Hybrid Model Recognition and Classification of Human Emotions in Thermal Images

Edmundo Bonilla-Huerta, Roberto Morales-Caporal, Eduardo Sánchez-Lucero, Crispín Hernández-Hernández and Yesenia N. González-Meneses

***Abstract* – The facial emotion detection is an area of opportunity to apply new methods of digital image processing, machine learning and deep learning. However, there are some facial features that are difficult to analyze by any classification system in RGB images. In this article, a hybrid facial recognition system is proposed which analyses images in the infrared plane to improve the recognition and classification of different facial expressions (happiness, tenderness, fear and sadness). The thermal images were processed by using Digital Processing Images to extract and segment main facial features. A Fuzzy Inference System is proposed to better categorize human facial expressions. The experiments obtained with our hybrid model are very promising to detect and classify basic human emotions, especially to categorize happiness in individuals with high classification performance.**

***Index Terms* — Emotion, Fuzzy Inference System, Image Processing, Thermal Images**

I. Introduction

Since the last decades, it has been questioned whether machines that can solve almost any activity are also capable of feeling emotions. Today it is known that the full range of human emotions are registered in specific regions of the brain. There are positive and negative emotions that correspond to stimuli that human beings perceive through of our senses of our environment. Therefore, providing emotions in machines is a computational challenge to be addressed from a discipline derived from artificial intelligence, which is responsible for the study of emotions, known as Affective Computing, to design more intelligent and more humanistic systems and machines [1], [2], [3], [4], [5], [6], [7].

A. Human Emotions

Human emotions can be categorized as physical and cognitive. Affective systems design requires an understanding of both [2]. An emotion generally describes a sensation or a feeling regarding a situation. Affective computing can be performed in a unimodal or multimodal way to have a better classification of emotional states. A psychological study conducted by [8] reveals that human emotions are expressed through visual signals of the face (55%), voice (38%) and language (7%).

TecNM/Apizaco. E. Bonilla-Huerta, R. Morales Caporal, E. Lucero Sánchez, C. Hernández Hernández and Y. González Meneses. Computer Science Department. Apizaco, Tlaxcala, México (edmundo.bh, eduardo.sl, crispin.hh roberto.mc,yesenia.gm@apizaco.tecnm.mx).

B. Detection, Recognition and facial emotions classification

In the literature, different algorithms have been proposed to detect and classify emotions in human faces; Most of this works addresses the facial recognition features problem based on regions of interest (ROI). These features are mainly the eyes, the lips, the mouth, the eyebrows, even the nose. These methodologies are carried out in the visible plane through cameras. In recent years, analyzing facial characteristics using thermal cameras has recently attracted attention. However, the algorithms proposed to analyze the visible plane of the human face do not work correctly in the infrared plane [9]. Thermal images offer new perspectives to propose algorithms that explore all the data contained in the infrared plane. Infrared images are an excellent alternative to analyze several kinds of facial expressions in human faces such as gestures, lexical features and micro expressions to classify emotions [10], [11], [12].

C. Thermal Images Dataset

Thermography is a very promising technology for face detection, especially when variations exist in light, in low light conditions or even in darkness. Different public datasets were reported in literature for thermal images [11], [13], [14], [15]. The Asian database [12] aim to study multiple emotions, consisting basically of a selection of movie clips from Asian regions. In contrast, LATEMO-E dataset [16] consist of several clips for Latinoamerica region selected from specialists. Due to the limitations of stimulus response from different video-clips in foreign languages, in this work, LATEMO-E dataset was selected for the human emotion elicitation stimuli and an own dataset composed of several images to induce basic emotions in students.

The following video-clips was selected to stimulate human emotions: Sadness (The Boy in the Stripped Pajamas, The Impossible, Never Let Me Go, My Sister's Keeper), Fun (Blended, The Hangover Part III, The Proposal), Fear (Mirrors, The Conjuring II, III and IV), Tenderness (The Notebook 1, Les Choristes, Pride and Prejudice, Her, He's Just Not That Into You). Video clips were downloaded from the following official web site:
https://doi.org/10.6084/m9.figshare.5372782.v1

II. Hybrid Model

In this section, we describe step by step our hybrid model, The following stages are involved in the classification of emotions for thermal images (Figure 1).

https://doi.org/10.47978/TUS.2022.72.03.004

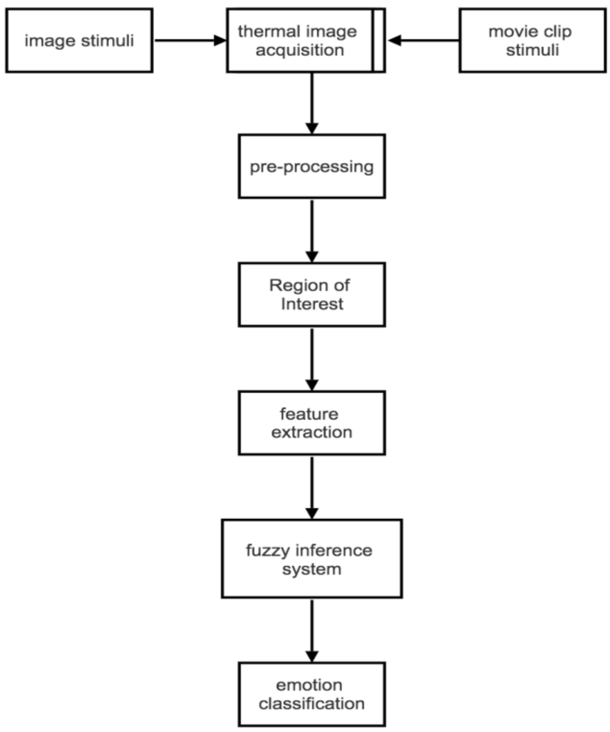


Fig. 1. Hybrid model for acquisition, detection and classification of emotions in thermal images.

A. Image and Video Stimuli

This work proposes different video clips and images stimuli to induce specific emotions in subjects. The subjects are students from TecNM University. Students were informed of the experiment objectives. A form was signed to indicate the consensus about the visual projection of sensitive and provocative image-sequences and movie-clips.

B. Thermal Image Acquisition

Temperature variation is a very sensitive task for thermal image acquisition. The experiment protocol consists of a room with illumination and controlled temperature. The image database has been gathered by a Fluke Thermal Camera model Ti32. Figure 2 displays a sample thermal image using this technology.

Fig. 2. Sample thermal image

C. Pre-Processing of Thermal Images

In this stage, two pre-processing techniques are done so that the image can be analyzed properly by our hybrid model. The best way to convert a grayscale image from a thermal image, each color component *Red(x,y)*, *Green(x,y)* and *Blue(x,y)* is transformed as follows [17]:

$$I(x,y) = K_{Red}(x,y) + K_{Green}(x,y) + K_{Blue}(x,y) \quad (1)$$

Where the following coefficients: k_{Red}, k_{Green} and k_{Blue} are the weighted sum of the RGB components, thus thermal image $I(x,y)$ is transformed into a greyscale image. This article proposes the chrominance values K_{Red}=0.2989, K_{Green}=0.5870 and K_{Blue}=0.1140. Figure 3 shows the results of this grayscale transformation.

Fig. 3. Grayscale thermal image

Cropping the face is done by the Viola-Jones algorithm [18][19], this method was originally conceived for face detection in images. One of its characteristics is its high capacity to detect facial features (face, nose, eyes, mouth), however it is not capable of recognizing such features. This is one of the main reasons to use this algorithm for classification activities.

The first stage of Viola-Jones method applies small detection windows (24x24, 20x15) into an image and determine if a feature is found, otherwise window size is increased by a factor of 1.25 and the process is repeated until the detection window is the same of the original image. The algorithm can find 160000 features approximately in an image with a window size of 24x24.

The second stage creates a new image representation by using an Integral Image. This integral is the sum of all pixels above and to the left in a single pixel (x,y) on the grayscale image:

$$II(x,y) = \sum_{x=1}^{n} \sum_{y=1}^{m} I(x,y) \quad (2)$$

The third stage uses the AdaBoost algorithm to extract the most relevant features and thus obtain a strong classifier from different weak classifiers with the lowest error on the training set.

There are many similar methods for facial recognition, but the Viola-Jones algorithm is reported in literature as highly efficient for face detection. Finally, the remaining parts of the image are removed to facilitate the following Region of Interest (ROI) stage.

D. Region of Interest (ROI)

Viola-Jones algorithm is proposed in our model to detect mainly the face and mouth of subjects in the training dataset to obtain the different region of interest. The procedure of training Haar cascade used by this algorithm sometimes fails to obtain the ROI's for visible images, this limitation was overcome by selecting the bigger bounding box automatically from multiple regions of interest (See Fig. 4).

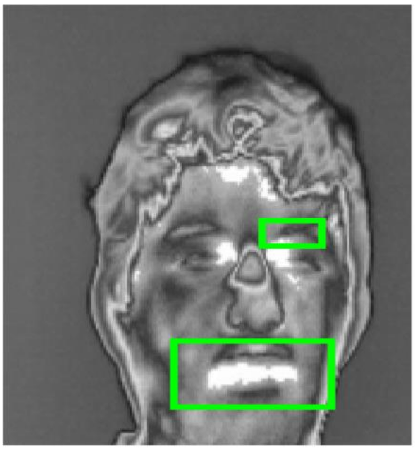


Fig. 4. Multiple mouth detection

E. Feature Extraction

In this stage, the most descriptive regions of interest are extracted from the facial thermal images: mouth and eyes. This article proposed to analyze the mouth region where emotion outlets arise such as happiness and sadness (See Fig. 5).

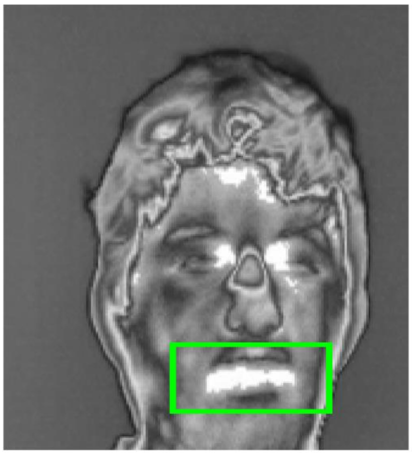


Fig. 5. Mouth detection

Several steps are involved for this stage. First, the ROI was binarized by using a threshold according to Otsu Method. This method of segmentation is proposed to select a threshold automatically that allows a better separation between object and background from an image [21, 22]. This method requires calculating the image histogram as shown in the following expression:

$$hist(k) = \sum_{x=1}^{n} \sum_{y=1}^{m} I(x,y) = k \tag{3}$$

The probability of each gray level is shown as follows:

$$p(k) = \frac{hist(k)}{nxm} \tag{4}$$

Where *nxm* is the number of pixels in x-axis and y-axis respectively of entire image. Thus, the image to be binarized after mouth detection contains two objects: the mouth and the background. Within-class variance is calculated as:

$$\sigma_w^2 = \mu_1 \times (t) \times \sigma_1^2(t) + \mu_2 \times (t) \times \sigma_2^2(t) \tag{5}$$

The weights represent the probability of being in the i[th] class, each being separated. Finally, σ_i^2 are the variances of these classes. The basic idea is to minimize the intra-class variance and to maximize the inter-class variance as indicated in the following equation:

$$\sigma_y^2 = \sigma^2 - \sigma_w^2 \tag{6}$$

The result of this process is shown in Figure 6.

Fig. 6. Binarized image of Mouth according Otsu method

Second, a Freeman chain code was applied to obtain the contour of ROI (See figure 6b). In order to obtain the mouth contour by using chain coding from the binary image *Ib(x, y)*, this procedure is based in general moments defined as follows:

$$m_{p,q} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p \, y^q \, f(x,y) dx \, dy \tag{7}$$

Where f*(x, y)* indicates mouth contour is equal to 1, in the otherwise 0 for the remaining plane of image (background). The integral curvilinear from expression can be calculated by using chain code from mouth contour:

$$m_{p,q} = \int \int_{Mouth}^{x^p} y^q \, dx \, dy =$$

$$\frac{1}{p+q+2} \int_{Contour}^{-} x^p \, y^{q+1} \, dx + x^{p+1} y^q dy \tag{8}$$

Thus, the contour can be expressed as a double summation. The chain codes indicate the direction taken by the set of segments that form the border of the object, in our case the mouth contour. Chain codes with neighborhoods 4 and 8 are generally proposed. This study proposed a simple 8-neighbor direction code (See Figure 7), starting at the upper right and moving clockwise.

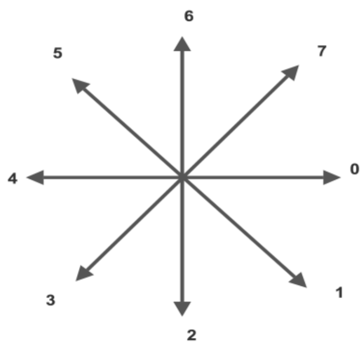


Fig. 7. 8-connected neighbour representation for chain code

In Figure 8, shows the mouth contour. Based in this chain code, height and width are calculated to determine how much the mouth is opening for smiling. Only width is used from this contour as measurement of human emotion to identify the happiness in a person. This parameter is evaluated by a FIS to categorize the happy emotion.

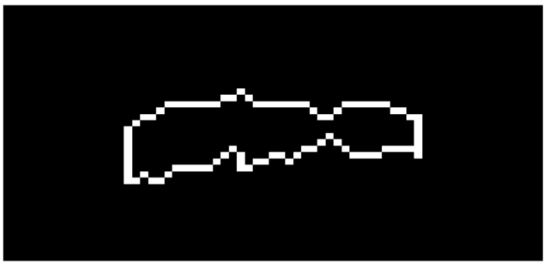


Fig. 8. Contour obtained from Chain Code

F. *Fuzzy Inference System*

A fuzzy Inference System (FIS) is defined basically of three parts: a Fuzzifier, a Fuzzy Rule Base and a Defuzzifier as shown in Fig. 9

Fig. 9. Architecture of a FIS.

G. *Fuzzifier*

Fuzzification is the process of conversion of a precise quantity to a fuzzy quantity [20]. Generally, this task performs a transformation of data into a suitable fuzzy set representation by using a membership function. This operation is achieved for input and output variables. This work proposed 3 fuzzy sets as input of a FIS as and 3 fuzzy sets as output respectively (See Figures 10 and 11).

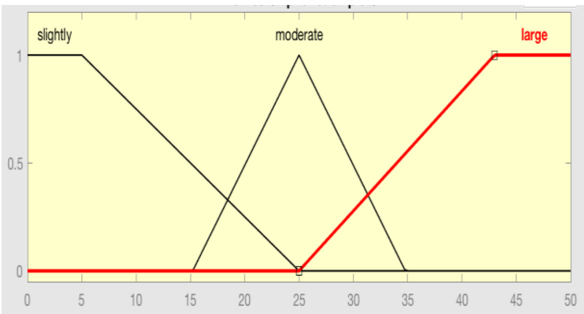


Fig. 10. Fuzzy partition for input Mouth

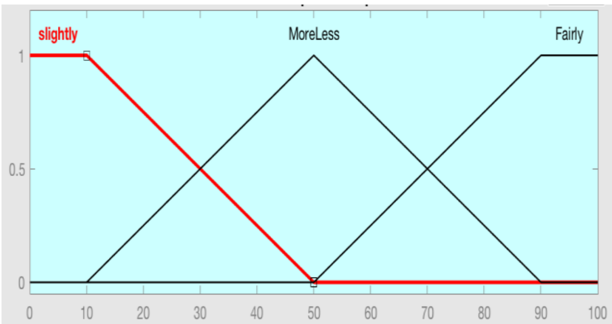


Fig. 11. Fuzzy partition for output Smile

Fuzzy Rule Base

The fuzzy rule base stores the knowledge about the problem. The fuzzy rules IF-THEN characterizes the input-output relation of the FIS. The general form of the fuzzy rules is:

Rule: If Input is A_i, ..., THEN Output is B_i, $_{i=1,2...n}$

The first part of the rule is a fuzzy proposition denoted antecedent and the second part is called consequent. The evaluation of the fuzzy rules for the proposal are defined as following:

1. IF Mouth-opening IS slight THEN Happy IS slight Happy
2. IF Mouth-opening IS moderate THEN Happy is More or Less Happy
3. IF Mouth-opening is large THEN Happy is Fairly Happy

Defuzzifier

Defuzzification is the conversion of a fuzzy quantity to a precise quantity [20]. The Centroid method is used in this study, also called center of gravity, given by the following general expression:

$$y = \frac{\int y \ \mu(x)dy}{\int \mu(x)dy} \tag{9}$$

In order to work with discrete output values, the integrals can be replaced by summations, as follows:

$$y = \frac{\sum_{k=1}^{n} y_k \ \mu(y_k)}{\sum_{k=1}^{n} \mu(y_k)} \tag{10}$$

In order to classify thermal images and detect human emotion, the selected measurement from feature extraction is analysed by a Fuzzy Inference System (FIS) to fuzzify and categorize the input value (Mouth-opening) as shown in Figure 12.

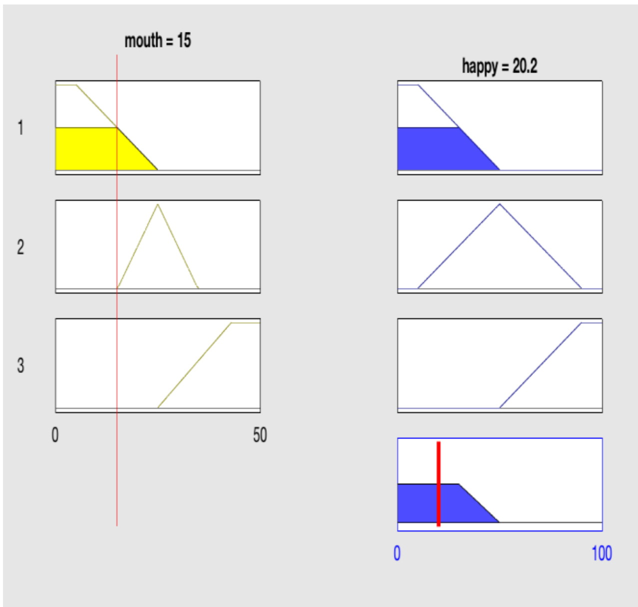
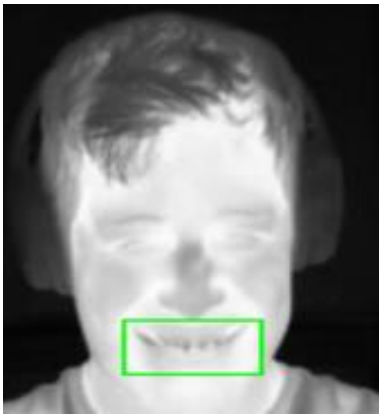


Fig. 12. Mouth emotion categorised as happy=20.2%.

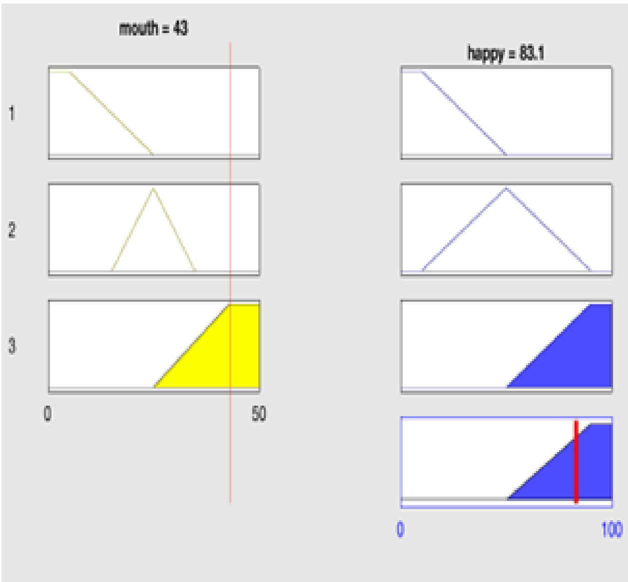
Regarding the happy smile expression in Figure 13(a), the binarized mouth is presented in Figure 13(b). In Figure 13(c) mouth contour is shown. Finally, in Figure 13 (d) FIS shows the accuracy precision of 83.1% for emotion happy.



a)

b)

c)



d)

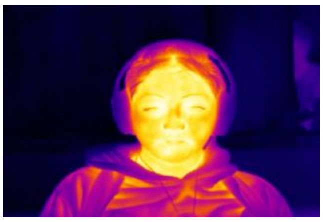
Fig. 13. a) Segmented mouth emotion, b) binary mouth emotion, c) chain code contour, and d) emotion categorised as happy=83.1%.

III. Experimental Results

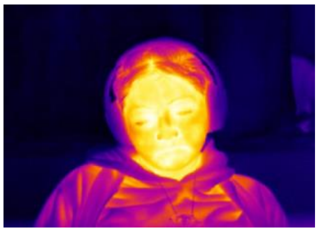
In this section, we present the results obtained with our hybrid model. The proposed method is simulated using MATLAB and Fluke thermal imaging software. Images in this study were obtained using a camera Fluke Ti32 model, Handled Infrared Thermal Imager. The resolution of each image captured is 450x350 pixels. The focus distance in the experiment was defined as 2 meters from camera to subject. MacOs High Sierra, processor 2.9 Ghz, Inte Core i7. Memory 16 GB. This experiment was conducted in a controlled environment, from April to September 2022.

For the thermal image dataset, a total of 70 students participated, ranging in age from 19 to 21, consisting of 38 males and 32 females. Each student was informed about the experiment and several images were obtained to detect basic emotions with thermal camera, under a controlled illumination. Figure 14 presents a sequence of thermal facial images for each emotional state.

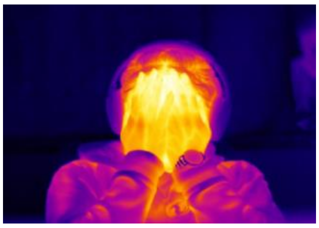
a)



b)



c)



d)

Fig. 14. Four basic emotions: a) happy, b) tenderness, c) sad and d) fear

Table 1, shows results the classification by using F1-score per class in a one vs rest through the harmonic mean of precision metrics and recall as follows [23]:

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall} = \frac{2\,TP}{2TP+FP+FN} \qquad (11)$$

Table 1. Classification of facial expressions

Emotions	F1-score
Happy	100%
tenderness	78%
sad	80%
fear	85%

It is observed that our hybrid model can be classified in 4 basic emotions (happiness, tenderness, fear and sadness). If the opening mouth distance height measurement is analysed by the fuzzy logic inference system, our hybrid approach is very robust for the happy emotion classification when a smile is detected in the mouth.

With respect to difficulties to classify facial expressions in thermal images, it is noted that sad and tenderness have a very low classification rate, due mainly that many students do not reflect correctly those emotions. In women the facial expression involves a different slightly changes in position head or body. Sometimes most women softly close their eyes or kept half-closed. In contrast in men the facial expressions are neutral, or they no shown any change in the facial muscles or body. To tackle this task, it is necessary to analyse facial micro-expressions in eyes, mouth, position of head and body simultaneously to achieve a better recognition rate.

It is very difficult to compare with other works reported in the literature, due to different environmental conditions such as experimental and equipment positions, natural and artificial illumination, humidity and temperature controlled room, thermal infrared camera, camera focal distance, selection of video clips stimulus to consider several number of emotion expressions (positive and negative) and target population during data collection, would affects directly the tasks of data acquisition, data analysis, pre-processing and classification.

Experimental results indicates that our model is very competitive in finding facial expressions of happiness and fear. These scores suggest that stimulus-based elicitation is very efficient and effective than other similar approaches reported in the literature.

IV. CONCLUSION

This paper proposes a hybrid model for the identification and classification of basic human emotions. As future work, we would combine multiple object detection techniques, firstly for thermal facial images for the visible image plane and second the infrared image plane respectively. The FIS is a complementary technique to classify a smiling face in thermal images instead of using convolutional neural networks. The computational challenge is to improve our model in order to obtain better results in the classification of basic emotional states. This hybrid model could be implemented for video analysis of customers, monitoring reactions in a school, airport, store, or in a hospital.

ACKNOWLEDGMENT

This work was supported by TecNM.

REFERENCES

[1] A. Paiva, R. Prada, and R. W. Picard. “Affective Computing and Intelligent Interaction” on Second International Conference, ACII 2007. Lisbon,vol. 4738. Springer.

[2] R. W. Picard. “Affective computing”. 2007. MIT press.

[3] J. Or. “Affective computing”. 2008. BoD–Books on Demand.

[4] L. Tian, S. Oviatt, M. Muszynski, B. Chamberlain, J. Healey, and A. Sano. “Future of Affective Computing and Applied Affective Computing”. 2022. Applied Affective Computing” 2022.

[5] L. Tian, S. Oviatt, M. Muszynski, B. C. Chamberlain, J. Healey, and A. Sano. “Emotion-aware Human–Robot Interaction and Social Robots”, 2022. Applied Affective Computing. DOI: 10.1145/3502398.3502407

[6] A. Kolak, M. Mirkovic, D. Nikolic, and I. Kalin. “Multimodal Techniques and Methods in Affective Computing–A Brief Overview Industrial Innovation” 2022. in Digital Age, pp. 518–525. DOI: 10.1007/978-3-030-97947-8_68

[7] Y. Wang et al., (2022). “A systematic review on affective computing: Emotion models, databases, and recent advances”, In Information Fusion, 2022. DOI: 10.1016/j.inffus.2022.03.009

[8] A. Mehrabian. “Communication without words”, In. Communication theory, Routledge, 2017. pp. 193–200.

[9] M. Kowalski, A. Grudzien, and W. Ciurapinski. “Detection of human faces in thermal infrared images”. 2021. Metrology and Measurement Systems, pp. 307-321.

[10] D. Aiordachioaie, A. Culea-Florescu and S. M. Pavel. “On Human Faces Thermal Image Processing for Classification Purposes”, 2019. In. 6th International Symposium on Electrical and Electronics Engineering (ISEEE), pp. 1-6. DOI: 10.1109/ISEEE48094.2019.9136103.

[11] M. F. H. Siddiqui, P. Dhakal, X. Yang, and A. Y. Javaid. “A survey on databases for multimodal emotion recognition and an introduction to the VIRI (visible and InfraRed image) database”, 2022. In. Multimodal Technologies Interaction, vol. 6, num. 6. DOI: 10.3390/mti6060047

[12] Deng, Y.; Yang, M.; Zhou, R. “A New Standardized Emotional Film Database for Asian Culture”. 2017. Front. Psychol. vol. 8. DOI: 10.3389/fpsyg.2017.01941

[13] M. Krišto and M. Ivasic-Kos, "An overview of thermal face recognition methods" 2018. In. 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1098-1103. DOI:10.23919/MIPRO.2018.8400200

[14] Y. He, B. Deng, H. Wang, L. Cheng, K. Zhou, S. Cai and F. Ciampa. “Infrared machine vision and infrared thermography with deep learning: A review”, 2021. In. Infrared Physics & Technology, Volume 116. DOI: 10.1016/j.infrared.2021.103754

[15] R. Ashrafi, M. Azarbayjani and H. Tabkhi. “Charlotte-Thermal Face: A Fully Annotated Thermal Infrared Face Dataset with Various Environmental Conditions and Distances”, 2022. In. Infrared Physics & Technology, Volume 124. DOI: 10.1016/j.infrared.2022.104209

[16] M. Yanina, et al. "LATEMO-E: a film database to elicit discrete emotions and evaluate emotional dimensions in Latin-Americans." 2019. In. Trends in Psychology Vol. 27, pp.473-490. DOI:10.9788/TP2019.2-13

[17] D. Mery and F. Pedreschi, F. “Segmentation of colour food images using a robust algorithm”. In. Journal of Food engineering, 66(3), 353-360. 2005. DOI: 10.1016/j.jfoodeng.2004.04.001

[18] P. Viola and M. Jones. “Rapid object detection using a boosted cascade of simple features”. In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001.Vol. 1, pp. I-I. 2001. DOI: 10.1109/CVPR.2001.990517

[19] Y. Q. Wang. “An analysis of the Viola-Jones face detection algorithm”. Image Processing On Line, 2014. Vol. 4, pp.128-148. DOI: 10.5201/ipol.2014.104

[20] T.J. Ross. Fuzzy logic with engineering applications. 2005. John Wiley & Sons.

[21] M.A.B. Siddique, R.B. Arif and M.M. R. Khan. "Digital Image Segmentation in Matlab: A Brief Study on OTSU’s Image Thresholding". In. International Conference on Innovation in Engineering and Technology (ICIET). 2018. DOI:10.1109/ciet.2018.8660942

[22] X. Xu, S. Xu, L. Jin and E. Song. “Characteristic analysis of Otsu threshold and its applications”. 2011. In. Pattern recognition letters, 32(7), 956-961. DOI: 10.1016/j.patrec.2011.01.021

[23] M. Grandini, E. Bagli and G. Visani. “Metrics for multi-class classification: an overview”. 2020. arXiv preprint arXiv:2008.05756. DOI: 10.48550/arXiv.2008.05756