A case study: Improving face detection in Cartagena Vessel Traffic Service Operators to detect fatigue onset using thermographic images through the YOLO5Face model.

F. Crestelo Moreno[1], J. Roca González[2], J. Suardíaz Muro[2], G. Boil[3]

[1] Marine Science and Technology Department, University of Oviedo, Gijón, Spain, fcrestelo@uniovi.es

[2] Department of Electronic Technology, Technical University of Cartagena, Cartagena, Spain, jroca.gonzalez@upct.es

[3] G. Boil, CITI University, Ulaanbaatar, Mongolia, gregory@citi.edu.mn

Abstract

Fatigue is a common physiological state that can have a significant impact on human performance and well-being, so its early detection is essential in a number of fields, such as maritime safety, which is the subject of this article. Traditional methods of fatigue assessment methods often rely on subjective self-reports or performance metrics, which can lack sensitivity and objectivity. In recent years, thermography has emerged as a promising non-invasive tool that has been validated to detect the onset of fatigue in humans. One way to detect the onset of fatigue is to monitor the temperature of the orofacial area using thermographic imaging. This article provides an overview of the principles of thermography, focusing on the analysis of orofacial thermographic images in real work situations in 16 port control operators in Cartagena. For this purpose, images of these subjects were taken during their watch and the points of interest mentioned above were identified using the YOLO5Face model, a version of YOLO5 trained to recognise faces.

As a result of using this model, it was possible to identify and monitor points of interest in thermographic images in an uncontrolled environment.

1. Introduction

Fatigue has always been a concern for the maritime industry. In 1993, the International Maritime Organisation (IMO) adopted an initial Resolution A.772(18) on fatigue factors in manning and safety, which was followed by a number of subsequent recommendations and guidelines. In this Resolution, the IMO recognises that there is no universally accepted definition of fatigue, but generally considers fatigue to be a state of exhaustion, overexertion, sleeplessness or excessive tiredness. Accurate and early detection of fatigue is therefore essential to prevent accidents, enhance safety, optimise productivity and improve quality of life.

According to the European Maritime Safety Agency (EMSA), out of a total of 823 accidents analysed between 2014 and 2020, 89.5% of all occurrences were related to human actions, either at the accident event level or at the contributing factor level [1].

Infrared thermal imaging has been used in medicine since the early 1960s. Working groups within the European Association of Thermology produced the first publications on the standardisation of thermal imaging in 1978 [2].

In the context of human fatigue research, thermography is used to detect changes in skin temperature, mainly in the orofacial region, with particular interest in the area of the nose, where temperature changes may indicate physiological stress or fatigue [3]. When a person experiences fatigue or physical or mental stress, blood flow patterns and heat distribution on the surface of the skin can change. These changes can be visualised and analysed using thermographic imaging.

The main objective of this paper is to describe the methodology for detecting the occurrence of fatigue in vessel traffic service operators (VTSOs) by using model training for thermographic images.

2. Materials

The primary tool employed in this study was the FLIR E-60 thermal imaging camera (FLIR Systems, Inc., Wilsonville, OR, USA). The FLIR E-60 is a state-of-the-art thermal imaging camera designed for high-resolution thermal imaging and temperature measurement applications.

Detector Type	Uncooled microbolometer
Spectral Range	7.5 - 13.0 µm
Temperature Range	-20°C to 650°C (-4°F to 1202°F)
Thermal Sensitivity	<0.05°C at 30°C (86°F)
Resolution	320 x 240 pixels
Image Refresh Rate	60 Hz
Accuracy	±2% or ±2°C (±3.6°F), whichever is greater

Table 1.- *Key features of the FLIR E60 thermal imaging camera*

The FLIR E-60 camera was chosen for its exceptional thermal sensitivity and high-resolution capabilities, which were crucial for capturing and analysing temperature variations in the subjects under investigation.

3. Experimental Setup and sample

The experiments were carried out in the Port of Cartagena, in the Traffic Control Room, a controlled environment to ensure accurate thermal data collection, following strict protocols such as those developed by the University of Glamorgan for the recording and analysis of thermographic images in humans [4], which take into account factors such as room size, ambient temperature, humidity, atmospheric pressure and radiation sources.

Sixteen active VTSOs from the Port Control of Cartagena (Spain), a centre attached to a cooperation agreement with the Polytechnic University of Cartagena, participated in this study. These participants have a wide range of age, work experience, or gender [N = 16 mean (±SD) age = 41 ± 9.27 yrs.; total experience in the workplace = 5 ± 5.85 yrs.; 69 % male].

The study design and procedures were previously reviewed and approved by the Scientific and Ethical Research Committee of the Technical University of Cartagena, and all the volunteers were free to withdraw from the study at any time.

This occupational sector was chosen because it is a complex socio-technical system [6-8], where factors such as interaction with technology, long on-call times, or the high responsibility are triggers for the onset of fatigue.

4. Method

The thermographic images were taken in real work situations using the previously mentioned thermographic protocols.

The camera is positioned at a distance of no more than 1 metre from the subject and at the appropriate angle to frame the orofacial area, but not more than 45 degrees [5](*Figure 1*).

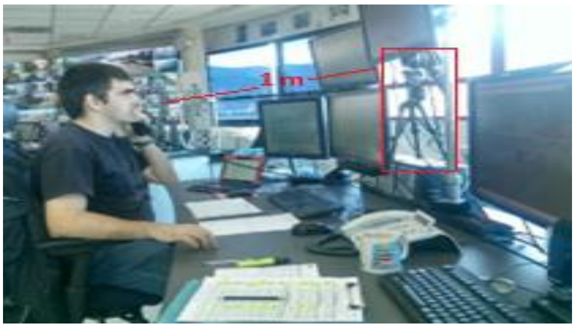

Figure 1.- *Placing the thermal imaging camera for an orofacial study*

The method we used to pinpoint the forehead / the nose is the following: we first localise the centre of the eyes and use it as the diagonal of a square whose two others vertices are the forehead and the nose. Thus, the challenge is to localise on each frame the eyes of the subject, with recording where the volunteers are constantly moving to perform their tasks, where some use glasses which are opaque in a thermal image, and with constant passage of other people, as it is a shared centre with the Port Authority. To automate this task, we choose the AI tool: a version of the YOLO5 model (fifth version of 'You Look Only Once' model) was used, more specifically the YOLO5Face model, a version of YOLO5 designed for face recognition [9].

This model can run on a variety of hardware configurations. We ran it on a local CPU (1.4GHz Intel Core i5 quad-core) and it only took a few seconds to a minute per video, depending on the length of the video. The recommended configuration by the developers of YOLO is a GPU with at least 8GB of memory.

The original YOLO5 model is trained to detect faces on RGB images. However, the face recognition from thermal images suffers from the lack of large annotated thermal datasets, especially for 'in the wild' images, meaning images in an uncontrolled environment. This difficulty has been overcome by Kuzdeuov et al. [10]. They constituted a dataset, Thermal Face in the Wild (TFW), which is a collection of thermal images in which faces are manually annotated with bounding boxes and the centres of the eyes, nose tips and corners of the mouth. It consists of 9,982 images and 16,509 labelled faces. The novelty of the dataset is the integration of a large amounts of images / faces captured in the wild (5,112 faces in a controlled environment, 1,748 faces in a semi-controlled environment, 9,649 faces in the wild). Faces in many situations were captured, such as occlusion, head pose variations, variations of scales, locations, weather condition, and wearing of a face mask. These situations are known to be challenging. Then the YOLO5Face model was re-trained using this dataset, resulting in a highly accurate model, with average precision scores of 97% on the test dataset of outdoor images. We used their model to our orofacial region identifying purpose (*Figure 2*).

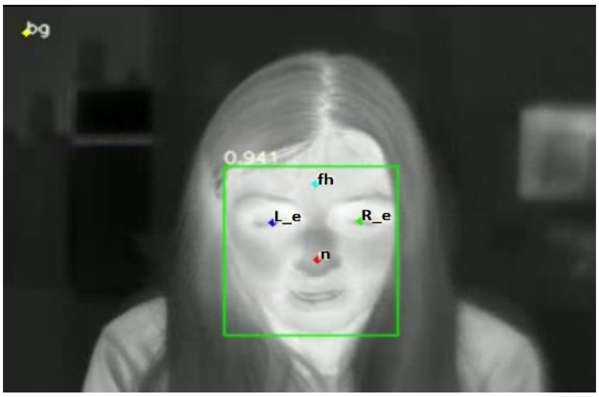

Figure 2.- *Capture of the orofacial region identifying the points of interest in one of the VTSOs (bg= background; n= nose; L_e = left eye; R_e= Right Eye; fh= forehead)*

To do so, we needed the FLIR thermal video specific format, the SEQ format, to be supported by the frames extracting function and then to be fed into the re-trained YOLO5Face model. It appeared that such a function wasn't provided by the FLIR company, nor available on the internet. To address this need, we rewrote the extracting function to support the SEQ format by integrating the fnv library provided by FLIR, library that process the SEQ videos. One of the challenges of this process is that thermal frames are given in two formats: the 8-grayscale format, that is then converted in RGB images and given to the YOLO5Face model for face recognition and localisation of Point Of Interests (coordinates in the frame), and the 16-grayscale format that encapsulates the temperature at each pixel of the frame, specifically at the POIs. By combining the use of these two formats, the resulting function can retrieve orofacial coordinates and temperatures from the frames.

A feature added to the extracting function is the ratio argument. It aims to handle heavy videos, allowing only a percentage of frames to be extracted, for instance a ratio of 0.1 meaning that one frame is extracted every 10 frames. Indeed, the FLIR camera used is designed to capture videos with a fps close to 7.5, which means one frame captured each 0.13 s, and some videos lasts for 5 or more hours, that results in a substantial number of frames to process. The temperature does not need such an accurate monitoring. Then the addition of the ratio feature was decided, allowing the temperature to be monitored once per second (ratio close to 0.1) or even once per minute (ratio close to 0.002).

As a result of this improves, the data processing sequence is as follows: first, the extraction function detects the SEQ videos and extracts and processes the frames from the videos in 8 and 16 grayscales format. Then the extracted 8-grayscale frames are analysed by the trained YOLO5Face model. Once we have the coordinates of the landmarks, we read the temperatures on the 16-grayscale frames and record it.

However, during the acquisition of the points of interest, it appeared that two problematic situations occur and must be filtered out.

In the first situation, one or more faces are detected by the model, in which case only the temperature of the first detected face is recorded. We don't have control on which face is detected at first by the model, because it depends on the position of the faces on the frame. Regarding such situations, we make the assumption that any other faces shown is sporadic. Thus, the recorded temperatures might 'jump' when a change of 'face' occurs. This can be filtered out by averaging temperatures over long enough periods (moving average technique), as discussed in the results section.

In the second situation, no face is detected. First, neither the forehead nor the nose coordinates are recorded (null value in the database). For the recorded temperature, several strategies are implemented: record a temperature of 0, record a temperature equal to the reference temperature in the background, record the last recorded temperature of the nose/forehead. The user can then decide which strategy fits better to his analysis. In this study we chose the first option, recording a temperature of 0 when no faces appear.

Finally, once all the videos have been processed, the data is stored in a csv file, one file for each video. The csv file contains a database of the coordinates of the left eye, right eye, forehead, nose and background and the temperatures measured at these points.

Results

To explain the applied method for proceeding the raw collected data, we showcase one of the results obtained. The watch displayed lasts more than 6 hours; it has been recorded at a ratio close to one frame each second. As shown in the figure (Figure 3), the raw data isn't readable, for the reasons raised above.

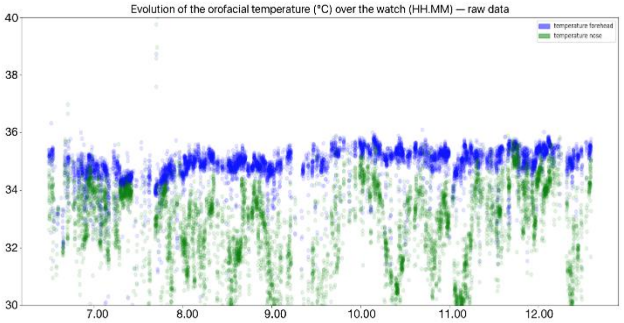

Figure 3.- *Nose and forehead temperature raw curves*

To make the data more understandable and instructive we will follow a three steps procedure:

First, we won't show the points where the temperature is 0. It means we cut out the points where no faces appear. And for averaging purposes we replace this temperature with the last temperature recorded for each passage where the face is out of the frame.

Second, we try to assess when another face is detected in the frame. To do so we monitor the motion of the left eye of the detected face. The assumption is that a high variation in the left eye position means another face have been detected. The limitation of this assumption is when the other face detected is close enough to the face of the subject so that it could be misinterpreted as a natural movement of the head of the subject. To find these high variations we do a statistical analysis of the movement of the eye: we calculate the variations of position between each two consecutive times 'dX' and then calculate its mean 'm' and its standard deviation 'std'. Then, each point beyond the 'm + 4std' value is considered an outlier in the variations (the number 4, or sometimes 3, is a thumb-rule in the outlier's detection). We then paint in red all the resulting points.

Third, we use a 'moving average' method: for each time, we calculate the average temperature of the temperatures recorded in the window of 5 min around this point (starting from 2.5 min before and until 2.5 min after this point). Under the assumption that each apparition of another face is sporadic (few seconds to one minute at most), the period of 5 minutes seemed reasonable filter out the little jumps observed in the raw data while still representing the temperature of the subject. It also allows to address any fluctuations of the camera recording process.

Following these three steps, we get a much smoother temperature curve for the nose and the forehead, as displayed in figure (Figure 4). We can then identify trends, for instance a decreasing of the temperature for the first half reflecting the fatigue of the subject and then high variations of the temperatures on the second half that could be interpreted as the subject resisting the fatigue. A few red points appear, meaning either that the used method to identify other faces recording isn't effective or that the showing of other faces is indeed sporadic. But looking at the regularity of the temperature curve, it doesn't seem to impact very much the study and the 'moving average technique' is enough to smooth these singularities.

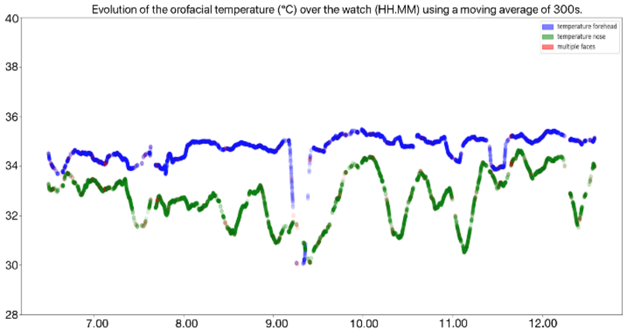

Figure 4.- *Nose and forehead temperature curves processed*

The results obtained were verified by manually determining the coordinates of the points of interest on the nose and forehead using geometry. For this purpose, anatomical reference structures were used, such as the lines connecting the lateral commissures of each eye to the perpendicular with respect to the nasal tip.

The temperature results obtained between the manual method and the algorithm did not differ by more than a tenth of a degree, which is within the tolerance in this field [11]. The test was performed by checking the temperature records every 15 minutes of recording for each volunteer and comparing them with those obtained every second by the algorithm.

As a result, 200 temperature points were manually extracted to represent the nasal temperature curve, and these results were crossed with the real-time temperature curve obtained by the algorithm, resulting in more than 8000 points, which is unattainable manually. For the showcased watch, the manually recorded temperatures fit well the above curves (Figure 5).

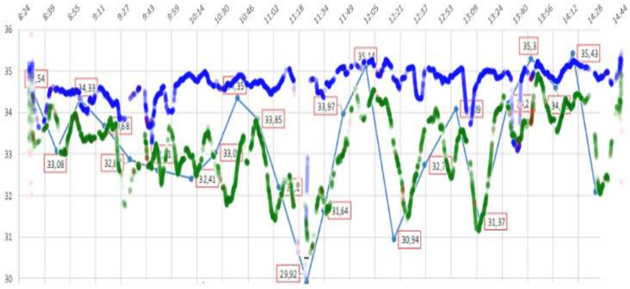

Figure 5.- *Manual and automatic temperature-comparing curves*

Conclusion

Despite the fact that the recordings were made in a real working situation, and the thermography protocols could not be strictly applied the results are quite consistent and the temperature trends are clear so that these temperature variations could be analysed to determine the occurrence of fatigue. The results showed that nasal temperature was a reliable indicator of the port coordinators' mental workload factor, which is closely related to fatigue. It decreased when decision making was required and increased during periods of relaxation. However, the results could be improved by making corrections in situations where more than one face is recognised. We cannot predict which face will be recognised first, but we can measure 'gaps' in the coordinates of the reference points. This can improve tracking and eliminate the assumption of intermittent situations with multiple faces.

Acknowledgements

We would like to thank the Spanish Search and Rescue Service and the volunteers of the Port Control of Cartagena for their collaboration in this study.

References

[1] EMSA, «Annual overview of marine casualties and incidents 2021», 2021. http://emsa.europa.eu/newsroom/latest-news/item/4266-annual-overview-of-marine-casualties-and-incidents-2020.html

[2] Aarts, N.J.M. et al. Thermographic terminology. Acta Thermographica, 1978. Suppl. 2. 1-30

[3] Marinescu, A.C., Sharples, S., Ritchie, A.C., Sánchez López, T., McDowell, M., Morvan, H.P., 2018. Physiological Parameter Response to Variation of Mental Workload. Human Factors: The Journal of the Human Factors and Ergonomics Society 60, 31–56. https://doi.org/10.1177/0018720817733101

[4] Ammer, K., 2008. The Glamorgan Protocol for recording and evaluation of thermal images of the human body. Thermology International 18, 125–129.

[5] Peterson, M.D., Joseph D. Bronzino, Donald R. (Ed.), 2013. Medical Infrared Imaging: Principles and Practices. CRC Press, Boca Raton. https://doi.org/10.1201/b12938

[6] Crestelo Moreno, F., Soto-López, V., Menéndez-Teleña, D., Roca-González, J., Suardíaz Muro, J., Roces, C., Paíno, M., Fernández, I., Díaz-Secades, L.A., 2023. Fatigue as a key human factor in complex sociotechnical systems: Vessel Traffic Services. Frontiers in Public Health 11.

[7] Nuutinen, M., Savioja, P., Sonninen, S., 2007. Challenges of developing the complex socio-technical system: Realising the present, acknowledging the past, and envisaging the future of vessel traffic services. Applied Ergonomics 38, 513–524.

[8] Relling, T., Lützhöft, M., Hildre, H.P., Ostnes, R., 2019. How vessel traffic service operators cope with complexity – only human performance absorbs human performance. Theoretical Issues in Ergonomics Science 21, 418–441. https://doi.org/10.1080/1463922X.2019.1682711

[9] YOLOv5, "Yolov5," https://github.com/ultralytics/yolov5. 1, 2, 3, 5, 6, 9

[10] TFW: Annotated Thermal Faces in the Wild Dataset Askat Kuzdeuov,, Dana Aubakirova, Darina Koishigarina, and Huseyin Atakan Varol, Senior Member, IEEE

[11] Lahiri, B.B., Bagavathiappan, S., Jayakumar, T., Philip, J., 2012. Medical applications of infrared thermography: A review. Infrared Phys Technol 55, 221–235. https://doi.org/10.1016/j.infrared.2012.03.007