QIRT'2022

10.21611/qirt.2022.1027

16th Quantitative InfraRed Thermography Conference

The use of machine learning for face regions detection in thermograms

by M. Kaczmarek*, J. Guzik*

* Gdansk Univ. of Technology, 80-233, Narutowicza Str., Gdansk, Poland, *mariusz.kaczmarek@pg.edu.pl*

Abstract

The aim of this study is to analyze the methods of detecting characteristic points of the face in thermographic images. As part of the implementation an extensive analysis of scientific publications covering similar issues both for the analysis of images made in visible light and thermographic images was carried out. On the basis of this analysis, 3 models were selected and then they were implemented and tested on the basis of test images. Finally, two models were selected, the operation of which was documented on thermographic images - DAN and AAM, as well as one model that was tested for RGB images - the ERT model. Satisfactory detection parameters were achieved in the experiment.

1. Introduction

As technology advances, thermal imaging cameras are becoming available to a wider and wider range of users as their price becomes more affordable. For this reason, more and more research is carried out on the use of thermograms for the broadly understood identification and verification of biometric features. Taking into account the characteristics of thermographic images and their difference compared to images made in visible light: their lower spatial resolution, contrast, much less information on texture and geometry, and the lack of colour diversity of elements in the image, the use of such image analysis methods as are used for RGB images cannot be used. It includes methods such as: colour analysis, image segmentation algorithms based on edges, or methods based on the analysis of symmetry and similarity of geometric shapes [1]. The use of the above-mentioned methods does not give satisfactory results in the case of face detection in infrared images. Therefore, for thermographic images of the face, it is possible to accurately determine the location of characteristic points of the face, such as eyes, nose and mouth, and to mark them, for example, by applying frames of a specific size and shape, using deep learning algorithms. A key element of machine learning and deep learning methods is the data set used in the model learning process. Such a collection should be of appropriate quality, variety and, in the case of most models, with a detailed description. At the moment, however, there are not too many databases of thermograms made on a sufficiently large group of people, which at the same time will be characterized by high resolution and will contain precisely marked labels of characteristic points of the face. The number of such databases is much smaller than in the case of databases of images made in visible light.

2. Material and method

Based on an extensive analysis of the state of knowledge, a database of thermographic images [2], created by M. Kopaczke and R. Kholk at the German University of Aachen, was selected to carry out research for the purposes of this article - hence, for the purposes of this paper, refer to it can also be used as Aachen base.

Three models were selected and then they were implemented and tested on the basis of test images – AAM, DAN and ERT. The chosen algorithms was initially trained on the training base consisting of 1500 images presenting faces of different people in many different positions and presenting various emotions. Intentionally, for the initial training of the algorithm, a group of thermograms consisting of approximately half of the total used training base was used to obtain the results of the algorithm in a relatively short time and to enable easier navigation through the preliminary results.

2.1. Active Appearance Model (AAM) and Ensamble of Regression Trees (ERT) models

For the AAM [3] model, initialize it by specifying the initial shape - in this case the face. This shape can be defined manually by adjusting the outline of the face. Then, the mean shape of the model is adjusted on the basis of the defined envelope, which results in the proposed arrangement of the characteristic points of the face. The resulting distribution of points is characterized by the best statistical fit for images showing the face in front. Since we want to analyse images showing faces in different positions, it became necessary to adjust the model to the changing position of the head - it was implemented by using the random forest regression method, which is trained with HOG (Histogram of Oriented Gradients) created on the basis of distributions of characteristic points previously detected by the AAM model. In this way, the estimated head position is transferred back to the AAM model, which on this basis is able to determine the distribution of the characteristic points of the face with greater accuracy for the face in any orientation. ERT [4] - the solution is based on the use of a cascade of regression algorithms and gradient boosting at the stage of each cascade in order to obtain higher algorithm efficiency and better classification results.

2.2. Deep Alignment Network (DAN) model [5]

It is a multi-level classifier that extracts facial features each time based on the overall image, not only image fragments in the vicinity of the point of interest. The first stage begins with the initial estimation of the face shape, which at successive levels of the classifier is better and better suited to the actual system of characteristic points of the face in the image. At each level, 3 attributes are received at the entrance: input image adjusted so that the current estimated system of characteristic points coincides with the face shape estimated at the beginning of the algorithm; map of the intensity of the selected characteristic points of the face (landmark heatmap); image from the previous stage, along with marked features.

3. Results and conclusions

For the tests, images from the Aachen database were used - the same base from which thermograms were used to train the implemented AAM, DAN and ERT models, and thermograms from own recordings were used during the experiments. Thermographic images made with the FLIR A320G thermographic camera, with a resolution of 320x240 available at the Gdańsk University of Technology for research purposes. Facial thermograms from 23 people were obtained. The images were taken against a white, thermally neutral wall, and the camera was distant from the photographed person by about 2 meters. The subjects were positioned with their faces in front of the camera, in a neutral position that did not express any specific emotions. The thermogram included the figure of the subject, ending below the shoulder line - the position of the subject was analogous to the Aachen thermogram base used to train the model. The second dataset consisted of recordings of people walking or walking towards a thermographic camera, as is the case with monitoring of people in public buildings or airports. Some sample images are shown below in the picture – Fig.1.

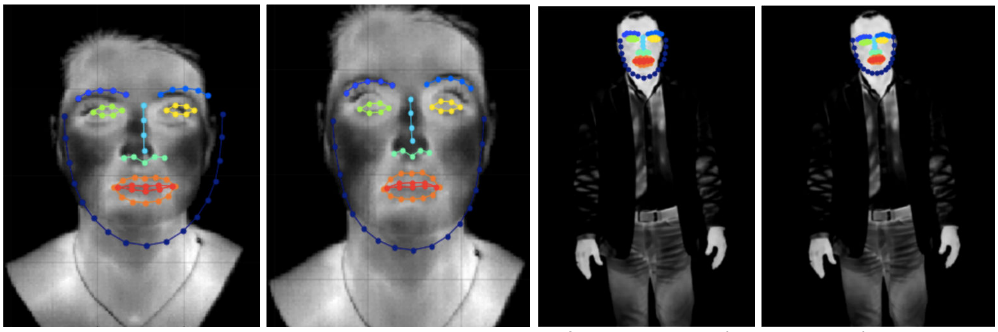

Fig. 1. *Thermographic images taken with the FLIR A320G camera with fitted mask of characteristic face points obtained in ERT and AAM model (left image – ERT fitted mask, right image – AAM fitted mask)*

The AAM model showed greater versatility of operation on various types of input images - different image dimensions and the way the test object was presented. It also required fewer extra steps to get the end result. The ERT model gave similar results when testing on the basis of images from the training image database, however, when testing images in a format different from them, complications were not infrequent, and the obtained values of errors of fit were higher. The DAN model is computationally very expensive and the results could not be obtained in real time on laptops. The use of the AAM model offers many opportunities for development and additional improvement. As part of further improvement of the accuracy of the model fit, it is possible to propose to enlarge the base of training images with thermograms from other thermographic cameras along with corresponding files with labels of the actual locations of characteristic points. Another interesting issue that can be used for the further development of the model is to propose an algorithm that allows for matching the characteristic points of the face on thermograms showing the subjects wearing glasses and masks covering the mouth and nose.

REFERENCES

[1] Arya S., Pratap N., Bhatia K., Future of Face Recognition: A Review, Procedia Computer Science, Volume 58, 2015, Pages 578-585, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2015.08.076.

[2] Kopaczka, M., Kolk, R., Schock, J., Burkhard, F., Merhof, D.: A Thermal Infrared Face Database With Facial Landmarks and Emotion Labels. IEEE Transactions on Instrumentation and Measurement, (2018).

[3] Kopaczka, M., Acar, K., Merhof, D.: Robust Facial Landmark Detection and Face Tracking in Thermal Infrared Images using Active Appearance Models., International Conference on Computer Vision Theory and Applications, pp. 150-158, 2016.

[4] V. Kazemi and J. Sullivan, One millisecond face alignment with an ensemble of regression trees, 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1867-1874, doi: 10.1109/CVPR.2014.241.

[5] Kowalski, M., J. Naruniec, T. Trzciński.: Deep Alignment Network: A Convolutional Neural Network for Robust Face Alignment. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 2034-2043, 2017.