[1] 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.

[2] Y. Sun *et al.*, "Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise," *J. Biomed. Opt.*, vol. 16, no. 7, 2011, Art. no. 077010.

[3] M.-Z. Poh *et al.*, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Opt. Express*, vol. 18, no. 10, pp. 10762–10774, 2010.

[4] 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.

[5] H. Monkaresi *et al.*, "Automated detection of engagement using video-based estimation of facial expressions and heart rate," *IEEE Trans. Affective Comput.*, vol. 8, no. 1, pp. 15–28, Jan.–Mar. 2017.

[6] W. Verkruysse *et al.*, "Remote plethysmographic imaging using ambient light," *Opt. Express*, vol. 16, no. 26, pp. 21434–21445, 2008. [Online]. Available: http://www.opticsinfobase.org/abstract.cfm?URI=oe-16-26-21434-1

[7] M. Lewandowska *et al.*, "Measuring pulse rate with a webcam—A non-contact method for evaluating cardiac activity," in *Proc. Federated Conf. Comput. Sci. Inf. Syst.*, 2011, pp. 405–410.

[8] T. Blöcher *et al.*, "An online PPGI approach for camera based heart rate monitoring using beat-to-beat detection," in *Proc. IEEE Sensors Appl. Symp.*, 2017, pp. 1–6.

[9] L. Iozzia *et al.*, "Assessment of beat-to-beat heart rate detection method using a camera as contactless sensor," in *Proc. IEEE 38th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2016, pp. 521–524.

[10] H. Nisar *et al.*, "Contactless heart rate monitor for multiple persons in a video," in *Proc. IEEE Int. Conf. Consum. Electron. Taiwan*, 2016, pp. 1–2.

[11] J. Spigulis *et al.*, "Multi-spectral skin imaging by a consumer photo-camera," in *Proc. Multimodal Biomed. Imag.*, 2010, vol. 7557, pp. 75570M-1–75570M-9.

[12] D. J. McDuff *et al.*, "The impact of video compression on remote cardiac pulse measurement using imaging photoplethysmography," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2017, pp. 63–70.

[13] E. B. Blackford and J. R. Estepp, "Effects of frame rate and image resolution on pulse rate measured using multiple camera imaging photoplethysmography," in *Proc. Med. Imag. Biomed. Appl. Mol., Structural, Functional Imag.*, 2015, vol. 9417, pp. 94172D-1–94172D-14.

[14] Y. Sun *et al.*, "Noncontact imaging photoplethysmography to effectively access pulse rate variability," *J. Biomed. Opt.*, vol. 18, no. 6, 2012, Art. no. 061205.

[15] R. Špetlík *et al.*, "Non-contact reflectance photoplethysmography: Progress, limitations, and myths," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit.*, May 2018, pp. 702–709.

[16] "FFmpeg," 2018. [Online]. Available: http://ffmpeg.org/. Accessed on: Oct. 18, 2018.

[17] G. J. Sullivan *et al.*, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[18] "Recommendation bt.601, studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios," ITU, Geneva, Switzerland, 2011.

[19] "Recommendation bt 709, parameter values for the HDTV standards for production and international programme exchange," ITU, Geneva, Switzerland, 2015.

[20] M. Rapczynski *et al.*, "How the region of interest impacts contact free heart rate estimation algorithms," in *Proc. 25th IEEE Int. Conf. Image Process.*, Oct. 2018, pp. 2027–2031.

[21] L. Wei *et al.*, "Automatic webcam-based human heart rate measurements using laplacian eigenmap," in *Proc. Asian Conf. Comput. Vis.*, Springer, 2012, pp. 281–292.

[22] M. A. Haque *et al.*, "Heartbeat signal from facial video for biometric recognition," in *Image Analysis*, R. R. Paulsen and K. S. Pedersen, Eds. Berlin, Germany: Springer, 2015, pp. 165–174.

[23] M. Rapczynski *et al.*, "Continuous low latency heart rate estimation from painful faces in real time," in *Proc. 23th Int. Conf. Pattern Recognit.*, 2016, pp. 1165–1170.

[24] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," *Int. J. Comput. Vis.*, vol. 46, no. 1, pp. 81–96, 2002.

[25] F. Saxen and A. Al-Hamadi, "Color-based skin segmentation: An evaluation of the state of the art," in *Proc. IEEE Int. Conf. Image Process.*, 2014, pp. 4467–4471.

[26] 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.

[27] L. Feng *et al.*, "Motion-resistant remote imaging photoplethysmography based on the optical properties of skin," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 5, pp. 879–891, May 2015.

[28] M. Huelsbusch, "An image-based functional method for opto-electronic detection of skin-perfusion," Ph.D. dissertation, RWTH Aachen Dept. Elect. Eng., Aachen, Germany, 2008.

[29] W. Wang *et al.*, "Robust heart rate from fitness videos," *Physiol. Meas.*, vol. 38, no. 6, pp. 1023–1044, 2017.

[30] R. Stricker *et al.*, "Non-contact video-based pulse rate measurement on a mobile service robot," in *Proc. IEEE 23rd Int. Symp. Robot Human Interactive Commun.*, 2014, pp. 1056–1062.

[31] M. Schmidt *et al.*, "A real-time QRS detector based on higher-order statistics for ECG gated cardiac MRI," in *Proc. IEEE Comput. Cardiol. Conf.*, 2014, pp. 733–736.

[32] Z. Zhang *et al.*, "Multimodal spontaneous emotion corpus for human behavior analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 3438–3446.

[33] "FFmpeg h.264 video encoding guide," 2018. [Online]. Available: http://trac.ffmpeg.org/wiki/Encode/H.264. Accessed on: Aug. 16, 2018.