Proc. 20. Workshop Farbbildverarbeitung (FWS14), Wuppertal, pp. 161-174

Comparison of Face Segmentation Methods for Non-contact Video-based Pulse Rate Measurement

Ronny Stricker, Steffen Müller, Horst-Michael Groß

Ilmenau University of Technology,
Helmholtz-Platz 5, D-98693 Ilmenau
eMail: ronny.stricker@tu-ilmenau.de
URL:http://www.tu-ilmenau.de/neurob

Abstract Robotic applications in the context of prevention and assistance for elderly people living alone in their home environment are of special interest in the near future. Measuring pulse rate on a mobile service robot aimed to motivate elderly users for physical exercise can be a valuable information in order to adapt to the users conditions. Non-contact image photoplethysmography has shown promessing results for remote pulse rate measurements of the face under ambient illumination. However, significantly less work can be found that aims to analyze different methods for face segmentation. Therefore, in this work, a typical processing pipeline was implemented, and a detailed comparison of methods for face segmentation was conducted, which is the key factor for robust pulse rate extraction even, if the subject is moving. A benchmark data set is introduced focusing on the amount of motion of the head during the measurement.

1 INTRODUCTION

Important functionality of a robot companion for the elderly is the motivation and guiding of the users to do physical and cognitive exercises, as well as offering the explicit measurement of vital parameters.
Fig. 1 shows the robot platform used for our research. It is equipped with interaction devices, mainly a touch-display, as well as a couple of additional sensors enabling autonomous navigation and perception of people and obstacles in the robots environment.

Focusing on the vital signals and parameters, in this work, the camera in center of the robot's face was used to capture a video stream of the user's face while sitting in front of the system.
The robot fulfills several tasks during which the user does sit relatively still in front of the robot focusing its screen (e.g. when checking mails or appointments, doing video conferencing, or during the explanation of physical exercises). These periods of time are supposed to be ideal for remote pulse rate measuring. Therefore, the user does not have to wear intrusive devices, and the pulse rate can be measured spread over the whole day without discomfort.

Figure 1: Scitos-G3 robot.

A lot of different approaches to measure the cardiac pulse rate are known from literature with electrocardiography (ECG) methods as the so called gold standard. However, ECG-based methods are very intrusive, since they require electrodes placed on the body, and the devices are expensive. Alternative approaches are covered by the term photoplethysmography and are measuring the cardio-vascular pulse wave traveling through the body by evaluating the blood volume changes in the microvascular bed of tissue. Since light is absorbed stronger by blood than by the surrounding tissue, the reflectance of the human skin changes over time [1].
One example of the class of contact-based photoplethysmographie is the pulse oximeter that can be traced back to the work of Goldberger at al. in 1987 [2]. Worn on a fingertip or earlobe, these devices are illuminating the tissue with a dedicated light source to monitor blood O_2 saturation and pulse rate.
Initial attempts to remote measurements of pulse rate have been presented in the first decade of this century using special cameras and active red and infrared lighting [3, 4]. These attempts are still being researched [5, 6] but are not suited for mobile robots due to the special lighting requirements. We want to point out, that there are developments in that community to separate noise and pulse rate and to detect false pulse signals from inanimate objects [7]. However, these methods are out of the scope of this paper.
The first attempt to remote measurement of pulse rate with ambient light have been made by Verkruysse et al. [1] in 2008. Since that time, the work done by [8, 9, 10] have had major influence and lead to mobile phone applications mostly based upon the same methods, e.g. [11].
Remote measurement of the pulse rate is usually based on the face region of the subject since this region is unusual to be occluded. Since tracking of different body parts lies beyond the scope of this paper, we are focusing on the pulse rate extraction from the human head.
Although, the negative effects of head movements during measurements have been noted by different authors, we have found significantly less work dealing with a systematic analysis of how and to which extend different movements disturb the measurements. The most detailed analysis of head movements and approaches to overcome artifacts introduced by them is given by Sahindrakar in [12].

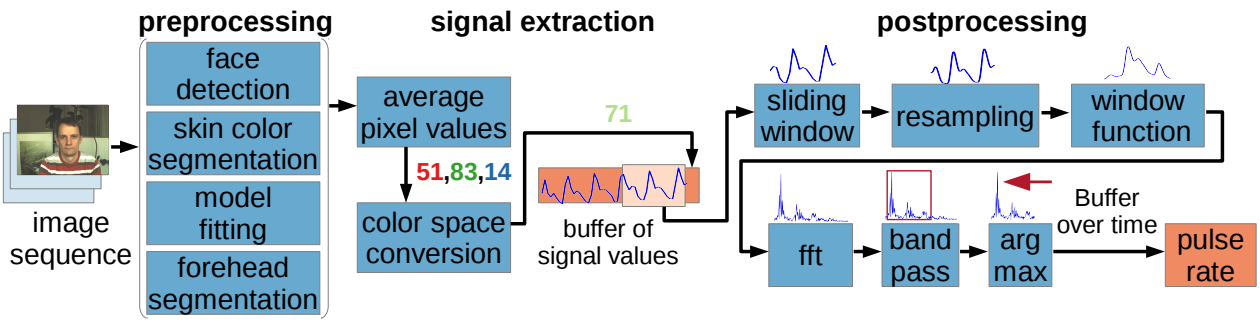

Figure 2: Processing pipeline for pulse rate extraction from video sequence. The pipeline is divided into *preprocessing*, *signal extraction* and *postprocessing*.

Therefore, in the following we want to work out the circumstances under which pulse measurements can be obtained on a mobile robot. Furthermore, we will give a comparison of different methods known from literature to increase motion robustness of non-contact photoplethysmography. Carrying on with the state-of-the-art, we introduce an approach to further increase motion robustness and computational complexity, which is crucial on a mobile robot that has to run a bunch of algorithms with limited resources.
At first in Sec. 2, we give a short overview of the principles involved in measuring pulse rate from images using ambient light. Different Region of Interest (ROI) extraction methods and our proposed method are explained in Sec. 3. In Sec. 4, we introduce our database to evaluate the influence of different types and strength of head motion. The comparison of the different methods is given in Sec. 5. In the last section we discuss our findings and give an outlook to future work.

2 Extracting Pulse Rate from Face Images

As already explained, the variation of blood amount in the skin tissue during a pulse rate wave leads to a change of the reflected light, which can be recorded with a standard RGB camera. Because pulse rate is a very slow signal (60 - 100 pulsations per minute for an adult during rest and $220 - Age$ at a max), the sampling rate of the camera is of minor influence to the results. Sun et al. [6] verified that fact, thus a usual frame rate of 15 to 30 per second is sufficient for our purposes.
Unfortunately, the amount of color change due to the pulse rate wave is very small compared to global shading effects, such that the signal is in the range of the pixel noise. To gain a useful signal, despite that bad conditions, averaging a couple of pixels is necessary, which reduces the influence of pixel noise compared to the signal. The quality of the raw signal mainly depends on the selection of the pixels used and can be contaminated by artifacts due to motion of the subject or changing illumination of the skin. Even, if the sampled area of skin is too large, the signal might be blurred since the pulse wave has different phase shifts in different parts of the body. Before further processing, therefore, a robust segmentation of the skin region –mostly used is the face– is an essential part of the algorithm, that is in focus of that paper. Fig. 2 gives an overview of the whole processing chain used in our approach. Various methods for **preprocessing** of the images have been suggested in literature starting with fixed

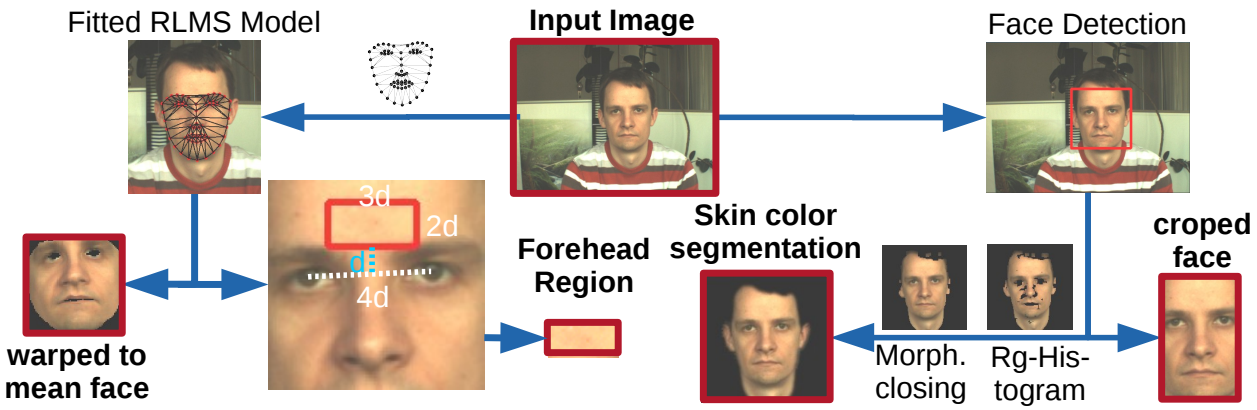

Figure 3: Overview of different ROI extraction algorithms.

boxes in the image, where the subject has to place her/his face inside. This was improved by rectangular face detection and later skin color segmentation inside the region of interest. More sophisticated methods try to compensate effects of changing projection of the face during head movements by tracking individual patches in the face [12]. We suggest to further improve the head pose invariant extraction of color signals by means of a face model that tracks the head over time and projects the shape of the face to a standard shape independent on the actual head orientation. In Sec. 3, these preprocessing alternatives are discussed in more detail.

Having a mask of pixels to be evaluated, the next step in the processing chain is the **signal extraction**. This is done by averaging the pixel values inside and combining the RGB channels in order to get a one-dimensional signal over time to be further processed. For that combination a simple selection of the green channel G or a normalization of intensity $g = G/(R+B+G)$ is common in literature. The normalization approach reduces global illumination changes and increases the influence of the color hue. Therefore, the normalization showed a tremendous improvement of the accuracy and thus is favored for the further discussion. In a real time application, the resulting values for each frame are pushed in a buffer, where a sliding time window of several seconds in length is used for frequency analysis in the **postprocessing**. On the buffer of all three available input channels, source separation techniques by means of Independent Component Analysis (ICA) or Principal Component Analysis (PCA) can be applied in order to remove artifacts in the combined signal. However, the use of ICA for that purpose seams to be questionable. On one hand, it requires measurements over 30 seconds to be robust [12] and on the another hand, it is computationally complex and can corrupt the data, if head motion is present [12, 13], thus we omit that step.

Based on the 1-D time signal, there are two possible ways for further analysis. The first one is peak detection in time domain, which is rather complicated with a signal of such poor quality. From the peak distances, several authors compute a pulse rate, which has the advantage of fast results compared to the second alternative, which is the analysis in the frequency domain by means of a Fourier Transform. In that approach, also applied by us, the reachable frequency resolution $\Delta f = \frac{1}{T}$ is proportional to the inverse duration T of the available signal, causing the usage of a 15 sec window in our case.

The signal values are re-sampled equidistantly using cubic spline interpolation. To enable proper analysis of the non-periodic signal part, a window function is applied in order to suppress side lobe responses in the frequency domain before transformation into a power spectrum is performed. In the resulting spectrum, the periodic pulse wave should show a significant peak, which can be found by a simple maximum search. Since the frequency resolution is limited by the sampling frequency, a quadratic interpolation between the maximum and its neighbor values is applied to find the maximum position with higher accuracy. To prevent from false classification of artifacts in implausible frequency ranges, the spectrum is multiplied with a band pass filter in before.
Result of the described procedure is a sequence of pulse rate values, one for each window processed. In order to use this for further decisions in an application, it would be reasonable to have a quantification of the signal quality and thus of the reliability of the values. This information can be extracted from the power spectrum as well. The signal to noise ratio (SNR) is useful here and has been evaluated for the different preprocessing alternatives in the experiments presented in Sec. 5.

3 Face Segmentation Methods

After the basic processing chain has been introduced in that section, several methods for preprocessing of the image have to be discussed. In the experiments conducted, these alternatives have been evaluated with respect to robustness against head motion and facial movements induced by talking and interaction.

3.1 Fixed Region in the Image

The most simple way to extract the raw signal is to integrate the pixel values of a fixed rectangle of the image. Therefore, the signal from all pixels (mainly containing skin area) are averaged reducing the pixel noise. However, a rich variety of artifacts in the background of the image are reducing signal quality. Furthermore, problems arise if multiple subjects are present in the image, and head movements change the average pixel values. Different parts of the background become visible, when the head is moving in front of non-uniform background.

3.2 Face Tracking

To reduce influence of background, most methods from literature rely on some form of face detection and tracking. During previous tests we found out, that simple face detection, e.g. using the well known Viola-Jones face detector [14], can lead to bad pulse rate estimates, since the face ROIs are bound to discrete scales and therefore jump when the subject moves the head. We suggest to use the Viola-Jones for initialization only and apply feature tracking afterwards. For feature tracking we have used the tracker described in [22] that tracks a sparse template-based feature point set and can track in plane object movements. We decided to use this tracker since it delivers state-of-the-art tracking performance, can run at hyper-real-time and, therefore, is suitable to be utilized on a mobile robot. Similar to [8, 9] the bounding box from

the tracker is reduced to 60% of the original width to avoid the border of the face, where background might be contained, while the vertical elongation of the ROI is used completely (see Fig. 3). With this procedure, the raw signal for further processing is extracted by averaging the pixel values of the ROI in the input image.

3.3 Model-free Segmentation Based on Skin Color

Based on the face tracking, it is possible to improve segmentation accuracy by taking into account the color features of the skin regions of interest. For our evaluation, a processing stack for the face ROI has been implemented relying on a powerful segmentation method called grabcut [16]. The grabcut algorithm has to be initialized with a region in the image that is labeled as foreground (the face) and a region known to be background. We suggest to use skin color segmentation to perform this step as described in the following paragraph. From that, in an iterative process the remaining pixels are assigned to one of the two classes based on the color histogram of both classes. The resulting foreground labeled pixels after two iterations of the algorithm are taken as improved face mask. To avoid artifacts from the border of the face, a geometric erosion is applied on the mask before all pixel values inside the mask are averaged to gain the raw signal for further processing. (Fig. 3)
The quality of the grabcut segmentation mainly depends on the quality of the initial mask fed into the grabcut algorithm. The ROI provided by the *face tracking* approach also contains parts of the hair and sometimes background. Therefore, it cannot serve as a mask for grabut. A segmentation based only on skin color followed by some morphologic processing does not deliver perfect face tracking. Hower, it yields a sufficient initial mask for crabcut.
In detail, first the image is converted to intensity normalized rg-color space, and a histogram of the pixels in the face tracker ROI is derived containing 64 bins for each of the r and g channels. We assume that the maximum peak in that histogram refers to skin color, thus next step is marking all pixels falling into that maximum bin or into the directly neighboring ones as foreground. Noise, resulting in single pixels marked as foreground, is removed by means of a morphologic opening (erosion followed by a dilation). To include inner face features in the initial foreground mask, like the eyes and shadows beneath the nose, a morphologic closing operation is performed afterward on the preliminary mask. The initial foreground mask for grabcut now results from eroding the preliminary mask by 5%, while the background follows from dilating the preliminary mask by 10%. This yields a undefined border region that has to be classified by the grabcut algorithm.
Even if this procedure is executed consecutively on every frame and yields a perfect mask of all skin pixels of the head, the signal extracted from that region might be problematic due to the spatial variation of the pulse rate wave. The signal might be blurred compared to signals extracted from smaller local regions, but on the other hand a large area reduces pixel noise in the signal sufficiently. The model free segmentation also does not take self occlusion of rotated heads into account. This kind of artifacts can be compensated only by means of more complex methods as described in the following.

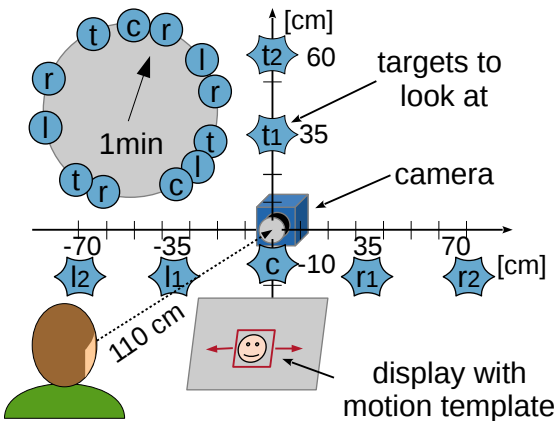

Figure 4: Setup for recording the benchmarking video sequences of controlled head motion: display with translating target, head direction targets for rotation sequences (l_1, l_2, t_1 for small elongation angles of 20°and l_2, r_2, t_2 for medium elongation of 35°.)

3.4 Model-based Approach

The methods presented up to now do not use an explicit representation of the exact shape of the face. Therefore, the pixel regions included in the ROI can vary over time, especially if the head is moving. To reduce these noise artifacts, we suggest application of a deformable model fitting approach for face extraction. One of the approaches providing reliable and accurate fitting results under varying illumination is the *Deformable Model Fitting by Regularized Landmark Mean-Shift* (*RLMS*) approach introduced by Saragih et al. in [17]. The approach tracks a number of landmarks located at different face locations by local optimization and enforcement of a global joint motion prior. Since detailed explanation lies above the scope of this paper, the interested reader is referred to [17]. The tracking code is based on the code provided by Saragih [18] and does use the default model provided by Saragih. This model delivers descent face tracking results with changing illumination. However, the tracking results do defer from person to person. The landmarks used by the tracker are located at fixed locations of the face and are connected by a triangulation. Therefore, a piecewise affine warp can be used to project model instances in a single frame onto the mean face obtained during model training. (see Fig. 3)

3.5 Model-based approach using the forehead region

The deformable model provides compensation of overall head movement, but it may be unable to cope with inner facial movements (e.g. moving eyes and mouth while talking). By extraction of pulse rate based on ROI placed on the forehead of the subject only, the moving parts of the face can be avoided. According to [19], we locate the forehead region based on the position and distance of the eyes. Since we already do have precise model of the face provided by the *RLMS* algorithm, we decided to use the eye position estimated by *RLMS*. As proposed by Lewandowska [19], we extract the forehead region as shown in Fig. 3.

4 Database

To compare the different approaches and to examine the artifacts introduced by head motion in more detail, we have recorded a benchmarking database. Head movements

were performed under controlled and well defined parameters. The database comprises 10 persons (8 male, 2 female) that were recorded in 6 different setups resulting in a total number of 60 sequences of 1 minute each.
The videos were captured with a eco274CVGE camera by SVS-Vistek GmbH at a frame rate of 30 Hz with a cropped resolution of 640x480 pixels and a 4.8mm lens. Reference data have been captured in parallel using a finger clip pulse oximeter (pulox CMS50E) that delivers pulse rate wave and SpO_2 readings with a sampling rate of 60 Hz.
The test subjects were placed in front of the camera with an average distance of 1.1 meters. Lighting condition was daylight trough a large window frontal to the face with clouds changing illumination conditions slightly over time.
Further details on the recorded data can by found at [20]. The database is available upon request and can be used for evaluation. To our knowledge, there is no free database available for ambient light image photoplethysmography so far. Therefore, we encourage other researchers to use this database as a reference benchmark for their own algorithms.
The six different setups were as follows:
i) **Steady (S)**; The subject was sitting still and looks directly into the camera avoiding head motion.
ii) **Talking (T)**; Simulated video sequence, where the subjects were asked to talk while avoiding additional head motion. This setup equals a video conference situation in a real robot application.
iii) **Slow translation (ST)**; These sequences comprise head movements parallel to the camera plane. Therefore, the images recorded by the camera were displayed on screen and shown to the subjects. A moving rectangle of the size of the face was added to the image, and the subjects were asked to keep their face inside. The rectangle was moving horizontally at a controlled speed and with a predefined pattern, thus the sequences of all individuals are repeatable. The average speed was 7% of the face height per second, where the average face height was 100 pixels.
iv) **Fast translation (FT)**; This dataset has the same setup as slow translation, except twice the speed of the moving target.
v) **Small rotation (SR)**; This setup comprises different targets that were placed at 35 cm around the camera. The subjects were told to look at these targets in a predefined sequence. They were asked to move not only there eyes but orient their head. See Fig. 4 for an impression of the setup. The one minute sequence of the targets is shown in the little clock in the figure. Random times ensure that the motion artifacts are not periodically. Depending on the distance between the camera and the subject, that roughly varies between 1 m and 1.3 m, the head rotation angles are round about 20°.
vi) **Medium rotation (MR)**; This sequences had the same setup as for small rotation, but with targets placed 70 cm around the camera resulting in average head angle of 35°.
The pulse rate of the test persons varies slightly between and during sequences that do have a length of 1 Minute each. Minimum pulse rate measured using the oximeter is at 42 BPM and the maximum rate was 148 BPM. Although, we have had very high pulse rates from one subject, all recording were taken during rest.

5 EXPERIMENTS

For our experiments, we extracted the pulse rate from all sequences using the method described in Sec. 2 and using the face segmentation methods described in Sec. 3. We used 512 equidistant supporting points for the cubic spline resampling, which is close to the framerate of the camera, if a window length of 15 seconds was used. These 512 samples also provide sufficient resolution for the FFT. The α parameter of the Kaiser-Besser window is of minor influence in our setup and was set to 18. Furthermore, we applied a bandpass that allows values from 0.67 to 3.3 Hz, which corresponds to a pulse rate of 40 to 200 beats per minute (BPM). For comparison, we used the photoplethysmogram (changes of the skin color reflectance over time) obtained from the pulse oximeter. Therefore, the same signal extraction and post-processing steps were applied on the reference signal. The actual evaluation was executed on the interval from second 5 to 60 for every sequence to give the methods some time for initialization. Further details on the parameters used for the different face segmentation methods can by found at [20].

5.1 Signal Extraction

As already noted by other authors, the hemoglobin absorptivity differs across the spectral range recorded by the camera [8]. Therefore, different color channels are suited to a different extend. The green channel is the one used most often for pulse rate extraction, if no component analysis is applied. As shown in Fig. 5a, we can confirm that the green channel performs best if all methods and sequences are averaged, whereas the normalized color channels (See Sec. 2) do perform significantly better. This significant boost of the performance is caused by changing shadows on the face due to head movements changing the intensity. The hue changes of the skin color is affected less on the other side.

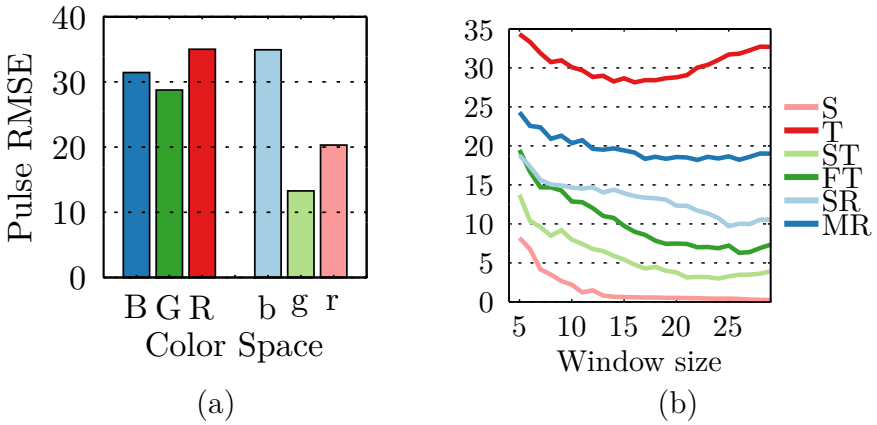

Figure 5: (a) Comparison of different color spaces. RMSE over all sequences and all algorithms.
(b) Comparison of window sizes used for FFT. RMSE values on normalized green channel sampled with 512 points.

5.2 FFT Window Size

The window size used for the FFT is a trade-off between accuracy, time resolution, and noise robustness. Fig. 5b shows the accuracy for window sizes between 5 and 30 seconds. It becomes obvious that pulse rate measurements get more reliable with an increasing window size. Nevertheless, we find window length between 15 and 20 seconds to be most robust. With more than 20 seconds, measurements can get unreliable due to the non-stationary nature of the pulse rate, which is visible in the plot of the talking sequences (Fig. 5b). Therefore, a FFT window length of 15 seconds has been used for the following experiments.

5.3 Comparing Different Sequence Types

Since the root mean square error (RMSE) is not a good measure, if the errors can become arbitrary large, we have analyzed the signal-to-noise ratios (SNR) (Fig. 6) of the computed power spectrograms as proposed in [12]:

$$SNR = \frac{\sum_{n=A}^{B} spect(n)}{\sum_{n=MinPulseBin}^{MaxPulseBin} spect(n) - \sum_{n=A}^{B} spect(n)} \tag{1}$$

A is the bin of the power spectrum at the reference pulse rate with an offset of -12 BPM and B is the bin with an offset of +12 BPM.
As a second measure of quality, we present histograms of the pulse rate deviation compared to the oximeter reference on the individual sequences. As expected and shown in Fig. 7a, measuring pulse rate on the *Steady* sequences could be performed robust and led to almost perfect measurements for all ROI extraction methods.
In plane head translation can be handled well by all the different head tracking methods with an slight advantage for *RLMS* and *Forehead*, that do deliver results close to the *Steady* case. Although, the skin detection produces decent segmentation results on a single image, the pixels classified as foreground vary over time, which introduces noise and does make the results a bit unpredictable.
The small rotation sequences delivered somewhat surprising results, since the *face tracking* method did perform better than the *RLMS* approach (Fig. 7e). The medium rotation sequences (Fig. 7f) are even harder, since greater rotations do have a stronger

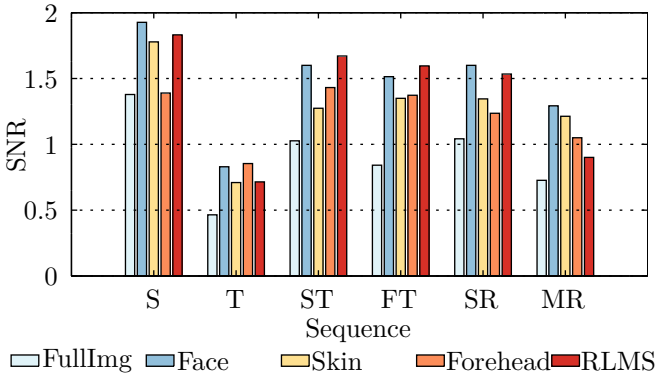

Figure 6: SNR of the different extraction methods on the different sequences.

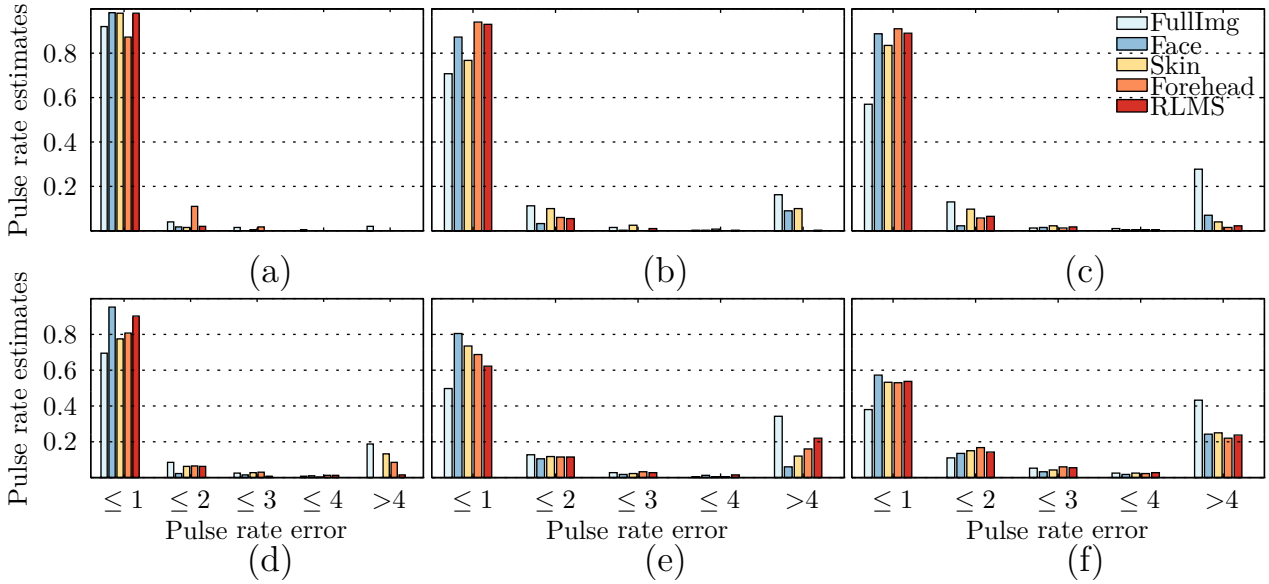

Figure 7: Plot of the pulse measurement error for the different approaches and sequences: (a) *Steady* (b) *Slow translation* (c) *Fast translation* (d) *Small rotation* (e) *Medium Rotation* (f) *Talking*

effect on shadows in the face leading to mediocre signal noise ratios (Fig. 6). Having a closer look at the face tracking generated by the *RLMS* approach, it becomes obvious that the outline of the face is not always tracked accurate if larger face rotation comes into play. Therefore, we suggest to interrupt pulse rate measurements, if stronger head rotations do occur, which can be detected if the rotation parameters of the *RLMS* model exceed a certain treshold.
The *Talking* sequences are the ones most challenging in our dataset (Fig. 7b). Pulse rate histograms as well as SNR are bad for these sequences and for all presented methods. However, on a real robot *RLMS* can be favored, if it is coupled with a talking classifier, which can be trained on the shape parameters of the model. By means of that, the unreliable time intervals, during which the subject is talking, can be left out.
It has to be mentioned that real head motions, in contrast to our motions predefined for measuring motion robustness, are a combination of translation and rotation.

5.4 Computational Complexity

Computational complexity is very important for a service robot system since it hast to fulfill different tasks (e.g. navigation and people tracking) at the same time with limited resources. The robot used is equipped with an Intel Core i7-2640m 2.8 GHz (2 Cores), where only one core was used for our runtime evaluation. The results are presented in Table 1. It is obvious that the skin segmentation has the highest computational complexity. The other approaches (if implemented in the suggested way) are running at a very high speed leaving enough CPU resources to perform additional HRI tasks on the robot.

Table 1: Runtime comparison

[ms]	FullImage	Face	Skin	Forehead	RLMS
ROI Extr.	0	10.4	10.4+57.4	12.9+0	12.9
Postproc.	0.5	0.5	0.5	0.5	0.5
Max FPS	2000	91.7	14.6	74.6	74.6

6 CONCLUSION

The experiments presented here showed that all methods are able to measure the pulse rate reliably, if no motion is present. However, the sensitivity to motion artifacts does differ. Since the *Skin* segmentation approach is computationally complex and does not deliver leading results in the comparison ranking of the different approaches, we do not recommend it for usage on a mobile robot. The *Forehead* approach seems promising but becomes unreliable if parts of the forehead are covered with hair. The two methods that are suited best to be applied on a mobile robot are the *RLMS* and *Face tracking* approaches. With both methods the pulse rate can be measured in an unobtrusive manner, if the user of the robot is sitting relatively still, e.g. when checking mails, appointments or doing video conferencing. Nevertheless, we do favour the *RLMS* approach since it enables the detection of different types of head motions using the model parameters. Therefore, critical head motions or talking can be detected easily and the pulse rate evaluation can be limited to time intervals with in plane translation or small rotation head motion. It should be possible to deliver good results even on *Talking* sequences since 50% of the measurements obtained are very good (Fig. 7b). Moreover, it is possible to adapt the time-window length used for analysis based on this motion information in order to maximize frequency resolution.
For this study, furthermore, a test dataset with pulse rate reference signal data has been recorded, which is available for other researchers on request.

7 OUTLOOK

Next step in our work will be to focus on the region used for signal extraction in more detail. First attempts of extracting single triangles from the *RLMS* mesh (see Fig. 3, top left) and performing pulse rate extraction for these individually have shown to be less reliable. Even the best triangle performs far worse than the combination of all triangles. Nevertheless, we are confident that a combination of different triangles (which can also be dynamically selected) can be found that can further reduce the error rates of the already well performing complete *RLMS* approach.
Another point for improvement is the face model itself. The *RLMS* tracker is based on a set of classifiers trained on an universal set of faces, not covering any individual properties of the subject of interest. Using the *RLMS* model for initialization of a face template, which is used for tracking afterward similar to [21], can make the face model fit more robust to the face of the current user during head motion.

References

[1] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, Remote plethysmographic imaging using ambient light, Opt. Express, vol. 16, 2008, pp. 21434–21445.

[2] D. S. Goldberger, J. E. Corenman and K. R. McCord, Durable sensor for detecting optical pulses, US Patent 4,685,464. 1987

[3] F. Wieringa, F. Mastik, and A. Steen, Contactless multiple wavelength photoplethysmographic imaging: A first step toward spo 2 camera technology. Annals of biomedical engineering vol. 33, nr. 8, 2005, pp. 1034–1041.

[4] K. T. Humphreys, T. Ward and C. Markham, Noncontact simultaneous dual wavelength photoplethysmography: A further step toward noncontact pulse oximetry. Review of Scientific Instruments, vol. 78, nr. 4, 2007, p. 044304.

[5] Y. Sun, S. Hu, V. Azorin-Peris, S. Greenwald, J. Chambers and Y. Zhu, Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise. Journal of Biomedical Optics, vol. 16, nr. 7, 2011, p. 077010.

[6] Y. Sun, S. Hu, V. Azorin-Peris, R. Kalawsky and S. Greenwald. Noncontact imaging photoplethysmography to effectively access pulse rate variability. Journal of biomedical optics, vol. 18, nr. 6, 2013, p. 061205.

[7] F. Zhao, M. Li, Y. Qian, and J. Z. Tsien, Remote measurements of heart and respiration rates for telemedicine. PLoS ONE vol. 8, nr. 10, 2013, p. e71384.

[8] M.-Z. Poh, D. J. McDuff and R. W. Picard, Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Optics Express, vol. 18, nr. 10, 2010, pp. 10762–10774.

[9] M.-Z. Poh, D. J. McDuff and R. W. Picard, Advancements in noncontact, multiparameter physiological measurements using a webcam. IEEE Trans. on Biomedical Engineering, vol. 58, nr. 1, 2011, p. 7.

[10] W. Verkruijsse and M.P. Bodlaender, A novel biometric signature: multi-site, remote (>100 m) photo-plethysmography using ambient light. Technical Report Philips Research, 2010, 00097.

[11] S. Kwon, H. Kim and K. S. Park, Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. in Proc. of International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, 2012, pp. 2174?-2177.

[12] P. Sahindrakar, G. de Haan and I. Kirenko, Improving Motion Robustness of Contact-less Monitoring of Heart Rate Using Video Analysis. Ma.S. thesis, Technische Universiteit Eindhoven, Department of Mathematics and Computer Science, 2011.

[13] S. Kwon, H. Kim and K. S. Park, Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. in Proc. of: International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, 2012, pp. 2174–2177.

[14] P. Viola and M. Jones, Robust real-time face detection, International Journal of Computer Vision, vol. 57, nr. 2, 2004, pp. 137-154.

[15] OpenCV: `http://opencv.org`

[16] C. Rother, V. Kolmogorov, and A. Blake, GrabCut: Interactive foreground extraction using iterated graph cuts, ACM Trans. Graph., vol. 23, 2004, pp. 309–314.

[17] J. M. Saragih, S. Lucey and J. F. Cohn, Deformable model fitting by Regularized Landmark Mean-Shift, In: International Journal of Computer Vision, vol. 91, nr. 2, 2011, pp. 200–215.

[18] J. M. Saragih, S. Lucey, and J. F. Cohn. Face alignment through subspace constrained Mean-Shifts. in Proc. of International Conference of Computer Vision, Kyoto, 2009, pp. 1034–1041.

[19] M. Lewandowska, J. Ruminski, T. Kocejko, and J. Nowak, Measuring pulse rate with a webcam - a non-contact method for evaluating cardiac activity, in Proc. of Federated Conference on Computer Science and Information Systems, Poland, 2011, pp. 405–410.

[20] Pulse rate detection database: `http://www.tu-ilmenau.de/neurob/data-sets/pulse`

[21] R. Stricker, Ch. Martin, H.-M. Gross, Increasing the robustness of 2D active appearance models for real-world applications, in Proc. Int. Conf. on Computer Vision Systems, Liege, 2009, pp. 364–373.

[22] A. Kolarow, M. Brauckmann, M. Eisenbach, K. Schenk, E. Einhorn, K. Debes, H.-M. Gross, Vision-based hyper-real-time object tracker for human-robot applications, in: Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, Vilamoura, Portugal, 2012, pp. 2108–2115.