# A COMPARISON OF THERMAL IMAGE DESCRIPTORS FOR FACE ANALYSIS

*Ricardo Carrapiço, André Mourão, João Magalhães and Sofia Cavaco*

NOVA LINCS, Dept. Computer Science, Faculdade de Ciencias e Tecnologia
Universidade Nova de Lisboa, Portugal

{r.carrapico, a.mourao}@campus.fct.unl.pt, {jm.magalhaes, scavaco}@fct.unl.pt

## ABSTRACT

Thermal imaging is a type of imaging that uses thermographic cameras to detect radiation in the infrared range of the electromagnetic spectrum. Thermal images are particularly well suited for face detection and recognition because of the low sensitivity to illumination changes, color skins, beards and other artifacts. In this paper, we take a fresh look at the problem of face analysis in the thermal domain. We consider several thermal image descriptors and assess their performance in two popular tasks: face recognition and facial expression recognition. The results have shown that face recognition can reach accuracy levels of 91% with Localized Binary Patterns. Also, despite the difficulty of facial expression detection, our experiments have revealed that Haar based features (FCTH - Fuzzy Color and Texture Histogram) offers the best results for some facial expressions.

***Index Terms***— Thermal images, face recognition, facial expressions, image descriptors.

## 1. INTRODUCTION

The use of thermal images is still a very new concept compared to the use of visible images mainly due to the price of thermal cameras. However, in recent years the interest in infrared images grew up. Along with that, cheaper but less sensitive camaras became available, which allows for more research in the area. These new generation of consumer level thermal cameras, Figure 1, will become pervasive and will leverage application domain.

The use of IR light to capture images also allows to obtain other types of information to analyze. Since the human body dissipates heat, it is possible to have a good contrast between the environment and a person, which makes this type of images very useful to detect several human actions. The images produced with this method are particularly good for face detection and recognition because of the low sensitivity to variations in face appearance caused by illumination changes. It is also possible to track the human breath exclusively using thermal imaging [1]. The periorbital regions, the zones between the bridge of the nose and the inner corner of the eyes, are one of the hottest zones in the human face. Due to this fact, these zones can be used as a feature for tracking the face.

Our work is set in the context of face analysis in the thermal domain. More specifically, we evaluate several thermal image descriptors in the tasks of face recognition and facial expression recognition.

**Fig. 1**. Example of a consumer-level thermal camera[1].

## 2. RELATED WORK

A few researchers have explored the possibility of using exclusively thermal images for face recognition [2–6]. Opposed to face recognition algorithms in the visible spectrum, that mainly use the eyes location, face recognition with thermal images has difficulties in determining the eyes' position. Several approaches extract thermal contours and match the shapes for identification. These techniques use shape matching and the eigenface method, which shows better results with thermal images than with visible spectrum images [7].

There has been very little work on face detection and recognition from infrared images when compared to images in the visible spectrum. Nonetheless, some work in this area has been proposed and Kong et al. [2] has presented a very complete review.

This work funded by the Portuguese Foundation for Science and Technology under projects VisualSpeech (CMUP-EPB/TIC/0075/2013) and NOVA-LINCS (PEest/UID/CEC/04516/2013).

[1] http://www.flir.com/flirone/

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

Facial expression recognition using infrared images has also been explored. Trujillo et al. [3] proposed a facial expression recognition feature extraction model for these images. This approach uses face localization to detect the facial features and then computes the eigenfeatures used for feature extraction. The features are then fed into a support vector machine to identify the facial expressions.

Because movement releases heat, repeated muscle contractions release heat. The amount of heat released from these contractions is sufficient to be detected by high-sensitivity thermal cameras. This provides a great alternative to be explored in facial expression detection due to the fact that thermal images are not affected by lighting variations.

To be able to identify facial expressions, we need to first detect individual Action Units of the facial action coding system (FACS) [8]. Jarlier et al. [4] discovered that thermal images present good results when used to detect these units.

Guzman et al. [6] proposed a thermal imaging framework for face recognition that extracts unique features and finds similarity in thermal images. The framework's protocol consisted of taking pictures of the person at four different times to accomodate for vascular changes over time that could affect the signature matching. The face is detected using localized-contouring algorithms. A thermal signature is then generated for each image and added to a template. Each signature can be compared to a template, or two templates can also be matched. The authors proposed to use the framework in biometric validation.

Thermal images also proved to be a good method to detect affective states. Nhan and Chau [5] used a genetic algorithm to search the best combinations of different features, obtained from facial thermal images, blood volume pulse data and respiration data.

## 3. THERMAL IMAGE DESCRIPTORS

We used four different image feature descriptors, tipically used in visible images, to assess their performance when applied to thermal images. Gabor and Local Binary Patterns (LBP) are two well known methods in face image analysis. The Color and Edge Directivity Descriptor (CEDD) and the Fuzzy Color and Texture Histogram (FCTH) are alternative methods that have shown good results in medical imaging. Table 1 presents a summary of these four feature descriptors.

### 3.1. Gabor Bank-Filters

Gabor filters are linear filters used in image processing for edge detection, which allow for texture representation and discrimination. These filters are particularly useful in facial expression recognition due to being good in edge detection (which is key to detect facial components like the mouth or the eyes) and for filtering most of the noise in the image [9]. Two-dimensional Gabor filters are Gaussians weighted by an exponentially decaying sinusoid that can be applied at a given orientation and scale (for more details see [10]). These filters are particularly interesting because their behaviour resembles the behaviour of cells in the primary visual cortex. In order to capture the different details of the human face, several Gabor filters with different scales and orientations can be used. We used four scales and six orientations as described in [10], making a total of 24 filters. Each feature vector is composed by both variance and standard deviation for each application of the filters. These filters were applied to six regions of the face to reduce the noise in the feature vectors.

| Descriptor | Descriptor description |
|---|---|
| Gabor | Bank of Gabor Wavelets filters with multiple directions and scales. Compact descriptors with mean and variance of each filter. |
| LBP | Histogram of frequency of binary patterns. |
| CEDD | Fuzzy-Linking histogram (color) and MPEG-7 edge histogram (texture). |
| FCTH | Equivalent to CEDD, but it uses the high frequency bands of the Haar Wavelet transform for the texture description. |

**Table 1**. Comparison of the texture information descriptor of the different feature types

### 3.2. Local Binary Patterns

LBP is a very efficient texture operator which labels the pixels of an image by thresholding the neighbors of each pixel and considering the result as a binary number. This method is very simple and it builds 8-digit binary numbers in the following manner: the image is divided into cells (e.g. 16x16 pixels for each cell). Each pixel in each cell is compared to each of its eight neighbors, following clockwise or counter-clockwise order. For each of the eight neighbours, 0 or 1 is written into the 8-digit binary number, depending on the pixel being larger or smaller than the neighbour, respectively. Then for each cell, the histogram of frequency of the 8-digit binary numbers is computed and normalized. Finally, the feature vector of the image is obtained by concatenating the histograms from all cells. This resulting feature vector can then be used in classification processes using machine-learning algorithms.

### 3.3. Color and Edge Directivity Descriptor

CEDD is a feature descriptor that can be extracted from images [11]. It joins the color and texture information of the image in a single histogram. The descriptor consists of six texture areas, with each area separated into 24 subregions and each subregion describing a color.

The color information is given by two fuzzy systems that

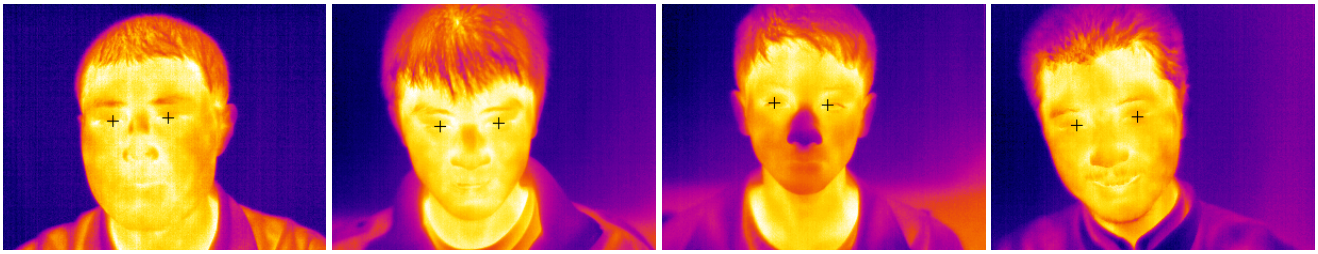

**Fig. 2**. Example images from the used dataset with eye landmarks (see section 4.1).

map the colors of the image into a 24 color palette. The texture information extraction is done using a fuzzy version of the digital filters proposed by the MPEG-7 EHD. The extraction method works by decomposing the image in 1600 images blocks (rectangular parts of the image) and submitting the block to two units, the color extraction unit and the texture extraction unit. The first unit classifies the image block into one of the 24 colors used by the system. The second unit classifies the image block in the respective texture area from 0 to 6. The results of the two units are combined and make part of the histogram. After repeating this process for every image block, the histogram is normalized and quantized for binary representation. (See [11] for more details.)

### 3.4. Fuzzy Color and Texture Histogram

FCTH is a feature descriptor that is very similar to CEDD [12]. It uses the same extraction method and the same color unit explained in section 3.3. The difference between the two types of features lies in the texture information extraction process. FCTH uses high frequency bands of the Haar wavelet Transform in a fuzzy system to form eight texture areas instead of MPEG-7 EHD digital filters and six texture areas in CEDD.

## 4. EVALUATION

### 4.1. Dataset

We used a dataset provided by the University of Science and Technology of China that offers both visible and infrared images of people performing facial expressions [13]. The visible and infrared images (in the band of 8-14 $\mu$m) were captured at the same time, i.e. with a regular camera and a thermal camera side-by-side. In addition, the dataset contains images captured under three illumination conditions and with different variations: frontal, left and right sides.

The dataset is organized into two subsets: a subset of posed and a subset of spontaneous facial expressions. The images in the *posed expression* subset, were obtained after asking the persons to do the expressions. The other subset was created by exposing the person to a situation intended to make her express an emotion and subsequently the corresponding facial expression. For this reason this last subset is composed of spontaneous facial expressions.

Both subsets consider the same six facial expressions, happiness, anger, sadness, fear, disgust and surprise. The posed subset has contributions from 107 persons, the spontaneous subset from 103 persons under front illumination, 99 under left illumination and 103 under right illumination. The dataset provides different annotations that include landmarks in some of the visible and infrared images. It is important to mention that for the infrared images with right side illumination in the spontaneous subset, there is a manual and automatic annotations of the eyes landmarks (Fig. 2).

### 4.2. Face Recognition and Retrieval Results

Face recognition is one of the oldest problems in signal processing and computer vision. In this section we evaluate the presented set of image descriptors in the task of face recognition and retrieval in the thermal domain. This experiment used the dataset and their labels for each individual. The face image registration relied on the eyes landmarks that are provided with the dataset. From these landmarks, we extracted the face region and applied the feature extractors to it.

In this setting, we first examined the structure of the feature space created by each descriptor. Our goal, is to discover the relation between the geometry of the feature spaces and the notion of a person's face visual similarity in the thermal domain. For each face image, we ranked the remaining ones and computed precision and recall at different positions of the rank (10, 20, 30, 40 and 50). Figure 3 characterizes the average across all face images in terms of: precision and recall. In terms of precision, we can see that the best descriptors are the Gabor and the LPB, being around 30% precision on the top 10 images. However, the most relevant result is the LBP recall, which reaches 92.75% with only 50 images.

These results indicate that the geometry of these feature spaces offers a solid ground to tackle other tasks, such as face recognition. Table 2 presents the face recognition results with a *k-nn* classifier. In this task, the LBP feature space is again the best one - it achieved 91% accuracy. We varied the considered number of neighbors from 1 to 9 and observed that using 3 neighbors was the best setting. This is highly relevant, because combining this result with the results of precision and recall, we can deduce that LPB is a high-precision

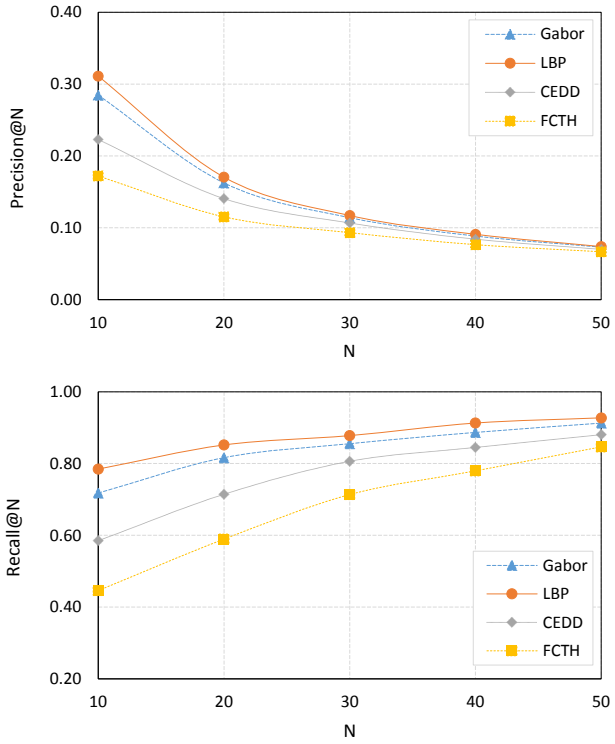


**Fig. 3**. Face retrieval precision and recall at different rank levels.

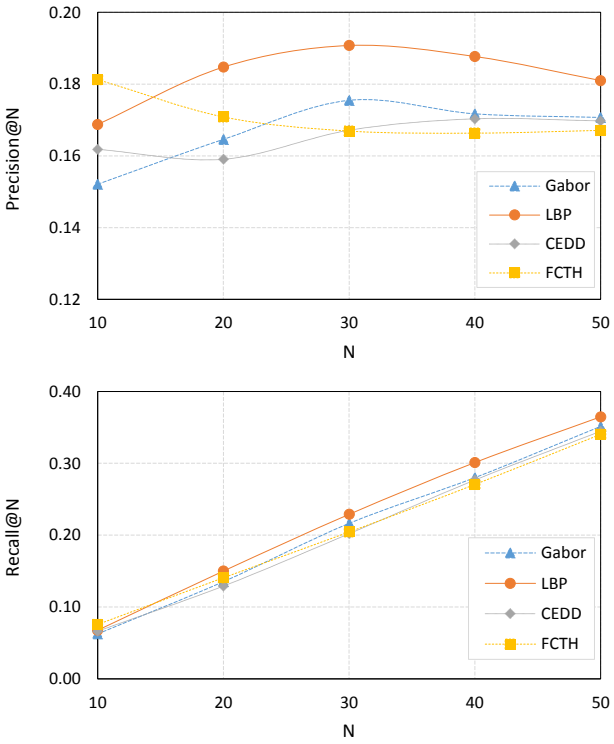


**Fig. 4**. Facial expression retrieval precision and recall at different rank levels.

| ***k*** | 1 | 3 | 5 | 7 | 9 |
|---|---|---|---|---|---|
| Gabor | 80.6 | 85.4 | 81.3 | 73.6 | 72.9 |
| LBP | **91.0** | 88.9 | 83.3 | 78.5 | 75.7 |
| CEDD | 60.4 | 70.1 | 63.2 | 58.3 | 52.8 |
| FCTH | 34.0 | 53.5 | 44.4 | 41.7 | 38.2 |

**Table 2**. Face recognition results with a *k-nn* baseline.

| ***k*** | 1 | 3 | 5 | 7 | 9 |
|---|---|---|---|---|---|
| Gabor | 15.3 | 23.6 | 19.4 | 18.8 | 19.4 |
| LBP | 13.9 | 29.9 | 22.2 | 21.5 | 22.2 |
| CEDD | 14.6 | 31.3 | 22.9 | 22.2 | 21.5 |
| FCTH | 11.1 | **34.0** | 24.3 | 31.9 | 27.1 |

**Table 3**. Facial expression classification results with a *k-nn* baseline.

feature space for face recognition in the thermal domain: although in the top 10 closest images to any given query only ∼30% belong to the correct person, Figure 3, these are in fact the majority (the remaining ∼70% belong to other persons).

### 4.3. Facial Expression Recognition and Retrieval Results

A more subtle, and challeging, task in face analysis concerns the identification of a person's facial expression. A facial expression is composed by several facial Action Units (AU), which are the movements of localized face muscles. These subtle changes in a person's face are usually detected by Gabor and LBP features in the visible domain. In the thermal domain these AU are less noticable due to the low variation of temperature caused by muscles movement.

Similarly to the previous section, we first analysed the geometry of the feature spaces from the perspective of facial expressions. The first observation we extract from Figure 4, where we present the precision and recall behavior, is that LBP and Gabor features are not the best features in the closest neighborhood. The precision graph shows that FCTH offers the best precision at the top 10 images and presents a monotonic behaviour, unlike the remaining feature spaces that exhibit some undesirable precision variations. The recall metric shows a constant increase and no noticable difference between the image descriptors.

Concerning the facial expression classification task, Table 3, the FCTH descriptor delivered the best facial expression classification accuracy (34%). This result adds knowledge to the face analysis literature that has worked with Gabor and LBP in the visible domain, but should now look at other image descriptors in the thermal domain such as the FCTH descriptor.

| | sad | disgust | angry | surprise | fear | happy |
|---|---|---|---|---|---|---|
| Gabor | **26.3** | 6.5 | 30.0 | 25.0 | 6.3 | 41.2 |
| LBP | 21.1 | 19.4 | 30.0 | **37.5** | 18.8 | **44.1** |
| CEDD | 21.1 | **48.4** | 15.0 | **37.5** | **31.3** | 26.5 |
| FCTH | 10.5 | 38.7 | **40.0** | 33.3 | **31.3** | 41.2 |

**Table 4**. Facial expression detection results per descriptor type.

To further understand which feature spaces work better with each expression, we analysed the accuracy of each descriptor in the detection of the different facial expressions. Table 4 presents the results. We can see that there is no clear winner in terms of feature descriptors. There isn't a descriptor that is good for every facial expression due to the subtle differences in each expression in the thermal domain. However, different descriptors showed good results for different expressions. CEDD had the highest accuracy (48.4%) while detecting disgust. FCTH also showed a decent result (38.7%) for this expression when compared to Gabor (6.5%) and LBP (19.4%). This result is surprising due to the fact that disgust is usually difficult to detect in the visible domain. The easiest expression to detect was happiness followed by surprise, which normally are also the easiest expressions to detect in visible images.

## 5. DISCUSSION

In this paper we compared different thermal image descriptors for facial image analysis. Two image descriptors, Gabor and LBP, widely used in facial image processing were compared to two other descriptors, FCTH and CEDD, that are more popular in medical images processing.

A key fact that we observed is that extracting facial expressions in the thermal domain is a very difficult task. However, it is possible to do person identification with such low-cost devices. This brings us to the first main contribution of this paper. Localized-Binary-Patterns (LBP) builds a feature space whose geometry allows recognizing a person with a very high accuracy (91%). Moreover, we used a simple *k-nn* classifier to allow a clean comparison of the feature spaces because since it does not make assumptions of the input data, it is closer to the real geometry of the face. However, we believe that these accuracies can be improved with more sophisticated classifiers.

Finally, the experiment concerning facial expression recognition revealed two interesting facts: first, temperature differences are too low to be accurately detected by the image descriptors (the best classification accuracy was 34%), and second, the best descriptor turned out to be the FCTH descriptor that has shown several success in the medical imaging domain.

## REFERENCES


[1] Z. Zhu, J. Fei, and I. Pavlidis, "Tracking human breath in infrared imaging," *IEEE Symposium on Bioinformatics and Bioengineering*, pp. 227–231, 2005.

[2] S. G. Kong, J. Heo, B. R. Abidi, J. Paik, and M. A. Abidi, "Recent advances in visual and infrared face recognitiona review," *Computer Vision and Image Understanding*, vol. 97, no. 1, pp. 103–135, Jan. 2005.

[3] L. Trujillo, G. Olague, R. Hammoud, and B. Hernandez, "Automatic Feature Localization in Thermal Images for Facial Expression Recognition," *IEEE CVPR'05 - Workshops*, 2005.

[4] S. Jarlier, D. Grandjean, S. Delplanque, K. N. Diaye, I. Cayeux, D. Sander, P. Vuilleumier, and K. R. Scherer, "Thermal Analysis of Facial Muscles Contractions," *IEEE Trans. on Affect. Comp.*, vol. 2, no. 1, 2011.

[5] B. R. Nhan and T. Chau, "Classifying affective states using thermal infrared imaging of the human face," *IEEE Trans. on Biomed. Eng.*, vol. 57, no. 4, 2010.

[6] A. M. Guzman, M. Goryawala, J. Wang, A. Barreto, J. Andrian, N. Rishe, and M. Adjouadi, "Thermal imaging as a biometrics approach to facial signature authentication.," *IEEE journal of biomedical and health informatics*, vol. 17, no. 1, pp. 214–22, Jan. 2013.

[7] D.A. Socolinsky, L.B. Wolff, J.D. Neuheisel, and C.K. Eveland, "Illumination invariant face recognition using thermal infrared imagery," *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[8] P. Ekman and W. Friesen, "Facial Action Coding System: A Technique for the Measurement of Facial Movement," *Consulting Psychologists Press*, 1978.

[9] M. Dahmane and J. Meunier, "Continuous Emotion Recognition Using Gabor Energy Filters," in *Affective Computing and Intelligent Interaction*. 2011.

[10] B. S. Manjunath, "Texture features for browsing and retrieval of image data," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, 1996.

[11] SavvasA. Chatzichristofis and YiannisS. Boutalis, "Cedd: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval," in *Computer Vision Systems*. 2008.

[12] S. A. Chatzichristofis and Y. S. Boutalis, "FCTH: Fuzzy Color and texture histogram a low level feature for accurate image retrieval," *WIAMIS 2008 - Int'l Workshop on Image Analysis for Multimedia Interactive Services*.

[13] S. Wang, Z. Liu, S. Lv, Y. Lv, G. Wu, P. Peng, F. Chen, and X. Wang, "A Natural Visible and Infrared Facial Expression Database for Expression Recognition and Emotion Inference," *IEEE Trans. on Multimedia*, vol. 12, no. 7, pp. 682–691, Nov. 2010.