Machine Vision and Applications (2019) 30:933–944
https://doi.org/10.1007/s00138-018-0968-1

SPECIAL ISSUE PAPER

Video-based discomfort detection for infants

Yue Sun[1] · Caifeng Shan[2] · Tao Tan[1] · Xi Long[1] · Arash Pourtaherian[1] · Svitlana Zinger[1] · Peter H. N. de With[1]

Received: 6 April 2018 / Revised: 20 June 2018 / Accepted: 29 June 2018 / Published online: 13 August 2018

Abstract

Infants are particularly vulnerable to the effects of pain and discomfort, which can lead to abnormal brain development, yielding long-term adverse neurodevelopmental outcomes. In this study, we propose a video-based method for automated detection of their discomfort. The infant face is first detected and normalized. A two-phase classification workflow is then employed, where Phase 1 is subject-independent, and Phase 2 is subject-dependent. Phase 1 derives geometric and appearance features, while Phase 2 incorporates facial landmark-based template matching. An SVM classifier is finally applied to video frames to recognize facial expressions of comfort or discomfort. The method is evaluated using videos from 22 infants. Experimental results show an AUC of 0.87 for the subject-independent phase and 0.97 for the subject-dependent phase, which is promising for clinical use.

Keywords Infant discomfort · Face detection · Discomfort/stress detection · Facial expression recognition

1 Introduction

Monitoring of the wellbeing of infants is an important topic for hospitals, parents and the infants themselves, for instance, for their brain development. Frequent pain or discomfort in newborn infants, who are at a time of physiological immaturity and undergo rapid brain development, can cause complications, such as delay in cognitive and motor development [17]. Cumulative pain-related discomfort may also contribute to abnormal brain development, which can yield long-term adverse neurodevelopmental outcomes [4,6,32,33,42]. Neurobiological vulnerability to pain in newborn infants is well established, due to their lower pain threshold, sensitization from repeated pain, and immature systems for maintaining homeostasis [12,13]. Therefore, in the Neonatal Intensive Care Unit (NICU), continuous discomfort or pain assessment for newborn infants is highly desired, since it helps caregivers understand the severity of the infants situation and develop appropriate treatments.

Self-reporting is currently considered to be the gold standard for pain assessment among patients and is the most reliable indicator of the existence and intensity of acute pain and discomfort [40]. Assessment scales that rely on patient self-reporting are commonly used to measure the intensity of pain. However, infants cannot report verbally their pain or discomfort and must rely on healthcare professionals to recognize their behavioral or physiological signs suggesting pain and discomfort [20]. In order to assist healthcare professionals in detecting the level of pain or discomfort in an infant or child, several pain/comfort-scale forms have been developed. For example, comfort scale [3] is developed for assessing distress in patients of pediatric intensive care units. The Objective Pain Scale (OPS) has been utilized with infants and children from birth to three years of age [30]. Premature Infant Pain Profile (PIPP) [39] is a multidimensional measure developed to assess acute pain in preterm and term infants.

✉ Caifeng Shan
caifeng.shan@philips.com

Yue Sun
Y.Sun1@tue.nl

Tao Tan
T.Tan1@tue.nl

Xi Long
xi.long@philips.com

Arash Pourtaherian
A.Pourtaherian@tue.nl

Svitlana Zinger
S.Zinger@tue.nl

Peter H. N. de With
P.H.N.de.With@tue.nl

1 Eindhoven University of Technology, De Zaale, 5612AZ Eindhoven, The Netherlands

2 Philips Research, High Tech Campus 34, 5656AE Eindhoven, The Netherlands

Each scale is scored by healthcare professionals after observing infants for about 2 minutes.

The current manual monitoring by healthcare professionals is of high cost, time-consuming and subjective in infant pain/discomfort assessment [9,34]. More crucially, infants are only observed a few times a day ("spot measurement") without continuous monitoring, which may leave many discomfort moments unnoticed. The intermittent assessment might lead to misdiagnosis and over/under treatment. In this regard, an automatic discomfort or pain assessment method for infants becomes a necessity. The objective of this paper is to develop an automated video-based discomfort detection system for infants that is able to, e.g., alert clinical staff immediately when infants start suffering from discomfort. Our system provides continuous monitoring for infants, which replaces the current intermittent manual observation.

Facial expression is one of the most common indicators of discomfort and pain. The Primal Face of Pain (PFP), shown in Fig. 1, is a universal facial expression, associated with pain, which is hardwired and presents at birth [35]. In this work, we propose an automatic discomfort detection system for infants by analyzing their facial expressions in videos. After face detection and normalization in videos, geometric and appearance features are extracted from the facial Region of Interest (ROI) for discomfort detection, using a Support Vector Machine (SVM) [8,15] classifier. In clinical practice, for a given infant, his/her comfort face images can be easily collected. Our contribution is in deploying two types of detection. Specifically, in addition to subject-independent detection, we further introduce template matching based features for *subject-dependent* detection, which provides more reliable performance. Tested on a dataset of 22 newborn infants collected at the hospital, the proposed method achieves promising performance exceeding the subject-independent method.

The remainder of this paper is organized as follows. In Sect. 2, related work on pain and discomfort detection is described. Sect. 3 elaborates our method, and Sect. 4 explains

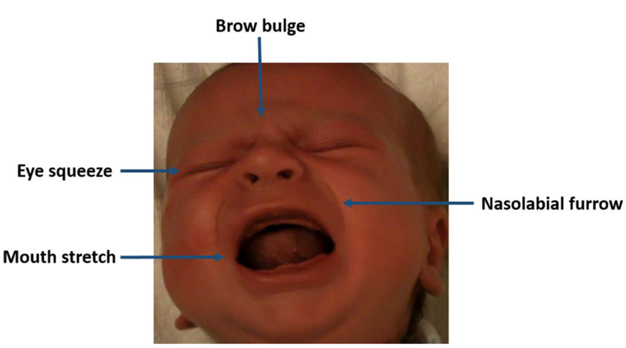

Fig. 1 Primal face of pain, involving brow bulge, eye squeeze, and a horizontally stretched open mouth with deepening of the nasolabial furrow

the experimental results, with some discussions in Sect. 5. Finally, Sect. 6 concludes the paper.

2 Related work

In the past years, there has been an increasing interest in pain assessment because of important applications [18]. Various approaches have been devoted to assess pain based on physiological indicators, for instance, vital signs such as heart rate (HR), heart rate variability (HRV), respiratory rate (RR), blood oxygen saturation (SpO2), body temperature, and blood pressure. The signs can be measured and used to assess the person's level of physical functioning. Lindh et al. [25] presented an approach to assess infants pain by frequency-domain analysis of HRV during heel lancing, which showed an increase in low-frequency power in the response of preterm infants to heel stick compared to baseline. Acharya et al. [1] detected cardiac abnormalities by classifying cardiac rhythms using an artificial neural network and fuzzy relationships, which achieved an accuracy level of 80–85%. However, vital signs such as HR and RR are currently measured using techniques including electrocardiograms (ECG) and pulse oximetry, which require contact with the patient's skin. Attaching the sensors to infant skin adds an extra burden to infants, compared to the contactless method using videos.

Besides the physiological-based approaches, there is another category of methods, which assess pain or discomfort based on behavior analysis. Existing behavioral-based approaches to evaluate infant pain can be based on facial expression, crying sound, and body motion. Infant cry is a common sign of discomfort, hunger, or pain. For classifying crying sound, Mima et al. [29] presented a method that analyzes baby cries in spectrography, and classifies them as cries due to pain, sleeping, hunger, etc. The overall accuracy of the proposed method was 85%. Up to this point, very few studies on pain assessment based on body movements have been published [43,44].

Lots of attention was paid to facial expressions in adults. Shan et al. [37] empirically evaluated facial representation based on statistical local features, Local Binary Patterns (LBP), for person-independent facial expression recognition and illustrated that LBP features are effective and efficient for facial expression recognition. They achieved a recognition rate of 91.4% for 7-class (anger, disgust, fear, joy, sadness, surprise, and neutral) facial expression by using Boosted-LBP based SVM with RBF kernel. Kotsia et al. [24] achieved a recognition accuracy of 99.7% for facial expression recognition, using the proposed multiclass SVMs and 95.1% for facial expression recognition based on a set of chosen facial Action Units (AUs). Kharghanian et al. [22] applied a hierarchical unsupervised feature learning approach to extract the features needed for pain detection from facial images using a convolutional deep belief network and achieved near 95% for the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC). Ashraf et al. [5] explored an approach for automatically recognizing acute pain with Active Appearance Models (AAM). The shape and appearance components represented by AAM were further decoupled to separate features. Finally, SVM was employed for classification. The method was evaluated on 15,761 frames from the UNBC-Mcmaster shoulder pain database, which showed a frame-level accuracy value of 82.4% for detecting pain frames, and a false positive rate of 30.1%. One of the challenges in this task is to align faces together for more effective feature extraction. Lucey et al. [28] showed that the AAM can deal with patient movements and can achieve significant improvements in both the facial AU and pain detection performance. Using the UNBC-McMaster database, the AUC was 0.847 when combining similarity-normalized shape features (SPTS) with similarity-normalized appearance features (SAPP) and canonical normalized appearance features (CAPP). Littlewort et al. [26] applied an automated facial expression recognition system to spontaneous facial expressions of pain. A set of 20 detectors from the facial action coding were extracted and passed on to a classification stage, in which a classifier was trained to detect the difference between expressions of real pain and fake pain. The automated system obtained an accuracy of 88% for subject-independent discrimination of real versus fake pain.

Most of the existing methods for automatic pain assessment based on facial expressions focus on adults. However, the methods designed for assessing adult pain might not have similar performance on the infants, because the facial morphology and dynamics vary between infants and adults as reported in [16]. Sikka et al. [38] presented a Facial Action Coding System(FACS)-based method to describe children's facial expressions of pain. The model detection of pain versus no-pain achieved the AUC of 0.84–0.94 in both ongoing and transient pain conditions on the videos from 50 children. One challenge of FACS-based methods is the extensive time required for labeling AUs in each video frame. Fotiadou et al. [14] proposed an infant discomfort detection system, based on the AAM. Their system was evaluated in 15 videos of 8 infants, yielding a 0.98 AUC performance. However, the AAM mesh has to be initialized for each individual baby by identifying a set of landmarks manually.

3 Methods

We propose a video-based automatic discomfort detection system. The method of discomfort detection involves 2 key steps: (1) face detection and face normalization, and (2) feature extraction, and facial expression classification to discriminate infant status into comfort or discomfort. Figure 2 shows the overall processing chain. Each block of the proposed system is described in more detail in the following subsections.

3.1 Face detection and normalization

We start with extracting infant faces from each frame of videos. The entire face detection and normalization workflow

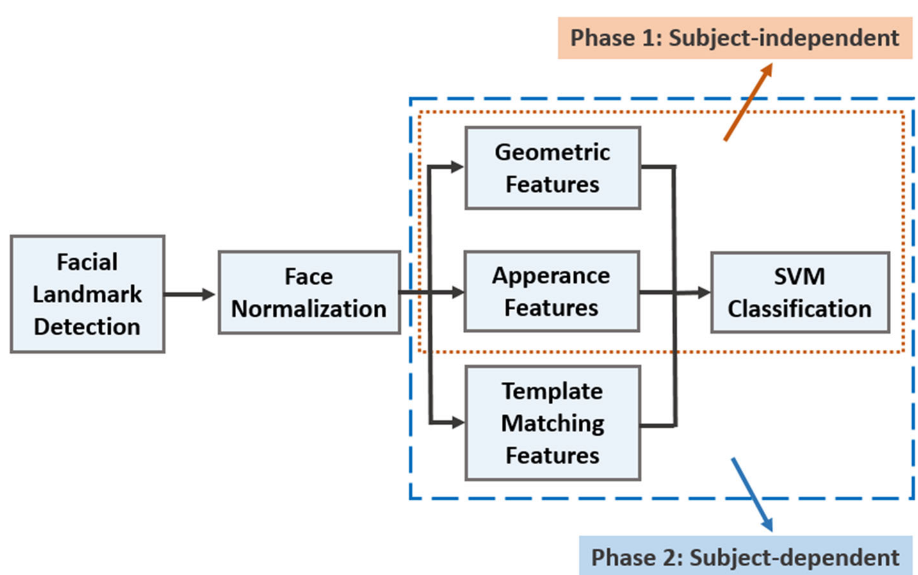

Fig. 2 Overall workflow for discomfort recognition, where the orange-dotted rectangle indicates subject-independent discomfort detection (Phase 1), and the blue-dashed rectangle is for subject-dependent discomfort detection (Phase 2)

generates the face ROI as input for the next step of discomfort/comfort classification. Given an input video frame of a face image, sixty-eight facial landmarks are first localized using the dlib face landmark detector [23], implemented by Kazemi *et al.* [21]. We utilize the dlib face landmark detector because it detects the face with a rich number (68) of landmarks [19] and its implementation is also efficient[10]. The 68 landmarks are points on the face such as the corners of the eyes, mouth, along the eyebrows, along the boundary of face, and so forth. See Fig. 3 for an original face image (Fig. 3a) with the corresponding normalized face ROI and 68 facial landmarks (Fig. 3b).

Once the 68 landmarks are identified, we select the central point between two inner eye corner points as the point to rotate the image. The image is rotated to the position that the two eye corner points are in a same horizontal line. Thus, this step corrects the rotation variance of faces. We select Landmark 1 as the leftmost point, Landmark 17 as the rightmost, and Landmark 9 as the bottommost points to define the left, right, and bottom boundaries of the face ROI (See Fig. 3b for the landmarks). We measure the distance between Landmark 9 and the mid point of two inner eye corners. The top boundary of the ROI is defined to be the horizontal line that has the same distance from the mid inner eye point as Landmark 9. A margin of 20 pixels are added to all the boundaries to cover the whole infant face, and avoid loss of facial information. Finally, all face images are cropped and resized to the size of 500×375 pixels.

3.2 Phase 1: subject-independent discomfort detection

Geometric and appearance features are extracted from the facial ROI for discomfort detection using the SVM classifier.

3.2.1 Geometric features

In general, when infants start suffering from discomfort, they tend to squeeze their eyes and stretch their mouths, as shown in Fig. 1. In order to extract relevant features, we calculate the areas of eye and mouth. We count the number of pixels inside the polygons surrounded by the landmarks of the left eye, right eye, outer lip contour and inner lip contour, respectively (See Fig. 4). These four area sizes are considered as four geometric features. Figure 5 shows the distributions of geometric features for comfort and discomfort cases of our dataset.

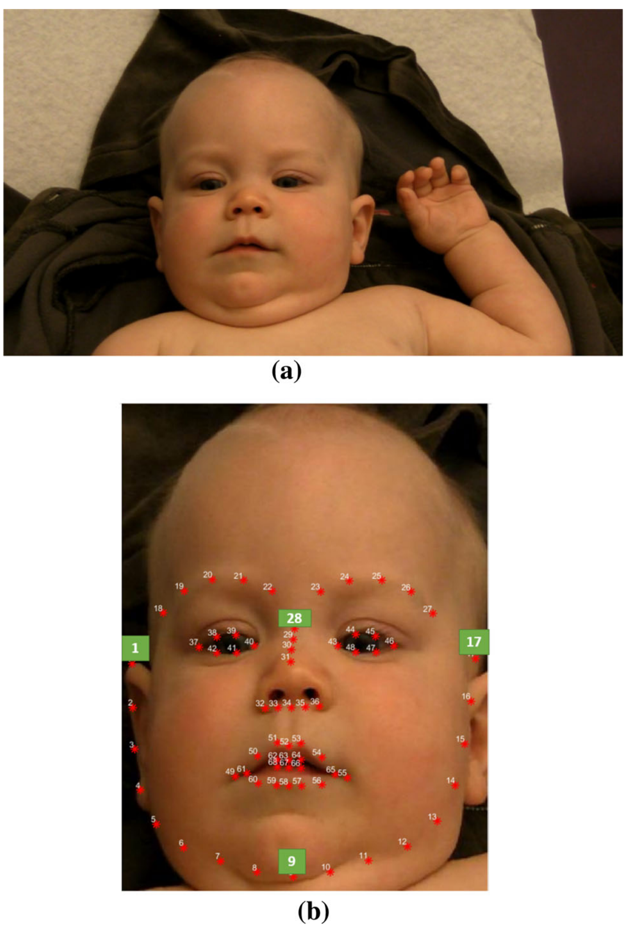

Fig. 3 **a** Example of face image, **b** corresponding normalized face ROI and 68 facial landmarks. The highlighted landmarks 1, 9, and 17 are used for defining the face ROI. The 28th landmark is employed in Phase 2 for further aligning sets of landmarks from different frames

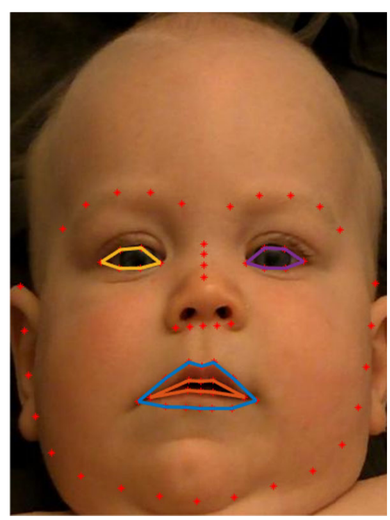

Fig. 4 Polygons consist of facial landmarks for eye and mouth area calculation, where yellow polygon outlines left eye, purple for right eye, blue for outer lip contour, and orange for inner lip contour (color figure online)

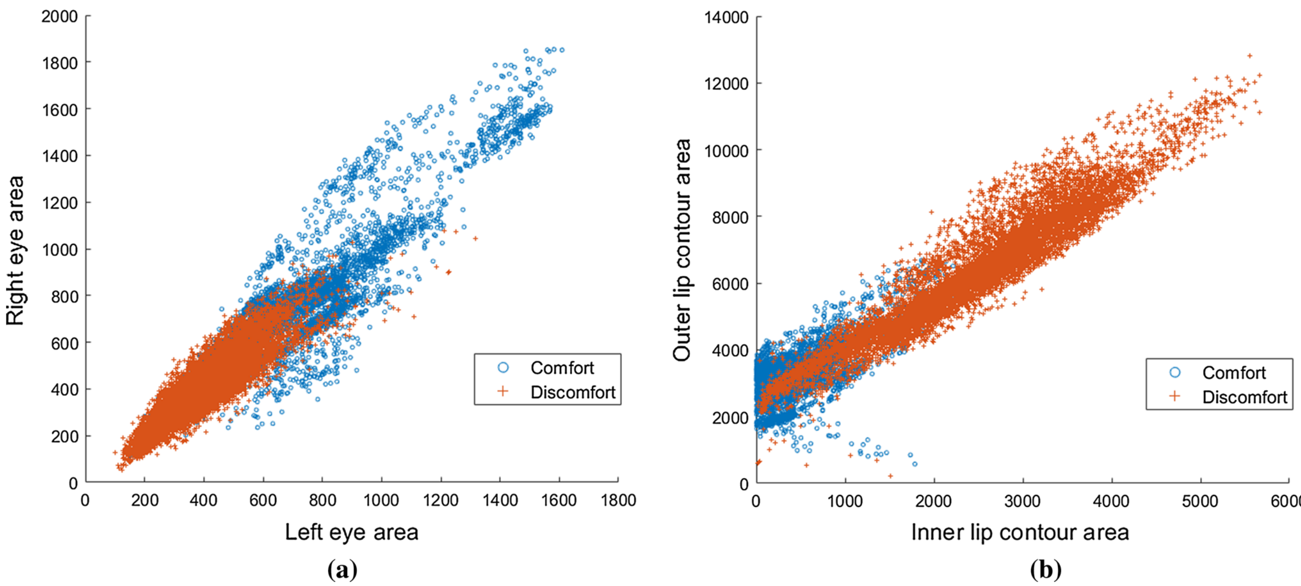

Fig. 5 Scatter plots of geometric feature values of comfort (blue-circled dot) and discomfort (orange plus sign) cases for all infant face images in our dataset. Subfigure **a** plots the feature values of left and right eye area. Subfigure **b** presents the values of the inner and outer lip contour area (color figure online)

Fig. 6 Example of a comfort (left) and a discomfort (right) case from an infant, which shows the face appearance difference between comfort and discomfort status

3.2.2 Appearance features

The appearance of infant face is changing when experiencing discomfort. For example, as shown in Fig. 1, brow bulge and nasolabial furrow become apparent in the discomfort faces, which leads to texture changes in the forehead and between the nose and upper lip area. Figure 6 compares the face appearance of comfort and discomfort status of one infant. Different texture descriptors have been exploited for facial expression recognition. Two methods that have proven to be effective for facial representation are Histogram of Oriented Gradients (HOG) [11,41] and Local Binary Patterns (LBP) [2,31,36]. As the histogram of gradient descriptor, HOG has been widely utilized to capture edge or local shape

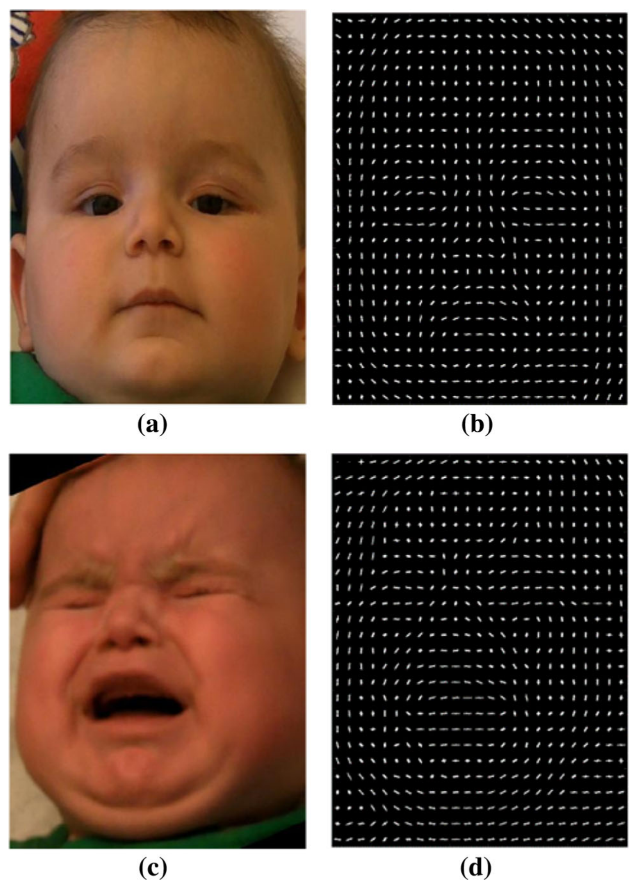

Fig. 7 Examples of HOG (8 orientations) processing on a comfort (**a**) and a discomfort face (**c**) images, where **b** and **d** are the corresponding HOG of **a** and **c**

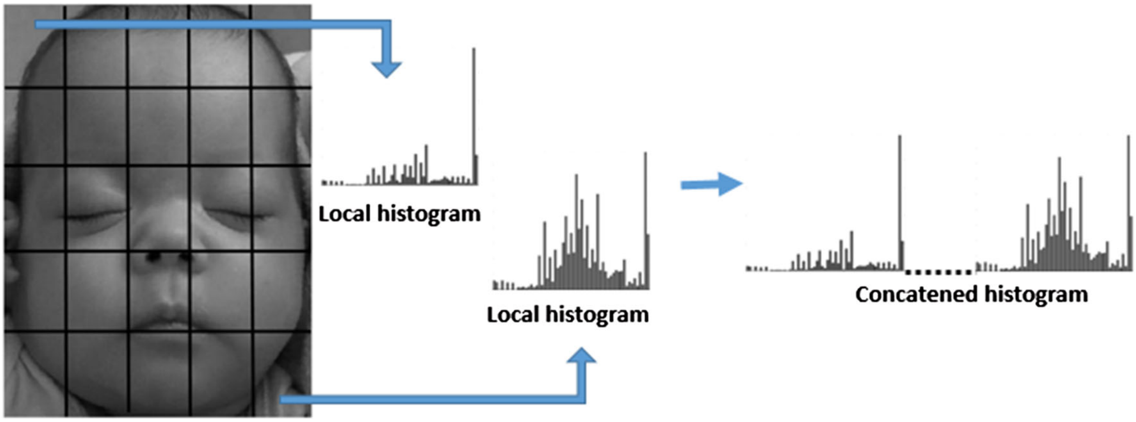

Fig. 8 The infant face image (500 × 375 pixels) is divided into 5 × 5 subregions from which LBP histograms are extracted and concatenated into a single, spatially enhanced feature histogram

information [27]. LBP is invariance to monotonic gray-level changes and is highly discriminative, which makes it suitable for demanding image analysis tasks such as object detection. In this work, HOG and LBP are combined as the feature set to describe faces for discomfort detection. In detail, for the HOG calculation, the input face is divided into 20 × 15 equal non-overlapping regions. For each region, HOG is computed (See Fig. 7). The final HOG vector features are formed by concatenation of HOG vectors of regions. The length of HOG feature vector is 5,625. For LBP, each face image is divided into blocks of 100 × 75 pixels to which the uniform LBP operator [31] is applied, and LBP histograms are extracted and concatenated into a single, spatially enhanced feature histogram (see Fig. 8). The length of LBP feature vector is 1,475.

3.2.3 Classification

The large size of HOG/LBP feature vector limits the number of training samples and increases the computation cost in classification. Feature selection is one of the important and frequently used techniques in data preprocessing for data mining [7]. The selection reduces the number of features and removes irrelevant, redundant data and even noisy components. In our work, feature selection is used to reduce the dimensionality of the input feature space and thus enable the subsequent use of classification algorithms. The criterion of AUC is chosen to identify relevant significant features, which means features selected by optimizing the AUC.

Finally, we adopt SVM [8,15] with a radial basis function (RBF) as the kernel, to recognize facial expressions using the selected features. We employ the SVM implementation in MATLAB (Mathworks, Natick, MA, USA) for the two-class classification. Leave-one-subject-out cross-validation is used for the experiments. The ROC is plotted to evaluate the performance with the value of the AUC. The labeling error rate is also measured and reported as an additional metric. Moreover, the quality of the classification is measured from a confusion matrix, which records correctly and incorrectly recognized faces for each class.

3.3 Phase 2: subject-dependent discomfort detection

Phase 2 is proposed to incorporate infant-dependent features in addition to Phase 1 subject-independent features mentioned above. In clinical practice, comfort moments of infants occur more often than discomfort, which means it is easier to capture. Thus, we obtain annotated comfort moments of infants, and use the faces as predefined parameterized templates to compare the similarity with the input frame.

All facial landmarks are further translated to achieve that the 28th landmark (the topmost point outlining nose) is the origin of each landmark point set. We calculate the Euclidean distance between the coordinates of eyebrow-, eye-, nose-, mouth-related facial landmarks of every template image and the corresponding coordinates of the input face image.

For each infant, we use 50 templates. The template images are randomly selected from the comfort face images of each infant as their own reference (See Fig. 9 for the schematic of comfort reference set and each input image). The total 51 points of eyebrow-, eye-, nose-, mouth-related facial landmarks are used for similarity calculation. The two-dimensional coordinates of the 51 facial landmarks of the face image are defined as the template. The Euclidean distance between the coordinates of the 51 facial landmarks of the input image and each template of the 50 face images, is computed. We pick the mean, median, maximum, minimum, 30th percentile and 80th percentile of the 50 obtained values of the Euclidean distance as a set of numerical features.

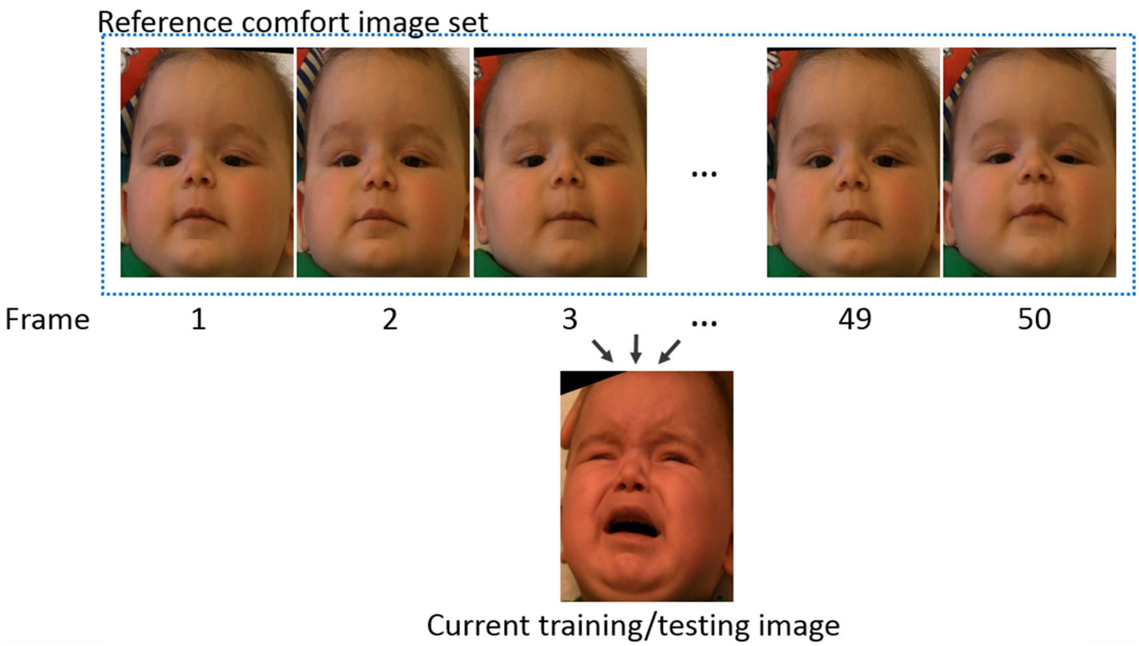

Fig. 9 The top row shows example frames from the 50 templates representing comfort of one infant in Phase 2: subject-dependent discomfort detection. The bottom image is an illustration of the input training/testing image

4 Experimental results

4.1 Materials

The study was conducted with videos recorded at the Maxima Medical Center in Veldhoven, The Netherlands, by a hand-held high-definition camera (Sanyo Xacti VPC-FH1BK). For all infants in the database, written consent was obtained from at least one of the parents. Twenty-two infants were filmed in total. Twenty infant faces were recorded when they were experiencing stressful moments, including treatment moments and special occasions, like: clinical treatment of heel prick, placing an intravenous (IV) line, venipuncture, vaccination or postoperative pain, and discomfort moments of a diaper change, feeling hungry or crying for attention. For 10 out of the 20 infants, the relaxed comfort state of resting or sleeping was also recorded. There were two infants only having their relaxed moments recorded. Thus, the image frames contain one to two emotions per subject. The number of infants regarding the recorded status of comfort/discomfort is summarized in Table 1. The duration of the video segments varies from less than 1 minute to several minutes.

The age of the 22 recorded infants ranged between 2 days and 13 months old. Three of the infants were born premature, and under 37 weeks at the time of recording. Example of video frames for all the infants in the dataset are shown in Fig. 10. The frames of the three premature infants are outlined by yellow dotted rectangles. The resolution of each video frame is 1920×1080 pixels, while the frame rate is 30 fps.

Table 1 Dataset summarization

Infant status	Number of videos
Comfort only	2
Discomfort only	10
Exhibiting both	10

The videos were recorded under uncontrolled, regular hospital lighting conditions. The labels of comfort/discomfort for each frame were annotated according to the consensus of two clinical experts.

We extracted video segments of which infants are in supine position. Finally, a total of 16,378 frames were obtained, on which facial landmarks were detected. From all of the frames, 6075 present comfort, and the rest 10,303 are discomfort frames.

4.2 Results for Phase 1

The subject-independent method (Phase 1) was evaluated by using the 22-infant dataset. Leave-one-out cross-validation was performed. The results of AUCs and labeling error rates for each feature type and for the combination of all features are summarized in Table 2.

When all features are combined, the average of the labeling error rates for all the infants is 0.15. The confusion matrix based on the selected 40 best features is shown in Fig. 11. The accuracy values for the three premature infants are 0.97, 0.61, and 0.79, respectively.

Fig. 10 Examples of frames in the database. Comfort frames are highlighted with a green box, and discomfort frames are indicated in red. Top 4 rows show the 10 infants with both comfort and discomfort moments recorded, where the first and third rows are the discomfort frames, and second and fourth rows are the comfort frames. The fifth and sixth rows are the 10 infants with only discomfort moments recorded. The two pictures at the bottom show the infants with only comfort moments. Three premature infants are outlined by yellow dotted rectangles (color figure online)

Figure 12 shows misclassified faces. Figure 12a shows a false negative case. The image is a discomfort frame. However, our system treats it as comfort because of the opened eyes and mildly opened mouth. Figure 12b portrays a false positive example. In this case, our system classifies this frame as discomfort, since the mouth size is large and the eye size is small. However, in fact, the infant is yawning at this moment and feeling comfortable.

Table 2 Performance measures including classification error rates, and AUCs of different categories of features

Feature category	AUC	Lab. err. rate
Geometric features	0.85	0.22
HOG	0.69	0.29
LBP	0.71	0.28
ALL	0.87	0.15

4.3 Results for Phase 2

The performance of Phase 2 subject-dependent method was evaluated by conducting experiments on the 10-baby dataset, of which both comfort and discomfort moments are present. Infant comfort/discomfort classification through the workflow of Phase 1 was executed and Phase 2 afterward. The

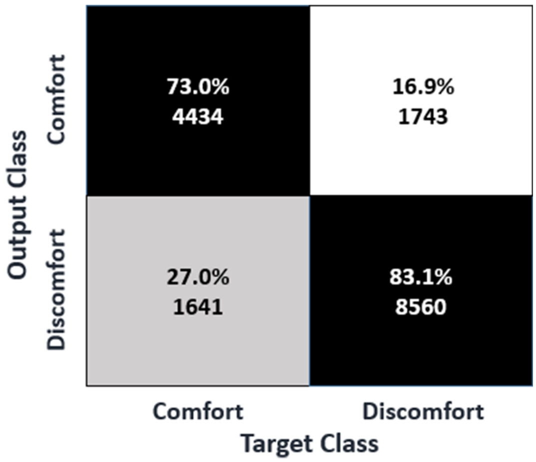

Fig. 11 Confusion matrix when all features are combined. Among all the comfort frames, 73.0% are correctly detected. Among all the discomfort frames, 83.1% are correctly detected

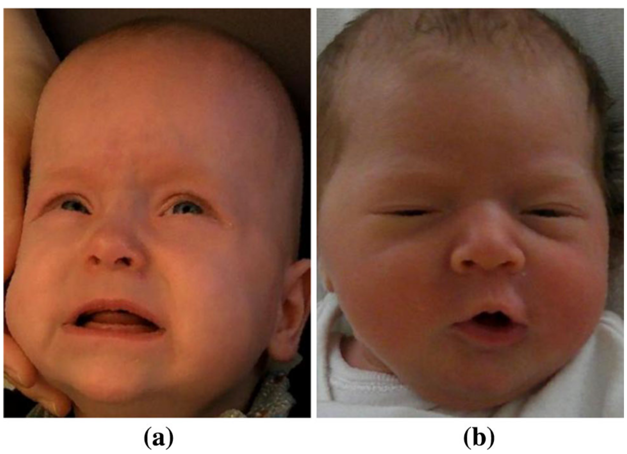

Fig. 12 Examples of misclassified frames. Figure **a** shows a false negative frame, and **b** is a false positive frame

experiment for Phase 1 (without subject-dependent methods) was based on the selected features with the best AUC. We evaluated the efficacy of Phase 2 template-matching features by adding these features to the selected features of Phase 1. Figure 13 shows ROCs for Phase 1 (Without template- matching features) and Phase 2 (With template-matching features), where AUC increases from 0.89 to 0.97. Figure 14 plots the accuracy values per infant when separately applying subject-independent, template matching, and the combined features. The overall accuracy of Phase 1 for the 10-infant dataset is 0.79, and 0.95 for Phase 2, which amounts to a significant accuracy increase of 20%.

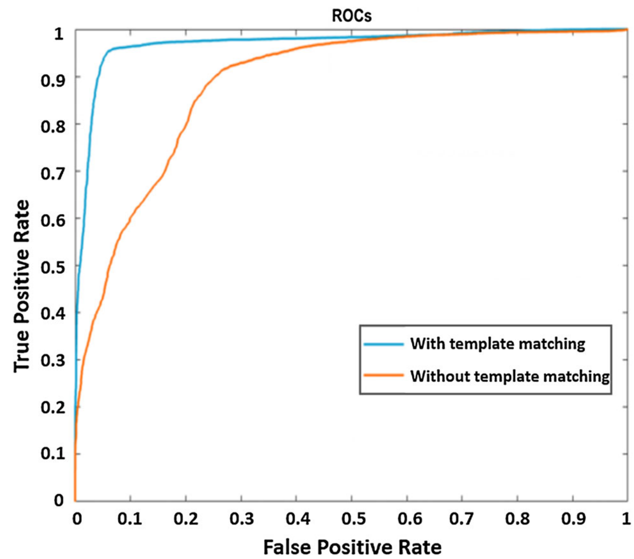

Fig. 13 ROC curves for classification with and without template-matching features, where AUC is 0.97 for the method with template matching, and 0.89 for without template matching

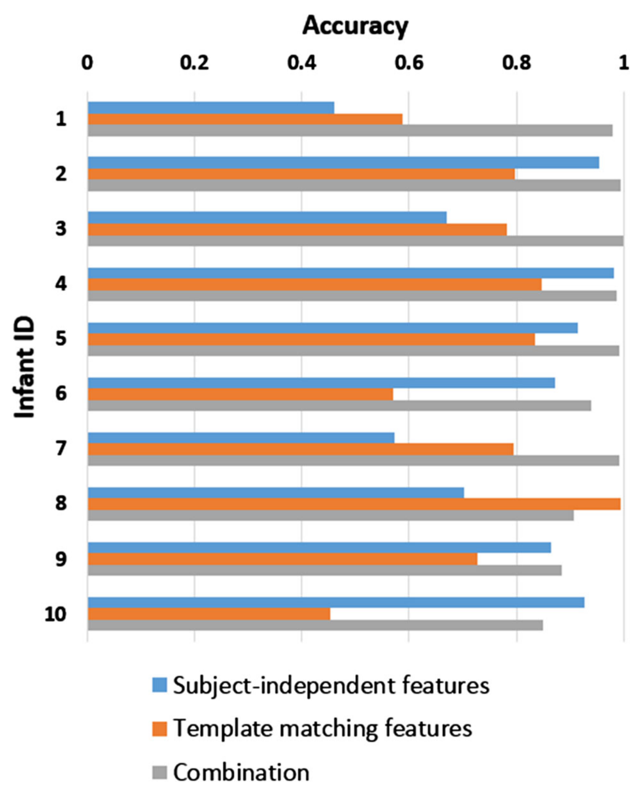

Fig. 14 Accuracies for each infant when applying subject-independent features, template-matching features, and the combination of both. The average accuracy values of all the 10 infants are 0.79 using subject-independent features (Phase 1), 0.74 using template-matching features, and 0.95 when combining all the features together (Phase 2)

5 Discussion

The automated classification on videos of infants helps to detect their discomfort. The system combines 2 phases of subject-independent and subject-dependent methods, respectively. An AUC of 0.97 is achieved, which is promising for clinical practice.

The highest AUC is achieved when combining Phase 2 subject-dependent features with selected subject- independent features of Phase 1, which proves that the features are sufficiently complementary. With our proposed landmark-based template matching, the AUC increased to 0.97, resulting in an overall accuracy of 0.95. The detection accuracies for comfort and discomfort frames were 73.0% and 83.1%, respectively. The average accuracy of the three premature infants in the dataset is 0.79, which shows that our system is also interesting to be considered for premature infant discomfort detection.

The high detection accuracy of our system indicates that the computer system has potential to be used as an alert system to notify doctors or nurses about the status of the babies. The medical staff can make a final decision after a quick inspection of the system result. Our contribution to this field is in three ways. First, this is one of the first automatic systems for facial expression recognition for infants in the hospital care. Second, we propose to normalize the faces for effective feature computation, especially for infants. Third, we incorporate landmark-based template matching involving Euclidean distance to boost our system performance.

To further improve the accuracy of our system, we will explore different settings (parameters) of HOG and LBP features in the future. Regarding geometric features, we could further exploit the rich information from landmarks, such as the distance between the eye and eyebrow, or the distance between upper and lower lips.

For the entire dataset of 22 infants, we extract 7,104 features per frame. After feature selection, the AUC ranges from 0.84 to 0.87. The best feature count is 40, extracted from each frame, which is feasible, particularly for a real-time video-based application.

For the 10-baby dataset with both comfort and discomfort moments, we showed the benefits of subject-dependent features, using landmark-based template matching. Since the reference landmarks are baby-specific and extracted from normalized comfort status, the extracted features are very effective for infant status discrimination. However, the disadvantage of using this type of features is that we need to record comfort status as an additional initialization step for each baby in real practice.

Our system can reach a labeling error rate of only 0.05. In the case that a discomfort expression was misclassified as a comfort expression by the automatic system, there is often no face feature that was linked to the status. For example, the face itself may also not be in a good position. However, it should be noted that the images were annotated within a video of a period and the annotator has the temporal information. When our system by mistake classifies a comfort expression as a discomfort status, it is frequently caused by a baby that is yawning (or aiming to) with opened mouth.

Our automatic system is selective. The area under the ROC curve is very high (0.97). From the ROC curve, we can see that we can keep the sensitivity of detecting discomfort status of our computer system to be almost 100%, while the specificity is 80%. It means that our system can identify about 80% comfort frames without missing any discomfort frames. The fraction of remaining frames is so small that the healthcare professionals can decide on the status.

For the occurrence of face occlusion or head turning, it is not feasible to detect the face and therefore analyze the face expression when using our system. In the future, we would like to add features based on body motion analysis to make our system robust. Finally, we would like to state that the used recordings could be made under nearly ideal conditions since we were allowed to capture the infant faces in many ways. However, in practical clinical conditions, e.g., during treatments, this may not be possible without interrupts, because of the occlusions from the nurses and doctors. This would lead to drops in the detection.

6 Conclusions

We have proposed a video-based automated system that can differentiate discomfort of infants from comfort status. The system aims at alerting pain and discomfort, with the aim to improve developmental outcomes of infants on a long term. The experimental results are promising, so that it has the potential for clinical usage.

One limitation of our study is the small size of the dataset. Collecting videos from infants at the pediatrics department in the hospital is very restrictive because of the ethical and legal issues. In the meantime, not many infants within our target group are available for recording. Moreover, some of the captured recordings are not usable due to interruptions from nurses or parents, face occlusion, etc. Currently, we are in the process of acquiring more data and include these data in our upcoming work. Another limitation of this work is that features used are computed from a static frame. The temporal dynamics of facial expressions are not taken into account. In the future work, we will analyze facial expression changes over time and also consider dynamic features based on body motion analysis. Furthermore, we may add features extracted from vital signs, such as heart rate and respiration rate, and evaluate those in combination with our visual features for improved robustness and reliability. Last, but not least, deep learning methods have shown superior performance recently in many computer vision applications. In our upcoming work, we will also investigate deep learning-based methods for our problem.

References

1. Acharya, R., Kumar, A., Bhat, P., Lim, C., Kannathal, N., Krishnan, S., et al.: Classification of cardiac abnormalities using heart rate signals. Med. Biol. Eng. Comput. **42**(3), 288–293 (2004)
2. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2037–2041 (2006)
3. Ambuel, B., Hamlett, K.W., Marx, C.M., Blumer, J.L.: Assessing distress in pediatric intensive care environments: the comfort scale. J. Pediatr. Psychol. **17**(1), 95–109 (1992)
4. American Academy of Pediatrics, and Fetus and Newborn Committee: Prevention and management of pain in the neonate: an update. Pediatrics **118**(5), 2231–2241 (2006)
5. Ashraf, A.B., Lucey, S., Cohn, J.F., Chen, T., Ambadar, Z., Prkachin, K.M., Solomon, P.E.: The painful face-pain expression recognition using active appearance models. Image Vis. Comput. **27**(12), 1788–1796 (2009)
6. Behrman, R., Butler, A.S.: Institute of Medicine Committee on Understanding Premature Birth and Assuring Healthy Outcomes Board on Health Sciences Outcomes: Preterm Birth: Causes, Consequences, and Prevention. Preterm Birth: Causes, Consequences, and Prevention. National Academies Press, Washington (2007)
7. Blum, A.L., Langley, P.: Selection of relevant features and examples in machine learning. Artif. Intell. **97**(1–2), 245–271 (1997)
8. 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)
9. Brown, S., Timmins, F.: An exploration of nurses' knowledge of, and attitudes towards, pain recognition and management in neonates. J. Neonatal Nurs. **11**(2), 65–71 (2005)
10. Chang, F.J., Tran, A.T., Hassner, T., Masi, I., Nevatia, R., Medioni, G.: Expnet: landmark-free, deep, 3D facial expressions. arXiv preprint arXiv:1802.00542 (2018)
11. Déniz, O., Bueno, G., Salido, J., De la Torre, F.: Face recognition using histograms of oriented gradients. Pattern Recogn. Lett. **32**(12), 1598–1603 (2011)
12. Fitzgerald, M.: The development of nociceptive circuits. Nat. Rev. Neurosci. **6**(7), 507–520 (2005)
13. Fitzgerald, M., Millard, C., McIntosh, N.: Cutaneous hypersensitivity following peripheral tissue damage in newborn infants and its reversal with topical anaesthesia. Pain **39**(1), 31–36 (1989)
14. Fotiadou, E., Zinger, S., Tjon A Ten, W., Bambang Oetomo, S., et al.: Video-based facial discomfort analysis for infants. In: IS&T/SPIE Electronic Imaging, pp. 90290F–90290F. International Society for Optics and Photonics (2014)
15. Gholami, B., Haddad, W.M., Tannenbaum, A.R.: Relevance vector machine learning for neonate pain intensity assessment using digital imaging. IEEE Trans. Biomed. Eng. **57**(6), 1457–1466 (2010)
16. Grunau, R., Craig, K.: Pain expression in neonates: facial action and cry. Pain **28**(3), 395–410 (1987)
17. Grunau, R.E., Whitfield, M.F., Petrie-Thomas, J., Synnes, A.R., Cepeda, I.L., Keidar, A., Rogers, M., MacKay, M., Hubber-Richard, P., Johannesen, D.: Neonatal pain, parenting stress and interaction, in relation to cognitive and motor development at 8 and 18 months in preterm infants. Pain **143**(1), 138–146 (2009)
18. Hammal, Z., Cohn, J.F.: Towards multimodal pain assessment for research and clinical use. In: Workshop on Roadmapping the Future of Multimodal Interaction Research Including Business Opportunities and Challenges (2014)
19. Hsieh, R., Mochizuki, Y., Asano, T., Higashida, M., Shirai, A.: Real baby-real family: Vr entertainment baby interaction system. In: ACM SIGGRAPH 2017 Emerging Technologies, p. 20. ACM (2017)
20. Johnston, C.C., Stevens, B.J., Yang, F., Horton, L.: Differential response to pain by very premature neonates. Pain **61**(3), 471–479 (1995)
21. Kazemi, V., Josephine, S.: One millisecond face alignment with an ensemble of regression trees. In: 27th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, United States, 23 June 2014 through 28 June 2014, pp. 1867–1874. IEEE Computer Society (2014)
22. Kharghanian, R., Peiravi, A., Moradi, F.: Pain detection from facial images using unsupervised feature learning approach. In: 2016 IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC), pp. 419–422. IEEE (2016)
23. King, D.E.: Dlib-ml: a machine learning toolkit. J. Mach. Learn. Res. **10**((Jul)), 1755–1758 (2009)
24. Kotsia, I., Pitas, I.: Facial expression recognition in image sequences using geometric deformation features and support vector machines. IEEE Trans. Image Process. **16**(1), 172–187 (2007)
25. Lindh, V., Wiklund, U., Håkansson, S.: Heel lancing in term newborn infants: an evaluation of pain by frequency domain analysis of heart rate variability. Pain **80**(1–2), 143–148 (1999)
26. Littlewort, G.C., Bartlett, M.S., Lee, K.: Automatic coding of facial expressions displayed during posed and genuine pain. Image Vis. Comput. **27**(12), 1797–1803 (2009)
27. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
28. Lucey, P., Cohn, J.F., Matthews, I., Lucey, S., Sridharan, S., Howlett, J., Prkachin, K.M.: Automatically detecting pain in video through facial action units. IEEE Trans. Syst. Man Cybern. Part B Cybern. **41**(3), 664–674 (2011)
29. Mima, Y., Arakawa, K.: Cause estimation of younger babies' cries from the frequency analyses of the voice-classification of hunger, sleepiness, and discomfort. In: International Symposium on Intelligent Signal Processing and Communications, 2006. ISPACS'06, pp. 29–32. IEEE (2006)
30. Norden, J., Hannallah, R., Getson, P., O'Donnell, R., Kelliher, G., Walker, N.: Reliability of an objective pain scale in children. J. Pain Symptom Manag. **6**(3), 196 (1991)
31. 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)
32. Petrini, J.R., Dias, T., McCormick, M.C., Massolo, M.L., Green, N.S., Escobar, G.J.: Increased risk of adverse neurological development for late preterm infants. J. Pediatr. **154**(2), 169–176 (2009)
33. Raju, T.N., Higgins, R.D., Stark, A.R., Leveno, K.J.: Optimizing care and outcome for late-preterm (near-term) infants: a summary of the workshop sponsored by the national institute of child health and human development. Pediatrics **118**(3), 1207–1214 (2006)
34. Riddell, R.P., Racine, N.: Assessing pain in infancy: the caregiver context. Pain Res. Manag. **14**(1), 27–32 (2009)
35. Schiavenato, M., Byers, J.F., Scovanner, P., McMahon, J.M., Xia, Y., Lu, N., He, H.: Neonatal pain facial expression: evaluating the primal face of pain. Pain **138**(2), 460–471 (2008)
36. Shan, C., Gong, S., McOwan, P.W.: Robust facial expression recognition using local binary patterns. In: 2005 IEEE International Conference on Image Processing, ICIP 2005, vol. 2, pp. II–370. IEEE (2005)

37. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis. Comput. **27**(6), 803–816 (2009)
38. Sikka, K., Ahmed, A.A., Diaz, D., Goodwin, M.S., Craig, K.D., Bartlett, M.S., Huang, J.S.: Automated assessment of children's postoperative pain using computer vision. Pediatrics **136**(1), 124–131 (2015)
39. Stevens, B., Johnston, C., Petryshen, P., Taddio, A.: Premature infant pain profile: development and initial validation. Clin. J. Pain **12**(1), 13–22 (1996)
40. US Department of Health and Human Services: Acute Pain Management in Infants, Children, and Adolescents: Operative and Medical Procedures. Agency for Health Care Policy and Research, Rockville (1992)
41. Vu, N.S., Caplier, A.: Face recognition with patterns of oriented edge magnitudes. In: European conference on computer vision, pp. 313–326. Springer, Berlin (2010)
42. Whit Hall, R., Anand, K.: Short-and long-term impact of neonatal pain and stress. NeoReviews **6**, 69–75 (2005)
43. Zamzmi, G., Pai, C.Y., Goldgof, D., Kasturi, R., Ashmeade, T., Sun, Y.: An approach for automated multimodal analysis of infants' pain. In: 23rd International Conference on Pattern Recognition (ICPR 2016)
44. Zamzmi, G., Pai, C.Y., Goldgof, D., Kasturi, R., Sun, Y., Ashmeade, T.: Automated pain assessment in neonates. In: 20th Scandinavian Conference on Image Analysis (SCIA 2017)

Yue Sun received her BSc degree in Biomedical Engineering from Huazhong University of Science and Technology, Wuhan, China in 2010. In 2014, she completed her MSc in Biomedical Engineering at the University of Western Ontario, London, Canada. She is currently working as a Ph.D. researcher in the Signal Processing Systems, Video Coding and Architectures research group at Eindhoven University of Technology, Eindhoven, The Netherlands. Her research interests include video/image processing, machine learning, facial expression/human behavior analysis.

Caifeng Shan is currently a Senior Scientist and Project Leader with Philips Research, Eindhoven, The Netherlands. He received the PhD degree in computer vision from Queen Mary, University of London. He is the recipient of the 2007 Chinese Government Award for Outstanding Students Abroad. His research interests include computer vision, pattern recognition, image and video analysis, machine learning, bio-medical imaging, and related applications. He has authored about 80 scientific publications (Google Scholar Citations: 4700+) and about 50 patent applications. He was Associate Editor of IEEE Transactions on Circuits and Systems for Video Technology (2011-2016), and has been Editorial Board Member of Journal of Visual Communication and Image Representation and Journal on Ambient Intelligence and Smart Environments. He has edited three books and has been the Guest Editor of IEEE Transactions on Multimedia, IEEE Transactions on Circuits and Systems for Video Technology, Signal Processing (Elsevier), and Machine Vision and Applications. He organized several international workshops at flagship conferences such as IEEE ICCV and ACM Multimedia. He has served as a Program Committee Member and Reviewer for numerous international conferences and journals. He is Senior Member of IEEE.

Tao Tan graduated from Zhejiang University (China) for his bachelor study with honor and obtained his master degree from Eindhoven University of Technology (TU/e) in 2009. He worked at Philips as a researcher from 2008 to 2009. He obtained his PhD degree from the world-renown computer-aided diagnosis group (DIAG), Radboud University Medical Center. He is currently a senior scientist in ScreenPoint Medical and an assistant professor at TU/e. He is also actively involved in medical imaging challenge organizations.

Xi Long holds a BSc with honor in Electronic Information Engineering from Zhejiang University (China) and obtained his MSc in Electrical Engineering from Eindhoven University of Technology (TU/e) in 2009. He worked at Philips as a researcher from 2008 to 2009 and at Tencent (China) as project manager and UX engineer from 2009 to 2011. He then went on to do his PhD at TU/e and Philips Research, on signal processing and machine learning in unobtrusive sleep monitoring, which he completed cum laude in 2015. Long currently wears several hats: he is a Scientist and Lead FPP Data Analytics Cluster at Philips Research, a Project Coordinator at Maxima Medical Center and UMC Utrecht, and an Assistant Professor at TU/e.

Arash Pourtaherian received his BSc. degree in Electrical Engineering jointly from Indiana University-Purdue University Indianapolis, IN, USA and University of Tehran, Iran in 2010. He then moved to the Netherlands, where he studied MSc. Electrical Engineering at the Eindhoven University of Technology (TU/e). There, he continued his research as a PhD candidate in the Video Coding and Architectures–Signal Processing Systems group at the Electrical Engineering faculty. As of 2018, he is a post-doc at TU/e, working on ultrasound imaging and analysis, 3D sensing and reconstruction, and medical data processing.

Svitlana Zinger is currently an assistant professor at Eindhoven University of Technology, The Netherlands. She received her MSc in computer science in 2000 from the Radiophysics faculty of the Dnepropetrovsk State University, Ukraine. In 2004, she received a PhD from the Ecole Nationale Superieure des Telecommunications, France, for her thesis on interpolation and resampling of 3D data. In 2005 she was a postdoctoral fellow in the Multimedia and Multilingual Knowledge Engineering Laboratory of the French Atomic Agency, France, where she worked on creation of a large-scale image ontology for content based image retrieval. In 2006-2008 Sveta was a postdoctoral researcher at the Center for Language and Cognition Groningen and an associated researcher at the Artificial Intelligence department in the University of Groningen, the Netherlands, working on information retrieval from handwritten documents.

Peter H.N. de With studied Electrical Engineering at Eindhoven University of Technology (TU/e) and obtained his PhD degree from Delft University of Technology in 1992. From 1984 to 1997, he researched video compression and was senior TV Systems Architect at Philips Research. From 1997 to 2000, De With was Full Professor at the University of Mannheim (Germany). He joined LogicaCMG in Eindhoven as principal consultant in 2000, and was then appointed part-time professor of Video Coding and Architectures at TU/e. De With was Vice President of VP Video (Analysis) Technology at Cyclomedia Technology from 2008 to 2010. In 2011, he was appointed Full Professor at TU/e, scientific director of Care and Cure Technology (C3Te) and theme leader of Smart Diagnosis at SA Health.