On Pain Assessment from Facial Videos Using Spatio-Temporal Local Descriptors

Ruijing Yang[1], Shujun Tong[2], Miguel Bordallo[3], Elhocine Boutellaa[4], Jinye Peng[2], Xiaoyi Feng[2] and Abdenour Hadid[2,3]
[1]Department of Information Science and Technology, Northwest University, Xi'an, China.
[2]School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China.
[3]Center For Machine Vision and Signal Analysis, University of Oulu, Finland.
[4]Centre de Développement des Technologies Avancées, Algeria.
e-mail: miguelbl@ee.oulu.fi

***Abstract*— Automatically recognizing pain from spontaneous facial expression is of increased attention, since it can provide for a direct and relatively objective indication to pain experience. Until now, most of the existing works have focused on analyzing pain from individual images or video-frames, hence discarding the spatio-temporal information that can be useful in the continuous assessment of pain. In this context, this paper investigates and quantifies for the first time the role of the spatio-temporal information in pain assessment by comparing the performance of several baseline local descriptors used in their traditional spatial form against their spatio-temporal counterparts that take into account the video dynamics. For this purpose, we perform extensive experiments on two benchmark datasets. Our results indicate that using spatio-temporal information to classify video-sequences consistently shows superior performance when compared against the one obtained using only static information.**

***Keywords*— Automatic pain assessment, facial expression, spatio-temporal features, LBP**

I. Introduction

Pain is defined as "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage" [1]. It is a feedback of the protection mechanism for the body against potential harmful behavior, as well as an indicator of health condition. Hence, pain monitoring is of great importance. The key to successfully monitoring pain is its accurate assessment [2], since overtreatment or overmedication may lead to various physical problems for the patients, while insufficient treatment may cause mental suffering and pathophysiological effects [3].

Currently, the golden standard of pain assessment is self-report, a convenient method that does not require any special skills. In self-report, the patient is asked by clinicians to quantify the experiencing pain level. However, this assessment usually suffers from the lack of reliability and subject variance, e.g. it may be affected by personal experiences. Moreover, self-report is not suitable for unconscious people who cannot express pain vocally (i.e. dementia, neonates, ICU patients).

In this context, automatically recognizing pain from spontaneous facial expression is of increased attention, since it can provide for a direct and relatively objective indication to pain experience. However, this is a very challenging problem that needs to deal with video sequences taken in various conditions that many times only show subtle changes in expression during time. In order to investigate the automatic pain assessment problem, researchers have collected several databases in constrained environments. These databases contain digital images or videos depicting subjects suffering different levels of pain. The baseline ground truth is then obtained from reported pain scores based on self-reports and measurements from observers.

However, most of the existing works have focused on analyzing pain from individual images or video-frames, hence discarding the spatio-temporal information that can be useful in the continuous pain assessment. In this paper, we investigate the role of the spatio-temporal information in pain assessment by comparing the performance of several baseline local descriptors used in their traditional spatial form against their spatio-temporal counterparts that take into account the video dynamics.

Our contributions can be summarized as follows:

- To assess the presence of pain in video sequences, we propose the extraction of spatio-temporal local descriptors from videos of painful subjects
- We perform extensive comparisons of local descriptors in two datasets, UNBC-McMaster Shoulder Pain Expression Archive and BioVid Heat Pain
- We show that a consistent improvement in the performance can be obtained when using spatio-temporal features compared with their spatial counterparts
- In addition, we obtain state-of-the-art accuracy for pain assessing from videos when applied to a 2-class problem

The rest of this paper is organized as follows: Section 2 discusses the related work that can be found in the literature. Section 3 describes our proposed methodology based on spatio-temporal features and classification. Section 4 depicts the experimental results and their analysis. Finally, Section 5 concludes the paper and offers some future directions.

II. Related work

During the past decade, the development of pain recognition from faces has moved from the assessment of static images [4] to dynamic video sequences [5], from the recognition of acted painful expression [6] to spontaneous pain recognition [7], [8] and from pain detection [9] to pain intensity estimation [10].

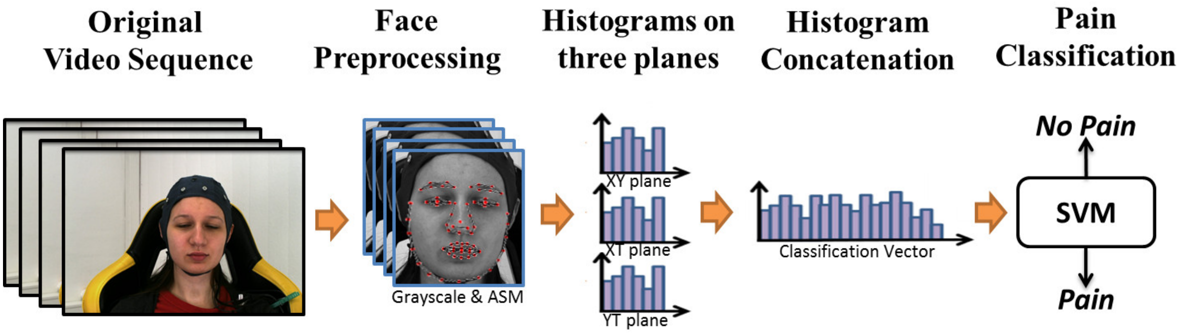

Fig. 1: Simplified scheme of the classification methodology.

Coding facial features and their relation to pain started with Craig et al. [11], who evidenced that the changes in facial appearance could be a useful cue for pain recognition. Following research has pinpointed the correlation of facial expressions with the activity of specific muscles. Since facial expressions are complex to quantify, numerous efforts have been made in the development observational systems. Facial Action Coding System (FACS) [12] is used to describe the corresponding correlation between different facial muscle movement and facial expressions described by 44 independent action units (AU). Prkachin and Solomon [13] found that four action units are correlated with pain and carry the bulk of pain information. They proposed the Prkachin and Solomon Pain Intensity Metric (PSPI), which defined pain as the sum of the intensities of brow lowering, orbital tightening, levator contraction, and eye closure. This metric is able to discern among 16 discrete pain levels and is the only one able to assess intensity of pain on a frame-by-frame level.

The methodology used in pain analysis from faces varies across studies. However, generally an automatic pain assessment system usually focuses on the extraction of meaningful features, the reduction of data dimensionality and the employment of classifiers that assign the estimated pain level.

Among the most typical used features, local descriptors have proven to be suitable to perform the task with statistically significant outcomes [14]–[16]. Other usual features include the ones encoding changes in the distances of facial points/landmarks [5], [8], [17], [18] or model-based features [7], [9], [19], [20].

Among the most used classifiers, Support Vector Machines (SVM) seem to show very good performance, either when utilizing linear kernel [18], [20] or radial basis function kernel [8], [21]. Other classification strategies include Multiple Instance Learning with boosting [22] or Random Forests [8]. Transfer Learning [23] or Regression Techniques [15] have also been utilized to reduce the inter-personal variation.

The majority of the work found in the literature for both static images and video-sequences is based on frame-level pain detection [7], [15], [16], [23] and does not assess the specific information obtained by the possible dynamic textures. Frame-level pain assessment datasets that only use frame-level ground truth (FACS, PSPI) are very costly in terms of labeling data. Hence they are prone to contain mistakes that can introduce bias in the assessment methods [9] [22]. On the other hand, using a coarser ground truth, such as observer labeling or self-report, video sequence-level pain assessment could be more efficient than frame-level assessment for a practical setup, since it is able to predict a meaningful overall pain level on a whole video sequence.

III. Adopted methodology

We propose a methodology for automatic pain assesment from facial videos of painful subjects. Our proposed approach consists of five main steps. It starts with detecting, segmenting and aligning the face images based on eye coordinates and other facial landmarks. Then, two types of descriptors are extracted: spatial texture features from video-frames and spatio-temporal features from video sequences. We extract local binary patterns (LBP), local phase quantization (LPQ) and binarized statistical image features (BSIF) which are all extracted from the videos using Three Orthogonal Planes (TOP). The resulting feature vectors are combined using early fusion to be then used as the input of a set of Support Vector Machines that predict the if the expression of pain is present in the sequence.

A. *Face detection and tracking*

To mitigate the influence of possible inconsistent color and pose across the videos included in the database, the first step of our approach consists on segmenting the face region from each video sequence. For that purpose, we have employed an active shape model (ASM) approach that detects 68 facial landmarks and is able to track them along the video. The regions containing faces are then cropped from every frame in the video using the detected landmarks. Finally, the face-regions are aligned using key landmark points, registering them to a predefined template that preserves the interpupillary distance. All cropped videos are then converted into grayscale.

B. *Spatio-temporal local descriptors*

Local texture descriptors are highly discriminative grayscale texture descriptors. Extracted from small regions of face images, they encode local texture information, clustering them into uniform regions that are pooled in local histograms. Local descriptors are robust to illumination changes and invariant to global illumination variations. Additionally, these representations are computationally simple and allow for real-time operation [24] [25]

Originally designed for grayscale images, local descriptors are able to encode video sequences on a frame by frame basis. However, they can be extended to consider spatio-temporal information when the features are extracted from Three Orthogonal Planes (TOP). In our approach, we analyze the spatio-temporal texture of pain videos depicting faces of painful individuals using three local descriptors and their extensions: Local Binary Patterns (LBP-TOP) [26] [27], Local Phase Quantization (LPQ-TOP) [28] [29] and Binarized Statistical Image Features (BSIF-TOP) [30] [31].

LBP and LBP-TOP - In Local Binary Patterns, for each pixel in an image, a binary code is computed by thresholding a circularly symmetric neighbourhood with the value of the central pixel. The occurrences of the different binary patterns are collected into histogram to represent the image texture information. Face images are divided into several regions from which the local binary patterns are computed, concatenating them into a spatial feature vector used as a face descriptor. To incorporate dynamic temporal information, LBP features can be extracted independently from three orthogonal planes.

LPQ and LPQ-TOP - The Local Phase Quantization descriptor [28] was proposed for texture classification that is robust to image blurring. The LPQ descriptor is based on the insensitivity of the low-frequency phase components to centrally symmetric blur. Similarly to LBP, an LPQ describes a local neighbourhood with an integer ranged in [0,255]. Codes computed for all image pixel neighborhoods are collected into a histogram. The local histograms simply describes the texture.

LPQ is computed using four complex low frequencies: $u_0 = (\alpha, 0)$, $u_1 = (\alpha, \alpha)$, $u_2 = (0, \alpha)$, $u_3 = (-\alpha, -\alpha)$ where α is a small scalar frequency ($\alpha << 1$) ensuring the blur is centrally symmetric. Each pixel x of the image is characterized by a vector F_x of complex frequencies:

$$\begin{aligned} F_x =& [Re\{F(x, u_0), F(x, u_1), F(x, u_2), F(x, u_3)\}, \\ & Im\{F(x, u_0), F(x, u_1), F(x, u_2), F(x, u_3)\}], \end{aligned} \tag{1}$$

$Re\{.\}$ and $Im\{.\}$ denotes the real part and the imaginary part of a complex number.

To maximize the information preservation by the quantization process, the coefficients should be statistically independent. Therefore, a decorrelation step based on a whitening transform is applied in LPQ before the quantization process. Subsequently, the vector of whitened coefficients is quantized via a simple shareholding scheme:

$$q_i = \begin{cases} 0 \text{ if } f'_i < 0 \\ 1 \text{ otherwise} \end{cases} \tag{2}$$

where f'_i is the ith whitened coefficient. Finally, the resulting binary quantized coefficients are represented as integer value in [0-255] as follows:

$$LPQ(x) = \sum_{i=0}^{8} q_i 2^{i-1} \tag{3}$$

For the spatio-temporal case, the phase information is computed locally in a window in three orthogonal planes.The histogram is obtained by accumulating the occurrence of quantized phase code in each of them, concatenating these histograms from the three directions into a single classification feature for classification tasks.

BSIF and BSIF-TOP - Inspired by other descriptors such as LBP and LPQ, Binarized Statistical Image Features (BSIF) compute a binary code for each pixel via binarizing the responses of filters applied on local image patches which are generated through linearly projecting local image patches onto a subspace. Within the subspace, the basis vectors are learned from a small set of natural images using independent component analysis (ICA).

Given a patch X of size $l \times l$ pixels and a linear filter W_i of the same size, the filter response s_i is obtained by:

$$s_i = \sum_{u,v} W_i(u, v) X(u, v) = w_i^T x, \tag{4}$$

where w_i and x are vectors containing the pixels of W_i and X, respectively. A series of binary digits b can be obtained by binarizing each response s_i as follows:

$$b_i = \begin{cases} 1, if & s_i \geq 0 \\ 0, & otherwise \end{cases} \tag{5}$$

here, b_i is the ith element of b.

In order to learn a useful set of n filters W_i, the statistical independence of the responses s_i should be maximized. To achieve that, a sufficient number of patches are randomly sampled from the training images. The patches are then normalized to zero mean and principal component analysis (PCA) is applied to minimize their dimension to n. Finally, the filters are derived by applying the independent component analysis algorithm. Once the filter matrix W is computed, it can be directly utilized for calculating BSIF features from any image.

Following the Three-Orthogonal-Planes approach, Arashloo et al. [31] extended the work on original BSIF operator. In this case, the filters are learnt in a similar way as for the static BSIF, using independent component analysis (ICA) on each of the three planes and allowing their use in the spatio-temporal domain. In the phase of filter learning, a whitening transformation on pixel level is applied in various regions of spatio-temporal support. The filter responses can be binarized independently to produce a binary code for each temporal pixel. The descriptor is able to capture the spatio-temporal content of sequence at multiple scales by varying the sizes of the dynamic BSIF filter set.

C. Classification and Fusion

After the extraction of features, the next step is the fusion of features and classification. In this work, we adopted early fusion, fusing features directly after the feature extraction process, and directly concatenating the different feature sets. The fused features are then fed to the classifier.

To estimate the pain level of sequences and frames, we use a bi-class linear Support Vector Machine (SVM) as our classifier. For both pain detection (two-class problems) and pain intensity estimation (multi-class problems), we have used the one-vs-one strategy, i. e. one SVM is trained for each pair of classes. For testing, a mean voting method is applied to obtain the final decision.

The predictive performance of SVMs depends on parameter selection. In order to select the best error penalty parameter range (C), we split the training data and obtain the accuracy of 12 different sets of SVM, which corresponds to a search range from -6 to 16 with an interval of 2. The selected model is the one that presents the best accuracy on the training data. The final accuracy value is obtained after applying a mean pooling strategy over all different data splits.

IV. Experimental results and analysis

To gain insight into the effect of dynamic textural information in pain automatic pain assessment, we performed extensive experiments on two different benchmark pain assesment databases, namely the UNBC-McMaster Shoulder Pain Expression Archive (ShoulderPain) and the BioVid Heat Pain Database (BioVid). We compared the performance of the three different local descriptors in both their spatial and spatio-temporal form on three orthogonal planes (TOP). In this section, we describe the experimental data and setup, and then discuss the obtained results.

A. Experimental data

*1) **ShoulderPain**:* The UNBC-McMaster Shoulder Pain Expression Archive Database [32] is one of the most used databases for the tasks of automatic pain assessment. This database contains video sequences of patient's faces when they were actively and passively moving their affected or unaffected shoulders. The database is distributed to research community and aims to facilitate researchers conducting experiments on pain recognition tasks using facial expressions. The video sequences in this database are annotated both on frame-level with PSPI (ranges from 0-1) scores, and sequence-level with Observed Pain Intensity (OPI) scores (ranges from 0-5) where score 0 refer to no pain, while a value greater than 0 indicates a certain pain intensity. Figure 2 shows an example sequence.

The UNBC-McMaster Shoulder Pain Expression Archive Database [32] recorded 200 video sequences from 129 volunteers (63 male, 66 female) from various occupations and age group when experiencing a series of active and passive motion of their affected and unaffected limbs on two separate occasions. These participants are self-identified as suffering

Fig. 2: An example from the Shoulder Pain Expression Archive Database [32]

from shoulder pain. In this database, for frame-level assessment, each frame is AU-coded by certified FACS coders. For sequence-level assessment, the database includes self-report and observer ratings. The publicly available portion of this database contains:

- 200 video sequences containing spontaneous facial expressions of pain.
- 48,398 FACS coded frames.
- Corresponding frame-by-frame pain scores (PSPI) and sequence-level scores (VAS, OPI).
- 66 point AAM landmarks.

*2) **BioVid**:* The BioVid Heat Pain Database (BioVid) [33] is a recent dataset created to improve the reliability and objectivity in measuring pain. To advance automated pain recognition systems, the BioVid database records, in addition of videos showing the expression of pain (visual signals), a set of biological signals.

BioVid depicts a total number of 90 subjects, recruited from three age groups with equal number of males and females: 1) age 18-35 (30 subjects), 2) age 36-50 (30 subjects), 3) age 51-65 (30 subjects). Pain stimuli was introduced to each subject on their right arm using a thermode with four individual and specific levels of pain. Additionally, biopsychological data such as the skin conductance level (SCL), the average action potential of the heart (electrocardiogram, ECG), the electrical muscle activity (electromyography, EMG), and electroencephalogram (EEG) were recorded by a Nexus-32 amplifier. Additionally, a Kinect Sensor was utilized from above the frontal Pike camera to record depth maps, color images and the associated timestamps provided by the Kinect. However, only the visual information is available for the research community. Figure 3 shows an example of the data contained in the database.

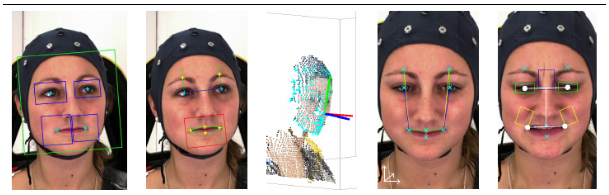

Fig. 3: An example from the BioVid Heat Pain Database and its associated data [33]

All the videos from 87 subjects are labeled with corresponding pain stimulus intensity. Videos are labeled either with no pain or with each one of the other four pain intensities: low pain (PA1, pain threshold), severe pain (PA4, pain tolerance) and two intermediate intensities (PA2 and PA3). For each subject, there are 20 video samples per class, thus giving a total number of 100 samples per person for a total 8,700 videos.

B. Database protocols and experimental setup

In our experiments, we followed the defined 2-class protocol of the ShoulderPain dataset. In this context, for the Shoulder pain dataset, following the protocol proposed in [9], [34], only subjects who had at least one trial with $OPI = 0$ and one trial with $OPI \geq 3$ were included. Our experiments are designed for pain detection, using the classification of baseline (no pain, BLN) vs pain (PA). Videos that report intermediate pain intensity with $OPI = 1$ and $OPI = 2$ were omitted. Hence, 147 sequences (53 pain, 92 no pain) from 21 subjects are used in our experiment.

Pain classification is then computed at both frame level and sequence level. Frame level experiments use spatial local descriptors and the PSPI score as the ground truth to detect pain, where frames with $PSPI \geq 1$ are deemed as positive sample (Pain), and with $PSPI = 0$ as negative samples (No pain). Sequence level experiments utilize the spatio-temporal variants of local descriptors and the OPI scores as the ground truth. Analogously, a sequence with $OPI \geq 3$ is defined as the positive class (Pain) and $OPI = 0$ as the negative class (No pain).

For the BioVid dataset, for comparative purpose, we have generated a simple 2-class protocol. The ground truth of the BioVid database is not based on self-report, and it is instead correlated with the intensity of a painful heat stimulus applied to the subjects. In this context, videos and frames recorded with no stimulation (No Pain) are compared to the ones recorded while applying a high heat painful stimulous (Heat levels 3 and 4, Pain).

To evaluate the performance of the different methods, we applied a leave-one-subject-out strategy in both datasets, ShoulderPain and BioVid. In this context, we use videos and frames from 1 subject as test data, and videos and frames from the other subjects as training data. Hence, 21 classification tasks on ShoulderPain and 87 classification tasks on BioVid were conducted and their results averaged. To ensure a fair comparison between methods and databases, the results are reported with the mean accuracy measured on the Receiver Operating Characteristic (ROC) curves.

C. Experimental Results

Table 1 shows the obtained results on the ShoulderPain database, comparing the performance of the three local descriptors at both frame and sequence levels. To explore the complementarity of the features, we also report the combination of the descriptors by fusing them at feature level. From these results, we can clearly see that the use of the spatio-temporal information of video-sequences consistently improves the pain assessment when compared with the corresponding spatial features obtained on a frame-by-frame basis.

Table 1:
Mean Accuracy (%) on the UNBC-McMaster dataset

Features	Accuracy (%) Frame level	Accuracy (%) Sequence level
Lucey et al. [34]	68.31	80.99
LBP	73.89	79.29
LPQ	67.94	77.98
BSIF	68.70	79.83
LBP+LPQ	68.47	80.41
LPQ+BSIF	68.05	**83.42**
LBP+BSIF	**73.04**	79.55

When comparing different descriptors individually, we can see that LBP features perform the best. However, the combination of several features further improves the results. This indicates that the combination of pairs of spatial features could provide complementary information. However, the high dimensionality of the resulting vectors (up to 517 dims. for frame level, 34,260 dims. for sequence level) suggest that dimensionality reduction techniques could be beneficial.

When comparing our proposed spatial and spatio-temporal features with others in the literature based on the combination of geometrical and model-based features [34], we can see that our results based on texture descriptors offer similar (although slightly better) performance. In addition, since texture and geometrical information are complementary, these results suggest that both strategies could be combined to further improve the results, highlighting that integrating temporal information within video sequences is indeed of importance.

Table 2 shows the results obtained on the BioVid database. We compare spatial and spatio-temporal descriptors with the ones provided for the head-movement obtained by Kinect and the descriptors created from time-windows [33].

Table 2:
Mean Accuracy (%) on the BioVid (part A) dataset

Features	Accuracy (%) Frame level	Accuracy (%) Sequence level
Head-movement [33]	-	67.00
Time-windows [33]	-	**71.00**
LBP	59.08	63.72
LPQ	58.82	63.19
BSIF	59.25	65.17
LBP+LPQ	60.23	63.48
LPQ+BSIF	59.83	64.51
LBP+BSIF	60.23	64.80

Again, it can be seen that the information obtained by assessing pain at a sequence level shows an improved performance when compared with its static counterpart. Spatio-temporal texture information shows comparable results to other facial-dynamics information such as the head-movement processed from 3-D data or features obtained by studying specific and known time-windows.

V. Conclusion

In this paper we presented the first study that analyzes and quantifies the role of spatio-temporal information in automatic pain assessment. We performed experiments on two different benchmark datasets, UNBC-McMaster Shoulder Pain and BioVid. We compared the performance of three spatial local descriptors and their combinations against the performance of their spatio-temporal counterparts. The obtained results indicate that temporal information obtained considering whole video sequences is of great importance when assessing pain from facial images and sequences. In addition, textural information seems to show the potential to be fused with other methods, further improving the overall assessment.

The choice of the descriptors used and the evaluated datasets is mainly motivated by the popularity and the availability of the chosen material. This is important for reproducible research and fair comparison with state of the art. However, the work is by no mean complete as our results and findings should be considered as preliminary and hence should be further validated utilizing other methods and investigating possible complementarities with other geometrical and model-based facial dynamics features.

References

[1] H Merskey, DG Albe-Fessard, and JJ Bonica. International association for the study of pain. pain terms: a list with definitions and notes on usage. *Pain*, 6:249–252, 1979.

[2] Michael Serpell. *Handbook of pain management*. Springer, 2008.

[3] Henry McQuay, Andrew Moore, and Douglas Justins. Treating acute pain in hospital. *BMJ: British Medical Journal*, 314(7093):1531, 1997.

[4] Sheryl Brahnam, Loris Nanni, and Randall Sexton. *Introduction to neonatal facial pain detection using common and advanced face classification techniques*. Advanced Computational Intelligence Paradigms In Healthcare - 1. Springer-Verlag, Berlin Heidelberg, 2007.

[5] Zakia Hammal, Miriam Kunz, Martin Arguin, and Frederic Gosselin. Spontaneous pain expression recognition in video sequences. In *International Academic Conference on Visions of Computer Society (BCS)*, pages 191–210. British Computer Society, 2012.

[6] Gwen C Littlewort, Marian Stewart Bartlett, and Kang Lee. Faces of pain: Automated measurement of spontaneous facial expressions of genuine and posed pain. In *The 9th International Conference on Multimodal Interfaces*, pages 15–21. ACM, 2007.

[7] Patrick Lucey, Jeffrey F Cohn, Kenneth M Prkachin, Patricia E Solomon, Sien Chew, and Iain Matthews. Painful monitoring: Automatic pain monitoring using the unbc-mcmaster shoulder pain expression archive database. *Image and Vision Computing*, 30(3):197–205, 2012.

[8] Philipp Werner, Ayoub Al-Hamadi, Robert Niese, Steffen Walter, Sascha Gruss, and Harald C Traue. Automatic pain recognition from video and biomedical signals. In *The 22nd International Conference on Pattern Recognition (ICPR)*, pages 4582–4587. IEEE, 2015.

[9] Ahmed Bilal Ashraf, Simon Lucey, Jeffrey F Cohn, Tsuhan Chen, Zara Ambadar, Kenneth M Prkachin, and Patricia E Solomon. The painful face pain expression recognition using active appearance models. *Image and Vision Computing*, 27(12):1788–1796, 2009.

[10] Philipp Werner, Ayoub Al-Hamadi, Robert Niese, Steffen Walter, Sascha Gruss, and Harald C Traue. Towards pain monitoring: Facial expression, head pose, a new database, an automatic system and remaining challenges. In *British Machine Vision Conference*, pages 1–13, 2014.

[11] K. D. Craig, K. M. Prkachin, and Rve Grunau. *The facial expression of pain*. Guilford Press, 1992.

[12] Paul Ekman and Wallace Friesen. *Facial Action Coding System: A technique for the measurement of facial movements*. Consulting Psychologist Press, Plato Alto, 1978.

[13] Kenneth M Prkachin and Patricia E Solomon. The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain. *Pain*, 139(2):267–274, 2008.

[14] Loris Nanni, Sheryl Brahnam, and Alessandra Lumini. A local approach based on a local binary patterns variant texture descriptor for classifying pain states. *Expert Systems with Applications*, 37(12):7888–7894, 2010.

[15] Sebastian Kaltwang, Ognjen Rudovic, and Maja Pantic. Continuous pain intensity estimation from facial expressions. In *The 8th International Symposium on Advances in Visual Computing (ISVC)*, volume 7432 of *Lecture Notes in Computer Science*, pages 368–377. Springer, 2012.

[16] Ognjen Rudovic, Maja Pantic, and Vladimir Pavlovic. Heteroscedastic conditional ordinal random fields for pain intensity estimation from facial images. *CoRR*, abs/1301.5063, 2013.

[17] Zakia Hammal and Miriam Kunz. Pain monitoring: A dynamic and context-sensitive system. *Pattern Recognition*, 45(4):1265–1280, 2012.

[18] 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 *12th International Workshop on Multiple Classifier Systems (MCS)*, volume 9132 of *LNCS*, pages 220–230. Springer, 2015.

[19] Zakia Hammal and Jeffrey F Cohn. Automatic detection of pain intensity. In *The 14th ACM International Conference on Multimodal Interaction (ICMI)*, pages 47–52. ACM, 2013.

[20] Henrik Pedersen. Learning appearance features for pain detection using the unbc mcmaster shoulder pain expression archive database. In *The 10th International Conference on Computer Vision Systems*, pages 128–136. Springer, 2015.

[21] Marian Stewart Bartlett, Gwen C Littlewort, Mark G Frank, and Kang Lee. Automatic decoding of facial movements reveals deceptive pain expressions. *Current Biology*, 24(7):738–743, 2014.

[22] Karan Sikka, Abhinav Dhall, and Marian Stewart Bartlett. Classification and weakly supervised pain localization using multiple segment representation. *Image and vision computing*, 32(10):659–670, 2014.

[23] Jixu Chen, Xiaoming Liu, Peter Tu, and Amy Aragones. Person-specific expression recognition with transfer learning. In *International Conference on Image Processing (ICIP)*, pages 2621–2624. IEEE, 2012.

[24] Miguel Bordallo López, Alejandro Nieto, Jani Boutellier, Jari Hannuksela, and Olli Silvén. Evaluation of real-time lbp computing in multiple architectures. *Journal of Real-Time Image Processing*, pages 1–22, 2014.

[25] Abdenour Hadid, Juha Ylioinas, and Miguel Bordallo López. Face and texture analysis using local descriptors: a comparative analysis. In *2014 4th International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–4. IEEE, 2014.

[26] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Trans. on pattern analysis and machine intelligence*, 28(12):2037–2041, 2006.

[27] Guoying Zhao and Matti Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *Trans. on Pattern Analysis and Machine Intelligence*, 29(6):915–928, 2007.

[28] Ville Ojansivu and Janne Heikkilä. Blur insensitive texture classification using local phase quantization. In *International conference on image and signal processing*, pages 236–243. Springer, 2008.

[29] Bihan Jiang, Michel F Valstar, and Maja Pantic. Action unit detection using sparse appearance descriptors in space-time video volumes. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 314–321. IEEE, 2011.

[30] Juho Kannala and Esa Rahtu. Bsif: Binarized statistical image features. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 1363–1366. IEEE, 2012.

[31] Shervin Rahimzadeh Arashloo and Josef Kittler. Dynamic texture recognition using multiscale binarized statistical image features. *IEEE Transactions on Multimedia*, 16(8):2099–2109, 2014.

[32] Patrick Lucey, Jeffrey F Cohn, Kenneth M Prkachin, Patricia E Solomon, and Iain Matthews. Painful data: The unbc-mcmaster shoulder pain expression archive database. In *The 9th IEEE International Conference on Automatic Face and Gesture Recognition and Workshops (FG 2011)*, volume 1, pages 57–64. IEEE Computer Society, 2011.

[33] Steffen Walter, Sascha Gruss, Hagen Ehleiter, Junwen Tan, Harald C Traue, Philipp Werner, Ayoub Al-Hamadi, Stephen Crawcour, Adriano O Andrade, and Gustavo Moreira da Silva. The biovid heat pain database: Data for the advancement and systematic validation of an automated pain recognition system. In *International Conference on Cybernetics (CYBCONF)*, pages 128–131. IEEE, 2013.

[34] Patrick Lucey, Jessica Howlett, Jeff Cohn, Simon Lucey, Sridha Sridharan, and Zara Ambadar. Improving pain recognition through better utilisation of temporal information. In *International Conference on Auditory-Visual Speech Processing*, pages 167–172. ISCA, 2008.