Post-print of: M. Szankin, A. Kwasniewska and J. Ruminski, "Influence of Thermal Imagery Resolution on Accuracy of Deep Learning based Face Recognition," 2019 12th International Conference on Human System Interaction (HSI), Richmond, VA, USA, 2019, pp. 1-6, doi: 10.1109/HSI47298.2019.8942636.

Influence of Thermal Imagery Resolution on Accuracy of Deep Learning based Face Recognition

Maciej Szankin
Artificial Intelligence Products Group
Intel AI
San Diego, CA, USA
maciej.szankin@intel.com

Alicja Kwasniewska
Faculty of Electronics, Telecommunications and Informatics
Gdansk University of Technology
Gdansk, Poland
Intel AI
San Diego, CA, USA
alicja.kwasniewska@intel.com

Jacek Ruminski
Faculty of Electronics, Telecommunications and Informatics
Gdansk University of Technology
Gdansk, Poland
jacek.ruminski@pg.edu.pl

***Abstract*—Human-system interactions frequently require a retrieval of the key context information about the user and the environment. Image processing techniques have been widely applied in this area, providing details about recognized objects, people and actions. Considering remote diagnostics solutions, e.g. non-contact vital signs estimation and smart home monitoring systems that utilize person's identity, security is a very important factor. Thus, thermal imaging has become more and more popular, as it does not reveal features that are often used for person recognition, i.e. sharp edges, clear changes of pixel values between areas, etc. On the other hand, there are much more visible light data available for deep model training. Taking it into account, person recognition from thermography is much more challenging due to specific characteristics (blurring and smooth representation of features) and small amount of training data. Moreover, when low resolution data is used, features become even less visible, so this problem may become more difficult. This study focuses on verifying whether model trained to extract important facial embedding from RGB images can perform equally well if applied to thermal domain, without additional re-training. We also perform a set of experiments aim at evaluating the influence of resolution degradation by down-scaling images on the recognition accuracy. In addition, we present deep super-resolution (SR) model that by enhancing donw-scaled images can improve results for data acquired in scenarios that simulate real-life environment, i.e. mimicking facial expressions and performing head motions. Preliminary results proved that in such cases SR helps to increase accuracy by 6.5% for data 8 times smaller than original images. It has also been shown that it is possible to accurately recognize even 40 volunteers using only 4 images per person as a reference embedding. Thus, the initial profiles can be easily created in a real time, what is an additional advantage considering a solution setup in a new environment.**

***Index Terms*—face recognition, thermal imagery, deep neural networks, image enhancement**

I. Introduction

Human-system interactions frequently require a retrieval of the key context information about the user and the environment. Apart from information acquired using various sensors, the context can be also provided by applying computer vision algorithms, e.g. person, objects, or actions detection and recognition [1]–[5]. The most challenging problems in solutions that use the vision context are often associated with poor lighting conditions [6] and security concerns [7]. Huang and Bian [8] addressed the illumination variations by adopting Gamma correction, Difference of Gauss filtering (DoG) and contrast equalization. Different approach proposed in [6] applied illuminance-invariant features, such as edge maps, Local Binary Patterns (LBP), Gabor wavelets, and local autocorrelation filters. It has been also shown that face recognition using the skin model represented in the HSV V color space works robustly regardless of the lighting conditions [9]. The face overlap can be further improved by using brightness control or by rejecting pixels with low channel values [10].

Yet, the preprocessing step used to address various illuminations lead to the increase of the computational overhead and in many cases do not work for all lighting conditions [11]. Moreover, in visible light images, the privacy and ethical concerns are still valid, especially for medical applications and person monitoring solutions in smart environments. Thus, thermal imagery is often utilized in these systems [12], as it helps to secure privacy by representing objects as the temperature distribution instead of high level features.

The important research question, though, is whether person's identity can be revealed from thermal images by using Deep Neural Networks (DNN). Although feature representation in thermography is different than in visible light data, DNNs have recently gained a lot of popularity due to their human-like capabilities, so it is important to verify their robustness in the thermal domain. Also, it is important to determine whether various values of the image resolution have influence on the accuracy of the face recognition task. Thus, the aim of this study is to 1) verify the accuracy of deep learning based person recognition using facial features embedding extracted from thermal data and 2) evaluate the influence of resolution enhancement and degradation on the utilized biometrics. For embedding extraction, a model designed and trained on data acquired in visible light spectrum was used.

This work has been partially supported by Statutory Funds of Electronics, Telecommunications and Informatics Faculty, Gdansk University of Technology and by Intel Corporation, USA. We thank all our colleagues from Intel Corp., who provided insight and expertise that greatly assisted the research.

In this way, we determine whether it is possible to generate a useful facial features representation from thermal data by utilizing high frequency features learnt from RGB images.

The remaining of the work is organized into the following parts: Section II overviews the work related to the face recognition tasks using deep learning techniques. In Section III we present methods used in the study, including utilized datasets and face recognition models. The preliminary results are presented in Section IV and further discussed in Section V. Finally, the paper is concluded in Section VI.

II. Related work

Face recognition is a well studied problem. Since 2000s it has been actively developed reducing the training error rates by 3 orders of magnitude [13]. Various algorithms have already been proposed to solve face recognition problems, from relatively simple geometric feature based SVM [14], Principal Component Analysis (PCA) [15] and Histograms of Oriented Gradients (HOG) [16] to deep neural networks like DeepID3 [17] and DeepFace [18]. In this research we focus on the artificial intelligence field of this study, deep learning specifically. Thanks to the recent technological developments, both in software and hardware, the systems nowadays can achieve human-like performance in various cognitive tasks, including person recognition [19]. One of the main breakthroughs of face recognition that became state-of-the-art architecture was developed by F. Schroff et al. [20]. Contrary to the previous solutions [21] [18] based on fully connected classification layers FaceNet directly utilizes embedding optimizing the model using triplet loss function. As the model is based on learning the Euclidean embedding per image, the similarity between various faces can be calculated as a L2 distance between vectors representing faces - the lower the distance the more similar the two faces are. Once the vectors are generated they can be clustered together using standard machine learning algorithms, like k-NN, or classified using k-means.

The face recognition methods verified for visible light images were applied to recognize faces from thermal images (acquired in Short/Medium/Long Wave Infra Red ranges - SWIR/MWIR/LWIR). The traditional, two-step approach was typically used. First, characteristic features were extracted and next, the feature vectors were used used in classification/recognition. In [22] authors extracted texture features using local binary descriptors (LBP), local ternary descriptors (LTP), or differential LTP (DLTP). They applied their method for the Equinox multimodal facial image dataset [23] obtaining 96% accuracy for SWIR images. The LBP features were also used in [24] producing best results using Linear discriminant analysis (LDA) on LWIR images. Thermal images represent unique facial features related to temperature distribution. Due to the heat flow in objects, we can observe smooth changes between facial areas, resulting in low contrast and lack of high frequency components. Thus, deep neural networks trained on visible light images may not be sufficient for thermal data. In this study we would like to evaluate if it's possible to re-apply these networks to another image domain.

Some studies on the impact of image resolution on face recognition in visible light have been already conducted. As studies done in [25] [26] indicate, for existing face recognition algorithm to work the input resolution should have the resolution of at least 32 by 32 pixels. To the best of our knowledge, no such study exists for thermal data. In our experiments we evaluated different scales (100%, 50%, 25% and 12.5%) of the same image. The scaling was done after the image was cropped with SSD detection network to show only the facial region. As a result, the average resolution for the lowest scale (12.5%) was $13.14 \pm 1.47px$ by $15.57 \pm 1.96px$. This study further evaluates whether it's possible to recognize person from thermal image of low resolution and if it can be improved be utilizing Super Resolution algorithm. Super Resolution already has been proven to be beneficial for non-contact estimation of respiratory rate from thermal data [27]. We propose to further improve this pipeline by providing face recognition system working on very small input images achieving high accuracy, even when volunteers perform small head rotations or mimic various emotions.

III. Methodology

A. Datasets

Experiments were carried out on two thermal face datasets. First dataset that we evaluated, was created by us. This dataset (hereafter referred as SC3000-DB) contains of 766 images grouped into 40 categories, each category representing a single volunteer from a group of 19 males, 21 females, age: 34.11±12 (19-20 images per volunteer). The sequences were captured with the use of FLIR SC3000 thermal camera capable of recording 30 Frames Per Second (FPS) with spatial resolution of 320x240 in a noise reduction mode. The range of temperature measurement is from -20°C to +80°C. Data received with this camera is in 14-bit radiometric format, which in this experiment was linearly down-scaled to 8-bit greyscale PNG image format. The camera was placed on a stationary tripod at the height of 112cm from the ground and in a distance of 1.2m from the volunteer's face. The volunteers were asked to look directly in the camera for the time period of 2 minutes. From gathered sequences every 180^{th} frame was saved to the SC3000-DB database to limit the number of images that are too similar to each other.

To make the results of the experiment less biased towards a single dataset, the IRIS [28] set was chosen as a second database to reinforce the conclusions. The IRIS dataset consists of images recorded with a help of 30 individuals (total of 4190 thermal images). What sets this dataset apart from our own is that the volunteers were not focused on the camera, instead they performed subtle head movements. On average there are 11 images per head pose per volunteer. Additionally, there might be differences in facial features representation, as the dataset captures three different facial expressions: angry, surprised, and laughing. The IRIS dataset was recorded with the Raytheon Palm-IR-Pro camera with a spatial resolution of 320x240 and uncompressed bitmap as an output format.

Fig. 1: Sample images from both datasets: SC3000-DB (top row) and IRIS (bottom row).

Examples of thermal images from both datasets are presented in Fig. 1

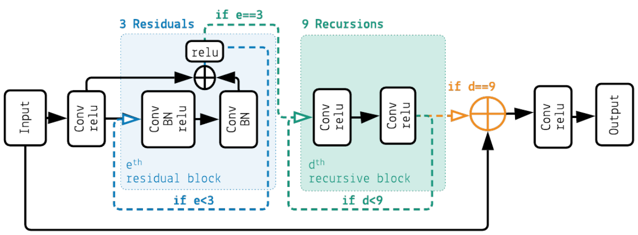

Fig. 2: Architecture of the applied CNN-based model for enhancing resolution of thermal images. Letter e denotes e^{th} residual block (out of 3) applied for extracting features that are then used to restore High Resolution image from Low Resolution input. Letter d denotes d^{th} recursion (out of 9) used in the part of the network which aims at performing non-linear mapping of extracted features to a vector used for image reconstruction.

B. Face detection

As the images in the datasets used for evaluation of face recognition often show more than just the face of the volunteer, in the first step it was instrumental to crop the visible area of the images to show only the region of face. To achieve this, the SSD model with Inception V2 backbone [29] was re-trained using transfer learning [30] on images from our database (the images were split in 70:15:15 proportion into training, validation and test sets). In order to tune hyperparameters, we used the random search approach [31] to find the best training configuration. It turned out that the best accuracy was achieved for 40k training steps, batch size 32, initial learning rate 0.004, learning rate decay steps 5000 and decay factor 0.95. The trained network achieved score of $84.1 \pm 6\%$ in Intersection over Union (IoU) metric on SC3000-DB test set. IoU is a popular metric used for evaluating object detection system performance by comparing similarity between ground-truth (GT) area and detected region (DR), defined as:

$$IoU = \frac{GT \bigcap DR}{GT \bigcup DR} \tag{1}$$

For comparison we trained second model on IRIS dataset using the same split proportions as for the SC3000-DB. On testing set model achieved lower score of $79.4 \pm 14\%$. For further processing the first model was chosen as it promised achieving better results. For next steps in the experiment we extracted cropped faces from whole SC3000-DB and IRIS sets. From all cropped faces 20% of images for each volunteer were used to create personal profiles via facial feature extraction, and the rest 80% was used to test face recognition accuracy.

C. Resolution degradation and enhancement

This study aims at evaluating the accuracy of person recognition from thermal imagery of various resolutions. To achieve this, we simulated resolution degradation by generating downscaled versions of original images from both databases. After cropping images to facial areas, the images were scaled down 2, 4 and 8 times. Resulting images were as small as $13.14 \pm 1.47px$ x $15.57 \pm 1.96px$ for the scale of 1/8. In this way, we created similar images, as would be acquired by using lower resolution thermal cameras.

Image enhancement was implemented by using Super Resolution (SR) Convolutional Neural Network (CNN). SR is a task of generating a high resolution (HR) output from the given low resolution (LR) image. The architecture of the model used in this work to improve image resolution is presented in Fig. 2. Similarly to the Deeply-recursive Convolutional Network (DRCN) [32], our model uses recursive blocks with shared weights to create a deeper representation, known to achieve a better performance [33] without introducing new parameters. Yet, to address the problem of feature blurring in thermal imagery, we decided to widen the receptive field by introducing residual blocks to the utilized CNN network. Weights are also shared across residual blocks, so the number of model parameters remains constant. By applying a set of residuals, the receptive wield is widened, so theoretically the model is able to solve the problem of contextual information spread over larger image regions, what is present in thermography due to the heat flow in objects. In addition, we also correlate input to the network with the restored HR output by using the skip connection. In practice, LR inputs are highly similar to HR images, except some detailed features, so preserving this relation by using the skip connection is helpful for enhancing images [34]. To choose the number of residuals and recursions, we randomly generated various network configurations and trained it on both IRIS and SC3000-DB train subsets (70% of images from each dataset used as a training set). The training set was downscaled with a scale of 2 before feeding images to the model. The hyperparameters were set as proposed in the DRCN model [32], i.e. Adam [35] optimizer, momentum 0.9, weight decay 0.0001, 96 filters of a size 3x3 each in all convolutional layers, all weights initialized using the Xavier algorithm. Following [34], training data were cropped to 41x41 patches with a stride of 21. From all trained networks, we then selected architecture that was producing the highest Peak-Signal-to-Noise Ratio (PSNR), i.e. 9 recursions and 3 residuals. This model achieved PSNR of 47.49dB for SC3000-DB and 34.93dB for IRIS. Corresponding LR sets (after downscaling them with a factor of 2 using bicubic interpolation) produced PSNR of 27.91dB and 29.4dB for SC3000-DB and IRIS, respectively. For further experiments, the trained model was utilized to generate enhanced versions of low resolution thermal images.

D. Facial feature extraction

Similarly to [19] we utilize facial features represented as the vector of embeddings extracted from cropped images with the use of FaceNet DNN architecture [20]. Yet, contrary to this work we use larger embedding's vector of size 512, which improved by small margin the ability to capture more subtle differences in the facial features as seen in the [20]'s implementation in TensorFlow [36].

In this experiment we used model trained on VGGFace2 [37] facial images database. It is worth highlighting that this database consists of images captured in the visible light spectrum. In this research we wanted to validate whether a model designed and trained on data with high frequency features (e.g. edges, corners) clearly present in visible light spectrum can be used to recognize face from thermal imaging.

E. Face recognition

In this step we tested two methods of comparing facial features vectors. First approach used Support Vector Machines (SVM) with linear kernel to find the class (person profile) for a given input image. Profiles in the database are represented as a list of all vectors calculated from the training images. Second approach was based on the Euclidean Distance between vector representations from the database profile and the input image. This method required each of the user profiles to be stored as a single vector. Thus, all vectors obtained when creating the given user's profile were averaged.

IV. Results

TABLE I: Accuracy [%] of person recognition from test set images (80% of all images); the reference embedding generated with 20% of images

svm.LinearSVC (all)								
test set	SC3000 (ours) train set				IRIS train set			
	orig.	bicub. 50%	bicub. 25%	SR (ours)	orig.	bicub. 50%	bicub. 25%	SR (ours)
orig.	99.5	-	-	-	82.14	-	-	-
bicub. 50%	-	99.17	-	-	-	81.33	-	-
bicub. 25%	-	-	96.36	-	-	-	74.01	-
SR (ours)	-	-	-	99.33	-	-	-	81.87
Euclidean (avg)								
test set	SC3000 (ours) train set				IRIS train set			
	orig.	bicub. 50%	bicub. 25%	SR (ours)	orig.	bicub. 50%	bicub. 25%	SR (ours)
orig.	99.66	-	-	-	63.48	-	-	-
bicub. 50%	-	98.67	-	-	-	58.01	-	-
bicub. 25%	-	-	90.42	-	-	-	57.68	-
SR (ours)	-	-	-	98.84	-	-	-	60.85

The Fig. 3 shows a 2D representation of embedding vectors acquired from images of original size and after down-scaling them with a factor of 2, 4 and 8 for chosen ten volunteers. Table I presents the accuracy of person recognition from thermal images by comparing the input image from the test set with with persons representations stored as a facial feature embedding vector in the database. Various dataset transformation were evaluated in this study.

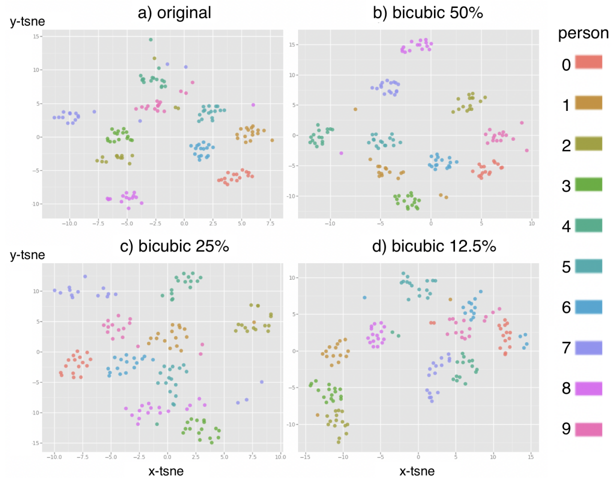

Fig. 3: 2D visualization of embedding vectors extracted from images of a) original size, b) downscaled to 50%, c) 25% and d) 12.5% of original size for first ten categories from SC3000-DB test set. High dimensionality was reduced using t-Distributed Stochastic Neighbouring Entities (t-SNE) [38] technique.

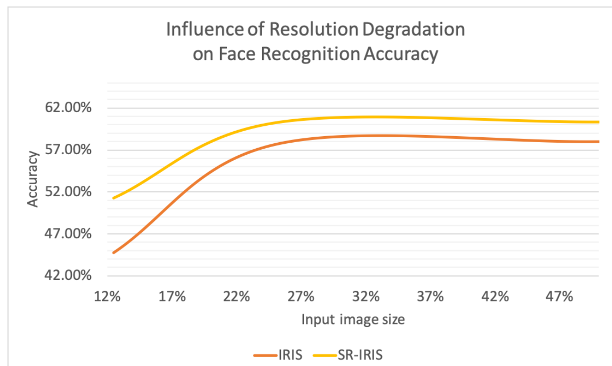

Fig. 4: Influence of resolution degradation and enhancement on face recognition accuracy measured with Euclidean distance on IRIS dataset.

The relation between the input image quality and the face recognition accuracy is presented in Fig. 4 and Fig. 5. The changes in image quality were introduced by first down-scaling the images be a given factor, then up-scaling them back to the original resolution. In this way we simulated low resolution inputs that were fed to Super Resolution model for resolution enhancement.

V. Discussion

Preliminary results of the experiment presented in this paper showed the feasibility of applying DNNs to the face recognition task in the thermal domain. In this research we have re-purposed the FaceNet model trained on RGB data to a new imaging domain. Even though it was trained to extract high frequency components, it was able to generalize well on the thermal images. As a result of applying this model on the chosen dataset, unique facial embeddings were extracted, allowing to distinguish various volunteers with the accuracy of 99.5% and 82.14% on SC3000-DB and IRIS datasets, respectively, using SVM classifier.

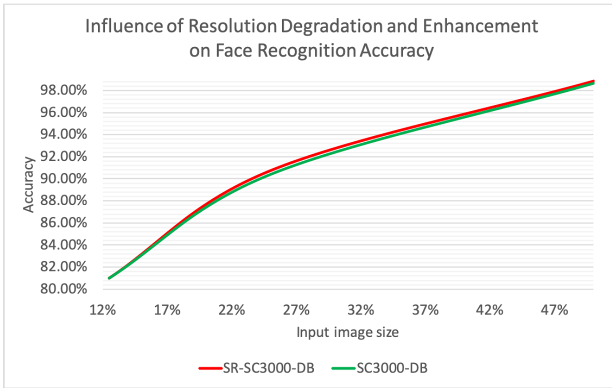

Fig. 5: Influence of resolution degradation and enhancement on face recognition accuracy measured with Euclidean distance on SC3000-DB dataset.

Satisfactory accuracy was achieved for Euclidean distance as well but only for SC300-DB set (accuracy $>$ 90% for images downscaled by a factor of 4). For IRIS dataset, results were much worse, producing accuracy of 63.5% for original data and only 57.7% for bicubic interpolation with a scale of 1/4. Also, Euclidean distance turned out to be worse than SVM, while evaluating IRIS dataset. For our dataset, results were similar regardless of the applied classification method, what can be explained by a fact that images in SC3000-DB were collected in strictly defined conditions resulting in a reduced influence of involuntary movements and background noise. Furthermore, our data have uniform projection from original 14bits to 8 bits, on which the experiment was performed, which partially explains the accuracy difference between these two datasets. In the future work we will assess the accuracy on data of higher precision format.

The SVM method offers a variety of kernels for pattern analysis. The choice of a proper kernel heavily depends on the structure of data and the factors like the number of features and the number of samples in the dataset. Our training dataset consists of relatively small number of samples (only 20% of images in each of the datasets were selected for volunteer's profiles - avg. of 4 images in SC3000-DB and avg. of 42 in IRIS set), while containing a large number of facial features in the extracted embedding.

It is noteworthy that both datasets were divided into 2-8 proportion for training and testing sets, respectively. In this work we were able to prove that even with a small number of images used for creating user profiles the DNN is able to recognize the same person from testing images with high accuracy of 99.5% and 82.14% for SC3000-DB and IRIS, respectively. Also, both datasets contain images that were gathered within short time span. This might have had an impact while measuring accuracy of facial recognition, as the data that was used for creating user profiles may be uniform with test data. As a part of the future work we would like to include experiments that include data gathered over longer time spans, which may also result in creation of a new thermal database for facial recognition.

Furthermore, as can be seen on Fig. 3 different classes can be linearly separable when considering multiclass classification with a one-against-one approach. Linear kernel also have the performance advantage over the non-linear ones, as it's simpler in terms of computations. Thus, we used a linear kernel that is often recommended when number of features is larger than number of observations. The performed experiments proved the robustness of the chosen kernel. The achieved accuracy of person recognition was very high, especially for our dataset that even after reducing resolution 4 times still produced results $>$ 96%.

From the Fig.3 we can also observe influence of resolution degradation on classes separability. In can be seen on the approximation that the original data can be easily divided into separate clusters. With the resolution decrease some of the features get smoothed resulting in a class boundary that often overlaps with other clusters, making it sometimes harder to achieve accurate classification. To help counter the issues with inputs of low resolution we applied Super Resolution model to all inputs smaller than the original one. As can be seen in Fig. 4 the accuracy of face recognition has improved when Super Resolution was applied to the up-scaled input images from the IRIS dataset that contain motion, facial expressions and some irregularities. On the SC3000-DB dataset that included less to no dynamic movements the improvements were minimal. This observation has been proved by the results presented in the Table I. However, considering real-life scenarios, it's hard to enforce on the potential users to remain motionless for a longer period of time. The real life scenarios could include remote monitoring of vital signs [39] and user's authentication [40] e.g. to smart home devices. In this cases application of SR seems crucial to achieve high accuracy. Moreover, as shown in [27] SR can help not only with improving accuracy of the face recognition, but also of the breathing rate evaluation. By combining both systems, we could potentially build a home-based patient monitoring system that is able to track user-specific vital signs changes during daily activities.

VI. Conclusion

This study aimed at verifying whether image resolution has influence on person recognition task using facial features embedding gathered from thermal imagery. In addition, it was evaluated if model trained to extract high frequency components (on RGB images) will be able to generate a meaningful person embedding from thermal data that is characterized by blurring and smoothness due to the heat flow in objects. The preliminary results proved the need to enhance image resolution in order to achieve a high accuracy of person recognition. The presented Super-Resolution solution allowed to improve results of person recognition from images downscaled with bicubic interpolation by 8 % for the resizing scale of 4 on the IRIS dataset. Yet, for our database that assumed strictly defined measurement conditions (no movements, volunteer loking toward the camera) we did not observe any gain of performance.

Thus, we would like to perform similar research on data collected by us but during various measuring scenarios, e.g.

volunteers turning their head horizontally and vertically. In future study, we would also like to evaluate the proposed approach of person recognition on images with original bit resolution (14-bit represented as 16-bit image to avoid lossy conversion to 8-bit data).

References

[1] C. Cruz, L. E. Sucar, and E. F. Morales, "Real-time face recognition for human-robot interaction," in *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*. IEEE, 2008, pp. 1–6.

[2] H. Meng, N. Pears, and C. Bailey, "A human action recognition system for embedded computer vision application," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–6.

[3] J. Choi, Y.-i. Cho, T. Han, and H. S. Yang, "A view-based real-time human action recognition system as an interface for human computer interaction," in *International Conference on Virtual Systems and Multimedia*. Springer, 2007, pp. 112–120.

[4] S. R. Fanello, I. Gori, G. Metta, and F. Odone, "Keep it simple and sparse: Real-time action recognition," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 2617–2640, 2013.

[5] G. Gkioxari, R. Girshick, P. Dollár, and K. He, "Detecting and recognizing human-object interactions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8359–8367.

[6] P. Kalaiselvi and S. Nithya, "Face recognition system under varying lighting conditions," *IOSR Journal of Computer Engineering (IOSR-JCE)*, vol. 14, no. 3, pp. 79–88, 2013.

[7] S. Prabhakar, S. Pankanti, and A. K. Jain, "Biometric recognition: Security and privacy concerns," *IEEE security & privacy*, no. 2, pp. 33–42, 2003.

[8] F. Huang and H. Bian, "Identity authentication system using face recognition techniques in human-computer interaction," in *Proceedings of the 32nd Chinese Control Conference*. IEEE, 2013, pp. 3823–3827.

[9] M. Li, Y. Sun, B. Yuan, and Y. You, "Research on the automatic human face recognition system," in *Proceedings 7th International Conference on Signal Processing, 2004. Proceedings. ICSP'04. 2004.*, vol. 3. IEEE, 2004, pp. 2584–2587.

[10] J. B. Martinkauppi, M. N. Soriano, and M. V. Laaksonen, "Behavior of skin color under varying illumination seen by different cameras at different color spaces," in *Machine Vision Applications in Industrial Inspection IX*, vol. 4301. International Society for Optics and Photonics, 2001, pp. 102–113.

[11] A. Kwaśniewska, J. Klimiuk-Myszk, J. Ruminski, J. Forrier, B. Martin, and I. Pecci, "Quality of graphical markers for the needs of eyewear devices," in *2015 8th International Conference on Human System Interaction (HSI)*. IEEE, 2015, pp. 388–395.

[12] S. Kido, T. Miyasaka, T. Tanaka, T. Shimizu, and T. Saga, "Fall detection in toilet rooms using thermal imaging sensors," in *2009 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 2009, pp. 83–88.

[13] P. J. Phillips, J. R. Beveridge, B. A. Draper, G. Givens, A. J. O'Toole, D. S. Bolme, J. Dunlop, Y. M. Lui, H. Sahibzada, and S. Weimer, *An introduction to the good, the bad, & the ugly face recognition challenge problem*. IEEE, 2011.

[14] B. Heisele, P. Ho, and T. Poggio, "Face recognition with support vector machines: Global versus component-based approach," in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 2. IEEE, 2001, pp. 688–694.

[15] M. Krišto and M. Ivasic-Kos, "An overview of thermal face recognition methods," in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 2018, pp. 1098–1103.

[16] A. Albiol, D. Monzo, A. Martin, J. Sastre, and A. Albiol, "Face recognition using hog–ebgm," *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1537–1543, 2008.

[17] Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," *arXiv preprint arXiv:1502.00873*, 2015.

[18] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1701–1708.

[19] M. Szankin, A. Kwasniewska, T. Sirlapu, M. Wang, J. Ruminski, R. Nicolas, and M. Bartscherer, "Long distance vital signs monitoring with person identification for smart home solutions," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 1558–1561.

[20] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.

[21] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2892–2900.

[22] M. A. Akhloufi and A. Bendada, "Infrared face recognition using texture descriptors," in *Thermosense XXXII*, vol. 7661. International Society for Optics and Photonics, 2010, p. 766109.

[23] A. Selinger and D. A. Socolinsky, "Appearance-based facial recognition using visible and thermal imagery: a comparative study," EQUINOX CORP NEW YORK NY, Tech. Rep., 2006.

[24] M. A. Akhloufi and A. Bendada, "Probabilistic bayesian framework for infrared face recognition," in *proc. International Conference on Machine Vision, Image Processing, and Pattern Analysis*, 2009, pp. 66–70.

[25] E. Mostafa, R. Hammoud, A. Ali, and A. Farag, "Face recognition in low resolution thermal images," *Computer Vision and Image Understanding*, vol. 117, no. 12, pp. 1689–1694, 2013.

[26] Y. M. Lui, D. Bolme, B. A. Draper, J. R. Beveridge, G. Givens, and P. J. Phillips, "A meta-analysis of face recognition covariates," in *2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*. IEEE, 2009, pp. 1–8.

[27] A. Kwasniewska, M. Szankin, J. Ruminski, and M. Kaczmarek, "Evaluating accuracy of respiratory rate estimation from super resolved thermal imagery," in *2019 41th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019 in print.

[28] "Iris thermal/visible face database," http://vcipl-okstate.org/pbvs/bench/Data/02/download.html, accessed: 2018-09-30.

[29] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[30] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*. IGI Global, 2010, pp. 242–264.

[31] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, no. Feb, pp. 281–305, 2012.

[32] J. Kim, J. Kwon Lee, and K. Mu Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1637–1645.

[33] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1.

[34] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1646–1654.

[35] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[36] "Facenet: A unified embedding for face recognition and clustering implementation in tensorflow," https://github.com/davidsandberg/facenet, accessed: 2019-04-15.

[37] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 67–74.

[38] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[39] J. Ruminski and A. Kwasniewska, "Evaluation of respiration rate using thermal imaging in mobile conditions," in *Application of Infrared to Biomedical Sciences*. Springer, 2017, pp. 311–346.

[40] S. Athira and O. R. Murthy, "Face authentication using thermal imaging," in *Computational Vision and Bio Inspired Computing*. Springer, 2018, pp. 1006–1014.