This CVPR2015 workshop paper is the Open Access version, provided by the Computer Vision Foundation.
The authoritative version of this paper is available in IEEE Xplore.

Pain Recognition using Spatiotemporal Oriented Energy of Facial Muscles

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

Abstract

Pain is a critical sign in many medical situations and its automatic detection and recognition using computer vision techniques is of great importance. Utilizes this fact that pain is a spatiotemporal process, the proposed system in this paper employs steerable and separable filters to measures energies released by the facial muscles during the pain process. The proposed system not only detects the pain but recognizes its level. Experimental results on the publicly available pain database of UNBC show promising outcome for automatic pain detection and recognition.

1. Introduction

Pain is an unpleasant sensation that informs us about some (potential) damages or danger in the structure or the function of the body. It causes emotional effects like anger and depression and may even impact on the quality of life, social activities, relationships and our job. Yet pain is one of the most common reasons for seeking medical care, over 80% of patients complain about some sorts of pain [16]. So, for clinical trials and physicians, pain, similar to blood pressure, body temperature, heart-beat rate and respiration, is an important indicator of health. Therefore, reliable assessment of pain is essential for health related issues. That is why in 1995 Dr. James Campbell called the pain assessment as the fifth vital sign and suggested that quality care means that pain is measured and treated [17].

The most popular technique for pain assessment is Patient self-report. It is convenient and does not require special skills, but has some limitations. It includes inconsistent metrics, reactivity to suggestions, efforts at impression management and differences in conceptualizations of pain between clinicians and sufferers [15]. Moreover, self-reporting cannot be used, e.g., with children and those patients who cannot communicate properly due to neurological impairment or those who require breathing assistant. Craig et al. in [6] evidenced that changes in facial appearance can be a very useful cue for recognizing the pain. In Atul Gawandes recent book [9], it has been shown that periodically monitoring of patients pain level by medical staff improves patients treatment. However, sustained monitoring of patients by this way is difficult, unreliable and stressful. To solve this issue, automatic recognition of pain using computer vision techniques, mostly from facial images, has received great attention over the past few years [3]-[19]. Brahman et al. [3] proposed a binary pain detection approach (pain versus no-pain) using Principal Component Analysis (PCA) and Support Vector Machines (SVM). Ashraf et al. [1] detected the pain using Appearance Active Model (AAM). Littlewort et al. [13] employed a two-layer SVM-based approach in order to detect real pain or posed pain. The above mentioned systems implement a binary classifier, meaning they recognize only two cases of pain versus no-pain, while based on the Prkachin and Solomon Pain Intensity metric [18], pain can be quantized into 16 discrete levels ranging from no-pain (0) to maximal pain (15).

To the Best of our knowledge, there are only few research articles that have estimated the pain level automatically, like those in [11-14]. In [14] a system has been developed which can detect three levels of pain intensity. It uses geometry-based and appearance-based features with a separate SVM classifier for each intensity level of pain. Kaltwang et al. [11] proposed an approach using a combination of appearance-based features, Local Binary Pattern (LBP), and Cosine Discrete Transform (DCT), for detecting intensity levels of pain. They applied a Relevance Vector Regression (RVR) model to predict the pain intensity from each feature set. The above mentioned systems use handcrafted features like LBP and try different classifiers like PCA, SVM, and RVR to detected and recognize the pain. Though they produce interesting results, they do not consider the dynamics of the face. We have observed during our experiments that pain is exposed on the face through changes and motions of some of the facial muscles. These motions obviously release some energy. The level of the released energy is in direct relationship with the level of the pain. This is exactly the point that we want to exploit in this

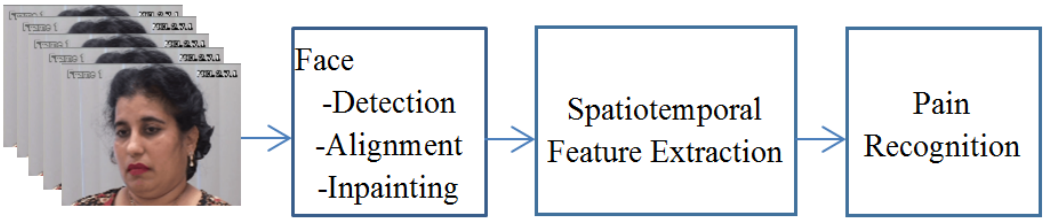

Figure 1. The block diagram of the proposed system.

paper: we develop a system for pain recognition that measures the level of the released energy of the facial muscles over the time. Changes (activation) of facial muscles during the pain have been previously used for pain recognition in Prkachin and Solomon [18]. However, they do not consider the released energy of the facial muscles, but detect the facial Action Units (AU)s and combine them to measure the pain.

There is not that many research work neither on exploiting the temporal axis nor on exploiting the released energy of the facial muscles for detecting and recognizing the pain. For example, [19] measures the pain over the temporal axis. However, it does not use the released energy of the muscles and is more focused on developing a classifier for pain recognition, which is based on Conditional Ordinal Random Fields (CORF). The only system that uses the released energy of facial muscles is the one developed by Hammal et al. [10]. This system uses a combination of AAM and an energy based filter, Log-normal filter, to estimate four intensity levels of pain. Though this system exploits the released energy of the facial muscles, it does that only on a frame by frame basis, in a spatial domain. The proposed system in this paper exploits the released energy of the facial muscles not only on the spatial domain, but also in the temporal one. To do that, we use a specific type of spatiotemporal filter which is shown to be very useful for extracting information in both spatial and temporal domains at the same time, for other applications, like region tracking in [7], [4].

The rest of this paper is organized as follows: the employed filter and the other details of the proposed system are given in the following section. Section 3 explains the performed experiments and discusses the results that are obtained on a public facial database. Finally, section 4 concludes the paper.

2. The Proposed System

The block diagram of the proposed system is shown in Fig. 1. Following the diagram, given an input video sequence, the faces are first detected, simply using the provided landmarks [1]. Then, an Active Appearance Model (AAM) algorithm is used to align the detected faces in different frames of the video to a fixed framework using the provided landmarks. This registration to the fixed framework will cause losing some of the areas of the face, in some of the frames, which appear as holes or lines on the registered faces. To compensate for this, we use an inpanting algorithm. Then, the spatiotemporal filtering is performed in both x, y, and t dimensions to detect the energy released by the facial muscles motion of the aligned faces. Finally, the pain is detected and its level is recognized. These steps are explained in the following subsections.

2.1. Face Detection and Alignment

Detecting the face is an essential step in any facial analysis system, including, pain recognition. The employed database in this paper [18] provides the position of facial landmarks in all the frames of the dataset. We simply use these landmark positions to extract the facial regions in each frame. To do so, as it is shown in Fig. 2a, the facial landmarks are used as vertices of triangles which cover the entire face area, as it is done in [12]. This detected face needs to be segmented from the rest of the image. For this purpose, first, a binary mask (Fig. 2b) is generated such that:

$$Mask = \bigcup_{k=1}^{K} I_k \tag{1}$$

where:

$$I_k = \begin{cases} 1 & P_{ij} \in T_k \\ 0 & Otherwise \end{cases} \tag{2}$$

where T_k is the kth triangle created by landmark points, P_{ij} is a pixel on the image located at (i, j), I_k is a binary image corresponding to T_k and U is a union function. Finally, the face can be segmented from the rest of the image by applying the mask on the image (Fig. 2c).

As mentioned before, the proposed system measures the energy that is released due to the motion of the facial muscles. However, in a video sequence, such motions are not the only type of motion. For example, Fig. 3a shows the positions of 66 facial landmarks in a video sequence of 100 frames. If there was no motion in the video at all, one could only see 66 facial landmarks, but as it can be seen in Fig. 3a,

[1] The employed database in this paper provides the positions of the facial landmarks for all the images.

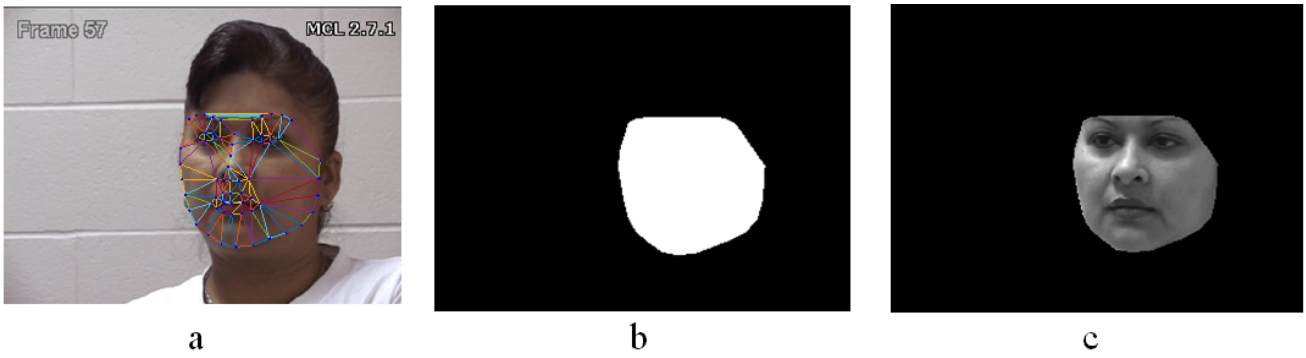

Figure 2. a) Triangles generated from facial landmarks using the algorithm of [12] for face detection, b) the used mask for segmenting the face, and c) the segmented face.

the position of each landmark is changing from one frame to another. This indicates the presence of other motions on the face, like motions resulting from the head pose. Such motions should be filter out. To do that, we employ the face alignment algorithm of [8]. The faces in this algorithm are aligned using the facial landmarks. The results of this alignment, applied to Fig. 3a, can be seen in Fig. 3b.

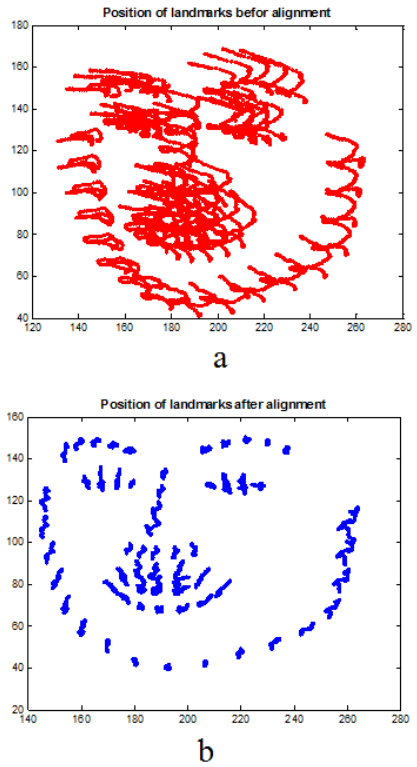

Figure 3. Facial landmarks of 100 face images of a sequence: a) before and b) after alignment.

The alignment algorithm first finds the alignment parameters using the facial landmarks. Then, it uses these alignment parameters to warp the face images of the input video sequence into a common framework using the warping algorithm of [5]. The reader is referred to [5] for the details of the warping algorithm. The result of this warping for a face image is shown in Fig. 4a. It can be seen from Fig. 4a that there are usually some holes (or even lines) in the results of the warped image, which indicate unknown pixel values. This is due to the warping of the facial images that are of different head poses. To deal with this, we use the inpainting algorithm of [2] which uses a series of up-sampling followed by down-sampling. The results of this algorithm applied to Fig. 4a can be seen in Fig. 4b.

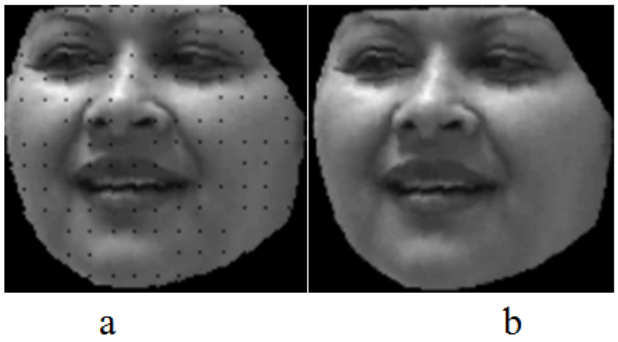

Figure 4. The warped aligned face image: a) before and b) after inpainting.

Having aligned the facial images of the input video sequence and generating an aligned facial video, using the above mentioned steps, the next step is to extract the spatiotemporal features. These features extract the direction and the level of the energies released by the facial muscles. These directions and levels are different for different facial expressions. For example, for a neutral face one should not expect too much energy to be released, while for a laughing face or a face suffering from pain, different levels of energy will be released by the facial muscles in different directions. Extracting of orientation and level of the released energy of

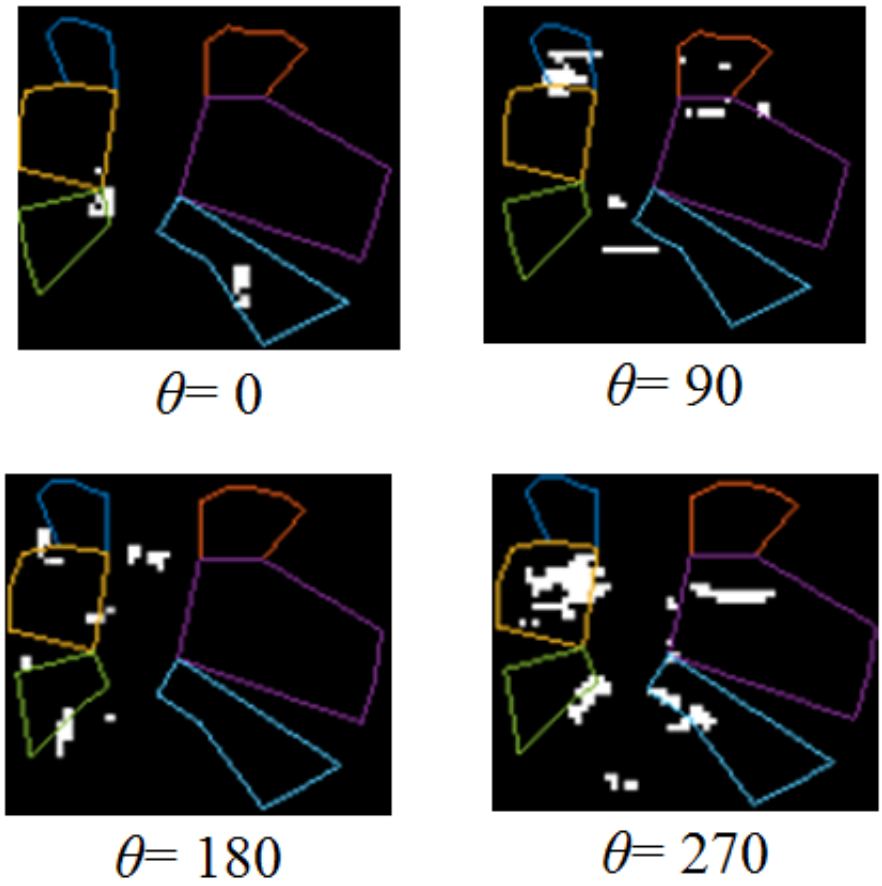

Figure 5. Pixel-based energies (white regions) obtained for the image shown in Fig. 4b computed at four different orientations. The different colors represent different facial regions.

the facial muscles are explained in the following subsection.

2.2. Spatiotemporal Feature Extraction

The extraction of the orientation and the level of the energies released by the facial muscles are done through steerable and separable filters of [4]. These filters compose of a second derivative Gaussian $G_2(\theta, \gamma)$ followed by a Hilbert transform $H_2(\theta, \gamma)$, in different directions of θ, and scales of γ. We do not use a multiscale method, because the level of the energy is not that much visible in coarse scales, hence $\gamma = 1$. During the pain, however, the facial muscles can move in any directions, but such motions can be decomposed into four main directions. Therefore, we measure the released energies in four main directions corresponding to $\theta = 0$, 90, 180, and 270 degrees. The released energy from every pixel is then calculated by:

$$E(x, y, t, \theta, \gamma) = [G_2(\theta, \gamma) * I(x, y, t)]^2 + [H_2(\theta, \gamma) * I(x, y, t)]^2 \quad (3)$$

where * stands for a convolution operator, (x, y, t) shows the pixel value located at the position of x and y of the tth frame (temporal domain) of the aligned video sequence of I, and $E(x, y, t, \theta, \gamma)$ shows the energy released by this pixel at the direction of θ and the scale of γ. To make the above obtained energy measure comparable in different facial expressions, we normalize it using:

$$\hat{E}(x, y, t, \theta, \gamma) = \frac{E(x, y, t, \theta, \gamma)}{\sum E(x, y, t, \theta_i, \gamma) + \epsilon} \quad (4)$$

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 [4]:

$$\dot{E}(x, y, t, \theta, \gamma) = \hat{E}(x, y, t, \theta, \gamma).z(x, y, t, \theta) \quad (5)$$

where

$$z(x,y,t,\theta) = \begin{cases} 1 & \sum_{\gamma_i} \hat{E}(x,y,t,\theta,\gamma_i) > Z_\theta \\ 0 & Otherwise \end{cases} \quad (6)$$

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. 5 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. Fig. 5 shows these pixel-based energies for a facial image computed at the four different orientations.

The above obtained pixel-based energies can be converted into a more understandable form, if we study the regional changes/motions of the facial muscles of different parts of the face. Based on our observations, different regions of the face contribute differently to the level and direction of the energy in different facial statuses. We have observed that facial muscles that are actively participating to the facial motions during the pain are coming from the three regions that are highlighted in Fig. 6. Besides this, the facial muscles on the left side and the right side of each of these three regions are participating differently in motions during the pain. Because of this, inside each region we have used different colors to distinguish between the left and the right sides.

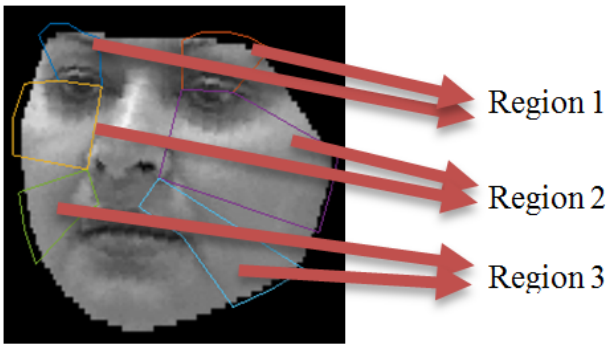

Figure 6. The facial muscles of these three regions are actively contributing to the facial motions during the pain.

To convert the pixel-based energies into region based energies, we obtain the histograms of the directions of the pixel-based energies, computed above, for all the three mentioned facial regions. We calculate the histogram of the directions, $H_(R_i)$, by:

$$H_{R_i}(t,\theta_i,\gamma) = \sum_{R_i} \dot{E}(x,y,t,\theta_i,\gamma) \quad (7)$$

where R_i, $i = 1, 2$, or 3 is the ith region of the face. Fig. 7a to Fig. 7c show three histograms of directions of the weighted normalized energies that are obtained using Eq. 7 at three different stages of a pain process. Fig. 7a (top) shows a neutral face, therefore, there is not that much energy in either of the directions. It can be verified by Fig. 7a (bottom). Fig. 7b (top) shows just the beginning of a pain. The corresponding histogram in Fig. 7b (bottom) it can be seen that muscles in region 1 release energy in direction 270 degree (downwards), but those in region 2 release energy in direction 90 degree (upwards) and those in region 3 release some energies to sides (here direction 180 degree). Fig. 7c (top) shows the face just before revealing from the pain. It can be seen from its corresponding histogram shown in Fig. 7c (bottom) that muscles are releasing energy in the opposite direction of Fig. 7b (bottom) to get back to their original locations.

The above obtained energies can inform us only about some muscles activities (motions), but we need a specific interpretation to see if these motions are due to the pain or not. To do that, we need to study the effect of the pain on the motions of the muscles in the temporal domain. To consider the time domain, we simply obtain the histograms of the directions for each facial region in the aligned input video (Fig. 8(left)). However, as mentioned, since the muscles will move back to their original locations at the end of the pain, instead of the measured directions, we simply consider the **changes** of the released energies of the muscles in two main orientations: up-down (UD) and left-right (LR). For UD we use $UD_{R_i} = H_{R_i}(t,0,\gamma) - H_{R_i}(t,180,\gamma)$, and for $LR_{R_i} = H_{R_i}(t,90,\gamma) - H_{R_i}(t,270,\gamma)$. These will convert the histograms of directions in (Fig. 8(left)) into **changes** in histograms of orientations, as shown in Fig. 8(right).

2.3. Pain Recognition

Having obtained the histograms of orientations from each region, the final step is to combine them by considering the temporal domain and recognize the pain. To consider the temporal domain for monitoring the changes in the released energy we take the integral of the two UD and LR histograms of orientations in each region, using:

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

and

$$A_{R_{i_{LR}}} = \sum_{t=1}^{n} LR_t \quad (9)$$

where $A_{R_{i_{UD}}}$ and $A_{R_{i_{LR}}}$ are the integrals of UD and LR for the ith region ($i = 1, 2, 3$), respectively, and n is the number of the frames in the aligned video. Finally, the pain intensity, PI, is obtained by calculating the above two integrals for each of the three regions:

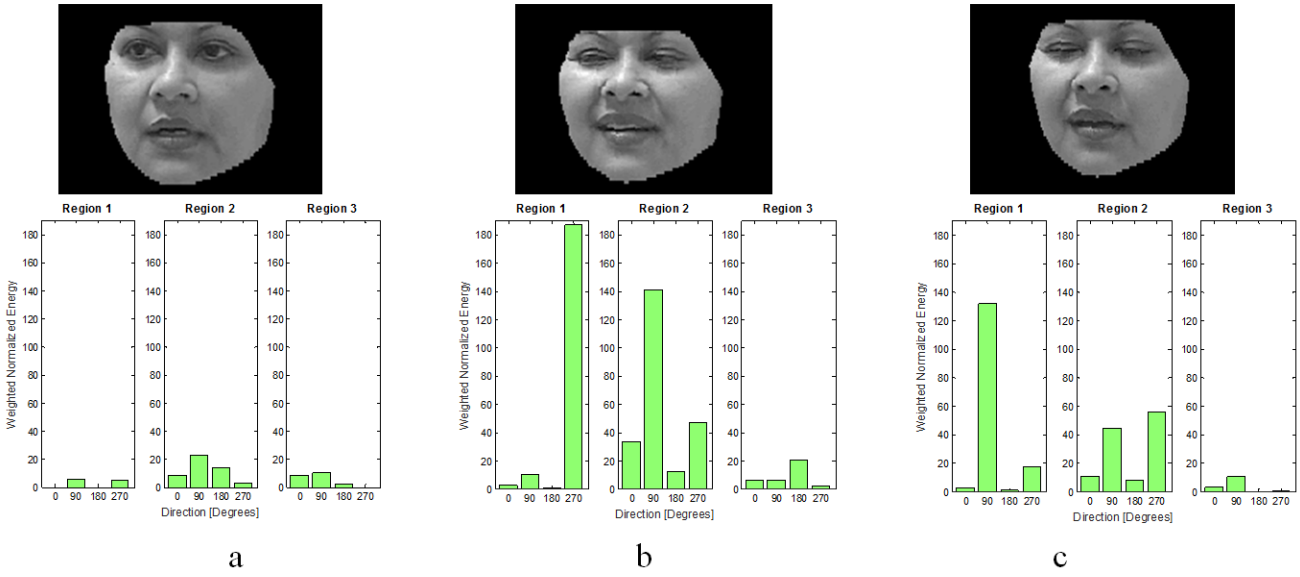

Figure 7. Histograms of directions of the normalized released energies from different facial regions (bottom row), for three different images (top row) at different stages of a pain process: a) neutral face, b) the beginning of the pain, and c) just before the end of the pain.

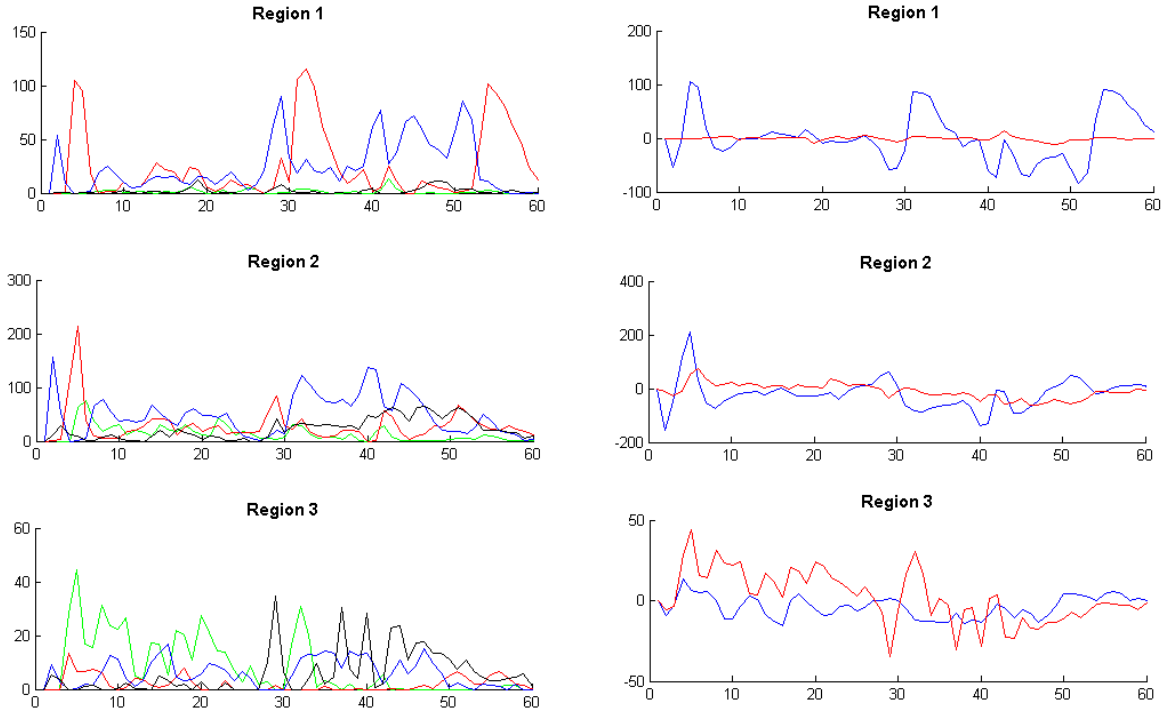

Figure 8. Histograms of directions (left) and changes in histograms of orientations (right) for three facial regions for the aligned facial images of a video sequence of 60 frames. The red, green, blue, and dark colors in the left column show the released energies at directions, 0, 90, 180, and 270, respectively. The red and blue colors in the right column show the up-down (UD) and left-right (LR) histograms of orientations, respectively.

where $w_{R_{i_{UD}}}$ and $w_{R_{i_{LR}}}$ define some experimentally

$$PI = \sum_{i=1}^{3} w_{R_{i_{UD}}} A_{R_{i_{UD}}} + \sum_{i=1}^{3} w_{R_{i_{LR}}} A_{R_{i_{LR}}} \qquad (10)$$

obtained weights of the corresponding regional histograms of orientations. The PI gives us an indication of the presence of the pain in each frame of the video. Depending on the value of PI we find some experimentally achievable thresholds to classify the pain into three class of no-pain, weak, and strong. The experimental results are given in the next section.

3. Experimental Results

The proposed system has been implemented in Matlab 2014b. We have used the publicly available UNBC-MacMaster Shoulder Pain Expression Archive Database [18] for evaluating the proposed system. This database is composed of 25 participants who suffer from pain in their shoulders. They have been filmed during series of movements in two different scenarios (active and passive). In the active scenario participants move their arms themselves, but in the passive scenario a physiotherapist is responsible for this. Videos were captured at a resolution of 320×240. The total number of the recorded frames is 48398. In this database, the ground truth pain information has been provided using AUs (see Fig. 9), by:

$$Pain = AU4 + \max(AU6, AU7) + \max(AU9, AU10) + AU43 \quad (11)$$

Figure 9. Active AUs of Pain, the image is from the UNBC database of [18].

For each frame in the database, the AU intensities were coded on a 6 level scale except the AU number 43 which was coded on two levels [6].

To evaluate our system we selected randomly 50 sequences from 12 participants, containing 4926 frames. Table 1 shows the results of the proposed system against the results of the system developed in [10]. The last columns of the table show the percent by which the system have been able to recognize the pain in that specific level. It can be seen from Fig. 10 and this table, that our system not only detects the pain but also recognizes three different levels of the pain. These three levels are no-pain where $PI \leq 0$, weak pain where $1 \leq PI \leq 2$, and strong pain where $PI \geq 3$. The proposed system actually outperforms the system of [10] in terms of the accuracy of recognizing the level of the pain. It should be mentioned that system of [10] is the only energy based system in the literature for calculating the pain, but it is working in the spatial domain. Outperforming this system by our proposed system means that including the temporal information in an energy-based pain recognition system results in better outcomes.

We should take into account that the pain intensity in [10] has been classified into four levels. Two pain levels of Trace and Weak in [10] correspond to the level Weak in our system. Therefore, for comparison purposes mean of the Trace and Weak levels has been reported as Weak in Table 1.

Finally, Fig. 10 shows the PI values obtained by the proposed system (using Eq. 10) for two different video sequences each containing100 facial images, along with their ground truth data taken from the database employed. It can be seen from this figure, that there is a good overlap between the peaks of the estimated pain intensity curve and the ground truth. It should be noted that the negative values in the estimated values will be considered as zero (no-pain).

4. Conclusion

The proposed system in this paper uses separable steerable filters for automatic detection and recognition of pain. To do that, it applies these filters in both spatial (x, and y axises) and temporal (time axis) domains and measures the energies released by the facial muscles that are active during the pain process. The proposed system has produced promising experimental results on the publicly available dataset of UNBC [18]. The results can be improved more by employing a better warping algorithm to compensate for the variations of head pose, for which we plan to use 3D information of facial landmarks in our future works.

References

[1] A. Ashraf, S. Lucey, T. Chen, K. Prkachin, P. Solomon, Z. Ambadar, and J. Cohn. The painful face: pain expression recognition using active appearance models. In *Proceedings of the 9th International Conference on Multimodal interfaces*, pages 9–14, 2007. 1

[2] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 417–424, 2000. 3

[3] S. Brahnam, C. Chuang, F. Shih, and M. Slack. Machine recognition and representation of neonatal facial displays of acute pain. *Artificial Intelligence in Medicine*, 36:211–222, 2006. 1

[4] 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. 2, 4

Semantic Ground Truth	Pain Index Ground Truth	Number of Frames	System of [10] (in %)	Proposed System (in %)
No Pain	0	4230	65	77
Weak	1, 2	387	36	62
Strong	$\geqslant$3	309	70	70

Table 1. Comparing the results of the proposed system (PS) against the system of [10] for different levels of pain from the images of the UNBC database [18].

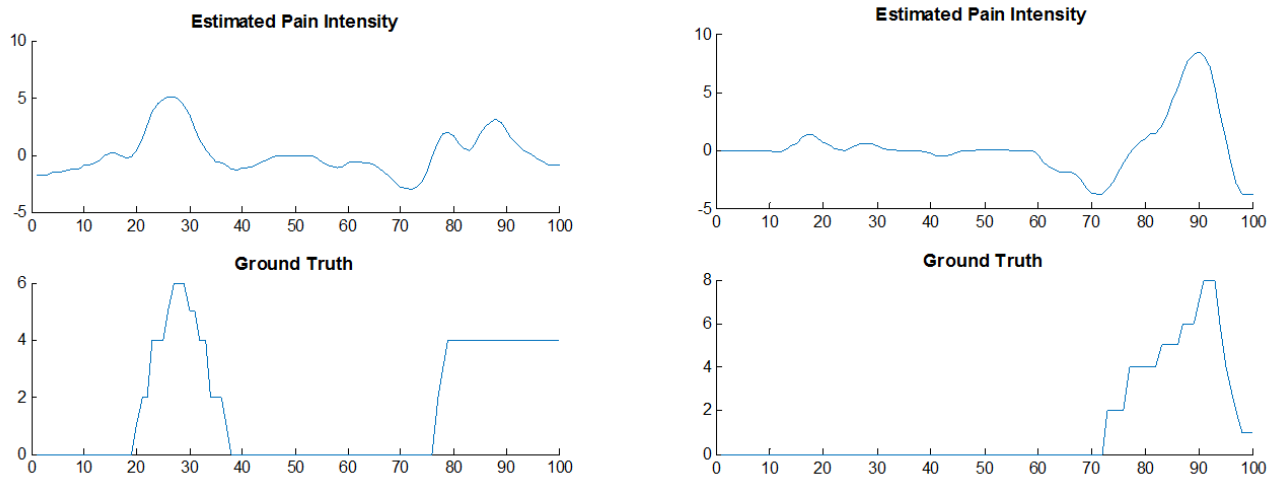

Figure 10. Comparing the pain levels obtained by the proposed system (top row) against the Ground truth (bottom row), for two different pain scenarios.

[5] T. Cootes and C. Taylor. Statistical models of appearance for computer vision. *Technical report, Imaging Science and Biomedical Engineering, University of Manchester*, 2004. 3

[6] K. Craig, K. Prkachin, and R. Grunau. The facial expression of pain. *Handbook of pain assessment, Guilford, New York*, 2001. 1, 7

[7] K. Derpanis and J. Gryn. Three-dimensional nth derivative of gaussian separable steerable filters. In *IEEE International Conference on Image Processing*, pages 553–556, 2005. 2

[8] R. Donner, M. Reiter, G. Langs, P. Peloschek, and H. Bischof. Fast active appearance model search using canonical correlation analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(10):1690–1694, 2006. 3

[9] A. Gawande. The checklist manifesto: How to get things right? *Metropolitan Books, New York*, 2010. 1

[10] Z. Hammal and J. Cohn. Automatic detection of pain intensity. In *Proceedings of the 14th ACM international conference on Multimodal interaction*, pages 22–26, 2012. 2, 7, 8

[11] S. Kaltwang, O. Rudovic, and M. Pantic. Continuous pain intensity estimation from facial expressions. In *International Symposium on Advances in Visual Computing*, pages 368–377, 2012. 1

[12] D. Lee and B. Schachter. Two algorithms for constructing a delaunay triangulation. *International Journal of Computer and Information Sciences*, 9(3):219–242, 1980. 2, 3

[13] G. Littlewort, M. Bartlett, and K. Lee. Automatic coding of facial expressions displayed during posed and genuine pain. *Image and Vision Computing*, 27:1797–1803, 2009. 1

[14] P. Lucey, J. Cohn, K. Prkachin, P. Solomon, S. Chew, and I. Matthews. Painful monitoring: Automatic pain monitoring using the unbc-mcmaster shoulder pain expression archive database. *Image and Vision Computing*, 30:197–205, 2012. 1

[15] P. Lucey, J. Cohn, K. Prkachin, P. Solomon, and I. 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, 2011. 1

[16] W. M., D. S., D. D., H. L., and R. J. The fifth vital sign-what does it mean? *Pain Practice*, 8(6):417–422, 2008. 1

[17] C. Miaskowski, M. Bair, and R. Chou. Principles of analgesic use in the treatment of acute pain and cancer pain. *6th ed. Chicago: American Pain Society*, 2008. 1

[18] K. Prkachin and P. Solomon. The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain. *Pain*, 139:267–274, 2008. 1, 2, 7, 8

[19] O. Rudovic, V. Pavlovic, , and M. Pantic. Automatic pain intensity estimation with heteroscedastic conditional random fields. *Advances in Visual Computing*, 8034:234–243, 2013. 1, 2