Learning Pain from Emotion: Transferred HoT Data Representation for Pain Intensity Estimation

Corneliu Florea[1], Laura Florea[1], and Constantin Vertan[1]

Image Processing and Applications Laboratory,
University "Politehnica" of Bucharest,
{`corneliu.florea; laura.florea; constantin.vertan`} `@upb.ro`

Abstract. Automatic[1] monitoring for the assessment of pain can significantly improve the psychological comfort of patients. Recently introduced databases with expert annotation opened the way for pain intensity estimation from facial analysis. In this contribution, pivotal face elements are identified using the Histograms of Topographical features (HoT) which are a generalization of the topographical primal sketch. In order to improve the discrimination between different pain intensity values and respectively the generalization with respect to the monitored persons, we transfer data representation from the emotion oriented Cohn-Kanade database to the UNBC McMaster Shoulder Pain database.

Keywords: Histograms of Topographical features (HoT); Spectral regression; Transfer learning; Pain intensity estimation

1 Introduction

When discussing the necessity of an automatic pain assessment system, several facts are to be considered: 1. Adult patients, typically, self-assess the pain intensity using a no-reference system, which leads to inconsistent properties across scale, reactivity to suggestion, efforts at impressing unit personnel etc. [10]. 2. Patients with difficulty in communication (e.g. newborns) cannot self report and assessment by specialized personnel is demanded. 3. Correct pain intensity assessment is a critical factor for psychological comfort in the periods spent waiting at emergency units [9]. A solution is automatic appraisal of pain with choices of automatic response of medical personnel alert. Straightforward extensions envisage remote healthcare surveillance or impaired person assistance.

Although other means of investigation (e.g. bio-medical signals) were discussed [25], in the last period significant efforts have been made to identify reliable and valid facial indicators of pain [20], in an effort to develop non-invasive systems. One approach is to analyze the patients face using the Facial Action Coding Systems (FACS) [6], thus identifying action units (AUs) intensity and computing the pain score via the Prkachin - Solomon formula [21].

In this paper we propose a system for face analysis and, more precisely, for pain intensity estimation. To surmount the variations in the face images, we propose the use of the here–introduced Histogram of Topographical (HoT) features. The limitations of the pain annotated database are addressed within a clustering oriented transfer learning procedure that uses emotion portrait data to identify the internal representation of the face descriptors and augments the pain intensity estimation performance.

Further, in section 2 we give a short review of the face based pain intensity estimation systems. Section 3 contains the description of the used features, pointing out the difference with respect to the state of the art. The procedure for transfer learning is presented in section 5, followed by implementation details and results in section 6. The paper ends with discussions and conclusions.

2   Prior art

The majority of the face–based pain estimation methods exploits the Action Unit (AU) face description previously used in emotion detection. A detailed review of the emotion detection methods is in the work of Zeng et al. [26].

The pain recognition from facial expressions was referred in the work of Littlewort et al. [17], who used a previously developed AU detector complemented by Gabor filters, AdaBoost and Support Vector Machines (SVM) to separate fake versus genuine cases of pain. Lucey et al. [20] used Active Appearance Models (AAM) to track and align the faces based on manually labelled key-frames and further fed them to a SVM for frame-level classification. A frame is labelled as "with pain" if any of the pain related action units found earlier by Prkachin [21] are present. Chen et al. [3] transferred information from other patients to the current patient, within the UNBC database, in order to enhance the pain classification accuracy over Local Binary Pattern (LBP) features and AAM landmarks provided by Lucey et al. [20].

Hammal and Kunz [11] measured the nasal wrinkles and used the Transferable Belief Model over subsets of the STOIC database in order to predict the pain for each frame. Kaltwang et al. [14] jointly used LBP, Discrete Cosine Transform (DCT) and AAM landmarks in order to estimate the pain intensity either via AU or directly at a sequence level processing. Werner et al. [25] fused data acquired from multiple sources and information from a head pose estimator to detect the triggering level and the maximum level of supportability of pain.

3   Histogram of Topographical Features

Global/Local Image Descriptors - State of the Art. Many types of local image descriptors are used across the plethora of computer vision applications [23]. Most of the solutions computed in the image support domain are approachable within the framework of the Taylor series expansion of the image function, namely with respect to the order of the derivative used.

Considering the zero-order coefficient of the Taylor series, i.e. the image values themselves, one of the most popular descriptors is the histogram of image values. Next, relying on the first derivative (i.e. the directional gradient), several histogram based descriptors such as HOG [4] or SIFT [18] gained popularity.

The second-order image derivative (i.e. the Hessian matrix) is stable with respect to image intensity and scale and was used in SIFT [18] and SURF [1] image key-points detectors. Deng et al. [5] used the dominant eigenvalue of the Hessian matrix to describe the regions in terms of principal curvature, while Frangi et al. [8] deployed a hard classification of the Hessian eigenvalues in each pixel (thus identifying the degree of local curviness) to describe tubular structures (e.g. blood vessels) in medical images.

Summarizing, we stress that all the mentioned state of the art systems rely on information gathered form a *single* Taylor coefficient of order zero, one or two in order to describe images globally, or locally.

The approximation of the image in terms of the first two Taylor series coefficients is the foundation of the topographical primal sketch introduced by Haralick [12], which was further used for face description by Wang and Jin [24]. In this approach, the description of the image is limited to a maximum number of 12 (or 16) classes which correspond to the basic topographical elements. Further extension lays in the work of Lee and Chen [16], who used the Hessian for locating key points and described their vicinity with the histogram of color values (order zero) and with the histogram of oriented gradients (order one).

We consider that all pixels from a region of interest carry important topographic information which can be gathered in orientation histograms or normalized magnitude histograms. In certain cases, only a combination of these may prove to be informative enough for a complete description of images.

Features. In a seminal work, Haralick et al.[12] introduced the so-called topographical primal sketch. The gray-scale image is considered as a function $I : \mathbb{R}^2 \to \mathbb{R}$. Given such a function, its approximation in any location (i, j) is done using the second-order Taylor series expansion:

$$I(i + \Delta_i, j + \Delta_j) \approx I(i, j) + \nabla I \cdot \langle \Delta_i, \Delta_j \rangle + \frac{1}{2} \begin{bmatrix} \Delta_i\ \Delta_j \end{bmatrix} \mathcal{H}(i, j) \begin{bmatrix} \Delta_i \\ \Delta_j \end{bmatrix} \quad (1)$$

where ∇I is the two-dimensional gradient and $\mathcal{H}(i, j)$ is the Hessian matrix.

Eq. (1) states that a surface is composed by a continuous component and some local variation. A first order expansion uses only the ∇I term (the inclination amplitude) to detail the "local variation", while the second order expansion (i.e. the Hessian), $\mathcal{H}(i, j)$ complements with information about the curvature of the local surface. Based on the gradient and Hessian eigenvalues a region can be classified into several primal topographical features. This implies a hard classification and carries a limitation burden as it is not able to distinguish, for instance, between a deep pit or a shallow pit. We further propose a smoother and more adaptive feature set by considering the normalized local histograms extracted from Hessian eigenvalues and orientation and respectively gradient.

Derivation. Frangi et al. [8] used the concepts of linear scale space theory [7] to elegantly compute the image derivatives. Here, the image space is replaced by the scale space of an image $L(i, j, \sigma)$:

$$L(i, j, \sigma) = G(i, j, \sigma) \otimes I(i, j); \tag{2}$$

where $G(i, j, \sigma)$ is a Gaussian kernel with variance σ^2. The differentiation is computed by a convolution with the derivative of the Gaussian kernel:

$$\frac{\partial}{\partial i} L(i, j, \sigma) = \sigma I(i, j) \cdot \frac{\partial}{\partial i} G(i, j, \sigma) \tag{3}$$

In the scale space, the Hessian matrix $\mathcal{H}(i, j, \sigma)$ at location (i, j) and scale σ is defined as:

$$\mathcal{H}(i, j, \sigma) = \begin{pmatrix} L_{ii}(i, j, \sigma) & L_{ij}(i, j, \sigma) \\ L_{ji}(i, j, \sigma) & L_{jj}(i, j, \sigma) \end{pmatrix} \tag{4}$$

where $L_{ii}(i, j, \sigma)$ is the convolution of the Gaussian second order derivative $\frac{\partial^2}{\partial i^2} G(i, j, \sigma)$ with the image I at location (i, j), and similarly for $L_{ij}(i, j, \sigma) = L_{ji}(i, j, \sigma)$ and $L_{jj}(i, j, \sigma)$. Further analysis requires the computation of the eigenvalues and eigenvectors of the Hessian matrix.

The decomposition of the Hessian in eigenvalue representation acquiesce the principal directions in which the local second order structure of the image can be decomposed. The second order hints to the surface curvature and, thus, to the direction of the largest/smallest bending. The two eigenvalues of the Hessian matrix $\mathcal{H}(i, j, \sigma)$ are $\lambda_1(i, j, \sigma) \leq \lambda_2(i, j, \sigma)$. The eigenvector corresponding to the largest eigenvalue is oriented in the direction with the largest local curvature; this direction of the principal curvature is denoted by $\theta_\lambda(i, j, \sigma)$.

Local Descriptors for Pain Description. In the remainder of the work, for each region of interest Ω, the following HoT descriptors will be used:

– Second order data (Hessian):
  • The histogram of hard voting of image surface curvature orientation. For each pixel in Ω, "1" is added to the orientation of the ridge/valley extracted by computing the angle of the first Hessian eigenvector, if $\lambda_2 > T_\lambda$.

$$H_1^H([\theta]) = \frac{1}{Z_1} \sum_{(i,j) \in \Omega} (\theta_\lambda(i, j) == [\theta]) \cdot (\lambda_2(i, j) > T_\lambda) \tag{5}$$

  • The histogram of soft voting ridge orientation adds, instead of "1", the difference between the absolute values of the Hessian eigenvalues.

$$H_2^H([\theta]) = \frac{1}{Z_2} \sum_{(i,j) \in \Omega} (\theta_\lambda(i, j) == [\theta]) \cdot (\lambda_2(i, j) - \lambda_1(i, j)) \tag{6}$$

The H_1^H and H_2^H histograms produce, each, a vector of length equal with the number of orientation bins (the preferred choice being 8) and describe the curvature strength in the image pixels.

- The range–histogram of the smallest eigenvalue, given a predefined range interval (e.g. $[0, M_{\lambda2} = 30]$). The length of this histogram is typically constructed over $N_{bin} = 8$ bins.

$$H_3^H(k) = \frac{1}{Z_3} \sum_{(i,j)\in\Omega} \left(\lambda_2(i,j) \in \left[(k-1)\frac{M_{\lambda2}}{N_{bin}}; \ k\frac{M_{\lambda2}}{N_{bin}} \right] \right) \qquad (7)$$

- The range–histogram of the differences between the eigenvalues given a predefined differences range interval (e.g. $[0, M_{\lambda12} = 50]$).

$$H_4^H(k) = \frac{1}{Z_4} \sum_{(i,j)\in\Omega} \left((\lambda_1(i,j) - \lambda_2(i,j)) \in \left[(k-1)\frac{M_{\lambda12}}{N_{bin}}; \ k\frac{M_{\lambda12}}{N_{bin}} \right] \right)$$
$$(8)$$

- First order data (gradient):
  - Histogram of orientation, H_1^G [4]; each pixel having a gradient larger than a threshold, T_G casts one vote;
  - Histogram of gradient magnitude, H_2^G. The magnitudes between 0 and a maximum value (100) are accumulated in 8 bins.

The constants $Z_1, \ldots, Z_4$ ensure that each histogram is normalized. Experimentally chosen values for the thresholds are: $T_\lambda = 0.1$ and $T_G = 5$.

4   Databases

Pain Database. We test the performance of the proposed system over the publicly available UNBC-McMaster Shoulder Pain Expression Archive Database [19], which contains face videos of patients suffering from shoulder pain as they perform motion tests of their arms. The movement is either voluntary, or the subject's arm is moved by the physiotherapist. Only one of the arms is affected by pain, but movements of the other arm are recorded as well to form a control set. The database contains 200 sequences of 25 subjects, totalling 48,398 frames.

The pain intensity is computed with the Prkachin-Solomon pain score [21], resulting in 16 discrete levels (0 to 15) obtained from the quantization of the elementary face AUs :

$$Pain = AU_4 + \max(AU_6, AU_7) + \max(AU_9, AU_{10}) + AU_{43} \qquad (9)$$

This score for the pain intensity is provided by the database creators, therefore acting as a reliable ground-truth for the pain intensity estimation. While in our work AUs are not computed separately for pain intensity estimation, yet eq. (9) explicitly confirms that databases build for AU recognition are relevant for pain intensity estimation.

Non Pain Database. Noting the limited number of persons available within the UNBC database, we extend the data used for learning with additional examples from a non-pain specific database, more precisely, the Cohn-Kanade database [15]. This contains 486 sequences from 97 persons and each sequence begins with a neutral expression and proceeds to a peak expression. The peak expression for each sequence is coded in the FACS system and is given an emotion label.

4.1   Landmark Localization and Areas of Interest

The UNBC-McMaster database is delivered with a set of 66 face landmarks extracted with an AAM tracker initialized in manual annotated key frames and applied on each image. In contrast, Cohn-Kanade was manually annotated[2].

We consider here that the specificity of the pain-related AUs is represented with only 22 landmarks from the given set of each of the databases. The reduced set is showed within figure 1 (a).

The UNBC landmarks are very accurate [19], yet their information is insufficient to provide robust pain estimation. In this sense, Kaltwang et al. [14] reported that using only points, for direct pain intensity estimation, a mean square error of 2.592 and a correlation coefficient of 0.363 is achieved.

Due to the specific nature of the AUs contributing to pain, and based on the 22 landmarks, we have selected 5 areas of interest, showed in figure 1 (a), as carrying potentially usefull data for pain intensity estimation.

Due to the variability of encountered head poses, we started by roughly normalizing the images: we ensured that the eyes were horizontal and the inter–ocular distance was always the same (i.e. 50). Out–of–plane rotation was not dealt with explicitly, but implicitly by the use of the histograms as features.

5   Transfer learning

The target database of the proposed system, UNBC, is highly extensive as number of frames, but is also rather limited with respect to the number of persons (only 25) and inter-person similarity. To increase the robustness of the proposed algorithm, a new mechanism for transfer learning is proposed.

We have inspired our work from the "self-taught learning" paradigm [22]. A source database, described by the un-labelled data $\mathbf{x}_u^{(1)}; \mathbf{x}_u^{(2)}; \ldots; \mathbf{x}_u^{(k)} \in \mathbb{R}^n$ is used to learn the underlying data structure so to enhance the classification over the labelled data of the target database: $\left\{ (\mathbf{x}_l^{(1)}; y^{(1)}); (\mathbf{x}_l^{(2)}; y^{(2)}); \ldots; (\mathbf{x}_l^{(m)}; y^{(m)}) \right\}$, where $\mathbf{x}$ is the data and y are labels. According to [22], the data structure could be learned by solving the following optimization problem:

---

[2] The landmarks were made public by G. Lipori, "Manual annotations of facial fiducial points on the Cohn Kanade database", LAIV laboratory, University of Milan, web url: lipori.di.unimi.it/download/gt2.html.
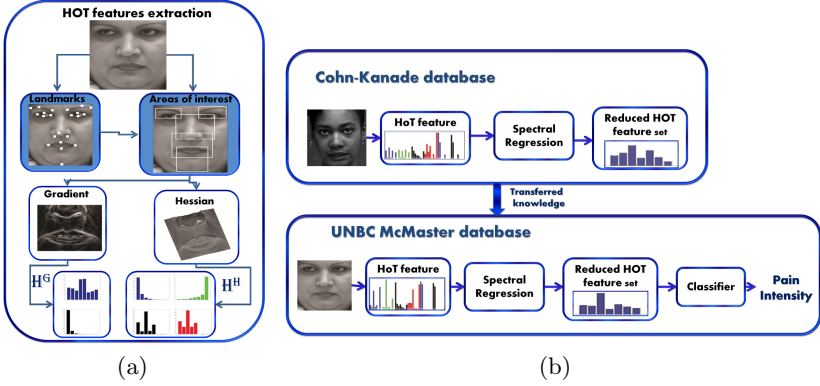
(a)   (b)

Fig. 1. (a) The features extraction procedure. (b) The transfer knowledge system. Data internal representation is computed on unlabelled data from Cohn-Kanade database to make use of the larger number of persons. The reduced data is fitted in order to predict pain intensity.

$$minimize_{\mathbf{b},a} \sum_i \left[\|\mathbf{x}_u^{(i)} - \sum_j a_j^{(i)} \mathbf{b}_j\|_2^2 + \beta\|\mathbf{a}^{(i)}\|_1 \right]; \;\; s.t.\|\mathbf{b}_j\|_2 \leq 1, \forall j \quad (10)$$

The minimization problem from eq. (10) may be interpreted as a generalization of the Principal Component Analysis concept[3] as it optimizes an overall representation, with the purpose of identifying the best minimum set of linear projections. Taking into account that the interest is in classification, we consider that: 1. the source database should be relevant to the classification task over the target database and 2. original features should form relevant clusters such that, 3. the optimization over the source database preserves local grouping. A modality to preserve the original data clustering is to compute the Locality Preserving Indexing with the similarity matrix $\mathbf{W}$:

$$W_{i,j} = \begin{cases} \frac{\mathbf{x}_i^T \mathbf{x}_j}{\|\mathbf{x}_i\|\|\mathbf{x}_j\|} & \text{if } \mathbf{x}_i \in N_p(\mathbf{x}_j) \vee \mathbf{x}_j \in N_p(\mathbf{x}_i) \\ 0 & otherwise \end{cases} \quad (11)$$

where $N_p(\mathbf{x}_i)$ contains the $p = 8$ closest neighbors of $\mathbf{x}_i$. The optimization ran over the similarity matrix, such that we solved the following regularized least squares problem over the unlabelled source database:

$$minimize_{\mathbf{B}=[\mathbf{b}_1...\mathbf{b}_t]} \sum_i \left(\left(\mathbf{b}_j^T \mathbf{x}_u^{(i)} - u_i^j\right)^2 + \alpha\|\mathbf{b_j}\|_2^2 \right); \;\; i = 1, \ldots, k \quad (12)$$

---

[3] PCA is retrieved by solving $minimize_{\mathbf{b},a} \sum_i \|\mathbf{x}_u^{(i)} - \sum_j a_j^{(i)} \mathbf{b}_j\|_2^2$ s.t. $\|\mathbf{b}_j\|_2 = 1$ and $b_1, \ldots b_T$ - orthogonal.

where u_i^j is the j-th element of the eigenvector $\mathbf{u}_i$ of the symmetrical similarity matrix $\mathbf{W}$. This process of extracting the data representation (eq. (11) and (12)) is known as *spectral regression* and it was introduced by Cai et al. [2]. A similar transfer learning method was proposed by Jiang et al. [13], with two core differences: data similarity is computed using a hard assignment compared to the soft approach from eq. (11) and unsupervised clustering was performed on the target database.

Finally, the labelled new data is obtained by classification of the projected vectors $\mathbf{z}_l^{(i)}$, determined as:

$$\mathbf{z}_l^{(i)} = \mathbf{B}\mathbf{x}_l^{(i)}, \forall i = 1, \ldots m \qquad (13)$$

where $\mathbf{B} = [\mathbf{b}_1 \ldots \mathbf{b}_t]$.

In our algorithm, the neutral image and respectively the images with the apex emotion from Cohn-Kanade database were the unlabelled data from the source database, while the UNBC McMaster was the target, labelled, database. The transfer learning process and the projection equation, (13), were applied independently on the Hessian based histograms, $[H_1^H, \ldots H_4^H]$ and, respectively, on the gradient based histograms $[H_1^G, H_2^G]$.

The overall proposed system, including the transfer learning procedure, is visually presented in figure 1 (b) and the method for HoT features extraction is presented in figure 1 (a).

6   Results

6.1   Testing and training

The used training-testing scheme was the same as in the works of Lucey et al. [20] or Kaltwang et al. [14]: leave one person out cross-validation. At a time, data from 24 persons was used for training and 1 person was used for testing.

As the number of images with positive examples (with a specific AU or with Pain label) is much lower than the one containing negative data, for the actual training the two sets were made even; the negative examples used were randomly selected. To increase the robustness of the system, 3 classifiers were trained in parallel with independently drawn examples and the system output was taken as the average of the classifiers.

For the actual discrimination of the pain intensity, we used the same model as in the case of similar works, [20], [14]. We used two levels of classifiers (late fusion scheme): first, each category of features was input into a Support Vector Regressor (SVR) (with radial basis kernel function, cost 4 and $\Gamma = 2^{-5}$). Landmarks were not spectrally regressed (i.e were not re-represented with eq. (13)). The results were fused together within a second level SVR.

Given a new UNBC image and the relevant landmarks position, the query to determine the pain intensity for that image takes approximately 0.15 seconds using single thread Matlab implementation on an Intel Xeon at 3.3 GHz, with classification performed using LibSVM.
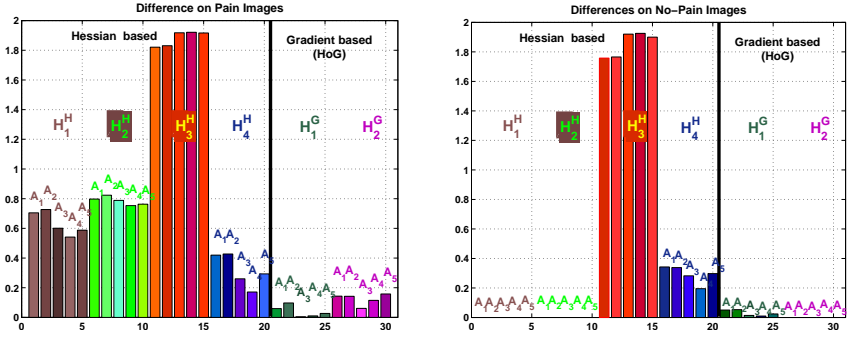
Fig. 2. Sum of absolute differences when comparing all images without pain and respectively with intense pain to a chosen no-pain reference image. Ideally, we aim for large values in the left plot and zeros in the right one. A_1 refers to the first area of interest (i.e. around the left eye), A_2 to the second one (around right eye), etc.

6.2   Experiments

First, we investigate the capabilities of the *HoT features* by considering the following *example*: we took the first frontal image without pain for each person and we considered its HoT features as reference; next, we computed the HoT features of all the images with a pain intensity higher than 5 and of all the images without pain for each person separately. We plotted the sum of absolute differences between the set considered as reference to the mentioned images with and without pain respectively. The results are presented in figure 2. Ideally, large values are aimed in the left plot and zeros in the right one. We noted that, for this particular example, the largest contribution in discriminating between pain and no-pain cases was due to Hessian based H_1^H and H_2^H histograms. Gradient based histograms lead to inconclusive differences in the case of intense pain, while H_3^H and H_4^H produced large values also for the no-pain case. The histogram type contribution on the entire database is presented in table 1.

Next, we evaluated the performance of the proposed approach for the task of continuous pain intensity estimation. The system performance was measured with the mean squared error $(\overline{\varepsilon^2})$ and with the Pearson correlation coefficient (ρ) between the predicted intensity and ground truth intensity.

The preferred implementation was by direct estimation of pain and the achieved values may be followed in table 2. Alternatively, one may consider as intermediate step the AU estimation, followed by pain prediction using equation (9); yet state of the art [14] showed that this method produced weaker results.

Feature contribution. In table 1 we reported the achieved relative accuracy increase (decrease) when only specific combinations of histogram types were used. Landmarks were not used for this experiment. As one can see, all the histograms contributed positively.

Table 1. Contribution of each of the histogram types used. We report the Pearson correlation coefficient when the mentioned type of histogram was removed. The reference is the right-most result (all histograms used). Thus, smaller is the value (i.e. larger is the decrease), higher is the contribution of the specific type of histogram.

Histogram removed	H_1^H	H_2^H	H_3^H	H_4^H	H_1^G	H_2^G	None - HoT
Correlation, ρ	0.330	0.368	0.355	0.358	0.350	0.224	0.41

Table 2. Accuracy of pain intensity estimation. We report the achieved results for various versions of features used and different methods of learning relevant features. Features used contained only Hessian based histograms (H_i^H - Hess), only gradient based histograms (H_i^G - Grad), both of them forming the so called Histogram of Topographical (HoT) features; the complete version (denoted by "All") contained landmarks and HoT. The relevant features were learned either by means of Spectral Regression (SR), or by PCA on the Cohn-Kanade database (CK) - via transfer learning or on the UNBC McMaster database. The Pain is estimated directly by the classifiers which were trained accordingly.

Learning	SR - CK				PCA-CK	SR-UNBC	PCA-UNBC	None	[14]
Feature	Hess	Grad	HoT	All	All	All	All	All	[14]
Measure	Mean Square Error ε^2								
Pain	3.76	4.67	3.35	**1.18**	1.17	1.26	1.16	1.21	*1.39*
Measure	Correlation, ρ								
Pain	0.25	0.34	0.41	**0.55**	0.48	0.53	0.49	0.53	*0.59*

The Influence of the Transfer Learning Method. In table 2, the overall performance, when various possibilities of transfer learning are considered, is also presented. The internal data representation was learned on the source database either by Spectral Regression (SR), or by PCA. Other considered alternatives were to perform no transfer at all, or to extract inner data representation using the data from the labelled UNBC database. The results showed that specifically relying on the similarity measure and taking into account a larger number of persons, the discrimination capability increased.

Comparison with state of art. While, as mentioned in section 2, there exist several methods that presented results on UNBC McMaster database, yet only Lucey et al. [20] and Kaltwang et al. [14] reported results on the entire database and with total separation between users when testing/trainig. However, as Lucey et al. [20] reported only binary detection results, and we are interested in intensity estimation, we will compare with the method proposed by Kaltwang et al.[14]. Their best reported performance is for the combination of DCT coefficients with histogram of LBP; the means square error was 1.386 while the correlation coefficient of 0.590. As one can see, while the correlation values are smaller, our mean square error is also smaller therefore reducing the chance for large false positives.

7   Discussion

In this paper we introduced the Histogram of Topographic features to describe faces. The addition of Hessian based terms allowed separation of various face movements and thus pain intensity levels. The robustness of the system was further enhanced by a new transfer learning method; it was inspired from the self taught learning paradigm and relied on preserving local similarity of the feature vectors as learned over a more consistent database in terms of persons, to ensure that relevant dimensions of the features are used in the subsequent classification process.

Regarding the addition of the actual features, while their individual contribution was rather small, they complemented each other well, as showed by the increase of the overall performance when all feature types were used.

The transfer learning from a database with larger number of persons increased the system robustness. More precisely, the solution that did not use the transfer procedure on some persons lead to better results, with the cost of providing smaller results on others that are more different from the remainder used for training. The transfer provided more consistent results overall, a fact which was proved by the entropy of the correlation coefficient increase from 9.01 to 9.26, enhancing the generalization with respect to person change.

The system provides indeed a number of failures. While the AU 43 (closing eyes), according to eq. (9), contributes to pain intensity, not all blinks are pain-related; the system, as in the case of [14], mistakenly associate all blinks with pain. These errors are rather small and can be easily filtered if the temporal sequence is considered. Other failures are in cases where the person's method of expressing pain is rather different from most of the others; for instance, the second person widely opens the eyes, instead of closing them, leading the system to produce false negatives. Other errors are related to the fact that the person is speaking during the test; false positives are, than, produced. At the end, we must point to the fact that an unexperienced user's opinion, if considering only individual, discontinuous, frames, produces the same kind of errors.

References

1. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-up robust features (SURF). CVIU 110(3), 346–359 (2008)
2. Cai, D., He, X., Han, J.: Spectral regression for efficient regularized subspace learning. In: ICCV. pp. 1–8 (2007)
3. Chen, J., Liu, X., Tu, P., Aragones, A.: Person-specific expression recognition with transfer learning. In: ICIP. pp. 2621–2624 (2012)
4. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR. pp. 886–893 (2005)

5. Deng, H., Zhang, W., Mortensen, E., Dietterich, T., Shapiro, L.: Principal curvature-based region detector for object recognition. In: CVPR. pp. 2578–2585 (2007)
6. Ekman, P., Friesen, W., Hager, J.: Facial action coding system (2nd ed.). Research Information; Salt Lake City (2002)
7. Florack, L., Haar-Romeny, B.M., Koenderink, J., Viergever, M.: Scale and the differential structure of images. Imag. and Vis. Comp. 10(6), 376–388 (1992)
8. Frangi, A., Niessen, W., Vincken, K., Viergever, M.: Multiscale vessel enhancement filtering. In: MICCAI. pp. 130–137 (1998)
9. Gawande, A.: The Checklist Manifesto: How to Get Things Right. Metropolitan Books (2004)
10. Hadjistavropoulos, T., Craig, K.: Pain: Psychological perspectives, chap. Social influences and the communication of pain, pp. 87–112. Erlbaum; New York (2004)
11. Hammal, Z., Kunz, M.: Pain monitoring: A dynamic and context-sensitive system. Pat. Rec. 45(4), 1265–1280 (1983)
12. Haralick, R., Watson, L., Laffey, T.: The topographic primal sketch. The Intl. J. of Robotics Research 2(1), 50–71 (1983)
13. Jiang, W., Chung, F.: Transfer spectral clustering. In: ECML PKDD. pp. 789–803 (2012)
14. Kaltwang, S., Rudovic, O., Pantic, M.: Continuous pain intensity estimation from facial expressions. In: ISVC. pp. 368–377 (2012)
15. Kanade, T., Cohn, J.F., Tian, Y.: Comprehensive database for facial expression analysis. In: FG. pp. 46–53 (2000)
16. Lee, W.T., Chen, H.T.: Histogram-based interest point detectors. In: CVPR. pp. 1590 – 1596 (2009)
17. Littlewort, G., Bartlett, M., Lee, K.: Faces of pain: Automated measurement of spontaneous facial expressions of genuine and posed pain. In: ICMI. pp. 15–21 (2007)
18. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV 60(2), 91–110 (2004)
19. Lucey, P., Cohn, J., Prkachin, K., Solomon, P., Matthews, I.: Painful data: The UNBC McMaster shoulder pain expression archive database. In: FG. pp. 57–64 (2011)
20. Lucey, P., Cohn, J., Prkachin, K., Solomon, P., Chew, S., Matthews, I.: Painful monitoring: Automatic pain monitoring using the UNBC-McMaster shoulder pain expression archive database. Imag. and Vis. Comp. 30, 197–205 (2012)
21. Prkachin, K., P. Solomon, P.: The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain. Pain 139, 267–274 (2008)
22. Raina, R., Battle, A., Lee, H., Packer, B., Ng, A.: Self-taught learning: Transfer learning from unlabeled data. In: ICML. pp. 759–766 (2007)
23. Tuytelaars, T., Mikolajczyk, K.: Local invariant feature detectors: A survey. Foundations and Trends in Computer Graphics and Vision 3(3), 177–280 (2008)
24. Wang, J., Yin, L.: Static topographic modeling for facial expression recognition and analysis. CVIU 108(1-2), 19–34 (2007)
25. Werner, P., Al-Hamadi, A., Niese, R., Walter, S., Gruss, S., Traue, H.: Towards pain monitoring: Facial expression, head pose, a new database, an automatic system and remaining challenges. In: BMVC (2013)
26. Zeng, Z., Pantic, M., Roisman, G., Huang, T.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE T. PAMI 31(1), 39–58 (2009)