Recognition of Neonatal Facial Expressions of Acute Pain Using Boosted Gabor Features

Lei Yuan
Dept. of Computer Science and Engineering
Arizona State University
Tempe, Arizona, USA
Email: lyuan0388@gmail.com

Forrest Sheng Bao
Dept. of Computer Science / Dept. of Electrical & Computer Engineering
Texas Tech University
Lubbock, Texas, USA
Email: forrest.bao@gmail.com

Guanming Lu
College of Telecommunication and Information Engineering
Nanjing Univ. of Posts & Telecom.
Nanjing 210003, China
Email: lugm@njupt.edu.cn

Abstract—Facial expressions are considered a critical factor in neonatal pain assessment. This paper proposes a pain expression recognition method using boosted Gabor features. Each neonatal facial image is convoluted with the 2D Gabor filters to extract 412,160 Gabor features. Since the high-dimension Gabor feature vectors are quite redundant, we employ a modified version of AdaBoost algorithm to select and combine the most informative features for classification. The "pain vs. non-pain" problem is treated as two sub-problems by using a coarse-to-fine hierarchical classifier. Experiments with 510 neonatal expression images show that the proposed method is quite effective. Only 30 Gabor features are enough to achieve good classification performance. The recognition rate of pain versus non-pain is up to 88% (i.e. error rate $\epsilon = 0.12$). Compared with one existing algorithm for neonatal pain recognition, our approach can reach similar accuracy in much lower time complexity.

Index Terms—Neonatal pain recognition, AdaBoost, Feature Selection, Float Search

I. INTRODUCTION

In recent years, the question of long-term behavioral effects of repetitive pain in neonates has attracted the attention of health professionals [1]–[3]. The need for prevention and management of pain in neonates has gained much acceptance [4]. Various pain assessment measures (tools, instruments, etc.) based on facial expressions of pain have been developed [5]–[7]. The main controversies surrounding the use of such assessment tools are the subjectivity of the observer. Thus, an automated recognition of facial expressions of pain not relying on health professionals has potential medical significance [8].

The objective of this study is to bypass the observational problems by developing a machine classification system to diagnose neonatal facial expressions of pain. Besides Brahnam et al. [8], [9], few works have been conducted in this field. In [8], [9], the features are extracted by applying PCA directly onto raw image pixels. Those features are then classified using various methods such as linear discriminant analysis (LDA) and support vector machines (SVMs). A total of 70 features are used, and the minimum error rate achieved by those classifiers is 0.12. In this paper, we will employ a more expressive kind of feature, the Gabor feature, to the representation of neonatal faces. Then, a more efficient way of combining these features for the classification will be developed. Our classifier can also achieve an error rate of 0.12. But our complexity is lower than the existing approach.

The rest of the paper will be organized as follows: in section II, background knowledge of automated facial expression analysis will be given. After that, we will present our novel version of boost learning algorithm with hybrid strategy in section III. Experimental methods and results will be given in section IV.

II. AUTOMATIC FACIAL EXPRESSION ANALYSIS

Over the last decades, automatic facial expression analysis has become an active research field that shows high potential in many areas, such as the automated assessment of neonatal pain.

As stated in [10]–[12], an automated facial expression analysis system consists of three parts: face detection, facial feature extraction/representation and classification. In this paper, we will focus on the last two parts.

Feature extraction is the first part of our proposed system. Various methods [12] have been developed to construct suitable features for face expression analysis. In general, facial feature extraction methods can be categorized according to whether they focus on deformation or motions of faces, respectively whether they act locally or holistically [12].

The last stage of this system is classification. After the feature vectors are extracted, various methods from pattern classification and recognition [13], [14] could be adopted to handle the multi-label classification problem. Here we deploy the boost learning algorithm [15], a method that could perform the feature selection and classification simultaneously.

III. PROPOSED SCHEME

The general structure of our proposed scheme is illustrated in Fig. 1. There are two working modes of our proposed scheme, the training mode and the testing mode. In the training mode, a full set of 40 Gabor filters will be applied to the pre-processed training images to extract the features. Then, the over-complete features are sent to our novel boost learning algorithm, the HybridBoost, to construct a strong classifier by selecting a small set of them. In the testing mode, only
the selected feature will be extracted, then the trained strong classifier is used for facial expression recognition.

![Diagram of facial expression recognition process]

In the following subsections, we will discuss these parts in detail.

A. Gabor Filters

Gabor filters have attracted much attention since firstly pioneered by L. Wiskott et al. [16] in face recognition applications. Considering its excellent capacities and great success in facial recognition [17], we choose Gabor features as the representation of neonatal faces.

In spatial domain, the 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal wave:

$$\psi_{u,v}(z) = \frac{1}{\sigma^2} e^{-\frac{|k_{u,v}|^2 + |k_{u,v}|^2 z^2}{2\pi}} [\cos(\varphi_{u,v} + z^2)]$$  \hspace{1cm} (1)

where $k_{u,v} = k_{u,v} e^{i\phi_{u,v}}$, $\varphi_{u,v} = \frac{2\pi u}{v}$ and $\phi_{u,v} \in [0, \pi)$.

The Gabor features are simply obtained by convoluting the image with a series of Gabor filters. Since our Gabor filters have 40 different parameter combinations, 5 v’s and 8 u’s, we have 400 sets of convolution results to one single image.

B. AdaBoost Algorithm

The dimension of the Gabor feature vector is very high. For example, if the size of an image is $112 \times 92$, and 5 scales and 8 orientations are selected, the dimension of the Gabor feature vector will be $112 \times 92 \times 5 \times 8 = 412,160$. It is difficult to operate feature vectors of such a high dimension.

With the success of boosting in the face detection system [10], AdaBoost [15], shows great potential in selecting the most discriminative features as well as combining them to solve the classification problem at the same time.

AdaBoost algorithm is implemented by maintaining the distribution $D_t$ over the training sets during the iteration. $D_t$ can be considered as the weights applied to the training samples. During each iteration, the weights are updated according to the error rate currently achieved. The weights of the misclassified samples increase while those of the others decrease. Thus, the following iteration will focus on those "hard" samples.

However, the strategy behind AdaBoost minimizes a function of the margin over the training set, while the ultimate goal in automated neonatal pain expression analysis is to minimize a cost directly (usually linearly) associated with the error rate. This motivated us to develop a more effective boosting learning algorithm for this application. Thus, a strong classifier learned by AdaBoost could be suboptimal for current application in terms of error rate [18]. Li et al. [19] proposed a novel boosting algorithm, the FloatBoost, to solve this problem. However, in our application. This might lead to potential in over-fitting. In our algorithm, we consider to hybrid this strategy with the original one in AdaBoost which minimizes the margin function. With this modification, we developed our novel boost learning algorithm, called the HybridBoost, which combines both the strategies from original AdaBoost and FloatBoost. We attempts to balance the theoretical objective function of the margin and the intuitive one of the error rate. The HybridBoost will be explained below.

C. Modified Boosting with Hybrid Strategy

The first step of a Boosting algorithm in one iteration is to train and select the weak classifier with the minimum weighted error. This is the Forward Inclusion step. In this step, we adopt the original strategy described in AdaBoost [15], to select the trained weak classifier with minimum weighted error defined by current distribution, and assign the classifier with a factor $\alpha$ as the weight of its hypotheses. This can be seen as a strategy that prevents the learning from over-fitting.

After the classifier is selected and put into a "pool" of chosen classifiers, a procedure called "Conditional Exclusion" is carried out. In this procedure, each classifier in the pool is taken out alone, and the error rate on the training set achieved by the rest of them is evaluated. For example, If we current have selected 10 weak classifiers, then we can get 10 results by taking each of them out to test the error rate. If the best results is better than the minimum error rate achieved so far, the corresponding classifier will be excluded, and the procedure continues. If not, the iteration ends and the next round of forward inclusion begins. This is the strategy that tries to make the strong classifier optimal in term of error rate, with less selected features.

The pseudo-code of Modified Boosting with hybrid strategies (HybridBoost) is given in Algorithm 1.

D. The Course to Fine Classifier Hierarchy

There are three categories of neonatal facial expressions for us to deal with here, calm, cry and pain. Because the intention of this study is to distinguish pain expressions from non-pain expressions, it is intuitive to consider this as a two-class classification job and build only one strong classifier. However, the classifiers (features) needed to distinguish calm faces from non-calm faces might be different from the ones needed to distinguish crying faces from painful faces. So a hierarchy of classifiers would help the proper use of each feature and enhance the overall accuracy. The structure of the hierarchy is given in Fig. 2.

IV. RESEARCH METHOD

In our experiments, we tried to distinguish cry expressions that were in response to pain from those cry expressions that
The HybridBoost Algorithm

1) (Input)
   a) A set of \( m \) samples \( \{(x_1, y_1), \ldots, (x_m, y_m)\} \), with labels \( y_i \in Y = \{+1, -1\} \), \( m = a + b \), of which \( a \) samples have \( y_i = +1 \) and \( b \) samples have \( y_i = -1 \);
   b) The maximum number \( T \) of weak classifiers;
   c) The error rate \( \epsilon(H_t) \), and the acceptance threshold \( \epsilon^* \).

2) (Initialization)
   a) \( D_0(i) = \frac{1}{m} \) for those samples with \( y_i = +1 \) or \( D_0(i) = \frac{1}{m} \) for those samples with \( y_i = -1 \);
   b) \( \epsilon_{min} = 1, t = 1, \mathcal{H}_0 = \{\}, \mathcal{A}_0 = \{\} \), where \( \mathcal{H} \) and \( \mathcal{A} \) is the sets of currently selected weak hypotheses \( h \) and \( \alpha \), respectively;
   c) Train all the weak classifiers, one for each feature attribute, using FindAttrTest.

3) (Forward Inclusion)
   a) \( t \leftarrow t + 1 \);
   b) Select the weak classifier \( h_t \) with the minimum weighted error;
   c) Update \( \alpha_t, \beta_t, D_t \) according to AdaBoost.M1 [15];
   d) \( \mathcal{H}_t = \mathcal{H}_{t-1} \cup h_t, \mathcal{A}_t = \mathcal{A}_{t-1} \cup \alpha_t \)

4) (Conditional Exclusion)
   a) \( h^* = \arg \min_{h \in \mathcal{H}_t} \epsilon(\mathcal{H}_t - h) \), with corresponding \( \alpha' \);
   b) If \( \epsilon(\mathcal{H}_t - h^*) < \epsilon_{min} \), then:
      i) \( \mathcal{H}_{t-1} = \mathcal{H}_t - h^*, \mathcal{A}_{t-1} = \mathcal{A}_t - \alpha' \), \( \epsilon_{min} = \epsilon(\mathcal{H}_t - h^*) \);
      ii) \( t \leftarrow t - 1 \);
      iii) goto 3a;
   c) else:
      i) If \( t = T \), or \( \epsilon_{min} < \epsilon^* \), goto 5;
      ii) otherwise goto 3a

5) (Output)
   \[
   h_{fin}(x) = \text{sign} \left[ \sum_{h(x) \in \mathcal{H}_T, \alpha \in \mathcal{A}_T} \alpha h(x) \right]
   \]

---

**Algorithm 1 The HybridBoost Algorithm**

---

**Fig. 2. The hierarchy structure of strong classifiers**

---

**Fig. 3. Examples of the three facial expressions in neonate data set**

---

**A. Experimental procedures**

The experimental procedures can be divided into the following stages: pre-processing, feature extraction, feature selection and classification.

The pre-processing stage includes sub-image segmentation, scale normalization and gray-scale normalization. The original 510 images are manually cropped to the size of \( 112 \times 92 \) pixels, rotated and scaled.

In the feature extraction stage, 2D Gabor filter is applied to extract the expression features from facial images. We use 5 scales and 8 orientations, and the number of Gabor features for one image is \( 112 \times 92 \times 5 \times 8 = 412,160 \).

Finally, in the feature selection and classification stage, the proposed HybridBoost is applied to select the most informative features/weak classifiers, and a hierarchy of strong classifiers are constructed. In this stage, two experiments has been done to evaluate the performance.

In Experiment 1, we randomly selected 370 images as the training set, and 140 images as the testing set. A hierarchy of 2 strong classifiers are trained on the training set, with a minimum error rate \( \epsilon_{min} \) requirement, \( \epsilon_{min} = 0.01 \). The number of selected features to achieve this requirement, \( T \), is recorded. Then, we trained the strong classifiers with different number of weak classifiers, and tested their error rate on the test set.

In Experiment 2, we put all the images together, and a 10-fold cross-validation technique was applied. The error rates were averaged to obtain a final performance evaluation.

**B. Experimental Results and Analysis**

1) Experiment I: The results of this experiment are illustrated in Fig. 4, 5 and Table I. In Table I, the numbers were in response to a less noxious stimulus. Thus, two stimuli are included in this study: (1) heel puncture, (2) transporting the neonate from one crib to another.

The images used in the experiments, some of which as shown in Fig. 3, were divided into two sets: training set and testing set, based on facial expression categories, not subjects. As a result, the training set or testing set contained multiple samples of each subject in each category pair.
The error rate for the "Calm vs. non-Calm" sub-problem.

The error rate for the "Pain vs. Cry" sub-problem.

2) Experiment II: The experimental results of this experiment is shown in Table II. In this experiment, the overall performance of our proposed hierarchy structure is evaluated. With a total of 30 weak classifiers, the error rate for the "Pain vs. non-Pain" problem is 0.12, a quite good results comparable to [9], with much less complexity.

REFERENCES


<table>
<thead>
<tr>
<th>Problem</th>
<th>$T$</th>
<th>$\epsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain vs. non-Pain</td>
<td>30</td>
<td>0.12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sub-problem</th>
<th>$T$</th>
<th>$\epsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calm vs. non-Calm</td>
<td>20</td>
<td>0.03</td>
</tr>
<tr>
<td>Cry vs. Pain</td>
<td>22</td>
<td>0.19</td>
</tr>
</tbody>
</table>

TABLE I

SUB-PROBLEM PERFORMANCE

TABLE II

OVERALL PERFORMANCE