Multi-modal emotion recognition using recurrence plots and transfer learning on physiological signals

ELALAMY, Rayan, FANOURAKIS, Marios Aristogenis, CHANEL, Guillaume

DOI: 10.1109/ACII52823.2021.9597442
Multi-modal emotion recognition using recurrence plots and transfer learning on physiological signals

Rayan Elalamy  
Computer Science Department  
University of Geneva  
Geneva, Switzerland  
Rayan.Elalamy@unige.ch

Marios Fanourakis  
Computer Science Department  
University of Geneva  
Geneva, Switzerland  
Marios.Fanourakis@unige.ch

Guillaume Chanel  
Computer Science Department  
University of Geneva  
Geneva, Switzerland  
Guillaume.Chanel@unige.ch

Abstract—In this paper we propose to use Recurrence Plots (RP) to generate 2D representations of physiological activity which should be less subject dependent and better suited for non-stationary signals such as EDA. The performance of spectrograms and RPs are compared on two publicly available datasets: AMIGOS and DEAP. Transfer learning is employed by using a pre-trained ResNet-50 model to recognize emotional states (high vs low arousal and high vs low valence) from the two types of representations. Results show that RPs reach a similar performance to spectrograms on periodic signals such as ECG and plethysmography (F1 of 0.76 for valence and 0.74 for arousal on the AMIGOS dataset) while they outperform spectrograms on EDA (F1 of 0.74 for valence and 0.75 for arousal). By combining the two sources of information we were able to reach a F1 of 0.76 for valence and 0.75 for arousal.

Index Terms—Emotion recognition, recurrence plots, spectrograms, physiology, deep transfer learning.

I. INTRODUCTION

In recent years, automatic emotion recognition has gained a lot of attention as it can have many applications in various fields such as health [1], human-computer interaction [2] or user profiling [3]. In general, this problem is approached by using facial expressions or bio-sensing features such as electroencephalogram (EEG), electrodermal activity (EDA) and electroencephalogram (ECG). Moreover, the wide adoption of real-time wearable devices for monitoring physiological signals [4], [5] allows to collect data more easily in a non-intrusive way. These devices will further facilitate the use of physiological signals in the human computer interaction area. Recently, the use of deep learning techniques which outperform other more traditional emotion recognition classification algorithms such as Support Vector Machines (SVM), and k-Nearest Neighbors (KNN) [6] increased the interest in emotion recognition as it is more adapted to deal with bio-signals and videos.

Choosing signal representations remains one of the first step to build an emotion classification model and this choice has a significant impact on the model architecture and its performance. For example, the three conventional approaches are building models using raw signals [7], extracting engineered features [6] or converting them into spectrograms.

One of the advantages of converting signals to spectrogram images, is that they can be used in very powerful pretrained neural networks such as ResNet [8] or VGG [9] as feature extractors. Moreover, you can use inputs of different length as long as your converted images have the same dimensions. This process has been widely used for audio classification tasks [10]. However, spectrograms raise some issues. Indeed, for non-periodic and non-stationary physiological signals such as EDA, spectrograms are not well suited. To deal with it we propose to use recurrence plots (RPs), an image extraction method. This graphical representation allows to detect non-stationarities that are not captured in a traditional spectrogram [11]. Recurrence plots have been widely used in a wide range of areas such as material research [12] or financial market prediction [13]. The most widely used affect space to perform emotion recognition tasks is the emotion circumplex model [14]. This model is a circular configuration representing emotions in a two dimensional space along two axes representing arousal and valence. The space can be divided into four different parts High Valence High Arousal (HVHA), High Valence Low Arousal (HVL), Low Valence Low Arousal (LVL), Low Valence High Arousal (LVA). The overall contribution of this work is to study the performance of transfer learning with ResNet applied on recurrence plots to do emotion recognition task and compare it with the more conventional spectrogram.

II. RELATED WORK AND HYPOTHESES

The interest of physiological signals applied to emotion recognition is that they are difficult to control unlike facial expressions that can be disguised intentionally [15]. Even when attempting to mask facial expressions, physiological signals can still convey the inner emotional state [16]. Many studies demonstrated that ECG and EDA are well suited for emotion recognition [17]. Table I shows the different approaches and results in the context of physiological emotion recognition on the AMIGOS and DEAP datasets. As can be seen, several approaches have been taken for emotion recognition including shallow (e.g. SVM, Naive Bayes) and deep methods (e.g. CNN, LSTM).

Recent works showed that deep learning models can extract relevant discriminative features that are able to capture com-
plex non-linear dependencies between signals and the emotion to be predicted. Some of the most successful ones are Convolutional Neural Networks [6]. In general, these deep learning models perform better than manual feature engineering approaches and classical machine learning methods for emotion recognition. In [18], researchers combined spectrogram based features extracted by using pretrained VGG16 and 1D statistical features to obtain state of the art results on the DEAP and AMIGOS datasets. More recently, researchers applied in [19] an attention-based bidirectional LSTM-RNN on bio-signals spectrograms only. Transformation of 1D signals to 2D images has several advantages such as the possibility of representing signals of different length in a fixed image resolution.

Alternatively to spectrograms, recurrence plots are starting to be more and more present in image extraction from bi-signals. State of the art results were obtained by using recurrence plots in time series classification [20] and several studies used them to classify bio-signals [21], [22]. Recurrence plots can be particularly relevant in this context since they are able to inherently cope with non-stationary and non-periodic signals. To our knowledge, all studies investigating emotion recognition from physiological RPs first compute features from the plots using a technique called recurrence quantification analysis (RQA) [23]. For instance, a k-nearest neighbor classifier was used on 13 RQA features extracted from EEG signals in [21]. In [22], public speaking anxiety recognition is performed by applying RQA on EDA signals. In this case only two features were considered (the entropy and the recurrence rate) reaching an accuracy of 79% on their own dataset. These studies demonstrate that RPs are relevant for emotion recognition but they to not provide comparisons with more classical approaches such as frequency analyses.

RQA have been compared to power spectral densities (PSD) in [24]. In this paper, the authors propose to compute entropy using RQA on EEG signals. More precisely, an image is constructed for each sample by combining the computed entropy on several frequency bands and channels. A CNN is then applied to perform classification of the obtained images. This approach is tested on five participants independently and compared to PSD features classified with an SVM. The results demonstrates that the RQA entropy outperforms PSD features with a kappa score of 0.88 and 0.59 respectively. However it is not clear if this performance difference is due to the change of feature (RQA vs. PSD) or method (CNN vs. SVM). A statistical analysis was also performed in [25] showing significant differences for several RQA features between mental stress and rest conditions. RQA was applied on the inter-beat intervals of an ECG signals and similar results were obtained when doing a frequency analysis.

To sum up, although few studies attempted to analyse the emotion recognition performance of RPs and to compare it with more traditional features such as PSD, the potential advantages of RPs remains unclear. In addition, most studies rely on RQA without exploiting the RPs structures directly, despite of the impressive performance of CNNs on spectrograms and other images. Finally, RPs have been employed on uni-modal physiological signals only. This study thus aims at (i) providing insights on the feasibility of using CNN transfer learning for the classification of RPs, (ii) comparing the efficiency of RPs to spectrograms on two different physiological measures (EDA and heart activity) and their fusion. Concerning the first objective, we hypothesise that a pre-trained CNN such as RESNet-50 should be able to extract features similarly to RQA and thus to reach state of the art performances. Concerning the second objective, our hypothesis is that recurrence plots will be particularly efficient for EDA signals, since they tend to show non-cyclic and non-stationary patterns, while the heart activity is more periodic and well represented by a spectrogram.

### III. DATASETS

In this section we describe the two publicly available datasets that we used to evaluate our model. We only mention ECG/Photoplethysmogram (PPG) and EDA descriptions as we focus on these types of signals in this paper.

#### A. DEAP

The DEAP dataset is a multi-modal dataset for the analysis of human affective states presented in 2012 [29]. EDA and PPG signals of 32 participants were collected with a sampling rate of 512Hz as each watched 40 music videos of same duration (60 seconds). Participants rated each video in terms of the levels of arousal and valence (1 to 9 continuous values for each dimension). By considering that a low value was smaller or equal to 5 and a high value was greater than 5, we mapped each rating to one of the four spaces high vs low arousal and high valence vs low arousal.

![Figure 1. Repartition of arousal/valence labels in DEAP](image)

#### B. AMIGOS

The AMIGOS Dataset is a more recent multi-modal database collected in 2018 [26]. EDA and ECG signals of 40 participants were collected as they watched 16 short videos of various length (shorter than 250 seconds). ECG was samaleed at 256 Hz and EDA was sampled at 128 Hz. Participants then rated each video as in the DEAP dataset. We mapped the ratings to one of the four spaces as before. Figure 2 shows the
Table I
REPORTED STUDIES ON THE AMIGOS AND DEAP DATASETS

<table>
<thead>
<tr>
<th>Study</th>
<th>Modalities</th>
<th>Classifier</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>AMIGOS</td>
</tr>
<tr>
<td>[26]</td>
<td>EEG, ECG, EDA</td>
<td>SVM</td>
<td>Valence Acc: 0.57, Arousal Acc: 0.59</td>
</tr>
</tbody>
</table>
| [18]  | EEG, ECG, EDA, Face             | ELM        | Valence F1: 0.80 (EEG), 0.80 (ECG), 0.79 (GSR), 0.77 (Face), 0.82 (ECG/EDA fusion)  
       |                                  |            | Arousal F1: 0.74 (EEG), 0.76 (ECG), 0.74 (GSR), 0.76 (Face), 0.76 (ECG/EDA fusion)  |
| [19]  | EEG, ECG, EDA                   | Attention-based BLSTM | Valence Acc/F1: 0.83 / 0.72  
       |                                  |            | Arousal Acc/F1: 0.78 / 0.70                                                  |
| [6]   | ECG, EDA                        | 1D-CNN     | Valence Acc/F1: 0.75 / 0.71 (GSR), 0.71 / 0.68 (ECG)  
       |                                  |            | Arousal Acc/F1: 0.71 / 0.67 (GSR), 0.81 / 0.76 (ECG)                       |
|       |                                  |            | DEAP                                                                    |
| [27]  | EEG                             | SVM        | Valence (4 classes) Acc: 0.51, Arousal Acc: 0.76                         |
| [28]  | EEG, PPG, EDA, EOG, EMG, Temp., Respiration | MESAE   | Valence Acc: 0.77, Arousal Acc: 0.76                                     |
| [29]  | EEG                             | Naïve Bayes | Valence Acc: 0.62, Arousal Acc: 0.58                                     |
| [30]  | EEG, EMG, EOG                   | DBN        | Valence Acc: 0.51, Arousal Acc: 0.61                                     |
| [31]  | EEG                             | SVM        | Valence Acc: 0.73, Arousal Acc: 0.73                                     |

IV. Feature extraction

In this section we explain what are spectrograms and recurrence plots and how we processed data to extract them. To have a rigorous explanation we introduce a few notations. We denote by $x$ a signal and by $x[n]$ the $n_{th}$ value of $x$.

A. Pre-processing

For each signal to have the same sampling rate, we down-sampled them all at 128Hz. To remove high-frequency noise, we applied a low-pass filter of 45Hz for ECG and PPG and 15 Hz for EDA as it is a slow changing signal. Then, we applied min-max normalization as we are more looking for variations in the signal than absolute value.

B. Spectrograms

Spectrograms are one of the main tools in spectral analysis. They can be defined as an intensity plot of the Short-Time Fourier Transform (STFT). The complex spectrogram $X(n,k)$ is computed as the following:

$$X(n,k) = \sum_{l=-\frac{W}{2}}^{\frac{W}{2}} w(l) \cdot x[l + n] \cdot e^{2\pi jlk/W},$$

with $n$ being the sample index, $k$ the frequency bin index, $W$ the overlapping window length and $w$ being a standard ham-
Figure 3. EDA (AMIGOS), ECG (AMIGOS / first 2000 points from signals for visibility), PPG (DEAP) raw signals and their corresponding recurrence plots and spectrograms. (Plots has been generated with matplotlib using the veridis color map)

As ECG, PPG and EDA are low frequency signals we only consider the 0-15Hz frequency range. We compute the spectrogram of each signal and we convert it to an RGB image of dimension 224x224. Figure 3 shows the different spectrograms extracted from signals. Spectrograms have a great disadvantage when you want to use convolutional neural networks. On a more conventional image like a photo taken with a camera, close pixels have a great chance to belong to the same object and the fact that this local relationship is very clear makes convolutional neural networks very effective to deal with such images. However, as the two axes of a spectrogram do not represent the same concept (time and frequency) local relationships features are much harder to extract through convolutional layers. So, to reach the full potential of convolutional neural networks it can be worth using other kinds of images.

C. Recurrence Plots

Recurrence plots are a visualization tool used to analyze recurrent behaviors of time series. These plots represent the distance relationship among each part of the signal. The main idea is to show how a time series returns or moves away from previous states. To calculate a recurrence plot from a signal of length $N$, we can use the following equation:

$$ R_{i,j}(\epsilon) = \Theta(\epsilon - ||x[i] - x[j]||), \ i, j = 1, ..., N $$

where $|| \cdot ||$ is the euclidean norm, $\epsilon$ is the recurrence threshold $\Theta$ is the Heaviside function that is used to binarize the distance matrices. However, to prevent information loss, this
discretization step is usually omitted [32]. Then, we can simplify the previous formula as:

\[ R_{i,j} = ||x[i] - x[j]||, \quad i, j = 1, \ldots, N \]

In general, we consider four types of recurrence plot topologies:

- **Homogeneous**: corresponds to stationary time series. The RP has a strong main diagonal and some white noise in the off-diagonal pixels.
- **Periodic**: for oscillating systems. On such RP we can see a small pattern repeating an all the graph.
- **Drift**: typical of signal with a slow varying trend. Pixels further away from the diagonal are less present in this type of RP.
- **Disperse**: for signals with abrupt changes in the dynamics. Such RP show local and less repetitive patterns.

ECG and PPG signals are rather periodic signals although some drifts might be observed since we did not remove low frequencies. EDA signals can be either seen as drifts when there is a slow decay of skin conductance without any electrodermal response, and disperse when there is a response. You can see on Figure 3 periodic RP for ECG/PPG and disperse RP for EDA. Contrary to spectrograms, local relationship is very clear in recurrence plots as the two axes represent the same concept (time). Then, recurrence plots are very effective to detect recurrent behavior, such as periodicities or irregular patterns.

V. EMOTION RECOGNITION MODEL

A. Transfer Learning to deal with small datasets

AMIGOS and DEAP are relatively small datasets in the context of deep learning (respectively 640 and 1280 different samples). To achieve their full potential, deep neural networks must be trained on very large datasets such as ImageNet [33] which is an image database designed for use in visual object detection research. It contains more than 14 million images that are classified in more than 20 000 categories. To cope with our small datasets, we use transfer learning techniques with ResNet, one of the most effective pretrained model on ImageNet.

B. ResNet

ResNet [8] is a deep convolutional neural network that introduced residual blocks which consist in layers with shortcut connections which solved the problem of vanishing gradients. In our model we use ResNet-50 as feature extractor. Even though ResNet is much deeper than other models such as VGG16 [9], the model size is actually significantly smaller due to the usage of global average pooling rather than fully-connected layers.

C. Model Architecture

We perform transfer learning techniques with ResNet50 that means that we use the network pretrained for object detection task on the ImageNet dataset. That means weights are fixed and are not modified through our training. The goal is to take advantage of the efficient features representation of an image at the end of ResNet to encode our images. To be more precise, we use the output of the last residual block of dimension \(7 \times 7 \times 2048\) and we apply a max pooling layer to extract a 2048-dimensional vector for each image. Then, to reduce the dimensionality of the features representation of each image we use Principal Component Analysis (PCA) and select the 30 most important components as suggested in [18]. To combine the two modalities a feature level fusion is performed by concatenating the two sets of 30 extracted features. Then we perform logistic regression with L2 regularization on the extracted features with binary cross-entropy loss to classify Low/High Arousal/Valence. Figure 4 shows the global architecture of our model that takes as input two \(224 \times 224 \times 3\) RGB images and return the emotion class corresponding to the input.
D. Training and evaluation

We train and evaluate the logistic regression model on one dataset at a time and we perform a leave-one-out subject cross validation. An independent binary classification is performed for each emotion dimension (arousal and valence). Note that only the logistic regression step must be done twice as ResNet features extraction and PCA do not depend on labels.

VI. RESULTS

A. Single modality

Table II shows the performance of the model for each modality by using the different features extraction methods. We obtained performance results comparable to [18] and to [6] that are state of the art on the AMIGOS dataset as you can see on Table I. For EDA, F1 scores are between 0.67 and 0.74 for arousal; and 0.71 and 0.79 for valence. Our RP plots approach reached a F1 score of 0.75 for arousal and 0.74 for valence. It thus obtained the highest performance for arousal, and is in the state of the art for valence. This demonstrates the possibility to employ a transfer learning approach for emotion recognition from physiological signals images.

In addition, it can be seen that for ECG/PPG signals recurrence plots and spectrograms obtain equivalent results but that recurrence plots outperform spectrograms for EDA signals. A paired samples Wilcoxon test was done to check if the difference between RP and spectrograms is significant. The difference for EDA is strongly significant for the AMIGOS dataset \((p − value < 0.0001)\) and the DEAP dataset \((p − value < 0.001)\), while there is no significant difference between spectrogram and RP for ECG. That confirms the fact that RPs are as suited as spectrograms for periodical signals. The results show that RPs are more effective than spectrograms on signals that are non-stationary such as EDA. Nevertheless, it remains possible to use RPs for periodic signals too as it seems there is no performance loss compared to spectrograms based models.

B. Fusion of modalities

Table III shows the performance of the model with modality fusion. We can see that in all the cases, it slightly improves models performance. Again we obtain results comparable to the state of the art and we still notice that recurrence plots outperform spectrograms in this case. However, there is no significant difference between spectrograms and RP results. That is due to the fact that ECG spectrograms are effective enough to obtain good classification results and compensate the lack of performance of EDA spectrograms. Moreover, the difference between models with single modality and models with fusion of modalities is not significant. That can be due to the fact that only one simple scheme of fusion was used.

VII. CONCLUSION

In this paper we performed emotion classification from physiological signals. The proposed method consists of transforming the one dimensional signals into two dimensional images which captures temporal global pattern. We evaluated our proposed method on two publicly available datasets: DEAP and AMIGOS. Our first hypothesis was validated by showing that RPs can be effective as image representation of ECG, PPG and EDA signals. Moreover it could reach performance comparable to the state of the art. Experimental results showed that recurrence plots can outperform spectrograms, especially when dealing with non-periodic and non-stationary signals such as EDA, thus validating our second hypothesis. To the best of our knowledge, this study was the first attempt to use deep transfer learning techniques on recurrence plots for physiological emotion recognition.

These results indicate that recurrence plots can be a real asset for physiological signal analysis applied to emotion recognition. In future work, it can be valuable to explore different ways of modality fusion. For example, we could merge the 2048 features from each modality and then apply PCA on the 4096 obtained features. Analysing the fusion of RP images with spectrograms would also allow to leverage advantages of both approaches. The obtained results also need to be validated with other architectures than ResNet-50 to ensure that they are agnostic to the type of model. Finally, further investigations must be done on cross datasets learning.

REFERENCES


