Pain Expression as a Biometric: Why Patients’ Self-Reported Pain doesn’t Match with the Objectively Measured Pain?

Anonymous ISBA 2016 submission

Abstract

Developing a vision-based efficient and automatic pain intensity measurement system requires understanding the relationship between self-reported pain intensity and pain expression in the facial videos. In this paper, we first demonstrate how pain expression in facial video frames may not match with the self-reported score. This is because the pain and non-pain frames are not always visually distinctive; though the self-report tells different story of having pain and non-pain status. On the other hand previous studies reported that general facial expressions can be used as biometrics. Thus, in this paper we investigated the relevance of pain expression from facial video to be used as a biometric or soft-biometric trait. In order to do that, we employed a biometric person recognition scenario by using features obtained from the pain expression pattern found in the temporal axis of subjects’ videos. The results confirmed that the pain expression pattern has distinctive features between the subjects of the UNBC McMaster shoulder pain database. We concluded that as the pain expression pattern have subjective features as a biometric, this can also cause the difference between self-reported pain level and the visually observed pain intensity level.

1. Introduction

“Pain is an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage” - this is how ‘pain’ was defined by the International Association for the Study of Pain (IASP). It is a prevalent medical problem and needs to be managed effectively as a moral imperative, a professional responsibility and a duty of medical practitioners [1]. The widely used technique to measure pain level is ‘self-report’. However, self-reported pain level assessment does not always effectively apt in practical scenarios due to inconsistent metric properties across dimensions, efforts at impression management or deception, and differences between clinicians’ and sufferers’ conceptualization of pain [2]. Moreover, it requires cognitive, linguistic and social competencies that make self-report unfeasible to use for young children and patients with limited ability to communicate [3]–[5].

Visual pain expression, revealed in the face, can be considered as a subset of facial expression and expresses emotion valley regarding to experiencing pain [6]. It can also provide the information about the severity of pain that can be assessed by using the Facial Action Coding System (FACS) of Ekman and Friesen [7], [8]. The FACS has long been used for measuring facial expression appearance and intensity. Thus, vision-based approaches came into scene to measure pain by using features from facial appearance change. Prkachin first reported the consistency of facial pain expressions for different pain modalities in [9] and then together with Solomon developed a pain metric called Prkachin and Solomon Pain Intensity (PSPI) scale based on FACS in [10].

Several studies were conducted to find the correlation between self-reported pain and facial expression changes observed visually, as it is necessary to understand this relationship to develop a vision-based efficient and automatic pain intensity detection system. Many of them reported that self-report and pain expressions are largely unrelated [9], [11]–[13]. On the other hand, some others found significant relationship between these two [14]–[18]. Prkachin et al. provided an explanation for such discrepancies among these studies [10]. They brought forward a psychometric problem exhibited by the methods of [9], [11]–[13] by stating that these methods used very few measures of subjective reports of pain levels. On the other hand, Kunz et al. showed that visual analysis of pain becomes more difficult to be correlated with self-report in the presence of external factors like ‘smiling in pain’ and social motives [19]. The relationship of gender (male’s vs female’s way of experiencing) to pain was reported in [20], [21]. A glimpse of the reason why pain expression may not match with the self-reported score can be found in Figure 1. From the facial images in the figure, we can see that the pain and non-pain frames are not visually distinctive so much; however the self-report tells different story of having pain and non-pain status.
In this study, we use a database created by Lucey et al. with the UNBC-McMaster shoulder pain database [22]. The pain frames are at the left and the non-pain frames are at the right.

Recent studies reported that general facial expressions like sad, anger, happy, etc. translated by FACS can be used as a biometric or soft-biometric trait in person identification [23]–[26]. As pain expression in the face is a subset of facial expression, pain expression may also have some distinctive biometric property to identify subjects. Thus, in addition to the aforementioned three reasons from [10], [19]–[21] of reporting the lack of relationship between self-report and pain expression, there can be another reason that pain expression in the face is subjective and varies from person to person even though self-reported pain levels are same. However, this reasoning needs to be justified and this is the first concern of this paper.

Facial expressions are different for different emotional state like sad, happy, disgust. Study showed that general facial expressions including sad, happy, disgust, anger, fear and surprise of different people for the same emotional state also vary [23]. Thus, like many other biometric traits such as Electrocardiogram (ECG), Phonocardiogram (PCG), gait, gesture, etc. [27]–[29] general facial expressions can be used as biometric or soft-biometric for authentication or forensic investigations, as shown in [23]–[26]. Though pain can be considered as a subset of facial expressions, it is not investigated in the literature that whether pain expression patterns of different persons are distinctive or not. Thus, we can investigate whether or not the expression patterns are distinctive between the subjects. This is the second concern of this paper.

The contributions of this paper are to address these two concerns mentioned in the previous two paragraphs. We analyze different subjects pain expression pattern exhibited in the temporal axis of video frames and find whether pain expression patterns are distinctive between the subjects. If we find that they are distinctive between the subjects, then we can conclude as follows:

- Along with other reasons, the varying pattern of pain expression in temporal domain with respect to subjects’ identity is a reason of finding self-report and pain expressions are largely unrelated.
- Like other facial expression patterns obtained from facial video; the pain expression pattern is so distinctive between the subjects that it can be used as a biometric/soft-biometric.

In order to do that, we employ a biometric person recognition scenario by using features obtained from the pain expression pattern found in the temporal axis of subjects’ videos. The outcome of the paper can be used in further research to understand the difference between self-reported pain level and visually observed pain level from facial expression. Understanding of this relationship will in turns helps to develop more accurate automatic pain detection system using visual features that will match with self-reported pain levels by considering subject-specific patterns of pain level reporting.

The rest of the paper is organized as follows. Section 2 describes the methodology of our experiment. Section 3 demonstrates the experimental results and discussions. Finally, Section 4 concludes the paper.

2. Methodology of the Experiment
Employing pain expression pattern in a biometric person recognition scenario requires a multi-step procedure. In this section, we first describe a shoulder pain expression database, the UNBC-McMaster database [22], to be used in our experiment. We then demonstrate the procedure of extracting pain expressions from each frame of video sequences and employing these expressions in the temporal axis of the video sequences as pain pattern for a biometric authentication experiment.

2.1. The database
In this study, we use a database created by Lucey et al. with
the title “PAINFUL DATA: The UNBC-McMaster Shoulder Pain Expression Archive Database” [22] and the database is hereafter referred to the UNBC-McMaster database. The database contains facial video sequences of participants who had been suffering from shoulder pain and were performing a series of active and passive range of motion tests to their affected and unaffected limbs on multiple occasions. The database also contains FACS information of the video frames, self-reported pain scores in sequence level and facial landmark points obtained by Active Appearance Model (AAM) [30], [31]. The database was widely used in the literature including [32]–[35].

Figure 2 Some example video frames from the UNBC-McMaster shoulder pain database [22].

The database was created by capturing facial videos from 129 participants (63 males and 66 females). The participant had a wide variety of occupations and ages. During data capturing the participants underwent eight standard range-of-motion tests: abduction, flexion, and internal and external rotation of each arm separately as suggested in [36]. Participants’ self-reported pain score along with offline independent observers rated pain intensity were recorded. Figure 2 shows some example video frames from the database.

2.2. Extracting pain expressions from the frames

Pain expression in a face can be observed by analyzing different facial actions such as eyebrow-raising, cheeks-raising, nose-crinkling, lip-raising, lips-pulling, etc. [3]. These facial actions can be described by 44 different facial action units defined in [7]. A vast body of literature described which units out of these 44 action units represent pain-information. A list of the relevant action units is provided in Table 1. Except AU43, all of these action units are coded on a 5-levels intensity dimension (A-E or a-e) by a human FACS coder in a frame-by-frame basis. The maximum intensity is denoted by E/e and the slightest indication of AU’s existence is denoted by A/a. The AU43 is coded by 2-levels closure status.

<table>
<thead>
<tr>
<th>Action Unit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU1</td>
<td>Raising inner eyebrow corners</td>
</tr>
<tr>
<td>AU2/2L/2R</td>
<td>Raising outer eyebrow corners</td>
</tr>
<tr>
<td>AU4</td>
<td>Lowering eyebrows</td>
</tr>
<tr>
<td>AU5</td>
<td>Raising upper eyelids</td>
</tr>
<tr>
<td>AU6</td>
<td>Raising cheeks</td>
</tr>
<tr>
<td>AU7</td>
<td>Pulling up eyelids</td>
</tr>
<tr>
<td>AU9</td>
<td>Crinkling nose</td>
</tr>
<tr>
<td>AU10</td>
<td>Raising upper lip</td>
</tr>
<tr>
<td>AU12/12L/12R</td>
<td>Pulling up lip corners obliquely</td>
</tr>
<tr>
<td>AU14/14L/14R</td>
<td>Tightening lip corners</td>
</tr>
<tr>
<td>AU15</td>
<td>Pulling down lip corners</td>
</tr>
<tr>
<td>AU16</td>
<td>Pulling down lower lip</td>
</tr>
<tr>
<td>AU17</td>
<td>Pulling up chin boss</td>
</tr>
<tr>
<td>AU18</td>
<td>Pulling lips together</td>
</tr>
<tr>
<td>AU20</td>
<td>Pulling lips horizontally</td>
</tr>
<tr>
<td>AU22</td>
<td>Funneling lips</td>
</tr>
<tr>
<td>AU23</td>
<td>Tightening lips</td>
</tr>
<tr>
<td>AU24</td>
<td>Pressing lips against each other</td>
</tr>
<tr>
<td>AU28</td>
<td>Sucking lips into the mouth</td>
</tr>
<tr>
<td>AU32</td>
<td>Biting lip</td>
</tr>
<tr>
<td>AU37</td>
<td>Wiping lip</td>
</tr>
</tbody>
</table>
Around two decades ago Prkachin reported that only four action units- AU4, AU6/AU7, AU9/AU10 and AU43—carry the majority information about pain. This report was later confirmed in a recent investigation and a pain scale called PSPI was developed based on the FACS information of facial pain expression [10]. This PSPI metric is defined by a sum rule as follows:

\[ PSPI = AU4 + (AU6|AU7) + (AU9|AU10) + AU43 \]  

where, \((\ . \ | . )\) operator refers to the greater one among the two arguments. The summation result yields a 16-point scale. The details of this scale can be found in [10]. The authors of the UNBC-McMaster database provides FACS coded information for the video frames in the database [7], [8]. By employing the aforementioned sum rule on these FACS values for the frames we can calculate the pain intensity level of each frame in PSPI scale. If we consider \(\text{PSPI}_{\text{score}}\) of one frame, it provides us the instantaneous pain intensity level in that frame. However, in a video sequence we can obtain the frames \(\text{PSPI}_{\text{score}}\) or FACS values as time-series. As our interest is to investigate whether the pain expression patterns are distinctive between subjects, we obtain time-series of \(\text{PSPI}_{\text{score}}\) and FACS values to generate pain expression patterns to be employed in a biometric authentication framework. The details of the time series configuration will be provided in the experimental environment section.

2.3. Biometric authentication framework

A biometric authentication framework consists of four basic building blocks: a) data acquisition module, b) feature extractor, c) training module and d) testing module [37]. Figure 3 shows the structure of the framework used in our experiment. We accomplished the first two steps of the authentication system by using the off-the-shelf UNBC-McMaster shoulder pain database. While creating the database, the data acquisition phase was accomplished by using simple digital camera and the features were extracted as the FACS values using certified human FACS coder as discussed before.

The rest of the two modules require train/test partition of the database. The training and testing module also require a machine learning approach to accomplish biometric authentication as a classification task. In order to do that, we employ an Artificial Neural Network (ANN) [38]. A basic ANN contains sets of neuron divided into input layer, hidden layers and output layer. When input layer receive the input data, it calculates the weights by employing an activation function to generate the outputs in the neuron(s) of the output layer. The detailed parameter values regarding to our experimental setup will be provided in the experimental environment section.

![Biometric authentication framework](image)

Figure 3: The biometric authentication framework used in our experiment.

3. Experimental Results

3.1. Experimental environment

We used the UNBC-McMaster shoulder pain database to evaluate the performance of pain expression as a biometric/soft-biometric trait. The original paper of the database reported 48398 FACS coded facial video frames [22]. However, the online portal of the database (http://www.pitt.edu/~jeffcohn/PainArchive/) does not contain all of these data mentioned in the original paper. Currently, we have 31971 frames from 16 subjects with FACS codes among which 4922 frames have pain intensity levels 1-12 in PSPI scale. The distribution of the pain frames with all the frames for the subjects are listed in Table 2. Exploiting temporal axis information from pain expression in a video sequence requires considering the FACS values from more than one frame. Thus, we generate the feature vector by aggregating the FACS values of a frame and 30 previous frames along with their calculated values.
PSPI score. Our objective is not to distinguish between pain and non-pain frames by using the FACS values. Instead we would like to realize the whether the patterns of FACS values as the representation of pain in video sequences of different subjects are distinctive to each other.

Table 2 Subject-wise pain/non-pain frames in the experimental UNBC-McMaster shoulder pain database.

<table>
<thead>
<tr>
<th>Subject ID</th>
<th>No. of pain frames</th>
<th>Total no. of frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>239</td>
<td>2134</td>
</tr>
<tr>
<td>2</td>
<td>92</td>
<td>1120</td>
</tr>
<tr>
<td>3</td>
<td>64</td>
<td>1608</td>
</tr>
<tr>
<td>4</td>
<td>84</td>
<td>894</td>
</tr>
<tr>
<td>5</td>
<td>522</td>
<td>2752</td>
</tr>
<tr>
<td>6</td>
<td>95</td>
<td>2609</td>
</tr>
<tr>
<td>7</td>
<td>98</td>
<td>773</td>
</tr>
<tr>
<td>8</td>
<td>160</td>
<td>1612</td>
</tr>
<tr>
<td>9</td>
<td>512</td>
<td>2474</td>
</tr>
<tr>
<td>10</td>
<td>1120</td>
<td>2038</td>
</tr>
<tr>
<td>11</td>
<td>471</td>
<td>1502</td>
</tr>
<tr>
<td>12</td>
<td>498</td>
<td>809</td>
</tr>
<tr>
<td>13</td>
<td>181</td>
<td>2361</td>
</tr>
<tr>
<td>14</td>
<td>148</td>
<td>3360</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>2819</td>
</tr>
<tr>
<td>16</td>
<td>638</td>
<td>2706</td>
</tr>
</tbody>
</table>

A feed-forward ANN based classification framework was implemented in Matlab as shown in Figure 4. The number of hidden layers for the ANN was 5, the number of neurons in the input layer was 223 (based on the number of input non-zero FACS values), and the number of output neuron was 1. When training data is fed to the network, the ANN learns the weights to transform the inputs to an output. We then feed the testing data to get the testing results. The output neuron provides a subject ID automatically calculated by the neural network from the weights (w), and feature values (b) where the pain pattern is expressed in the FACS values. If the subject ID matches with the ground truth ID value, then it is a success. We randomly divided the experimental data by employing a test/train ratio of 0.05 to 0.50, where 0.05 refers to 95% training data and 5% testing data from the total database. Whole process was iterated 10 times to ensure multifold validation in each test/train configurations.

3.2. Performance evaluation

The ANN training validation errors and testing accuracies obtained in 10-fold executions of a test/train configuration 0.05 are listed in Table 3. In addition, the authentication results for the testing frames of all 16 subjects from one execution of test/train configuration 0.05 are shown in a confusion matrix (a row matrix) at Table 4. The true positive detections are shown in the first diagonal of the matrix, false positive detections are in the columns, and false negative detections are in the rows. From the results of Table 3, we can observe that randomly dividing the database into testing and training set with a test/train configuration may yield different testing accuracies in different executions; however the network learns some distinctive features in every attempt. The testing accuracy also showed proportional consistency with the validation error generated by the ANN for the train data. In addition, execution time for 31571 frames is around 152 frames per second in the worst case scenario of 9th execution cycle. The confusion matrix also shows that a good number of true positive identifications were achieved for the most of the frames.

Figure 4 The feed-forward ANN implemented in Matlab.

Table 3 Results of the 10-folds execution of a test/train configuration 0.05.

<table>
<thead>
<tr>
<th>Execution turn</th>
<th>Validation error</th>
<th>Testing accuracy (%)</th>
<th>Time to execute (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.023</td>
<td>98.94</td>
<td>134.07</td>
</tr>
<tr>
<td>2</td>
<td>0.015</td>
<td>100.0</td>
<td>198.87</td>
</tr>
<tr>
<td>3</td>
<td>0.354</td>
<td>64.84</td>
<td>93.60</td>
</tr>
<tr>
<td>4</td>
<td>0.113</td>
<td>86.87</td>
<td>76.10</td>
</tr>
<tr>
<td>5</td>
<td>0.042</td>
<td>99.32</td>
<td>143.94</td>
</tr>
<tr>
<td>6</td>
<td>0.051</td>
<td>96.59</td>
<td>205.31</td>
</tr>
<tr>
<td>7</td>
<td>0.313</td>
<td>66.68</td>
<td>184.27</td>
</tr>
<tr>
<td>8</td>
<td>0.013</td>
<td>100.0</td>
<td>152.71</td>
</tr>
<tr>
<td>9</td>
<td>0.091</td>
<td>90.41</td>
<td>207.53</td>
</tr>
<tr>
<td>10</td>
<td>0.266</td>
<td>70.56</td>
<td>203.32</td>
</tr>
</tbody>
</table>

In order to explore the identification accuracy for different test/train configurations of the UNBC-McMaster database, we listed the results of 10-fold execution of
different test/train configurations from 0.05 to 0.50 in Table 5. From the results we can observed that when the network get big number of training samples in 0.05 test/train configuration, the testing accuracy is very high. When the training data is reduced the accuracy also reduces slightly, until when the network does not get sufficient training data (e.g. the case of 0.50 configuration). The standard deviation of the 10-fold execution also increases when training data is reduced.

3.3. Discussions

The primary objective our investigation was to clarify whether or not the pain expression patterns can distinguish between the subjects of the UNBC-McMaster shoulder pain database. We used the FACS values of facial video frames in temporal axis as pain expression pattern and obtained very high accuracy in distinguishing between the subjects. Thus, the results reasonably lead us to the conclusion that like other facial expression patterns obtained from facial video [24]; the pain expression pattern is also distinctive between the subjects and it can be potential candidate to be used as a biometric/soft-biometric trait. In addition, along with many other reasons [19], the varying pattern of pain expression in temporal domain with respect to subjects’ identity can be a reason of finding self-report and pain expression based PSPI scores are largely unrelated.

Table 4 Confusion matrix for distinguishing between the subjects in frame levels by using pain expression pattern in a test/train configuration 0.05.

| Subjects | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 | S10 | S11 | S12 | S13 | S14 | S15 | S16 | Total testing frames |
|----------|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|---------------------|
| S1       | 71 | 32 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 103                 |
| S2       | 0  | 47 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 47                  |
| S3       | 0  | 0  | 86 | 0  | 0  | 0  | 0  | 0  | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 86                  |
| S4       | 0  | 0  | 0  | 40 | 0  | 0  | 0  | 0  | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 40                  |
| S5       | 0  | 0  | 0  | 0  | 134| 0  | 0  | 0  | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 134                 |
| S6       | 0  | 0  | 0  | 0  | 0  | 135| 0  | 0  | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 135                 |
| S7       | 0  | 0  | 0  | 0  | 0  | 0  | 40 | 0  | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 40                  |
| S8       | 0  | 0  | 0  | 0  | 0  | 0  | 72 | 0  | 0  | 0   | 0   | 0   | 0   | 0   | 0   | 72                  |
| S9       | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 130| 0  | 0   | 0   | 0   | 0   | 0   | 0   | 130                 |
| S10      | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 108| 0   | 0   | 0   | 0   | 0   | 0   | 108                 |
| S11      | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 79  | 0   | 0   | 0   | 0   | 0   | 79                  |
| S12      | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0   | 43  | 0   | 0   | 0   | 0   | 43                  |
| S13      | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0   | 124| 2   | 0   | 0   | 0   | 126                 |
| S14      | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0   | 178| 0   | 0   | 0   | 0   | 178                 |
| S15      | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0   | 0  | 134| 0   | 0   | 0   | 134                 |
| S16      | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0   | 0  | 0  | 123| 123| 0   | 246                 |

4. Conclusions

In this paper, we first pointed out that pain expression in facial video frames may not match with the self-reported score. This is because the pain and non-pain frames are not always visually distinctive enough; though the self-report tells different story of having pain and non-pain status. On
the other hand previous studies reported that facial expression patterns can be used as a biometric. Bearing these in mind, in this paper we investigated the relevance of pain expression from facial video to be used as a biometric or soft-biometric trait. In order to do that, we employed a biometric person recognition scenario using ANN with features obtained from the pain expression pattern found in the temporal axis of subjects’ videos. The results confirmed that the pain expression patterns have distinctive features between the subjects of the UNBC McMaster shoulder pain database. As the pain expression patterns have subjective features to be used as biometric, this can cause the difference between self-reported pain level and the PSPI score.

Our present study has the limitations that the database with 16 different subjects is not big enough and the database only contains shoulder pain expressions. However, the outcome of the paper is expected to be used in the future research to understand the difference between self-reported pain level and visually observed pain level in the facial pain expression. Understanding of this relationship will in turns helps to develop more accurate automatic pain detection system using visual features.

Table 5 Multifold identification results with different test/train configurations.

<table>
<thead>
<tr>
<th>Test/train configuration</th>
<th>Avg. testing accuracy</th>
<th>Standard deviation of 10 fold execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>90.67</td>
<td>11.87</td>
</tr>
<tr>
<td>0.10</td>
<td>79.01</td>
<td>15.12</td>
</tr>
<tr>
<td>0.20</td>
<td>77.61</td>
<td>12.19</td>
</tr>
<tr>
<td>0.30</td>
<td>77.48</td>
<td>12.87</td>
</tr>
<tr>
<td>0.40</td>
<td>78.70</td>
<td>18.32</td>
</tr>
<tr>
<td>0.50</td>
<td>67.57</td>
<td>17.72</td>
</tr>
</tbody>
</table>

References


