Spatiotemporal Analysis of RGB-D-T Facial Images for Multimodal Pain Level Recognition

Irani, Ramin; Nasrollahi, Kamal; Oliu Simon, Marc; Corneanu, Ciprian; Guerrero, Sergio Escalera; Bahnsen, Chris; Lundtoft, Dennis Holm; Moeslund, Thomas B.; Pedersen, Tanja; Klitgaard, Marie-Louise; Petrini, Laura

Published in:
IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW), 2015

DOI (link to publication from Publisher):
10.1109/CVPRW.2015.7301341

Publication date:
2015

Document Version
Early version, also known as pre-print

Spatiotemporal Analysis of RGB-D-T Facial Images for Multimodal Pain Level Recognition

Ramin Irani, Kamal Nasrollahi
Visual Analysis of People (VAP) Laboratory
Rendsburggade 14, 9000 Aalborg, Denmark
{ri,kn}@create.aau.dk

Marc O. Simon, Ciprian A. Corneanu, Sergio Escalera
Computer Vision Center, UAB
Edificio O, Campus UAB, 08193, Bellaterra (Cerdanyola), Barcelona, Spain
Dept. Applied Methematics, University of Barcelona
Gran Via de les Corts Catalanes, 585, 08007, Barcelona
{moliusimon, cipriancorneanu}@gmail.com, sergio@maia.ub.es

Chris Bahnsen, Dennis H. Lundtoft, Thomas B. Moeslund
Visual Analysis of People (VAP) Laboratory
Rendsburggade 14, 9000 Aalborg, Denmark
{cb,tbm}@create.aau.dk

Tanja L. Pedersen, Maria-Louise Klitgaard, and Laura Petrini
Dept. of Communication and Psychology
Fredrik Bajers Vej 7, 9220 Aalborg, Denmark
lap@hst.aau.dk

Abstract

Pain is a vital sign of human health and its automatic detection can be of crucial importance in many different contexts, including medical scenarios. While most available computer vision techniques are based on RGB, in this paper, we investigate the effect of combining RGB, depth, and thermal facial images for pain intensity level recognition. For this purpose, we extract energies released by facial pixels using a spatiotemporal filter. Experiments on a group of 12 elderly people applying the multimodal approach show that the proposed method successfully detects pain and recognizes between three intensity levels in 82% of the analyzed frames, improving by more than 6% the results that only consider RGB data.

1. Introduction

Pain plays an essential role as part of a complex system for dealing with injury [27]. Distinguishing harmful from harmless situations, prompting avoidance of harm and its associated cues, giving a high priority to escape from danger, and promoting healing by inhibiting other activities that might cause further tissue damage has great adaptive value [4]. Pain serves to promote the organisms health and integrity, to the extent that congenital absence of pain on injury significantly shortens human life [10]. There are rare cases of people with no pain sensation. An often-cited case is that of F.C., who did not exhibit a normal pain response to tissue damage. She repeatedly bit the tip of her tongue, burned herself, did not turn over in bed or shift her weight while standing, and showed a lack of autonomic response to painful stimuli. She died at the age of 29 [25].

Physiological measures of pain vary significantly from person to person, failing to reflect its intensity [24]. Indicators of pain include changes in heart and respiratory rates, blood pressure, vagal tone, and palmar sweating [8]. In addition to physiological responses, facial expressions play a critical role for communicating pain [30, 28, 12].

The main way of assessing pain in clinical contexts is by self report. This can be sometimes problematic because of

the lack of reliability and consistency and in the case of persons with limited communication abilities (infants, young children, patients with certain neurological impairments, intubated and unconscious persons) this option is simply not available [14]. One alternative is to have human experts to asses pain. Unfortunately a considerable amount of training is needed and the process is burdensome. Another alternative is the automatic assessment of pain. Even though there is no consensus about the physiological measures of pain [6] brow-lowering, tightening the eyelids, raising the cheeks (orbit tightening), nose wrinkling or upper-lip raising and eye closure were identified as a core set of actions during facial expression of pain [29, 26].

Pain is a complex behavioural expression. While until now research has mostly concentrated on facial pain recognition from RGB, in the case of human affect perception there is strong evidence supporting the integration of multiple modalities over using a single modality [2, 32, 35]. For instance, depth information offers important advantages. It is more invariant to rotation and illumination and it captures more subtle changes on the face. It also facilitates recognizing a wider range of expressions which would be more difficult to detect from only RGB. Pain experience is highly correlated with cardiovascular changes such as increase in heart rate, blood pressure, cardiac output and blood flow [17], which in turn has a direct effect on skin temperature. Radiance at different facial regions, captured through thermal infrared imaging, varies according to emotions [38]. While we could not find a similar study in the case of pain, thermal imaging could add valuable information in the automatic assessment of pain. Because the different modalities can be redundant, concatenating features might not be efficient. A common solution is to use fusion. Many studies have demonstrated the advantage of classifier fusion over the individual classifiers [19]. Finally, another important aspect of pain assessment from facial expressions is temporal information [9, 1]. For instance, it has been shown that temporal dynamics of facial behavior represent a critical factor for distinction between spontaneous and posed facial behavior [3, 11, 36] and for categorization of complex behaviors like pain [11, 42].

In this paper, we present a multimodal dynamic pain recognition method from RGB, depth and thermal facial images. We extract energies released by facial pixels in these three modalities using a spatiotemporal filter, and test the methodology in a real case scenario consisting of 12 subjects, obtaining high recognition rates measuring the level of pain, and showing the benefits of including multimodal information.

The rest of this paper is organized as follows. In Section 2 we review related methods in the field. Methodology is described in Section 3. Section 4 presents the experimental results. Finally, Section 5 concludes the paper.

2. Related Work

While the vast majority of related work described in the literature focuses on pain recognition from RGB (e.g. [21, 3, 31], works based on 3D [39] or multimodal [40, 41] also exist. A first category of RGB methods do not include temporal information [13, 22, 21, 23]. In [13] a method for automatically detecting four levels of pain intensity is proposed based on SVM trained with the responses of Log-normal Filters. In [22], Littlewort et al. classify real and fake pain with 88% accuracy compared to the 49% obtained by naive human subjects. In [3], a pain no pain classification is proposed, achieving a hit rate of 81% using AAM and SVM. Similar to [22], [21] obtains above naive human discrimination between posed and genuine facial expression of pain (72% compared to 52%) by using boosted Gabor filters and SVM. Finally in [23], Lucey et al. show that detecting pain by fusing pain associated AUs is more efficient than using extracted features to directly detect pain/no-pain. A distinct group of RGB methods use temporal information [31, 40, 18, 15]. In [18] Kaltwang et al. propose what they claim to be the first fully automatic continuous pain intensity method. It is based on a late fusion of a set of regression functions learned from appearance (DCT and LBP) and geometric (shape) features. In [31] facial expressions of pain intensity is detected by using *Conditional Ordinal Random Fields* (CORF). In [15] an approach based on the Transferable Belief Model are proposed capable of obtaining above human observers performance when recognizing the pain expression among the six basic facial expressions and neutral on acted and spontaneous sequences. Examples of using other modalities include 3D [39] and physiological signals in a multimodal context [41, 40]. [39] is based on a SVM classifier and a function model for intensity rating. The intensity model is trained using Comparative Learning, a technique that simplifies labelling of data. In [40, 41] it is proposed a multimodal (RGB+physiological) dataset (BioVid Heat Pain Database) and dynamic methods for recognizing pain by combining information from video and biomedical signals, namely facial expression, head movement, galvanic skin response, electromyography and electrocardiogram. In contrast to previous works, the method we propose here is the first one to combine RGB, depth and thermal facial images for pain recognition.

3. Methodology

The block-diagram of the proposed system is shown in Fig. 1. Having a trimodal input video, first the RGB modality is used to detect the face and facial landmark positions. The registration information between the three modalities is used to estimate the positions of the landmarks in the other two modalities. Afterwards, an energy-based method using steerable separable spatiotemporal filters, which uses

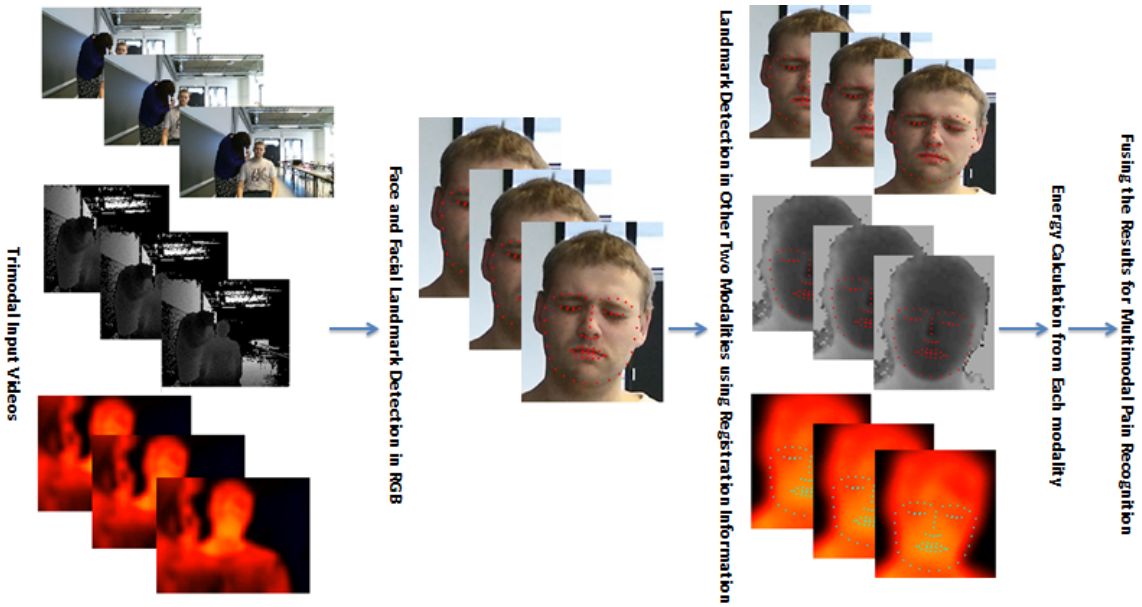


Figure 1: The block diagram of the proposed system.

the landmark positions, is applied to each modality. This gives an indication of visible pain in each of these modalities. Finally, a fusion unit is used to combine the results of the three modalities to recognize the pain level. These steps are explained in the following subsections.

3.1. Landmark detection in RGB

In order to develop a fully automatic pain recognition method, first a landmark detection approach is applied over the RGB modality. Two steps are required in order to efficiently detect landmarks in video sequences. Firstly, the Viola&Jones [37] face detection algorithm is applied to the first frame. Landmarks are then located inside the facial region by using the Supervised Descent Method (SDM) [43]. In the subsequent frames, the facial region is obtained from the previous frame geometry, applying SDM inside that region to estimate the new landmark locations. The SDM algorithm consists on a custom implementation trained for the detection of 68 landmarks (see Figure 2(a)). For training, a combination of the LFPW [5], HELEN [20], AFW [44] and IBUG [33] datasets is used, amounting to 3837 instances. The ground truth of the 68 facial landmarks over these datasets is obtained from the 300 Faces In-The-Wild Challenge (300-W) [34].

Since the tracking approach uses the previous frame landmarks to select the facial region, it is important to have a robust algorithm for landmark localization. SDM is less prone to local minima when compared to other minimization methods [43], learning the descent direction and step size towards the global minima regardless of the gradient at the current estimate. This is achieved with a cascaded approach, where an initial shape estimate S^0 is iteratively adjusted to the image though linear regressors. At each step a simplified version of SIFT is used to extract features from the landmarks. These are concatenated into a feature vector F^t_{SIFT}, where the dimensionality has been reduced by using PCA to keep 95% of the original variance. A linear regression W^t estimates the displacement between the current shape estimate S^t and the face geometry, as shown in Equation 1.

$$S^{t+1} = S^t + F^t_{SIFT} \cdot W^t \tag{1}$$

This robustness can be further improved by using multiple initializations. For this purpose, $n = 10$ plane rotations of the mean shape are homogeneously sampled from the range $[-\pi/2, +\pi/2]$ and fit to the image during test. The distance between each pair of fits $D_{i,j} = d(S_i, S_j)$ is stored into a matrix $D^{<n \times n>}$ of distances, being $d(x, y)$ the the sum of euclidean distances between corresponding landmarks. The fit minimizing the sum of distances to the others, i.e. the centroid fit, is selected as the best one. This criterion is used because it corresponds to the fit towards which most other alignments, regardless of the initialization orientation, tend to converge, thus having a higher probability of corresponding to the global minima.

3.2. Landmark detection in depth and thermal

The landmarks obtained from the RGB frames are translated to the corresponding frames in the depth and thermal modalities by first finding a registration between the three modalities. Once the registration is found, we transform

the landmarks coordinate system from RGB to both depth and thermal, representing the geometry in those spaces. The registration of these modalities is explained in the following subsection.

3.2.1 Registration of different modalities

The registration between the RGB and depth modalities uses the built-in calibration tool of the Kinect^{TM} for Windows 2.0 SDK. Although the calibration parameters used for the registration are not directly visible, it is possible to obtain an accurate registration of each depth image to the corresponding RGB frame. Registration of the thermal modality to RGB requires considering two modalities captured by two separate devices, whose relative positions are not know beforehand. Therefore, we obtain a separate calibration of the thermal and RGB modalities by moving a custom-made multimodal checkerboard in the region where the upper body of test participants is located. The multimodal calibration board consists of a white, A3-sized 10 mm polystyrene backdrop which is heated by a heat gun immediately before the calibration, and thick card board plate where a chessboard pattern is cut out. The differences in temperature and color of the two boards enable the detection of point correspondences between RGB and thermal.

The point correspondences from the calibration stage is used to obtain a homography which is accurate for points near the face of the participants.

3.3. Feature extraction

Having found the positions of the landmarks in all the three modalities, the next step is to use these positions to extract a feature that can give us an indication of the pain in each modality. The following steps should be performed for all the three modalities similarly, thus we will explain them only for RGB modality.

Since changes due to pain in facial expression are spatiotemporal phenomena, we need to employ a descriptor that considers both spatial and temporal domains, and can be independently applied to all three modalities. For these reasons a steerable separable spatiotemporal filter has been chosen, which considers the second derivative of a Gaussian filter and their corresponding Hilbert transforms. This filter measures the orientation and level of energy in the $3D$ space of x, y, and t, representing the spatial and time spaces, respectively. The spatial responses of the filter describe the spatial texture of the face, while the temporal responses describe the dynamic of the features, e.g., the velocity. For each pixel, the energy is calculated by:

$$E(x,y,t,\theta,\gamma) = [G_2(\theta,\gamma) * I(x,y,t)]^2 + [H_2(\theta,\gamma) * I(x,y,t)]^2, \quad (2)$$

where '*' stands for a convolution operator, (x,y,t) shows the pixel value located at the position (x,y) of the tth frame (temporal domain) of the aligned video sequence I, and $E(x,y,t,\theta,\gamma)$ shows the energy released by this pixel at the direction θ and scale γ. To make the above obtained energy measure comparable in different facial expressions, we normalize it using:

$$\hat{E}(x,y,t,\theta,\gamma) = \frac{E(x,y,t,\theta,\gamma)}{\sum E(x,y,t,\theta_i,\gamma) + \epsilon}, \quad (3)$$

where θ_i considers all the directions and ϵ is a small bias used for preventing numerical instability when the overall estimated energy is too small. Finally, to improve the localization, we weight the above normalized energy using [7]:

$$\dot{E}(x,y,t,\theta,\gamma) = \hat{E}(x,y,t,\theta,\gamma).z(x,y,t,\theta), \quad (4)$$

where:

$$z(x,y,t,\theta) = \begin{cases} 1 & \sum_{\gamma_i} \hat{E}(x,y,t,\theta,\gamma_i) > Z_\theta \\ 0 & Otherwise \end{cases}, \quad (5)$$

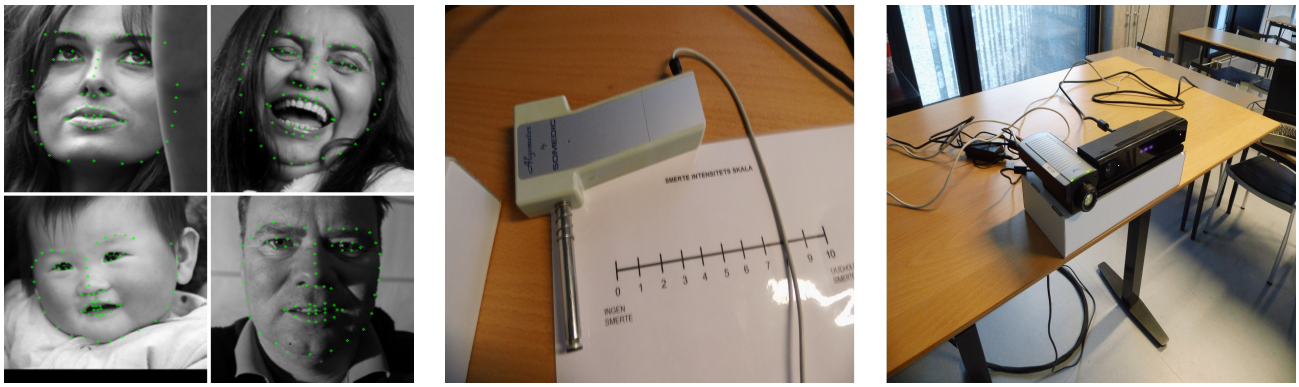
in which Z_θ is a threshold for keeping energies at the direction θ, as too small energies are likely to be noise. The weighted normalized energy obtained in Eq. 4 assigns a number to each pixel (corresponding to the level of the released energy by that pixel) in each of the four chosen directions of $\theta = 0, 90, 180$, and 270. Following [16] these pixel based energies are then combined into region based energies using their histograms of directions by:

$$H_{R_i}(t,\theta_i,\gamma) = \sum_{R_i} \dot{E}(x,y,t,\theta_i,\gamma), \quad (6)$$

where H_{R_i} is the histogram of the directions, and R_i, $i = 1, 2$, or 3 is the ith region of the face [16]. Since the muscles are moving back to their original locations, after they are moved due to, e.g., pain, we need to combine the regional histograms, by considering the regions that are directly related to each other during the pain process. Following [16] two directions of up-down (UD) and left-right (LR) are considered for combining the histograms. These directional histograms are obtained for each modality of RGB, depth, and thermal. Then, they will be separately used to obtain the pain level, which is explained in the next section.

3.4. Pain recognition

In the previous subsection two histograms where obtained for the energy orientation of facial regions of each modality, resulting in six directional histograms of energy. The two histograms of the RGB modality are combined by:



(a) (b) (c)

Figure 2: (a) Sample images from the 300-W dataset labeled with 68 facial landmarks. (b) The hand-held device used for introducing the pain. (c) The cameras used for capturing the three modalities.

$$PI_{RGB} = \sum_{i=1}^{3} w_{R_{i_{UD}}} A_{R_{i_{UD}}} + \sum_{i=1}^{3} w_{R_{i_{LR}}} A_{R_{i_{LR}}}, \quad (7)$$

in which PI_{RGB} is the pain index in RGB modality and $A_{R_{i_{UD}}}$ and $A_{R_{i_{LR}}}$ are defined as the integrals of UD and LR for the ith region $(i = 1, 2, 3)$, respectively [16]:

$$A_{R_{i_{UD}}} = \sum_{t=1}^{n} UD_t, \quad A_{R_{i_{LR}}} = \sum_{t=1}^{n} LR_t, \quad (8)$$

where n is the number of the frames in the video. Similarly pain indexes are determined for the other two modalities, resulting in PI_D and PI_T, representing pain indexes obtained for depth and thermal modalities, respectively. These three pain indexes are then fused together to recognize the pain level using:

$$PI = w_{RGB} PI_{RGB} + w_D PI_D + w_T PI_T, \quad (9)$$

where w_{RGB}, w_I, w_T are the weights associated to corresponding modalities, and PI is the fused pain index. It should be noted that $w_{RGB} + w_D + w_T = 1$. In the following subsection, it is explained how PI is used to determine the pain level based on experimentally found thresholds.

4. Experimental results

In order to present the results, we first discuss the setup and data considered for the experiments, and evaluation measurements and parameters.

4.1. Setup and data

12 healthy elderly volunteers (all females) between the ages of 66 and 90 years (mean age 73.6 years) participated in the study. Participants were screened with an interview prior participation to exclude conditions that could affect pain perception and pain report. Exclusion criteria were, if the participant reported the presence of severe ongoing pain, neuropsychological and psychiatric disorders, diabetes, or had signs of a rheumatic or arthritic disease, especially on the neck/shoulders. During the interview, subjects were also tested with the Mini Mental State Examination (MMSE) in order to ensure intact cognitive capabilities.

All subjects were pain-free and none of them had taken any analgesic or sedative for at least 48 hours prior to the experiment. The study protocol was approved by the regional ethics committee. Experimental pressure pain was applied on the subjects trapezius muscle. Eight stimuli of different intensities: No-Pain, Light-Pain, Moderate Pain and Strong Pain were applied on left and right trapezius muscles of the participants.

An electronic hand-held pressure algometer (Somedic AB, Stockholm, Sweden) was used to produce noxious mechanical pressure (Fig. 2(b)). A force gauge fitted with a rubber disk with a surface of 1 cm^2 was used in this study. Pain and no-pain stimuli were determined for each subject on the base of the individual pain detection threshold (PDT). Pain stimuli were calculated as follow: No-Pain: 0.2 X PDT, Light Pain: 1.10 X PDT, Moderate Pain: 1.30 X PDT, and Strong Pain: 1.5 X PDT.

Subjects pain self-reports were recorded using a numerical rating scale (NRS) that measured the perceived intensity of the stimulation. The NRS ranges from 0 (no pain) to 10 (the worst pain you can imagine). Participants NRS was recorded after each stimulus.

During each pain and no-pain stimulation subjects face was video-recorded in order to identify specific pain behaviors on the participants face.

During the process the subjects were filmed using a de-

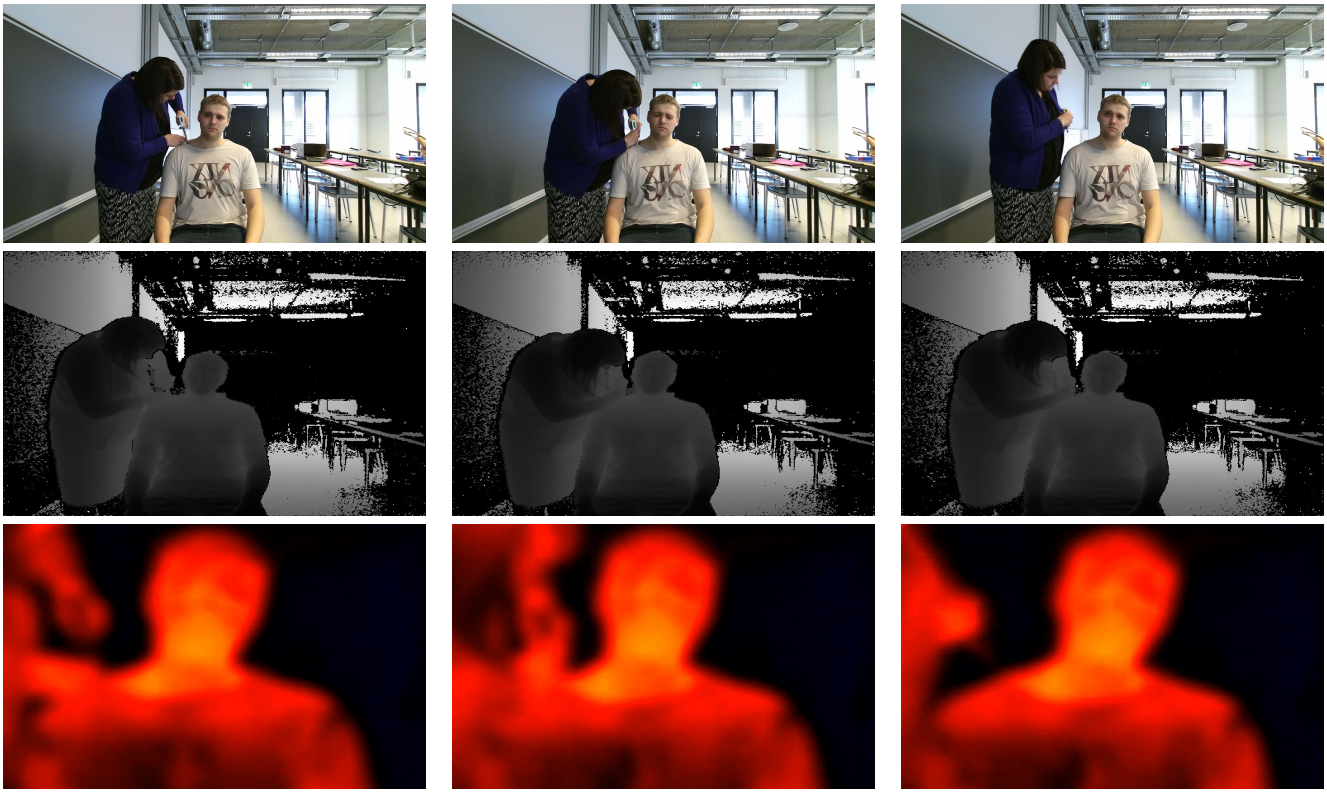


Figure 3: RGB (first row), depth (second row), and thermal (third row) of a test subject during the experimental process: first column (before introducing pain), second column (under pain), and third column (after the pain).

vice capturing all the three modalities of RGB, depth, and thermal. The first two modalities were recorded by a Microsoft Kinect for Windows V2 device and the thermal modality by an AXIS Q1921 thermal camera. Setup is shown in Fig. 2(c).

Fig. 3 shows a test subject in these three modalities during our experimental process. The first, second, and third rows show the RGB, depth, and thermal modalities of this test subject.

4.2. Evaluation measurements and parameters

Having obtained pain index PI for each of the subjects using Eq. 9, if this pain is smaller than 1, it is considered as no-pain, if it is between 2 and 5, it is considered as a weak pain, and if it is larger than 6, it is considered as a strong pain. These thresholds were experimentally defined in agreement with the team of psychologists. Furthermore, the weights w_{RGB}, w_D and w_T in Eq. 9 have been set via cross-validatio, being 0.6, 0.35, and 0.05, respectively. The obtained weights indicate that RGB provide the primary source of information, followed by depth features and finally by the thermal ones. The obtained weights greater than zero show that the use of all three modalities can be useful and complementary for pain recognition.

4.3. Results and discussion

Table 1 shows the results of the proposed system compared to [16], which considers the same recognition approach only using the RGB modality. One can observe that the proposed system achieves high recognition rates for the three levels of pains, improving the results provided in [16], and showing the benefits of the multimodal approach over just considering RGB data.

More specifically, one can see in Table 1 that for the first two levels of pain (no pain and weak pain) the proposed system outperforms the previous system of [16] with a large margin of 16% and 8%, respectively. However, there is no difference in detecting pains of the strong level among the two systems for the considered data. It is mainly produced because for the strong level of pain almost all the details and changes of the facial expressions can be observed in the RGB modalities, and thus adding the other two modalities, at least for the collected video sequences, do not provide any further improvement. However, for the two levels of no pain and weak pain, depth and thermal features are useful to complement visual information from RGB modality and extract more subtle visual features, being useful to discriminate among categories with lower inter-class variability.

Finally, Fig. 4 shows the results of the proposed system

Semantic Ground Truth	Pain Index Ground Truth	Number of Frames	System of [16] (in %)	Proposed System (in %)
No Pain	0 and 1	757	72	88
Weak	2,3,4,5	427	79	87
Strong	$\geqslant$6	1204	76	76
		sum: 2388	weighted avg:75.26	weighted avg: 81.77

Table 1: Comparing the results of the proposed system against the system of [16] applied to our RGB-D-T facial database.

against the RGB-based one of [16] for a small clip within a sequence. Once can observe that the results of our multimodal system is closer to the ground truth compared to the results of [16].

5. Conclusion and future works

Pain is a temporal process that can usually be detected from facial images. The proposed system in this paper uses a spatiotemporal approach using a filter which extracts released energies of facial pixels in three modalities, RGB, depth, and thermal, and groups them into histograms of orientations for different facial regions. The integrals of each of these histograms of orientations over time are then used to find a pain index for each modality. These different pain indexes are then fused into a final pain index. The experimental results on a group of 12 elderly people show that the proposed system can accurately detect the pain and recognize its level into three classes of no-pain, weak and strong pain, improving results of single RGB sequence analysis by more than 6%.

References

[1] Z. Ambadar, W. Schooler, and J. Cohn. The importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science*, 16(5):403–410, 2005.

[2] N. Ambady and R. Rosenthal. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bull.*, 111(2):256–274, 1992.

[3] 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 and vision computing*, 27(12):1788–1796, 2009.

[4] P. Bateson. Assessment of pain in animals. *Animal Behaviour*, 52:827–839, 1991.

[5] P. N. Belhumeur, D. W. Jacobs, D. Kriegman, and N. Kumar. Localizing parts of faces using a consensus of exemplars. In *CVPR*, pages 545–552. IEEE, 2011.

[6] S. Brahnam, C.-F. Chuang, F. Y. Shih, and M. R. Slack. Machine recognition and representation of neonatal facial displays of acute pain. *Artificial Intelligence in Medicine*, 36(3):211–222, 2006.

[7] K. Cannons and R. Wildes. The applicability of spatiotemporal oriented energy features to region tracking. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(4):784–796, 2014.

[8] S. Coffman, Y. Alvarez, M. Pyngolil, R. Petit, C. Hall, and M. Smyth. Nursing assessment and management of pain in critically ill children. *Heart Lung*, 26(221), 1997.

[9] J. Cohn and K. Schmidt. The timing of facial motion in posed and spontaneous smiles. *Journal of Wavelets, Multiresolution & Information Processing*, 2(2):121–132, 2004.

[10] A. R. Damasio. *The feeling of what happens: Body, emotion and the making of consciousness*. 1999.

[11] P. Ekman and E. Rosenberg. *What the face reveals:basic and applied studies of spontaneous expression using the facial action coding system, 2nd edn.* Oxford University Press, London, 2005.

[12] T. Hadjistavropoulos. Social influences and the communication of pain. *Pain: psychological perspectives*, page 87, 2004.

[13] Z. Hammal and J. F. Cohn. Automatic detection of pain intensity. In *ACM-ICMI*, pages 47–52. ACM, 2012.

[14] Z. Hammal and J. F. Cohn. Towards multimodal pain assessment for research and clinical use. In *Roadmapping the Future of Multimodal Interaction Research, Business Opportunities and Challenges*, pages 13–17. ACM, 2014.

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

[16] R. Irani, K. Nasrollahi, and T. B. Moeslund. Pain recognition using spatiotemporal oriented energy of facial muscles. In *Computer Vision and Pattern Recognition Workshop, 2015 IEEE Conference on*, pages 679–692. IEEE, 2015.

[17] W. Janig. The sympathetic nervous system in pain. *Eur. J. Anaesthesiol*, (10):53–60.

[18] S. Kaltwang, O. Rudovic, and M. Pantic. Continuous pain intensity estimation from facial expressions. *Advances in Visual Computing*, pages 368–377, 2012.

[19] L. I. Kuncheva. *Combining Pattern Classifier: Methods and Algorithms*. John Wiley & Sons, 2004.

[20] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang. Interactive facial feature localization. In *Computer Vision–ECCV 2012*, pages 679–692. Springer, 2012.

[21] G. C. Littlewort, M. S. Bartlett, and K. Lee. Faces of pain: automated measurement of spontaneous facial expressions of genuine and posed pain. In *Proceedings of the 9th international conference on Multimodal interfaces*, pages 15–21. ACM, 2007.

[22] G. C. Littlewort, M. S. Bartlett, and K. Lee. Automatic coding of facial expressions displayed during posed and genuine pain. *ICV*, 27(12):1797–1803, 2009.

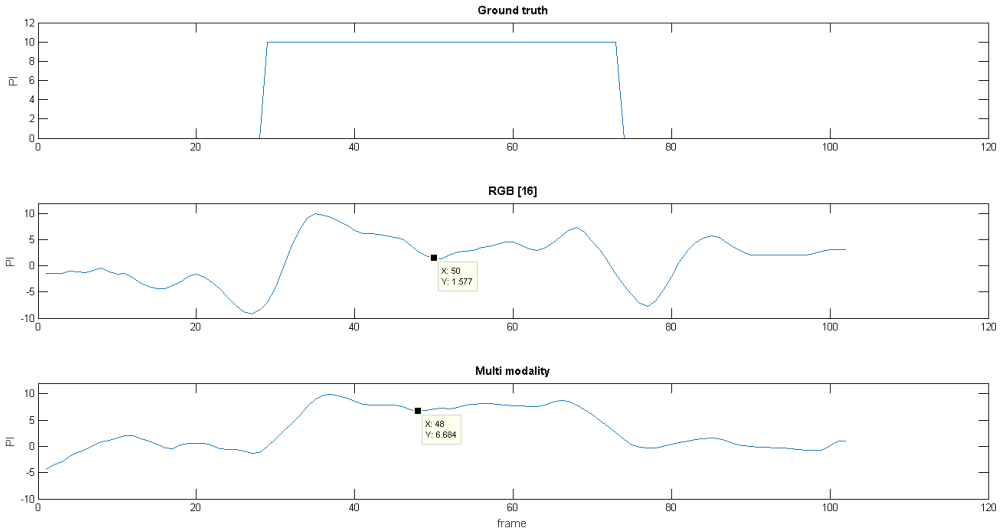


Figure 4: Comparing the results of the proposed system against that of [16] aganist the ground truth.

[23] P. Lucey, J. Cohn, S. Lucey, I. Matthews, S. Sridharan, and K. M. Prkachin. Automatically detecting pain using facial actions. In *Affective Computing and Intelligent Interaction*, pages 1–8. IEEE, 2009.

[24] P. McGrath. *Pain in children: nature, assessment and treatment.* Guildford Press, New York.

[25] N. B. Patel. Physiology of pain. *Guide to pain management in low-resource settings*, page 13, 2010.

[26] K. M. Prkachin. The consistency of facial expressions of pain: a comparison across modalities. *Pain*, 51(3):297–306, 1992.

[27] K. M. Prkachin. Assessing pain by facial expression: facial expression as nexus. *Pain Research & Management: The Journal of the Canadian Pain Society*, 14(1):53, 2009.

[28] K. M. Prkachin and K. D. Craig. Expressing pain: The communication and interpretation of facial pain signals. *Journal of Nonverbal Behavior*, 19(4):191–205, 1995.

[29] 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, 2008.

[30] K. M. Prkachin, P. E. Solomon, and J. Ross. Underestimation of pain by health-care providers: towards a model of the process of inferring pain in others. *CJNR (Canadian Journal of Nursing Research)*, 39(2):88–106, 2007.

[31] O. Rudovic, V. Pavlovic, and M. Pantic. Automatic pain intensity estimation with heteroscedastic conditional ordinal random fields. *Advances in Visual Computing*, pages 234–243, 2013.

[32] J. A. Russell, J. Bachorowski, and J. Fernandez-Dols. Facial and Vocal Expressions of Emotion. 54:329–349, 2003.

[33] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*, pages 397–403. IEEE, 2013.

[34] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. A semi-automatic methodology for facial landmark annotation. In *CVPR Workshops (CVPRW), 2013*, pages 896–903. IEEE, 2013.

[35] K. R. Scherer. Appraisal theory. *Handbook of cognition and emotion*, pages 637–663, 1999.

[36] M. Valstar, M. Pantic, Z. Ambadar, and J. Cohn. Spontaneous versus posed facial behavior: automatic analysis of Brow actions. *Proc eight intl conf multimodal interfaces (ICMI06)*, pages 162–170, 2006.

[37] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001*, volume 1, pages I–511. IEEE, 2001.

[38] S. Wang, Z. Liu, S. Lv, Y. Lv, G. Wu, P. Peng, F. Chen, and X. Wang. A natural visible and infrared facial expression database for expression recognition and emotion inference. *T. Multimedia*, 12(7):682–691, 2010.

[39] P. Werner, A. Al-Hamadi, and R. Niese. Pain recognition and intensity rating based on comparative learning. In *ICIP*, pages 2313–2316. IEEE, 2012.

[40] P. Werner, A. Al-Hamadi, R. Niese, S. Walter, S. Gruss, and H. C. Traue. Towards pain monitoring: Facial expression, head pose, a new database, an automatic system and remaining challenges. In *BMVC*, pages 119–1, 2013.

[41] P. Werner, A. Al-Hamadi, R. Niese, S. Walter, S. Gruss, and H. C. Traue. Automatic pain recognition from video and biomedical signals. In *ICPR*, pages 4582–4587. IEEE, 2014.

[42] A. C. d. C. Williams. Facial expression of pain, empathy, evolution, and social learning. *Behavioral and brain sciences*, 25(04):475–480, 2002.

[43] X. Xiong and F. De la Torre. Supervised descent method and its applications to face alignment. In *CVPR*, pages 532–539. IEEE, 2013.

[44] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *CVPR*.