An Approach for Skeleton Fitting in Long-Wavelength Infrared Images
First Results for a Robust Head Localisation using Probability Masks

Julia Richter, Christian Wiede and Gangolf Hirtz
Department of Electrical Engineering and Information Technology,
Technische Universität Chemnitz, Reichenhainer Str. 70, 09126 Chemnitz, Germany

Keywords: Human Skeleton Extraction, Thermal Imaging, Head Localisation.

Abstract: Human skeleton extraction has become a key instrument for motion analysis in the fields of surveillance, entertainment and medical diagnostics. While a vast amount of research has been carried out on skeleton extraction using RGB and depth images, far too little attention has been paid to extraction methods using long-wavelength infrared images. This paper provides an overview about existing approaches and explores their limitations. So far, extant studies have exploited thermal data only for silhouette generation as a preprocessing step. Moreover, they make strong assumptions, such as T-pose initialization. On this basis, we are developing an algorithm to fit the joints of a skeleton model into thermal images without such restrictions. We propose to find the head location as an initial step by using probability masks. These masks are designed to allow a robust head localisation in unrestricted settings. For the future algorithm design, we plan to localise the remaining skeleton joints by means of geometrical constraints. At this point, we will also consider sequences where persons wear thick clothes, which is aggravating the extraction procedure. This paper presents the current state of this project and outlines further approaches that have to be investigated to extract the complete skeleton.

1 INTRODUCTION

The observation of humans plays a pivotal role in surveillance, entertainment and medical diagnostics. A fundamental feature for such observations is the human skeleton, because it is a unique description of the human body. Therefore, it has been used in a variety of extant research, especially for human activity recognition (Yao et al., 2011), (Wang et al., 2015) and motion analysis (Huang et al., 2013), (Su et al., 2014), (Khan et al., 2014).

Humans are warm-blooded beings, which means they try to maintain a constant body temperature. Long-wavelength infrared (LWIR) thermography allows the measurement of the radiating energy released from the human body. This information can contribute to the localisation of human skeleton joints. In our work, we used a sensor with a spectral range of 7.5 to 13 micrometers. Compared to RGB and monochrome cameras, thermal cameras have the general advantage that they can sense infrared radiation emitted from humans. However, reflections, emissions due to sun illumination, non-human warm objects, environmental influences, individually varying differences, and non-homogeneously heat distributions over the body, e. g. caused by clothes, can cause problems.

To date, very few studies have investigated skeleton fitting in LWIR images. This paper explores the limitations of existing research and introduces a method that aims at fitting human skeletons in unconstrained 2-D LWIR images. This study makes a major contribution to research on skeleton fitting by demonstrating the exploitation of the measured surface temperatures. Furthermore, this is the first study in the field of skeleton fitting by means of LWIR images that undertakes a quantitative accuracy evaluation. Skeleton detection in LWIR images could contribute in a variety of applications that are especially related to medicine and sports diagnostics. As an example, Richter et al. used thermal images to measure the skin temperature of the biceps brachii during sports exercises (Richter et al., 2017). They mapped the skeleton provided by a Kinect sensor to a thermal image. In this way, they could automatically locate the muscle, which was a manual or semi-manual procedure in previous work, e. g. (Formenti et al., 2013), (Neves et al., 2014), (Bartuzi et al., 2012). To avoid the sensor calibration, which was necessary for the mapping, the skeleton could be extracted directly from the thermal

Richter, J., Wiede, C. and Hirtz, G.
An Approach for Skeleton Fitting in Long-Wavelength Infrared Images - First Results for a Robust Head Localisation using Probability Masks.
DOI: 10.5220/0006645103730379
In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018) - Volume 5: VISAPP*, pages 373-379
ISBN: 978-989-758-290-5

image. Medical applications that could profit from skeleton extraction are the automatic detection of inflammations and cancer, for instance.

While the first section of this paper already gave an insight into advantages, disadvantages and several applications of LWIR thermography, the second section examines extant work already carried out on skeleton extraction and formulates the research gap. Thereupon, Section 3 is concerned with the proposed method and the current state of the algorithm. Since the work is still in progress, we present the detection of the head by means of probability masks. Thereupon, Section 4 introduces the evaluation methodology that we used to determine the performance of the proposed method. In Section 5, the results are presented and discussed while Section 6 closes the paper by contemplating concepts and challenges concerning further development steps for a final skeleton extraction algorithm.

2 RELATED WORK

Our literature review focuses on research that is concerned with general aspects and applications of human body segmentation in thermal images. Moreover, we devote further attention to skeleton fitting in RGB images, because the existing concepts could be applicable to thermal imaging as well.

Much of the research that was carried out on detecting persons in thermal images utilised a foreground-background segmentation in the first step. Han et al. applied a Gaussian model to detect persons in thermal images (Han and Bhanu, 2005). The thereby obtained silhouettes were used to calculate gait energy images in order to recognise repetitive activities, i. e. walking patterns. Davis and Sharma used a background subtraction as well (Davis and Sharma, 2004). At this point, they especially focused on classic problems involved with thermal imagery, e. g. halo effects that cause inadequate results for commonly used statistical background subtraction techniques. In another work, which aimed at classifying sport types performed in a gym, Gade et al. employed an automatic threshold model to segment persons. These persons were then represented by bounding boxes (Gade and Moeslund, 2013). The bottom centres of these boxes were converted to world coordinates by means of a homography. The evaluation of resulting occupancy patterns in the top view allowed the sport type determination.

The previously mentioned approaches concentrated on processing several persons' silhouettes, i. e. the whole body, rather than on segmenting and analysing specific body parts. In contrast to that, numerous studies seeked to estimate the location of explicit body parts in thermal data. While some of them were concerned with the detection of only one part, such as the face (Wong et al., 2012), (Buddharaju et al., 2006), (Buddharaju et al., 2007), (Yu et al., 2010), a variety of studies investigated the segmentation of the whole body into refined parts. A rather coarse segmentation was introduced by Pham et al. (Pham et al., 2007). They introduced a 2-D human shape model consisting of an ellipse (head) and two rectangles (torso, legs) to detect if persons are lying down in a crowded area. After a background subtraction, they generated head hypotheses by using an elliptical template and by assuming that the highest gradients often occur around exposed body parts, especially in the case of faces. In addition to this, their algorithm detects the head-shoulder part with a cascade of several cascade classifiers employing histograms of oriented gradients. If the person was determined to be standing, the algorithm segmented the silhouette into the remaining parts. Hereby, the segmentation process was formulated as a maximum posteriori problem. Bhanu et al. fit a 3-D kinematic model with twelve parts to a 2-D silhouette that was calculated using a simple difference between an image and a background image (Bhanu and Han, 2002). For this fitting process, they presumed that the observed person is walking and viewed from the side. Their algorithm requires that the hand that is not facing the camera is periodically occluded in the video. A threshold was used to segment the face and the hands. They projected the 3-D model to the 2-D thermal image by means of camera parameters. Subsequently, they obtained the optimally matching 3-D model by performing a least square fit that minimises the difference between projected model and the 2-D silhouette.

Only few approaches, however, can be found so far for skeleton joint fitting in almost unconstrained settings. The approach that is the most similar with regard to our aim is the work of Iwasawa et al. (Iwasawa et al., 1997). They estimated ten joints in a sequence of thermal images. In a first step, they calculated the silhouette by applying a threshold on the image. Thereupon a distance transformed image was determined from the silhouette. This was followed by the determination of the center of gravity and of the upper body orientation. Subsequently, significant points were detected by a heuristic contour analysis of the silhouette to find the head top, the hand and foot tips as well as elbow and knee joints. In their approach, a T-pose initialization is required. Moreover, the thermal image is used for silhouette generation only. The possibility to exploit thermal informa-

tion for joint localisation was not taken advantage of. Since a large number of studies investigating skeleton fitting in RGB images is based on silhouettes as well, we briefly present common approaches at this point. Similar to (Iwasawa et al., 1997), Vignola et al. fitted a skeleton to silhouettes using distance transform (Vignola et al., 2003). For evaluation, they calculated the averaged joint-wise error to ground truth 2-D joint coordinates for scenarios with a different level of difficulty. Da et al. fitted ellipses to the upper body silhouette using contour curves (Da Xu and Kemp, 2009). Ding et al. calculated skeleton joints from critical points, which were determined by calculating the gradient of the distance transformed image (Ding et al., 2010). Recently, a learning-based approach using convolutional neural networks was presented by Wei et al. (Wei et al., 2016).

The studies presented thus far mainly used thermal images for segmenting persons from the background without performing a body part segmentation afterwards (Han and Bhanu, 2005), (Davis and Sharma, 2004), (Gade and Moeslund, 2013). Those studies that subsequently performed a body part segmentation either

- used a coarse model (Pham et al., 2007) and do not generate a refined model with skeleton joints,
- simplify the fitting process by making strong assumptions, such as T-pose initialisation, thin clothes and limited occlusions or
- are restricted to the processing of the obtained silhouette without further employing thermal information (Iwasawa et al., 1997).

Except the studies that are connected to face detection, the studies that come closest to our approach have not further exploited the available thermal information. Moreover, the only study that generates a refined model with specific joints (Iwasawa et al., 1997) exploits the available thermal data only for foreground segmentation. Furthermore, they do not present a quantitative accuracy evaluation for all the localized joints.

For these reasons, our work that is still in progress aims at fitting a refined skeleton model with 15 joints to thermal image sequences without assuming a constrained setting. These 15 joints are illustrated in Figure 1. We thereby include thermal data in the fitting process in order to take advantage of the thermal information. In this paper, we present the localisation of the head, i. e. j_5, by using probability masks.

3 METHOD

Although one might expect that the head can be easily localised in a thermal image because of the comparatively high temperature and the typical location, there are a variety of challenging scenarios, which we would like to address. These scenarios are:

- The person is bent forwards, so that only a part of the face is visible and the head is surrounded by other warm regions.
- The person is viewed from behind, so that no face is visible.
- The person is viewed from the side, so that only a part of the face is visible.
- The person raises the arms, so that they merge their area with the face.

In this paper, we assume that only one person is present in a recorded image sequence.

The algorithm for head localisation is presented in the following sections. Figure 2 illustrates the overview of the algorithm.

3.1 Foreground Segmentation

We assume that the person is the warmest object in the image. As a first step, temperatures lower than 60 % of the highest temperature were defined as background. After a linear auto contrast adjustment, the person's silhouette was extracted by using Otsu's method (Otsu, 1979). For scenarios with thick clothes, Otsu's method was unsuitable to obtain a closed silhouette. Here, a segmentation algorithm that is based on SLIC superpixels (Achanta et al., 2012) and DBSCAN clustering (Ester et al., 1996) would be more appropriate.

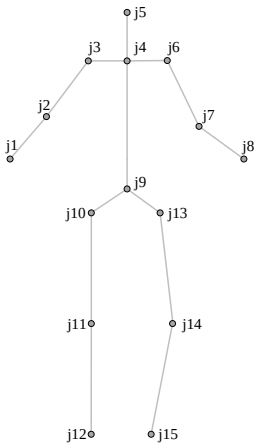

Figure 1: Skeleton model with fifteen joints.

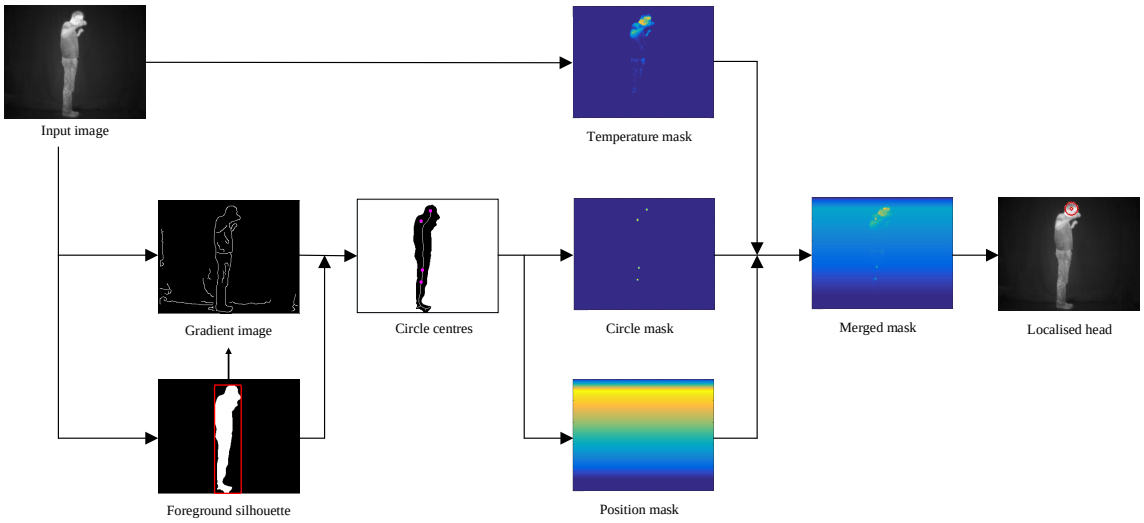

Figure 2: Overview of the proposed head localisation algorithm.

3.2 Mask Generation

Based on the input image and the segmented foreground, probability masks were calculated that represent the probability of each pixel to be the head centre position. In the following, the calculation of these masks and the final decision about the head location are described.

3.2.1 Circle Mask

To obtain possible head locations, circles of an appropriate head size were detected on the obtained foreground image and on its gradient image by means of the Canny Edge detector and Hough Transformation. Only circles with centres on the foreground were selected for the further processing. All mask pixels that exceed a fixed distance from these circle centres, which was set to three pixels, were set to zero. We choose a very small distance, because the circle centres are a very accurate indication for the head position. The remaining pixel values were then weighted according to their proximity to the closest centre. In this way, pixels on a centre position obtained the probability of one. Towards the edge of the circles, the probability linearly decreases towards zero. In the current work, we use a fixed head size, which should be adapted according to the size of the foreground blob in future.

3.2.2 Temperature Mask

The head is assumed to have a relative high temperature. Therefore, the temperature values in the infrared image can be used as a probability measure for possible head locations and to filter out improbable circle centres from the previous step. To obtain the temperature mask, all temperature values lower than 80 % of the maximum temperature were set to zero. Subsequently, a quadratic function was applied to map the remaining values to a probability between zero and one.

3.2.3 Position Mask

We introduced a further mask that is influenced by the vertical position of the circle with the topmost position in the image, because we assume the topmost circle more likely to be the head. The probability of each image row linearly rises from zero to one while starting from the first image row until the row of the topmost circle. The probabilities of the following rows are linearly decreasing again towards the last row.

3.2.4 Head Localisation on Merged Mask

In a final step, the masks were merged by an element-wise addition to obtain an overall probability measure for each pixel. The preliminary head position h_{pre} was denoted as the pixel with the highest probability. Since this position differed from the actual head in several cases, we rather used the circle centre of the closest circle h_{cir} if the distance between h_{pre} and h_{cir} exceeded the radius of this circle.

3.2.5 Tracking

The obtained head position was tracked by using a Kalman filter. The tracked head position will be the input for the calculation of further joints.

4 EVALUATION METHODOLOGY

The following sections present the data and parameters that were used to evaluate the head localisation.

4.1 Sensor

The employed sensor is a thermal camera of the type FLIR A35sc. This camera measures long-wavelength infrared emission with a spectral range of 7.5 to 13 μm. It provides a thermal sensitivity of 50 mK at 30 $^{\circ}C$ and has a spatial resolution of 320×256 pixel. The measured temperature is encoded with 14 bit.

4.2 Ground Truth Data Acquisition

In order to acquire ground truth data, we recorded sequences with a frame rate of 30 frames per second and manually labelled the joint positions in the thermal images. The ground truth position of the head was estimated to be the centre of the head. Overall, we recorded nine probands of different sex, body size and proportions and different hair styles ranging from bald head to long hair. The persons were standing approximately three meter away from the camera. The scenarios cover simple scenarios in T-poses and star-like poses as well as complex scenarios as described at the beginning of Section 3. Overall, 332 sample images were used for this evaluation.

4.3 Evaluation Parameters

Vignola et al. (Vignola et al., 2003) calculated the mean Euclidean distance between the labelled and the determined joint positions and the standard deviation to evaluate the accuracy of their algorithm. In this way, information about the error direction is lost, however. Therefore, in our work, the signed mean errors $\overline{e}_x$ and $\overline{e}_y$ as well as the standard deviations σ_x and σ_y with respect to the x and y coordinates were calculated for every joint in terms of pixels according to Equations 1 to 4. N denotes the number of tested images, which was 332 in our experiments, n is the image index, $(x_{o,n}, y_{o,n})$ corresponds to the output coordinate and $(x_{t,n}, y_{t,n})$ to the ground truth coordinate of one joint. In this paper, we apply these equations for the head joint only.

$$\overline{e}_x = \frac{1}{N} \cdot \sum_{n=1}^{N} (x_{o,n} - x_{t,n}) \tag{1}$$

$$\overline{e}_y = \frac{1}{N} \cdot \sum_{n=1}^{N} (y_{o,n} - y_{t,n}) \tag{2}$$

$$\sigma_x = \sqrt{\frac{1}{N} \cdot \sum_{n=1}^{N} \left((x_{o,n} - x_{t,n}) - \overline{e}_x\right)^2} \tag{3}$$

$$\sigma_y = \sqrt{\frac{1}{N} \cdot \sum_{n=1}^{N} \left((y_{o,n} - y_{t,n}) - \overline{e}_y\right)^2} \tag{4}$$

5 EXPERIMENTAL RESULTS

To illustrate the performance of the head localisation, we run the algorithm on example scenarios that we defined as challenging at the beginning of Section 3. A selection of the results for these scenarios is shown in Figure 3.

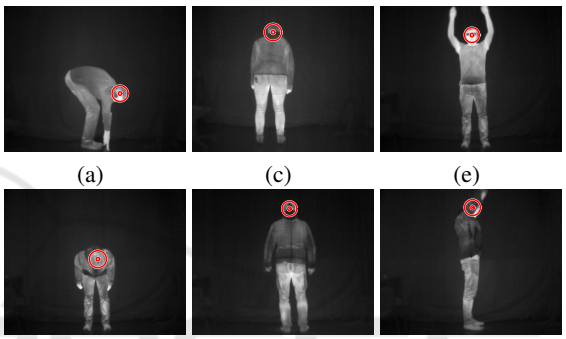

Figure 3: Example scenarios with persons viewed from the side (a), (f), bent forwards (a), (b), viewed from behind (c), (d) and with raised arms (e), (f).

The following table lists the signed mean error in row direction $\overline{e}_y$ and in column direction $\overline{e}_x$ as well as the standard deviation σ_y and σ_x in both directions.

Table 1: Signed mean errors in x and y direction and the corresponding standard deviation for the head joint. All numbers in pixels.

Head			
$\overline{e}_x$	$\overline{e}_y$	σ_x	σ_y
-0.57	-1.57	3.23	5.01

To visualise these results, both the mean error vector $\overline{e} = (\overline{e}_x, \overline{e}_y)$ and the standard deviations σ_x and σ_y were plotted with respect to the labelled head joint position as can be seen in Figure 4. For this visualisation, the labelled head position of one of the probands standing in T-pose was selected.

The results show small values for $\overline{e}_x$, $\overline{e}_y$ and the standard deviations σ_x and σ_y. That means that the determined head position does not deviate much from the labelled head position. This demonstrates the robustness of the head localisation even for challenging scenarios. Nevertheless, there were certain scenarios

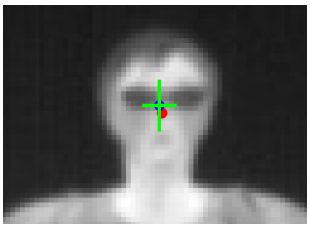

Figure 4: Visualised results. The red dot indicates the labelled joint position, the blue point the mean error with respect to the labelled joint and the green lines the standard deviations.

that still should be further considered: During our experiments, we noticed that shirts with a round neck can lead to the detection of additional circles, which are lower than the actual head circle. Especially in cases of uncovered necks, the higher temperature of the neck leads to the selection of the lower circle. A refinement of the head position by finding head signatures that include the typical curvature around the head could be a solution. Further incorrect detections occurred when the person was viewed from behind and the back showed a similar temperature than the head itself. In such cases, detected circles on such regions resulted in a higher probability than the detected circle on the head.

6 FUTURE WORK

In future work we aim at localising the remaining joints of the presented skeleton model on the basis of the head joint position and geometrical constraints. At this point, the centre of gravity (Iwasawa et al., 1997) and the distribution of end points and branches of the skeleton that was extracted by using bending potential ratio (Shen et al., 2011) can be relevant geometric clues to limit the search regions. Moreover, we plan to extract meta information about a person, such as the orientation with respect to the camera and whether the person is bent, standing straight or sitting, for example. Based on this meta information, different algorithms with different skeleton configurations can be used. Furthermore, the use of SLIC superpixels and DBSCAN clustering has to be investigated with regard to foreground extraction in cases of thick clothes, which reduce the emitted infrared radiation. In addition to this, the algorithm should be finally adapted to detect more than one person in the image. Further research will also involve varying distances to the camera, the occurrence of occlusions, reflections and other warm objects.

In conclusion, we would like to stress that skeleton extraction in LWIR images will contribute in a variety of applications. We intend, for instance, to measure temperature changes of a selection of body parts by means of the located skeleton joints. In this way, the course of joint inflammations after an injury could be diagnosed and the therapy could be adapted accordingly. Besides this, further application fields, such as security and surveillance, could benefit from such kind of automatised temperature measurements.

ACKNOWLEDGEMENTS

This project is funded by the European Social Fund (ESF). Moreover, we would like to thank all the persons who participated during the recordings.

REFERENCES

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Susstrunk, S. (2012). SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *PAMI. Vol 34 No 116*, pages 2274–2281.

Bartuzi, P., Roman-Liu, D., and Wiśniewski, T. (2012). The influence of fatigue on muscle temperature. *International Journal of Occupational Safety and Ergonomics*, 18(2):233–243.

Bhanu, B. and Han, J. (2002). Kinematic-based human motion analysis in infrared sequences. In *Applications of Computer Vision, 2002.(WACV 2002). Proceedings. Sixth IEEE Workshop on*, pages 208–212. IEEE.

Buddharaju, P., Pavlidis, I., and Tsiamyrtzis, P. (2006). Pose-invariant physiological face recognition in the thermal infrared spectrum. In *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, pages 53–53. IEEE.

Buddharaju, P., Pavlidis, I. T., Tsiamyrtzis, P., and Bazakos, M. (2007). Physiology-based face recognition in the thermal infrared spectrum. *IEEE transactions on pattern analysis and machine intelligence*, 29(4):613–626.

Da Xu, R. Y. and Kemp, M. (2009). Multiple curvature based approach to human upper body parts detection with connected ellipse model fine-tuning. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 2577–2580. IEEE.

Davis, J. W. and Sharma, V. (2004). Robust detection of people in thermal imagery. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, pages 713–716. IEEE.

Ding, J., Wang, Y., and Yu, L. (2010). Extraction of human body skeleton based on silhouette images. In *Education Technology and Computer Science (ETCS), 2010 Second International Workshop on*, volume 1, pages 71–74. IEEE.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in

large spatial databases with noise. *Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pages 226–231.

Formenti, D., Ludwig, N., Gargano, M., Gondola, M., Dellerma, N., Caumo, A., and Alberti, G. (2013). Thermal imaging of exercise-associated skin temperature changes in trained and untrained female subjects. *Annals of biomedical engineering*, 41(4):863–871.

Gade, R. and Moeslund, T. (2013). Sports type classification using signature heatmaps. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 999–1004.

Han, J. and Bhanu, B. (2005). Human activity recognition in thermal infrared imagery. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops*, pages 17–17. IEEE.

Huang, T.-C., Cheng, Y.-C., and Chiang, C.-C. (2013). Automatic Dancing Assessment Using Kinect. In *Advances in Intelligent Systems and Applications-Volume 2*, pages 511–520. Springer.

Iwasawa, S., Ebihara, K., Ohya, J., and Morishima, S. (1997). Real-time estimation of human body posture from monocular thermal images. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 15–20. IEEE.

Khan, N. M., Lin, S., Guan, L., and Guo, B. (2014). A visual evaluation framework for in-home physical rehabilitation. In *Multimedia (ISM), 2014 IEEE International Symposium on Multimedia*, pages 237–240. IEEE.

Neves, E. B., Vilaça-Alves, J., Krueger, E., and Reis, V. M. (2014). Changes in skin temperature during muscular work: a pilot study. *Pan American Journal of Medical Thermology*, 1(1):11–15.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66.

Pham, Q.-C., Gond, L., Begard, J., Allezard, N., and Sayd, P. (2007). Real-time posture analysis in a crowd using thermal imaging. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE.

Richter, J., Wiede, C., Kaden, S., Weigert, M., and Hirtz, G. (2017). Skin Temperature Measurement based on Human Skeleton Extraction and Infra-red Thermography - An Application of Sensor Fusion Methods in the Field of Physical Training. In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 6: VISAPP, (VISIGRAPP 2017)*, pages 59–66.

Shen, W., Bai, X., Hu, R., Wang, H., and Latecki, L. J. (2011). Skeleton growing and pruning with bending potential ratio. *Pattern Recognition*, 44(2):196–209.

Su, C.-J., Chiang, C.-Y., and Huang, J.-Y. (2014). Kinect-enabled home-based rehabilitation system using Dynamic Time Warping and fuzzy logic. *Applied Soft Computing*, 22:652–666.

Vignola, J., Lalonde, J.-F., and Bergevin, R. (2003). Progressive human skeleton fitting. In *Proceedings of the 16th Conference on Vision Interface*, pages 35–42.

Wang, Y., Sun, S., and Ding, X. (2015). A self-adaptive weighted affinity propagation clustering for key frames extraction on human action recognition. *Journal of Visual Communication and Image Representation*, 33:193–202.

Wei, S.-E., Ramakrishna, V., Kanade, T., and Sheikh, Y. (2016). Convolutional pose machines. *arXiv preprint arXiv:1602.00134*.

Wong, W. K., Hui, J. H., Desa, J. B. M., Ishak, N. I. N. B., Sulaiman, A. B., and Nor, Y. B. M. (2012). Face detection in thermal imaging using head curve geometry. In *Image and Signal Processing (CISP), 2012 5th International Congress on*, pages 881–884. IEEE.

Yao, A., Gall, J., Fanelli, G., and Van Gool, L. J. (2011). Does Human Action Recognition Benefit from Pose Estimation?. In *BMVC*, volume 3, page 6.

Yu, X., Chua, W. K., Dong, L., Hoe, K. E., and Li, L. (2010). Head pose estimation in thermal images for human and robot interaction. In *Industrial Mechatronics and Automation (ICIMA), 2010 2nd International Conference on*, volume 2, pages 698–701. IEEE.