This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2019.2958341, IEEE Transactions on Pattern Analysis and Machine Intelligence

Automatic Detection of Pain from Facial Expressions: A Survey

Teena Hassan, Dominik Seuß, Johannes Wollenberg, Katharina Weitz, Miriam Kunz, Stefan Lautenbacher, Jens-Uwe Garbas, Ute Schmid

Abstract—Pain sensation is essential for survival, since it draws attention to physical threat to the body. Pain assessment is usually done through self-reports. However, self-assessment of pain is not available in the case of noncommunicative patients, and therefore, observer reports should be relied upon. Observer reports of pain could be prone to errors due to subjective biases of observers. Moreover, continuous monitoring by humans is impractical. Therefore, automatic pain detection technology could be deployed to assist human caregivers and complement their service, thereby improving the quality of pain management, especially for noncommunicative patients. Facial expressions are a reliable indicator of pain, and are used in all observer-based pain assessment tools. Following the advancements in automatic facial expression analysis, computer vision researchers have tried to use this technology for developing approaches for automatically detecting pain from facial expressions. This paper surveys the literature published in this field over the past decade, categorizes it, and identifies future research directions. The survey covers the pain datasets used in the reviewed literature, the learning tasks targeted by the approaches, the features extracted from images and image sequences to represent pain-related information, and finally, the machine learning methods used.

Index Terms—automatic pain detection, facial expressions of pain, pain datasets, pain feature representation, facial expression analysis, machine learning, survey.

1 Introduction

The International Association for the Study of Pain (IASP) [1, p. 209] defines pain as "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage." Pain has the function of increasing attention, and initializing and maintaining mechanisms such as self-protection, recovery, and healing [2]. Without pain, human life would be significantly shorter [3]. The expression of pain triggers social reactions such as empathy, care, and nursing [2]. However, untreated pain is known to be a major contributor to reduced quality of life [4], to a progressive decline of functional and mental capacity [5], loss of appetite [6], reduced sleep [7], and behavioral disturbances including agitation, depression, and anxiety [8]. Therefore, timely detection and adequate treatment of pain is important.

Reliable assessment of pain is necessary for determining appropriate analgesics (pain-relieving medication) and their dosage. Self-reports or observational scales are used to assess pain. Self-reporting methods include rating scales (e.g., Visual Analogue Scale for Pain (VAS) [9], Numeric Rating Scale (NRS) [10]), pain diaries [11], or verbal descriptions (e.g., [12]). Patients who are noncommunicative due to a critical illness, narcotic medication, cognitive impairment, or infancy, cannot use self-reporting methods to communicate the pain they are experiencing. Therefore, assessment by other people, especially caregivers and nursing staff, is necessary. For this, different observational pain scales such as Behavioral Pain Scale (BPS) [13], Pain Assessment in Advanced Dementia (PAINAD) [14], or Neonatal Infant Pain Scale (NIPS) [15], are used in clinical settings. Facial expressions, body movements, and vocalizations are part of such observational pain scales. In research settings, other assessment tools are also used to study these observable dimensions of pain expression in greater detail. For example, the Prkachin-Solomon-Pain-Intensity (PSPI) scale [16] is used quite frequently by human coders (cf. [17]) as well as by computer scientists (cf. [18]) to annotate intensities of facial expressions of pain.

Pain assessment through observation is very challenging, and is affected by the subjective biases and errors in beliefs of the observer [19]. Studies such as [20] and [21] have found that pain is underestimated by nursing

- *T. Hassan is currently a Research Associate at Bielefeld University, and an external researcher at the Electronic Imaging Department, Fraunhofer IIS, 91058 Erlangen, Germany. The contributions in this paper are part of her research at Fraunhofer IIS.*
  E-mail: teena.c.hassan@gmail.com
- *D. Seuß and J. Garbas are with the Electronic Imaging Department, Fraunhofer IIS, 91058 Erlangen, Germany.*
- *J. Wollenberg was with the Electronic Imaging Department, Fraunhofer IIS, 91058 Erlangen, Germany, at the time his contributions to the paper were created.*
- *K. Weitz is currently with the Department of Computer Science, University of Augsburg, Germany. She was with the Electronic Imaging Department, Fraunhofer IIS, 91058 Erlangen, Germany, at the time her contributions to the paper were created.*
- *M. Kunz is with the Faculty of Medical Sciences, University of Augsburg, Germany.*
- *S. Lautenbacher is with the Faculty of Human Sciences and Education, University of Bamberg, Germany.*
- *U. Schmid is with the Faculty of Information Systems and Applied Computer Science, University of Bamberg, Germany.*
  E-mail: ute.schmid@uni-bamberg.de

Manuscript received Month dd, yyyy.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2019.2958341, IEEE Transactions on Pattern Analysis and Machine Intelligence

TABLE 1
Summary of datasets containing facial expressions of pain that are available upon request via email to first author or through a website.

Reference	Diagnostic Status	Pain Stimulus	Demographics	Sample Size	Annotation Granularity	Annotation (Labels)
Infant COPE [50] [55]	healthy	heel lancing	26 neonates (18–72 hours); 13 male, 13 female; Caucasian	204 facial images (pain images: 60)	frame-level	pain, crying, heel friction, nasal air stimulus, rest
UNBC-McMaster Shoulder Pain Expression Archive Database [18]	shoulder pain	range of motion tests on shoulders	129 adults; 63 male, 66 female	200 image sequences (total frames: 48,398; pain frames: 8369)	frame-level	12 AUs and their intensities (A–E), 66 facial landmarks, PSPI score
					sequence-level	self-report via VAS, sensory scale, affective-motivational scale; observer report via Observer Rated Pain Intensity (OPI)
BioVid Heat Pain Database [52]	healthy	heat	Part A: 87 adults (18–65 years); 44 male, 43 female	8700 videos (pain videos: 6960)	sequence-level	baseline (no pain), 4 pain stimulus intensity levels
			Part B: 86 adults (18–65 years); 42 male, 44 female	8600 videos (pain videos: 6880; partial facial occlusion due to facial EMG electrodes)	sequence-level	baseline (no pain), 4 pain stimulus intensity levels
			Part C: 87 adults (18–65 years); 44 male, 43 female	87 videos (long version of Part A with one video per subject)	segment-level	pain stimulus
		case vignette	Part D: 90 adults (18–65 years); 45 male, 45 female	630 videos (posed pain videos: 90)	sequence-level	7 posed expressions: neutral, pain, anger, disgust, fear, happy, sad
Hi4D-ADSIP [68]	healthy	none	80 adults (18–60 years); 32 male, 48 female; diverse ethnicities	3360 3D sequences (pain sequences: 240)	sequence-level	posed pain, 6 posed emotions (anger, disgust, surprise, fear, sadness, happiness), and 7 other facial articulations at 3 intensity levels each: mild, normal, extreme
BP4D-Spontaneous [69]	healthy	cold	41 adults (18–29 years); 18 male, 23 female; 20 Euro-American, 11 Asian, 6 African-American, 4 Hispanic	328 2D and 328 3D videos (pain videos: 41 2D, 41 3D)	frame-level	27 AUs (max. 20 sec segments), 3D head rotation angles, facial landmarks (83 for 3D, 49 for 2D)
					sequence-level	pain, anger, startle, fear, sadness, disgust, embarrassment, happiness

Spatial features consist of *geometric* or *textural* features. Geometric features describe the shape of the face in terms of point-based shape description schemes. These define point placements on facial features such as eyes, eyebrows, cheek, nose, lips, chin, and/or facial boundary. The locations of these facial feature points or higher-order features such as distances and angles between the facial feature points, are used as geometric features. Textural features describe the appearance of the face and facial features. Textural features include a description of the edges of facial features, and the wrinkles or folds that appear on or around them. Textural features used in literature range from raw pixel intensities to hand-crafted or self-learned features. The commonly used hand-crafted textural feature descriptors are Gabor filters [133], Local Binary Patterns (LBP) [134] [135], and Histogram of Oriented Gradients (HOG) [136]. Geometric features were rarely used alone. Textural features, either alone or in combination with geometric features, are the most widely used features in automatic pain detection. The combination of geometric and textural features are denoted in Tables 4 and 5 as *hybrid* features.

Spatiotemporal features describe changes in spatial features over time, and can be categorized in a similar fashion as spatial features into geometric, textural, and hybrid categories. In other words, spatiotemporal geometric and textural features were used either independently or in combination (hybrid). Geometric features extracted from a sequence of images were summarized using mathematical and statistical operators. Spatiotemporal textural features such as LBP-TOP [137] and HOG-TOP [138] were extracted from Three Orthogonal Planes (TOP), one of which covers the temporal dimension that spans a temporally ordered sequence of images. Yang et al. [104] compared the performance of several spatiotemporal textural features, such as LBP-TOP, LPQ-TOP[11], BSIF-TOP[12], and their combinations.

In cases where hybrid, mixed, or multiple features of the same type were used, the fusion of features was performed either before the learning step (cf. [75], [123]), or by fusing the decisions of classifiers trained separately for each feature (cf. [107], [116]). The former is commonly referred to as "early fusion", and the latter is commonly referred to as "late fusion".

In two-step approaches, pain detection is done based on an intermediate representation of the face in terms of

11. LPQ-TOP is the spatiotemporal variant of Local Phase Quantization (LPQ) [139]
12. BSIF-TOP is the spatiotemporal variant of Binarized Statistical Image Features (BSIF) [140]

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2019.2958341, IEEE Transactions on Pattern Analysis and Machine Intelligence

TABLE 2
Summary of datasets whose availability is unknown at the time of writing this survey. The authors of the respective papers might be contacted for potential access. Note: 'elderly' denotes adults aged over 65 years.

Reference	**Diagnostic Status**	**Pain Stimulus**	**Demographics**	**Sample Size**	**Annotation Granularity**	**Annotation (Labels)**
Wilkie [58]	lung cancer	Activities of Daily Living (ADL)	43 adults; 27 male, 16 female; 31 Caucasian, 12 others	43 videos	segment-level	9 AUs (20 sec segments)
					sequence-level	self-report via VAS and State-Trait Anxiety Inventory (STAI)
Roy et al. [70]	healthy	none	34 adults	1088 videos (pain videos: 136)	sequence-level	pain, neutral, 6 basic emotions (anger, sadness, happiness, disgust, fear, surprise)
Kunz et al. [53]	demented	mechanical pressure	42 elderly (mean: 76.7 years); 20 male, 22 female	42 videos (pain stimulus sessions: 840)	segment-level	for each 5 sec stimulus session: 44 AUs and their intensities (A–E), self-report of pain level via verbal category scale
	healthy		54 elderly (mean: 74.2 years); 11 male, 43 female	54 videos (pain stimulus sessions: 1080)		
Lu et al. [51]	healthy	heel lancing	57 neonates; 30 male, 27 female	510 images (pain images: 160)	frame-level	pain, cry, calm
Hammal et al. [71]	healthy	heat	20 adults	20 videos (pain stimulus sessions: 40)	segment-level	for each 5 sec stimulus session: 44 AUs and their intensities (A–E)
Kunz et al. [57]	healthy	heat	44 young adults (18–30 years); 22 male, 22 female	44 videos (pain stimulus sessions: 352)	segment-level	for each 5 sec segment after stimulus reached peak: 44 AUs and their intensities (A–E), self-report via VAS
Littlewort et al. [56] [72]	healthy	cold	26 adults; 6 male, 20 female	78 one-minute videos (real pain videos: 26; faked pain videos: 26)	sequence-level	baseline (no pain), real pain, faked pain
Niese et al. [54]	healthy	hand movements with tourniquet attached	21 adults (20–30 years); 10 male, 11 female	21 image sequences (total frames: 966000; pain frames: 31500)	segment-level	self-report of pain intensity via NRS
EmoPain [27]	chronic lower back pain	physical exercises	22 adults (19–67 years); 7 male, 15 female; 18 Caucasian, 4 others	44 videos (total frames: 585,487; pain frames: 50,071)	frame-level	pain, no pain
					segment-level	self-report of pain and anxiety on 1–10 scale
	healthy		28 adults (mean age: 37.1 years); 14 male, 14 female; 26 Caucasian, 2 Asian	–	–	no pain
Irani et al. [66]	healthy	mechanical pressure	12 elderly females (66–90 years)	96 videos (total frames: 2388; pain frames: 1631)	sequence-level	self-report of pain intensity via NRS
Pediatric Pain Dataset [59] [73]	after appendectomy	endogenous and exogenous (manual pressure at surgical site)	50 youth (5–18 years); 27 male, 23 female; 35 Hispanic, 9 non-Hispanic white, 5 Asian, 1 Native American	300 videos (endogenous pain: 150 exogenous pain: 150)	sequence-level	self and observer reports of pain intensity via NRS
Singh [60]	back/neck/knee pain	manual pressure on affected area	21 adults; 12 male, 9 female	21 image sequences (total frames: 336)	frame-level	7 AUs and their intensities
Tsai et al. [28]	emergency cases with pain or headache	endogenous	117 adults	205 videos	sequence-level	self-report via NRS

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2019.2958341, IEEE Transactions on Pattern Analysis and Machine Intelligence

TABLE 3
Summary of the learning approaches that have been developed and tested for automatic pain detection from facial expressions.

Learning Task	Temporal Information	References
One-Step Approaches		
pain and no-pain	no	Brahnam et al. [79], Monwar and Rezaei [80], Brahnam et al. [81], Lu et al. [51], Ashraf et al. [82], Lucey et al. [83], Siebers et al. [84], Nanni et al. [85], Gholami et al. [86], Monwar and Rezaei [87], Wei and Li-min [88], Lucey et al. [18], Lucey et al. [89], Werner et al. [90], Chen et al. [91], Khan et al. [92], Pedersen [93], Neshov and Manolova [94], Rathee and Ganotra [95], Aung et al. [27], Kharghanian et al. [96], Roy et al. [97], Rupenga and Vadapalli [98], Meawad et al. [99], Alphonse and Dharma [100]
	yes	Werner et al. [101], Meng and Bianchi-Berthouze [102], Werner et al. [29], Kächele et al. [103], Yang et al. [104]
pain and emotions	no	Niese et al. [54]
	yes	Hammal et al. [71], Hammal and Kunz [105]
pain and states (crying, calm/rest)	no	Brahnam et al. [79], Lu et al. [51], Yuan et al. [106]
pain and distress (via heel friction or air stimulus on nose)	no	Brahnam et al. [79]
pain intensity (continuous)	no	Werner et al. [90], Kaltwang et al. [107], Romera-Paredes et al. [108], Neshov and Manolova [94], Wang et al. [109], Liu et al. [110]
	yes	Kächele et al. [103], Florea et al. [111], Zhou et al. [112], Kaltwang et al. [113], Zhao et al. [114], Rodriguez et al. [115], Egede et al. [116], Egede and Valstar [117], Lopez-Martinez et al. [118], Tavakolian and Hadid [119]
pain intensity (discrete)	no	Gholami et al. [86], Lucey et al. [89], Hammal and Cohn [120], Singh [60], Rathee and Ganotra [95], Roy et al. [97], Alphonse and Dharma [100]
	yes	Rudovic et al. [121], Irani et al. [122], Irani et al. [66], Werner et al. [123], Tsai et al. [28], Lopez-Martinez et al. [118]
pain event in sequence	yes	with localization: Sikka et al. [124], Sikka et al. [125], Lo Presti and La Cascia [126], Lo Presti and La Cascia [127]
		without localization: Chen et al. [75]
Two-Step Approaches		
pain and no-pain	no	Lucey et al. [83], Lucey et al. [128], Zafar and Khan [129]
	yes	Schmid et al. [77], Sikka et al. [59], Siebers et al. [78]
pain intensity (continuous)	yes	Sikka [73], Sikka et al. [59], Zhang et al. [76], Lopez-Martinez et al. [130]
pain intensity (discrete)	no	Zafar and Khan [129]
	yes	Ghasemi et al. [74]
posed and genuine pain	yes	Littlewort et al. [72], Littlewort et al. [56], Bartlett et al. [131]

AUs. Features used for learning pain-related targets were therefore extracted from the AU labels or AU scores[13] provided by the first step. We categorize these features that are indirect representations of the input image or image sequence into *non-temporal* and *temporal* features. Table 6 provides an overview of the indirect features that have been used for automatic pain detection. Non-temporal features refer to the AU representations for a single image or a *single timestep* in an image sequence. In this case, the AU labels or scores for the image are used as features for pain detection (e.g. [77], [78], [83]). Temporal features refer to AU representations for a sequence of images spanning *multiple timesteps*. In this case, AU scores provided by the first learning stage are aggregated using statistical operators (cf. [59]) or dynamic features are extracted using temporal filters (cf. [131]). Note that the categorization into non-temporal and temporal features is based purely on whether the pain detection in the second step used AU detection outputs for a single image/timestep or for multiple timesteps. It does not take into account whether temporal information was used in the first step for AU detection. It was noted that the two-step approach followed by Lopez-Martinez et al. [130] used a combination of direct and indirect features for continuous pain intensity estimation (see Table 6).

The extracted features are often post-processed to increase their discriminative power or to extract the most important information. Principal Component Analysis (PCA) is a commonly used method to select the most important feature dimensions and thereby transform the features into a lower-dimensional space (cf. [76], [79], [114]). Rathee and Ganotra [95] proposed multiview distance metric learning to fuse LBP, HOG, and Gabor features, and to increase the discriminative power of the new set of features. Florea et al. [111] used a semi-supervised transfer learning method based on spectral regression to learn the most discriminative feature dimensions of the extracted Histogram of Topological (HoT) features and to reduce the dimensionality of the feature space. An exhaustive survey of the feature post-processing methods is outside the scope of this paper. The reader is advised to refer to other surveys on facial expression analysis (e.g. [132]) to obtain an overview about

13. The term "scores" is used in this paper to broadly refer to scores/probabilities/intensities of AUs.