Deep Multimodal Pain Recognition: A Database and Comparison of Spatio-Temporal Visual Modalities

Mohammad A. Haque, Ruben B. Bautista, Fatemeh Noroozi, Kaustubh Kulkarni, Christian B. Laursen, Ramin Irani, Marco Bellantonio, Sergio Escalera, Golamreza Anbarjafari, Kamal Nasrollahi, Ole K. Andersen, Erika G. Spaich, Thomas B. Moeslund

VAP, CVC, SMI, CNAP, UT, AAU

Corresponding author's email: mah@create.aau.dk

Abstract

Managing pain is not only a duty but also highly cost prone. Automatic pain assessment systems from facial videos are rapidly evolving due to the need of managing pain in a robust and cost effective way. Among different challenges of automatic pain assessment from facial video data two issues are increasingly prevalent: first, exploiting both spatial and temporal information of the face to assess pain level, and second, incorporating multiple visual modalities to capture complementary face information related to pain. Most works in the literature focus on merely exploiting spatial information on chromatic (RGB) video data on shallow learning scenarios. However, employing deep learning techniques for spatio-temporal analysis considering Depth (D) and Thermal (T) along with RGB has high potential in this area. In this paper, we present the first state-of-the-art publicly available database, 'Multimodal Intensity Pain (MIntPAIN)' database, for RGBDT pain level recognition in sequences. We provide the first baseline results including 5 pain levels recognition by analyzing independent visual modalities and their fusion with CNN and LSTM models.

Introduction

The most primitive state of pain management is the assessment of pain. Self-report in pain assessment is not always practical due to difference in concept of pain and requiring cognitive, linguistic and social competencies. Thus, different objective measure were proposed, such as Prkachin and Solomon Pain Intensity (PSPI) scale based on Facial Action Coding System (FACS) [2] and Visual Analog Scale (VAS) [3]. Regardless of these objective measures, there are two major challenges in video-based pain detection. First, exploiting both spatial and temporal information of the face to assess pain level. Second, incorporating multiple visual modalities to capture complementary face. But we need a database too for performance evaluation while proposing methods to address the above challenges. Thus, in this work we present:

- The first publicly available database, 'Multimodal Intensity Pain (MIntPAIN)' database, for RGBDT (color, depth and thermal) pain level recognition in sequences.
- Baseline results including 5 pain levels recognition by analyzing independent visual modalities and their fusion
- Employed state-of-the-art deep learning CNN+LSTM model [4] to exploit spatio-temporal information

The Database

Two notable databases that are publicly available for video pain experiments are the UNBC-McMaster database [1] and BioVid database [5]. However, to the best of our knowledge, there is no publicly available RGBDT database that focuses on pain analysis from face. Thus, the main contribution of this work is creating a multimodal pain intensity database. It is to be noted that electrical stimulation is a highly reproducible and noninvasive method to elicit experimental pain or discomfort; and both Functional Electrical Stimulation (FES) of muscle nerves and electrical stimulation of the Nociceptive Withdrawal Reflex (NWR) are easy to manipulate using graded stimulation intensities to generate pain [6]. Thus, we have collected the new RGBDT database, named as Multimodal Intensity Pain (MIntPAIN) database, by employing controlled electrical stimulation to generate pain to the subjects' muscles. **Table I** shows the distinction between the new MIntPAIN database and the other two available pain databases. Before proceeding to the pain detection, some preprocessing steps are necessary, such as: frame by frame synchronization using time-stapms, Markus's face detection [7], and 8-points homography. The preprocessing steps are shown in **Figure 1**.

Exploiting Multimodal Spatio-Temporal Information

Figure 2 shows the cropped faces obtained after preprocessing from all three modalities. **Figure 3** shows the approach of fusing the modalities in a hybrid deep learning framework from [4] by using VGG-Face CNN with LSTM.

Modalities	CNN-RGB	CNN-T	CNN-D	LSTM-RGB	LSTM-D	LSTM-T
Mean Frame(%)	**18.17**	18.08	16.71	15.36	14.72	13.13
Mean Sequence (%)	**18.55**	18.33	17.41	15.36	14.72	13.13

Table 2: Results of the independent modalities

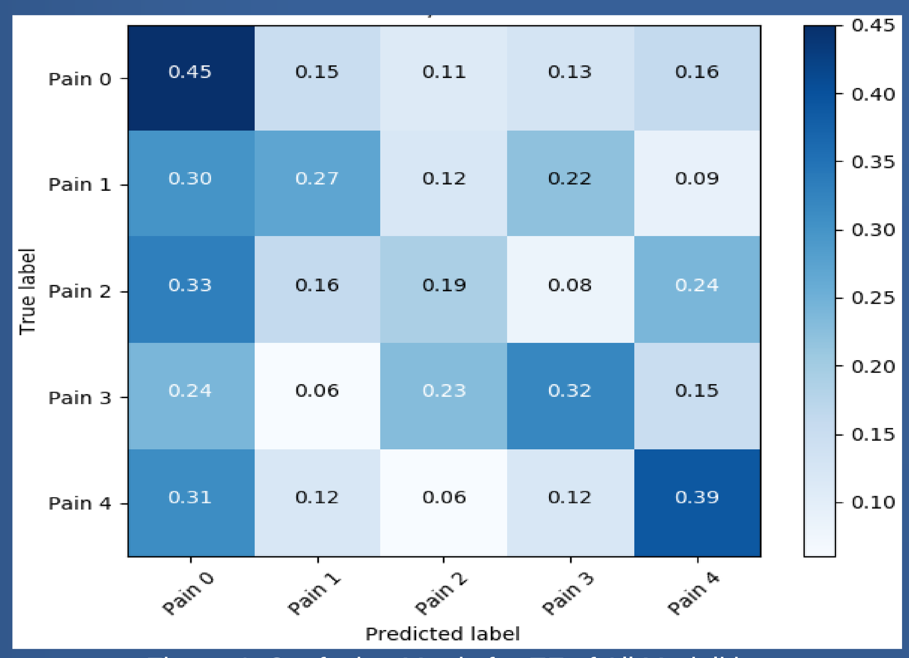

Figure 4: Confusion Matrix for EF of All Modalities

Results and Discussions

Table 2 shows the VGG-Face CNN and LSTM results on independent modalities. On the other hand, **Table 3** shows Early Fusion (EF) and Late Fusion (LF) results for different combinations of the modalities. **Figure 4** shows the confusion matrix corresponding to the early fusion of all three modalities. From our evaluations, we observed that fusion of modalities are more discriminative than training the classifiers with independent ones. The early fusion of all three modalities provided the highest performance. These results support the usability of the different visual data modalities. We also observed that the usage of LSTM to learn long term dependencies in our data achieves poor performance for the considered input fine-tuned VGG-features.

Attribute	UNBC-McMaster database (2011)	BioVid database (2013)	MIntPain database (2018)
No. of subjects	129 (16 are available)	90 (87 are available)	20
Subject's type	Self-identified pain patient	Healthy voluteers	Healthy volunteers
Pain type	Natural sholder pain	Stimulated heat pain	Stimulated electrical pain
Pain levels	0-16 (PSPI) and 0-10 (VAS)	1-4 (Stimuli)	0-4 (Stimulai)
Modalities	RGB	RGB	RGB, Depth, Thermal
Size of the database	200 variable length videos with 31,571 frames	17,300 5s videos with 25 fps	9366 variable length videos with 1,87,939 frames

Table 1: The new MIntPAIN database and the other public visual databases focusing on video pain.

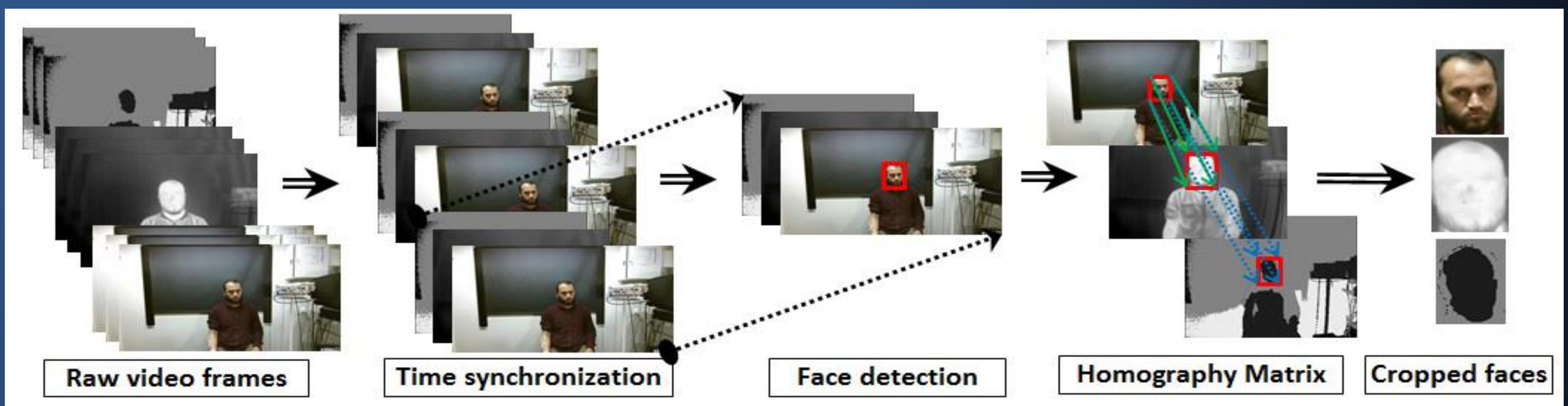

Figure 1: Preprocessing steps employed on the raw video frames of different modalities to crop facial regions before pain assessment

Figure 2: Faces from two subjects for all 5 pain levels (Level0 to Level4 from left to right) for all different RGBDT modalities

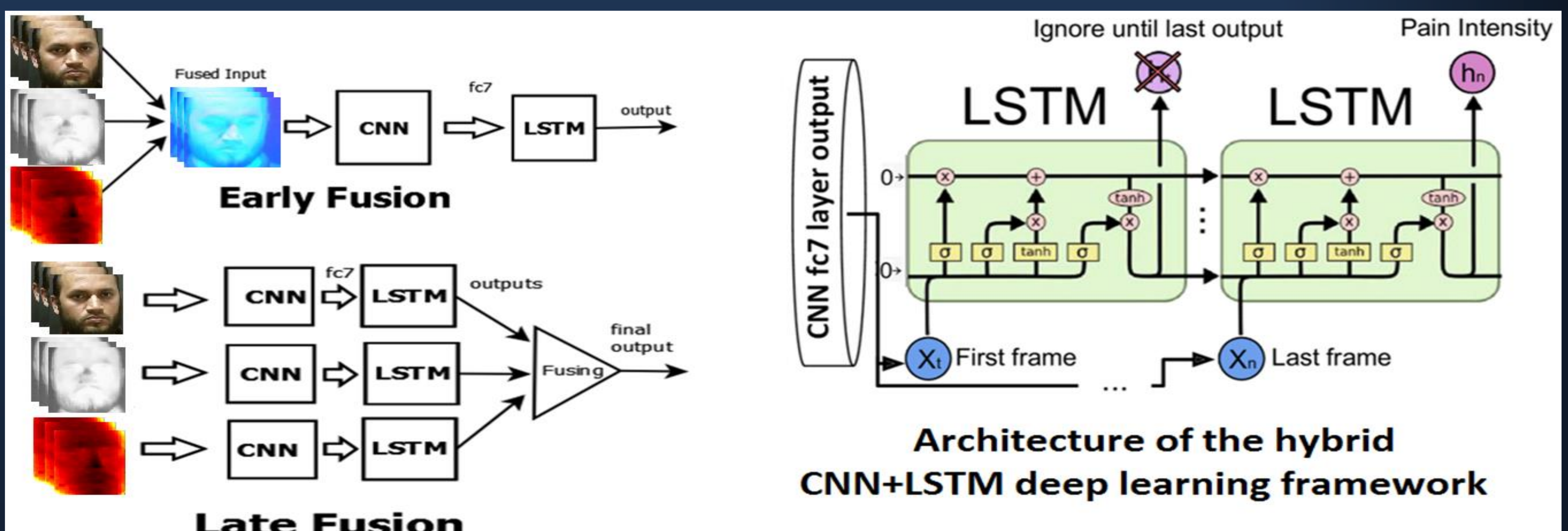

Figure 3: The block diagram of fusion strategies along with the deep hybrid classification framework based on a CNN and LSTM

Fusion	EF-RGB-T	EF-RGB-D	EF-D-T	EF-RGB-DT	LF-RGB-T	LF-RGB-D	LF-D-T	LF RGB-D-T
Mean Frame (%)	23.85	24.62	23.12	**32.40**	21.80	23.20	22.50	25.20
Mean Sequence(%)	30.77	27.92	25.30	**36.55**	22.10	22.30	22.70	25.40

Table 3: Early (EF) and late fusion (LF) results for different combinations of the modalities

MISC

The database is available here:
http://www.vap.aau.dk/mintpain-database/

An implementation of CNN+LSTM available here:
https://github.com/prlz77/LSTM-on-CNN

REFERENCES:

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

[2] K. M. Prkachin and P. E. Solomon. The structure, reliability and validity of pain expression: evidence from patients with shoulder pain. Pain, 139(2):267–274, Oct 2008.

[3] Ken Steffen Frahm, Carsten Dahl Mørch, Warren M. Grill, and Ole Kæseler Andersen. Experimental and model-based analysis of differences in perception of cutaneous electrical stimulation across the sole of the foot. Medical & Biological Engineering & Computing, 51(9):999–1009, Sep 2013.

[4] Marco Bellantonio, Mohammad A Haque, Pau Rodriguez, Kamal Nasrollahi, Taisi Telve, Sergio Escarela, Jordi Gonzalez, Thomas B Moeslund, Pejman Rasti, and Gholamreza Anbarjafari. Spatio-temporal pain recognition in cnn-based super-resolved facial images. In International Workshop on Face and Facial Expression Recognition from Real World Videos, pages 151–162. Springer, 2016.

[5] Mohammad A Haque, Kamal Nasrollahi, and Thomas B Moeslund. Pain expression as a biometric: Why patients' self-reported pain doesn't match with the objectively measured pain? In Identity, Security and Behavior Analysis (ISBA), 2017 IEEE International Conference on, pages 1–8. IEEE, 2017.

[6] Ole K Andersen, Finn A Sonnenborg, and Lars Arendt-Nielsen. Reflex receptive fields for human withdrawal reflexes elicited by non-painful and painful electrical stimulation of the foot sole. Clinical Neurophysiology, 112(4):641 – 649, 2001.

[7] Markus Mathias, Rodrigo Benenson, Marco Pedersoli, and Luc Van Gool. Face detection without bells and whistles. In EuropeanConference on Computer Vision, pages 720–735. Springer, 2014.