Algorithms for Face Detection on Infrared Thermal Images

Ricardo Ribeiro

DETI/IEETA
University of Aveiro
3810-193 Aveiro, Portugal
Email: `rfribeiro@ua.pt`

António J. R. Neves

DETI/IEETA
University of Aveiro
3810-193 Aveiro, Portugal
Email: `an@ua.pt`

Abstract—Face detection on digital images is an area with immense practical potential which includes a wide range of commercial and research applications, and it continues to be one of the most active research areas on computer vision. Even after several decades of intense research, the state-of-the-art in this topic continues to improve, benefiting from advances in a wide range of different fields. In an attempt to overcome some of the limitations in face detection using visible light, the use of thermal imaging has emerged as a particularly promising research direction. This is possible because nowadays the infrared sensors have reached a new technological level. In this work we propose several methods for face detection on infrared thermal images. The well known algorithm developed by Paul Viola and Michael Jones, using Haar feature-based cascade classifiers, is used to compare the traditional algorithms developed for visible light images when applied to thermal imaging. Moreover, we present three algorithms for face detection, using image segmentation a pre-processing step to obtain a binary image. Our first approach is based on the use of an edge detection algorithm applied to the binary image. The face detection is based on the analysis of the obtained contours. As a second approach, the use of a template matching method searches and try to find the best location of a template image with the shape of human head in the binary image. In a third approach, a matching algorithm is used. This algorithm correlates a template with the distance transform of the edge image. This algorithm incorporates edge orientation information resulting in the reduction of false detection and the cost variation is limited. We performed tests taking into consideration different environments. Two of them were events at the University of Aveiro with the objective to simulate real environments, were several images were recorded while people were looking to posters. Some other laboratory tests were performed with image captured processed in real time. The results show that the proposed methods have promising outcome, but the second method is the most suitable for the performed experiments.

Keywords–Face detection; Infrared images; Image processing; Robotics; Object detection.

I. INTRODUCTION

This manuscript is an extended version of the original paper presented at the Second International Conference on Advances in Signal, Image and Video Processing, SIGNAL 2017 [1]. This extended version provides a deep overview on thermal cameras, includes more detailed information about the proposed algorithms for face detection on thermal images and more experimental results using these algorithms. As far as we know, there is limited research on this topic so we believe this work is an important contribution to the field.

In this work, we propose algorithms for face detection using thermal infrared cameras. The main goals of this work

are the use of these type of sensors in service robots and in monitoring people attention, taking into consideration of the temperature of the face over time. We are also working on emotional analysis through thermal images.

Infrared thermal cameras, also called thermographic cameras or simply thermal cameras, are used to capture radiation from the electromagnetic spectrum that the human eyes cannot see. This radiation called by infrared radiation was discovered in 1800, by William Herschel (1738-1822) [2].

Since infrared radiation was discovered, the infrared technology is in constant development. This radiation has been used in many applications in different scientific areas, for example medicine, astronomy, military, computer vision, among others. The development of the first infrared cameras began with an infrared line scanner. These line scanners were built for military requirements in the late 1940s. In the early 1950s, infrared thermal viewers took one hour to produce a single image [3].

The infrared radiation is captured by the sensor and converted to a thermal image. Figure 1 shows an thermal image complemented with an image in the visual spectrum from the same scene (typically a Red, Green and Blue - RGB - image).

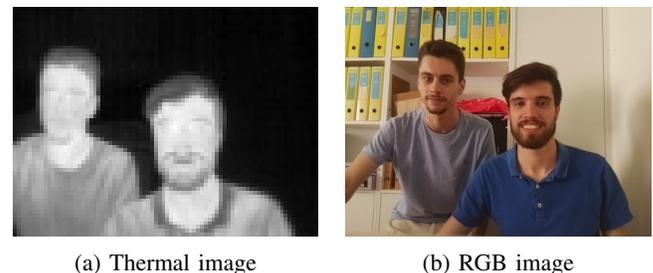

(a) Thermal image   (b) RGB image

Figure 1. An example of a thermal image and the same scene acquired by a RGB camera.

Thermal cameras are becoming a common tool for a wide range of science applications. A survey providing an overview of the current applications is presented in [4]. These applications began in the military field but were increasingly expanded into other fields due to the lower cost, resolution and quality increase of the image, being more accessible for academic applications, industrial, research, agricultural, security, personal use, among others applications. In robotics, for computer vision, thermal image analysis and processing is in constant development, being used more often in systems for

detection, recognition and tracking of objects, humans, among others.

This paper is organized as follows. An introductory section that provides a brief introduction to the paper and presents the objective. In Section II, we provide concepts and a literature review of relevant content of this work. Infrared thermal radiation is discussed, as well as the nature of infrared thermal cameras. Also in this section, we describe some related work. In Section III, we present in detail the proposed algorithms for face detection on infrared thermal images. In Section IV, we provide experimental results of the tests performed in different environments, conditions and scenes. Finally, a summary of work done, comparison of the algorithms, concluding remarks and our future work are made in Section V.

II. THERMAL IMAGES

The infrared radiation is often referred as thermal radiation, due to all objects with a temperature above absolute zero emit infrared radiation [4]. In the electromagnetic spectrum, the infrared spectral region is located from the wavelengths of 0.7 to 1000 micrometers (μm), as can be seen in Figure 2.

The infrared band lies below of the red light of the visible spectrum. This wavelength band is often subdivided into different spectral regions depending on the characteristics of this radiation in each region. Figure 2 also shows this different regions that are [5][6]:

- **Near-infrared (NIR, wavelength 0.7-1.4 μm)** - Used in fiber optic telecommunications and night vision devices.
- **Short-wavelength infrared (SWIR, wavelength 1.4-3 μm)** - Used in long distance telecommunications.
- **Mid-wavelength infrared (MWIR, wavelength 3-8 μm)** - Used in military application like guided missile technology. Sometimes called thermal infrared.
- **Long-wavelength infrared (LWIR, wavelength 8-15 μm)** - Used in thermal imaging that uses sensors to obtain images of objects. It is based on thermal emissions. This subdivision is also called the thermal infrared.
- **Far-infrared (FIR, wavelength 15-1000 μm)** - Used in military and astronomy applications.

For infrared or thermal imaging not all subdivisions of the infrared spectrum are used, only a small part called thermal infrared. Figure 3 shows an expanded view of this small region.

According to [7], commercial cameras are available for three spectral ranges that are defined for thermography. These three ranges are: the long-wave (LW) region between 7 to 14 μm, the mid-wave (MW) region between 3 to 5 μm and short-wave (SW) region between 0.9 to 1.7 μm. The reason of the restriction to these wavelengths are the physics of detectors, the transmission properties of the atmosphere and considerations of the amount of thermal radiation to be expected. The thermal motion of charged particles in matter generate the thermal radiation.

As previously stated, every object that has a temperature above of 0 K (-273.15 °C) emits thermal radiation in a different wavelength depending on the temperature and material properties. A blackbody is a perfect emitter of thermal radiation. As definition "a blackbody allows all incident radiation to pass into it (no reflected energy) and internally absorbs all the incident radiation (no energy transmitted through the body). This is true for radiation of all wavelengths and for all angles of incidence. Hence the blackbody is a perfect absorber for all incident radiation" [8].

Planck's law in general says that every physical body emits electromagnetic radiation. Due to its dependence on temperature, Planck radiation is considered thermal radiation. Figure 4 shows the spectral nature of Planck's law, where the emissive power of a blackbody at several temperatures is related with the wavelength.

The concept of emissivity is important to understand the thermal radiation emissions of a physical body. In a simple way, two objects with different emissivities at the same physical temperature will not have the same temperature in a thermal image. The emissivity also depends on the temperature of the surface as well as wavelength and angle. Kirchhoff's law states that the emissivity of a ideal blackbody is 1, this means that a blackbody absorbs all the radiation.

Figure 5 shows a thermal image with a Leslie's cube and the respective visible spectrum image in grayscale. All faces of the cube are at the same temperature. In grayscale image the face that has been painted black has more emissivity than the polished face. So the polished face of the cube emits less thermal radiation than the black face and reflects the radiation emitted by the hand. The polished faces are composed of shiny metal, such as aluminium, that are metals with low emissivity. The emissivity of the polished face of the images of the left is higher than the polished face of the right images and the white painted surface is nearly as emissive as a black surface.

Leslie's cube is used to demonstrate and measure the variations in emissivities for different materials.

The transmission of radiation in the atmosphere is not the same in all wavelengths due to the existence of several gases that absorb some radiation. Molecules of the atmosphere, as CO_2 and H_2O, are able to absorb infrared radiation. This radiation is not transmitted in some wavelength around 2.7 μm (H_2O and CO_2), around 4.2 μm (CO_2), between 5.5 and 7 μm (H_2O) and above 14 μm (H_2O, CO_2). This absorption defines the wavelength that the thermal cameras are sensitive. Figure 6 illustrates this absorption over these wavelengths and demonstrates that CO_2 and H_2O dominate attenuation of the infrared radiation.

Michael Vollmer and Klaus-Peter Mollmann have a book "Infrared Thermal Imaging: Fundamentals, Research and Applications" [7] that explains, among of others, physical principles and laws for infrared or thermal radiation.

A. Infrared Thermal Cameras

Nowadays, infrared cameras are scanning devices that capture one point or one row of an image at a time, or use a staring array. This array uses a two-dimensional focal plane array (FPA) where all image elements are captured at the same time with each detector element in the array [4]. The FPA techonology is the most used in infrared detectors. Infrared cameras also include optical lens, electronic circuits, possibly a cooler for the detector or a shutter and software for processing and displaying images to store and transmit captured data in addition to the FPA detector. A simple diagram of an infrared camera is represented in Figure 7.
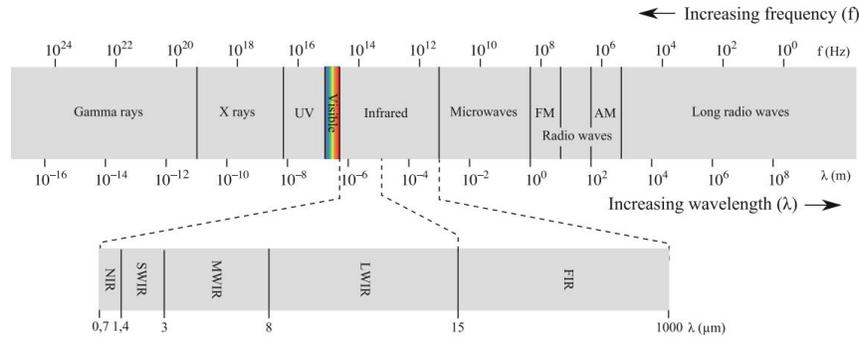
Figure 2. Electromagnetic spectrum with different divisions of infrared spectrum [4].
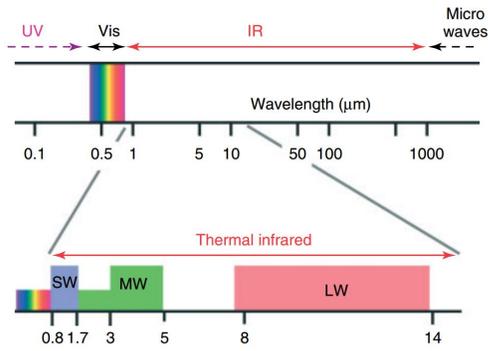

Figure 3. Infrared and adjacent spectral regions with expanded view of thermal infrared region [7].
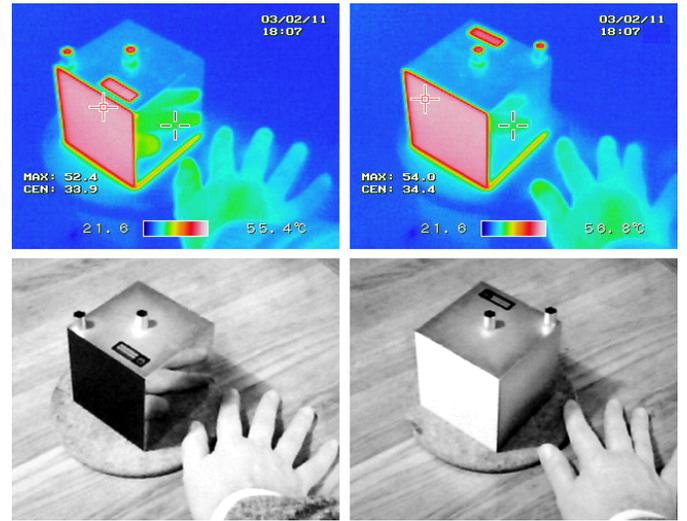

Figure 5. Leslei's cube with four faces at the same temperature represented in thermal and grayscale image [9].
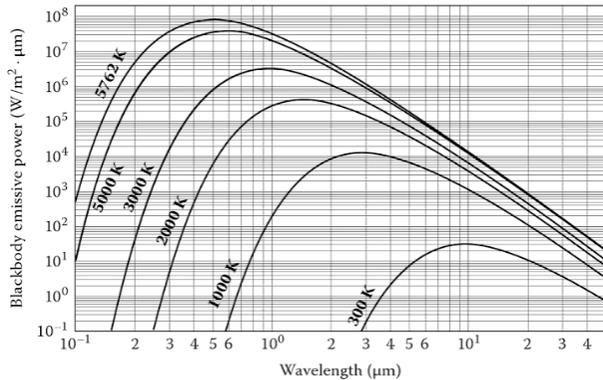

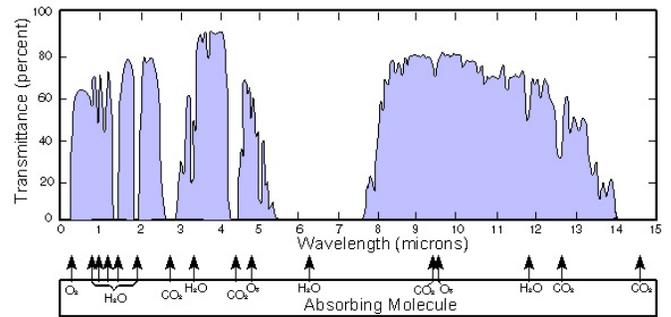Figure 4. Emissive power of a blackbody according to wavelength at several temperatures [8].

Figure 6. Atmospheric transmission in different infrared wavelengths [10].

Thermal cameras are generally divided into two types: cooled or uncooled. Cooled infrared detectors normally operate in a wavelength of 2 to 5.6 μm range that corresponds to the SWIR and MWIR band [11]. These detectors are called photon detectors or quantum detectors and convert the absorbed infrared radiation directly into a change of the electronic energy distribution in a semiconductor by changing the free charge carrier concentration [4]. Cooled thermal cameras have typical operating temperature of around 77 kelvins (K). Detectors

without cooling are not sensitive to the infrared radiation, being flooded or blinded by their own radiation [11]. A camera with these type of detectors has a high sensitivity and can deliver hundreds of high-definition (HD) resolution frames per second.

Uncooled cameras are usually composed of thermal detectors. Thermal detectors convert the absorbed infrared radiation into thermal energy causing a rise in the detector tempera-
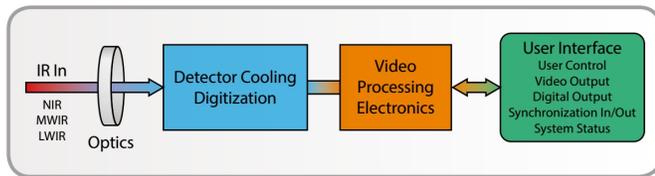
Figure 7. Simplified block diagram of an infrared camera [11].

ture. Most infrared cameras of this type have microbolometer detectors. Microbolometer FPAs are thermal sensors and can be created from metal or semiconductor material. Figure 8 shows the basic anatomy of a typical microbolometer pixel. Uncooled infrared detectors operate in a wavelength of 8 to 14 μm range, the LWIR band. These detectors have lower sensitivity and slower response time than cooled detectors but in compensation they are smaller, silent and low cost. In this work, the infrared thermal camera used is uncooled and uses microbolometers that are the most used in commercial infrared cameras [12].
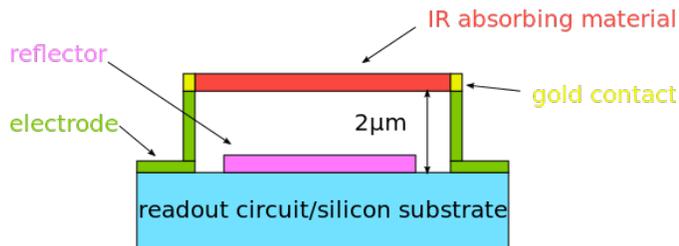

Figure 8. Cross-sectional view of a microbolometer [13].

The output data of a thermal camera is typically an image in grayscale with depth form 8 to 16 bit per pixel. Radiometric cameras usually provide output data in a raw 16-bit intensity value. These cameras are considered calibrated if they can measure temperatures through the output data. Some cameras also have a mode to visualize the output data in pseudo colors, 24 bit per pixel, due to the details of an image being more perceptible by the human eye to color images than grayscale images. Figure 9 shows an image captured by an infrared thermal camera using different colorizations.
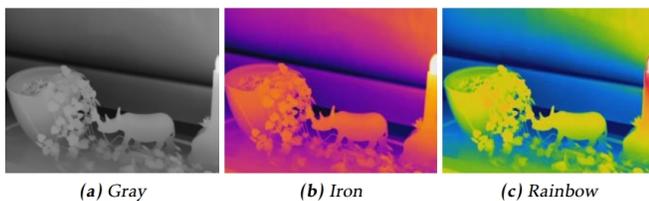

Figure 9. Thermal images in grayscale(a) and pseudo color (b)(c) [14].

B. Related Works

In color or RGB cameras, the colors and visibility of the scene depend on an energy source, such as the sun or artificial light. The captured RGB images depend on the illumination, with changing intensity, color balance, direction and among others, which sometimes affects image processing in some

applications. Furthermore, nothing can be captured in total darkness.

Thermal cameras are advantageous in many applications due to their ability to capture invisible radiation seeing in total darkness, their robustness to illumination changes and shadow effects, and less intrusion on privacy [15]. Thereby they do not depend on any external energy source. The human body emits infrared radiation in the form of heat and its temperature tends to remain constant. Generally, the part of the human body, which is not covered, is the face and it stands out from the background. These cameras, when calibrated, are advantageous in temperature measurement compared to point-based methods since temperatures over a large area can be compared, although contact methods are more efficient [14]. In this work, the thermal camera used is not calibrated, so it is not possible to know the exact temperature in the region of interest.

Some algorithms for face detection using color or grayscale images currently have a very high efficiency rate but it is not possible to use them directly in the thermal images. Using these algorithms may result in some problems regarding thermal imaging, that can reduce this efficiency significantly, such as occlusion of the face with objects that emit infrared radiation, the uniform temperature in the face, objects with the shape and same temperature of the face, the face away from the camera, among others. For this reason, only one feature-based algorithm is used, Haar Cascade [16]. The Opencv library provides utilities and functions to use the Haar Cascade algorithm and training cascade classifiers to use in thermal imaging for face detection. This algorithm has an efficient feature selection, a scale and location invariant detector [16]. The face detection in this work is represented on a shape of a bounding box obtaining not only the face, but also some background noise that can affect the detected face for future work.

Comparing with visible light cameras, thermal cameras have a considerable higher cost. For example, a thermal camera with low resolution 80×60, such as the camera used in this work, costs hundreds of euros and increasing the resolution comes with a higher prices of many thousands of euros. These cameras need more training than visual cameras for correct usage. The emissivity and reflectivity of different materials and the impact of the atmosphere affects the captured thermal image. In order to provide accurate measurements of temperature, these cameras need to be calibrated according to the physical principles, which is a complex process.

The face is one of the zones of the human body that is more suitable for the body temperature extraction and posteriorly the emotion analysis. The face detection in this work is represented on a shape of a bounding box obtaining the face and also some surrounding background noise that can affect the detected face for the next modules. The temperature measurement and emotional analysis of the face is being developed, not being included in this current work.

III. PROPOSED APPROACH

In this section, different algorithms and methods for face detection are described. The goal is to implement the algorithms on a real-time application. Therefore, these algorithms will be tested on a real-time system. The development of this work is made in C++ programming language, through the

OpenCV library, which is an open source computer vision and machine learning software library.

The resources used for the acquisition of the thermal image and the way of how the image is generated are also described in this section. Figure 10 shows a block diagram with the process from the incidence of infrared radiation on the thermal camera to the detection represented in the image to be displayed, using the proposed algorithms. The first two blocks are part of the integrated camera module. In order to receive the data, the Serial Peripheral Interface (SPI) port of Raspberry Pi is configured. The large block encompasses all the steps for data processing, obtaining the thermal image. This image is processed by the proposed algorithms and it is displayed with face detection in shape of a rectangle. In the next section, each block of the diagram is described in detail.

A. Image Acquisition

The acquisition of the thermal image is made in real-time. This image acquisition process is based on a developing project by the company Pure Engineering [17]. A FLIR LEPTON Long Wave Infrared (50 shutterless) camera module is used, with a focal plane array of 80×60 active pixels [18]. This camera is a non-radiometric version. This camera is also not calibrated, therefore it is not possible to obtain temperature values of each pixel. The output value also changes with the temperature value of the infrared sensor of the camera and the temperature of the scene. It is also used a Raspberry Pi 3 model B for communication, such as SPI communication, and image processing. This camera also supports a command and control interface (CCI) hosted on a Two-Wire Interface (TWI) similar to Inter-Integrated Circuit (I2C) for software interface [18].

The incident infrared radiation passes through the lens assembly and it is focused from the scene onto an 80×60 array of thermal detectors (microbolometers), which integrate the complete FPA. The serial stream from the FPA is received by a system on a chip (SoC) device, which provides signal processing and output formatting. The data is sent via the Lepton Video over Serial Peripheral Interface (VoSPI) protocol to the Raspberry Pi. When the camera is turned on, the SPI port of the Raspberry Pi must be opened for communication, in order to receive the data sent by the camera. It is needed to set the SPI mode, the bit per word and the SPI bus speed, so that there is no communication failure and data is received correctly through the SPI port selected.

Through the VoSPI, the output data are received in an 14-bits data. To be able to display a visible thermal image, this data is converted into an image matrix arranged in an 8-bits data with one channel provided by OpenCV. In the 14-bits data, each pixel can assume a wide range of values and this range corresponds to temperature values between -10 °C to +65 °C. As in this case it is only important the temperature of the scene captured by the camera, the data received has to be filtered.

The process of converting the 14-bits data into 8-bits data starts with the conditioning of the data, searching for the maximum and minimum values of the frame pixels received. Generally, in an 8-bits image with one channel the pixel values are between 0 and 255. In this image, the minimum value corresponds to 0 and the maximum value to 255. The columns and rows of each pixel are calculated and it is created the

visible thermal image. The first image of Figure 1 shows a thermal image obtained. Darker areas correspond to colder regions in the scene. This thermal image is the input image for the proposed algorithms.

B. Haar Cascades

Haar Cascades is a machine learning approach, using Viola and Jones algorithm [16], for visual object detection where a cascade function is trained from positive images (images of faces) and negative images (images without faces). The performance of the trained classifier will be better, as more images are used.

Haar Cascades is one of the algorithms implemented by OpenCV library. This algorithm was studied by Mekyska et al. [19] showing a machine learning approach for face detection and it requires a high number of images of the object to be detected for the cascade training. A similar study was made in this work, but the results of the face detection are considered of low accuracy in a first stage, as it can be seen in [1]. In this paper, these results were improved by training a new cascade classifier, using face annotations. The thermal images with faces that were used for the cascade training can be obtained on the online dataset [20]. The results of the cascade training for Haar Cascade are shown in the results section.

C. Implementation Details

As far as we know, face detection on thermal images did not received too much attention on computer vision as the counterpart on visible light images. There are some possible ways of using some functionalities of the OpenCV library.

Figure 11 shows an approach for thermal image pre-processing before the different face detection algorithms is applied. Thermal image is, on a first stage, segmented and filtered with morphological operators in order to obtain a binary image for the later use of the three algorithms proposed for face detection using thermal image segmentation. The algorithms Template Matching and Chamfer Matching use, in addition to the binary image, a template for recognizing this pattern in the thermal image.

Segmentation uses the Otsus method that is a thresholding binarization method [21] and filtering is performed using morphological operators, such as dilation, erosion, opening and closing [22]. In this work, the following algorithms are developed/adapted and implemented:

- **Face Contours** - Acquisition and filtering of the contours in order to obtain the longest contour in the binary image and detect the face through it [23].
- **Template Matching** - Technique for finding areas of an image that match to a template image [24].
- **Chamfer Matching** - Technique to find the best alignment between two edge maps [25].

1) Segmentation and Filtering: The thermal image obtained goes through a pre-processing in order to create a binary image. Before being segmented, the thermal image is processed using Gaussian blur [26] to reduce image noise and smooth the image. Then Otsus method is used. Otsus method is a parameterless global thresholding binarization method [21]. This method involves iterating through all the possible threshold values, assuming that the image contains two classes
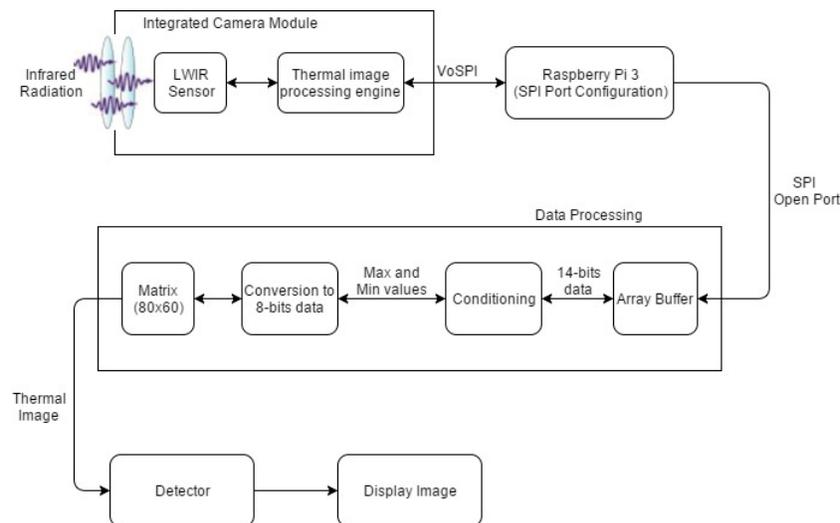
Figure 10. Processing of the infrared radiation emitted by the scene until the image is displayed.
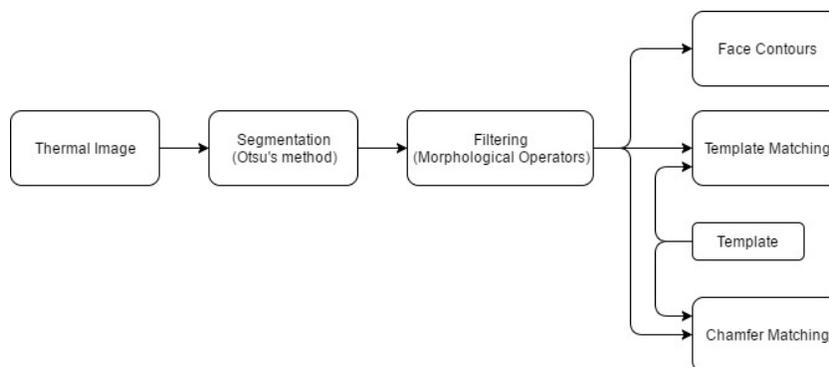

Figure 11. Thermal image processing approach to be used by the algorithms.

of pixels following bi-modal histogram (foreground pixels and background pixels) and calculating the optimum threshold separating the two classes, i.e., the pixels that either fall in foreground or background. The aim is to find the threshold value where the sum of foreground and background spreads is at its minimum [27].

The binary image generated using this method may still contain small black parts that are not relevant for face detection using the proposed algorithms. So this image is filtered with morphological operators in order to remove this noise. The word morphology concerns shapes: in mathematical morphology we process images according to shape, by treating both as sets of points. In this way, morphological operators define local transformations that change pixel values that are represented as sets. The ways in which pixel values are changed is formalized by the definition of the hit or miss transformation [26]. The two basic morphological operators are Erosion and Dilation. The variant forms are Opening and Closing. In this paper, these transformations are used in order to obtain a clean and uniform binary image.

2) Face Contours: Through the binary image an edges map is created, using the Canny edge detector algorithm [28], where the contours are found [23]. The Canny algorithm is one of the most popular edge detection technique that obtains struc-

tural information from the image and reduces the data to be processed. Among the edge detection techniques, Canny edges detection algorithm performs better under different scenarios and under noisy conditions [29].

The edges image does not contain any information about the edges as entities in themselves. The next step is to assemble the edge pixels into contours [30]. These contours are filtered in order to obtain just the contour of the larger area. Due to the contour of having some parts of the human body that are not relevant for face detection, for example neck and shoulders, the highest point of the contour is found. This point matches to the highest point on the face. Starting at this point, the two points that correspond to the largest width of the face are found. The detection of the face is made with this two points and the highest point in the face contour.

3) Template Matching: Template Matching is one of many techniques that uses a template to search and find the best match of this template, in an image. This is usually made by "sliding" a template with a similar pattern to the one or more that it is trying to find, across the image in horizontal scan and at each pixel, it calculates the distance between the template and the region of interest.

There are different matching methods to perform the template matching technique. The method Normalized Cross-

Correlation (NCC) is used, which according to [24] remains a viable choice for some if not all applications. According to [31], considering the image f with size M_x x M_y at the point (x, y), the intensity value of the image f is $f(x, y)$. The template t has a size of N_x x N_y. A common way to calculate the position (u_{pos}, v_{pos}) of the template t in the image f is to evaluate the NCC value γ at each point (u, v), which has been shifted by u steps in the x direction and by v steps in the y direction. The following equation gives a basic definition for the NCC [31]:

$$\gamma = \frac{\sum_{x,y} \left(f(x,y) - \overline{f}_{u,v} \right) \left(t(x-u, y-u) - \overline{t} \right)}{\sqrt{\sum_{x,y} \left(f(x,y) - \overline{f}_{u,v} \right)^2 \sum_{x,y} \left(t(x-u, y-u) - \overline{t} \right)^2}} \tag{1}$$

Where $\overline{f}_{u,v}$ is the mean value of $f(x, y)$ within the area of the template t shifted to (u, v). With similar notation $\overline{t}$ is the mean value of the template. Due to this normalization, $\gamma(x, y)$ is independent to changes in brightness or contrast of the image, which are related to the mean value and the standard deviation. In template matching, the position of the template in an image is determined by searching the maximum of $\gamma(x, y)$ [31].

Authors of [32] use an Image Pyramid to perform the template matching for objects with different sizes. In this work image pyramid is used to detect faces of different scales increasing the performance of the template matching. Due to the image captured by the camera being small in resolution, the pyramid image is created from a template with approximately the size of the image. Each level of the pyramid is generated downsizing the original template. Figure 12 shows the original template, that will be used, in the first image followed by images of some levels of the pyramid created. A good template is needed to obtain better results in the template matching.

Figure 12. Example of the template in different scales.

4) Chamfer Matching: The Chamfer Matching algorithm lies in shape matching using the edges maps and the respective Distance Transform [33] of a query image and a template image. The edges maps are created using the Canny edge detector algorithm [28] for simplicity and accuracy.

According to [25], the edge map of the query image is considered as $U = u_i$ and the edge map of the template as $V = v_j$, where u_i and v_j are the points of the edges of the image and the template, respectively. The distance transform is applied, creating an image where each pixel value denotes the distance to the nearest point of U. The Distance Transform algorithm computes an approximation of the Euclidean distance.

To increase the efficiency, the matching cost can be computed via a distance transform image $DT_V(x) = min_{v_j \in V} \mid x - v_j \mid$, which specifies the distance from each pixel to the

nearest edge pixel in V. The chamfer distance $d_{CM}(U, V)$ is determined by the average of distances between each point $u_i \in U$, and its nearest edge in V, as in the following equation

$$d_{CM}(U, V) = \frac{1}{n} \sum_{u_i \in U} DT_V(u_i) \tag{2}$$

where $n = \mid U \mid$.

In [25], to improve robustness, several variants of chamfer matching have been introduced by incorporating edge orientation information into the matching cost. The better match between image and template is the location of the minimal chamfer distance. In this case, the used template is considered to be matched at locations where the chamfer distance is below a defined threshold. To improve the detection for faces with different sizes, not only one template is used. The original template is resized in different scales, as shown in Figure 12.

IV. EXPERIMENTAL RESULTS

The infrared thermal camera used, presented in Figure 13, is a long-wavelength infrared (LWIR) camera module. This camera is connected to a Raspberry Pi 3 single board to be used as processing device.
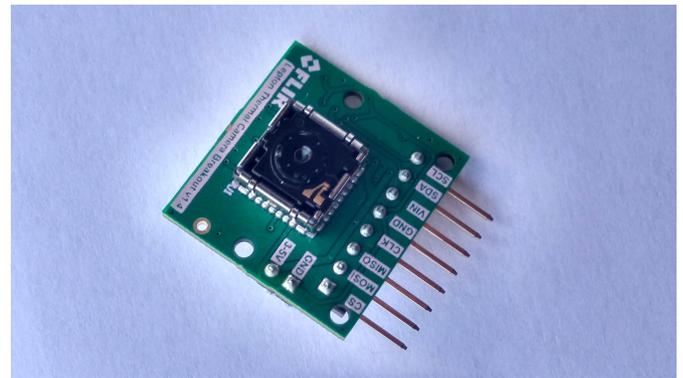

Figure 13. FLIR Lepton thermal camera module with breakout board [1].

In this section, the experimental results to verify the effectiveness of the implemented algorithms are present. In order to compare and analyze results, the variables precision and recall are calculated for an understanding and measure of relevance [34]. Figure 14 represents how to the calculation of precision and recall are made.

Using the Haar Cascades algorithm, tests were performed using captured images at the University of Aveiro event, "Teaching Day". The other algorithms were tested using images from another event organized by the Department of Electronics, Telecommunications and Informatics (DETI) of the University of Aveiro, called "Students & Teachers @deti 2017". Since the goal is the real time use, the proposed algorithms were tested under controlled conditions over time. The thermal images obtained in 8-bit grayscale in the image acquisition process were used as input and subsequently processed in each proposed algorithm. Examples of this type of images is represented in Figure 15.
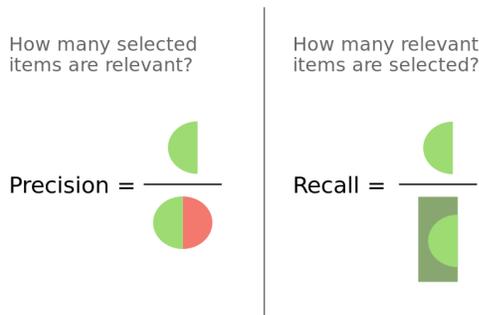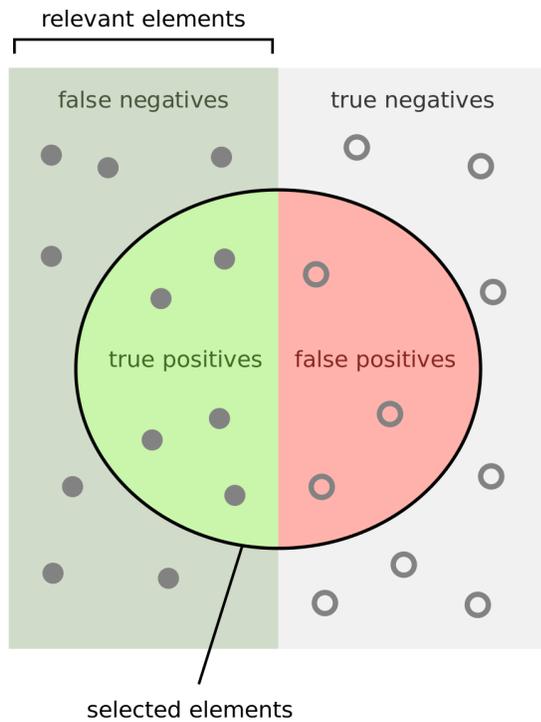
Figure 14. Precision and Recall [35].

Figure 15. Examples of images captured in Teaching Day.

A. Haar Cascades

In this work, the well-known traditional Viola-Jones algorithm for images of the visible spectrum was tested in thermal images using the same cascade classifiers for frontal face detection used for color images in order to show that this algorithm needs a well-trained classifier for the desired situation. Although it is intended to detect faces in both images, the image type is a very important factor to be considered. In visible images, the image is created based on the intensity of the primary colors, such as Red, Blue and Green, and the set of these colors generate a RGB image. In the case of the thermal image, the image is generated based on the emitted infrared radiation in the form of heat.

In [1] the Haar Cascade algorithm was used for a simple test. Figure 16 shows the preliminary example of face detection using HaarCascade. In this work, new tests were performed using the classifier use in [1] and improving this classifier in order to obtain better results.
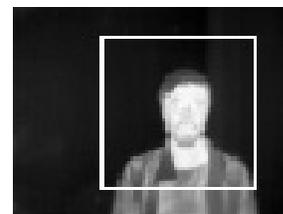

Figure 16. Face detected using Haar Cascades [1].

1) Teaching Day: At the University of Aveiro, there are initiatives that have been carried out in the last years, of critical and shared debate around teaching and learning, such as Teaching Day. In the course of this work, it was carried out the 5th edition of the Teaching Day with the theme "Technology at the service of learning: opportunities and constraints".

In the Teaching Day, students and teachers share practices and experiences of using technologies in the classroom and beyond, reflecting together the multiple ways in which technology contributes to new ways of being, teaching and learning, as well as of communicating and experiencing.

During the presentation of posters, the thermal camera used in this work was capturing frames of people who stopped to read the posters. Thousands of images were captured during this day. These images were chosen in order to reduce the number of data to be analyzed, this is, some repeated images without people for face detection were eliminated. The total of images that were used in the tests were 8130 images. Figure 15 shows some examples of images captured during the Teaching Day. In all these images there are as many faces as images. Since the camera captures low resolution images, some of these faces appear as a few pixels less than 20×20. The classifiers provided by Opencv library were trained for an input pattern size of 20×20 or 24×24 in [36]. There are more or less 1602 frontal faces that can be detected. The faces were counted manually, seeing each frame of these images. Only faces that have a dimension of approximately equal to or greater than 20×20 pixels have been counted, because the classifiers were trained for an input pattern of this size.

The first test using the Haar Cascade algorithm was made with these classifiers. The Opencv library provides three different classifiers for frontal face detection on visible light images

but only two of the three classifiers were used, which were trained for an input pattern size of 20×20. The other classifier was trained for an input pattern size of 24×24. According to [36], 20×20 is the optimal input pattern size for frontal face detection. Table I shows the number of detections, the number of false positives, the precision and recall of the detection of the classifier (a) and (b). Examples of the face detection using theses classifiers are shown in Figure 17. The last image of each row is considered as false positive.

Figure 17. Examples of detection using classifiers (a) and (b) trained for visible light images.

Using cascade classifiers for visible light images in thermal images, the recall of the detection is very low. This means that the algorithm using these classifiers returns almost none of the faces present in the images. Therefore it is not possible to use this algorithm directly in thermal image. In order to improve the results, it was made a Haar Cascade training for face detection in thermal images.

With the purpose of demonstrating that using a machine learning for training a new classifier, it is possible to improve the face detection using thermal images. The OpenCV library provides an easy way for creating training samples from positive and negative images. It also provides the tool for training classifiers through these samples. Before creating the samples, the positive and negative thermal images were collected .

The dataset used, in this attempt of training classifiers for face detection on thermal images, is OTCBVS Benchmark Dataset Collection [37]. This dataset provides among of thermal images that can be used as positive and negative images. For positive images, the Terravic Facial IR Database [20] is used, that has thermal images of faces. There are images of 18 different faces with variations, such as rotation of the face, images captured indoor and outdoor, people with glasses and hat. For the training of the classifiers not all the images of the database were used, being selected only 1154 positive images. Figure 18 shows examples of these positive images. 757 negative images were collected from other database of the OTCBVS Benchmark Dataset Collection [37] and thermal images on the web. An example of these negative images are shown in Figure 19.

The second test using the Haar Cascade algorithm was made with the classifier trained in the previous section. In this test, the thermal images captured during the Teaching Day presented in Figure 15 were used. Face detection is unstable since it does not detect the face if the image does not contain the neck and shoulders of the person. For this reason, the number of faces increase to 4173, being the size 20×20 for the whole face, neck and shoulders. Examples of the face detection using the trained classifier are presented in Figure 20. In last row of the figure with false positive detections are presented. The number of detections are 5049 and 2251 of

Figure 18. Examples of positive images.

Figure 19. Examples of negative images.

these detections contain a face (true positives). The other 2798 detections are false positives. Therefore, the calculated precision and recall are 44.6% and 53.9%, respectively.

The study presented on [19] used the Viola and Jones algorithm [16] applied to thermal image. The results of this study show that in order to obtain a good accuracy, a large amount of training data is needed. In this study, the positive images only contain the face and not other members of the human body. However, the results of the study made in [19] have better performance comparing with the results presented in this first stage.

Figure 20. Examples of the detection using Haar Cascade algorithm with the trained classifier for thermal images.

In other to attempt improving the performance of the detection, the positive images of the Figure 18 were edited and cropped to train a classifier with positive images containing only faces. Figure 21 shows examples of this new positive images. The parameters used in this training are the same of the previous training by changing only the positive images. 5770 samples with window size of 20×20 were also created.

For this test, there are more or less 1602 faces that can be

TABLE I. Results of the test using classifiers trained in [36] for frontal face detection.

Cascade Classifiers	Number of Faces	Number of Detections	False Positives	Precision (%)	Recall (%)
a	1602	17	3	82.4	0.8
b	1602	16	1	93.8	0.9

Figure 21. Examples of positive images containing only faces.

detected. Examples of the face detection using the improved classifier are presented in Figure 22. Results of images with false positives detection are presented in the last row of the figure. 3829 possible faces were detected, but only 1464 corresponded to true positive. Consequently, the number of false positives is 2375. The precision and recall are 38.1% and 91.4%, respectively.

Figure 22. Examples of the detection using Haar Cascade algorithm with the improved classifier for thermal images.

For the purpose of better understanding the results of all tests performed using the Haar Cascade algorithm, Table II presents the results for face detection of each classifier used. In this table, the cascade classifiers are represented by letters, being the letters "a" and "b" for the classifiers trained by [36], "c" and "d" the first and last classifiers trained in this paper. The results shows that the training made to improve the first training detects most of the faces in the images. Although it has a lower precision than the first training, the detection only contains the face and not other members of the human body.

B. Proposed Methods for Face Detection

In this section, experimental results are presented using a pre-processing of the thermal images for the three proposed

algorithms. The Face Contours algorithm is an algorithm, using the contours of the resulting binary image, made for single detection that helps to understand how the detection of edges and the manipulation of theses contours in an image works. The Template Matching and Chamfer Matching involve other concepts besides the detection of edges and contours. These algorithms use a template by searching the binary image for this pattern, calculating errors or distances to find the best match.

One of the tests using Template Matching and Chamfer Matching algorithms was made using images captured in the event "Students & Teachers @deti 2017". The first test using Face Contours algorithm was made with images captured for single face detection.

1) Students & Teachers @deti 2017: Students & teachers @deti is an open day of DETI of the University of Aveiro, where the students show their project work and the teachers show their research and development (R&D) activities. In this day, it is shown, to the whole academic community (students, teachers) and business companies, the work developed in the department during the academic year.

The work developed in this work was exhibited in this event. It was taken the initiative to capture thermal images during this event, in order to test these images with the proposed algorithms. 18072 thermal images were captured and 2935 of these images were selected. In order to reduce the data to be processed, some repeated images were removed. Examples of images captured during this event are shown in Figure 23. These images were used in the next sections to test the proposed algorithms.

Figure 23. Examples of images captured in the event students & teachers @deti.

2) Pre-Processing: Preliminary results have been presented in [1] as shown in Figure 24(b) and Figure 24(c) an example of the segmentation obtained in this type of images and the use of morphological operators, respectively.

The next algorithms use a binary image similar to that shown in Figure 24(c) as the input image. Using the thermal images captured in the event "Students & Teachers @deti

TABLE II. Results of all tests made in this work using the Haar Cascade algorithm.

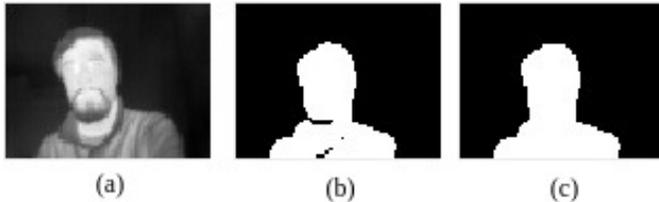| Cascade Classifiers	Windows Size	Number of Faces	Number of Detections	True Positives	False Positives	Precision (%)	Recall (%)
a	20x20	1602	17	14	3	82.4	0.8
b	20x20	1602	16	15	1	93.8	0.9
c	20x20	4173	5049	2251	2798	44,6	53.9
d	20x20	1602	3839	1464	2375	38.1	91.4

Figure 24. Thermal image(a) followed by segmentation(b) and the morphological operation(c) [1].

2017", examples of pre-processing results of the thermal images of Figure 23 are shown in Figure 25.

Figure 25. Examples of binary images using images captured in the event.

3) Face Contours: The first results obtained using this algorithm are presented in [1]. Figure 26 shows the preliminary example of face detection using this algorithm.

This algorithm was developed for single face detection, therefore the images of Figure 23 are not used for this test. Figure 27 shows the sequence of images processed by this algorithm, from image capture to final face detection. These results can be influenced if the contours of the face are discontinued for some reason. This algorithm is not developed for multiple face detection as can be seen in the last row of the Figure 27.

4) Template Matching: Figure 28 shows some examples of the tests performed previously in [1]. Using the images of Figure 23, a test was made for this algorithm. When using template matching algorithm a good template is needed. Figure 12 shows the used template in different scales. This algorithm slides the template over the binary images of Figure 25. Figure 29 shows some examples of the face detection using this algorithm. The last row of this figure represents examples with false positive detections.

5) Chamfer Matching: This algorithm uses the same template of the Template Matching algorithm shown in Figure 12,

but it is converted to an edge using the Canny edge detector algorithm to be used later in the construction of the matching cost image. Examples of distance transform image are shown in Figure 30. Preliminary results using Chamfer Matching have been presented in [1]. Figure 31 shows examples of face detection of previously performed tests. Some results applying the Chamfer Matching are shown in Figure 32. Examples of false positive detections are shown in the last row of the figure.

Table III shows the results of the algorithms, Template Matching and Chamfer Matching, that used images of the Figure 23. These two algorithms are capable to detect multiple faces in a thermal image. Comparing the both results, Template Matching is the algorithm that has more precision and recall. This means that Template Matching can detect more faces than fake faces on all detected faces and detects most of the faces in the images.

C. Real-Time Results

In order to test and compare the performance of the different algorithms proposed over time, a test was made within a controlled environment. This environment was a classroom where students work on their computers. The thermal camera was placed in front of a students with a constant background and without movement. The student's face was at a distance of about 50 to 100 centimeters from the camera. During the test, the student acts normally working on his computer and talking to other students in the classroom, varying the pose and the rotation of the face, not constantly looking towards the camera. The test duration for each algorithm was approximately 20 minutes in a total of 80 minutes. The captured thermal images was limited to a frame rate of 2 frames per second.

Figure 33 shows examples of the face detection of these tests using the four algorithms implemented in this work, Haar Cascade, Face Contours, Template Matching and Chamfer Matching, respectively from left to right of this figure. The two last rows of the Figure 33 shows examples of false positives and bad detections.

The results of the test over the time can be seen in Table IV. Analyzing this table, the precision of the algorithms is very similar but the Haar Cascade is the one that has less precision and recall. Face Contours algorithm has the best results for this test. This algorithm for application of single face detection has a good performance, having some limitations, such as the contours discontinuation of the face or a background with objects at the same temperature as the face.

Comparing Template Matching and Chamfer Matching, the number of false positives is practically the same but the number of faces and frames in 20 minutes of the test is lower, due to the processing time of the Chamfer Matching algorithm. Table V shows the average of the processing time of the

Figure 26. Some examples of face detection in different conditions using contour detection [1].

TABLE III. Results of the test using thermal images of the event with multiples detection.

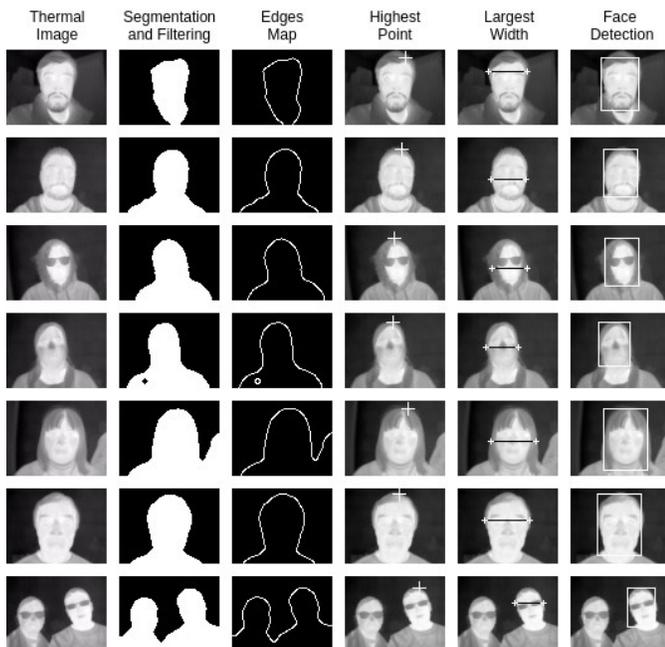| Proposed Algorithms	Number of Faces	Number of Detections	True Positives	False Positives	Precision (%)	Recall (%)
Template Matching	3229	3535	2770	765	78.4	85.8
Chamfer Matching	3229	3361	1810	1551	53.9	56.1

Figure 27. Examples of the results using Face Contours algorithm.

test for each algorithm. Chamfer Matching has the highest processing time, being almost 3 times slower comparing with others algorithms. This algorithm involves several steps, which require more processing time, for example distance transform and edge orientation.

V. CONCLUSIONS

In this paper we present a study regarding the development and adaptation of several algorithms for face detection in thermal images. Face detection on thermal imaging is a possible approach to overcome the challenges on face detection in visible light images, even after decades of active research.

Despite the thermal camera used is not calibrated, being impossible to know the exact temperature in the region of interest, the detection results were satisfactory and opens new directions on this field. As future work, we intend to improve the implemented algorithms for face detection and develop algorithms for calibration of this type of sensors, in order to measure absolute temperatures.

In terms of experimental results, Haar Cascade using Viola and Jones algorithm has better performance and accuracy based on [19]. However, this algorithm needs a large amount of data and time for training to obtain good results. Face detection in thermal image using Haar Cascade can be improved using one of the proposed methods for image segmentation and create several thermal images that contain only the face for training database in order to obtain better detection. The use of classifiers for visible light images in thermal images is not at all possible, due to infrared thermal cameras using sensors to create an image from the emitted radiation of the scene in the form of heat and the RGB cameras use sensors for captures intensity of light. The last trained classifier has better results than the other classifiers used in this work.

Face detection through thermal imaging using segmentation has a good accuracy in single face detection. For multiple detections, the results show that the better used algorithm is Template Matching. Comparing the algorithms used in this paper, Template Matching is the most suitable. Although Face contours and Haar Cascade have a processing time similar to Template Matching, the Haar Cascade has a lower recall, detecting more false positives and Face Contours has the disadvantage of not detecting multiple faces. Chamfer Matching has a similar detection to Template Matching for single detection but the processing time of this algorithm is almost 3 times slower and the detection performance, when applied to multiple face detection, decreases significantly.

Finally, we are also working on image registration algorithms, for the simultaneous use of these cameras and RGB cameras, in order to obtain a multimodal detection system.

REFERENCES

[1] R. F. Ribeiro, J. M. Fernandes, and A. J. Neves, "Face detection on infrared thermal image," in Proc. SIGNAL, 2017, pp. 38–42.

Figure 28. Some examples of face detection in different conditions using Template Matching [1].

TABLE IV. Results of the test in a classroom for proposed algorithms.

Proposed Algorithms	Number of Faces	Number of Detections	True Positives	False Positives	Precision (%)	Recall (%)
Haar Cascade	2035	1530	1465	65	95.8	72.0
Face Contours	1981	1965	1960	5	99.7	99.2
Template Matching	2026	1819	1817	2	99.9	89.7
Chamfer Matching	1614	1323	1320	3	99.8	81.8

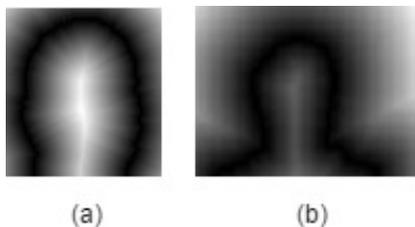Figure 29. Face detection based on Template Matching.

(a)   (b)

Figure 30. Example of Distance Transform image of template(a) and query image(b) [1].

TABLE V. Processing time of each proposed algorithm.

Proposed Algorithms	Number of Frames	Average Processing Time (ms)
Haar Cascade	2035	87.92
Face Contours	1981	90.11
Template Matching	2027	90.41
Chamfer Matching	1616	242.93

[2] M. Rowan-Robinson, Night vision: Exploring the infrared universe. Cambridge University Press, 2013.

[3] H. Kaplan, Practical applications of infrared thermal sensing and imaging equipment. SPIE press, 2007, vol. 75.

[4] R. Gade and T. B. Moeslund, "Thermal cameras and applications: A survey," Machine vision and applications, vol. 25, no. 1, 2014, pp. 245–262.

[5] J. Byrnes, Unexploded ordnance detection and mitigation. Springer Science & Business Media, 2008.

[6] B. MikaélA, Infrared Radiation: A Handbook for Applications. Springer Science & Business Media, 2013.

[7] M. Vollmer and K.-P. Möllmann, "Fundamentals of infrared thermal imaging," Infrared Thermal Imaging: Fundamentals, Research and Applications, 2010, pp. 1–72.

[8] J. R. Howell, M. P. Menguc, and R. Siegel, Thermal radiation heat transfer. CRC press, 2010.

[9] P. Kuiper, "Lesliescube — Wikipedia, the free encyclopedia," 2011, [accessed November 22, 2017]. [Online]. Available: https://en.wikipedia.org/wiki/Emissivity#/media/File:LesliesCube.png

[10] Wikipedia, "Atmospheric windows in the infrared — Wikipedia, the free encyclopedia."

[11] FLIR Systems, AB, "The ultimate infrared handbook for r&d professionals." [accessed November 22, 2017]. [Online]. Available: http://www.flirmedia.com/

[12] A. Schaufelbuhl, N. Schneeberger, U. Munch, M. Waelti, O. Paul, O. Brand, H. Baltes, C. Menolfi, Q. Huang, E. Doering et al., "Uncooled low-cost thermal imager based on micromachined cmos integrated sensor array," Journal of Microelectromechanical systems, vol. 10, no. 4, 2001, pp. 503–510.

[13] FeuRenard, "Very generalized view of the layers of a microbolometer — Wikipedia, the free encyclopedia," 2016, [accessed November 22, 2017]. [Online]. Available: https://commons.wikimedia.org

[14] A. Berg, "Detection and tracking in thermal infrared imagery," Ph.D. dissertation, Linköping University Electronic Press, 2016.

[15] A. Berg, J. Ahlberg, and M. Felsberg, "A thermal object tracking benchmark," in Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on. IEEE, 2015, pp. 1–6.

[16] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, vol. 1. IEEE, 2001, pp. I–I.

[17] Pure Engineering LLC, "Flir lepton breakout board," [accessed November 22, 2017]. [Online]. Available: http://www.pureengineering.com/projects/lepton

Figure 31. Some examples of face detection in different conditions using Chamfer Matching [1].

Figure 32. Examples of the face detection using Chamfer Matching algorithm.

Figure 33. Examples of the face detection of the four implemented algorithms.

[18] FLIR, "Flir lepton long wave infrared (lwir) datasheet," 2016, [accessed November 22, 2017]. [Online]. Available: http://www.flir.com

[19] J. Mekyska, V. Espinosa-Duró, and M. Faundez-Zanuy, "Face segmentation: A comparison between visible and thermal images," in Security Technology (ICCST), 2010 IEEE International Carnahan Conference on. IEEE, 2010, pp. 185–189.

[20] Roland Miezianko, Terravic Research Infrared Database, "Dataset 04: Terravic facial infrared database," [accessed November 22, 2017]. [Online]. Available: http://vcipl-okstate.org/pbvs/bench/Data/04/download.html

[21] R. F. Moghaddam and M. Cheriet, "Adotsu: An adaptive and parameterless generalization of otsu's method for document image binarization," Pattern Recognition, vol. 45, no. 6, 2012, pp. 2419–2431.

[22] H. J. Heijmans, "Connected morphological operators for binary images," Computer Vision and Image Understanding, vol. 73, no. 1, 1999, pp. 99–120.

[23] S. Suzuki et al., "Topological structural analysis of digitized binary images by border following," Computer vision, graphics, and image processing, vol. 30, no. 1, 1985, pp. 32–46.

[24] J. P. Lewis, "Fast normalized cross-correlation," in Vision interface, vol. 10, no. 1, 1995, pp. 120–123.

[25] M.-Y. Liu, O. Tuzel, A. Veeraraghavan, and R. Chellappa, "Fast directional chamfer matching," in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010, pp. 1696–1703.

[26] M. S. Nixon and A. S. Aguado, Feature extraction & image processing for computer vision. Academic Press, 2012.

[27] N. Otsu, "A threshold selection method from gray-level histograms," IEEE transactions on systems, man, and cybernetics, vol. 9, no. 1, 1979, pp. 62–66.

[28] J. Canny, "A computational approach to edge detection," IEEE Transactions on pattern analysis and machine intelligence, no. 6, 1986, pp. 679–698.

[29] R. Maini and H. Aggarwal, "Study and comparison of various image edge detection techniques," International journal of image processing (IJIP), vol. 3, no. 1, 2009, pp. 1–11.

[30] A. Kaehler and G. Bradski, Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library. " O'Reilly Media, Inc.", 2016.

[31] K. Briechle and U. D. Hanebeck, "Template matching using fast normalized cross correlation," in Proc. SPIE, vol. 4387, 2001.

[32] L. Ferreira, A. Neves, A. Pereira, E. Pedrosa, and J. Cunha, "Human detection and tracking using a kinect camera for an autonomous service robot," Advances in Aritificial Intelligence-Local Proceedings, EPIA, 2013, pp. 276–288.

[33] G. Borgefors, "Distance transformations in digital images," Computer vision, graphics, and image processing, vol. 34, no. 3, 1986, pp. 344–371.

[34] D. M. Powers, "Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation," Journal of Machine Learning Technologies, vol. 2, no. 1, 2011, pp. 37–63.

[35] Walber, "Precision and recall — Wikipedia, the free encyclopedia," 2014, [accessed November 22, 2017]. [Online]. Available: https://commons.wikimedia.org/wiki/File%3APrecisionrecall.svg

[36] R. Lienhart, A. Kuranov, and V. Pisarevsky, "Empirical analysis of detection cascades of boosted classifiers for rapid object detection," Pattern Recognition, 2003, pp. 297–304.

[37] IEEE OTCBVS WS Series Bench, "Otcbvs benchmark dataset collection," [accessed November 22, 2017]. [Online]. Available: http://vcipl-okstate.org/pbvs/bench/index.html