A Topological Approach for Facial Region Segmentation in Thermal Images

Michael Lilley
Department of Computer and Information Science
University of Michigan-Dearborn
Dearborn, USA
mlilley@umich.edu

Kapotaksha Das
Department of Computer and Information Science
University of Michigan-Dearborn
Dearborn, USA
takposha@umich.edu

Kais Riani
Department of Computer and Information Science
University of Michigan-Dearborn
Dearborn, USA
kriani@umich.edu

Mohamed Abouelenien
Department of Computer and Information Science
University of Michigan-Dearborn
Dearborn, USA
zmohamed@umich.edu

2022 IEEE International Symposium on Multimedia (ISM) | 978-1-6654-7172-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISM55400.2022.00042

***Abstract*—The use of thermal images as a non-contact modality has expanded drastically in recent years, including applications such as biometrics, detection of specific human behaviors, and extraction of physiological signals. For instance, using thermal images to monitor physiological signals was found to provide better information compared to RGB images, especially in situations where using physical contact sensors is not a possibility. In this paper, we present a novel topological approach that can segment the cheeks and forehead of a human face in a thermal image, with the key benefit that explicit prior information about the cheeks and forehead is unneeded. Our approach leverages topological properties present in a thermal image to provide a non-rectangular bounding curve for the cheeks and forehead, which we compare against a dataset of 1000 thermal images that were manually annotated for this task. By generating a Vietoris-Rips complex on a thermal image filtered by the Canny edge detector, our approach can segment the cheeks and forehead with recall scores as high as 90.4% and 78.4%, respectively.**

***Index Terms*—thermal modality, region of interest segmentation, topological approach**

I. Introduction

Thermal imaging has become ubiquitous in many fields of computational science; particularly so in healthcare, engineering, and manufacturing. In applications where humans play a critical role, understanding the psychological conditions which they may be exhibiting is invaluable, both for reasons of safety and tailored functionality. For instance, thermal faces were analyzed in order to model the alertness levels of drivers by mapping their facial temperature distribution [1], [2]. Moreover, thermal images have been used recently for biometrics. For example, research by Hu et al. [3] and Szankin et al. [4] explored the use of thermal imgaging for biometric identification, especially in low-light scenarios. Furthermore, thermal images have been utilized as a non-contact way of estimating latent variables that are challenging to predict with traditional RGB images; in research by Pavlidis et al., thermal images were used to detect anxiety within individuals [5].

Despite this slew of benefits, it is important to note that thermal images suffer from key drawbacks compared to RGB images. This includes the inability to effectively use well-established techniques for RGB images, such as the Viola-Jones algorithm [6]. Furthermore, when compared to RGB data, the annotated thermal data available in the scale and quality required to implement deep learning models is quite restricted. The challenges associated with thermal imagery have compelled us to explore segmentation methods that do not require explicit prior information about the facial structure.

This research is motivated by the need of identifying optimal regions of interest (ROIs) in the thermal faces, such as the cheeks and forehead. In tasks such as detecting heart rates and respiration rates from thermal images, research by Barbosa et al. [7] and Elphick et al. [8] showed that it is necessary to have some form of ROI identification prior to any form of signal prediction. Other research also indicated the importance of utilizing specific regions in the thermal faces to model certain human behaviors, such as sleepiness, discomfort, and alertness, among others [9]. Moreover, having bounding curves over bounding boxes that more accurately reflect the shape of an ROI further increases the quality of information obtained [10], giving us further reason to present a methodology for non-rectangular thermal segmentation.

In particular, we make the following contributions: first, we use a dataset of high-resolution thermal images in order to segment the cheeks and forehead of the face without using the visual RGB modality. Secondly, we utilize a novel topological approach that uses Shi-Tomasi corner detection as the basis to generate a Vietoris-Rips complex, with the holes contained in the complex corresponding to non-rectangular regions inside the thermal faces. Finally, our approach toward thermal segmentation requires no prior information about the subject whose face is being segmented, leveraging unsupervised training techniques in conjunction with various computer vision algorithms in a complete pipeline for segmentation.

978-1-6654-7172-5/22/$31.00 ©2022 IEEE
DOI 10.1109/ISM55400.2022.00042

II. Related Work

While there exists several approaches to image segmentation as discussed in the Survey by Yuheng and Hao [11], these techniques apply primarily to RGB images. Even in the RGB image space, face detection is also often a challenging problem to solve. Despite this, it has attracted a lot of interest in recent years due to its applications in a variety of disciplines. Face detection and recognition is challenging under varying light conditions which could be addressed using thermal cameras, as the sensitivity to illumination changes is low [12].

More recently, Knapik et al. [13] proposed a new approach to detect eyes using thermal images. The work by Tashakori et al. in 2021 [14] used facial thermal imaging in a driver simulator to predict the drowsiness status of a driver. As discussed by Mekyska et al. [15], despite the benefits thermal images could provide, the usage of well established RGB segmentation algorithms could simply not be used effectively in thermal images. Abouelenien et al. [16] also showed that the Viola-Jones algorithm does not perform well at segmenting thermal regions of interest.

This leads us towards research that approaches human thermal segmentation by using a mixed modality approach such as in Coreneau et al. [17] where thermal signatures, 3D data, and RGB images were used in conjunction for isolating regions in the face, or in Palmero et al. [18] where they presented a joint RGB and thermal dataset and approach for segmentation of the human body. However, these approaches required the presence of other modalities, primarily RGB, which meant the need for at least two cameras to monitor a subject. The use of transfer learning principles in deep learning is an approach to facial recognition that has been investigated by Saxena et al. [19], where they explored the potential of using a pre-trained features from CNN on visible spectrum face images to perform heterogeneous face recognition in the thermal domain. Like before, these approaches were limited by their necessity of requiring either a multimodal dataset or a dataset of RGB images, not being thermal-only approaches to this task.

A thermal image's latent information allows one to detect emotional state in subjects, providing a means for them to be used in applications where an RGB camera might not have a suitable environment to capture information in. The relevance of segmentation to this was described in a study published by Or and Duffy [20], where it was found that various parts of the face correlated differently with stress, with the nose showing a consistent magnitude of temperature changes under periods of increased mental load. Despite studies like these providing credence to the notion that thermal imaging generates data otherwise unattainable by traditional methods, this has been afflicted by the the fact that images of this type are of very low contrast, making the isolation of regions of interest particularly difficult. Other research employed deep learning for thermal image segmentation tasks [21], [22], but these methods suffered from the need of large quantities of data that are often not available in practice, bringing us to the need of finding segmentation techniques that can be less dependent on prior data.

III. Methodology

Our proposed method builds on the observation that there exists a difference in the average temperature gradient between distinct regions of the face. This is particularly apparent when considering flat regions of the face with little structural variation, and the edges of facial features with intricate structure, such the eyes and nose. The former exhibits a relatively consistent temperature across the entire region, whereas the latter has low temperature gradient entropy. We aim to take advantage of this property in order to perform segmentation of the cheeks and forehead within thermal images of the face.

A. Bag of Visual Words Generation

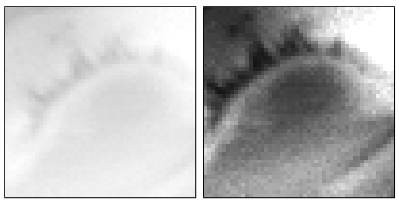


Fig. 1: Left - Before VHDR, Right - After applying VHDR

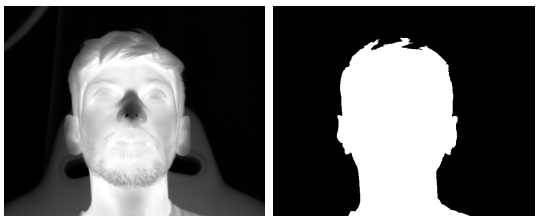
We first apply the virtual high dynamic range (VHDR) filter [23] to the images in our dataset of facial features in order to enhance the details of the extracted images as seen in Fig. 1. The VHDR filter applies nonlinear tone adjustment curves to bring out a wider dynamic range in an image to increase the likelihood of identifying features that would otherwise be imperceptible. Following this, we collect the descriptor objects which result from applying dense Scale Invariant Feature Transform into one set [24]. Given that some descriptors are representative of the same feature and thus could result in redundant vector representations, we then reduce the size of the descriptor set through K-means clustering. This compresses the dataset by reducing redundancy without removing important features from our bag-of-visual-words.

B. Feature Detection

Segmentation of the face is achieved utilizing Otsu's thresholding on the non-VHDR images, as the thermal image background is distinctively lower in temperature than the subject. Following this, we apply a floodfill algorithm to close any gaps in the binarized image, a representative example of which is seen in Fig. 2.

Next, we estimate the centerpoint of the binarized image; this is computed as the center of mass of the white pixels in the image. After this, we mark the left, right, and top boundaries of the set of white pixels. We interpolate a bottom boundary by measuring the y-distance between the top boundary and the centerpoint, then draw an equidistant boundary underneath.

Once again, we then apply the VHDR filter to the entire (original) image. With our external boundaries established, we can now extract our features of interest. We generate a set of SIFT descriptors for a pixel lattice of the same density over



(a) Original Thermal Image (b) Otsu's Thresholding and Floodfill applied

Fig. 2: Example for Face Isolation using Otsu's Thresholding

our test image. We can then generate a normalized colormap as seen in Fig. 3. corresponding to the minimum Euclidean distance between a descriptor $f(i,j)$ and the vectors in Λ_K.

Following this, we perform salient point clustering by picking the centroid of the largest clusters as the location of our features. In the case of the eyes, the centers of the two largest clusters are selected, whereas we only select the center of the largest cluster when searching for the nose.

After we locate the eyes and nose in the image, we utilize these to segment the face into three bounding boxes which are meant to separate the left and right cheeks, as well as the forehead as seen in Fig. 5. We then apply our algorithm to each of these three sub-regions individually.

C. Segmentation of Cheeks and Forehead

To the original thermal image we apply a Canny filter in order to get the relevant edges of the face. The two threshold parameters associated with the Canny algorithm are determined experimentally, with the goal of generating a binary image I_c which identifies the large temperature gradients on the edges of facial features, but ignores the subtle ones present in the skin as seen in Fig. 3.

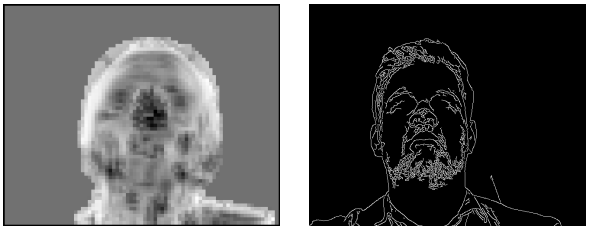


Fig. 3: Left: Example of Generated Colormap using SIFT. Right: Example of image I_c generated using properly tuned Canny settings

Once the optimal Canny settings are found, we formulate the task of segmenting the cheeks and forehead as a problem of estimating the boundaries of three large regions in the face that contain only zero-valued pixels in the Canny image I_c. For this we use tools meant for triangulating topological spaces.

We begin by applying the Shi-Tomasi (ST) corner detection algorithm [25]. Given the ST point cloud V, we can use the Vietoris-Rips complex generation algorithm [26] to generate a simplicial complex which is defined in Algorithm 1. Here, r is the radius used to compute the Vietoris-Rips complex, which determines the formation of lines between any two given points in the point cloud by applying an upper limit on the line length. Fig. 4 shows the effect of varying r for a sample set of points on the produced Vietoris-Rips complex.

Algorithm 1: Vietoris-Rips Complex Generation

Data: $V, r \in \mathbb{N}$
Result: $\mathcal{K}$
$\mathcal{K} \leftarrow \varnothing$;
for $(v_1, v_2) \in V \times V$ **do**
  if $||v_1 - v_2|| < r$ **then**
    $\mathcal{K} \leftarrow \mathcal{K} \cup \{(v_1, v_2)\}$

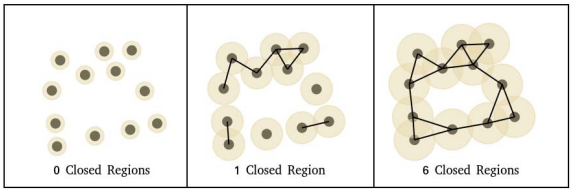


Fig. 4: Generation of Closed Partition on varying r in the Vietoris-Rips Complex Generation Algorithm

Once these points are found, we apply a heuristic to further improve our process: In the four corners of the bounding box surrounding the face, there are small black regions which intersect with the background; the pixels contained in these are exactly those inside of the bounding box described in section B6 not contained in the segmented face computed in section B1. We tile these regions with equidistant points so that the algorithm does not mistake these as one of the large contours.

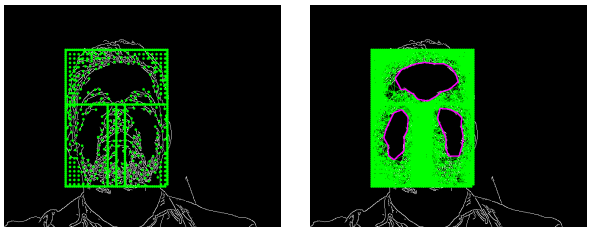


Fig. 5: Left: Bounding Boxes and Point Cloud V Generated From Canny Image I_c Right: Contour Regions mapped from Generated Closed Partitions from V

Finally, using the generated set of closed partitions, we apply a contour finding algorithm to detect the contours for the three regions as seen in Fig. 5.

IV. Experimental Setup

For our experiments, we sampled thermal images collected from across 45 subjects of varying ethnicities in a simulated setting. The dataset includes 30 males and 15 females ranging in age from 20 to 33 years old. All the thermal images were captured frames from a recording of the subject's face taken at 100 frames per second using a FLIR SC6700 thermal camera with a resolution of 640x512 pixels and 7.2M electrons.

For experimentation, we sampled 1000 frames randomly across the whole dataset, excluding those used for building

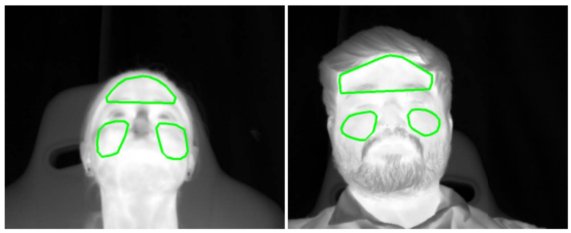


Fig. 6: Examples of Manually Annotated Images

the bag of visual words. We then used Supervisely [27], an image annotation tool to manually build a ground truth dataset of the cheeks and forehead across all the images. Fig. 6 shows some examples of the annotations on an image.

We have optimized our experiments for the following hyperparameters: *Cluster Count*, which is the total number of individual clusters during the process of K-means clustering. *Point Count*, which is the total number of points contained in all clusters. *Cluster Point Distance*, which is the maximum distance a point contained in a particular cluster may be from the cluster's center. *Vietoris-Rips (VR) Radius*, which is the set radius r used when computing the Vietoris-Rips complex. *Canny Low Threshold*, which is the lower threshold used to prune edges in the Canny edge detector. *Canny High Threshold*, or the upper threshold used to prune edges in the Canny Edge detector.

V. Results & Analysis

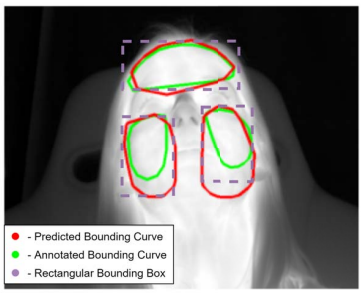


Fig. 7: Example of Bounding Curves for a Thermal Image

Fig. 7 represents an example of how our approach generates a bounding curve, shown in red, with the annotated curve shown in green. In results that we discuss further in Table III below, we see that the cheeks reaches a recall score of 90.4% while the cheeks reaches a precision of 71.6%.

We first place our focus on the effect the heuristic parameters have on our results, looking at the effect of each parameter on the algorithm's performance by taking the mean recall across each parameter value as seen in Table I. It appears that the heuristic parameters alone are not a significant contributing factor to the performance. However, this ignores the effect of the parameters on each other as dependent variables, which we analyze further in Table III. We also observe that the Canny threshold values have the largest effect on the performance of the algorithm.

To understand the effects of the thresholding parameters on performance, we consider Table II, which contains the best

TABLE I: Average Recall Seen across Cluster Count, Point Count and Distance Parameters

Parameter Name	Parameter Value	Mean Recall - Cheeks	Mean Recall - Forehead
Cluster Count	10	0.756	0.603
	20	0.748	0.609
Point Count	2500	0.751	0.608
	3000	0.752	0.604
Cluster Point Distance	35	0.752	0.606
	55	0.752	0.606
Vietoris-Rips Radius	25	0.756	0.624
	33	0.748	0.588
Canny Low Threshold	20	0.754	0.614
	40	0.809	0.67
	50	0.801	0.66
Canny High Threshold	70	0.867	0.726
	100	0.824	0.71
	150	0.688	0.508

TABLE II: Highest Recall Seen across each Region when varying the Initial Radius and Upper Canny values

Canny Low Threshold	Canny High Threshold	Max Recall - Cheeks	Max Recall - Forehead
20	70	0.893	0.776
	100	0.878	0.734
	150	0.721	0.558
40	70	0.911	0.794
	100	0.878	0.75
	150	0.726	0.511
50	70	0.912	0.783
	100	0.875	0.744
	150	0.733	0.516

computed recall for all combinations of Canny parameters that were experimented with. We see that the best recall values for the lower and upper thresholds were 50 and 70 respectively for the cheeks region with a score of 0.912, and 40 and 70 respectively for the forehead region with a score of 0.794. The values used in the hysteresis thresholding procedure are following non-maximum suppression filtering within the Canny edge detector.

We find that the measured precision for the cheeks tends to be lower than that of the forehead as seen in Table III. This is because in many subjects, a strong edge is not produced for the jawline, causing the boundaries for the cheeks to extend below the mouth as seen in Fig. 7. However, compared to a rectangular bounding box for identifying a region which will also inadvertently include the background, the algorithm's bounding curve is always within the subject's face. Across all 1000 images we used, 99.3% of all the pixels contained by the estimated bounding curve lies entirely within the subject's face. The ability to draw non-rectangular regions greatly reduces the number of wasted background pixels which could be useful in different scenarios, such as vision models where better region estimation will allow for a model to train solely on the temperature gradients present on the face rather than the sharp gradient between the face and the background.

Finally, we take a look at an effect of the other non-Canny parameters in an optimal Canny threshold setting. Table III has the highest recalls seen for the cheeks and forehead as well as the corresponding precision per region. Fixing the Canny

TABLE III: Highest Recall seen when using 40 as the Canny low threshold and 70 as the Canny high threshold

	Cluster Count	Point Count	Cluster Point Distance	Vietoris-Rips Radius	Recall - Cheeks	Recall - Forehead	Precision - Cheeks	Precision - Forehead
Cheeks	10	3000	35	33	0.904	0.752	0.454	0.716
Forehead	20	2500	35	25	0.88	0.794	0.376	0.585

values shows that the cluster point and VR radius parameters affect the performance of a region, with the cheeks preferring a lower cluster count and higher VR radius, and the converse for the forehead. This might also open up the potential of future work where applying different parameters to the images based on the region of interest could allow for further improvements of thermal region detection.

VI. Conclusion

With the expansion of the usage of thermal imaging for a variety of purposes, the need arises for methods that go beyond detecting thermal faces to approaches that reliably segment thermal ROIs in the faces. In this paper, we introduced a novel pipeline that showed to be effective in isolating specific ROIs in the thermal faces. We built a thermal image dataset in order to develop a novel topological approach towards non-rectangular segmentation of the cheeks and forehead from a thermal image of a subject. By taking advantage of the temperature gradients present on the face, we could employ topological algorithms that allowed for us to design an approach that requires no prior knowledge of the subject beforehand. Furthermore, our approach allowed us to generate non-rectangular bounding curves to segment regions that exclude the background entirely. The proposed thermal segmentation pipeline used Canny edge detection and the Shi-Tomasi corner detection algorithm, followed by applying the Vietoris-Rips complex generation algorithm that could generate closed regions around the cheeks and forehead. Finally, using our dataset of 1000 thermal images, our algorithm obtained recall figures as high as 90.4% for the cheeks and 79.4% for the forehead, indicating the effectiveness of our proposed pipeline and contributing to an area of research that still has not been extensively explored.

References

[1] K. Riani, S. Sharak, K. Das, M. Abouelenien, M. Burzo, R. Mihalcea, J. Elson, C. Maranville, K. Prakah-Asante, and W. Manzoor, "Towards classifying human circadian rhythm using multiple modalities," in *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2021, pp. 1–8.

[2] K. Das, S. Sharak, K. Riani, M. Abouelenien, M. Burzo, and M. Papakostas, "Multimodal detection of drivers drowsiness and distraction," in *Proceedings of the 2021 International Conference on Multimodal Interaction*, ser. ICMI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 416–424.

[3] S. Hu, N. Short, K. Gurton, and B. Riggan, "Overview of polarimetric thermal imaging for biometrics," in *Polarization: Measurement, Analysis, and Remote Sensing XIII*, D. B. Chenault and D. H. Goldstein, Eds., vol. 10655, International Society for Optics and Photonics. SPIE, 2018, pp. 1 – 8. [Online]. Available: https://doi.org/10.1117/12.2299761

[4] M. Szankin, A. Kwasniewska, and J. Ruminski, "Influence of thermal imagery resolution on accuracy of deep learning based face recognition," in *2019 12th International Conference on Human System Interaction (HSI)*, 2019, pp. 1–6.

[5] I. Pavlidis, J. Levine, and P. Baukol, "Thermal imaging for anxiety detection," in *Proceedings IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications (Cat. No.PR00640)*, 2000, pp. 104–109.

[6] K. Reese, Y. Zheng, and A. Elmaghraby, "A comparison of face detection algorithms in visible and thermal spectrums," in *Int'l Conf. on Advances in Computer Science and Application*. Citeseer, 2012.

[7] C. Barbosa Pereira, M. Czaplik, V. Blazek, S. Leonhardt, and D. Teichmann, "Monitoring of cardiorespiratory signals using thermal imaging: A pilot study on healthy human subjects," *Sensors (Basel, Switzerland)*, vol. 18, no. 5, p. 1541, May 2018.

[8] H. E. Elphick, A. H. Alkali, R. K. Kingshott, D. Burke, and R. Saatchi, "Exploratory study to evaluate respiratory rate using a thermal imaging camera," *Respiration*, vol. 97, no. 3, pp. 205–212, 2019.

[9] K. Das, M. Papakostas, K. Riani, A. Gasiorowski, M. Abouelenien, M. Burzo, and R. Mihalcea, "Detection and recognition of driver distraction using multimodal signals," *ACM Trans. Interact. Intell. Syst.*, Feb 2022.

[10] P. Kapoor, S. Patni *et al.*, "Image segmentation and asymmetry analysis of breast thermograms for tumor detection," *International Journal of Computer Applications*, vol. 50, no. 9, 2012.

[11] S. Yuheng and Y. Hao, "Image segmentation algorithms overview," 2017.

[12] L. B. Wolff, D. A. Socolinsky, and C. K. Eveland, "Quantitative measurement of illumination invariance for face recognition using thermal infrared imagery," in *Proceedings of SPIE*, vol. 4820, 2002.

[13] M. Knapik and B. Cyganek, "Fast eyes detection in thermal images," *Multimedia Tools and Applications*, vol. 80, no. 3, pp. 3601–3621, 2021.

[14] M. Tashakori, A. Nahvi, and S. Ebrahimian Hadi Kiashari, "Driver drowsiness detection using facial thermal imaging in a driving simulator," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of engineering in medicine*, vol. 236, no. 1, pp. 43–55, 2022.

[15] J. Mekyska, V. Espinosa-Duró, and M. Faundez-Zanuy, "Mekyska," in *44th Annual 2010 IEEE International Carnahan Conference on Security Technology*, 2010, pp. 185–189.

[16] M. Abouelenien, V. Pérez-Rosas, R. Mihalcea, and M. Burzo, "Detecting deceptive behavior via integration of discriminative features from multiple modalities," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 5, pp. 1042–1055, 2017.

[17] C. A. Corneanu, M. O. Simón, J. F. Cohn, and S. E. Guerrero, "Survey on rgb, 3d, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 8, pp. 1548–1568, 2016.

[18] C. Palmero, A. Clapés, C. Holmberg Bahnsen, A. Møgelmose, T. Moeslund, and S. Escalera, "Multi-modal rgb–depth–thermal human body segmentation," *International Journal of Computer Vision*, vol. 118, 2016.

[19] S. Saxena and J. Verbeek, "Heterogeneous face recognition with cnns," in *European conference on computer vision*. Springer, 2016.

[20] C. K. Or and V. G. Duffy, "Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement," *Occupational ergonomics*, vol. 7, pp. 83–94, 2007.

[21] A. Elmahmudi and H. Ugail, "Deep face recognition using imperfect facial data," *Future Generation Computer Systems*, vol. 99, 2019.

[22] W.-T. Chu and Y.-H. Liu, "Thermal facial landmark detection by deep multi-task learning," in *2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP)*, 2019, pp. 1–6.

[23] M. Knapik and B. Cyganek, "Fast eyes detection in thermal images," *Multimedia Tools and Applications*, vol. 80, no. 3, pp. 3601–3621, Jan 2021. [Online]. Available: https://doi.org/10.1007/s11042-020-09403-6

[24] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–, 11 2004.

[25] J. Shi and Tomasi, "Good features to track," in *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1994.

[26] H. Edelsbrunner and J. Harer, *Computational Topology: An Introduction*, ser. Applied Mathematics. American Mathematical Society, 2010. [Online]. Available: https://books.google.com/books?id=MDXa6gFRZuIC

[27] "Supervise.ly," https://supervise.ly/, accessed: 2022-02-20.