Face Detection in Thermal Infrared Images: A Comparison of Algorithm- and Machine-Learning-Based Approaches

Marcin Kopaczka[1], Jan Nestler[1], and Dorit Merhof[1]

Institute of Imaging and Computer Vision, RWTH Aachen University, Germany
`marcin.kopaczka@lfb.rwth-achen.de`

Abstract. In recent years, thermal infrared imaging has gained an increasing attention in person monitoring tasks due to its numerous advantages such as illumination invariance and its ability to monitor vital parameters directly. Many of these applications require facial region monitoring. In this context, several methods for face detection in thermal infrared images have been developed. Nearly all of the approaches introduced in this context make use of specific properties of facial images in the thermal infrared domain, such as local temperature maxima in the eye area or the fact that human bodies usually have a higher temperature radiation than the backgrounds used. On the other side, a number of well-performing methods for face detection in the visual spectrum has been introduced in recent years. These approaches use state-of-the-art algorithms from machine learning and feature extraction to detect faces in photographs and videos. So far, only one of these algorithms has been successfully applied to thermal infrared images. In our work, we therefore analyze how a larger number of these these algorithms can be adapted to thermal infrared images and show that a wide number of recently introduced algorithms for face detection in the visual spectrum can be trained to work in the thermal spectrum when an appropriate training database is available. Our evaluation shows that these machine-learning based approaches outperform thermal-specific solutions in terms of detection accuracy and false positive rate. In conclusion, we can show that well-performing methods introduced for face detection in the visual spectrum can also be used for face detection in thermal infrared images, making dedicated thermal-specific solutions unnecessary.

Keywords: thermal infrared, face detection

1 Introduction

Thermal imaging, also referred to as long-wave infrared (LWIR) imaging uses the thermal subband of the electromagnetic spectrum (7 to 14 m) to acquire illumination invariant images, allowing LWIR imaging sensors to operate even in absolute darkness. This sensor property makes LWIR imaging especially interesting for applications with high variance in lighting conditions. Currently,

two main focus areas of thermal imaging resarch are illumination-invariant face recognition for survelliance purposes [3][14][2] and the extraction of vital parameters such as respiratory and heart rate (RR/HR) from facial images [10]. Furthermore, emotion detection [6][23] and affective state analysis such as stress detection and polygraphy using thermal infrared images [21][19] are emerging research areas. All of these applications require analyzing facial regions of interest (ROIs) which need to be defined first. In most of the published cases, ROIs are defined manually due to the lack of reliable face and facial landmark detection algorithms for thermal infrared images. In the other cases, ROI definition is performed automatically or semi-automatically, however always under the constraint that certain prerequisites such as a full frontal view of the face or having the face restricted to a certain image area are given. Only a small number of publications [22][18] use a state-of-the-art face detection algorithm, in all reviewed literature the detection method was the Haar-based Viola-Jones face detection algorithm [25]. To the best of our knowledge there has been no application of more recent face detection methods to thermal infrared images. Furthermore, there is a lack of literature comparing face detection methods developed especially for the thermal infrared spectrum with well-established approaches that have been developed for the visual spectrum. We believe that the reason for lack of relevant publications is the fact that all current methods are based on automated classification using machine learning. Regardless of the actual classifier and feature descriptor used, image classification based on machine learning requires extensive amounts of manually annotated training data which is not available for the thermal infrared domain. Therefore, no quantitative comparison of domain-specific versus machine learning-based face detection algorithms has been performed so far. Our work aims at addressing this issue by focusing on several aspects: First, we introduce an extensively annotated thermal face database that can be used to train face detection algorithms that have been originally developed for the visual spectrum and have not been applied to infrared images due to the lack of sufficient training data. We use the database to train a selection of face detection algorithms that have been proven to perform well on images in the visible spectrum. Finally, we thoroughly evaluate the performance of these algorithms and compare it to the results of reference algorithms developed especially for face detection in thermal infrared images.

The structure of the paper is the following: Our database is described in section 2.1, followed by an overview of the algorithms that we trained with it in section 2.2. A description of the performed experiments and their results is given in section 2.3. The paper is concluded by a result discussion (section 4) and a conclusion (section 5).

2 Materials and Methods

In this section we first describe the image database created to train and evaluate face detection algorithms. Subsequently, all evaluated face detectors are briefly introduced.

2.1 Thermal Face Database

Since most algorithms for facial landmark detection require a ground truth database for training, the first required step is the creation of a database with face and non-face images in the thermal infrared domain. Many current face detection and tracking algorithms developed for the visual spectrum are trained using large sets of in-the-wild images such as the LFPW [1] or HELEN [15] databases which are acquired under unconstrained settings with a wide variety of head poses, lighting conditions and facial expressions. Up to now there existed no corresponding database meeting these requirements in the LWIR domain. Available thermal infrared datasets offer images with limited temperature contrast and spatial resolution and only a small set of pre-defined facial expressions and head poses. To overcome this shortcoming and allow to test the applicability of established algorithms in the thermal infrared, we have created a database of thermal videos of currently 82 persons in which each participant performed both arbitrary and protocol-specified head motions and facial expressions. All our images were acquired with an InfraTec 820 HD microbolometer array camera at its native 1024x768 pixels with a relative thermal resolution of 0.03K. The videos were taken from a distance of 0.8 m to 1.0 m and all participants were placed in front of a thermally homogeneous backdrop. From the video sequences, 2935 single frames that cover a wide and realistic range of head positions and facial expressions have been selected. A set of 68 facial landmark points has been manually marked in each image (Fig. 1). The points selected for landmarking are consistent with the landmarks used in a number of current face databases in the visual spectrum including LFPW and HELEN, allowing an efficient training of algorithms developed for the visible spectrum on thermal infrared data. Our fully annotated face database surpasses available thermal infrared face databases such as [26] or the databases listed in [11] in terms of spatial and thermal resolution as well as variety of head poses and number of annotated landmark points. The database will be made available to the public in the near future.
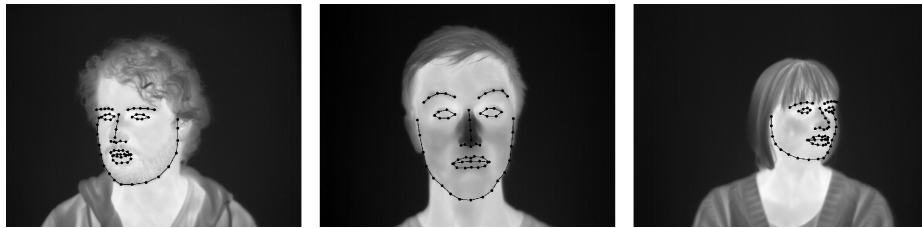

Fig. 1: A set of sample images from the thermal face database. The manually created landmark annotations are shown as markers.

2.2   Face Detection Algorithms

In this work, we compare five machine-learning based algorithms that have already been successfully applied to face detection in the visible spectrum to two approaches that have been developed especially for thermal infrared images. It should be noted that only one of the five classifiers (namely the Viola-Jones-Method) has been used in research literature for thermal infrared face detection. To the best of our knowledge, this work is the first contribution to evaluate the performance of the remaining four algorithms in LWIR. The algorithms transferred from the visible spectrum are:

- The Haar cascade classifier **(VJ)** as presented by Viola and Jones [25]. This work was the first robust face detection algorithm capable of delivering real-time performance even on handheld machines with limited computing power such as consumer-grade digital photo cameras. VJ detects faces by systematically analyzing subimages extracted using a multiscale sliding window and applying differently sized Haar feature detectors in a cascading manner. In this approach, the subimage is matched with the most relevant Haar features first. If the feature response is positive then further analysis using more refined features is performed by following the cascade. Should the feature response be negative at any point, then the subimage is rejected as containing no face. The algorithm returns a face detection only if the subimage passes all steps of the cascade. The detector itself is trained by providing it with a set of face and non-face images to which a variety of Haar descriptors is applied. The algorithm learns the most significant features that allow differentiating between face and non-face images and arranges the best performing feature combinations in a fast decision tree using a boosting algorithm such as AdaBoost or GentleBoost. Using boosting for feature selection and efficient coding techniques for feature computation the algorithm is capable of performing face detection in photographs and videos in real-time. As stated above, this is the only algorithm that has already been applied successfully to thermal infrared images by different authors such as [24] and more currently [4]. The implementation used in our work is based on the OpenCV [12] variant of the cascade classifier, a method that improves the original work by Viola and Jones by combining the Haar descriptors in the cascade into groups instead of applying them individually at each step.
- The Haar cascade classifier **(VJ-LBP)** with local binary patterns (LBPs) as feature descriptor as presented by Liao et al. in [16]. This approach is similar to the first method, however in VJ-LBP local binary patterns are used instead of Haar features. Since they only require integral computations, LBPs can be computed more efficient than Haar features. Furthermore, using cascade classifiers with LBPs instead of Haar features has been reported to result in improved detection rates. As with the above method, we used the LBP cascade classifier available in OpenCV.
- Histograms of Oriented Gradients **(HOG)**, an image descriptor and object detection algorithm introduced by Dalal & Triggs [7] and currently being one of the most widely used feature descriptors. HOG features are computed by

Fig. 2: Possible Haar features for face detection

analyzing image gradients and grouping them into local histograms. To use HOG for face detection, the feature vector is computed for a training set of face and non-face images and the results are used to train a classifier - usually a support vector machine (SVM) - that learns how to distinguish the HOG feature representation of a face from background features. In our work, we used the implementation available in the dlib library [13] to train and test a HOG-based face detector.

– The Deformable Parts Model **(DPM)**, introduced by Felzenszwalb in [8]. This approach uses optimized HOG features for object description, however for classification and detection the method assumes that the sought objects are composed of different components (in our case face parts such as mouth or eyes) and that these components may be arranged differently in different images. Instead of assuming a fixed spatial configuration as in regular HOG, the method learns the feature representations of different object parts as well as their spatial distribution. The resulting detector can adapt to different spatial configurations induced by changes in head pose or camera position.

– Pixel Intensity Comparisons Organized in Decision Trees **(PICO)**, presented by Markus et al. in [17]. This method uses binary pixel intensity comparisons for object detection. Similar to VJ, the algorithm performs face detection by scanning sliding subimages and applying a decision tree boosted by Gentle-Boost. Each decision is defined by direct comparison of the intensity values of two pixels in the current subimage. Similar to Haar feature selection in VJ, pixel locations chosen for comparison are optimized by applying the best discriminating combinations first. This method drastically increases the tree's detection speed since feature computation can be omitted.

We compare these approaches to two algorithms developed primarily for the detection of faces in thermal infrared images. The two algorithms were chosen as they both represent common approaches presented for thermal infrared face detection. Generally, many algorithms designed for thermal face detection apply thresholding for pre-segmentation of the image or feature localization. The aim of the thresholding is either a foreground-background-segmentation under the assumption that the person is the dominant heat source in the image, or a threshold operation with a threshold close to the image's maximum temperature to locate the hottest spots in the image. The idea behind hottest spot localization

is the fact that in many cases the inner corners of the eyes are the hottest regions of the face and therefore the eyes can be easily located with thresholding. Either one or both of the approaches are the basic idea behind the two chosen methods as well as behind current algorithms such as [5] [20] or [27]. The methods that we selected for implementation were:

– Eye Corner Detection **(ED)**, presented by Friedrich and Yeshurun in [9]. In this work, the initial step is also a silhouette extraction followed by a facial feature detection. The original paper states that temperature-based feature detection may be performed on the areas around the eyes, with the eyes themselves being the warmest and the eyebrows being the coldest regions of the face. The authors of [9] have tested different approaches and have identified the eyebrows to be the most reliably detectable face part on their dataset. We performed the same tests on our database and found out that in our case and similar to the results of [20], the inner corners of the eyes can be located more reliably than eyebrows since there tend to be several cold regions in the face. Therefore we implemented the method as described in the paper, however with the difference that our approach used a feature better suited to our dataset (Fig. 3).
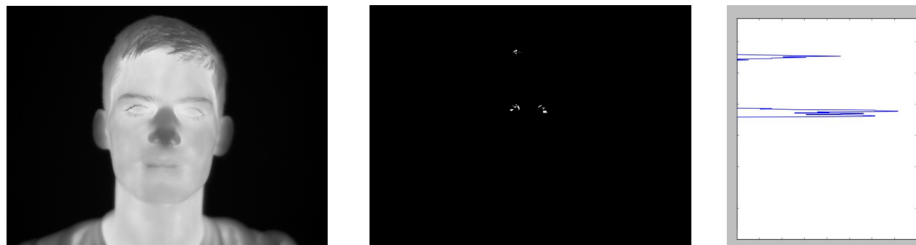

Fig. 3: The Eye Corner Detection algorithm. Left: Original input image. Center: image after thresholding. Right: Sum of all pixels per row of the thresholded image. The vertical coordinate of the eye centers is the row with the biggest sum.

– Projection Profile Analysis **(PPA)**, an algorithm described by Reese et al. in [22]. This method performs a silhouette extraction by foreground-background thresholding and computes projection histograms of the resulting binary image (Fig. 4) on the horizontal and vertical image axes. Subsequently, the face is localized by defining extreme points in the profile curves extracted from the projection histograms. These points are then localized by analyzing the extracted profile curves and their 1st and 2nd order derivatives. The approach assumes that the person is the dominant object in the image and that the image displays the whole head and the upper part of the torso. These prerequisites are given in our database.
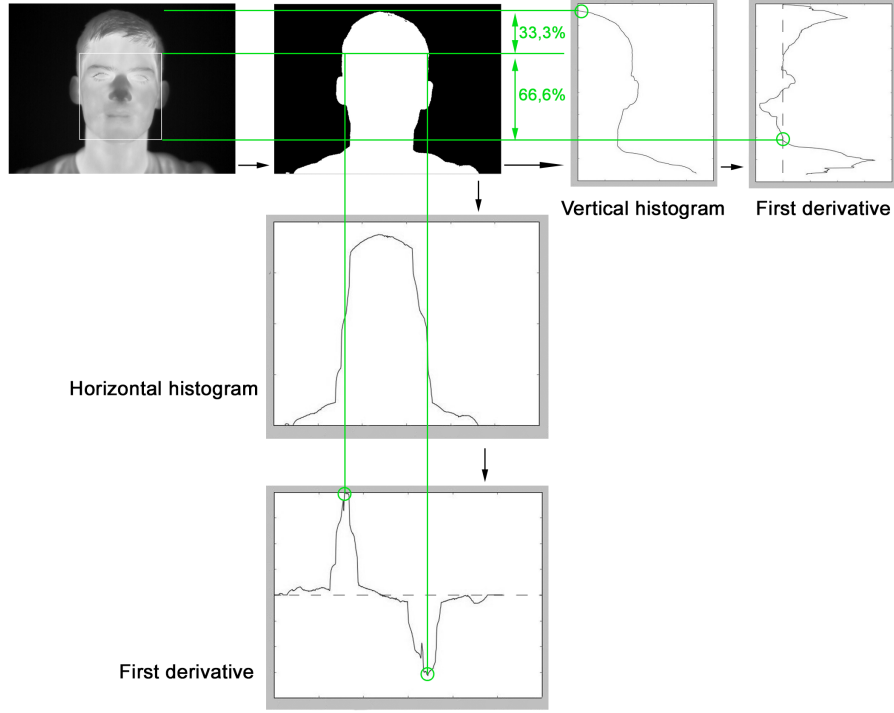
Fig. 4: The Projection Profile Analysis (PPA) approach: The algorithm performs a foreground-background segmentation and subsequently defines the face's bounding box by finding extrema in the derivatives of the profile curves. Note that for comparison with the ground truth we crop away the top 1/3 of the result to match the manual annotations better.

2.3 Evaluation Methodology

Depending on the algorithm, the results returned by the different methods are not directly comparable. The five machine learning-based methods all return a face bounding box that can be compared with the manually defined ground truth bounding box in the test images. However, the results computed by the two thermal infrared face detection methods need to be processed for a quantitative evaluation as they do not return a bounding box directly. For ED which returns eye corner coordinates, the returned coordinates x_1, y_1, x_2, y_2 and their distance $d = x_2 - x_1$ were used as base for bounding box computation. The final coordinates for the upper left and lower right corner of the bounding box are then defined as $x_{ul} = x_1 - 2d, y_{ul} = y_1 - d, x_{lr} = x_2 + 2d, y_{lr} = y_1 + 4d$. For PPA, the returned bounding box is refined by cropping off the top 1/3 of the algorithm result since the algorithm returns a bounding box for the whole head including face, forehead and hair.

3 Experiments and Results

For algorithm training and evaluation using cross-validation, the 2935 images of the database were split into 10 subsets by sequentially assigning all images of an individual to the subset with currently least images in it. This way, we could achieve approximately equally sized subsets in which no subject's images were split between different sets, thereby ensuring no overlap between training and test sets. We then performed algorithm training using leave-one-subset-out cross-validation, i.e. each image was used for testing once and for algorithm training in the other nine iterations.

The ground truth bounding box was defined by the bounding box of each image's landmark annotations, expanded by 5% in each direction (Fig. 5a). We have decided to enlarge the original bounding boxes since preliminary results have shown that the feature descriptors used by our face detection algorithms perform better on images that contain some background area around the face. The error metric used for evaluation was the Intersection over Union (IoU), defined for the overlapping area of ground truth G and detection bounding box D returned by the detector as the ratio of overlapping area and united area (see also Fig. 5a):

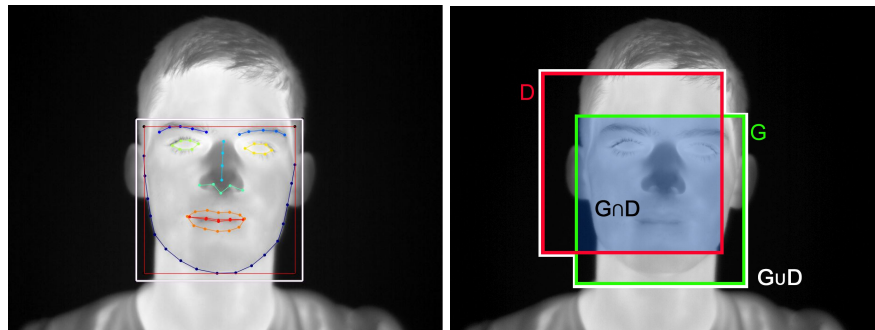$$\mathrm{IoU} = \frac{\mathrm{G} \cap \mathrm{D}}{\mathrm{G} \cup \mathrm{D}}.$$

Fig. 5: Left: ground truth definition by expanding the bounding box of the landmarks by 5%. Right: A visualization of the definitions of the intersection I = G ∩ D (blue) and the union U = G ∪ D (white) of two results for IoU computation.

Commonly, a successful detection is defined as a result with IoU > 0.5. The results for the different algorithms are shown in Fig. 6. It can be seen that all learning-based approaches outperform the algorithm-based methods by a considerable margin.
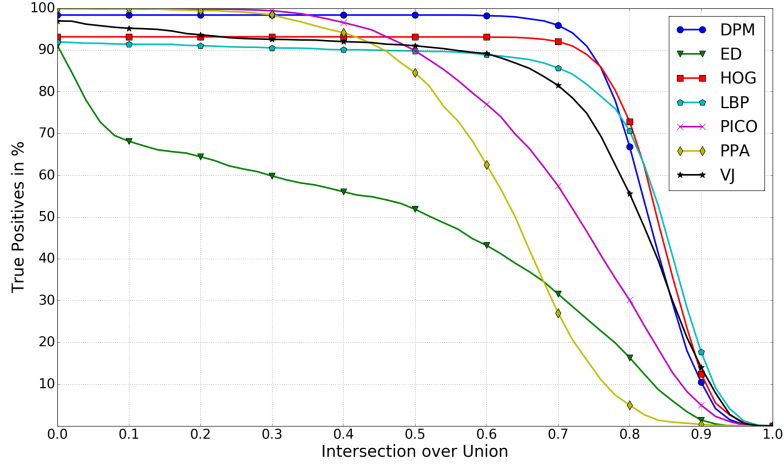
Fig. 6: Percentage of faces that are detected with a given IoU.

Table 1: Performance overview of the tested algorithms. Positive/negative rates, precision and recall are given for an IoU of 0.5. Detection and training times are given for an Intel i5-2500K CPU with 3.3 GHz.

	DPM	ED	HoG	LBP	PICO	PPA	VJ
Training Time (h)	19	-	0.5	50	2	-	73
Avg. Detection Time (s)	1,13	0,01	0,18	0,16	0,02	0,09	0,09
True Positive Rate	0,98	0,52	0,93	0,90	0,90	0,85	0,91
False Positive Rate	0,00	0,48	0,00	0,03	0,10	0,15	0,07
False Negative Rate	0,02	0,00	0,07	0,08	0,00	0,00	0,02
Precision	1,00	0,52	1,00	0,97	0,90	0,85	0,93
Recall	0,98	0,99	0,93	0,92	1,00	1,00	0,98

4 Discussion

Results show that classifier-based detection methods that are commonly used for face detection in the visual spectrum can be applied to thermal infrared images if trained with a well suited database. Each of the tested algorithms has shown better detection performance than the domain-specific solutions that were implemented and tested as well. At the same time, the performance indicators given in table 1 show different methods are more recommended than others depending on the given detection task. ED shows the weakest performance. An inspection of the values returned by the detector reveals that the algorithm is extremely sensitive to changes in pose and facial expression. A slight tilt of the head or opening the mouth usually resulted in failed detections. PPA shows a significantly better performance, however still behind the other algorithms.

VJ - the only classifier-based face detection algorithm that has been reportedly used for face detection in thermal infrared images before - as well as VJ-LBP and HOG show good similar detection rates. DPM shows the best detection and false positive rates of all algorithms at the cost of having the longest detection time of the tested methods, making it the algorithm of choice for all cases except those with very strict time constraints. In cases where detection time is crucial, PICO is the method of choice since it is as fast as the two domain-specific algorithms while yielding better results and remaining robust to pose changes.

5   Conclusion and Future Work

In this paper, we performed a performance comparison between face detection algorithms presented for the visual spectrum and applied to thermal infrared images on one side and approaches developed especially for thermal infrared images on the other. We have shown how machine-learning based face detectors can be trained to detect faces in thermal infrared images and that these algorithms outperform specialized approaches. We conclude that due to the better performance of trainable face detection algorithms there is generally no need to develop specialized face detection approaches for thermal infrared images. The only exception are highly controlled scenarios with high performance requirements in which the specialized approaches would performed better due to their lower computational cost.

Since we have shown that algorithms for face detection can be transferred from the visual to the thermal infrared spectrum with good results, we will analyze the applicability of other algorithms that have been developed for facial images in the visual spectrum to thermal images. Furthermore we plan to test the algorithms on images taken in less constrained environments to test the robustness of the different approaches in more realistic settings.

References

1. Belhumeur, P.N., Jacobs, D.W., Kriegman, D.J., Kumar, N.: Localizing parts of faces using a consensus of exemplars. In: IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). vol. 35, pp. 2930–2940 (December 2013)
2. Bhowmik, M.K., Saha, K., Majumder, S., Majumder, G., Saha, A., Sarma, A.N., Bhattacharjee, D., Basu, D.K., Nasipuri, M.: Thermal infrared face recognitiona biometric identification technique for robust security system. Reviews, refinements and new ideas in face recognition pp. 113–138 (2011)
3. Buddharaju, P., Pavlidis, I.T., Tsiamyrtzis, P., Bazakos, M.: Physiology-based face recognition in the thermal infrared spectrum. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(4), 613–626 (April 2007)
4. Chakraborty, M., Raman, S.K., Mukhopadhyay, S., Patsa, S., Anjum, N., Ray, J.G.: High precision automated face localization in thermal images: oral cancer dataset as test case. Proc. SPIE 10133, 1013326–1013326–7 (2017), http://dx.doi.org/10.1117/12.2254236

5. Chakrabortya, M., Ramanb, S., Mukhopadhyaya, S., Patsac, S., Anjumc, N., Rayc, J.: High precision automated face localization in thermal images: Oral cancer dataset as test case. In: SPIE Medical Imaging. pp. 1013326–1013326. International Society for Optics and Photonics (2017)

6. Cruz-Albarran, I.A., Benitez-Rangel, J.P., Osornio-Rios, R.A., Morales-Hernandez, L.A.: Human emotions detection based on a smart thermal system of thermographic images. Infrared Physics and Technology 81, 250–261 (2017)

7. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. vol. 1, pp. 886–893. IEEE (2005)

8. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. IEEE transactions on pattern analysis and machine intelligence 32(9), 1627–1645 (2010)

9. Friedrich, G., Yeshurun, Y.: Seeing people in the dark: face recognition in infrared images. In: International Workshop on Biologically Motivated Computer Vision. pp. 348–359. Springer (2002)

10. Gade, R., Moeslund, T.B.: Thermal cameras and applications: a survey. Machine vision and applications 25(1), 245–262 (2014)

11. Ghiass, R.S., Arandjelovi?, O., Bendada, A.H., Maldague, X.: Infrared face recognition: A comprehensive review of methodologies and databases. Pattern Recognition, vol. 47, no. 9 pp. 2807–2824 (2014)

12. Itseez: Open source computer vision library. https://github.com/itseez/opencv (2015)

13. King, D.E.: Dlib-ml: A machine learning toolkit. Journal of Machine Learning Research 10, 1755–1758 (2009)

14. Kong, S.G., Heo, J., Abidi, B.R., Paik, J., Abidi, M.A.: Recent advances in visual and infrared face recognitiona review. Computer Vision and Image Understanding 97(1), 103 – 135 (2005), http://www.sciencedirect.com/science/article/pii/S1077314204000451

15. Le, V., Brandt, J., Lin, Z., Bourdev, L., Huang, T.S.: Interactive facial feature localization. In: Computer Vision–ECCV 2012, pp. 679–692. Springer (2012)

16. Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S.: Learning multi-scale block local binary patterns for face recognition. Advances in biometrics pp. 828–837 (2007)

17. Markuš, N., Frljak, M., Pandžić, I.S., Ahlberg, J., Forchheimer, R.: Object detection with pixel intensity comparisons organized in decision trees. arXiv preprint arXiv:1305.4537 (2013)

18. Mostafa, E., Hammoud, R., Ali, A., Farag, A.: Face recognition in low resolution thermal images. Computer Vision and Image Understanding 117(12), 1689–1694 (2013)

19. Park, K.K., Suk, H.W., Hwang, H., Lee, J.H.: A functional analysis of deception detection of a mock crime using infrared thermal imaging and the concealed information test. Frontiers in Human Neuroscience 7 (2013)

20. Paul, M., Blanik, N., Blazek, V., Leonhardt, S.: An efficient method for facial component detection in thermal images. In: The International Conference on Quality Control by Artificial Vision 2015. pp. 95340P–95340P. International Society for Optics and Photonics (2015)

21. Pavlidis, I., Tsiamyrtzis, P., Shastri, D., Wesley, A., Zhou, Y., Lindner, P., Buddharaju, P., Joseph, R., Mandapati, A., Dunkin, B., et al.: Fast by nature-how stress patterns define human experience and performance in dexterous tasks. Scientific Reports 2 (2012)

22. Reese, K., Zheng, Y., Elmaghraby, A.: A comparison of face detection algorithms in visible and thermal spectrums. In: Intl Conf. on Advances in Computer Science and Application (2012)
23. Salazar-López, E., Domínguez, E., Ramos, V.J., de la Fuente, J., Meins, A., Iborra, O., Gálvez, G., Rodríguez-Artacho, M., Gómez-Milán, E.: The mental and subjective skin: Emotion, empathy, feelings and thermography. Consciousness and cognition 34, 149–162 (2015)
24. Sumriddetchkajorn, S., Somboonkaew, A.: Face detection in thermal imagery using an open source computer vision library. Proc. SPIE 7299, 729906–729906–6 (2009), http://dx.doi.org/10.1117/12.819996
25. Viola, P., Jones, M.J.: Robust real-time face detection. International journal of computer vision 57(2), 137–154 (2004)
26. Wang, S., Liu, Z., Lv, S., Lv, Y., Wu, G., Peng, P., Chen, F., Wang, X.: A natural visible and infrared facial expression database for expression recognition and emotion inference. Multimedia, IEEE Transactions on 12(7), 682–691 (Nov 2010)
27. Wong, W.K., Hui, J.H., Desa, J.B.M., Ishak, N.I.N.B., Sulaiman, A.B., Nor, Y.B.M.: Face detection in thermal imaging using head curve geometry. In: Image and Signal Processing (CISP), 2012 5th International Congress on. pp. 881–884. IEEE (2012)