

ROI Detection and Tracking for Physiological Monitoring Based on Calibration between RGB-D and Thermal Cameras

Feng Zhao¹, Serhan Cosar², Nicola Bellotto², and Shigang Yue²

¹ Department of Computer Science, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, United Kingdom, f.zhao@ljmu.ac.uk

² School of Computer Science, University of Lincoln, Lincoln, LN6 7TS, United Kingdom, {scosar,nbellotto,syue}@lincoln.ac.uk

Abstract. The subject’s body movements are still the main challenge to physiological monitoring using thermal imaging. Traditional thresholding, edge detection, and tracking techniques can handle small head motions, but have difficulty in coping with large movements. In this paper, we propose a calibration method integrating the RGB-D camera and thermal camera to allow free body movements in practice. Especially, two smart schemes are designed to make the chessboard pattern with black and white squares visible to both cameras during the calibration process. To accurately evaluate the facial temperature evolution of a subject, the RGB-D sensor is utilized to guide the localization of the ROIs in the thermal data by mapping the extracted face and landmark points of the RGB image to the thermal image. As a result, we achieve accurate and robust ROI detection and tracking, thus leading to better estimation of physiological parameters.

Keywords: Physiological Monitoring · Calibration · Thermal Imaging.

1 Introduction

The thermal infrared sensor [15] is capable of recording small facial temperature variations because of its high sensitivity to thermal radiations emitted from the skin. Compared with traditional contact and wearable devices [25, 27], the thermal imaging process can be carried out in a contactless and unobtrusive way without discomfort. More importantly, the collected thermal data can be used to measure the physiological parameters such as the body temperature, respiration rate, and heartbeat rate. Thus, thermal imaging makes continuous health monitoring of human beings convenient and comfortable but not noticeable, especially benefiting the neonatal infants, children, and elderly people.

Many approaches have been proposed for the physiological monitoring based on thermal imaging [2, 6, 8, 11, 16, 23]. The first key step is to identify the regions of interest (ROIs) like forehead, nose, mouth, cheeks, ears, and neck from the detected face, and track them in each frame image appropriately. Usually, the mean thermal signal of the nostril region carries the breathing information, and

the average temperature change along the superficial blood vessel in the neck area reveals the heartbeat information. The accurate and robust segmentation of these ROIs is crucial to the overall performance. The main challenge comes from the head movements of the subjects. In most previous works [4, 6–8, 11, 23, 28, 29], the subjects were instructed to stay still during the course. However, they usually exhibit small head motions. More practically, large movements can often be found.

To overcome the inaccuracy caused by small movements of the subjects, Fei et al. [12] applied a tracker based on iterative image registration [20] to the nasal tip area. Aiming at automatically following the breathing signal, a robust and stable Bayesian tracker utilizing the mean shift localization (MSL) [14]-based particle filtering was proposed in [30]. This tracker can cope with significant head movements, but it fails in the scenario that more than one window of the tracking ROI is lost or blocked during the process. In addition, the particle filtering is computational expensive. In [11, 22], the authors adopted the coalitional tracking algorithm [10] to deal with the subject’s motion during the breathing measurement. It employs a cluster of collaborative particle filter trackers and optimizes the multi-tracker interaction through game theory to gain both accuracy and robustness. It is worth noting that the behavior of the coalitional tracker is not absolutely deterministic due to the stochastic components of the particle filter trackers. The nostril ROI is then extracted by conducting integral projections on the edge image obtained from the original thermal face image applying a 3×3 Sobel edge detector. The leftmost and rightmost maxima of the vertical integral projection generate the respective left and right outer edges of the nostril, and the peak of the horizontal integral projection produces the nostril’s base edge.

To actively track the mouth-nose region, Chauvin et al. [6] modified the tracking, learning, and detection (TLD) predator approach [17] to make it adaptive to the thermal data. With a manually initialized ROI, the TLD algorithm produces the coordinates and size of the bounding box for each frame using a minimum set of features. If the tracker module loses the ROI, the detector module will reinitialize the tracker with a new bounding box obtained by an exhaustive search using a sliding-window method. Therefore, this technique can handle relatively large head movements. Similarly, they implemented the 3×3 Sobel operator to extract the contours of the mouth-nose ROI and then yielded the mask by setting all the pixels with a gradient value less than the specified threshold to zero. The ROI is attained by employing the mask operation on the contour image.

In [29], Yang et al. manually initialized a region for the first frame, which is then divided into a series of nested regions (R^i) based on an incremental temperature step (ΔT) and a reference temperature (T_{ref}) which is the mean temperature of all the pixels on the bounding box of the initialization region. By selecting the optimal threshold ($T^* = T_{ref} + i^* \Delta T$) that makes the area of region R^{i^*} drop sharply against the area of region R^{i^*-1} , the ROI of the first frame is chosen as the region with largest connected component in R^{i^*} . Correspondingly,

the ROIs of the subsequent frames are attained by employing the same T^* to compute the isotherm and tracking the contour utilizing energy minimization.

In practice, traditional thresholding (e.g., Otsu’s adaptive thresholding [4, 19, 23]) and edge detection (e.g., Prewitt operator [1–3]) are also widely used for the ROI segmentation. In some investigations [5, 16, 28], the ROIs are even manually placed to ensure the high accuracy of segmentation, which is time-consuming and susceptible to subjective errors.

Ideally, the person under estimation should be able to move around without constraints. To cope well with such free head movements, in this work, we propose to calibrate the thermal sensor with the RGB-D sensor for producing more reliable ROI segmentation results, taking advantage of the RGB-D imaging in face detection and landmark points extraction.

2 Methodologies

Most of previous approaches for measuring the physiological data generally work under certain environments such as temperature controlled room, subjects sitting comfortably in a chair (remaining still or with small movements), facing to the camera, and within a specific distance from the camera. Sudden temperature changes and body movements are the main challenges to the physiological monitoring. Although researchers investigated different methods [2, 4–6, 11, 16, 22, 23, 28, 29] to alleviate small head motions of the subjects, large or free head movements in real applications are still a major factor affecting the overall accuracy. The direct segmentation of the thermal images using thresholding and face landmark points detection applying the Dlib algorithms [18] usually produce results with noises. To accurately evaluate the facial temperature evolution of the subjects, a RGB-D camera can be potentially utilized to guide the localization of the ROIs in the thermal data. This necessitates a stereo calibration between the thermal and RGB-D sensors using a chessboard pattern with black and white squares.

As illustrated in Fig. 1, the thermal and RGB-D sensors are mounted on top of a robot at 1.3 m and 1.1 m from the floor, respectively. The comfortable distance between the robot and the subject is about 1.0 m. The chessboard can be easily detected by the RGB camera, however the black and white squares cannot be differentiated by the thermal camera as their heat radiations are more or less the same. To make the chessboard pattern visible to both sensors during the calibration process, we design two smart schemes.

2.1 Calibration using combined chessboard

The 8×6 chessboard pattern with 23.5 mm black and white squares is used for the calibration, where all the black squares in a thick white cardboard of A4 paper size are first removed, and then heated using a 300 W incandescent lamp for several minutes (see Fig. 2). At the same time, a thin black sheet of the same size is cooled down in a fridge for a few minutes to bring it to a low



Fig. 1. RGB-D and thermal sensors mounted on top of a robot.

temperature. The two sheets are then combined together so that both cameras can see a black and white chessboard pattern, as the white cardboard and black sheet appear as white and warm, and black and cold in the RGB-D camera and thermal camera, respectively (see Fig. 3).

2.2 Calibration using electrically-heated chessboard

The combined chessboard is visible to the thermal camera for at most 5 minutes as the temperature difference between the white board and black sheet diminishes, and the operation procedure is not very convenient. Sometimes, the heating

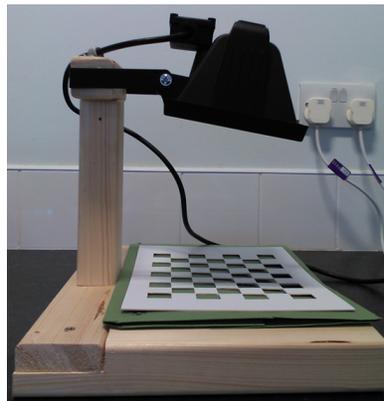


Fig. 2. Lamp for heating the white board.

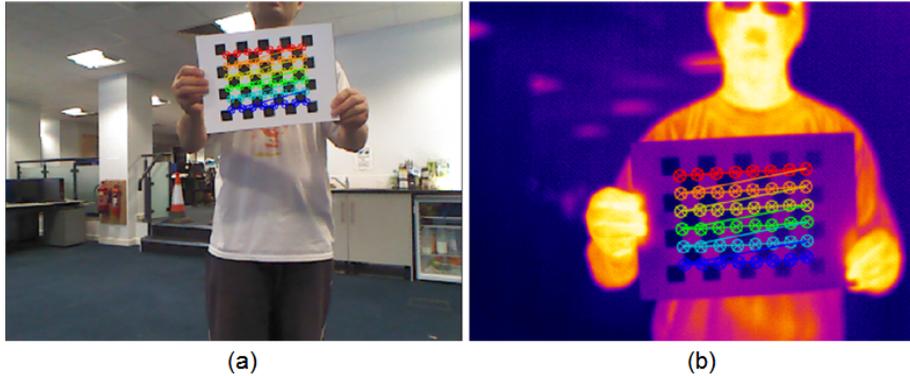


Fig. 3. Calibration of the RGB-D camera and thermal sensor using a combined chessboard. (a) Detected pattern in RGB image and (b) detected pattern in thermal image.

and cooling processes need to be carried out twice for one calibration. Each time before calibration, it also takes some time to finish the heating and cooling. To overcome these shortcomings, we design an electrically-heated chessboard, which consists of a 3 mm black acrylic sheet at the back and a 1.5 mm bi-color (black on white) laminate sheet at the front. The chessboard pattern is manufactured by precisely carving the black squares using a laser cutter. Each corner of the white and black squares is then associated with a resistor. When the power is on, the resistors emit radiations, which can be detected by the thermal camera, as depicted in Fig. 4.

In comparison with the combined chessboard, the calibration procedure by this pattern can be conducted immediately after turning on the power. Moreover, the resistors can be heated for a much longer time, enabling to perform multiple calibrations continuously. It is worthy to note that the resistors may be overheated after too long time (e.g., more than 30 minutes), which could affect the calibration results.

3 Experimental Results

As demonstrated in Fig. 1, the RGB-D camera used in the experiments is Microsoft Kinect V1 [21] with an 8-bit RGB resolution of 640×480 pixels and an 11-bit monochrome depth of the same resolution. Its frame rate is 9-30 Hz. The higher the resolution, the lower the frame rate. In addition, the sensor has a practical distance ranging from 1.2 m to 3.5 m, and an angular field of view (FOV) of 57° (horizontal) and 43° (vertical). Meanwhile, the employed thermal sensor is Optris PI 450 [24], which has a high sensitivity to temperature (0.04°C) and an image resolution of 382×288 pixels. Its frame rate ranges from 27 Hz to 80 Hz. The sensor lens has a relatively narrow FOV and a fixed focus length (38° horizontally and 29° vertically with 15 mm lens).

It takes about 3-5 minutes to complete the calibration process, where 80 or 150 sets of 48 corner points are recorded. Based on the pinhole camera model,

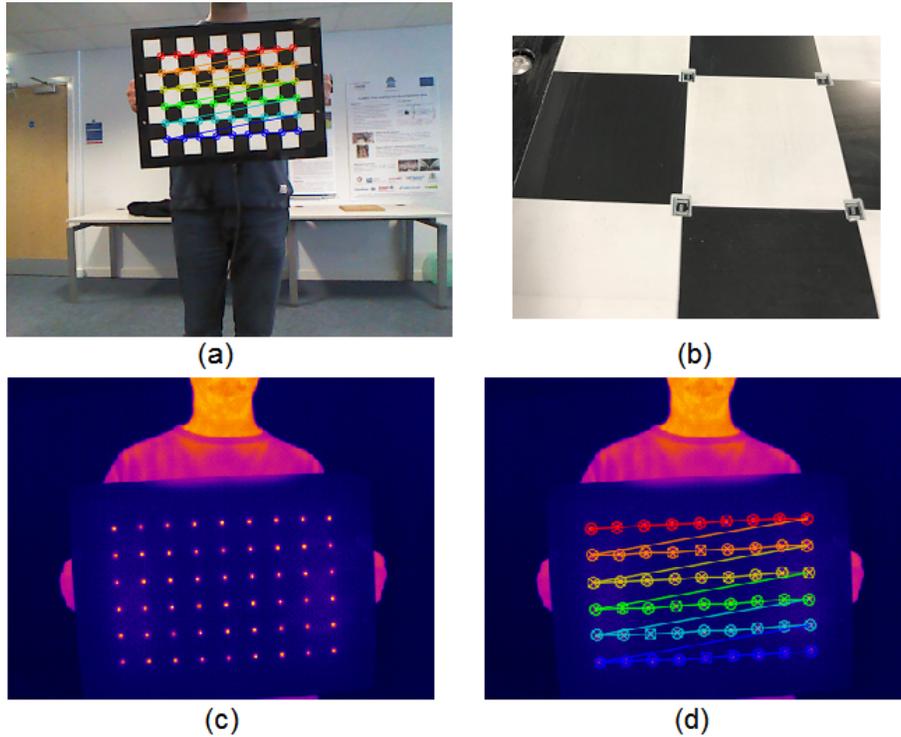


Fig. 4. Calibration of the thermal camera with the RGB-D sensor using an electrically-heated chessboard. (a) Detected pattern in RGB image, (b) resistors for heating the chessboard, (c) thermal image of the heated chessboard, and (d) detected pattern in thermal image.

the intrinsic parameters of both cameras together with the rotational and translational transformation between cameras are estimated via stereo calibration. Face detection and landmark points extraction in the RGB images are obtained by utilizing the Dlib machine learning toolkit [18] based on the histogram-of-oriented-gradient (HOG) features [9]. Since the target person is assumed not to move significantly between two consecutive frames, the face detection task is performed in a local window corresponding to the face previously detected, thus speeding up the detection process. If this cannot be done, the face detection is carried out completely in the new frame image.

As a consequence of the calibration process, the translation vector and rotation matrix are utilized to map the detected face and landmark points in the RGB image to the thermal image (see Fig. 5). It can be seen that the person can move around freely and the calibration results are very encouraging. The robust detection results in the thermal data lead to accurate ROIs, which will produce much better physiological measurements.

As described in Fig. 6, the facial ROIs (forehead, nose, and cheeks) in the thermal image can be properly located with respect to the calibrated landmark



Fig. 5. Calibration results overlaying on the RGB and thermal images in three orientations.

points. For instance, the forehead center can be considered as the middle of the two eyebrow corner points. Accordingly, the average temperature in the forehead area is computed as the body temperature. The evolutions of the mean temperature in the nose and cheek areas along the timeline are used for the measurement of the respective respiration and heartbeat rates applying harmonic analysis. Fig. 7 displays an example of the physiological monitoring results by fast Fourier transform (FFT) [6, 16], where the dominant frequency with the highest peak magnitude in the temperature signal's spectrum is identified, and then multiplied by 60 to obtain the respiration or heartbeat rate in cycles per minute. Compared with the ground truth data obtained by a contact thermometer and a wearable oximeter, the calculated body temperature error is within 1°C , the estimated respiration rate of 15 breaths per minute (BPM) is accurate, while the measured heartbeat rate of 95 beats per minute (BPM) needs to be further

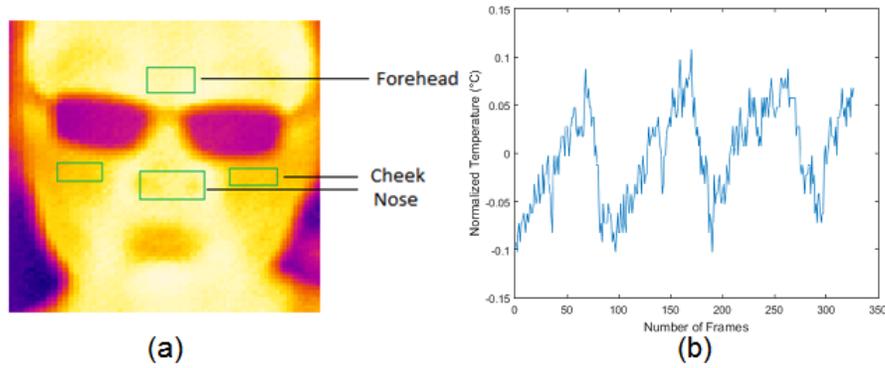


Fig. 6. Physiological monitoring. (a) Facial ROIs and (b) normalised temperature signal of the nose ROI.

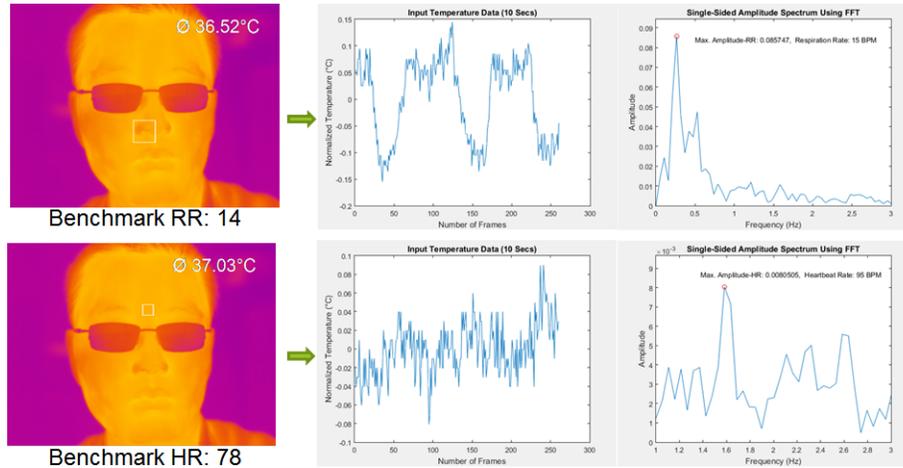


Fig. 7. Measured body temperature, respiration rate (RR), and heartbeat rate (HR) from different facial ROIs based on 10-sec thermal data.

improved. The relatively big bias of the heartbeat is because of its larger variation range than that of the respiration, and it can be reduced by denoising the input temperature signal.

4 Discussion

During the calibration, two conditions need to be satisfied: i) both cameras detect a complete chessboard, and ii) all the corner points on the chessboard can be identified by both cameras. In presence of more than one subjects, the calibration procedure can still be done, as demonstrated in Fig. 8. Furthermore, the calibration approach can address the issue of prohibiting subjects from wearing glasses mentioned in many existing methods [1–5, 30]. Nevertheless, we noticed that the calibration can only be successful when the chessboard is placed within

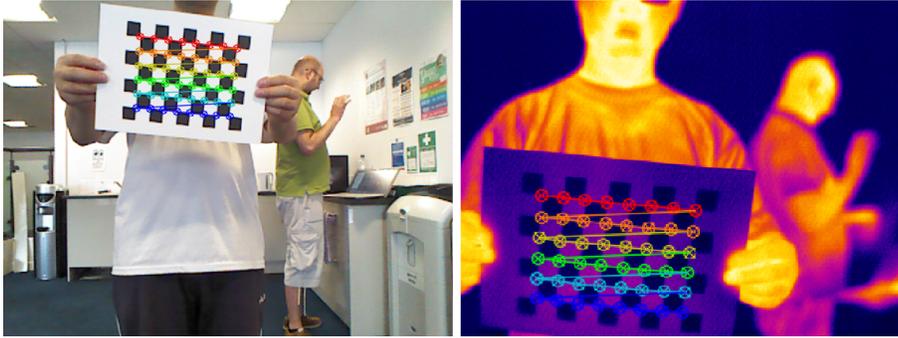


Fig. 8. Calibration process in presence of multiple subjects.

an effective distance range of 60-90 cm because of the different FOVs of the two cameras, especially the limited FOV of the thermal sensor. This means that many adjustments have to be conducted to match the FOV of the thermal camera with the other RGB camera (Kinect V1 in this work). In practice, having a motorized pan-tilt unit (PTU) to mount the two cameras could help with this difficulty.

Another reason is that the lens of the thermal camera has a fixed focus length, which means that the acquired image is very clear at a specific distance range, but becomes a bit blurring outside of this range. For average temperature measurement within a ROI, it is not really a problem since it already acts as a low-pass filter, but this could be an issue for extracting a very small feature (e.g., chessboard corner points) on a wide range of distance. Since thermal cameras of high cost with lens of adaptive focus lengths are available on the market, this problem can be alleviated.

To speed up the calibration, a chessboard with bigger square size (e.g., 6×3 with 35.0 mm squares) can be implemented. Particularly, it is more convenient and faster to complete the calibration procedure using the combined chessboard. Moreover, the range of the effective distance between the chessboard and the sensors is enlarged to 70-150 cm. Additionally, the frame rates of the RGB and thermal cameras may not coincide with each other. However, this will not greatly influence the ROIs detection and tracking, as the subjects usually keep still within a very short period of time. Thus, the same calibration results are valid for a few consecutive frames. In real applications, this can also be used to accelerate the calibration process.

Due to the narrow FOV and fixed focus length of the thermal sensor, the effective distance between the subjects and the cameras for physiological monitoring after calibration is within a limited range as well. For the 8×6 chessboard, it is about 1.0 m. In case that the 6×3 chessboard is used for calibration, this distance becomes 1.0-1.1 m. For both cases, the subject's moving forward or backward from the sensors will make the calibration results on the thermal data shifted (see Fig. 9). This problem can be solved by keeping the distance between

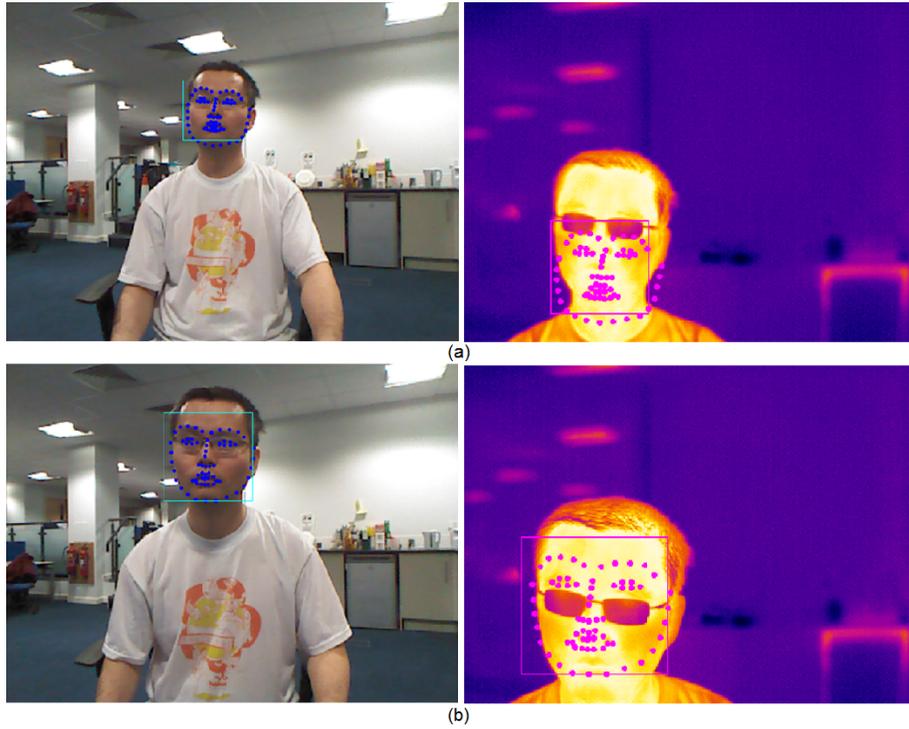


Fig. 9. Deviations of face landmark points on thermal images when the subject moves (a) backward and (b) forward.

the subject and the cameras through robot navigation, or employing a costly thermal sensor with multiple lens.

5 Conclusion

With the advent of portable and affordable thermal cameras of high performance (e.g., FLIR ONE [13] and Seek CompactPRO [26]) that can be easily attached to smartphones, thermal imaging is becoming more and more popular and it plays an important role in contactless physiological monitoring. Taking advantage of both the RGB-D and thermal imaging, we develop a calibration approach to tackle large or even free head movements occurring in thermal-based measurements of physiological parameters. The experimental results demonstrate that the calibration of thermal sensor with RGB camera is feasible and the proposed two calibration strategies benefit the ROIs segmentation and tracking, thus leading to more promising estimation results than existing methods. In future work, we will apply the circle-pattern chessboard instead of traditional square-pattern chessboard to further improve the calibration accuracy and finally produce better physiological monitoring results.

6 Acknowledgement

This work was fully supported by the EU H2020 project ENRICHME (no. 643691, <http://www.enrichme.eu>).

References

1. AL-Khalidi, F., Saatchi, R., Burke, D., Elphick, H.: Tracking human face features in thermal images for respiration monitoring. In: Proc. IEEE/ACS Int. Conf. Computer Systems and Applications. pp. 1–6 (2010)
2. AL-Khalidi, F., Saatchi, R., Elphick, H., Burke, D.: An evaluation of thermal imaging based respiration rate monitoring in children. *American J. Engineering and Applied Sciences* **4**(4), 586–597 (Jan 2011)
3. AL-Khalidi, F., Saatchi, R., Elphick, H., Burke, D.: Tracing the region of interest in thermal human face for respiration monitoring. *Int. J. Computer Applications* **119**(4), 42–46 (Jun 2015)
4. Alkali, A.H., Saatchi, R., Elphick, H., Burke, D.: Facial tracking in thermal images for real-time noncontact respiration rate monitoring. In: Proc. European Modelling Sym. pp. 265–270 (2013)
5. Alkali, A.H., Saatchi, R., Elphick, H., Burke, D.: Short-time Fourier and wavelet transform analysis of respiration signal obtained by thermal imaging. In: Proc. Int. Sym. Communication Systems Networks and Digital Signal Processing. pp. 183–187 (2014)
6. Chauvin, R., Hamel, M., Brière, S., Ferland, F., Grondin, F., Létourneau, D., Tousignant, M., Michaud, F.: Contact-free respiration rate monitoring using a pan-tilt thermal camera for stationary bike telerehabilitation sessions. *IEEE Systems Journal* **10**(3), 1046–1055 (Sep 2016)
7. Chekmenev, S.Y., Farag, A.A., Essock, E.A.: Multiresolution approach for non-contact measurements of arterial pulse using thermal imaging. In: Proc. IEEE Conf. Comput. Vis. Pattern Recog. Workshop. pp. 129–136 (2006)
8. Chekmenev, S.Y., Farag, A.A., Essock, E.A.: Thermal imaging of the superficial temporal artery: An arterial pulse recovery model. In: Proc. IEEE Conf. Comput. Vis. Pattern Recog. pp. 1–6 (2007)
9. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proc. IEEE Conf. Comput. Vis. Pattern Recog. vol. 1, pp. 886–893 (2005)
10. Dowdall, J., Pavlidis, I., Tsiamyrtzis, P.: Coalitional tracking. *Comput. Vis. Imag. Under.* **106**(2-3), 205–219 (May-Jun 2007)
11. Fei, J., Pavlidis, I.: Thermistor at a distance: Unobtrusive measurement of breathing. *IEEE Trans. Biomed. Eng.* **57**(4), 988–998 (Apr 2010)
12. Fei, J., Zhu, Z., Pavlidis, I.: Imaging breathing rate in the CO_2 absorption band. In: Proc. Annual Int. Conf. IEEE Engineering in Medicine and Biology Society. pp. 700–705 (2005)
13. FLIR ONE: <http://www.flir.com/flirone> (2018)
14. Fukunaga, K., Hostetler, L.: The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. Info. Theory* **21**(1), 32–40 (Jan 1975)
15. Gade, R., Moeslund, T.B.: Thermal cameras and applications: A survey. *Machine Vision and Applications* **25**(1), 245–262 (Jan 2014)

16. Garbey, M., Sun, N., Merla, A., Pavlidis, I.: Contact-free measurement of cardiac pulse based on the analysis of thermal imagery. *IEEE Trans. Biomed. Eng.* **54**(8), 1418–1426 (Aug 2007)
17. Kalal, Z., Mikolajczyk, K., Matas, J.: Face-TLD: Tracking-learning-detection applied to faces. In: *Proc. IEEE Int. Conf. Image Processing*. pp. 3789–3792 (2010)
18. King, D.E.: Dlib-ml: A machine learning toolkit. *J. Machine Learning Research* **10**, 1755–1758 (Jul 2009)
19. Lampo, T., Sierra, J., Chang, C.: Two algorithms for measuring human breathing rate automatically. In: *Proc. Int. Sym. Visual Computing*. pp. 686–697 (2009)
20. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: *Proc. Int. Joint Conf. Artificial Intell.* pp. 674–679 (1981)
21. Microsoft Kinect V1: <https://en.wikipedia.org/wiki/Kinect> (2018)
22. Murthy, J.N., van Jaarsveld, J., Fei, J., Pavlidis, I., Harrykissoon, R.I., Lucke, J.F., Faiz, S., Castriotta, R.J.: Thermal infrared imaging: A novel method to monitor airflow during polysomnography. *Sleep* **32**(11), 1521–1527 (2009)
23. Murthy, R., Pavlidis, I.: Noncontact measurement of breathing function. *IEEE Eng. Med. Biol. Mag.* **25**(3), 57–67 (May-Jun 2006)
24. Optris PI 450: <http://www.optris.com/thermal-imager-pi400> (2018)
25. Park, S.B., Noh, Y.S., Park, S.J., Yoon, H.R.: An improved algorithm for respiration signal extraction from electrocardiogram measured by conductive textile electrodes using instantaneous frequency estimation. *Med. Bio. Eng. Comput.* **46**(2), 147–158 (Feb 2008)
26. Seek CompactPRO: <http://www.thermal.com/products/compactpro> (2018)
27. Storck, K., Karlsson, M., Ask, P., Loyd, D.: Heat transfer evaluation of the nasal thermistor technique. *IEEE Trans. Biomed. Eng.* **43**(12), 1187–1191 (Dec 1996)
28. Sun, N., Garbey, M., Merla, A., Pavlidis, I.: Imaging the cardiovascular pulse. In: *Proc. IEEE Conf. Comput. Vis. Pattern Recog.* vol. 2, pp. 416–421 (2005)
29. Yang, M., Liu, Q., Turner, T., Wu, Y.: Vital sign estimation from passive thermal video. In: *Proc. IEEE Conf. Comput. Vis. Pattern Recog.* pp. 1–8 (2008)
30. Zhu, Z., Fei, J., Pavlidis, I.: Tracking human breath in infrared imaging. In: *Proc. IEEE Sym. Bioinformatics and Bioengineering*. pp. 227–231 (2005)