IEICE TRANS. INF. & SYST., VOL.E101–D, NO.7 JULY 2018

PAPER

Infants' Pain Recognition Based on Facial Expression: Dynamic Hybrid Descriptions

Ruicong ZHI[†,††a)], **Ghada ZAMZMI**[†††], **Dmitry GOLDGOF**[†††], **Terri ASHMEADE**[††††], ***Nonmembers***, **Tingting LI**[†,††], ***Member***, ***and*** **Yu SUN**[†††], ***Nonmember***

SUMMARY The accurate assessment of infants' pain is important for understanding their medical conditions and developing suitable treatment. Pediatric studies reported that the inadequate treatment of infants' pain might cause various neuroanatomical and psychological problems. The fact that infants can not communicate verbally motivates increasing interests to develop automatic pain assessment system that provides continuous and accurate pain assessment. In this paper, we propose a new set of pain facial activity features to describe the infants' facial expression of pain. Both dynamic facial texture feature and dynamic geometric feature are extracted from video sequences and utilized to classify facial expression of infants as pain or no pain. For the dynamic analysis of facial expression, we construct spatiotemporal domain representation for texture features and time series representation (i.e. time series of frame-level features) for geometric features. Multiple facial features are combined through both feature fusion and decision fusion schemes to evaluate their effectiveness in infants' pain assessment. Experiments are conducted on the video acquired from NICU infants, and the best accuracy of the proposed pain assessment approaches is 95.6%. Moreover, we find that although decision fusion does not perform better than that of feature fusion, the False Negative Rate of decision fusion (6.2%) is much lower than that of feature fusion (25%).

key words: *infants pain assessment, temporal geometric descriptor, LBP-TOP, decision fusion, video analysis*

1. Introduction

Pain can be defined as a protective mechanism that alerts about damage or injury that is occurring or potentially occurring [1]. Accurate pain assessment is important for understanding patients' medical conditions and developing suitable treatments. Studies have found that poor treatment of infants' pain might cause permanent neuroanatomical changes, developmental, and learning disabilities [2]–[4]. A significant proportion of sick infants receive poor pain assessment and management when they receive medical procedures and treatment [5].

Pain is a subjective experience, and traditionally self-report is considered to be the best measurement for pain. The most common self-report pain scales are verbal numeric scale [6], verbal descriptive scale, and visual analogue scale [7]. The main limitation of these scales is that they are not reliable, and easily affected by physical, cognitive, emotional state, and the pain history or pain expectation [8], [9]. In addition, they are only useful for patients who are able to communicate verbally.

For infants or individuals with communicative impairment, pain assessment is measured by physiological and/or behavioral indicators such as facial movements and body postures. NIPS (Neonatal Infant Pain Scale) and NPASS (Neonatal Pain, Agitation and Sedation Scale) [10] are two popular indicator-based scales for infants' pain assessment for acute pain and chronic pain respectively. The indicator-based scales are utilized by trained nurses at different time intervals for adequate pain monitoring. It is laboring for long-term pain assessment, and the nurses' judgement is highly biased due to the observer's experience, culture, and state [11]. Therefore, an automatic pain assessment system that can analyze infants' pain behaviors intelligently has recently attracted increasing interest, to provide continuous and accurate pain assessment.

1.1 Related Work

Facial expression is the most specific pain indicator, which is more sensitive to noxious procedures than cry, body movements, and heart rate [12], [13]. Moreover, the caregivers could judge the pain facial expression more salient and consistent than cry [15]–[17]. The facial expression of pain is unique and different from the six basic emotions [14] which are widely accepted by psychologists. The importance of face has been acknowledged in all multidimensional pain instruments [18].

Facial expression analysis of pain has attracted increasing attention in the last decades. However, related research on infants' pain expression analysis is limited. Only a few facial representations have been applied in automatic infants' pain expression recognition, such as appearance-based features acquired by Discrete Cosine Transform [19], Elongated Local Ternary Pattern and Elongated Local Binary Pattern [20], and Principle Component Analysis and Linear Discriminant Analysis [21]. The main limitation of these works is they deal with static images which ignore the dynamic information. Only a few studies reported the tem-

Manuscript received August 25, 2017.
Manuscript revised February 3, 2018.
Manuscript publicized April 20, 2018.

[†]The authors are with the School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, P.R. China.
[††]The authors are with Beijing Key Laboratory of Knowledge Engineering for Materials Science, Beijing 100083, P.R. China.
[†††]The authors are with Department of Computer Science and Engineering, University of South Florida, Tampa, Florida 33620, USA.
[††††]The author is with the College of Medicine Pediatrics, University of South Florida, Tampa, Florida 33620, USA.
a) E-mail: zhirc_research@126.com
DOI: 10.1587/transinf.2017EDP7272

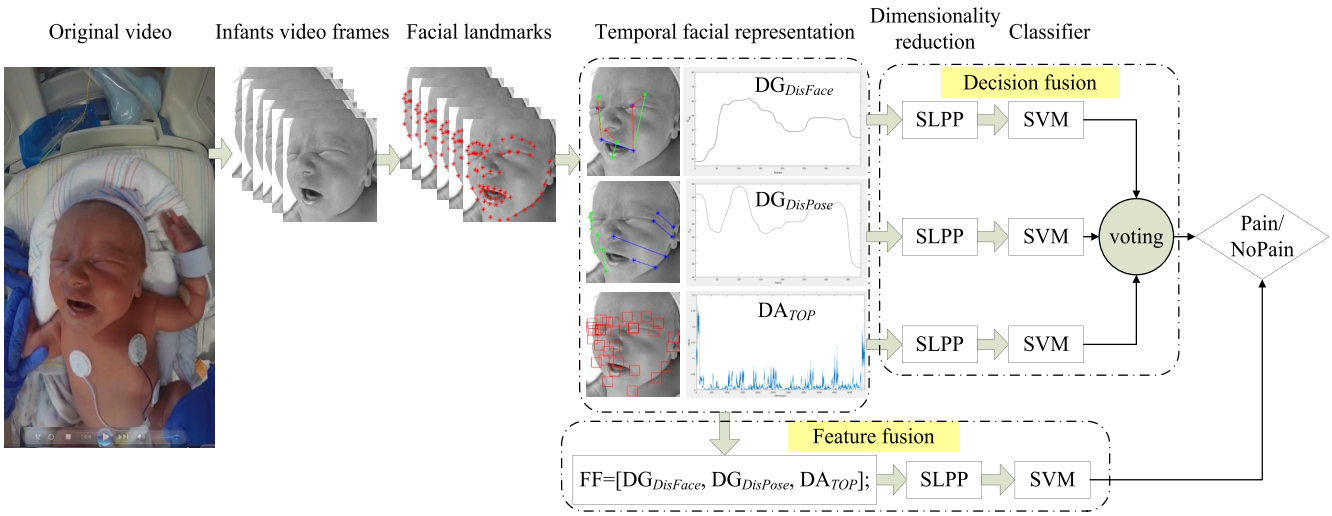


Fig. 1 Illustration of the automatic dynamic pain facial expression recognition system.

poral facial representation for infants' video during pain experience. Zamzmi et al. [22]–[24] utilized an optical flow method to spot the pain expression in the video and described the facial strain changes over time to classify pain facial expressions of infants undergoing acute painful procedures. Fotiadou et al. [25] presented Active Appearance Model to extract facial features of each video frame and the global motion information, and the evaluation was conducted for eight infants using SVM classifier. However, it is frame-level based method which treats the video frames as static images.

1.2 Overview of Contributions

In this paper, we propose a new set of pain facial activity features to describe the infants facial expressions of pain, and it is successfully applied to infants' pain assessment. The main contributions of our work include the following three aspects:

- Both facial texture feature and geometric feature are extracted from video sequences. Geometric representation is calculated simply with low dimensionality and could depict the facial configuration intuitionally. To our best knowledge, this work is the first to perform geometric-based feature for infants' pain facial expression analysis.
- Dynamic analysis is applied to pain facial expression recognition. We perform spatiotemporal domain representation for texture features and time series representation (i.e. time series of frame-level features) for geometric feature. Dynamic characteristics are important to capture how the expression evolves over time, e.g. eye squeeze during pain is hard to be distinguished in a single static image. On the other hand, it is time-consuming for pain labeling for every frame in the video, and there is no need to do so since usually pain lasts for a specific time-interval, it is better to segment and annotate video with proper labels.
- Feature fusion and decision fusion schemes are conducted to integrate multiple facial representations and pain class labels, to fully exploit the advantages of multiple facial features. Both fusion schemes are utilized to classify facial expressions of NICU infants as pain or no pain. The experimental comparison shows that fusion strategy could noticeably enhance the pain assessment accuracy. An overview of the system is shown in Fig. 1.

2. Dynamic Pain Facial Expression of Infants

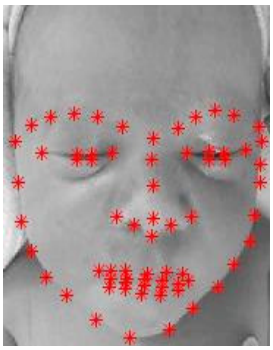
First, the well-known Active Appearance Model [26] was employed to locate the infant's face and track 68 facial landmarks (see Fig. 2 (a)) in the image sequences, including landmarks of facial organs and boundary of the face region. These landmark points were used to generate temporal geometric feature and temporal appearance feature related to pain facial activities. The approach was described as follows.

2.1 Temporal Geometric Feature Representation

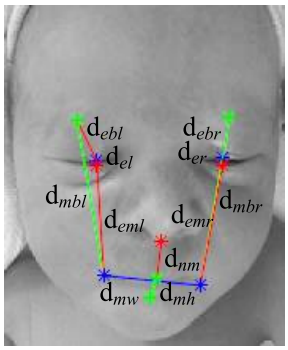
The facial geometric information could describe the changes of facial organs, and the configuration parameters were derived from a set of fiducial points. First, several distance parameters were extracted from each video frame to capture pain related facial changes. Then the static parameters comprised to a series of temporal signals, from which several typical signal characterization descriptors were extracted to obtain the temporal features of facial geometric representation.

2.1.1 Frame-Level Facial Configuration Measurement

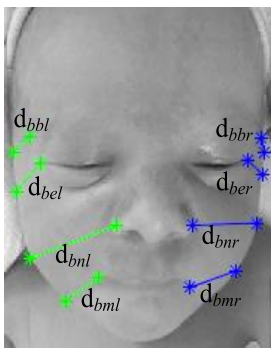
The facial configuration elicited by pain could be measured



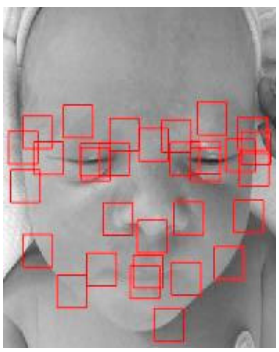
(a) facial landmarks



(b) facial configuration parameters



(c) head pose parameters



(d) landmark patches

Fig. 2 Frame-level parameters for infants pain facial expression.

by sign judgment method, such as Neonatal Facial Coding System (NFCS) which is designed for newborns to 2 months of age. There are ten discrete facial actions, including brow bulge, nasolabial furrow, eye squeeze, chin quiver, open lips, lip purse, horizontal mouth stretch, vertical mouth stretch, taut tongue, and tongue protrusion [27].

Based on the facial actions defined in NFCS, several frame-level geometric distance parameters were designed to capture the facial activities of pain. The distance parameters include: distance between eyebrows and eyes (d_{ebl} and d_{ebr}), distance between upper eyelid and lower eyelid (d_{el} and d_{er}), distance between eyebrow and mouth (d_{mbl} and d_{mbr}), distance between eye and mouth (d_{eml} and d_{emr}), distance between nose and mouth (d_{nm}), and the width (d_{mw}) and height (d_{mh}) of the mouth. Therefore, there are 11 distance parameters for geometric description of pain facial configuration.

Head movement is one of the major sign judgements for infants' pain assessment. According to the observation, head movement happened very commonly during pain experience. Therefore, a simple method was employed for head pose description by exploiting pose parameters from distances between key facial landmarks to face boundary landmarks (d_{bbl}, d_{bel}, d_{bnl}, d_{bml}, d_{bbr}, d_{ber}, d_{bnr}, d_{bmr}). The distance parameters were calculated for left side of face as well as right side of face, if the head shakes, the distances of two face sides would change related inversely. The facial activity parameters and pose motion parameters are illustrated in Fig. 2 (b) and Fig. 2 (c), respectively.

2.1.2 Temporal Descriptors for Video Sequence

The facial activity descriptors, consisting of facial configuration parameters and head pose parameters, were extracted from each image frames. Next, the corresponding temporal series were employed for each facial descriptor obtained from all video frames, which is inspired by [28].

Let $x(\bullet)$ denote the descriptor signal, and the signal is firstly smoothed by a Butterworth filter (first order with cut-off 1 Hz for temporal signals) temporally. Then, the first and second temporal derivations of the smoothed descriptor signal were estimated for feature description. These signals could be treated as the state signal, speed signal and acceleration signal of the descriptor signal, respectively. Subsequently, several parameters could be extracted from the temporal signals to better depict the characteristics of the signal variance over time [28]. A total of 16 temporal geometric facial features were extracted for each video sequence, and we categorized these parameters into 6 groups as follows:

— State parameters: maximum value (MAX), minimum value (MIN), mean value (MEAN), and median value (MEDIAN);

— Variability parameters: range (RANGE), standard deviation (SD), inter-quartile range (IQR), inter-decile range (IDR), and median absolute deviation (MAD);

— Peak parameters: instant of time when the amplitude is at its maximum (TMAX);

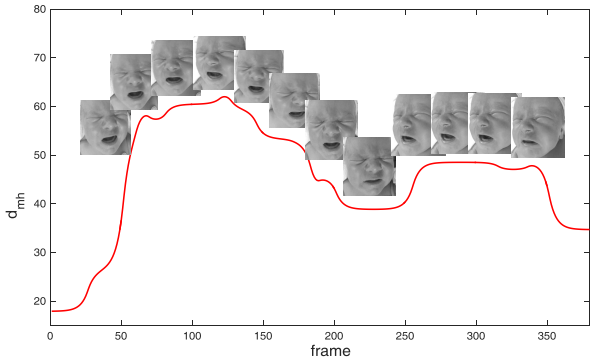
— Duration parameters: duration of when the amplitude is greater than mean (DGM), and duration of when the amplitude is greater than the average of mean and min (DGA);

— Segment parameters: number of segments where the amplitude is greater than mean (SGM), and number of segments where the amplitude is greater than the average of mean and min (SGA);

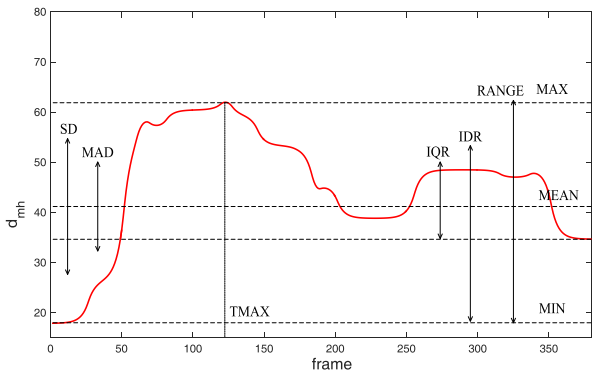
— Area parameters: area between signal and its minimum (AREA), and quotient of AREA and (the difference between MAX and MIN) (AREAR).

The temporal features extracted from the smoothed temporal signals and its derivations could reflect the facial actions related to pain, for example, the state parameters are related to the facial activity intensity; the variability parameters are measure of statistical dispersion of the signal, e.g. the inter-decile range is the difference between the first and the ninth deciles; the peak parameter depicts the time of the apex and the duration parameters are related to the time interval of the facial motion lasts; other parameters capture additional information, e.g. MAX and TMAX of the speed signal mean the speed and its timing. The temporal features provide facial motion information related to amplitude, speed, duration, and variability, which are helpful to depict the geometric variance in the video. Figure 3 demonstrates some of the parameters extracted from the temporal signal.

Each facial activity descriptor (11 facial configuration



(a) State signal (d_{mh}) with video frames



(b) Some temporal signal parameters

Fig. 3 An example of the temporal signal feature.

descriptors, and 8 head pose descriptors) could compose a temporal signal and the previously described discrete parameters were extracted from the smoothed signal and its derivation signals. Therefore, the dimensionality of geometric feature would be $n \times (3 \times 16)$, where n is the number of frame-level facial descriptor ($n = 19$ in this study).

2.2 Temporal Appearance Feature Representation

The histogram based static appearance descriptors could be extended to three dimensions for dynamic appearance information encoding. The resulting spatiotemporal representation could boost the accuracy compared to their static counterparts. The most well-applied histogram representation is Local Binary Patterns (LBP) and Local Phase Quantisation (LPQ), and their dynamic variances are LBP-TOP [29] and LPQ-TOP [30], which compute features from three orthogonal planes of X-Y, X-T, and Y-T, individually. The X-plane and Y-plane denote the spatial dimensions and T-plane denotes the time dimension.

It has been reported that spatiotemporal facial appearance descriptors perform better in small patches than holistically in the entire face [31] since it could better capture the skin texture at different facial regions, such as eyebrow corner, nasolabial furrow and mouth corner. In addition, local patch derived features could significantly reduce the computational complexity comparing to that extracted from the entire face (usually the whole face is divided into small blocks, and the descriptors are extracted from all the blocks). Therefore, we extracted the dynamic facial appearance features, namely LBP-TOP and LPQ-TOP, from 32×32 landmark patches located around the 31 facial landmarks demonstrated in Fig. 2 (d). For each landmark patch in a pain video, we obtained a feature vector with 177 dimensions for LBP-TOP, and 768 dimensions for LPQ-TOP. The dynamic appearance feature for each pain video is 5487 (177×31) dimensions (for LBP-TOP)/ 23808 (768×31) dimensions (for LPQ-TOP) by concatenating the local appearance descriptors over 31 facial landmark patches to represent the whole face.

A second scheme for temporal appearance feature extraction was also employed due to the simplicity and low dimensionality, i.e. calculating the frame-level mean gradient magnitude for each landmark patch firstly. The static parameters comprised sequence signals, and then the signal characterization features were extracted using the similar procedure described in Sect. 2.1. Therefore, the feature dimension of dynamic gradient features is 1488 (48×31).

2.3 Dimensionality Reduction Using SLPP

The original dynamic geometric features and appearance features yield a high-dimensional feature space, and the performance of entering the original feature matrix to classifier directly is not ideal due to the redundancy of features. The intrinsic features usually lie in a lower dimensional subspace which could represent the useful information from raw feature matrix. There are a number of dimensionality reduction methods, and manifold learning is one of nonlinear dimensionality reduction algorithms which have been widely applied in pattern recognition tasks. However, there is no explicit mapping expression in traditional manifold learning method, and it could not deal with new test samples for classification. Therefore, we utilized the Supervised Locality Preserving Projections (SLPP) [32], which is a linear approximation of Laplacian Eigenmaps (LE), to obtain a more compact feature subspace with much fewer parameters. The SLPP inherits the advantages of nonlinear manifold learning and also provides transformation function explicitly. The principle of the SLPP is illustrated as follows.

Let $X = [x_1, x_2, \ldots, x_N]$ denote the original feature matrix in $\mathbf{R}^D$. The SLPP aims to seek a transformation $\mathbf{A}$ to map the high-dimensional input data into a low-dimensional subspace $Y = [y_1, y_2, \ldots, y_N]$ in $\mathbf{R}^L$ through $y_i = \mathbf{A}^T x_i$, such that the local structure is preserved.

The local relationship between data points is described by weight matrix $\mathbf{W}$, which is determined through the k-nearest-neighbors of each point. In this study, the heat-kernel was utilized to construct the weight matrix, i.e. if x_j is the neighbor of x_i or x_i is the neighbor of x_j, the weight coefficient is set to $W_{ij} = exp(-\|x_i - x_j\|^2/t)$, where t is a constant parameter; otherwise, $W_{ij} = 0$.

The key point is to find the optimal mapping to remain that the neighbors in original data space are also close in

the projected space, by minimizing the objective function: $\min_A \sum_{i,j}^{N} \|y_i - y_j\|^2 W_{ij}$. According to the linear transformation defined as $y_i = \mathbf{A}^T x_i$, the minimization problem could be converted to

$$\arg\min_{\mathbf{A}^T\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{A}=1} \mathbf{A}^T\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{A} \quad (1)$$

where $\mathbf{D}$ is a diagonal matrix with $D_{ii} = \sum_j W_{ij}$, and $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the Laplacian matrix. The optimization problem of Eq. (1) could be solved by a generalized eigenvalue problem. Thus, the projection is implemented according to the eigenvectors corresponding to the eigenvalues obtained from Eq. (1).

The compact and effective features extracted by SLPP were fed to classifier to enhance the classification performances. Moreover, feature matrix with low dimensionality could reduce the computational complexity significantly, and less storage resources are needed.

2.4 Fusion and Classification

To best exploit the superiority of facial representations, we applied two fusion schemes which are feature fusion and decision fusion, and compared the pain assessment performances. **Feature fusion** method combines multiple temporal geometric representations and temporal appearance representations to form a single feature vector. **Decision fusion** method made the final decision according to the output of single classifier via majority voting, as a specific classifier is learned for a type of feature.

The dimensionality-reduced facial features (single type of feature or joint feature) were fed into Support Vector Machine (SVM) classifier for pain recognition. SVM is a powerful statistical classifier for binary classification. In feature level fusion, multiple facial features are integrated into one feature vector, and one SVM classifier is trained by the reduced features utilizing SLPP, and the classifier output the pain label of the system. In decision level fusion, each type of reduced facial feature trained a SVM classifier individually, and it contributes one vote (i.e., class label) to the final classification, and the major class in the combination is the final label of pain assessment. If there is a tie for different indicators, the class with highest confidence score is chosen as the final decision of pain assessment.

3. Experimental Results and Discussion

In this section, we evaluated the proposed scheme in **IPAD** (Infants Pain Assessment Database), using both dynamic geometric features and dynamic appearance features described above. Multiple feature integration methods and fusion methods (feature fusion and decision fusion) were evaluated. Moreover, the importance of descriptors was also discussed.

3.1 Database

The experiments were conducted on our IPAD database which contains videos acquired for NICU infants at Tampa General Hospital. The collected videos have facial expression, body movement, and sound for 12 infants. The ratio between male and female is 1:1. The age of the infants ranged from 32 to 40 gestational weeks, with a mean age of 35.9 (±2.8). Infants are also racially diverse with White (10) and Black (2). The infants were recorded during routine painful procedure (e.g. heel lancing). Each infant has one or two recordings for the pain procedure.

The ground truth of the pain assessment was obtained by trained nurse using NIPS. Two experienced nurses (worked in the NICU) of Tampa General Hospital are in charge of the pain scoring by NIPS, they score the pain indicators individually. The kappa coefficient of their scores is 0.86 which indicates very good agreement. The results that they agree on are utilized as ground truth. The pain scale is composed of six indicators including facial expression, cry, breathing patterns, arms, legs, and state of arousal. The total pain score is obtained by summing up all the scores of pain indicators. The label of "pain" or "no pain" was assigned to samples for training and classifying, which was also called as "gold standard" in clinic.

Each of the procedure video was segmented into seven time periods (T0 ~ T6) for subsequent analysis, and each of the segmentation is labeled with pain or no pain according to the NIPS score. There are 81 instances in total in the database. These seven epochs are:

T0: 5 minutes pre-procedure to provide the baseline state.

T1: actual pain procedure.

T2: 1 minute after the completion of the painful procedure.

T3: 2 minute after the completion of the painful procedure.

T4: 3 minute after the completion of the painful procedure.

T5: 4 minute after the completion of the painful procedure.

T6: 5 minute after the completion of the painful procedure.

3.2 Fusion Scheme for Multiple Features

In this section, the facial representation descriptors were divided into three categories, i.e. dynamic geometric features for facial configuration representation ($DG_{DisFace}$), dynamic geometric features for head pose representation ($DG_{DisPose}$), and dynamic appearance features including LBP-TOP (DA_{LBPTOP}), LPQ-TOP (DA_{LPQTOP}), and temporal gradient features ($DA_{gradient}$). The symbols denoted the low-dimensional features that were fed to classifiers. The output of classifier was one of two classes, i.e. pain and no pain. Cross-validation is widely adopted strategy which harnesses the maximum data without losing significant modeling or testing capability. The experiments were conducted by leave-one-subject-out cross-validation for subject independent evaluation.

Table 1 Comparison of recognition accuracies (%) of single feature and multiple features (The feature dimension is in the brackets).

	Feature fusion			Decision fusion
Features	PCA	LDA	SLPP	Majority voting
$DG_{DisFace}$	84.2 (4)	89.4 (1)	87.1 (2)	-
$DG_{DisPose}$	76.6 (10)	78.4 (1)	85.6 (2)	-
$DA_{gradient}$	79.9 (14)	81.1 (1)	84.3 (1)	-
DA_{LPQTOP}	79.9 (4)	81.3 (1)	85.4(3)	-
DA_{LBPTOP}	90.9 (5)	91.1 (1)	92.8 (2)	-
$DG_{DisFace}$ & $DG_{DisPose}$	79.9 (15)	88.0 (1)	88.8 (1)	86.4
$DG_{DisFace}$ & $DA_{gradient}$	81.3 (11)	83.8 (1)	87.5 (3)	87.7
$DG_{DisPose}$ & $DA_{gradient}$	79.9 (13)	76.4 (1)	85.4 (2)	84.0
$DG_{DisFace}$ & $DG_{DisPose}$ & $DA_{gradient}$	79.9 (13)	81.9 (1)	87.6 (2)	86.4
$DG_{DisFace}$ & DA_{LPQTOP}	79.9 (4)	81.3 (1)	85.4 (3)	85.2
$DG_{DisPose}$ & DA_{LPQTOP}	78.9 (4)	81.3 (1)	84.1 (1)	81.5
$DG_{DisFace}$ & $DG_{DisPose}$ & DA_{LPQTOP}	79.9 (4)	81.3 (1)	87.6 (3)	85.2
$DG_{DisFace}$ & DA_{LBPTOP}	90.9 (5)	88.4 (1)	95.1 (2)	92.8
$DG_{DisPose}$ & DA_{LBPTOP}	92.3 (4)	88.3 (1)	94.3 (3)	91.4
$DG_{DisFace}$ & $DG_{DisPose}$ & DA_{LBPTOP}	92.8 (4)	90.2 (1)	**95.6 (3)**	**93.8**

The overall accuracy measure was utilized to report the performances which were shown in Table 1. Three types of dimensionality reduction methods including PCA (Principle Component Analysis), LDA (Linear Discriminant Analysis) and SLPP were employed for comparison, and the results illustrated that SLPP performed best for most of the cases. The dimension of PCA was determined by PCAratio (percentage of variance), while in LDA and SLPP, the PCAratio is set to 0.95. The superiority of SLPP is due to the two properties, i.e. supervised and locality preserved, which can reflect the underlying nonlinear manifold that the samples lie on. The experimental comparison is summarized as follows:

- **Appearance & Geometry**: First, single facial feature of temporal texture description or temporal geometric description was evaluated, and DA_{LBPTOP} achieved the highest recognition accuracy of 92.8%. The evaluation result of $DG_{DisFace}$ (87.1%) was better than that of $DG_{DisPose}$ (85.6%), which meant that the facial configuration parameters were more effective than head pose parameters. For different dynamic appearance features, DA_{LBPTOP} (92.8%) outperformed $DA_{gradient}$ (84.3%) and DA_{LPQTOP} (85.4%) significantly. Moreover, the accuracies of geometric features $DG_{DisFace}$ and combination of ($DG_{DisFace}$ + $DG_{DisPose}$) (88.8%) were higher than appearance features $DA_{gradient}$ and DA_{LPQTOP}, while $DG_{DisPose}$ performed poorly comparing to other features.
- **Feature fusion**: The evaluation results of a variety of combinations of temporal appearance features and temporal geometric features were compared. It can be seen that the joint feature could promote the assessment performance, and the recognition accuracy of the joint feature was higher than a single feature. The best performance was obtained by the combination of ($DG_{DisFace}$ + $DG_{DisPose}$ + DA_{LBPTOP}), with the highest recognition accuracy of 95.6%. The evaluation results were close to the best when $DG_{DisFace}$ was concatenated with dynamic appearance feature, while it was not ideal when only $DG_{DisPose}$ was utilized for feature fusion with dynamic appearance feature. Significant difference was found between the recognition accuracy of three-feature fusion of ($DG_{DisFace}$ + $DG_{DisPose}$ + DA_{LBPTOP}) and that of single-feature in statistical T-test. Although the three-feature combination achieved the highest recognition accuracy, no significant difference was found between three-feature combinations and two-feature combinations. Similar result was observed for decision fusion. In spite of it, multiple dynamic facial features provide sufficient information for infants' pain assessment and tolerate to data missing caused by hospitalization environment.
- **Decision fusion**: The decision fusion accuracy of the combination of ($DG_{DisFace}$ + $DG_{DisPose}$ + DA_{LBPTOP}) reached the highest recognition accuracy (93.8%) comparing to other combinations, while it was lower than the best accuracy of feature fusion. Furthermore, the confusion matrices of feature fusion and decision fusion for ($DG_{DisFace}$ + $DG_{DisPose}$ + DA_{LBPTOP}) with SLPP were illustrated in Table 2 (each row of the matrix denotes the instances in an actual class, and each column represents the instants in a predicted class). The results depicted that decision fusion got higher False Positive Rate (FPR, type I error) (6.2%) than that of feature fusion (0%), while the False Negative Rate (FNR, type II error) of decision fusion (6.2%) was much lower than that of feature fusion (25%). The AUC of decision fusion (0.9144) was greater than the AUC of feature fusion (0.8798). For infants pain assessment, it is expected to minimize FNR rather than

Table 2 Confusion matrices of the best results for feature fusion and decision fusion.

	Feature fusion		Decision fusion	
	Pain	No pain	Pain	No pain
Pain	75%	25%	93.8%	6.2%
No pain	0	100%	6.2%	93.8%

FPR when the goal is to correctly detect pain when it occurs, even if there is a degree sacrifice of misclassification for the non-pain state. From this point of view, although the accuracy of feature fusion is higher than that of decision fusion, decision fusion is more suitable for infants' pain assessment, and vice versa.

- **Instance variation**: The infants videos were captured under real condition of NICU, the instances differ from light condition and occlusions. For the decision fusion of ($DG_{DisFace}$ + $DG_{DisPose}$ + DA_{LBPTOP}), the accuracies of ten infants instances achieved 100%, and the other two infants instances were 85.7% and 40%. The lowest accuracy due to frequent head shaking and mouth movement during no-pain procedure. Although sometimes there is side-effect for pose motion related features as infants in normal state may also have head movement, we could see from Table 1 that pose motion related features could enhance the overall accuracy of pain recognition. Some pain instances were misclassified to no-pain as the infants did not show pain expression during the painful procedure. The LBP-based histogram is gray scale invariant and rotation invariant, therefore, the facial features could tolerate the light condition to a degree. However, the accurate recognition for low intensity of pain facial expression is still challenge.

3.3 Correlation of Facial Features to Pain

The contributions of diverse facial descriptors were analyzed for temporal geometric features and temporal appearance features. The influence scores were calculated from the eigenvectors of the optimization problem in SLPP for the joint feature of ($DG_{DisFace}$ + $DG_{DisPose}$ + DA_{LBPTOP}), which could reflect the contributions of each facial descriptor for the best discriminant directions. The optical eigenvector could be obtained by solving the minimization problem expressed by Eq. (1). The eigenvector with the maximum eigenvalue is corresponding to the weight for each facial descriptor parameter. The scores for the same facial descriptor were grouped by summing up the values of all the parameters, and the left side and right side facial descriptors were grouped into one score. The feature contribution scores for temporal geometric feature were shown in Fig. 4.

The figure depicted that some of the facial descriptors of $DG_{DisFace}$ were negatively correlated with pain, such as the eyebrow-to-eye distance ($d_{ebl/r}$), upper-to-lower eyelid distance ($d_{el/r}$), eyebrow-to-mouth distance ($d_{mbl/r}$), and eye-to-mouth distance ($d_{eml/r}$); the other facial descriptors were positively correlated with pain, such as nose-to-mouth distance (d_{nm}), mouth width (d_{mw}) and mouth height (d_{mh}). The most influenced geometric facial descriptors were eye blinking related parameter $d_{el/r}$ and mouth stretch related parameter d_{mh} and d_{mw}. This is consistent with the observation that the infants are awake during the pain experience, and the eye squeeze and mouth move significantly as they

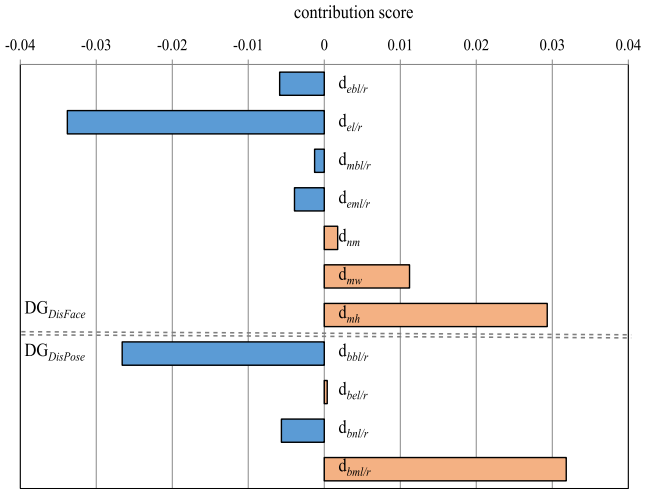


Fig. 4 Contribution scores of temporal geometric facial descriptors.

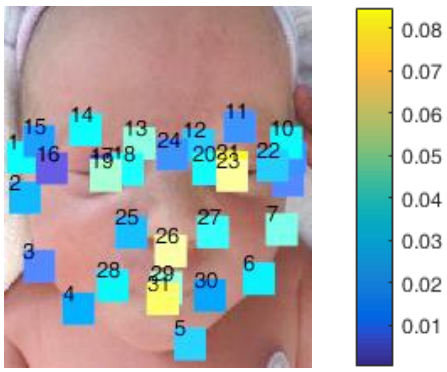


Fig. 5 Visualization of landmark patches importance for DA_{LBPTOP}.

usually cry when feeling pain. Other eye and mouth ($d_{ebl/r}$, $d_{mbl/r}$, and $d_{eml/r}$) related distances reflect similar information instead of providing a new clue for pain measurement, therefore low contribution scores are assigned for these facial descriptors.

The head pose descriptors provided supplemental information for pain assessment. Figure 4 demonstrated that distances between eyebrow/mouth and face boundary played an important role in pain classification, since head movement was very commonly happened during pain experience. The recognition accuracies obtained by using $DG_{DisFace}$ or $DG_{DisPose}$ were 87.1% and 85.6% respectively, which evidenced the high contribution of head pose descriptors.

The importance of landmark patches for the facial descriptors DA_{LBPTOP} was visualized in Fig. 5. The contribution scores of patches were illustrated by different colors. The warmer the color is, the more important the landmark patch is. The illustration depicted that the landmark patches around eyes and mouth were more important than the patches around eyebrow and face boundary. It is consistent with the description of NFCS. The top three important facial descriptors of DA_{LBPTOP} came from patch 21 (eye lid),

31 (mouth lip), and 23 (eye lid). It indicated that the texture changes of eye area and mouth area could more effectively reflect the pain facial reaction.

3.4 Comparison with State-of-Art

We compared our proposed dynamic facial representation to state-of-art researches for infants' pain assessment based on video processing. Our dataset is the same as that exploited in [23] and [22] with increasing number of infants. The previous studies utilized optical flow method to spot and recognize pain facial expression for infants. In [23], an overall accuracy of 96% was obtained for 9 acute pain infants videos with KNN, and the ten-folds cross validation was utilized for evaluation. Comparing to the overall accuracy of 95.6% of our proposed method, the evaluation performance is not significantly improved due to the following two aspects: firstly, the leave-one-subject-out cross-evaluation is more challenging as it is subject independent, and it can lead to lower classification accuracy [33]; secondly, there are more infants videos adopted in our experience, which increase the subject variation and lead to a capability degeneration of the automatic pain recognition system.

The study in [22] is the latest research for infants' pain assessment, on the basis of the same infants video dataset, the recognition accuracy of our scheme was promoted more than 7% comparing to the evaluation accuracy of 88% (facial expression only) reported in [22], which also utilized the leave-one-subject-out cross-evaluation. Therefore, the superiority of the multi-feature fusion for infants' pain assessment strategy is obvious, since the temporal appearance features and temporal geometric features provide various clues for depicting infants pain facial expression characteristics, and the fusion scheme could promote the evaluation accuracy significantly.

4. Conclusion

In this paper, we presented a new set of dynamic pain facial representations by jointly utilizing temporal geometric facial features and temporal appearance facial features. The facial geometric configuration descriptors and head pose descriptors were yielded from the time series of frame-level features. The temporal texture descriptor LBP-TOP was utilized to describe the facial changes over time. Both feature fusion and decision fusion schemes were applied for infants' pain assessment. Experiments were carried out on the video acquired from NICU infants, and the best accuracy of the automatic pain assessment system achieved 95.6% by merging all three types of features. Moreover, we found that although decision fusion did not perform better than that of feature fusion, the FNR of decision fusion (6.2%) was much lower than that of feature fusion (25%). Due to different requirements in clinic application, it is not suitable to conclude which fusion scheme is the best. If the sensitivity of pain assessment is more important for infants' pain monitoring, decision fusion is more suitable for clinic application, even if there may be a degree of misclassification for non-pain state, and vice versa. Besides, although our dataset size is larger than the state-of-art researches, it is still a limited infants pain dataset. Next step more feature selection methods will be employed to further identify the influential features. Moreover, we will keep recording the NICU infants facial reaction videos, and apply the multi-feature fusion system to a larger infants pain dataset.

Acknowledgments

This research is funded by the National Research and Development Major Project (2017YFD0400100), the National Natural Science Foundation of China (NO. 61673052), the grant from Chinese Scholarship Council (CSC), and USF Women's Health Collaborative Grant. We are grateful to the research coordinators (Marcia Kneusel and Judy Zaritt) at Tampa General Hospital for their help in the data collection. We are especially grateful to the parents who had agreed to allow their children to take part in this study.

References

[1] B. Gholami, W.M. Haddad, and A.R. Tannenbaum, "Relevance vector machine learning for neonate pain intensity assessment using digital imaging," IEEE Transactions on Biomedical Engineering, vol.57, no.6, pp.1457–1466, 2010.
[2] J. Vinall, S.P. Miller, V. Chau, S. Brummelte, A.R. Synnes, and R.E. Grunau, "Neonatal pain in relation to postnatal growth in infants born very preterm," Pain, vol.153, no.7, pp.1374–1381, 2012.
[3] American Academy of Pediatrics and Committee on Fetus & Newborn & Section on Surgery and Section on Anesthesiology & Pain Medicine and Canadian Paediatric Society and Fetus and Newborn Committee, "Prevention and management of pain in the neonate: an update," Pediatrics, vol.118, pp.2231–2241, 2006.
[4] A.H. Gee, R. Barbieri, D. Paydarfar, and P. Indic, "Predicting bradycardia in preterm infants using point process analysis of heart rate," IEEE Transactions on Biomedical Engineering, vol.64, no.9, pp.2300–2308, 2017.
[5] J. Lian and Y. Wang, "A review of pain assessment of newborns," Journal of Nursing Science, vol.30, pp.17–40, 2015.
[6] E. Kremer, H.J. Atkinson, and R.J. Ignelzi, "Measurement of pain: patient preference does not confound pain measurement," Pain, vol.10, pp.241–248, 1983.
[7] K. Ho, J. Spence, M.F. Murphy, "Review of pain-measurement tools," Annals of Emergency Medicine, vol.27, no.4, pp.427–432, 1996.
[8] R.C. Coghill, J.G. McHaffie, and Y.-F. Yen, "Neural correlates of interindividual differences in the subjective experience of pain," Proceedings of the National Academy of Sciences of the United States of America, vol.100, no.14, pp.8538–8542, 2003.
[9] R. Rojo, J.C. Prados-Frutos, and A. López-Valverde, "Pain assessment using the Facial Action Coding System. A systematic review," Medicina Clinica, vol.145, no.8, pp.350–355, 2015.
[10] N. Witt, S. Coynor, C. Edwards, and H. Bradshaw, "A guide to pain assessment and managementin the neonate," Current Emergency and Hospital Medicine Reports, vol.4, no.1, pp.1–10, 2016.
[11] R.P. Riddell and N. Racine, "Assessing pain in infancy: the caregiver context," Pain Research and Management, vol.14, no.1, pp.27–32, 2009.
[12] H.D. Hadjistavropoulos, K.D. Craig, R.E. Grunau, and M.F. Whitfield, "Judging pain in infants: behavioural, contextual, and developmental determinants," Pain, vol.73, no.3, pp.319–324, 1997.

[13] H.D. Hadjistavropoulos, K.D. Craig, R.V.E. Grunau, and C.C. Johnston, "Judging pain in newborns: facial and cry determinants," Journal of Pediatric Psychology, vol.19, no.4, pp.485–491, 1994.
[14] J. Kappesser and A.C. de Williams, "Pain and negative emotions in the face: judgements by health care professionals," Pain, vol.99, no.1, pp.197–206, 2002.
[15] B. Goodenough, L. Addicoat, G.D. Champion, M. Mclnerney, B. Young, K. Juniper, and J.B. Ziegler, "Pain in 4- to 6-year-old children receiving intramuscular injections: a comparison of the faces pain scale with other self-report and behavioral measures," The Clinical Journal of Pain, vol.13, no.1, pp.60–73, 1997.
[16] C.C. Johnston and M.E. Strada, "Acute pain response in infants: a multidimensional description," Pain, vol.24, no.3, pp.373–382, 1986.
[17] K.D. Craig, R.V. Grunau, and J. Aquan-Assee, "Judgement of pain in new-borns: facial activity and cry as determinants," Canadian Journal of Behavioural Science, vol.20, no.4, pp.442–451, 1988.
[18] J.W.B. Peters, H.M. Koot, R.E. Grunau, J. de Boer, M.J. van Druenen, D. Tibboel, and H.J. Duivenvoorden, "Neonatal facial coding system for assessing postoperative pain in infants: Item reduction is valid and feasible," The Clinical Journal of Pain, vol.19, no.6, pp.353–363, 2003.
[19] S. Brahnam, C.-F. Chuang, R.S. Sexton, and F.Y. Shih, "Machine assessment of neonatal facial expressions of acute pain," Decision Support Systems, vol.43, no.4, pp.1242–1254, 2007.
[20] L. Nanni, S. Brahnam, and A. Lumini, "A local approach based on a Local Binary Patterns variant texture descriptor for classifying pain states," Expert Systems with Applications, vol.37, no.12, pp.7888–7894, 2010.
[21] S. Brahnam, C.-F. Chuang, F.Y. Shih, and M.R. Slack, "Machine recognition and representation of neonatal facial displays of acute pain," Artificial Intelligence in Medicine, vol.36, no.3, pp.211–222, 2006.
[22] G. Zamzmi, C.-Y. Pai, D. Goldgof, R. Kasturi, T. Ashmeade, and Y. Sun, "An approach for automated multimodal analysis of infants pain," 23rd International Conference on Pattern Recognition (ICPR), pp.4143–4148, 2016.
[23] G. Zamzmi, G. Ruiz, D. Goldgof, R. Kasturi, Y. Sun, and T. Asheade, "Pain assessment in infants: towards spotting the pain expression based on the facial strain," 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), pp.1–5, 2015.
[24] G. Zamzmi, C.-Y. Pai, D. Godgof, R. Kasturi, Y. Sun, and T. Ashmeade, "Automated pain assessment in neonates," Scandinavian Conference on Image Analysis, vol.10270, pp.350–361, 2017.
[25] E. Fotiadou, S. Zinger, W.E. Tjon a Ten, S.B. Oetomo and P.H.N. de With, "Video-based facial discomfort analysis for infants," Visual Information Processing and Communication V, 90290F, pp.1–6, 2014.
[26] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance models," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.23, no.6, pp.681–685, 2001.
[27] J.W.B. Peters, H.M. Koot, R.E. Grunau, J. de Boer, M.J. van Druenen, D. Tibboel, and H.J. Duivenvoorden, "Neonatal facial coding system for assessing postoperative pain in infants: item reduction is valid and feasible," Clinical Journal of Pain, vol.19, no.6, pp.353–363, 2003.
[28] P. Werner, A. Al-Hamadi, K. Limbrecht-Ecklundt, S. Walter, S. Gruss, and H.C. Traue, "Automatic pain assessment with facial activity descriptors," IEEE Transactions on Affective Computing, vol.8, no.3, pp.286–299, 2017.
[29] G. Zhao and M. Pietikäinen, "Dynamic Texture recognition using local binary patterns with an application to facial expressions," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.29, no.6, pp.915–928, 2007.
[30] B. Jiang, M. Valstar, B. Martinez, and M. Pantic, "A dynamic appearance descriptor approach to facial actions temporal modeling," IEEE Transactions on Cybernetics, vol.44, no.2, pp.161–174, 2014.
[31] K. Zhao, W.-S. Chu, F. De la Torre, J.F. Cohn, and H. Zhang, "Joint patch and multi-label learning for facial action unit and holistic expression recognition," IEEE Transactions on Image Processing, vol.25, no.8, pp.3931–3946, 2016.
[32] X. He and P. Niyogi, "Locality preserving projections," Advances in Neural Information Processing System (NIPS), vol.16, pp.153–160, 2003.
[33] S. Brahnam, L. Nanni, and R. Sexton, "Introduction to neonatal facial pain detection using common and advanced face classification techniques," Adavanced Computational Intelligence Paradigms in Healthcare-1, vol.48, pp.225–253, 2007.

Ruicong Zhi received the Ph.D. degree in signal and information processing from Beijing Jiaotong University in 2010. From 2016~2017, she visited the University of South Florida as a visiting scholar. She visited the Royal Institute of Technology (KTH) in 2008 as a joint Ph.D. She is currently an associate professor in School of Computer and Communication Engineering, University of Science and Technology Beijing. She has published more than 50 papers, and has six patents. She has been the recipient of more than ten awards, including the National Excellent Doctoral Dissertation Award nomination. Her research interests include facial and behavior analysis, artificial intelligence, and pattern recognition.

Ghada Zamzmi received the MS degree in Computer Science from University of South Florida, 2015. She is currently working toward the PhD degree at Computer Science and Engineering, University of South Florida. Her research interest includes computer vision, image/video analysis, and emotion recognition for healthcare and human - computer interface.

Dmitry B. Goldgof has received the M.S. degree in Computer and Systems Engineering from the Rensselaer Polytechnic Institute and the Ph.D. degree in Electrical Engineering from the University of Illinois at Urbana-Champaign. He is currently Professor in the Department of Computer Science and Engineering at the University of South Florida and a Professor, Department of Oncological Sciences, USF Health. His research interests include Medical Image Analysis, Image and Video Processing, Computer Vision and Pattern Recognition, Ethics and Computing, Bioinformatics and Bioengineering. He has published 94 journal and over 200 conference publications (with high citations, h=50, g=88), 20 books chapters and edited 5 books. His work has been funded by numerous agencies including NIH, NSF, ONR, DOD, ARDA (IARPA), DARPA, NIST, FDOT etc. Professor Goldgof is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE) "for contributions to computer vision and biomedical applications", Fellow of the International Association for Pattern Recognition (IAPR) "for contributions to computer vision, pattern recognition, and biomedical engineering", Fellow of the American Association for the Advancement of Science (AAAS) "for distiguished contribution to the fields of computer vision, pattern recognition and biomedical applications, particularly in biomedical image analysis", and Fellow of American Institute of Medical and Biomedical Engineering (AIMBE). Professor Goldgof has served as IEEE Distinguished Visitor 2004-2006. In 2008 Professor Goldgof was selected for USF Theodore and Venette Askounes-Ashford Distinguished Scholar Award.

Terri Ashmeade is a neonatologist in Tampa, Florida and is affiliated with multiple hospitals in the area, including Johns Hopkins All Children's Hospital and Tampa General Hospital. She is also a professor in College of Medicine Pediatrics, University of South Florida. Dr. Ashmeade received her medical degree from University of Connecticut School of Medicine and has been in practice for more than 20 years. She is one of 32 doctors at Johns Hopkins All Children's Hospital and one of 12 at Tampa General Hospital who specialize in Neonatal-Perinatal Medicine.

Tingting Li received the Bachelor degree in Computer Science from University of Science and Technology Beijing in 2016. She is currently pursuing the MS degree at School of Computer and Communication Engineering, University of Science and Technology Beijing. Her research interest includes computer vision and emotion analysis.

Yu Sun received his Ph.D. degree in computer science from the University of Utah in 2007, B.S. and M.S. degrees in electrical engineering from Dalian University of Technology, Dalian, China, in 1997 and 2000, respectively. He was a Postdoctoral Associate at Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA from Dec. 2007 to May 2008 and a Postdoctoral Associate in the School of Computing at the University of Utah from May 2008 to May 2009. His research interests include robotics, haptics, computer vision, human computer interaction (HCI), and medical applications. He currently serves as a founding Co-Chair of the new Technical Committee on Robotic Hands, Grasping, and Manipulation of the IEEE Robotics and Automation Society (IEEE-RAS). He also co-chairs the Membership Services Committee of IEEE-RAS and sits on the Member Activity Board of IEEE-RAS.