Neonatal Facial Pain Assessment Combining Hand-Crafted and Deep Features

Luigi Celona(✉) and Luca Manoni

Department of Informatics, Systems and Communication,
University of Milano-Bicocca, viale Sarca, 336, 20126 Milano, Italy
luigi.celona@disco.unimib.it, l.manoni@campus.unimib.it

Abstract. In this paper we evaluate the combination of hand-crafted and deep learning-based features for neonatal pain assessment. To this end we consider two hand-crafted descriptors, i.e. Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG), and features extracted from two pre-trained Convolutional Neural Networks (CNNs). Experimental results on the publicly available Infant Classification Of Pain Expressions (COPE) database show competitive results compared to previous methods.

Keywords: Neonatal pain assessment · Hand-crafted features
Convolutional Neural Networks · Transfer learning · Features reduction
Feature fusion

1 Introduction

Pain is defined as the unpleasant sensory emotional experience caused by trauma, disease or injury. Valid and reliable assessment of pain is essential for both clinical trials and effective pain management. However, pain is a subjective, multifaceted experience that varies between individuals due to: personality, age, gender, social class, past experience, individual coping strategies, culture and appraisal of current circumstances. The best source of information for inferring pain is to examine infant's facial expressions.

Automatic systems for the detection of neonatal pain are proposed. Some of the previous methods try to predict pain focusing on heart rate variability or infant cry vocalizations [16,21]. However, such systems are quite impractical as the neonates would need to be tethered to sensors. A more practical approach is to use cameras and develop machine vision systems that unobtrusively and constantly scan facial expressions [1,13]. Brahnam *et al.* [5] proposed the first method for infant pain detection and developed the Infant Classification Of Pain Expressions (COPE) database of neonatal facial images. Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Support Vector Machines (SVM) were applied for the classification of four noxious stimuli: transport from one crib to another, air puff on the nose, friction from cotton and

S. Battiato et al. (Eds.): ICIAP 2017 International Workshops, LNCS 10590, pp. 197–204, 2017.
https://doi.org/10.1007/978-3-319-70742-6_19

alcohol rubbed on the lateral surface of the heel, and the puncture of a heel lance. However, in that paper experiments were conducted using a best-case scenario where samples of individual subjects were used both in testing and training sets. Brahnam *et al.* presented additional studies using the Infant COPE database [4,6,7]. These studies considered a more realistic scenario assuming that the classifiers would be trained on a set subjects and then applied to an unknown set of subjects. In [4], Brahnam *et al.* used raw pixel values of the grayscale image as feature vector, then they applied PCA for feature reduction and finally PCA, LDA SVMs and Neural Network Simultaneous Optimization Algorithm (NNSOA) were used to predict whether the neonatal face image contains a pain or nonpain expression. In [6], Brahnam *et al.* transformed raw pixel values of the gray-scale image using PCA or Discrete Cosine Transform (DCT). Then two methods were adopted for feature reduction: sorting by variance and Sequential Forward Floating Selection (SFFS); and four classifiers were evaluated: PCA, LDA, SVMs and NNSOA. Nanni *et al.* [18] compared several texture descriptors based on LBP in terms of AUC and proposed the ELongated Ternary Pattern (ELTP) and the Improved Local Ternary Pattern (ILTP) for infant pain classification. Recently, Mansor *et al.* [17] developed a system robust under different illumination levels. They achieved this results by altering illumination in original images, by estimating illumination thanks to the Multi Scale Retinex (MSR) algorithm for shadow removal, and finally by using LBP as features. Then Gaussian or Nearest Mean Classifier has been used as classifiers.

Feature fusion have been demonstrated to be very effective for various computer vision tasks. These methods generally involve the use of multiple handcrafted features, such as Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG). However, features extracted from pre-trained Convolutional Neural Networks (CNNs) have recently been proven to be robust than the handcrafted features. This paper investigates the use of feature fusion for infant pain assessment. The experiments are conducted on the *infant Classification Of Pain Expression (COPE)* database which contains facial images of several neonates captured after distinct stimuli.

2 Infant Classification of Pain Database

The Infant COPE (Classification Of Pain Expressions) database [4–6] is the only public available database for infants pain detection. It contains a total of 204 color facial images of 26 Caucasian neonates (13 boys and 13 girls) ranging in age from 18 h to 3 days. The facial expressions of the newborns were photographed in one session while the infants were experiencing four distinct stimuli in the following sequence:

1. **Rest/Cry** - After being transported from one crib to another
2. **Air puff** - Exposition to a puff of air emitted from a squeezable plastic camera lens cleaner
3. **Friction** - Friction on the heel using cotton wool soaked in alcohol
4. **Pain** - Puncture of a heel lance

(a) Rest (b) Cry (c) Air puff (d) Friction

(e) Pain

Fig. 1. Sample images from the COPE database.

Among the 204 images in the database, 67 are rest, 18 are cry, 23 are air puff, 36 are friction, and 60 are pain. Figure 1 shows some samples from the COPE database.

For the pain assessment problem, all images of the stimuli are divided into two categories: nonpain and pain. The set of nonpain images combines rest, cry, air puff and friction stimuli, and consists of 140 images. The set of pain images being a collection of the remaining 60 images.

3 The Proposed Method

The main steps of the proposed method for the facial neonatal pain assessment are shown in Fig. 2. Given an image: (1) we detect and align the original facial image using an affine transformation; (2) we extract both hand-crafted and deep features; (3) we apply Principal Component Analysis (PCA) as feature reduction algorithm and concatenate features for feature fusion; (4) finally, we train a linear Support Vector Machine (SVM) using fused features. What follows in this section is a detailed description of the previous steps.

3.1 Pre-processing

In the pre-processing step, we use a state-of-the-art algorithm for face detection and landmarks estimation [14]. The detected faces are rotated and aligned thanks to an affine transformation based on five landmarks, i.e. eyes corners, nose tip and mouth corners. Facial images are then obtained by cropping and scaling the transformed images to 224×224 pixels.

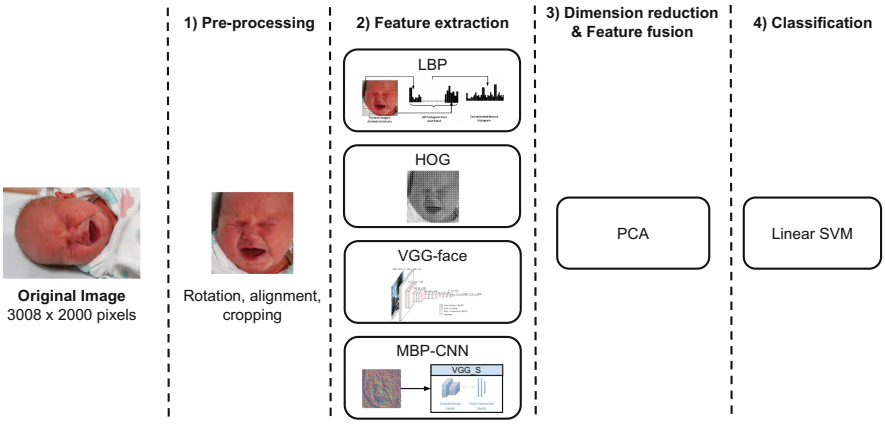

Fig. 2. Pipeline of the proposed method.

3.2 Feature Extraction

Hand-crafted features based methods achieved state-of-the-art performance on many machine vision tasks, such as object classification, object detection, texture classification and visual search in past few years [2,8–11].

Local Binary Patterns (LBP) [19] is a powerful local texture operator with a low computational complexity, and low sensitivity to changes in illumination. The LBP operator is computed by thresholding the value of the neighbors of each pixel with the value of the pixel itself. More in detail, it consists in the binary difference between the gray value of a pixel x and the gray values of its P neighborhood placed on a circle of radius of R. In this work we use a *uniform* LBP setting $P = 8$ and $R = 1$, resulting in a histogram of 59-bins. To better capture local texture, we divide the face into 25 (5×5) non-overlapping regions to extract LBP histograms. The LBP features extracted from each sub-region are concatenated into a single, spatially enhanced feature histogram. Furthermore we retain the color information by computing LBP histograms for each color channel and then by concatenating them. Thus, the resulting feature vector has length of 4,425 (59 bins $\times$ 25 regions $\times$ 3 channels).

Histogram of Oriented Gradients (HOG) [12] counts the occurrence of gradient orientation in local regions of an image. In this paper we use 2×2 blocks of 8×8 pixel cells with an overlap of half the block and histograms of 9 bins evenly spread from 0 to 180 degrees. The dimensions of such HOG descriptor on a 224×224 gray-scale image is 26,244 (729 regions $\times$ 4 blocks $\times$ 9 bins).

Transfer learning strategies by using CNN as feature extractor have proven to be more powerful than hand-crafted features on several application, e.g. object recognition and detection. They are based on the fact that a CNN pre-trained on a large dataset could be exploited for another task by reusing the learnt parameters of the source task [22]. Specifically, we feed the pre-trained CNN with a sample and then we use activation of a desired layer as features. In this paper,

we consider the features extracted from the *VGG face* [20] and from the *Mapped LBP+CNN* (MBPCNN) [15]. VGG face consists in a VGG-16 network architecture trained for the recognition among 2,622 celebrities. Instead, MBPCNN involves the use of mapped LBP features as input of a VGG_S network architecture for emotion recognition in the wild. For both CNNs we obtain the feature vector by removing the final last two fully connected layers and the softmax layer. The length of the feature vector is 4,096.

3.3 Feature Reduction and Fusion

In this step each feature vector is reduced to 175-dimension via Principal Component Analysis (PCA) and L2-normalized. Finally, we perform a feature fusion by a linear concatenation of the reduced features. The result is a single feature vector with length equal to $175 \times L$, where L is the number of considered feature vectors.

3.4 Classification

Support Vector Machine (SVM) with linear kernel is used for pain/nonpain classification. The regularization parameter C is determined using a grid search.

4 Experimental Setup

We evaluate the proposed method on the infant COPE database. The evaluation protocol we used consists in dividing the images by subject: the testing set contains images of a given subject and all the other subjects used for the training. This procedure has been repeated for each one of the 26 subjects of the COPE

Table 1. Results on the COPE database in terms of average accuracy over the 26 subjects.

Features	PCA	Accuracy (%)
LBP		77.52
HOG		81.75
VGG face		82.42
MBPCNN		81.53
HOG+LBP		82.80
HOG+LBP	✓	81.98
VGG face+MBPCNN		82.56
VGG face+MBPCNN	✓	**83.78**
HOG+LBP+VGG face+MBPCNN		81.96
HOG+LBP+VGG face+MBPCNN	✓	82.95

database. Table 1 reports the results for all of the experiments. The performance measure adopted in the experiments is the average accuracy over the 26 subjects.

We investigate the performances obtained considering a single feature at time. Specifically, we run a set of experiments using: LBP features extracted from a RGB image divided into sub-regions; HOG computed on gray-scale image; features extracted from the pre-trained VGG face; and features extracted from the pre-trained MBPCNN. According to the results it is possible to see that we achieve the best performance thanks to the 4096-dimension features extracted using the pre-trained VGG face. The average accuracy is equal to 82.42%.

In another set of experiments we fuse the pair of hand-crafted features (i.e. LBP+HOG) and the pair of deep features (i.e. VGG face and MBPCNN). More in detail, we run an experiment fusing LBP and HOG features by simply concatenating the two feature vectors and obtaining a vector of length 30,669. For this experiment the resulting accuracy is 82.80%. In the second experiment we first reduce the dimensionality of both feature vectors to 175 and then we concatenate the two vectors obtaining a vector of 350 features. The resulting average accuracy is equal to 81.98%. For the remaining two experiments we fuse the 4096-dimension feature vectors extracted from the VGG face and MBPCNN: in the first case we only concatenate the two feature vectors and obtain an accuracy of 82.56%; in the second case, at first each feature vector is reduced to 175 dimensions and then they are concatenated into a 350-dimension feature vector. This last experiment obtains the best average accuracy of 83.78%. Figure 3 depicts some misclassified faces obtained in the aforementioned experiment: Fig. 3a reports faces labeled as nonpain in the COPE database, but classified as pain; instead Fig. 3b shows some examples of faces labeled as pain, but

Fig. 3. Some misclassified samples: (a) Faces labeled as nonpain in the COPE database that the proposed method classifies as pain, (b) Faces labeled as pain in the COPE database that the proposed method classifies as nonpain.

classified as nonpain by our approach. From these images it is possible to see that some classification errors are due to incorrect labels in the dataset.

The last set of experiments consists in fusing all the considered features. For the first experiment we concatenate features and obtain a 38861-dimension feature vector. Instead, for the second experiment we apply PCA on each feature vector and then concatenate the four 175-dimension feature vectors. These experiments obtain an accuracy respectively of 81.96% and 82.95%.

To the best of our knowledge the best performance on pain assessment using the COPE database is obtained by Brahnam *et al.* [4]. For the sake of comparison we have reimplemented their method obtaining an average accuracy equal to 78.94%.

5 Conclusion

In this work we proposed a pipeline for pain assessment in newborn face images. The proposed method involves the fusion of both hand-crafted and deep features. More in detail, faces are detected using a face detector and then aligned using an affine transformation. LBP and HOG have been considered as hand-crafted features, pre-trained VGG face and MBPCNN have been used for deep feature extraction. Experimental results on the COPE database show the effectiveness of the proposed solution. As a future work, we plan to investigate other features and more powerful mechanisms [1,3] for feature selection and fusion.

References

1. Bianco, S., Celona, L., Schettini, R.: Robust smile detection using convolutional neural networks. J. Electron. Imaging **25**(6), 063002 (2016)
2. Bianco, S., Mazzini, D., Pau, D.P., Schettini, R.: Local detectors and compact descriptors for visual search: a quantitative comparison. Digital Sig. Process. **44**, 1–13 (2015)
3. Bianco, S., Schettini, R.: Adaptive color constancy using faces. IEEE Trans. Pattern Anal. Mach. Intell. **36**(8), 1505–1518 (2014)
4. Brahnam, S., Chuang, C.F., Sexton, R.S., Shih, F.Y.: Machine assessment of neonatal facial expressions of acute pain. Decis. Support Syst. **43**(4), 1242–1254 (2007)
5. Brahnam, S., Chuang, C.F., Shih, F.Y., Slack, M.R.: Machine recognition and representation of neonatal facial displays of acute pain. Artif. Intell. Med. **36**(3), 211–222 (2006)
6. Brahnam, S., Nanni, L., Sexton, R.: Introduction to neonatal facial pain detection using common and advanced face classification techniques. In: Yoshida, H., Jain, A., Ichalkaranje, A., Jain, L.C., Ichalkaranje, N. (eds.) Advanced Computational Intelligence Paradigms in Healthcare-1, pp. 225–253. Springer, Heidelberg (2007)
7. Brahnam, S., Nanni, L., Sexton, R.S.: Neonatal facial pain detection using NNSOA and LSVM. In: IPCV, pp. 352–357 (2008)
8. Cusano, C., Napoletano, P., Schettini, R.: Illuminant invariant descriptors for color texture classification. In: Tominaga, S., Schettini, R., Trémeau, A. (eds.) CCIW 2013. LNCS, vol. 7786, pp. 239–249. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-36700-7_19

9. Cusano, C., Napoletano, P., Schettini, R.: Intensity and color descriptors for texture classification. Proc. SPIE **8661**, 866113 (2013)
10. Cusano, C., Napoletano, P., Schettini, R.: Combining local binary patterns and local color contrast for texture classification under varying illumination. JOSA A **31**(7), 1453–1461 (2014)
11. Cusano, C., Napoletano, P., Schettini, R.: Local angular patterns for color texture classification. In: Murino, V., Puppo, E., Sona, D., Cristani, M., Sansone, C. (eds.) ICIAP 2015. LNCS, vol. 9281, pp. 111–118. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-23222-5_14
12. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 886–893. IEEE (2005)
13. Florea, C., Florea, L., Butnaru, R., Bandrabur, A., Vertan, C.: Pain intensity estimation by a self-taught selection of histograms of topographical features. Image Vis. Comput. **56**, 13–27 (2016)
14. King, D.E.: Dlib-ml: a machine learning toolkit. J. Mach. Learn. Res. **10**, 1755–1758 (2009)
15. Levi, G., Hassner, T.: Emotion recognition in the wild via convolutional neural networks and mapped binary patterns. In: Proceedings of ACM International Conference on Multimodal Interaction (ICMI), November 2015
16. Lindh, V., Wiklund, U., Håkansson, S.: Heel lancing in term new-born infants: an evaluation of pain by frequency domain analysis of heart rate variability. Pain **80**(1), 143–148 (1999)
17. Mansor, M.N., Rejab, M.N.: A computational model of the infant pain impressions with Gaussian and nearest mean classifier. In: 2013 IEEE International Conference on Control System, Computing and Engineering (ICCSCE), pp. 249–253. IEEE (2013)
18. Nanni, L., Brahnam, S., Lumini, A.: A local approach based on a local binary patterns variant texture descriptor for classifying pain states. Expert Syst. Appl. **37**(12), 7888–7894 (2010)
19. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
20. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: British Machine Vision Conference (2015)
21. Petroni, M., Malowany, A.S., Johnston, C.C., Stevens, B.J.: Identification of pain from infant cry vocalizations using artificial neural networks (ANNs). In: SPIE's 1995 Symposium on OE/Aerospace Sensing and Dual Use Photonics, pp. 729–738. International Society for Optics and Photonics (1995)
22. Razavian, A.S., Azizpour, H., Sullivan, J., Carlsson, S.: CNN features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 806–813 (2014)