

Chapter 5

Unobtrusive Vital Data Recognition by Robots to Enhance Natural Human–Robot Communication



Gerald Bieber, Marian Haescher, Niklas Antony, Florian Hoepfner and Silvio Krause

Abstract The ongoing technical improvement of robotic assistants, such as robot vacuum cleaners, telepresence robots, or shopping assistance robots, requires a powerful but unobtrusive form of communication between humans and robots. The capabilities of robots are expanding, which entails a need to improve and increase the perception of all possible communication channels. Therefore, the modalities of text- or speech-based communication have to be extended by body language and direct feedback such as non-verbal communication. In order to identify the feelings or bodily reactions of their interlocutor, we suggest that robots should use unobtrusive vital data assessment to recognize the emotional state of the human. Therefore, we present the concept of vital data recognition through the robot touching and scanning body parts. Thereby, the robot measures tiny movements of the skin, muscles, or veins caused by the pulse and heartbeat. Furthermore, we introduce a camera-based, non-body contact optical heart rate recognition method that can be used in robots in order to identify humans' reactions during robot-human communication or interaction. For the purpose of heart rate and heart rate variability detection, we have used standard cameras (webcams) that are located inside the robot's eye. Although camera-based vital sign identification has been discussed in previous research, we noticed that certain limitations with regard to real-world applications still exist. We identified artificial light sources as one of the main influencing factors. Therefore,

G. Bieber (✉) · M. Haescher · N. Antony · F. Hoepfner
Fraunhofer-Institut für Graphische Datenverarbeitung, Rostock, Germany
e-mail: gerald.bieber@igd-r.fraunhofer.de

M. Haescher
e-mail: marian.haescher@igd-r.fraunhofer.de

N. Antony
e-mail: niklas.antony@igd-r.fraunhofer.de

F. Hoepfner
e-mail: florian.hoepfner@igd-r.fraunhofer.de

S. Krause
Institut für Informatik, Universität Rostock, Rostock, Germany
e-mail: silvio.krause@uni-rostock.de

© Springer Nature Switzerland AG 2019
O. Korn (ed.), *Social Robots: Technological, Societal and Ethical Aspects of Human-Robot Interaction*, Human–Computer Interaction Series,
https://doi.org/10.1007/978-3-030-17107-0_5

we propose strategies that aim to improve natural communication between social robots and humans.

Keywords Vital data · Activity · Recognition · Autonomous computing · Optical · Camera · Webcam · Wearable computing · Assistive technology

5.1 Motivation

Nowadays, robots pervade many areas of technology and daily life. Examples of these, such as mechanical workers in the important field of assembly, manufacturing, and production show the possibilities that emerge by applying robots. Since developments in machine learning and artificial intelligence drive the capabilities of robots, new application fields arise, such as that of social assistance. Social robots will entertain, train, educate, or simply interact with users in the same way as humans do (Korn, 2018). During a conversation between a human and a social robot, much information has to be exchanged (even exceeding the modality of speech). This information includes optical, tactile, and acoustical modalities such as facial expressions, prosody of speech, and body language. Moreover, in a conversation, it is necessary to know if the counterpart is nervous or somehow affected by the discussion. To assist in this non-verbal communication, emotion recognition enabled by detecting human vital signs would be very beneficial. Usually, the recognition of vital signs is performed by applying wearable sensors attached to the human body. Depending on the parameters measured, these sensors tend to be obtrusive (e.g., wearable electrocardiography (ECG) sensors, heart rate chest straps) and require constant usage in order to provide gapless recording of data. Therefore, a touchless vital data recognition system for social robots is very beneficial for health care or communication purposes. In addition, or for special purposes, a light touch of the social robot on the user's body enables the robot to feel vital data, and this provides additional capabilities to enhance the natural communication (Fig. 5.1).

Due to the possible mobility of social robots, new environment-based requirements arise that lead to certain considerations with regard to the analysis of the transmitted signal quality. This leads to the following research questions:

- How can communication between a social robot and a human become more natural?
- Which of the vital data modalities can be assessed touchlessly?
- Which parameters influence the quality of touchless vital data recognition (via cameras) in the application field of social robots?
- In particular, what is the most relevant confounding factor for camera-based heart rate recognition and how can we deal with it?



Fig. 5.1 Cameras in the eye of the robot detect pulse rate and heart rate variability in order to enable natural communication between the social robot and the human

5.2 Related Work

Since robots are performing tasks in the home environment, the user desires the robots to have natural language capabilities (Goodrich, 2007). Speech interfaces support a natural communication modality and therefore support the identification of a user by recognizing individual speech habits or voice and language characteristics (Zissman, 1996). The combination of communication modalities enhances the total understanding and reliability of information exchange (e.g., McGurk effect) (Nath, 2012). We propose the identification of vital data for emotion detection since physiological signals show a strong correlation to emotions. However, whether emotions can be recognized reliably from physiological signals is still a matter of research (Jerritta, 2011). The most common signals for emotion detection are ECG signals for heart rate and heart rate variability, skin conductivity, respiration rate, and skin temperature. These parameters provide good results in terms of classification of emotions (Haag, 2004). Simple emotion detection can be achieved even with a reduced feature set (e.g., by analyzing ECG and respiratory signals only) (He, 2017).

The assessment of vital data by social robots is possible by direct body contact and by a touchless sensing. Such vital data assessment should be as unobtrusive as possible. This supports the natural communication and an agreeable feeling. During a normal conversation between humans, it is normal to touch the hand or arm of the interlocutor. It is of interest, if a robot is also accepted to assess relevant data to judge the human's feeling, mood, or emotion.

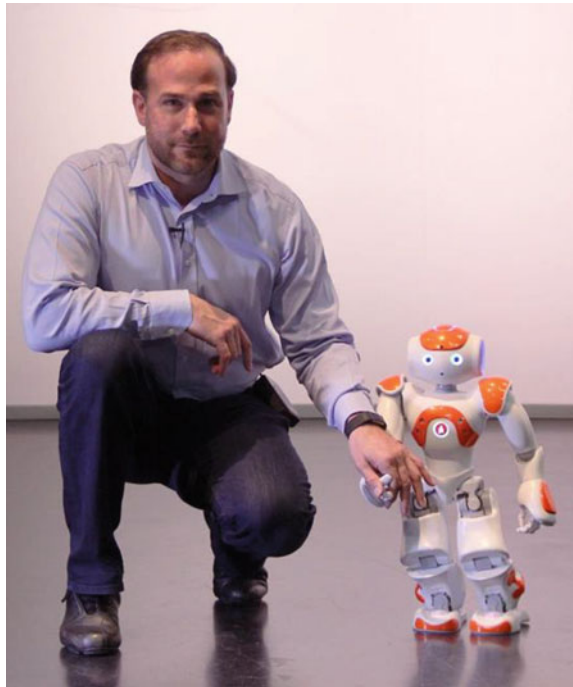
5.2.1 Touchy Sensors

Social robots might look like humans, but they can also be pet- or Muppet-like, comic figures, or even androids. The physical contact between robot and human is mainly by soft touch and body contact (Figs. 5.2 and 5.3).

Fig. 5.2 Care Robot Paro, it is touch sensitive and interacts with users while being fondled *Source* CC-BY-SA-2.0 (Biggs 2005)



Fig. 5.3 Toy-like robots



When a human is stroking the fur of a pet or is holding the hand of a robot, the integrated sensors of the social robot are able to measure simple but also complex vital data. A short contact is sufficient to measure the temperature and galvanic skin response very easily. In addition, the assessment of more advanced information is also possible. Direct contact with the skin enables electro-technical-based assessment methods. For example, an electrocardiogram (ECG) helps to detect the heart rate (HR) or heart rate variability (HRV) and provides basic stress parameters. Capacitive sensing might identify the respiration rate, and electromyography (EMG) is used for muscle activity detection.

Furthermore, remarkable parameters are the frequency and amplitude of muscle vibrations. Each muscle of a mammal performs tiny movements, resulting in a light vibration. This low-amplitude muscle vibration was first reported in the early 1960s (Rohracher, 1964), and this phenomenon is correlated to some body conditions, e.g., level of stress, medication, temperature distribution or hints of diseases such as Parkinson's or other neural degeneration diseases. Muscle activity can also be measured by electromyography (EMG) (Clancy, 2002), but recent accelerometry is also sensitive enough to detect muscle vibration (Bieber, 2013). While electro-technical methods usually need multiple body contacts to detect potential differences, accelerometry needs only one.

The measurement of a single point acceleration of the skin even provides information about the heart (Matthies, Haescher, Bieber, Salomon, & Urban, 2016). The physical movement of the heart and the blood flow through the body also cause movements of the body and skin. These tiny movements have characteristic patterns and may describe heart anomalies. This technique of measuring forces on the heart is called ballistocardiography or seismocardiography (Inan, 2015).

5.2.2 *Optical Sensors*

A touchless technology for the identification of vital data is the usage of optical information. The first non-invasive blood oxygen saturation meter SpO₂ was invented in 1935 (Matthes, 1935). With this, the skin of the ear was illuminated in order to measure the amount of light passing through the tissue. For the optical and volumetric measurement of the skin, only one frequency band (the color of the light spectrum) is needed (Fig. 5.4).

This technique is referred to as photoplethysmography (PPG) (Hertzman, 1937). It can be performed by analyzing the reflected light (reflective PPG) or light that shines through the tissue (transmissive PPG). Medical oxygen saturation meters (SpO₂) attached to the finger mainly apply the transmissive approach, while fitness trackers, smart bands, or smart watches, located on the wrist, mainly use the reflective approach. All of the devices use light-emitting diodes (LEDs) as the source of light for an appropriate illumination. The applied colors vary between the light of the green or red light diodes.

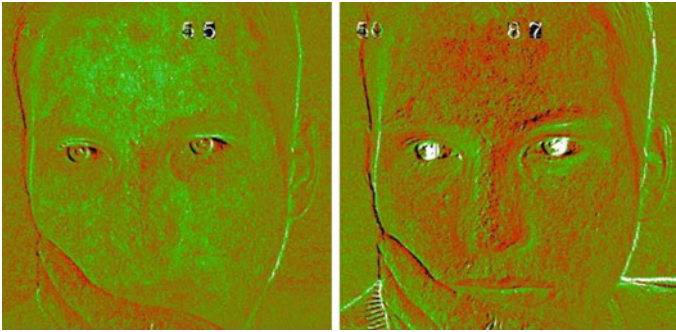


Fig. 5.4 False color image of a face with and without oxygen-enriched blood

The general concept for heart rate recognition with cameras is based on the identification of periodic change in skin color. Therefore, the camera detects skin-reflected light illuminated by sources within the surroundings of the user (Poh, 2011). An additional LED or comparable dedicated light source is not needed but provides better results. Blood with a higher oxygen saturation reflects light differently than blood with a lower oxygen saturation. With every heartbeat, the saturation changes and so does the light reflection (Kong, 2013). The facial skin shows a high degree of perfusion and therefore reflects light differently during the cardiac cycle. The effect of reflection characteristics is influenced by multiple factors, including varying tissue volume, tissue tension, and other side effects. Cameras detect the heart rate as a change in color, which is not visible to the human eye (Wu, 2012).

The optimal position for detecting changes in skin color is the forehead. This position is favorable because in a conversation with a robot, the head of the human is usually pointed toward the robot. Therefore, an integrated face detection algorithm in the social robot identifies the position of the eyes and the forehead region quite easily. The average of the green color channel values of the detected forehead region changes with every pulse cycle. The pulse rate can be determined by analyzing the resulting data stream. For the sake of data processing, we selected only part of the forehead image, the so-called region of interest (ROI). We recorded data with a camera (camera model IDS UI-306xCP-C) in a laboratory setting at constant lighting. The camera was mounted statically.

A social robot that is equipped with a camera for pulse rate detection should be able to move around in order to interact in different rooms or surroundings. Hence, the accuracy of pulse detection should be tolerant of user-specific effects (e.g., head movements) and environmental constraints. Therefore, we need to consider the main influencing effects of touchless vital data recognition via cameras.

The identification of heart rate by examination of the skin color depends on two general categories of parameter:

- Technical Parameters

- Environmental Parameters.

Both categories will be discussed in the following sections.

5.2.2.1 Technical Parameters

The quality of camera pictures depends on several factors. These include the image sensor, lens, processing hardware, and other factors.

Image sensor size: Digital cameras vary in design, size, energy consumption, and image quality. High-end cameras consist of an image sensor with a large physical size in comparison with compact cameras. When the image sensor is larger, more light can reach the individual pixel areas on the sensor. The Advanced Photo System type-C (APS-C) is an image sensor format approximately equivalent in size to the Advanced Photo System “classic” negatives of 25.1×16.7 mm. In contrast, cameras of compact devices such as the iPhone 5S have an image sensor with the size of 4.54×3.42 mm.

F-factor: Another parameter that determines how much light reaches the sensor is defined by the aperture size. The f-number of an optical system such as a camera lens is the ratio of the system’s focal length to the diameter of the entrance pupil (Smith, 2007). It is a dimensionless number that is a quantitative measure of shutter speed and therefore an important concept in photography. It is also known as the focal ratio, f-ratio, or f-stop (Smith, 2005). The higher the f-ratio, the better the exposure. This applies to most applications. An iPhone 5S camera has an $f/2.2$ lens.

Photosensitivity: Analog film provides specific sensitivity to light. This sensitivity is measured and numbered as an ISO speed. The product of ISO and shutter speed controls the brightness of the photo. The base ISO describes the speed of the highest image quality, minimizing as much noise as possible. Digital sensors only have a single sensitivity, which is mainly defined by the signal-to-noise ratio (SNR). The SNR is measured in decibels (dB). The higher the SNR, the better. A good value is about 40 dB (Baer, 2000).

Speed: The digital image sensor needs time to take a photo or to sense the frame of a video. For the recognition of pulse or respiration rate, at least a double sampling rate is necessary in order to meet the Nyquist–Shannon sampling theorem. Almost every digital video sensor is capable of providing 24 frames per second as the sampling rate (Etoh et al., 2001). Therefore, vital data recognition is possible.

Resolution: Higher pixel density is often correlated to better video quality. Since each camera has its own parameter set (screen size, field-of-view, etc.), we have to focus on the resolution of the face itself and not on the entire screen. Since we are focusing on the change in color of the green channel, the resolution of the ROI is relevant but is not of substantial importance. The number of pixels within the ROI might be 100, 1000, or even higher but is not the defining quality parameter. Hence, the low resolution of a standard video graphics array (VGA) video is sufficient for pulse recognition (Mestha, 2014).

Automatic functions: For analyzing the change in color within the region of interest, stable recording is required. Some cameras perform automatic white calibration or brightness adjustment for enhanced imaging (Weng, 2005). Due to discontinuity or changes in color, the automatic functions affect the vital data recognition and lead to errors or additional noise.

5.2.2.2 Environment Parameters

Vital data recognition via reflective PPG approaches with cameras works well in laboratory settings (Irani, 2014). Therefore, it has to be considered that real-world scenarios with social robots might involve additional challenges. These can be classified as follows:

Motion artifacts: The image sensor of the robot is not mounted in a fixed frame. Therefore, the camera might experience vibration caused by a cooling fan, a power transformer, or by the motions of the robot itself. Furthermore, the communicating counterparts' movements while speaking or performing natural body language have to be considered.

Optical considerations: During a conversation, the spatial constellation between robot and dialog partner might vary due to the change of distance or the optical angle. Hair or makeup might cover the region of interest. Moreover, glasses worn by the user might disturb the face recognition algorithm.

Light source: The communication between a social robot and a human can take place in an indoor environment. In that case, the brightness of the light and the light source itself may lead to signal noise or disturbances. Artificial light sources particularly influence the signal noise.

Temperature: The sensing of changes in color depends on the perfusion of the skin. In addition to this, the temperature of the environment also influences the blood flow. Other effects include physical parameters of the user (e.g., skin flexibility, drugs, coffee, etc.).

A camera-based touchless vital data recognition system must be aware of the influencing parameters. Moreover, the recognition system has to have implemented algorithms for identifying the major disturbances in order to adapt.

5.3 Detection Algorithm

In order to measure the pulse signal, a video stream has to be captured by a camera first. Therefore, it is necessary to detect the human's face, then identify the forehead and a suitable part of it (ROI). Subsequently, we determine the average intensity of the green color channel within the red-green-blue (RGB) signal in the ROI. This signal is the basis for recognizing the pulse wave. Therefore, we first need to track the face within the video stream. To accomplish this, the face-tracking algorithms of the OpenCV library can be applied (Bradski & Kaehler, 2008). In order to reduce

motion artifacts, one could perform face tracking on every frame. As this would result in reduced performance, the face-tracking frequency is reduced (once every 25 frames) and a larger main ROI is defined instead. This main ROI has the size of the whole forehead with another smaller region inside of it. This inner region moves with the head movement in each frame. This way, it follows the movement of the forehead without constant face tracking. This means a low requirement for computational power and ensures a stable sampling and frame rate.

Before evaluating the values of the ROI, it is necessary to remove motion artifacts with the help of filtering. Therefore, we check each pixel inside the ROI to determine if their green value was an outlier compared to the averaged value of the ROI from the preceding frame. Outliers are defined as pixel values that are beyond the $3 * \sigma$ (standard deviation) threshold within one frame.

Let now $f(t)$ be the recorded raw signal of the ROI at a time t and $m_{f(t)}$ the mean of $f(t)$. For the detection of outliers, we use low-pass and high-pass filters with sliding window. Butterworth filtering is also an option. For sliding window, we are not concentrating on just the current frame of the footage but on the mean of the last four frames.

Let ROI be a frame of $p * q$ pixels. Then the mean of one frame is calculated as:

$$m_{f(t)} = \frac{1}{p * q} * \sum_{i=0}^p \sum_{j=0}^q f(t)[i, j].$$

With these values, we can apply our filters:

- Low-pass filter:

$$m_{l(t)} = m_{f(t-1)} * (1 - \alpha) + m_{f(t)} * \alpha, \quad \text{with } \alpha = 0.05$$

- High-pass filter:

$$m_{h(t)} = m_{f(t)} - m_{l(t)}$$

Finally, we apply the sliding window to the filtered value to get our filtered mean value:

$$m = \frac{1}{4} * \sum_{i=0}^3 m_{h(t-i)}.$$

If a Butterworth filter is applied, we recommend cutting off frequencies below and above normal heart rates (sampling rate = frames per second (FPS), lower cutoff frequency = 0.52, upper cutoff frequency = 5.02).

After removing all outliers, the average green value of all leftover pixels is determined.

The resulting pulse curve allows us to determine a reliable pulse signal (Fig. 5.5). Subsequently, a fast Fourier transformation and peak detection serve to identify the heart rate and heart rate variability.

5.4 Optimization Strategies

Social robots usually apply cost-efficient cameras. These customary web cameras provide moderate resolutions and frame rates. They are usually optimized for video conferencing and reduced data traffic. In contrast to this, heart rate or heart rate variability detection scenarios require a focus on image quality.

Motion within a video sequence leads to major artifacts in the heart rate signal. Therefore, face recognition and head tracking technologies support the readjustment of the ROI and the assessment of a change in color. Furthermore, reference regions allow motion compensation as well as general changes in the color or brightness. Therefore, reference regions in the face might compensate for automatic functions or may stabilize the lighting situation (Fig. 5.5).

The change in color within the ROI leads to a periodic signal that corresponds to the heart rate, as presented in Fig. 5.6.

The frame rate of most customary cameras is sufficient for vital data recognition since less than 10 Hz are needed for sampling the heart rate and heart rate variability, or for providing respiration rate recognition (RR).

Natural daylight provides almost white light that consists of a sufficient amount of green light for our study. Furthermore, daylight is a continuous light source and provides a setting for very good measurements. In contrast, artificial light highly influences the recorded data and produces signal noise.



Fig. 5.5 Region of interest (blue) and reference region (red)

The analysis of artificial light in our measurements showed a tremendous change in brightness for higher frequencies. The normal power supply of the lights in our lab (located in Germany) is alternating current (AC) with a frequency of 50 Hz. This means that the voltage changes polarity 50 times per second. Thus, the light gets brighter with the maximum voltage and less bright in the zero-crossing zone. The zero-crossing happens 100 times per second so that the lights have a pulsation of 100 times per second. The intensity of the maximum brightness and the least brightness depends on the light technology. A neon light loses 50% of its intensity during the zero-crossing (Brundrett, 1974). Modern LED lights are affected even more by the pulsating current than neon lights or standard bulbs. The pulsating effect is also dependent on the ballast unit used. Summing up almost every artificial light pulsates (Fig. 5.7).

The changing brightness leads to an aliasing effect and influences the quality of data. In contrast to Figs. 5.4 and 5.5, Fig. 5.8 illustrates the high noise effect on the red and green zones caused by the pulsation of artificial light.

During our research with robots, we applied the robots' camera for heart rate and heart rate variability recognition. This research was performed mainly in indoor environments. We identified the surrounding light source and implemented a filter to minimize the aliasing effect. By using a digital filter, we could reduce the noise due to the very frequent changes in brightness. Furthermore, we modified the environmental lighting as soon as we noticed that the light conditions were insufficient for our measurements.

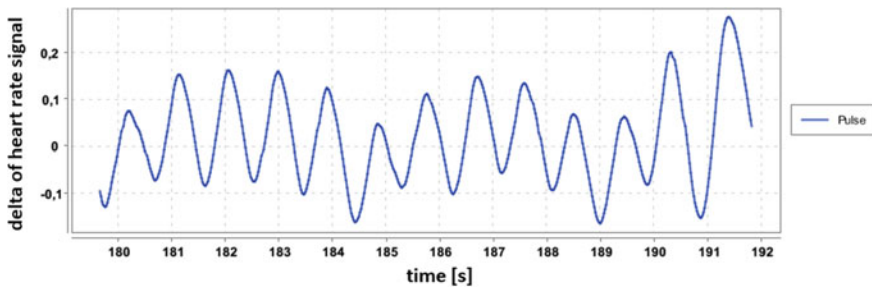


Fig. 5.6 Heart rate signal of the subject

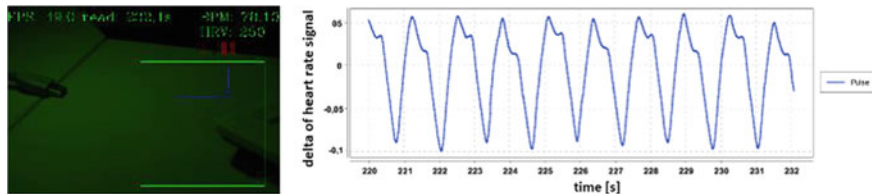
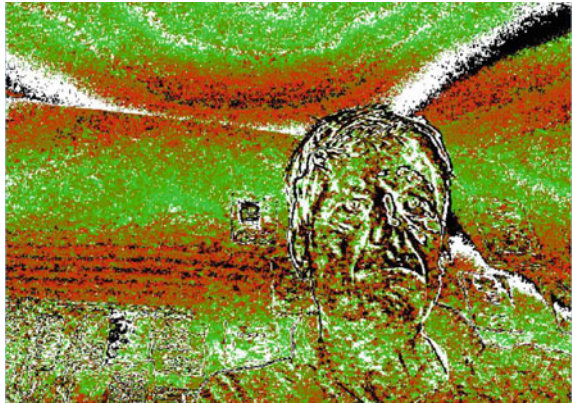


Fig. 5.7 Pulsating green LED light (left) and the received camera data (graph on the right)

Fig. 5.8 Aliasing effect of the signal due to the pulsation of artificial light



5.5 Study

In order to evaluate the performance of camera-based human–robot interaction, we conducted a lab study with eight participants. The study included two different camera systems: a Philips SPC 1300NC (webcam) and an IDS UI-306xCP-C (professional camera). The webcam was integrated into our social robot (Fig. 5.1) as an eye. The social robot was designed by us and originally used as a physical avatar (Sauer & Gobel, 2003). Although it was possible to measure a reliable pulse wave with the Philips camera and the social robot, we mostly applied the IDS since we could store video streams for later processing and analysis at higher resolutions and frame rates.

The setting for the study was a normal office workplace; the IDS camera was mounted on the monitor. The participants were advised to work at the computer. The average measurement duration was 15 min. The participants had to behave normally, as if they were not being recorded.

Our studies had two main purposes:

To identify which light intensity provides the highest heart rate accuracy

To estimate the percentage of time in which valid data is measurable using the heart rate recognition algorithm provided.

In total, we recorded about 100,000 samples under various light conditions. We found out that the heart rate accuracy is highly dependent on the brightness of the surrounding light (Fig. 5.9). Lower light intensity results in more dominant noise, which leads to varying light and color data. On the other hand, light which is too intense results in total reflectance and therefore overexposure of the skin. Optical saturation hinders a change in light intensity due to overexposure. The trend line (Fig. 5.7) indicates that, in order to achieve optimal results, the most useful light intensity is in the upper quartile.

In our study, we also investigated the amount of time required for valid identification of the pulse rate during camera surveillance. Therefore, we measured the total amount of time with valid and invalid pulse rates for all subjects. The subjects had to

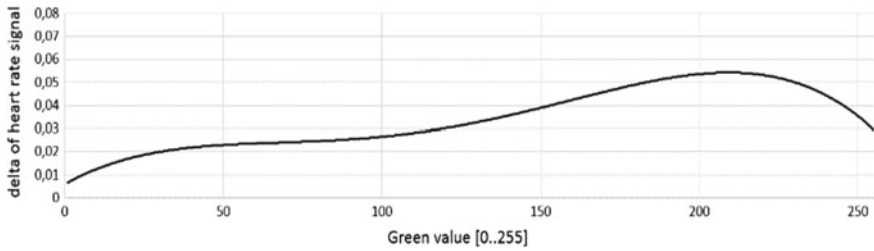


Fig. 5.9 Trend line of the delta of the heart rate signal (y) proportional to the measured green value of the pixels (x)

behave naturally while performing normal computer work. They were allowed to go to the restroom, talk to colleagues, or read printouts. In our tests, we found that for 50% of the time valid pulse detection was possible. 33% of the pulse detection was invalid because the OpenCV algorithm was not able to detect a face. This happened because the subject was unavailable, or the face could not be detected due to rotation, movement, or bad exposure. Unrealistic pulse rates were detected in the remaining 17% of the time and therefore excluded.

We also tested a scenario in which the subjects had to keep the head motionless while facing the camera the whole time. We identified that by using the aforementioned restriction under good light conditions, the pulse recognition was valid 95% of the time. Of course, keeping the head motionless is not a reasonable scenario for real-life applications since the many tasks one performs involve plenty of head motions. In addition to this, some of the subjects reported neck pain after several minutes.

5.6 Discussion

We were able to identify that face-to-face communication between a robot and a human enables a direct view of the subject's forehead. During our research with robots, we applied the robots' camera for heart rate and heart rate variability recognition. We found out that social robots could measure stress, strain, emotions, or medical parameters. This leads to the question of in which social situations this technology should be used. We think that social robots perfectly meet the care demands for elderly who are lonely or suffering from dementia. With the increasing potential of artificial intelligence, social robots will become very useful for entertainment but also as acquaintances or even friends. Their emotion detection leads to better understanding of the human by the robot, though of course, we hope that a robot will never be a better friend than a human is. The capability of vital data detection may also be very useful in hospital or care environments, so in future, rather than impersonal systems, nice robots will monitor patients.

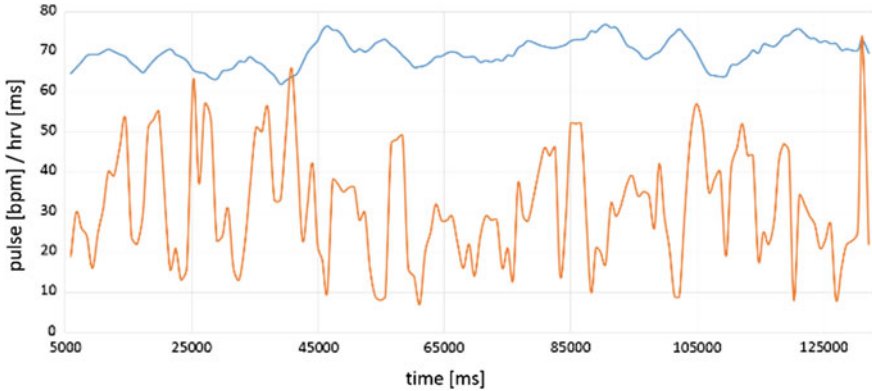


Fig. 5.10 Detected heart rate variability (HRV, red) and pulse rate (blue) via a camera-based reflective PPG approach

Unfortunately, this technology may have some social implications. Some people might feel uncomfortable with robots collecting their vital signals (either through or without touch). Furthermore, we can imagine that companies will use social robots to perform job interviews. The robot could ask specific questions and act like a polygraph, a lie detector. This scenario is also possible in a medical setting where doctors using social robots to obtain true answers from patients during the anamnesis. We should be aware that we are giving a further piece of capability to a robot that only a human had before. This might lead to the circumstance that a human can be assisted but can also be replaced.

Improving the face-tracking algorithm and lighting would greatly increase the amount of valid heart rate values. Our study shows that the recognition of heart rate and heart rate variability (Fig. 5.8) is possible. Therefore, camera-based vital data recognition allows touchless emotion recognition. In our study, we achieved assessment of a valid pulse rate for only half the time the measurement was performed. This is only a very rough estimation but indicates that the concept has a high potential and can be improved (Fig. 5.10).

We consider artificial light sources as well as movement artifacts and brightness change as the main noise in vital data recognition. A possible improvement might be achieved by transforming the RGB data into another color space, e.g., hue-saturation-value color space (HSV). This is currently under examination.

A simple recording of the environment in slow motion (e.g., iPhone 6s with 240 frames per second) demonstrates the varying light conditions. Social robots might also illuminate the person with whom they are communicating in the future. Furthermore, robots might use the invisible light spectra or infrared to extend their scanning possibilities.

The authors of “Emotion recognition using bio-sensors: First steps towards an automatic system.” (Haag, 2004) states that wearing biosensors is less disturbing

than being “watched” by a camera. We think that a friendly-looking social robot that is interacting with the interlocutor is not perceived as annoying or indiscrete.

5.7 Conclusion

In order to achieve natural communication between social robots and humans, important modalities have to be addressed. In the process of communication, social robots might apply a camera to identify the heart rate, heart rate variability or respiration rate of a user to enable the detection of emotional states. Since social robots will mostly be used indoors and in the homes of users, many sources of noise and disturbance might affect the camera-based vital data recognition. As one of the main noise factors, we identified the artificial light that surrounds the social robot. Aliasing filters can be used to reduce that noise in combination with an adapted frame rate to avoid side effects.

A sensitive conversational partner is capable of reacting to changing emotional states during a conversation. Our approach involves the integration of sensitivity in order to measure and understand the feelings of the interlocutor. In future applications, we envision social robots changing their facial expressions or skin color according to their emotional state to enable an exchange of emotional states with other social robots and humans. We are aware of the fact that social robots might receive more information about the emotional state than the interlocutors may want. This leads to interesting future scenarios that might involve social robots in job interviews, patient anamneses, and social or chaplain tasks, or even in polygraph (lie-detection) applications.

Our future work will focus on improving the vital sign recognition with cameras as well as the natural communication between social robots and humans.

References

- Baer, R. L. (2000). CCD requirements for digital photography. In *IS and TS Pics Conference, Society for Imaging Science & Technology*, pp. 26–30.
- Bieber, G. H. (2013). Sensor requirements for activity recognition on smart watches. In *Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments (PETRAE'13)*, ACM.
- Biggs, A. (2005). Paro therapeutic robot seal. Published on Flickr as CC-BY-SA-2.0, <http://www.flickr.com/people/ehjayb/ehjayb>.
- Bradski, G., & Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Inc.
- Brundrett, G. W. (1974). Human sensitivity to flicker. *Lighting Research & Technology*, 6(3), 127–143.
- Clancy, E. A. (2002). Sampling, noise-reduction and amplitude estimation issues in surface electromyography. *Journal of Electromyography and Kinesiology*, 12(1), 1–16.

- Etoh, T. G., Takehara, K., Okinaka, T., Takano, Y., Ruckelshausen, A., & Poggemann, D. (2001). Development of high-speed video cameras. In *SPIE 4183, 24th International Congress on High-Speed Photography and Photonics*.
- Goodrich, M. A. (2007). Human-robot interaction: a survey. *Foundations and Trends in Human-Computer Interaction*, 1(3), 203–275.
- Haag, A. G. (2004). Emotion recognition using bio-sensors: First steps towards an automatic system. In *Tutorial and Research Workshop on Affective Dialogue Systems* (pp. 36–48). Berlin, Heidelberg: Springer.
- He, C. Y. (2017). An emotion recognition system based on physiological signals obtained by wearable sensors. *Wearable Sensors and Robots* (pp. 15–25). Singapore: Springer.
- Hertzman, A. B. (1937). Photoelectric plethysmography of the fingers and toes in man. *Proceedings of the Society for Experimental Biology and Medicine*, 37(3), 529–534. <https://doi.org/10.3181/00379727-37-9630>.
- Inan, O. T. (2015). Ballistocardiography and seismocardiography: A review of recent advances. *IEEE Journal Biomedical and Health Informatics*, 19(4), 1414–1427.
- Irani, R. N. (2014). Improved pulse detection from head motions using DCT. In *2014 International Conference on Computer Vision Theory and Applications (VISAPP)* (pp. 118–124). IEEE.
- Jerritta, S. M. (2011). 7th International Colloquium on physiological signals-based human emotion recognition: A review. In *Signal Processing and its Applications (CSPA)* (pp. 410–415), IEEE.
- Kong, L. Z. (2013). Non-contact detection of oxygen saturation based on visible light imaging device using ambient light. *Optics Express*, 21(15), 17464–17471.
- Korn, O. B. (2018). Perspectives on social robots: from the historic background to an experts' view on future developments. In *Proceedings of the 11th Pervasive Technologies Related to Assistive Environments Conference*.
- Matthes, K. (1935). Untersuchungen über die Sauerstoffsättigung des menschlichen Arterienblutes. *Naunyn-Schmiedebergs Archiv für experimentelle Pathologie und Pharmakologie*, 179(6), 698–711.
- Matthies, D. J. C., Haescher, M., Bieber, G., Salomon, R., & Urban, B. (2016). SeismoPen: Pulse recognition via a smart pen. In *International Conference on Pervasive Technologies Related to Assistive Environments, PetraE 2016*, ACM, Corfu Island, Greece. <https://dx.doi.org/10.1145/2910674.2910708>.
- Mestha, L. K. (2014). Towards continuous monitoring of pulse rate in a neonatal intensive care unit with a webcam. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 3817–3820).
- Nath, A. R. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *Neuroimage*, 59(1), 781–787.
- Poh, M. Z. (2011). Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering*, 58(1), 7–11.
- Rohracher, H. (1964). Microvibration, permanent muscle-activity and constancy of body temperature. *Perceptual and Motor Skills*, 19(1), 198–198.
- Sauer, S., Gobel, S. (2003). Focus your young visitors: kids innovation–fundamental changes in digital edutainment. In *Museums and the Web 2003: Selected Papers from an International Conference (7th, Charlotte, NC, March 19–22, 2003)*.
- Smith, W. (2005). *Modern lens design*. Publisher McGraw-Hill Professional Engineering.
- Smith, W. (2007). *Modern optical engineering* (4th ed.). Publisher McGraw-Hill Professional Engineering.
- Weng, C. C. (2005). A novel automatic white balance method for digital still cameras. In *IEEE International Symposium on Circuits and Systems, ISCAS 2005* (pp. 3801–3804).
- Wu, H. Y. (2012). Eulerian video magnification for revealing subtle changes in the world. <https://doi.org/10.1145/2185520.2185561>.
- Zissman, M. A. (1996). Comparison of four approaches to automatic language identification of telephone speech. *IEEE Transactions on Speech and Audio Processing*, 4(1), 31.