Pain Recognition: dataset analysis and experimental validation

Manuel Pagliuca
Master student in Computer Science
University of Milan "La Statale"
Milan, Italy
manuel.pagliuca@studenti.unimi.it

***Abstract*—The aim of this project is to test the accuracy of early and late fusion approaches on a multimodal dataset to classify the presence of pain in patients. Participants were subjected to an external heat-pain stimulus through a physical device. Their facial expressions and biophysical signals were recorded through the use of cameras and the application of electrodes, then features were extracted. The descriptors came from two different modalities and will be combined by testing both fusion approaches. Finally, classifications and accuracy estimates were made, based on which it was possible to determine that early fusion is the most accurate approach for the dataset considered.**

***Index Terms*—affective computing, natural interaction, pain, classification, machine learning, features fusion, computer vision**

I. Introduction

Pain detection is a fundamentally important field in pain management, but the phenomenon itself is not yet fully understood. The difficulty in detection lies in the subjectivity of the experience itself, consequently, the need arises to empirically define what may be common characteristics. Following an introduction to automatic recognition methods from the paper [4] of the same name by *Philip Werner et al.*, an analysis began on how to broadly develop the model for this project. The goal is to create a system for pain detection by starting with a **multimodal** dataset (*biophysical* and *video* signals) and classifying it using supervised learning models (SVMs). The most complex part of this project is the extraction of video features, in which *computer vision* techniques were involved. Next, feature engineering techniques were used to perform the different classifications. Following the classifications, two labels are provided in the output to note the presence or absence of pain. Given the multi-modality of the dataset, two different fusion techniques (*early* and *late* fusion) were tested.

The project was developed following the work done by *Kächele et al.* [1]

II. Dataset and Feature Extraction

The database considered is the *BioVid Heat Pain* [2]. A total of 90 subjects participated in the experiment, recruited in the following age groups: (1) 18-35 (N = 30; split half man/women); (2) 36-50 (N = 30; split half man/women); (3) 51-65 (N = 30; split half man/women). The subjects received an expense allowance. The study was conducted according to the ethical guidelines of Helsinki (there was an ethics committee: 196/10-UBB/bal).

For the pain solicitation, a thermode has been used, and each of the 4 different stimulation strengths was applied 20 times to give rise to a total of 80 responses. From the electrodes applied on the participants, the following biophysical signals could be traced: *electromyography* (EMG), *electrocardiogram* (ECG), and *electrodermal activity* (GSR).

During the experiment, participants' faces were captured using 3 Pike F145C AVT cameras, one from the front and two from the side (only the front one was considered for the following project).

Unfortunately, for unknown reasons, the downloaded dataset was damaged and was missing multiple elements that could play a key role in the accuracy of the classifier (such as depth image and some biophysical signals). The actual number of participants in the dataset actually considered is 87.

A. Biophysiological Feature Extraction

Biophysical features belonging to the dataset are organized in a *.csv* file, and the various samplings are discriminated by a participant identifier and trial number (which is the same for video recording).

Within the file are the ECG, GSR, and EMG values, but also the corresponding recorded pain levels (the future labels for supervised learning).

There are 4 recorded pain levels and they correspond to the following strings in the dataset:

- BL0, when there is no pain present.
- PA1, PA2, and PA3 are the intermediate levels (in ascending order) of perceived pain.
- PA4 indicates the highest level of perceived pain.

For recording these pain levels, facial EMGs were not recorded, this is so that facial expressions could be recorded on screen (only the EMG of the trapezius muscle was recorded, and zygomatic and corrugator were lost).

Features are extracted from the file and loaded at runtime onto a data frame. In the feature engineering phase, since a simpler binary classification of pain presence is opted for, all intermediate pain levels are discarded from the table. Subsequently, the pain labels are encoded in binary for ease of future use by the classifier.

III. Video Feature Extraction

Video feature extraction involved the use of techniques from the world of computer vision and image processing to create multiple vectors containing the extracted features with an overall size of 19D.

How the video features are distributed on the vectors:

- 8D for facial landmark distances.
- 5D for facial expressions.
- 6D for head pose estimation.

Unfortunately, due to a loss of dataset data, the Kinect camera footage could not be accessed; head pose estimation was done using only the frontal recording.

Video feature extraction is performed for each frame of the video samples; each of which lasts 5 seconds. At the end of the video, a weighted average of the features is calculated (weighted against the number of frames considered for each feature, some frames may be discarded).

The video feature extraction method follows that of the paper [3] by *Philippe Werner et al.*

A. Facial distances

There are distances on the face that are considered empirically relevant for the characterization of pain. To extract these features, the Euclidean distance is calculated and saved in an 8D vector.

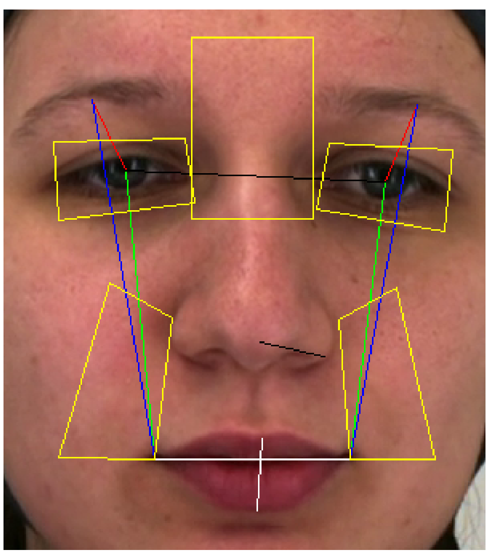


Fig. 1: Debug frame for video feature extraction.

The following distances between facial landmarks were tracked (the colors of the debug frame lines in the figure 1 are shown in brackets):

- Eyebrow-iris (red lines).
- Eyebrow-cleft lip (blue lines).
- Iris-cleft lip (green lines).
- Lip commissures (mouth width, horizontal white line).
- Upper lip-lower lip (mouth height, vertical white line).

B. Facial expressions

Feature extraction of facial expressions considers 5 polygonal regions defined on the participant's face (yellow polygons in Fig.1), which are empirically considered to be relevant in the study of pain.

For each of these regions, the *average gradient* (per frame) is calculated, in order to detect any changes (averaged out) in expression during stress. The regions concerned are:

- Nasolabial folds.
- Eye closure.
- Corrugator muscle.

The calculation of the gradient only takes place in the polygonal regions considered, and through *Sobel* operators.

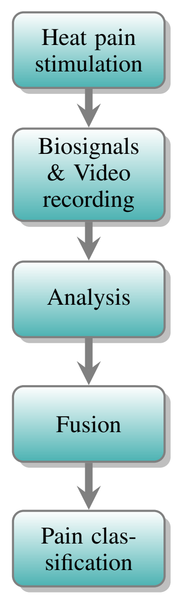
The horizontal and vertical variation is calculated in two separate variables, the absolute value is taken and the variables are merged with a bitwise OR.

C. Head pose estimation

For the calculation of head position, a method was used that, considering the facial landmark of the nose, inferred the angles of rotation of the neck. The feature vector for approximating the direction of the head is composed of the positions of the nose and the angles of rotation.

IV. Classification and Fusion Approaches

Given the multi-mode nature of the dataset, two conglomerate approaches were considered for the classification task.



Diag. 1: Workflow of the experimental procedure for pain recognition.

The diagram shows the high-level pipeline of the workflow. Following pain stimulation, audio and video features are obtained from *distinct* sources, and there will be two different analyses for these. Then fusion techniques will be applied, and in conclusion, the engineered features will be classified.

The first approach called "*early fusion*" consists of the concatenation of the biophysical and video features prior to classification. This is easily achieved with an inner join operation by a number of trials on the two data frames.

The second approach is called "*late fusion*", multiple classifiers are considered; each of these classifiers is trained on a specific feature of biophysical signals (GSR, ECG, ...) and video features.

Three SVMs are trained on the ECG, GSR, and video features respectively.

In the next step, a test dataset is iterated over and for each sample (*input*) its prediction is calculated using the different classifiers trained (obviously passing only the features). The mode of the 3 predictions obtained is calculated, i.e. whether or not the majority of the classifiers detect the presence of pain. If the mode of the predictions coincides with the ground truth, then the prediction is considered correct.

A. Support Vector Machine

Support vector machines (SVMs) are supervised learning models that analyze data for classification and regression analysis. It allows the separation of higher dimensionality data into at least two groups (*Maximal Margin Classifiers* and *Support Vector Classifiers* aren't able to handle complex data overlapping and defining a proper threshold for the classification). The SVM is a classifier that maximizes the margin between the positive and negative classes. The optimal hyperplane w is obtained by optimizing:

$$\min_{w,\xi} \frac{1}{2} w^T w + C \sum_i \xi_i$$

under the constraints $y_i(w^T x_i + b) \geq 1 - \xi_i, \forall i$ and $\xi_i \geq 0, \forall i$ where y_i is the label of sample x_i. For correctly classified points ξ will be equal to 0, and for all incorrectly classified it is the distance of that particular point from its correct hyperplane. C is a parameter that controls the penalty of samples that lie inside the margin region for linearly non-separable problems.

Kernel functions are a class of algorithms for scheme analysis, they are widely known for their application with SVMs. They map the data in a multidimensional feature space, transforming the feature space into a euclidean space. The most frequently used functions are the *radial basis*, *sigmoid*, *poly*, and *linear* functions.

V. Experimental Validation

The experimental approaches taken for classification have been multiple, distinct feature type classifications have occurred (Tab.I), but also multi-modal feature fusion approaches (Tab.II).

TABLE I: SVMs classifiers accuracies

Stimulus	ECG	GSR	ECG Trapezius	Video
BL0 vs. PA4	57.2413%	71.6091%	57.9310%	59.5402%

Considering the first approach, the kernel functions that have been used for training without fusion are only two, the RBF (*Radial Basis Function*) for biophysical signals, and the Poly (*Polynomial function*) function for video features.

SVM classifiers especially found good accuracy values with biophysical features, while compared to the paper [2] from which this project draws, there was a decrease in accuracy for the classification of video features.

A probable reason may be that it wasn't possible to access the totality of video samples that the original dataset had.

TABLE II: SVMs classifiers accuracies on fusion approaches

Stimulus	Early fusion	Late fusion
BL0 vs. PA4	72.4137%	60.2298%

Regarding the multimodal fusion approaches, there are good results given by the early fusion (better than the individual biophysical features), the same is not true for the results of the late fusion approach, which although they do not disappoint expectations fail to hold comparison with the previous approach.

For the results obtained with late fusion, *linear* kernel functions were used in all three SVMs.

VI. Conclusion

From the evaluation of the accuracies we obtained, we can state that in the case of the fusion approaches, the one that provides the most accurate prediction is the early fusion with a surplus of 12.1839% compared to the late approach.

The development of the project can be considered concluded as the results obtained are positive in both approaches, it must be said that as pain detection is an active field of computer science research (especially with fusion approaches), and there is a wide variety of approaches for classifying the latter that have not been considered (e.g. random forests).

References

[1] Markus Kächele, Philipp Werner, Ayoub Al-Hamadi, Günther Palm, Steffen Walter, and Friedhelm Schwenker. Bio-visual fusion for person-independent recognition of pain intensity. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9132 of *Lecture Notes in Computer Science*, pages 220–230, Cham, 2015. Springer International Publishing.

[2] Steffen Walter, Sascha Gruss, Hagen Ehleiter, Jun-Wen Tan, Harald C. Traue, Stephen Clive Crawcour, Philipp Werner, Ayoub Al-Hamadi, and Adriano O. Andrade. The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system. In *CYBCONF*, pages 128–131. IEEE, 2013.

[3] Philipp Werner, Ayoub Al-Hamadi, Robert L Niese, Steffen Walter, Sascha Gruss, and Harald C. Traue. Towards pain monitoring: Facial expression, head pose, a new database, an automatic system and remaining. In *British Machine Vision Conference*, 2013.

[4] Philipp Werner, Daniel Lopez-Martinez, Steffen Walter, Ayoub Al-Hamadi, Sascha Gruss, and Rosalind W. Picard. Automatic recognition methods supporting pain assessment: A survey. *IEEE Transactions on Affective Computing*, 13(1):530–552, 2022.