The Automatic Detection of Chronic Pain-Related Expression: Requirements, Challenges and the Multimodal *EmoPain* Dataset

Min S. H. Aung, Sebastian Kaltwang, Bernardino Romera-Paredes, Brais Martinez, Aneesha Singh, Matteo Cella, Michel Valstar, Hongying Meng, Andrew Kemp, Moshen Shafizadeh, Aaron C. Elkins, Natalie Kanakam, Amschel de Rothschild, Nick Tyler, Paul J. Watson, Amanda C. de C. Williams, Maja Pantic, and Nadia Bianchi-Berthouze

Abstract—Pain-related emotions are a major barrier to effective self rehabilitation in chronic pain. Automated coaching systems capable of detecting these emotions are a potential solution. This paper lays the foundation for the development of such systems by making three contributions. First, through literature reviews, an overview of how pain is expressed in chronic pain and the motivation for detecting it in physical rehabilitation is provided. Second, a fully labelled multimodal dataset (named *'EmoPain'*) containing high resolution multiple-view face videos, head mounted and room audio signals, full body 3D motion capture and electromyographic signals from back muscles is supplied. Natural unconstrained pain related facial expressions and body movement behaviours were elicited from people with chronic pain carrying out physical exercises. Both instructed and non-instructed exercises were considered to reflect traditional scenarios of physiotherapist directed therapy and home-based self-directed therapy. Two sets of labels were assigned: level of pain from facial expressions annotated by eight raters and the occurrence of six pain-related body behaviours segmented by four experts. Third, through exploratory experiments grounded in the data, the factors and challenges in the automated recognition of such expressions and behaviour are described, the paper concludes by discussing potential avenues in the context of these findings also highlighting differences for the two exercise scenarios addressed.

Index Terms—Chronic low back pain, emotion, pain behaviour, body movement, facial expression, surface electromyography, motion capture, automatic emotion recognition, multimodal database

1 Introduction

In recent years there has been a drive toward more accurate sensing and robust interpretation of activity within exercise and physical rehabilitation systems [1], [2], [3]. In part, this has been done to alleviate the high demands placed upon limited numbers of healthcare staff as well as to make rehabilitation more enjoyable (e.g., through the use of games). This has led research and industry to develop systems deployable in non-clinical settings such as the home or workplace, many with the objective of providing corrective biomechanical feedback [2]. However, in such systems, factors relating to the emotional states of the user have been largely ignored. For certain chronic conditions this is a crucial omission since emotions play a major role in the impediment of rehabilitation. It directly affects the efficacy of long term management strategies where a user can become anxious, discouraged and ultimately demotivated [4].

A particular case where emotional factors undermine adherence to successful rehabilitation is chronic pain (CP). CP is defined as pain that persists despite the resolution of injury or pathology or with no identified lesion or pathology [5]. It is attributed to changes in the central and peripheral nervous system resulting in amplified or uninhibited pain signals [6], [7]. These changes are closely linked with distress and affect behaviour, quality of life and daily function which can further result in depression, anxiety and social isolation [8].

Although management of all chronic conditions is generally subject to moderating factors that affect adoption and adherence to their respective therapies [9], CP differs in that pain conveys threat [10]. Emotionally, this generates anxiety and also contributes to catastrophic thinking. Untempered levels of anxiety can cause marked reluctance to undertake therapies which are perceived as potentially exacerbating pain to the extent of avoiding them [11], [12].

- *M.S.H. Aung, B. Romera-Paredes, A. Singh, H. Meng, M. Shafizadeh, and N. Bianchi-Berthouze are with the UCL Interaction Centre, University College London, London WC1E 6BT, United Kingdom. E-mail: {m.aung, ucabbro, aneesha.singh.10, n.berthouze}@ucl.ac.uk, Hongying.Meng@brunel.ac.uk.*
- *S. Kaltwang, B. Martinez, M. Valstar, A.C. Elkins, and M. Pantic are with the Department of Computing, Imperial College London, London SW7 2AZ, United Kingdom. E-mail: {sebastian.kaltwang08, b.martinez, a.elkins, m.pantic}@imperial.ac.uk, , michel.valstar@nottingham.ac.uk.*
- *M. Cella, N. Kanakam, A. de Rothschild, and A.C. de C. Williams are with the Department of Clinical, Educational & Health Psychology, University College London, London WC1E 6BT, United Kingdom. E-mail: matteo.cella@kcl.ac.uk, amanda.williams@ucl.ac.uk.*
- *P.J. Watson is with the Department of Health Sciences, University of Leicester, Leicester LE5 7PW, United Kingdom. E-mail: pjw25@le.ac.uk.*
- *A. Kemp is with the Physiotherapy Department, Maidstone & Tunbridge Wells NHS Trust, TN2 4QJ. E-mail: a.kemp@nhs.net.*
- *N. Tyler is with the Department of Civil, Environmental & Geomatic Engineering, University College London, London WC1E 6BT, United Kingdom. E-mail: n.tyler@ucl.ac.uk.*

Manuscript received 16 Apr. 2014; revised 7 Nov. 2014; accepted 9 Apr. 2015. Date of publication 29 July 2015; date of current version 5 Dec. 2016. Recommended for acceptance by R.A.Calvo. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TAFFC.2015.2462830

This work is licensed under a Creative Commons Attribution 3.0 License. For more information, see http://creativecommons.org/licenses/by/3.0/

In this paper, we focus on chronic musculoskeletal pain which affects an estimated one in ten adults globally [13]. For this common form of CP, avoidance results in a reduction of beneficial physical activity as well as the overuse of alternative parts of the body due to the reluctance to move perceived painful body regions. This could even lead to impairment in motor control where there is proprioceptive dysfunction [14]. The benefits of adherence to activity in rehabilitation are instead well understood. It protects against weakening, stiffness and inhibits the neurophysiological mechanisms underlying the spread of pain. It also increases confidence in physical capacity, underpins achieving valued goals [15] and improves quality of life [16].

Qualitative studies [17] showed how physiotherapists with cognitive behavioural training make use of patients' behaviour to decide upon the type and timing of encouragement during therapy. Such interventions can vary from breathing prompts to the partitioning of an activity into less daunting piecemeal tasks, or simply providing information and reassurance. Physiotherapists were also reported to use behaviour as a measure of a person's progress in learning to manage their own condition. This is used to pace the transfer of management responsibilities from the physiotherapist to the individual, eventually leading to fully effective self management.

Currently experts are unable to provide the ideal amount of continuous long-term monitoring and motivation given the large number of people with CP. This leads to a reliance on unsupervised self-management [18] which lacks emotional support and therefore risks limiting or even reversing treatment gains. Clearly, the deployment of automated systems with the capacity to recognise pain related expressions and behaviours would be a major step toward fulfilling this requirement gap. In principle, affect awareness integrated into self-use rehabilitation systems would allow for the development of systems that can provide tailored support and feedback during physical rehabilitation sessions.

This paper aims to progress in this direction in a threefold way. Through a discussion of literature in CP behaviour, we aim to provide an understanding of how CP and CP related emotions are expressed and the role they play in the exacerbation of the condition. Second, we fill a crucial empirical gap by supplying a multimodal fully labelled dataset for the most common musculoskeletal form of CP, namely chronic lower back pain (CLBP). This is a very disabling condition and often with high levels of chronicity [12]. We focus on this one form of musculoskeletal CP as mixed data from different types of musculoskeletal CP (e.g., neck or shoulder) would introduce added complexities and potential confounds within the dataset. However, it should be noted that, once a person has CLPB, the use of maladaptive body behaviour may lead to the emergence of pain in other parts of the body. The fully labelled multimodal dataset (named *'EmoPain'*) contains naturalistic pain-related affective expressions (facial and vocal expressions) and behaviours (movement and muscle activity) of people with CLBP while carrying out physical activity. Finally, we present the results of an analysis of this data with the aim to discuss some of the challenges that the automatic recognition of such expressions and behaviour presents; we also discuss possible avenues to address these challenges.

The paper is organized as follows: in the next section we review the literature on the non-verbal language of CP to ascertain the factors that contribute to pain behaviour expressivity. We also discuss how its detection can support rehabilitation. In Section 3, we review other existing datasets to clarify our contribution to the research community from this perspective. We also review the state of the art in affect recognition with respect to the modalities considered and the type of expressions we target. In Section 4, we present our data collection procedure, including details regarding patient recruitment, sensor set up and trial procedure. Section 5 describes two labelling procedures for face expression and body related behaviours. In Section 6, we report on the analysis of three of the modalities contained in this dataset with respect to the gathered labels providing the grounding for the discussion of the challenges they present. Section 7 concludes by discussing the findings and possible directions on how these could be addressed. Finally, we also provide information on access details to the *'EmoPain'* dataset in Section 8.

2 Affect in Chronic Pain

In psychology literature the catch-all term 'pain behaviour' [19] has been widely used in a variety of contexts for decades, though further refinements have been proposed [20]. In particular, a widely applied observational framework for CLBP was described by Keefe & Block [12] who identified five distinctive action-based categories (*guarding*, *bracing*, *rubbing*, *grimacing* and *sighing*) which incorporates all pain behaviour. A further categorisation was proposed by Sullivan et al. [21] who identified two functional categories. The first being *protective* behaviour defined as measures taken to avoid or minimize actual or anticipated pain. An example of this *protective* behaviour from the Keefe & Block framework is *guarding* [22], [23]. This is supported by a recent study by Tang et al. [24] where the authors showed that safety-seeking behaviour is highly prevalent among CLBP patients and has a high correspondence with anxiety and catastrophization.

In addition to such outward displays, *protective* behaviour is also measurable from anomalous levels of internal muscle activity. In a surface electromyographic study (sEMG) Watson et al. [25] showed psychological factors influenced the prolonged use of lumbar paraspinal muscles in CLBP patients during forward flexion exercises. Similarly, in a study by van der Hulst et al. [26], the authors showed high sEMG levels and a guarded mechanism during walking. In Geisser et al. [27], the authors examined the relationship between lumbar flexion, dynamic sEMG and pain related fear in people with CLBP. Correlations were found between pain-related fear and reduced lumbar flexion and increased sEMG in full flexion. Subsequent studies support this relationship between reduced lumbar motion and pain related fear [28], [29], [30].

Though called *protective* behaviour, such restriction of movement is often the cause of increased pain, worsens negative emotional states and effects withdrawal from physical activity. As discussed above, guarded movement is often produced by strong and prolonged activation of muscles. This also leads to a reduction of movements that interfere

with functioning, e.g., the reduction of arm swinging during walking reduces balance control thus exacerbating fear of falling. In addition, these movements may instigate empathic behaviour in others whose attempts to help usually discourage activity and increase focus on pain [31].

The second of Sullivan's functional categories is *communicative* behaviour where the predominant manifestations are facial expressions (i.e. *grimacing* from the Keefe & Block framework). Vocalizations and gestural body movement are also *communicative*. This is defined as deliberate or non-deliberate overt displays in order to communicate one's state and distress to observers [32], [33], and [34]. Moreover, it has no direct protective function. There has been a large volume of work on the facial expressions of pain [35]. Early work in general emotion research [36] showed distinct expressions that accompany acute episodes of pain. In [37] the authors characterized the pain faces based on the Facial Action Coding System (FACS) [38]; the consistency of these pain expressions when elicited by different stimulating modalities was demonstrated by Prkachin [39]. A later study by Kappesser & de C. Williams [40] showed that a specific pain face is distinguishable from other specific negative emotions. Studies have also shown a low correlation between observed pain behaviour and self-reported levels of pain [41]; but in contrast show a high correlation between self-reported emotional states and observed non-verbal pain behaviour for both acute and chronic pain [42]. Moreover, a review by de C. Williams [34] on the evolutionary perspectives of pain expression discussed evidence showing the voluntary modulation of some *communicative* and *protective* behaviours. However, like any other behaviour, with repeated association this can become habitual.

Unfortunately, in addition to negative effects on physical state, *communicative* and *protective* behaviours can also contribute to depression, social isolation [43]. From an emotion-embodied perspective, this can happen by the reinforcement of negative emotions and beliefs through the proprioceptive cues deriving from the unconscious or conscious enactment of such behaviour. Studies in neuroscience and psychology have shown that postures and body movements that are typically expressed during a particular emotional state such as fear or anxiety can bias the person towards that state even when the posture or movement was initially enacted for other reasons [44], [45], [46].

From a social perspective, *protective* behaviour appears to play an important role in evaluating personality traits [18] as well as physical capabilities. In [8], people showing pain-related *protective* behaviour were considered less ready to work, less likable and dependable than people expressing only *communicative* pain-related behaviour. At the same time, people exhibiting *communicative* behaviour were perceived as less likable and less dependable than people not exhibiting any pain-related behaviour or expression. Similar negative evaluations were obtained in [47] when observers were asked to evaluate the personality of a person after having observed the point-light displays of their *protective* behaviour during physical exercises.

Given these social and psychological findings, the design of systems to support people with CLBP should also consider the context of any interpersonal settings, e.g., sessions where a therapist is present versus self-directed rehabilitation where instructions are not given by another person. As such, there is a need for building a representative corpus to enable the affective computing community to pursue this. Up to now most of the studies have been based on observations mostly in very constrained settings. Given the variety of contexts where such technology could be of use (e.g., on the move [48], [49]) and the constraints imposed by such contexts on what can be sensed (e.g., with wearable technology), it is important that such a corpus should also allow for an in depth analyses of each modality. In particular, an exploration of the lesser studied *protective* behaviours is needed before embarking on system design. Little is known about how *protective* behaviour is exhibited in naturalistic settings and its relation with the more widely studied facial expressions.

3 Related Work

In this section we discuss existing public corpora related to pain and their contributions and limitations in the context of CP rehabilitation. We also review efforts to automatically recognize pain expression and emotions relevant to CLBP using the modalities contained in *EmoPain*.

3.1 Datasets

Facial expression analysis has received increasing attention over the past decade. As a result, many publicly available datasets exist that facilitate the study of facial expressions and principally consider the general emotional states and dimensions [41], [50], [51]. However, to our knowledge only one public dataset specifically focuses on pain expressivity. The UNBC-McMaster shoulder pain dataset [52] contains video sequences of spontaneous facial expressions from patients with shoulder pain. A temporally concurrent pain score based on the Prkachin and Solomon Pain Intensity (PSPI) measure [53] is assigned to each video sequence. FACS based descriptors to define an objective mapping between observable facial actions and associated pain levels were used with each frame fully FACS coded; additionally, the co-ordinates of 66 landmark points are also included along with sequence-level self reports of pain. With regard to CP this dataset is not multimodal and is comprised of participants with various diagnoses (arthritis, tendonitis, bursitis, subluxation and rotor cuff injury) and thus very different degrees of chronicity.

With regard to data specific to pain-related behaviour from full body motion capture, a public dataset has not been released to our knowledge. Nevertheless, corpora relating to the general basic affective states and body motion are available. In Ma et al. [54] the authors introduced a motion capture database of activities: walking, lifting, throwing and knocking acted in a happy, sad angry and neutral manner. The data contains 3D coordinates of 15 anatomical points across the whole body. However, this dataset only contains acted rather than naturalistic motions which are known to be more subtle and difficult to recognize. In contrast the AffectME corpus ([55], extended in [56]) does contain data from naturalistic physical activity. It contains whole body motion capture of subjects playing computer sport-games. A variety of affective expressions conveyed during game play or when re-assessing their

performance was captured and then labelled by multiple naïve observers. Though the non acted data in AffectME is valuable there is no specialization toward pain.

Whilst these corpora do contribute to the creation of an interactive system for self-directed physical rehabilitation to an extent; they were not collected with this specific setting in mind and none contains all of the pain-relevant modalities. Also, the set of activities in the above datasets do not relate to the type of activities observed in physical rehabilitations. Our dataset considers a scenario in which people with CLBP follow a set of exercises that are either typical in physical rehabilitation or are everyday activities that induces anxiety. This implies a free range of movements, self-occlusions, non-frontal viewpoints, and unsegmented recordings in which sparse episodes of pain appear within unsegmented exercises. We aim at extending existing resources by: 1) releasing a multimodal chronic pain-specific dataset, 2) considering a realistic physical rehabilitation scenario and 3) providing continuous ratings of facial pain expression and occurrences of pain body behaviour segmented by expert and naïve observers.

3.2 Automatic Recognition of Pain Expression and Behaviour

Although there is much research in the automated recognition of affect from facial expression (for surveys see: [41], [57]) there is a smaller body of work that has focused on the automated recognition of pain related expressions. The principal sensing modality for pain face recognition is computer vision. An early example is [58] who uses face shape features and artificial neural networks to classify images of subjects' faces in a normal mood versus images taken from a pain inducing task. Lucey et al. [52] (and in a subsequent study Ashraf et al. [59]) used Active Appearance Model (AAM) based features with support vector machine (SVM) classifiers to classify pain versus no pain within the aforementioned UNBC McMaster dataset, resulting in an area under ROC (AUC) score of 83.9 percent and a true positive hit rate of 82.4 percent respectively. However, in these studies the high scores could be attributed to the low threshold values used to categorise the pain label, thus setting the no pain category to principally include the least ambiguous frames.

Hammal & Cohn [60] took a different approach by stratifying the pain intensity into four levels and trained one-versus-all SVM classifiers for each class but with a reduction in performance (F1 score of 67 percent with subject dependent cross validation). Kaltwang et al. [61] took this further by estimating the full continuous pain intensity score using a framework of Relevance Vector Regression models (RVR). Tracked point locations, Discrete Cosine Transforms (DCT) and Local Binary Pattern in (LBP) based features were tested yielding an optimal correlation score of 0.59.

A shortcoming with the methods in [52], [60], and [61] is that only spatial features were used and dynamic information was not exploited, though in [60] a per sequence experiment was also done by clustering the per frame AAM based features, and with each sequence labelled by observers using a six point Likert scale. This served as a baseline for the study in [62] where the spatial AAM derived features are compacted using a DCT to compress in the spatial domain rather than the temporal domain, yielding on optimal classification rate of $\sim$81 percent.

One drawback in labeling an entire sequence is that the labels are weakly defined temporally when compared to labelling on a frame by frame basis, for example the onset and duration of a pain expression within a longer sequence is not known. To this end, Sikka et al. [63] proposed the multi segment multi-instance learning (MS-MIL) framework; this method determines subsequences which contain the expressive part of the original sequence. A set of contiguous subsequences were generated by temporal windowing or clustering and represent instances from a labelled sequence (or bag). Since labels are assigned to each bag and only positively labeled bags contain pain expressive subsequences, this can be viewed as a typical multi instance learning problem. Furthermore, this study also shows a way to determine a posterior probability of pain or no pain to each frame based on its proximity to the centre of each subsequence outperforming the methods in [62] and [59]. Romera-Paredes et al. applied a generalized multi task learning (MTL) approach to address the issue of a lack of specific facial action unit occurrences by certain subjects and show that leveraging occurrences from other subjects by learning on all subjects simultaneously can be of benefit [64]. A similar MTL framework was used in a Transfer Learning context to account for idiosyncrasies between subjects and improved on standard subtraction methods [65]. All the above studies are reduced to contain only neutral and pain expressions and thus do not resemble a naturalistic scenario where any other expression could be present. Additionally, they focus on strong acute pain expressions, whereas *EmoPain* contains expressions from unconstrained scenarios that would typically occur during a physiotherapy session for CP.

Aside from studies stemming from the release of the UNBC-McMaster dataset, there have been a variety of pain recognition studies based on other datasets. Werner et al. [66] attained a true positive rate of 92.9 percent in classifying pain expression using comparative learning on the Hi4D-ADSIP dataset but this was applied to pose data. Studies investigating differences in real and posed data include Littlewort et al. [67] who attained 88 percent accuracy by focusing on 20 specific facial action units. Bartlett et al. [68] showed that the use of face action dynamics with non linear SVMs can classify real versus fake pain expression better than human observers. Few studies outside of the UNBC-McMaster based works focused on specific clinical populations. Chen et al. [69] applied a rule based classification model to AAM features on a lung cancer patients' corpus where the pain related facial actions are more infrequent and subtle [70]; but again this used a pre-segmented constrained dataset.

Studies on body expression recognition have not been as widespread in comparison to face studies (for surveys see: [71], [72]) and to our knowledge no body related study focuses directly on CP. However, there have been works that have included affect categories that are relevant to CLBP populations such as fear, anxiety and depression [73]. Gunes & Piccardi [74] proposed a vision based bimodal system which tracked face and upper body motion to detect 12 affective states, two of which were fear and anxiety. Several static classifiers were tested at a frame level as well as

Hidden Markov Models (HMM) applied at a sequence level. This study demonstrated that a fused bimodal system was more successful than the use of face and body information separately. Another vision based face and upper body study by Joshi et al. [75] utilized spatiotemporal feature extraction with a Bag of words (BoW) framework. They demonstrated that head and body movements were as descriptive as facial expressions for detection of depression. Kleinsmith & Berthouze [76] included *avoidance* (defined as degree of attentiveness) among other standard affective dimensions: valence, arousal and potency (dominance), achieving recognition rates comparable to human observers using only postural descriptors.

With regard to the utilization of sEMG information in pain populations, the mainstay objectives have largely been for diagnostic aims. Most studies analysed statistical differences between the waveforms of people with CLBP and healthy control subjects [77], [78] but did not develop any predictive tools. However, Birrun-Manresa et al. [79] showed that a *k*-Nearest Neighbour (*k*-NN) approach successfully classified CLBP or Chronic Neck Pain patients from healthy controls using the nociceptive withdrawal reflex in the leg. Recent studies by Olugbade et al., using a small subset of *EmoPain*, demonstrated that a bimodal sEMG and motion feature set improved the accuracy in recognizing self reported levels of pain [80], [81]; similarly, Aung et al. [82] showed the possibility of detecting guarding behaviour using a similar subset.

Two studies by Huis in t' Veld et al. [83], [84] on muscle activation are also worth noting here. Although they did not explicitly develop recognition models they layed some groundwork for building a muscle based body action coding system (BACS I & II) which could facilitate recognition models. They investigated various muscle activations in relation to the expression of anger and fear by a subject and also activation responses from observers viewing the expressive subject. Their findings suggest that different muscles are active to a different extent according to passive viewing and active expression along with the type of emotion. All of these studies support the potential for CP specific recognition from body motion and muscle activity but they also indicate the need to initiate work that target the relevant emotions in the specific contexts that trigger them.

4 Data Collection

In this section we detail the acquisition and resultant content of the *EmoPain* dataset. We aimed to maximize naturality of the elicited data as well as resolution, quality and synchronization accuracy.

4.1 Patient Recruitment

Potential participants were initially identified by health care staff predominantly from the Pain Management Centre at the National Hospital for Neurology and Neurosurgery, United Kingdom as well as through pain charities such as 'Backcare'. Once identified they were informed about this study and referred to our team upon further interest.

Informed consent was sought from participants for taking part in the study and dissemination of the data including the sharing of data with other researchers. All identifiable information was anonymised (e.g., names and dates of birth). An exception to the anonymisation is the attributes within the video data. Only videos or images of those participants who provided written consent to disseminate and share video data is made available to the research community. Ethics approval was obtained through the NHS Ethics committee (11/LO/007) for people with CP recruited through the hospital and through the UCL ethics committee (10/9456) for people recruited through pain groups and for healthy participants.

For each potential participant a brief structured initial interview was carried out by a clinical psychologist trained in pain management. During this process eligibility was determined based on the Mini International Neuropsychiatric Interview (MINI) [85] to ascertain major psychiatric comorbidities other than depression and anxiety that may alter emotional expressivity (e.g., psychosis or substance abuse). Further inclusion criteria were: having CLBP for more than six months and we did not exclude those with depressed mood because its contribution to pain behaviour requires investigation.

From this superset, patients were excluded if: the principal pain was not located in the back, they had need of mobility aids, had joint replacement, arthrodesis or limb amputation, neuropathic pain, spinal stenosis, cardiovascular or respiratory disease, learning disability, poor understanding of English or were pregnant. A final set of 22 CLBP patients was determined (seven male, 15 female, mean age 50.5, 18 Caucasian, three black and one south-Asian). Though small, this group is typical of people with CP seeking treatment [16]: two thirds were female, they were mostly middle aged, and substantially disabled by their pain.

Furthermore, 28 healthy control subjects with no history of CLBP (14 male, 14 female, mean age 37.1, 26 Caucasian and two Asian) were also recruited from random volunteers from the local community as well as people known to the research team. The control participants were recruited to provide a variety of ways the recorded physical exercises would be executed in the absence of pain. Two main reasons have led to their inclusion. First, we assume that there is not a perfect way of executing an exercise, especially when not instructed, that can be taken as a model from which people with CP may deviate [17]. Second, people are idiosyncratic and hence the data should account for this to improve the effectiveness of the automatic recognition model. Although in this paper the control data is not analysed, it is included in the *EmoPain* dataset to allow for benchmarking in subsequent studies after public release.

4.2 Trial Procedure

Before recording, the CLBP group initially completed a questionnaire to ascertain pain experience, affective state and daily activity with questions based on established pain questionnaires: the Hospital Anxiety and Depression Scale (HADS) [86] and the Pain Catastrophizing Scale (PCS) [87]. The HADS score is a measure of anxiety and depression, together scored as distress, developed for use in populations with illness and disability, and widely used in CP. The PCS score assesses one of the pivotal cognitive-emotional variables in CP, with substantial predictive power in behaviour [10], [11], [15], [16]. These scores are provided in

TABLE 1
CLBP Participants' Profile Summary

Age	*Gender*	*HADS Score*	*PCS Score*	*Self Report Pain N*	*Self Report Pain D*	*Self Report Anxiety N*	*Self Report Anxiety D*
63	M	4	2	0	0	0	0
53	F	25	14	0	0.2	0	0
65	F	16	13	5.5	5.8	0.9	0.9
27	F	25	18	5.1	5.7	1.9	3.5
31	F	8	2	2.8	2.7	0	0
64	M	20	17	5	5.6	1.9	1.7
62	M	25	30	5.8	6.7	0	0
56	M	11	12	3.9	4.7	0	0
36	M	19	15	1.4	1.8	0	0
58	F	17	13	0.4	0.8	0	0
-	F	8	6	6.1	3.9	0	0
55	F	11	15	1.1	1.7	1.3	2.1
33	F	11	8	4.1	3.9	2.9	2.3
19	M	30	42	7.1	7.6	2.9	2.7
38	F	5	0	0	0	0	0
-	F	21	37	2.6	3	0	0
51	F	15	5	0.5	0.1	0.1	0
67	M	24	33	6.6	8.7	6.3	8
62	F	8	11	1.1	1.4	0.1	0.3
56	F	32	44	4.7	5.6	4	2.3
65	F	11	17	0	0	0	0
50	F	34	42	6.1	7.7	0	0
50.5		17.3	18	3.18	3.53	1.01	1.08

The scores shown are: sum of Hospital Anxiety and Depression Scores (HADS [86], scale: 0–42), sum of the Pain Catastrophizing Scores (PCS, scale: 0-52) [87] and mean levels of Self Reported Pain and Anxiety for all exercises in the normal (N) and difficult trials (D) (scale: 0-10). The final row contains the mean age and scores.

columns 3 and 4 in Table 1 with HADS ranging from 4 to 34 and PCS ranging from 0 to 44. The profiles (Table 1) were gathered to provide an understanding of the representativity of the dataset. Also, the profiles may be useful to improve the automatic recognition systems by considering these person specific factors (e.g., gender, level of depression).

Next, anthropometric measurements were manually taken using calipers: height, upper arm lengths, forearm lengths, thigh lengths, shank lengths, waist width and shoulder width. The subject's weight was also measured. Full body frontal and sagittal photographs were taken of each participant while standing inside a cube framework of a known size. These images were annotated to determine the skeletal proportions at later stage to calibrate the motion capture data.

Three sensory systems (detailed in Section 4.3) were then attached to the participant: four wireless surface electromyographic probes (Fig. 2b), a motion capture suit consisting of 18 microelectromechanical (MEMS) based inertial measuring units (IMU) (Fig. 2a) and a head mounted microphone. System initialization also included the adjustment of a camera rig supporting five face level cameras to the correct height (detailed in Section 4.3) and the calibration of the motion capture suit.

The exercises undertaken by the participants were a set of basic actions agreed by physiotherapists with expertise in treating CLBP. The exercises were varied yet consistent with known movements that generally place demands on the lower back. They are also functional activities that represent everyday tasks that those with CLBP may perceive as difficult and thus avoid for fear of increased pain [88].

For each exercise, two levels of difficulty were used and performed separately to elicit a wider range of pain-related behaviour. A minimum of two trials (one at each level of difficulty) were then conducted for each participant. The easier trial (*normal*) consisted of the following seven exercises: 1) standing on the preferred leg for 5 seconds initiated at the time of the subject's own choosing, repeated three times, 2) sitting still on a bench for thirty seconds, 3) reaching forwards with both hands as far as possible while standing, 4) standing still for thirty seconds, 5) sitting to standing and return to sitting initiated at the time of the subject's own choosing, repeated three times, 6) bending down to touch toes and 7) walking approximately 10 metres with one 180 degree turn.

In the *difficult* trial, four of the exercises were modified to increase the level of physical demand and possibly of anxiety: 1) standing on the preferred leg for 5 seconds initiated upon instruction repeated three times and then on the non-preferred leg in the same manner, 3) reaching forwards with both hands as far as possible while standing holding a 2 kg dumbbell, 5) sitting to standing and return to sitting repeated three times initiated upon instruction, and 6) walking as before while carrying one 2 kg weight in each hand, starting by bending down to pick up the weights.

After each exercise the CLBP group also self reported the level of pain and anxiety from a 0-10 scale, the mean value of these scores are shown in columns 5-8 in Table 1, the *N* and *D* descriptor indicates the normal and difficult exercise set respectively.

4.3 Recording Apparatus

As rehabilitation technology moves into non-clinical settings, an understanding of system requirements in terms of sensing modality, configuration and data granularity for affect aware systems is needed. We use apparatus that maximises fidelity and resolution; this will allow the research community to determine the minimum levels of data dimensionality, granularity and accuracy needed for robust recognition and further facilitate the design of wearable, cheaper and less invasive motion capture technology [80], [49] if the feature requirements are within the sensing limitations of the simpler devices. For example, with the advent of more accurate marker-less sensors (e.g., Kinect 2) there is a greater potential for such devices to be used.

4.3.1 Cameras and Audio

We configured eight video cameras as shown in Fig. 1. All cameras had a resolution of 1,024 × 1,024 pixels and a frame rate of 58 fps. Five of the cameras covered the frontal 90 degrees of a circle around the main exercise spot at ca. 1.5 m height, and were mounted together on an aluminium rig. Camera #8 pointed up from the floor so that the subject's face is captured when leaning forward. A long range camera was placed at the front right corner to capture a general overview of the scene. Another long range camera was placed at the front centre to capture facial expression during the walk exercise. The use of this multiple view camera set up allows for more unconstrained instruction during the exercises and therefore capturing natural movements. The main exercise area was walled by a series of 2 m

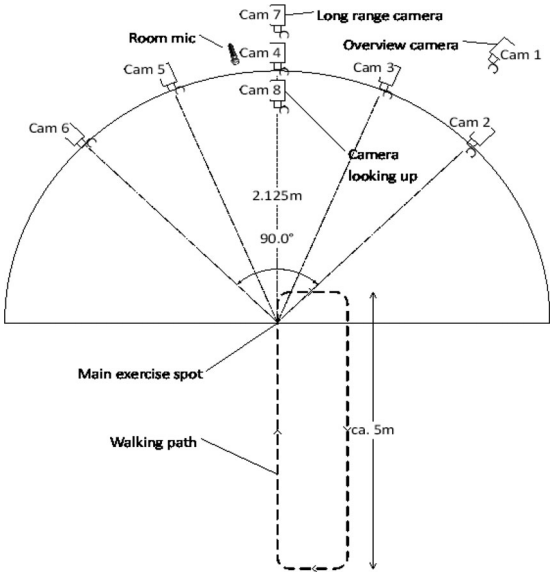


Fig. 1. Plan view of the configuration of eight high resolution cameras, five cameras mounted on a common rig to cover the frontal 90 degrees of a circle around the subject to allow for unconstrained natural movement. Two long range cameras for distance exercises and a floor camera to capture the face during forward flexion.

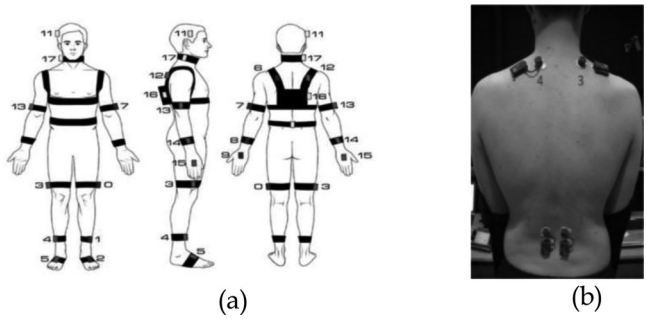


Fig. 2. IMU and EMG sensor attachments: (a) customized motion capture suit (Animzaoo IGS-190): 18 inertial measuring units attached with Velcro strapping on all main rigid body segments. The use of minimal attachment material reduces the sense of restrictiveness and to encourage naturalistic motion (diagram courtesy of Animazoo/Synertial), (b) Four fully wireless surface electromyographic sensors (BTS FREEEMG 300). Probes 3 and 4 are placed on the upper fibres of trapezius the muscles. Probes 1 and 2 are placed bilaterally on the lumbar paraspinal muscles approximately at the 4/5 lumbar vertebra.

whiteboards to improve the passive lighting conditions. In total, eight active lights were used: two pointed to the whiteboards behind the camera rig, two pointed from above the camera rig to the main exercise point, and two pointing from below the camera rig.

The tall whiteboard panels also created a more private space for the participants; only the participant and the physiotherapist or psychologist were allowed in this area. The audio signal was captured with two microphone channels, recorded at a rate of 48 kHz with 24 bit Pulse Code Modulation. The first channel was provided by an AKG C-1000S MKIII condenser microphone that was placed next to the centre camera on the rig and pointed towards the main exercise point. The second channel was recorded from a wireless AKG HC 577 L condenser headset microphone that was worn by the subject.

4.3.2 Motion Capture and Electromyography

A customized motion capture suit that specifically addresses the comfort requirements of CLBP patients based on the Animazoo IGS-190 system was used. Each sensor was a MEMS based IMU with Velcro attachment straps; this was done to minimize the amount of tight fitting material worn by the participants to enhance comfort, reduce the sense of restrictiveness and maximize naturalistic motion. Twelve sensors were placed on rigid limb segments (4 limbs $\times$ 3 segments); one on the hip, one on the centre of the torso, and one on each shoulder, neck and on the head totalling eighteen sensors (see Fig. 2a). The IMUs were connected in parallel and each returned 3D Euler angles sampled at 60 Hz. The whole body skeletal proportions of each subject (gathered as described in 4.2) combined with the rotational information from the Euler angle data were used to calculate the positional triplets of 26 anatomical points in 3D Cartesian space. This was done using the MoCap toolbox for Matlab [89]. Four wireless sEMG adhesive probes (BTS FREEEMG 300) were attached to the skin (Fig. 2b). Two probes were placed on the upper fibres of the trapezius muscles orientated along the alignment of the fibres of the muscle bilaterally. Two further probes were placed on the lumbar paraspinal muscles approximately at the lumbar 4/5 level bilaterally. The skin contact area was initially cleaned using isopropyl alcohol prior to attachment. Two disposable 24 mm silver/silver chloride electrodes containing integrated adhesive and conductive gel were snapped onto each sensor. The data was recorded at 1 kHz.

4.3.3 Synchronisation

The four recording systems (cameras, audio, motion capture and sEMG) were controlled by a single triggering script which starts the four systems in sequence. The start and end timestamp of each recording were recorded based on a common clock.

The cameras were synchronized between each other by a trigger signal that was sent by a master camera. This trigger signal was recorded as an additional audio channel, it provided further synchronisation between the cameras and the audio. Moreover, the motion capture system provides an external trigger signal which was also recorded as an audio channel. The sEMG system started with the first camera trigger; hence the synchronization between video and sEMG is given. This information is sufficient to align all modalities post recording with an extremely low error margin; the resulting audio-visual synchronization error is bounded to 25 μs. Specific details about this synchronization procedure can be found in [90].

5 Labelling

Labeling naturalistic data is a complex and challenging process, especially when current coding systems are not well established. In this section we describe the labeling process used for this dataset, the rationale behind it and discuss the issues that this raises through analysis. We describe two separate rater-based labeling procedures for: (i) pain expressions from face videos and (ii) pain related movement behaviours (Table 2) from videos with a full body perspective. Also, concurrency relationships between pain expression of the face and prevalent movement behaviours

TABLE 2
Behaviour Label Definitions

Type	Definition
Guarding or stiffness	Stiff, interrupted or rigid movement. It cannot occur while motionless
Hesitation	Stopping part way through a continuous movement with the movement appearing broken into stages
Bracing or support	Position in which a limb supports and maintains an abnormal distribution of weight during a movement which could be done without support.
Abrupt action	Any sudden movement extraneous to the intended motion; not a pause as in hesitation.
Limping	Asymmetric cadence, stride, timing and inequality of weight-bearing during movements.
Rubbing or stimulating	Massaging touching an affected body part with another body part, or shaking hands or legs.

are analysed. Acoustic information was not used in the labeling procedure due to the sparsity of utterances; similarly the sEMG information was also not used directly in labeling due to the difficulty in visually interpreting the waveforms and the fact that physiotherapists are not used to evaluating emotional state on the basis of this signal.

5.1 Pain Expression in the Face

The facial expressions of pain in the CLBP group were continuously labelled by eight independent naïve raters. The videos of four participants were not included in this procedure due to non consent for video release or synchronization error. The raters (five female and three male) were 22 to 30 years old and have no particular experience in rating pain. Naïve raters were used for facial expressions of pain to maximize the number of ratings (as FACS was not used). However, in order to familiarize the raters with pain expressions and the rating procedure, they were instructed to rate the UNBC-McMaster database [52] as a first step.

Once the training had been completed, the raters visually inspected the *EmoPain* videos showing a simultaneous dual view from two cameras: the central camera #4 and the camera pointing up from below camera #8 (see Fig. 1). The camera #8 footage was included as some of the exercise involved a forward flexion motion where only a camera pointing up from below would capture the face during these motions. Each video contained the entirety of one unsegmented trial (described in Section 4.2), the durations of which are the trial lengths in actual time, ranging from 3 to 6.5 minutes with an average of 4.6 minutes.

Each video was loaded into our self-developed annotation tool that uses a gaming joystick as an input device. To provide as natural a setting as possible, play back was done at real-time with 29 fps. The annotators were instructed to move the joystick according to their personal perception of pain. The neutral joystick position describes no pain and the maximum forward displacement represents the highest possible pain level. The current annotated pain level was visually reported as a bar on the side of the video in real time in order to give the annotators immediate feedback and thus locate the current pain level between no pain and the maximum possible level. This provides multiple ratings per trial from each rater, an example can be seen in Fig. 3. Each sequence contains continuous values between 0 and 1, where 0 represents the neutral position and 1 the maximum position of the joystick.

The rating procedure differs from [52], where pain is labelled by determining the discrete intensity labels of a pre-defined pain-related set of action units (as defined by FACS [38]), and then calculating pain according to these labels as indicated by [53] resulting in a 16 level discrete pain scale. In contrast to that, we directly measured pain by observer ratings which lead to a true continuous pain scale. To ascertain an accurate agreement level among the raters, we first accounted for the differences in rating tendency. Due to the real time nature of the rating tool, the first factor considered was the variation in reaction times. To this end, each rater initially underwent a latency test to measure reaction time before rating each sequence. The reaction test consisted of a cross-hair target on the screen that is displaced after a random time. For latency, we measured the time between the displacement event and the reaction of the user. This latency regarding a simple visual stimulus is assumed to be a lower bound of the real latency, which additionally includes a complex interpretation of affective states. The measured latency value was subtracted from the rating times of the corresponding sequence. The ratings were then smoothed with a moving average window of 0.05 sec, which is the minimum human reaction time regarding facial expressions [91].

Since pain expression occurs rarely relative to the whole length of the trial, we first identified the pain events for each rater. We define 'pain event' as the period where the rater moved the joystick from the neutral position. Since the joystick is analog, the output signal of the neutral position is

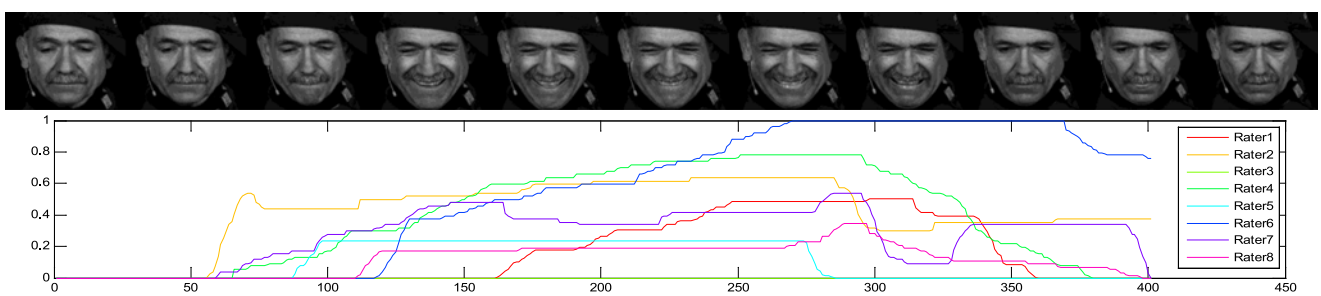


Fig. 3. Cropped video frames from Camera 4 showing an example grimace (above) with all eight temporally concurrent observer's ratings for pain (below). Vertical axis showing the rating and horizontal axis showing the time index.

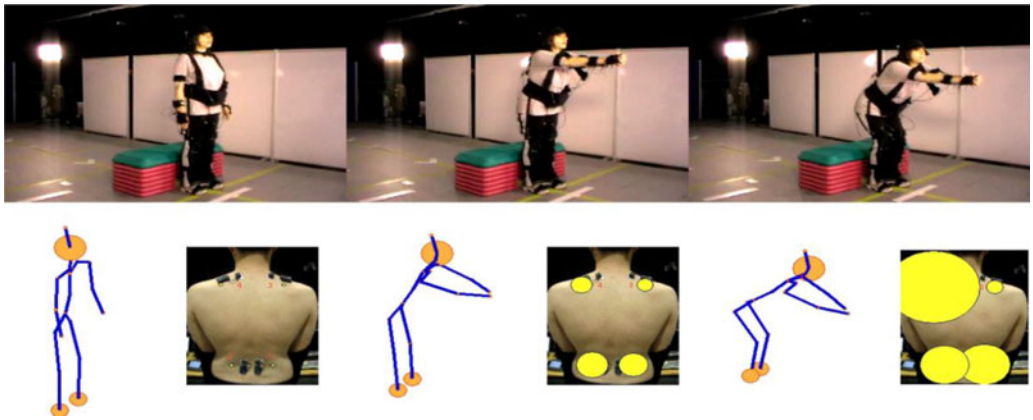


Fig. 4. Example of protective behaviour: the top row shows three frames from a CLBP patient undergoing a reaching forward exercise. The bottom row shows the concurrent motion captured avatar (left) and back muscle activity (right) visualized as circles with radii corresponding to the rectified sEMG amplitude. This instance of reaching forward was labelled as guarded by two raters and hesitative by a third. This participant executes a high knee bend and backward shift of the pelvis during the reaching phase as a compensation strategy to alleviate a perceived strain on the back.

not exactly 0, thus making it necessary to introduce a low threshold for detecting the pain events. We chose 0.02 as threshold which corresponds to 2 percent of the overall joystick range. The agreement for detecting pain on the full data set is low, as the Krippendorf's α score of 0.08 shows. To get a more reliable ground-truth for automatic recognition, frames where only 1 or 2 raters detected pain were discarded to clean out dubious examples. Per subject, we labelled on average 6.72 percent as pain (with a standard deviation of 4.69 percent) and 50.28 percent as no-pain (with a standard deviation of 18.50 percent) and the resulting Krippendorf's α and Fleiss' κ scores are both 0.393, which indicates a fair agreement [92].

This binary pain/no pain event ground-truth has been used for the experiments in Section 6. The average Intra-Class Correlation Coefficient ICC(C,1) [93] on the unprocessed continuous ratings is low at 0.102, which can be due to different reaction times of the raters and misinterpretation of non-pain related facial expressions. In order to get a reliable continuous ground-truth, we apply Dynamic Probabilistic CCA [94], which warps the annotations into a highly correlated space based on the assumption that the observations have been generated from a hidden shared space, which corresponds in our case to the true pain ground-truth. This procedure leads to new ratings with the ICC of 0.741.

5.2 Pain Related Body Movement Behaviour

Compared to facial expressions, definitive movement based pain-related behaviours are not as well established or as easily recognizable when observed. Consequently, a coding framework was determined through an iterative process by four physiotherapists, one psychologist with longstanding expertise in CLBP rehabilitation and one psychologist with expertise with clinical populations. Experts rather than naïve labelers were used for the labeling of the body behaviour given the difficulty of this task and the knowledge this requires.

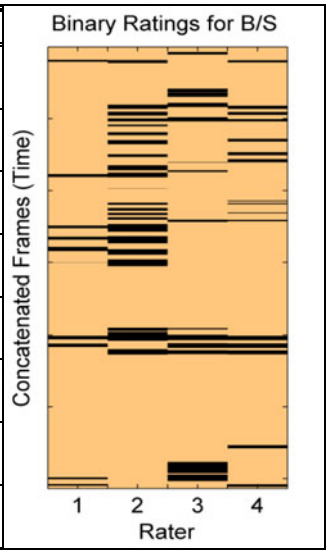
An initial set of nine behaviour categories was created based on the wide ranging Keefe & Block framework [12], the work of Romano et al. [95] on real time behaviour coding and from discussions among our experts. From this set, each expert initially viewed a small subset of videos independently. They then viewed the videos together and conferred to discuss differences in systematic rating tendency and uncertainty. After consensus the final six categories were defined (Table 2). Some categories were paired into one (*guarding/stiffness*, *bracing/support* and *rubbing/stimulating*). Three of the six experts: two physiotherapists and one psychologist, along with a further psychologist with exposure to clinical populations rated each trial by viewing footage from the overview camera #1 which contains a wide angle view of the whole setting (see Fig. 4), with a total of 35 trials. Using standard AVI viewing software with standard playback controls, each rater was asked to temporally segment all episodes where he or she deemed any of the *protective* behaviours (Table 2) to have occurred. A default null choice was labelled if no behaviour was observed. The raters were not given any forced choices and all labels could be assigned simultaneously.

The raters' temporal segmentations are considered as binary sequences with framewise values of 1 if there is the presence of a particular behaviour and 0 if there is an absence. We analyze the rater agreements within exercise specific subsets (the full trials were independently segmented according to exercise type by visual inspection from camera #1 by one experimenter with a biomechanics background). Collecting all segments with likewise exercises yielded the exercise specific subsets.

We calculated the simple matching coefficient (SMC) which is the proportion of correctly matching frames between two sequences; the means of the pairwise calculations are given in Table 3. Since SMC can be biased towards high values if there is a high prevalence of zeros for sparsely rated labels, we also calculated the Krippendorf's α for four raters. (see Table 3). Though the α values are low it should be noted that both metrics do not account for temporal proximity between positively labeled frames and only consider exact temporal matches. It can be seen in the example visualization in Table 3 (right) that many positive labels are in close temporal proximity due to the subjectivity of visual inspection. Therefore, the scores reported in Table 3 can be considered to be the strictest lower bound with no temporal tolerance afforded. The blank cells in Table 3 are due to subsets where the labels are not relevant (e.g., *limping* within

TABLE 3
Inter Rater Agreements for Each Exercise-Label Combination—Simple Matching Coefficient (Upper), Krippendorf's Alpha (Lower)—with a Visualised Example of the Ratings from One Exercise Subset (Right Figure)

	G/S	*B/S*	*A*	*L*
One Leg Stand	0.66 0.02	-	**0.76** 0.13	-
Reach Forward	**0.65** **0.29**	-	-	-
Sit to Stand (I)	0.55 0.06	**0.81** **0.26**	-	-
Stand to Sit (I)	0.52 0	**0.82** **0.18**	-	-
Sit to Stand (NI)	**0.65** 0.12	**0.78** **0.21**	-	-
Stand to Sit (NI)	0.59 0.05	**0.69** **0.26**	-	-
Bend Down	**0.56** 0.11	-	-	-
Walking	0.58 -0.1	-	-	0.58 0.04



Labels: G/S - guarding or stiffness, B/S- bracing or support, A – abrupt motion, L – limping, (I) – instructed, (NI) – not instructed. Blank cells represent combinations where too few or no labels were rated. The example visualization (right) shows the ratings for the Stand to Sit (I) subset for the B/S label at a frame rate of 25Hz with Krippendorf's Alpha = 0.18 and mean Simple Matching Coefficient =0.82.

the *reach forward* subset) or where there are too few positive labels. There were also few positive labels for *hesitation* and *rubbing or stimulation* in all subsets and also within the *sitting still* and *standing still* subsets for all labels.

5.3 Relation between Pain-Related Face and Body Expressions

Given the different roles of facial expressions as a *communicative* modality and body behaviours as both *communicative* and *protective* we investigated to what extent the two label sets were related. Work in the clinical literature has mainly looked at the relationship between self-reported ratings of pain in each of the modalities but not the concomitance and temporal relationship between them [96], [97]. Given that the face pain labels existed across the whole of the trials and that there are high levels of sparsity among the separate behaviour labels, we combined all of the six behaviour labels into one. For simplicity, we call this combined label 'protective'.

We compared the four expert raters' classifications of any protective behaviour over the full length of the trials with the binarised face label (see Section 5.1). Two types of measures were computed. First, four separate Krippendorf's α scores were calculated for each rater over all frames yielding: 0.13, 0.02, 0.09 and -0.01 respectively showing a low correlation between the two sets of labels. The low α scores could be expected given the different functional roles, this is supported by literature showing a weak relationship between protective behaviour and pain intensity [79]. However, it should also be noted that the low correlations are in part due to the sparsity of pain-related expressions and behaviour over the whole trial (73.5 percent of frames do not show either pain expressions nor pain behaviour). In fact, the total number of frames in the dataset is 585,487, of which only 127,567 (11.4 percent of the total) frames have been labelled as protective behaviour and only 50,071 (8.6 percent of the total) as pain expressions (see Fig. 5 left).

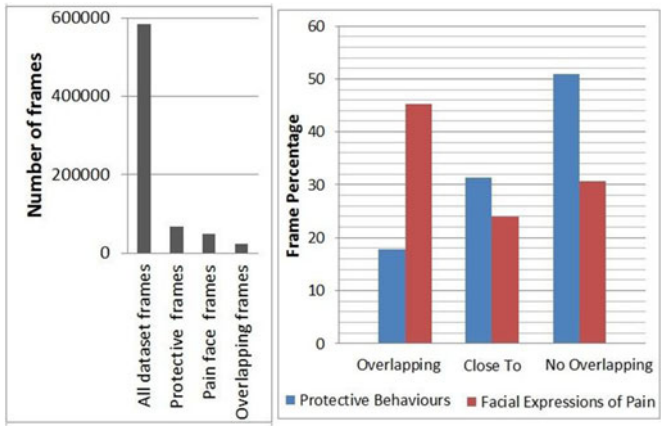


Fig. 5. (Left) distribution of frame type over the entire dataset. (Right) proportions of protective behaviour and facial expressions of pain that directly overlap each other, are within overlapping or close expressive segments or occur with no relation to each other.

Hence we computed a second measure. We counted the number of frames that had been labelled as both facial expression of pain and as protective behaviour. In the case of protective behaviour we consider frames that have been labelled by at least two raters as part of a protective behaviour to be on the conservative side. The results (Fig. 5 right, *'overlapping'* bins) show that only 17.83 percent of protective frames had been also labelled as pain expression. Instead, 45.3 percent of frames with facial expression of pain were also labelled as indicating protective behaviour. This second measure provides more insight as it shows that more facial expressions appear in concomitance with protective behaviour than vice-versa. We further explore the relationship by analyzing the overlap between episodes of facial pain expression and body behaviours, as their onset and offset could be misaligned. We define an episode as an interval of expressive frames with no temporal discontinuity. When an intersection between a face expression and protective behaviour episodes (i.e., a sequence of frames that has been labelled as both pain expression and protective behaviour) is found we counted the number of frames within the union of both episodes excluding the frames that have both labels. Fig. 5 shows that 31.3 percent of protective frames are temporally connected ('close-to' in Fig. 5 right graph) to facial expression episodes but not directly concurrent. Similarly, 23.9 percent of frames indicating a facial expression of pain were temporally connected to but not overlapping with protective episodes. Finally, this leaves 50.9 percent of protective frames and 30.8 percent of frames with facial expression of pain that were either not directly overlapping or temporally connected to each other.

These results show that both protective behaviour and facial expressions of pain frequently occurs independently of each other. However, this is more true for the protective behaviour (~60 percent of the time), whilst ~70 percent of facial expressions do occur in connection with a protective behaviour. Given the increase in percentages resulting by

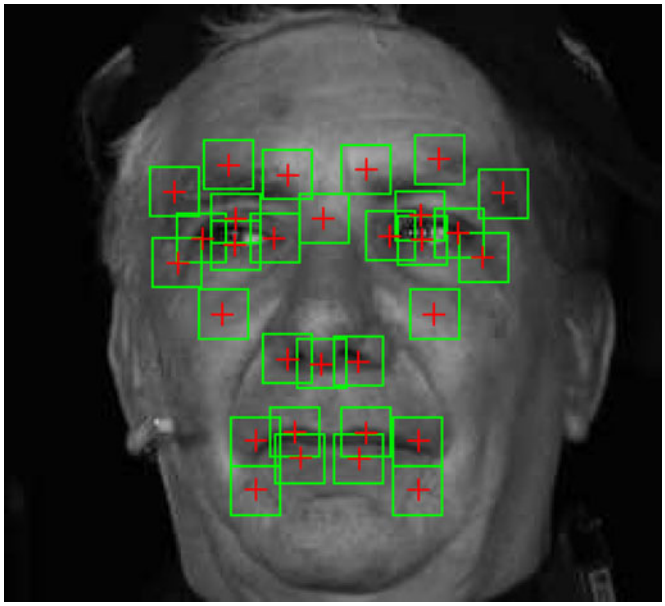


Fig. 6. Example of a normalized face image with highlighted regions from which the features were extracted.

exploring the frames related to but not overlapping with both modalities labelling, we explored the videos to understand why this may occur. Visual inspection of the videos (camera #1) shows that whilst the protective behaviour may start very early in the execution of the exercise, facial expression may occur in specific moment of it, possibly indicating points of the exercise that are perceived as more threating, or increased pain due to the protective behaviour. Facial expressions of pain do also appear after the end of the exercise possibly due to increased anxiety and increased pain; this could also be triggered by the presence of the instrcuctor. An in depth analysis of this is outside of the focus of this paper, however this results point to the richness of this dataset and the possibility that it offers not only to the affective computing community but also to the pain research community to better understand the relation between movement, exercise and pain experience.

6 Recognition Experiments

In this section we provide preliminary experimentations to investigate the possibility of automatically recognising the facial expression of pain and pain related body movement behaviours. We focus only on the CLBP participants and the labels discussed in Sections 5.1 and 5.2 (for initial comparisons with control participants' behaviour see [80]). We aim to identify some of the challenges that these two expressive modalities raise as discussed in each subsection in terms of data and in terms of the labelling issues. Other facets of the *'EmoPain'* dataset such as control participants' data, profiles, self ratings (Table 1) and acoustic modality are left for future work.

6.1 Facial Expressions of Pain

In this experiment we aimed to establish baseline recognition scores for the detection of pain versus no pain events as described in 5.1. To this end, we use standard front view imagery over the whole dataset and compare three established feature sets and a standard binary classifier. This will reveal how standard methods perform for our highly unconstrained and naturalistic imagery and sets a foundation for further directions.

The model will aim to detect frames where humans have rated a non-zero pain intensity. Facial point tracking as proposed in [98] was applied to the video sequences acquired by front view camera #4 (see Fig. 1). This yielded the tracked positions of 49 inner facial landmarks. However, due to the tracking method's applicability being dependent on frontal views, failure was likely to occur when head poses exceeded approximately 30 degrees of out-of-plane rotation. We manually removed the frames where at least half of the point locations were wrongly assigned and thus frames with minor errors remained in the data.

A mean shape was computed by registering all tracked shapes from all subjects using a non-reflective similarity transformation. This allowed for each image to be registered by first registering the facial landmarks to the mean shape, and then applying the resulting non-reflective similarity transformation to the image. This process assured that all face images were aligned accounting for head motions such as translations, scaling and in-plane rotations.

Three sets of standard features were tested in this experiment. The first set being the location of the tracked landmark points; in order to account for anthropomorphic differences between the subjects further normalization was done by subtracting the mean shape of each individual. The next two feature sets are based on local appearance features and were extracted from patches with a radius of 10 pixels centred on 30 fiducial landmarks depicted in Fig. 6. We calculated the uniform Local Binary Patterns [99] with 8 circular points and a radius of 2 pixels as a feature descriptor. This resulted in a 59-dimensional feature vector per patch. These were concatenated into a single vector to form the final feature vector. For the final feature set we applied the Discrete Cosine Transform [100] to the same patches and use the first 59 coefficients as features for each patch. Again, the final feature vector is the concatenation of all patches. LBPs and DCTs have been proven useful for facial expression recognition in general [101] and pain recognition specifically [62], [102], [103]. We conducted experiments on the 3 different sets of features explained above: (1) Points, (2) LBP and (3) DCT.

In order to detect the binarised pain/no pain states, we used the extracted features to train a linear support vector machine [104]. For evaluation, we employed a subject-independent cross-validation procedure by repeatedly leaving one subject out for testing and using the remaining subjects for training. Using the full length of 34 unsegmented trials (described in Section 4.2) from 17 patients with pain ground-truth leads to a total of 317,352 frames, where each frame is a training instance. At each cross-validation fold, we sub-sampled 10,000 frames for training, where 33.3 percent contained pain and 66.6 percent no-pain. We measured the performance using AUC and F1 scores. To obtain a single overall result, we calculated a weighted average over all cross-validation folds, where the weight is the number of positive (pain) examples contained in the respective fold. The SVM regularisation parameter was optimized using a grid-search.

TABLE 4
Pain/No-Pain Classification Results for Different Features, Measured by the Area under the ROC Curve and the F1-Score

	Points	*LBP*	*DCT*
AUC (std)	.658 (.170)	.645 (.106)	.628 (.151)
F1(std)	.446 (.189)	.386 (.234)	.395 (.210)

Overall the use of points yielded the best performance (Table 4), while LBP and DCT perform similarly to each other (LBP better with the AUC score and DCT better with the F1 score). These results are different from [62], where the points are outperformed by LBP and DCT. Although the appearance descriptors are able to capture more subtle changes in the face (e.g., wrinkles around the nose or eyes). We obtained the best performance from the coarser point descriptor. This indicates that there is a high appearance variation in the data which is not caused by pain (e.g., due to facial expressions not conveying pain, or head-pose and subsequent lighting changes), therefore challenging the learned appearance classifier. Although the extreme head poses that cause tracking failures were removed, out-of-plane rotations still cause a high variation within the landmarks and face appearance. Head-rotations and whole body movements also cause the lighting changes, since the face is illuminated from a different angle.

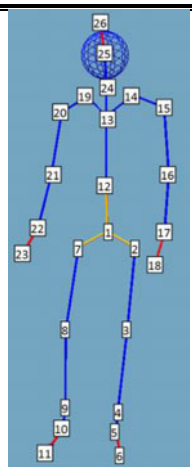
The results in Table 4 are lower in comparison with the results on the UNBC-McMaster data [52] or the Hi4D-ADSIP data [66]. However, this can be explained by the differences between the natures of the datasets: our database consists of subjects who suffer from CP and thus many of their expressions are subdued due to the long time exposure. Additionally, our data contains various other facial expressions (mainly smiles and speech), which further complicate the recognition tasks. The results in [52] and [61] are obtained on data which solely contains acute pain expressions in a more constraint scenario (no movement). In [66] acted expressions were used, which are in general easier to detect due to the exaggeration in comparison to naturalistic data.

6.2 Body Movement Based Behaviour Recognition

In this experiment we aimed to recognise the degree of body related behaviours within exercise specific subsets (Tables 2 and 3) using motion capture data and sEMG information. Since much less is known about movement based feature relevance, the main target of this analysis was to explore relative feature importance with regard to body part location and whether it was postural or velocity based information. For each of the behaviour types (Table 2), we implemented a regression model to predict the extent of occurrence for each behaviour type within exercise specific subsets where there exists a significant number of rated instances. The grouping by exercise was done to homogenize the movement context within which we aimed to recognise the presence of a behaviour (e.g., extent of guardedness during walking) rather than recognising the exercise itself (e.g., walking versus not walking). Moreover, where there is a significant number of instances of a motion being done without direct instruction we also consider them as a separate non instructed grouping. These are motions done entirely due to the participants' own accord, not to be confused with the self determined start times in the *normal* trials as described in Section 4.2.

TABLE 5
Description of Feature Angles

#	Vector Pairs	Angle Name
1	[26-1] [1-4]	Full body flexion (L)
2	[26-1] [1-9]	Full body flexion (R)
3	[12-1] [1-3]	Inner flexion (L)
4	[12-1][1-8]	Inner flexion (R)
5	[2-3] [3-4]	Knee (L)
6	[7-8] [8-9]	Knee (R)
7	[15-16] [16-17]	Elbow (L)
8	[20-21] [21-22]	Elbow (R)
9	[24-14] [14-15]	Shoulder (L)
10	[24-19] [19-20]	Shoulder (R)
11	[16-14] [14-2]	Lateral Bend (L)
12	[21-19] [19-7]	Lateral Bend (R)
13	[12-24] [24-26]	Neck



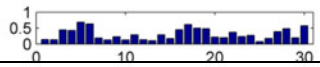
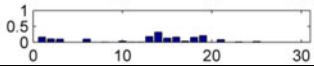
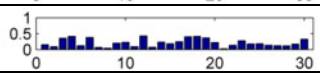
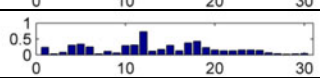
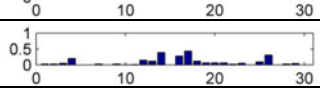
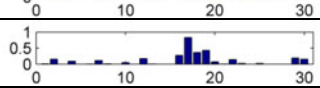
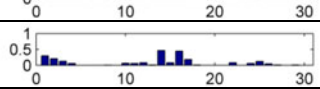
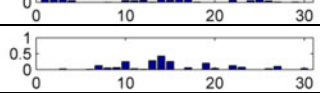
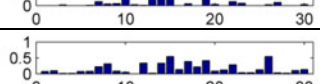
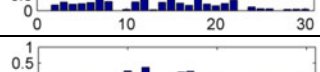
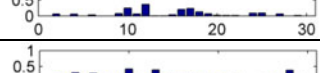
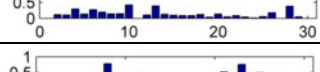
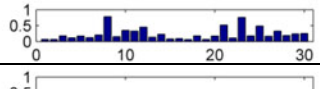
The inner angles between vectors pairs connecting various anatomical nodes are calculated, e.g. left knee is calculated by the angle between the two vectors: 2 to 3 and 3 to 4. The ranges are used for features 1-13, the mean of the corresponding 'energy' values are used for features 14-26.

Since each rater temporally segmented the overview video (camera #1) of each trial according to when they deemed each behaviour to have happened, we calculated the proportion of all positively labelled frames relative to the duration of each exercise instance. The target value to be predicted is the mean of these fractions from all four raters, (for brevity we will denote this value as $\tau \in [0, 1]$). Such a regression task has the advantage of preserving a measure for the overall 'extent of occurrence' of each behaviour since we retain the durational information. It also reflects a level of ambiguity of a label within an exercise instance; the low agreements shown in Table 3 indicate high levels of ambiguity even among experts and thus would not underpin a classification task.

In order to get an exploratory insight into feature importance we used three categories of features drawn from all parts of the body: (i) postural information described by the ranges of inner angles in 3D space at 13 angles (see Table 5), informed by [71], [82], (ii) velocity based information which we will refer to as 'energy' calculated from the sum square of the angular velocities at each of the 13 angles [105] and (iii) the muscle activity levels using the upper envelope of the rectified signal from each of the four sEMG probes (Fig. 2b). This leads to a 30-dimensional feature vector independent of skeletal proportions.

We used ensembles of decision trees: Random Forests (RF) [106] for prediction. Recent works by Griffin et al. [107], [108] compared a range of standard supervised learning models and showed RF to outperform others in recognising affect related styles in whole body motion. In addition, trained RFs can be further analysed to quantify the relative importance of each feature, one of our main objective in this analysis. Hence, as this is an exploratory study, no prior assumptions on feature relevance were made. All the models used the same feature set which contains whole body posture and velocity information as well as sEMG. The number of trees per ensemble and the number of features to be used for each tree was optimised using

TABLE 6
Evaluation of Random Forest Outputs Using Leave One Subject Out Validation and Relative Feature Importance

Label	Exercise	Instances	Correlation ↑	MSE ↓	Mean Feature Importance Histogram ↑
Guarding/ Stiffness	*One leg stand*	152	0.502	0.028	
Guarding/ Stiffness	*Reach for-ward*	39	0.337	0.020	
Guarding/ Stiffness	*Sit to stand I*	105	0.465	0.034	
Guarding/ Stiffness	*Stand to sit I*	84	0.365	0.032	
Guarding/ Stiffness	*Sit to stand NI*	41	0.508	0.028	
Guarding/ Stiffness	*Stand to sit NI*	72	0.471	0.028	
Guarding/ Stiffness	*Bend*	19	0.714	0.019	
Guarding/ Stiffness	*Walk*	36	0.164	0.023	
Bracing/ Support	*Sit to stand I*	105	0.609	0.019	
Bracing/ Support	*Stand to sit I*	84	0.496	0.016	
Bracing/ Support	*Sit to stand NI*	41	0.207	0.023	
Bracing/ Support	*Stand to sit NI*	72	0.271	0.041	
Abrupt motion	*One leg stand*	152	0.581	0.023	
Limping	*Walk*	36	0.114	0.038	

All exercise-label subsets with a significant number of labels are shown including the number of instances in each subset, the suffix I = instructed exercise and NI = not instructed. The Pearson's r correlation and mean-squared error (MSE) values between the models' output and the target values from person independent validation are shown. The '↑' arrow indicates higher values as better and '↓' indicates the reverse. The final column contains the averaged importance estimations of each variable using the out of bag instances during the training process. The feature indices are: 1-13 for angle ranges, 14-26 for the means of angle energies and 27-30 for the means of rectified sEMG (for angle definitions see Table 5).

a grid search. Each trained ensemble was evaluated using person independent cross validation. Table 6 shows the evaluation scores for the exercise/label combinations where a sufficient number of labels existed for learning (corresponding rater agreements levels can be seen in Table 3). Table 6 shows correlations > 0.5 in the bend, one leg stand and sit to stand (not instructed) subsets for *guarding/stiffness*. It is worth noting that for *guarding/stiffness* the RF models perform slightly better for the not instructed instances of sit to stand and stand to sit compared to their respective instructed subsets. This is the opposite case for *bracing/support* where there is a much lower correlation and higher mean-squared error (MSE) for the non instructed subsets. This could be attributed to the greater abstractness of *guarding/stiffness* compared to *bracing/support* which is defined by a more specific physical event which is repeated in a consistent way in the instructed subset.

In general the RF models tend to output conservative estimates of τ for all of the exercise-label combinations in Table 6. Also, the high correlation scores show that it is not the case that the low MSE values are simply attributed to consistent predictions close to the mean of τ as this would have also returned low correlation scores. The histograms in Table 6 reveal some notable feature importance estimates. *Guarding/ stiffness* in the bend exercise is well recognisable with high correlation and low MSE, also shown by initial results in [82].

This histogram shows high feature importance groupings for body flexion features (1-4) and their corresponding energies (14-17). There is also a difference between the instructed and not instructed sit to stand and stand to sit exercises for *guarding/stiffness*. Overall, there are uniform levels of feature importance between the postural features (1-13) and the velocity based energy features (13-26) for the instructed exercises. In comparison, only the velocity based group is important for the non instructed instances. Such an effect is not as succinct for *bracing/support*, where the non instructed distribution seems to reflect the instructed counterpart but at a lower magnitude. For the one leg stand, all knee related ranges (3-6) and energies (16-19) are important for *guarding/stiffness*, as are the sEMG features (27-30). However, this is not the case for *abrupt action*, where lateral, shoulder and elbow features show importance.

7 Discussion and Conclusions

This paper lays the groundwork for the development of much needed affect sensitive systems for CLBP rehabilitation. Since this research theme is little studied, we endeavoured to collect a multifaceted dataset with specifically selected participants and sensing modalities to acquire naturalistic pain related behaviour based on well established behavioural psychology frameworks [12], [20]. These first results provide a foundation for further specific investigations on the provided labels or on further label sets that can be generated from additional rating.

7.1 Face Pain Classification

The scores set in Table 4 can be taken as baselines for detecting face pain expressions during unconstrained exercise rehabilitation sessions. Three types of features (Points, LBP and DCT) and one prediction method (SVM) using data from the whole trial were used without distinction between exercise types. The results in Table 4 demonstrate a fair recognition capacity with the use of point locations as the most effective. However, there remain further options yet to be explored; experimentation with inference over temporal data seems promising: one option is temporal appearance features such as, Bags of Visual Words and LBP in three orthogonal planes (LBP-TOP), since they provide higher robustness than their static counterparts [109]. Another option are temporal models, such as Context Sensitive Conditional Ordinal Random Fields [110], which are able to infer facial expressions over sequences rather than single frames. These recognition scores reported here are not as high as the current state of the art for the UNBC-McMaster dataset. Part of the cause could be attributed to the unconstrained nature of the *EmoPain* data in terms of the various movements, presence of speaking and other non pain expressions. Moreover, our focus was on pain versus no pain with a binary classifier after thresholding the continuous ratings; this leaves open the potential use of regression models as well as models specific to each exercise subset.

With regard to the labels, the use of the FACS coding system is an alternative to our labelling approach. FACS provides an objective way to describe facial activity, which may lead to higher inter-rater reliability scores and a better temporal localisation of the facial activity.

However, such a system does not offer an interpretation of the facial expression if non-pain expressions are present, i.e. there is no univocal mapping from facial action unit (AU) labels to pain labels that can distinguish between different expressions. For example, some expressions of pain and the expression of happiness often share the same AU activation labels, and human interpretation of the event would still be necessary to produce reliable labels. Our approach bypasses the AU labelling and directly uses the human interpretation of the events. Further benefits in using the real time joystick approach is obtaining continuous labels and a much faster coding procedure, enabling the annotation of large amounts of data and the potential of a large number of raters given the expediency of the method.

7.2 Protective Behaviour Recognition

The exploratory analyses on the automatic detection of pain-related behavior raises a number of suggestions for the development of recognition systems:

1) The context of the exercise undertaken in relation to the behaviour of interest is an important factor. The relative importance estimates shown in Table 6 and detailed in Section 6.2 can inform new feature selections for exercise specific models. Future systems should exploit contextual information about expected movement, if this is not known a layer of activity recognition could easily be included prior to the behaviour recognition.
2) A notable result is the differences between the instructed and non-instructed subsets. For *guarding/stiffness*, only velocity based features were important for the non-instructed set, whereas both postural and velocity based features were important for the instructed set. This may be due to the irregularity of the motion strategies adopted in the non instructed cases. Interestingly, there is a slight trend toward the non-instructed instances of *guarding/stiffness* being easier to recognise; it is the opposite for the *bracing/support* label where the instructed instances had significantly higher correlation scores (Table 6). This raises a very interesting avenue for further investigation and may relate to the known differences in the recognition of acted versus natural movement.
3) The low agreement scoring in Table 3 re-iterates the challenge in the rating of labels defined by function rather than form, especially within unconstrained data. Given their physical foundation, form based descriptions tend to be more reliable than functional ones [111]. However, irrespective of the level of description reliability decreases with increasing complexity of the observed behavior. One possibility is to hone the definitions in Table 2 for a given type of exercise and whether it occurred due to an instruction. This could be done by stipulating important body parts within the definitions informed by the histograms in Table 6, which give an insight into important features in natural unconstrained motion. In addition, the labeling process could benefit by showing concurrent video and sEMG signals through some form of graphical visualization. This would probably require some form of training as this is not current practice during rehabilitation. Another direction for labeling that is emerging and worth considering is the use of crowdsourcing with a training phase; recent work by Park et al. [112] demonstrated that the inclusion of a training stage which naive raters must undertake before rating yielded consistent increases in Krippendorff's α scores. This could be supported by a recent set of images produced by Walsh et al. consisting of acted postures which yielded high observer agreement in the perception of pain [113].
4) The low concurrence between facial pain expressions and *guarding/stiffness* behaviour suggests face and

body expression may often occur separately, this supports previous findings that may point to independence [96]. This might suggest that care needs to be place when fusing these modalities and that the fusion should occur at later stages.

In addition, the results in Fig. 5 have shown the existence of a possible temporal relationship that may provide interesting insights in terms of pain experience. An in depth analysis of these relationships with pain experts and people with CP may further inform the timing and type of support that rehabilitation systems should provide. We hope that the opening of the dataset to the overall research community (including clinical research and HCI research) will benefit and foster interdisciplinary work in this area for more effective rehabilitation support.

8 *EmoPain* Dataset Availability

The dataset will be made available to the research community via a web-accessible interface linked from 'www.emopain.ac.uk'. The first release will contain eight continuous facial pain ratings and the temporal annotations for movement based pain behaviours from four raters, and the approximate onset and end timings of each exercise will also be provided for the patient set.

Within each trial the following four fully synchronized data streams will be included: high resolution videos of the face from seven viewpoints, head-mounted and a room ambient acoustic signals, the Cartesian co-ordinate triplets of 26 anatomical positions over the whole body and rectified upper envelope of the electromyographic signals from the upper and lower back. The processed features used in the experiments (Section 6) will also be available upon request. Finally, as mentioned in Section 6, the other facets of the dataset not investigated in this paper such as control data and questionnaire outcomes (Table 1) are also supplied where written consent was given.

The dataset will be maintained and updated as new data and labels become available; in particular further body behaviour labels are being updated. We are also gathering further data using sensing technologies that are more readily deployable in real world settings, e.g., Kinect and wearable devices [17], [114].

Acknowledgments

This work was funded by the EPSRC Emotion & Pain Project (EP/H017178/1, EP/H017194/1, EP/H016988/1). The authors would like to thank Derrick Boampong and the technical staff at the Pedestrian Accessibility Movement Environment Laboratory (PAMELA) at University College London. They also thank all participants for giving their time and support to the project. Nadia Bianchi-Berthouze is the corresponding author.

References

[1] M. Papastergiou, "Exploring the potential of computer and video games for health and physical education: A literature review," *Compu. Educ.*, vol. 53, no. 3, pp. 603–622, 2009.

[2] S. Patel, H. Park, P. Bonato, L. Chan, and M. Rogers, "A review of wearable sensors and systems with application in rehabilitation," *J. Neuroeng. Rehab.*, vol. 9, no. 12, 2012.

[3] D. Gonzalez-Ortega, F. J. Diaz-Pernaz, M. Martinez-Zarzuela, and M. Anton-Rodriguez, "A Kinect based system for cognitive rehabilitation exercises monitoring," *Comput. Methods Programs Biomed.*, vol. 113, pp. 620–631, 2014.

[4] M. Leeuw, M. E. J. B. Goossens, S. J. Linton, G. Crombez, K. Boersma, and J. W. S. Vlaeyen, "The fear-avoidance model of musculoskeletal pain: Current state of scientific evidence," *J Behav. Med.*, vol. 30, no. 1, pp. 77–94, 2007.

[5] D. C. Turk and T. E. Rudy, "IASP taxonomy of chronic pain syndromes: Preliminary assessment of reliability," *Pain*, vol. 30, no. 2, pp. 177–189, 1987.

[6] I. Tracey and M. C. Bushnell, "How neuroimaging studies have challenged us to rethink: Is chronic pain a disease?" *J. Pain* vol. 10, no. 11, pp. 1113–1120, 2009.

[7] M. C. Bushnell, M. Čeko, and L. A. Low, "Cognitive and emotional control of pain and its disruption in chronic pain," *Nature Neurosci. Rev.*, vol. 14, no. 1, pp. 502–51, 2013.

[8] M. O. Martel, T. H. Wideman, and M. J. L. Sullivan, "Patients who display protective pain behaviors are viewed as less likable, less dependable, and less likely to return to work," *Pain*, vol. 153, no. 4, pp. 843–849, 2012.

[9] M. Von Korff, J. Gruman, J Schaefer, S. Curry, and E. H. Wagner "Collaborative management of chronic illness," *Ann. Internal Med.*, vol. 127, pp. 1097–1102, 1997.

[10] G. Crombez, C. Eccleston, S. Van Damme, J. W. S. Vlaeyen, and P. Karoly, "The fear avoidance model of chronic pain: The next generation," *Clin. J. Pain*, vol. 28, pp. 475–483, 2012.

[11] J. W. S. Vlaeyen and S. J. Linton, "Fear-avoidance and its consequences in chronic musculoskeletal pain: A state of the art," *Pain*, vol. 85, no. 3, pp. 317–332, 2000.

[12] F. J. Keefe and A. R. Block, "Development of an observation method for assessing pain behaviour in chronic low back pain patients," *Behaviour Therapy*, vol. 13, no. 4, pp. 363–375, 1982.

[13] D. S. Goldberg and S. J. McGee, "Pain as a global public health priority," *BMC Public Health*, vol. 11, p. 770, 2011.

[14] R. della Volpe, T. Popa, F. Ginanneschi, R. Spidalieri, R. Mazzocchio, and A. Rossi, "Changes in coordination of postural control during dynamic stance in chronic low back pain patients," *Gait Posture*, vol. 24, no. 3, pp. 349–355, 2006.

[15] R. J. Gatchel, Y. B. Peng, M. L. Peters, P. N. Fuchs, and D. C. Turk, "The biopsychosocial approach to chronic pain: Scientific advances and future directions," *Psychological Bull.*, vol. 133, no. 4, pp. 581–624, 2007.

[16] A. C. de C. Williams, C. Eccleston, and S. Morley, "Psychological therapies for the management of chronic pain (excluding headache) in adults," *Cochrane Database Systematic Rev.*, vol. 11, no. 11, 2012.

[17] A. Singh, A. Klapper, J. Jia, A. Fidalgo, A. Tajadura-Jimenez, N. Kanakam, N. Bianchi-Berthouze, and A. Williams, "Motivating people with chronic pain to do physical activity: Opportunities for technology design," in *Proc. ACM Conf. Human Factors Comput. Syst.*, 2014, pp. 2803–2812.

[18] D. C. Turk and A. Okifuji, "Psychological factors in chronic pain: Evolution and revolution," *J. Consult. Clin. Psychol.*, vol. 70, no. 3, pp. 678–690, 2002.

[19] W. Fordyce, *Behavioural Method for Chronic Pain and Illness*. St Louis, MO, USA: Mosby, 1976.

[20] K. M. Prkachin, I. Schultz, J. Berkowitz, E. Hughes, and D. Hunt, "Assessing pain behaviour of low back pain patients in real time: concurrent validity and examiner *sensitivity*," *Behav. Res. Ther.*, vol. 40, pp. 595–607, 2002.

[21] M. J. L. Sullivan, P. Thibault, A. Savard, R. Catchlove, J. Kozey, and W. D. Stanish, "The influence of communication goals and physical demands on different dimensions of pain behaviour," *Pain*, vol. 125, pp. 270–277, 2006

[22] K. M. Prkachin, "Pain behaviour is not unitary," *Behavioral Brain Sci.*, vol. 9, pp. 754–755, 1986.

[23] P. D. Wall, *Pain: The Science of Suffering*. London, U.K.: Weidenfeld & Nicolson, 1999.

[24] N. K. Y. Tang, P. M. Salkovskis, E. Poplavskaya, K. J. Wright, M. Hanna, and J. Hester, "Increased use of safety-seeking behaviours in chronic back pain patients with high health anxiety," *Behav. Res. Ther*, vol. 45, no. 12, pp. 2821–2835, 2007.

[25] P. J. Watson, K. C. Booker, and C. J. Main, "Evidence for the role of psychological factors in abnormal paraspinal activity in patients with chronic low back pain," *J. Musculoskelet. Pain*, vol. 5, no. 4, pp. 41–56, 1997.

[26] M. van der Hulst, M. M. Vollenbroek-Hutton, J. S. Reitman, L. Schaake, K. G. Groothuis-Oudshoorn, and H. J. Hermens, "Back muscle activation patterns in chronic low back pain during walking: a "guarding" hypothesis," *Clin. J. Pain*, vol. 26, no. 1, pp. 30–37, 2010.

[27] M. E. Geisser, A. J. Haig, A. S. Wallbom, and E. A. Wiggert, "Pain-related fear, lumbar flexion, and dynamic EMG among persons with chronic musculoskeletal low back pain," *Clin. J. Pain.* vol. 20, no. 2, pp. 61–69, 2004.

[28] J. S. Thomas and C. R. France, "Pain-related fear is associated with avoidance of spinal motion during recovery from low back pain," *Spine*, vol. 32, no. 16, pp. 460–466. 2007.

[29] J. S. Thomas and C. R. France, "The relationship between pain-related fear and lumbar flexion during natural recovery from low back pain," *Europ. Spine J.*, vol. 17, no. 1, pp. 97–103, 2008.

[30] Z. Trost, C. R. France, M. J. Sullivan, and J. S. Thomas, "Pain-related fear predicts reduced spinal motion following experimental back injury," *Pain*, vol. 153, no. 5, pp. 1015–1021, 2012.

[31] J. Rainville, D. Ahern, L. Phalen, L. A. Childs, and R. Sutherland, "The association of pain with physical activities in chronic low back pain," *Spine*, vol. 17, no. 9, pp. 1060–1064, 1992.

[32] K. Prkachin and K. Craig, "Expressing pain: The communication and interpretation of facial pain signals," *J. Nonverbal Behavior*, vol. 19, no. 4, pp. 191–205, 1995.

[33] T. Hadjistavropoulos and K. D. Craig, *Pain: Psychological Perspectives*. New York, NY, USA: Psychology Press, 2003

[34] A. C. de C. Williams, "Facial expression of pain: An evolutionary account," *Behav. Brain Sci*, vol. 25, pp. 439–488, 2002.

[35] K. M. Prkachin, "Expressing pain by facial expression: Facial expression as nexus," *Pain Res. Manage.* vol. 14, no. 1, pp. 53–58, 2009.

[36] L. LaResche and S. F. Dworkin, "Facial expression accompanying pain," *Soc. Sci. Med.*, vol. 19, no. 12, pp. 1325–1330, 1984.

[37] K. D. Craig and C. J. Patrick, "Facial expression during induced pain," *J. Personality Social Psychol.*, vol. 48, no. 4, pp.1080–1091, 1985.

[38] P. Ekman and W. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto, CA, USA: Consulting Psychologists Press, 1978.

[39] K. M. Prkachin, "The consistency of facial expressions of pain – a comparison across modalities," *Pain*, vol. 51, no. 3, pp. 297–306, 1992.

[40] J. Kappesser and A. C. de C. Williams, "Pain and negative emotions in the face: Judgements by health care professionals," *Pain*, vol. 99, no. 1/2, pp. 197–206, 2002.

[41] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: audio, visual, and spontaneous expressions," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, Jan. 2009.

[42] K. Teske, R. L. Daut, and C. S. Cleeland, "Relationships between nurses' observations and patients' self-reports of pain," *Pain*, vol. 16, no. 3, pp. 289–296, 1983

[43] M. O. Martel, P. Thibault, and M. J. L Sullivan. "The persistence of pain behaviours in patients with chronic back pain is independent of pain and psychological factors," *Pain*, vol. 151, pp. 330–336, 2010.

[44] D. R. Carney, A. J. C. Cuddy, and A. J. Yap, "Power posing: Brief nonverbal displays affect neuroendocrine levels and risk tolerance," *Psychol. Sci.*, vol. 21, pp. 1363–1368, 2010.

[45] N. Bianchi-Berthouze, "Understanding the role of body movement in player engagement," *Human Comput. Interaction*, vol. 28, no. 1, pp. 42–75, 2013.

[46] P. M. Niedenthal, L. W. Barsalou, P. Winkielman, S. Krauth-Gruber, and F. Ric, "Embodiment in attitudes, social perception, and emotion," *Personal. Soc. Psychol. Rev.*, vol. 9, no. 3, pp. 184–211, 2005.

[47] C. E. Ashton-James, D. C. Richardson, A. C. de C. Williams, N. Berthouze, and P. H. Dekker, "The impact of pain behaviors on evaluations of warmth and competence," *Pain*, vol. 155, no. 12, pp. 2656–61, 2015.

[48] S. Felipe, A. Singh, C. Bradley, A. C. de C. Williams, and N. Bianchi-Berthouze, "Roles for personal informatics in chronic pain," in *Proc. Persuasive Health*, 2015, http://eudl.eu/doi/10.4108/icst.pervasivehealth.2015.259501

[49] A. Tajadura-Jiménez, M. Basia, O. Deroy, M. Fairhust, N. Marquardt, and N. Bianchi- Berthouze, "As light as your footsteps: altering walking sounds to change perceived body weight, emotional state and gait," in *Proc. ACM Conf. Human Factors Comput. Syst.*, 2015, pp. 2943–2952,

[50] A. Dhall, R. Goecke, J. Joshi, M. Wagner, and T. Gedeon, "Emotion recognition in the wild challenge 2013," in *Proc. ACM Int. Conf. Multimodal Interaction*, 2013, http://dl.acm.org/citation.cfm?id=2531923

[51] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, "The first facial expression recognition and analysis challenge," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog., Workshop Facial Expression Recog. Anal. Challenge*, 2011, pp. 921–926.

[52] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, and I. Matthews, "Painful data: The UNBC-McMaster shoulder pain expression archive database," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog., Workshop*, 2011, pp. 57–64.

[53] K. Prkachin and P. Solomon, "The structure, reliability and validity of pain expression: The evidence from patients with shoulder pain," *Pain*, vol. 139, pp. 267–274, 2008.

[54] Y. L. Ma, H. M. Paterson, and F. E. Pollick, "A motion capture library for the study of identity, gender, and emotion perception from biological motion," *Behavior Res. Methods*, vol. 38, pp. 143–141, 2006.

[55] A. Kleinsmith, N. Bianchi-Berthouze, and A. Steed, "Automatic recognition of non-acted affective postures," *IEEE Trans Sys., Man Cybern. B*, vol. 41, no. 4, pp. 1027–1038, Aug. 2011.

[56] N. Savva, A. Scarinzi, and N. Bianchi-Berthouze, "Continuous recognition of player's affective body expression as dynamic quality of aesthetic experience," *IEEE Trans. Comput. Intell. AI Games*, vol. 4, no. 3, pp. 199–212, Sep. 2012.

[57] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models method and their applications," *IEEE Trans. Affective Comput.*, vol. 1, no. 1, pp. 18–37, Jan.-Jun. 2010.

[58] M. Monwar, and S. Rezaei, "Pain recognition using artificial neural network," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol.*, 2006, pp. 28–33.

[59] A. B. Ashraf, S. Lucey, J. F. Cohn, T. Chen, Z. Ambadar, K. M. Prkachin, and P. E. Solomon, "The painful face—Pain expression recognition using active appearance models," *Image Vis. Comput.*, vol. 27, pp. 1788–1796, 2009.

[60] Z. Hammal and J. F. Cohn, "Automatic detection of pain intensity," in *Proc. Int. Conf. Multimodal Interaction.* 2012, pp. 47–52.

[61] S. Kaltwang, O. Rudovic, and M. Pantic, "Continuous pain intensity estimation from facial expressions," in *Proc. 8th Int. Symp. Adv. Vis. Comput.*, 2012, vol. 7432, pp. 368–377.

[62] P. Lucey, J. Howlett, J. Cohn, S. Lucey, S. Sridharan, and Z. Amdabar, "Improving pain recognition through better utilisation of temporal information," in *Proc. Int. Conf. Auditory-Visual Speech Process.*, 2008, pp. 167–172.

[63] K. Sikka, A. Dhall, and M. Bartlett, "Weakly supervised pain localization and classification with multiple segment learning," *Image Vis. Comput., Faces Gesture Recog.*, vol. 32, pp. 659–670, 2014.

[64] B. Romera-Paredes, M. S. H. Aung, N. Bianchi-Berthouze, and M. Pontil, "Multilinear multitask learning," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1444–1452.

[65] B. Romera-Paredes, M. S. H. Aung, M. Pontil, P. J. Watson, and N. Bianchi-Berthouze, "Transfer learning to account for idiosyncrasy in face and body expressions," in *Proc. 10th IEEE Int. Conf. Autom. Face Gesture Recog.*, 2013, pp. 1–6.

[66] P. Werner, A. Al-Hamadi, and R. Niese, "Pain recognition and intensity rating based on comparative learning," in *Proc. IEEE Int. Conf. Image Process.*, 2012, pp. 2313–2316.

[67] G. C. Littlewort, M. S. Barlett, and K. Lee, "Automatic coding of facial expressions displayed during posed and genuine pain," *Image Vis. Comput.*, vol. 27, no. 12, pp. 1797–1803, 2009.

[68] M. S. Bartlett, G. C. Littlewort, M. G. Frank, and K. Lee, "Automatic decoding of facial movements reveals deceptive pain expressions," *Current Biol.*, vol. 24, pp. 738–743, 2014.

[69] Z. Chen, R. Ansari, and D. J. Wilkie, "Automated detection of pain from facial expressions: A rule based approach using AAM," *Proc. SPIE*, vol. 8314, p. 8314, 2011.

[70] D. J. Wilkie, "Facial expressions of pain in lung cancer," *Analgesia*, vol. 1, no. 2, pp. 91–99, 1995.

[71] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: A survey," *IEEE Trans. Affective Comput.*, vol. 4, no. 1, pp. 15–33, Jan.–Mar. 2013.

[72] M. Karg, A. A. Samadani, R. Gorbert, K. Kuhnlenz, J. Hoey, and D. Kulic, "Body movements for affective expression: A survey of automatic recognition and generation," *IEEE Trans. Affective Comput.*, vol. 4, no. 4, pp. 341–359, Oct.-Dec. 2013.

[73] A Kapur, A. Kapur, N. Virji-Babul, G. Tzanetakis, and P. F. Dreissen, "Gesture-based affective computing on motion capture data," in *Proc. 1st Int. Conf. Affective Comput. Intell. Interaction*, 2005, pp. 1–7.
[74] H. Gunes and M. Piccardi, "Automatic temporal segment detection and affect recognition from face and body display," *IEEE Trans. Syst., Man Cybern. B*, vol. 39, no. 1, pp. 64–84, Feb. 2009.
[75] J. Joshi, R. Goecke, G. Parker, and M. Breakspear, "Can body expressions contribute to automatic depression analysis?" in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog.*, 2013, pp. 1–7.
[76] A. Kleinsmith and N. Bianchi-Berthouze, "Recognizing affective dimensions form body posture," in *Proc. 2nd Int. Conf. Affective Comput. Intell. Interaction*, 2007, pp. 48–58.
[77] C. L. Hubley-Kozey and M. Vezina, "Differentiating temporal electromyographic waveforms between those with chronic low back pain and healthy controls," *Clin. Biomech.*, vol. 17, pp. 621–629, 2002.
[78] C. Lariviere, D. Gagnon, and P. Loisel, "An application of pattern recognition for the comparison of trunk muscles EMG waveforms between subjects with and without chronic low back pain during flexion-extension and lateral bending tasks," *J. Electromyography Kinesiol.*, vol. 10, pp. 261–273, 2000.
[79] J. A. Biurrun-Manresa, G. P. Nguyen, M. Curatolo, T. B. Moeslund, and O. K. Andersen, "Probabilistic model for individual assessment of central hyperexcitability using the nociceptive withdrawal reflex: A biomarker for chronic low back and neck pain," *BMC Neurosci.*, vol. 14, p. 110, 2013.
[80] T. Olugbade, M. S. H. Aung, N. Bianchi-Berthouze, N. Marquardt, and A. C. de C. Williams, "Bi-modal detection of painful reaching for chronic pain rehabilitation systems," in *Proc. 16th ACM Int. Conf. Multimodal Interaction*, 2014, pp. 455–458.
[81] T. Olugbade, N. Bianchi-Berthouze, N. Marquardt, and A. C. de C. Williams, "Pain level recognition using kinematics and muscle activity for physical rehabilitation in chronic pain," in *Proc. 6th Conf. Affective Comput. Intell. Interaction*, 2015.
[82] M. S. H. Aung, N. Bianchi-Berthouze N, A. C. deC Williams, and P. Watson, "Automatic recognition of fear-avoidance behaviour in chronic pain physical rehabilitation," in *Proc. 8th Int. Conf. Pervasive Comput. Technol. Healthcare*, 2014, pp. 158–161.
[83] E. M. J. Huis in t'Veld, G. J. M. Van Boxtel, and B. de Gelder, "The body action coding system I: Muscle activations during perception and activation of emotion," *Social Neurosci.*, vol. 9, no. 3, pp. 249–264, 2014.
[84] E. M. J. Huis in t'Veld, G. J. M. Van Boxtel, and B. de Gelder, "The body action coding system II: Muscle activations during perception and activation of emotion," *Behavioural Neurosci.*, vol. 8, p. 330, 2014.
[85] Y. Lecrubier, D. V. Sheehan, E. Weiller, P. Amorim, I. Bonora, K Harnett Sheehan, J. Janavs, and G. C. Dunbar, "The mini neuropsychiatric interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI," *Eur. Psychiatry*, vol. 12, no. 5, pp. 224–231, 1997.
[86] A. S. Zigmond and R. P. Snaith, "The hospital anxiety and depression scale," *Acta Psychiatrica Scandinavica*, vol. 67, pp. 361–370. 1983.
[87] M. J. L. Sullivan, S. R. Bishop, and J. Pivik, "The pain catastrophizing scale: Development and validation," *Psychological Assessment*, vol. 7, pp. 524–532, 1995.
[88] C. J. Main and P. J. Watson "Guarded movements: Development of chronicity," *J. Musculoskeletal Pain*, vol. 4, no. 4, pp. 163–170, 1996.
[89] N. Lawrence, "Matlab motion capture toolbox," [Online], http://staffwww.dcs.shef.ac.uk/people/N.Lawrence/mocap/
[90] J. Lichtenauer, J. Shen, M. F. Valstar, and M. Pantic, "Cost-effective solution to synchronised audio-visual data capture using multiple sensors," *Image Vis. Comput.*, vol. 29, pp. 666–680, 2011.
[91] M. Sonnby-Borgström, P. Jonsson, and O. Svensson, "Emotional empathy as related to mimicry reactions at different levels of information processing," *J. Nonverbal Behaviour*, vol. 27, no. 1, pp. 3–23, 2003.
[92] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.
[93] K. O. McGraw and S. P. Wong, "Forming inferences about some intraclass correlation coefficients," *Psychol. Methods*, vol. 1, no. 1, pp. 30–46, 1996.
[94] M. Nicolaou, V. Pavlovic, and M. Pantic, "Dynamic probabilistic CCA for analysis of affective behaviour and fusion of continuous annotations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1299–1311, Jul. 2014.
[95] J. M. Romano, M. P. Jensen, J. A. Turner, A. B. Good, and H. Hops, "Chronic pain patient-partner interactions: Further support for a behavioral model of chronic pain," *Behavior Therapy*, vol. 31, no. 3, pp. 415–440. 2000.
[96] P. A. M. Peeters, and J. W. S. Vlaeyen, "Feeling more pain, yet showing less: The influence of social threat on pain," *J. Pain*, vol. 12, no. 12, pp. 1255–61, 2012.
[97] M. J. L. Sullivan, H. Adams, and M. E. Sullivan, "Communicative dimensions of pain catastrophizing: social cueing effects on pain behaviour and coping," *Pain*, vol. 107, pp. 220–226, 2004.
[98] X. Xiong and F. De la Torre, "Supervised descent method and its application to face alignment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2013, pp. 532–539.
[99] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
[100] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Trans. Comput.*, vol. C-23, no. 1, pp. 90–93, Jan. 1974.
[101] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image Vis. Comput.*, vol. 27, no. 6, pp. 803–816, 2009.
[102] S. Brahnam, L. Nanni, and R. Sexton, "Neonatal facial pain detection using NNSOA and LSVM," in *Proc. Int. Conf. Image Process., Comput. Vis. Pattern Recog.*, 2008, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.302.6295
[103] H. Meng and N. Bianchi-Berthouze, "Affective state level recognition in naturalistic facial and vocal expressions," *IEEE Trans. Syst., Man Cybern. B*, vol. 44, no. 3, pp. 315–328, Mar. 2014.
[104] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, 2008.
[105] D. Bernhardt, and P. Robinson, "Detecting affect from non stylised body motions," in *Proc. Int. Conf. Affective Comput. Intell. Interaction*, 2007, pp. 59–70.
[106] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[107] H. J. Griffin, M. S. H. Aung, B. Romera Paredes, C. McLoughlin, G. McKeown, W. Curran, and N. Bianchi-Berthouze, "Laughter type recognition from whole body motion," in *Proc. Human Assoc. Conf. Affective Comput. Intell. Interaction*, 2013, pp. 349–355.
[108] H. J. Griffin, M. S. H. Aung, B. Romera Paredes, C. McLoughlin, G. McKeown, W. Curran, and N. Bianchi-Berthouze, "Perception and automatic recogntion of laughter from whole body motion: Continuous and categorical perspectives," *IEEE Trans. Affective Comput.*, vol. 6, no. 2, pp. 165–178, Apr.–Jun. 2015.
[109] B. Jiang, M. F. Valstar, B. Martinez, and M. Pantic, "Dynamic appearance descriptor approach to facial actions temporal modelling," *IEEE Trans. Syst. Man Cybern. B, Cybern.*, vol. 44, no. 2, pp. 161–174, Feb. 2014.
[110] O. Rudovic, V. Pavlovic, and M. Pantic, "Context-sensitive dynamic ordinal regression for intensity estimation of facial action units," *IEE Trans Pattern Anal. Mach. Intell.*, vol. 37, no. 5, pp. 944–958, May 2015.
[111] N. Dael, M. Mortillaro, and K. R. Scherer, "The body action and posture coding system (BAP): Development and reliability," *J. Nonverbal Behavior*, vol. 36, no. 2, pp. 97–121, 2012.
[112] S. Park, P. Shoemark, and L.-P. Morency, "Toward crowdsourcing micro-level behaviour annotations: The challenges of interface, training and generalization," in *Proc. Int. Conf. Intell. User Interfaces*, 2014, pp. 37–46.
[113] J. Walsh, C. Eccleston, and E. Keogh, "Pain communication through body posture: The development and validation of a stimulus set," *Pain*, vol. 155, no. 11, pp. 2282–2290, 2014.
[114] A. Singh, S. Piana, D. Pollarolo, V. Gualtiero, G. Varni, A. Tajadura-Jimenez, A. Williams, A. Camurri, and N. Bianchi-Berthouze, "Go-with-the-flow: Tracking, analysis and sonification of movement and breathing to build confidence in activity despite chronic pain," *Human-Computer Interaction*, (in press).

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.