IEEE Int. Conference Neural Networks & Signal Processing
Zhenjiang, China, June 8~10, 2008

FACIAL EXPRESSION RECOGNITION FOR NEONATAL PAIN ASSESSMENT

Guanming Lu [1], *Xiaonan Li* [2], *Haibo Li* [3]
1. Nanjing University of Posts and Telecommunications, Nanjing, 210003, P.R.China
lugm@njupt.edu.cn
2. Nanjing Children's Hospital Affiliated to Nanjing Medical University, Nanjing, 210008, P.R.China
xnli@njmu.edu.cn
3. Department of Applied Physics and Electronics, Umeå University, Umeå, S-901 87, Sweden
haibo.li@tfe.umu.se

ABSTRACT

Facial expressions are considered a critical factor in neonatal pain assessment. This paper attempts to apply modern facial expression recognition techniques to the task of distinguishing pain expression from non-pain expression. Firstly, 2D Gabor filter is applied to extract the expression features from facial images. Then we apply Adaboost as a feature selection tool to remove the redundant Gabor features. Finally, the Gabor features selected by Adaboost are fed into the support vector machines (SVMs) for final classification. 510 facial images are investigated by using SVMs. The best recognition rates of pain versus non-pain (85.29%), pain versus calm (94.24%), pain versus cry (78.24%) were obtained from an SVM with a polynomial kernel of degree 3. The results of this study indicate that the application of SVM technique in pain assessment is a promising area of investigation.

Key Words —— Neonatal Pain, Expression Recognition, Support Vector Machine, AdaBoost, Gabor filer

1. INTRODUCTION

Facial expression recognition has attracted much attention in recent years[1, 2]. Much work has been done on classifying facial expressions by computer, but this has mostly been confined to Ekman's six basic emotions: anger, happiness, fear, disgust, sadness and surprise[3−5]. Pain expression recognition, however, appears to have been largely neglected. It has potential medical significance. In recent years, the question of long-term consequences of pain in neonates has prevalently attracted the attention of health professionals due to the fact that modern neonatal intensive care units (NICU) routinely employ procedures that involve acute pain and are often followed by local inflammation and hyperalgesia lasting for several hours or even days[6−7]. Today, the need for prevention and management of pain in neonates has gained universal acceptance[8].

Neonatal pain assessment is considered one of the most challenging tasks in neonatology because neonates cannot verbalize their pain experiences[9]. Various pain assessment measures (tools, instruments) have been developed, based on behavioral indicators of pain alone or a combination of behavioral and physiological indicators[10]. Facial expression is considered the most sensitive indicator of acute and postoperative pain in neonates[11]. Based on such research, Grunau and Craig developed Neonatal Facial Coding System (NFCS) in 1987[12].

However, the main controversies surrounding the use of such assessment tools are the subjectivity of the observer. Because these tools rely on the observations of health professionals, and studies have demonstrated that health professionals are not entirely impartial in their judgments[13]. The automatic recognition of facial expressions of pain has potential medical significance. But there appears to be little previous works on computer classification of pain expression[14], though Pantic and Rothkrantz speculate about the feasibility of doing so[15].

This paper attempts to tackle these problems by applying modern facial expression recognition techniques to the task of distinguishing pain expression from non-pain expression. In our work, 2D Gabor filter is applied to extract the expression features from facial images, then AdaBoost is trained on them to select those most informative features (named by AdaBoosted features) from all the original high-dimensional pixel-wise dense Gabor features, the resulting AdaBoosted features are fed into the SVM for final classification.

The paper is organized as follows. The research methods are described in Section 2. Section 3 shows the experimental results. Conclusions are drawn in Section 4.

2. RESEARCH METHODS

The initial intention of this study was to distinguish pain expressions from a variety of non-pain expressions.

According to our clinical observations, on average, there are about 2 to 5 times painful procedures, venipuncture for metabolic screening and serum bilirubin test, which need to be done during the first three days in hospital for a healthy neonate. In addition, transporting the neonate from one crib to another will provoke a cry expression that is similar to pain expressions but not in response to pain. We were thus afforded the opportunity of contrasting classification recognition rates of cry expressions that were in response to pain to those cry expressions that were in response to a less noxious stimulus. Thus, two stimuli are included in this study: (1) venipuncture, (2) transporting the neonate from one crib to another.

The images used in the experiments were divided into 2 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.

2.1. Subjects

Subjects were health, full term neonates and recruited in from the neonatal room of the Secondary Affiliated Hospital of Nanjing Medical University. The neonates were required to receive newborn metabolic disease screening, using heel puncture at third day of birth. Those who met the screening criteria when transferred into neonatal room will be randomly selected for this study. Informed consent was obtained from their parents.

2.2. Data collection

The facial images of 57 neonates (30 boys and 27 girls) were captured at different photo sessions while the neonates were experiencing two stimuli:

(1) Transport from one crib to another (calm/cry): after being transported from one crib to another, the neonate was swaddled and a series of photographs was taken during the procedure. The state of the neonate was noted as either cry or calm for each photograph taken in the series.

(2) Heel puncture (pain): after resting for at least 1 min, the external lateral surface of the heel was punctured for blood collection. Several continuous photographs of the neonate's face were taken starting immediately after introduction of the lancet and while the skin of the heel was squeezed for blood samples.

Of the 510 facial images, there are 170 calm images, 180 cry images, and 160 pain images. The non-pain image set consists of all the calm and cry images. Fig. 1 provides two example sets of the three neonate facial expressions of calm, cry, and pain.

Fig. 1 Examples of the three neonate facial expressions

2.3. Cross validation

Because the number of images in the dataset is small, in order to evaluate generalization performance, a 10-fold cross-validation testing scheme was adopted in all experiments. The cross-validation testing scheme included following four steps:

(1) The images were randomly divided into 10 segments in terms of the category pairs being examined.

(2) The first segment was used as the testing set, the remaining 9 segments were used as the training set, and an average classification recognition rate was obtained from the testing set.

(3) The second segment was used as the testing set instead of the first segment, step 2 was repeated. The third segment was used as the testing set, and so on. step 2 was repeated 10 times in total.

(4) The 10 classification recognition rates were averaged to obtain a final recognition rate.

As an example of this process for pain versus non-pain classification experiment, suppose there are 160 pain images and 350 non-pain images. In step 1, 510 images were randomly divided into 10 segments, and each segment included 16 pain and 35 non-pain images. In step 2, we would choose the one segment as the testing set, let the remaining 144 pain and 315 non-pain images become the training set. Then we would train the classifiers and obtain the recognition rate for the testing set. For example, if 6 images out of the 51 images in the testing set were wrongly classified, then the classification recognition rate for this run would be $(51-6)/51 = 88.24\%$. In step 3, we would use another segment as the testing set, and step 2 was repeated 10 times in total. In step 4, we would average the 10 classification recognition rate for the final recognition rate.

2.3. Experimental procedures

As illustrated in Fig. 2, the experimental procedures can be divided into the following stages: preprocessing, feature extraction, feature selection and classification.

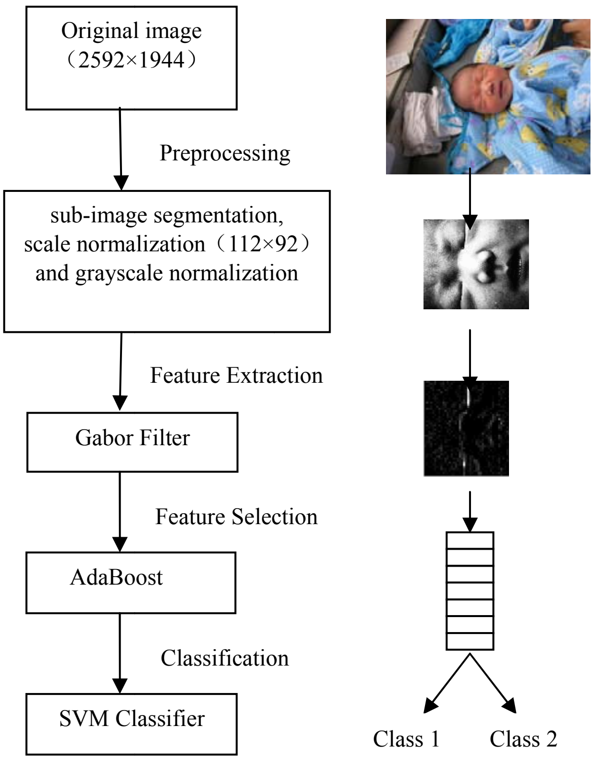

Fig. 2 Experimental procedures

The pre-processing stage includes sub-image segmentation, scale normalization and grayscale normalization. Preprocessing influences the efficiency of expression features extraction and the computation. In order to focus this study on the classifier performance, the original facial images are manually segmented, rotated and scaled using Adobe Photoshop 7 such that the eyes lie roughly along the same axis. Each original image with the size of 2592×1944 pixels is cropped to the size of 112×92 pixels. The task of grayscale normalization is to convert color images to grayscale ones and to adjust contrast and brightness of individual regions in the image. Histogram equalization is primarily used to improve overall contrast of the image to enhance foreground detail where the face is located. Image sharpening increases the intensity of the edges of the facial details.

In the feature extraction and selection stage, 2D Gabor filter is applied to extract the expression features from facial images[16], then AdaBoost is trained on them to select those most informative features (named by AdaBoosted features) from all the original high-dimensional pixel-wise dense Gabor features[17]. In our experiments, 900 AdaBoosted features are selected from 412160 original Gabor features.

Finally, in the classification stage, SVM with linear kernel, SVM with a polynomial kernel (degree=2, 3, 4), SVM with Radial Basis Function (RBF) are used to classify the feature vectors into the following category pairs: pain versus non-pain, pain versus calm, pain versus cry.

3. EXPERIMENTAL RESULTS

3.1. Feature selection

Our experiments examined the effect of feature selection by Adaboost on the classification performance. For each category pair, we searched for the number of features which gave high recognition accuracy. Experimental results show that 900 features which were selected from all 412160 Gabor features are enough. With more features exploited, the performance does not improve any longer. The scale (v) and orientation (u) distribution of the selected Gabor features is given in Table 1 for pain versus non-pain two-class recognition. Clearly, the higher frequency filters (smaller scales) are chosen much more often than low frequency ones, as we can see that the kernels with 0-scale are about 30% in the 900 Gabor features. According to the distribution of the selected features in different orientation, we discover that features in orientation $\pi/2$ (with u=4) has more discriminative power.

Table 1. The scale (v) and orientation (u) distribution of the selected Gabor features

v \ u	0	1	2	3	4	5	6	7	Total
0	38	32	20	31	62	35	39	21	278
1	26	25	15	17	38	26	27	18	192
2	14	14	11	12	27	14	11	6	109
3	15	19	14	20	29	18	17	10	142
4	23	24	13	18	40	25	25	11	179
Total	116	114	73	98	196	118	119	66	900

3.2. Classification

A face image is represented by a vector with 900 Gabor wavelet features selected by AdaBoost. Classification is performed on the training set and the testing set by using SVMs for each category pair. Tables 2 shows the classification recognition rates of linear SVM classifier, non-linear SVM classifier with a polynomial kernel of degree 2, 3, or 4, and non-linear SVM classifier with RBF kernel functions, for each of the following category pairs: pain versus calm, pain versus cry, pain versus non-pain.

Table 2. classification recognition rates

method \ category pair	pain vs. calm	pain vs. cry	pain vs. non-pain
SVM with linear kernel	90.30%	72.94%	83.73%
SVM with polynomial kernel (degree=2)	92.72%	75.88%	82.75%
SVM with polynomial kernel (degree=3)	94.24%	78.24%	85.29%
SVM with polynomial kernel (degree=4)	92.12%	75.29%	82.16%
SVM with RBF kernel	89.70%	71.47%	78.24%

4. CONCLUSIONS AND FUTURE DIRECTIONS

Most facial expression recognition algorithms have focused on adult faces. The recognition for expression of pain in neonates appears to have been largely neglected. However, it has potential medical significance. This study is one of the first to recognize expression of pain in neonates.

Gabor filters have attracted much attention and achieved great success in face recognition area, but the dimensionality of the Gabor feature vector is very high. This paper investigated the dimensionality reduction of Gabor features using AdaBoost and analyzed the scale and orientation distribution of the selected Gabor features. The experiment results have shown that AdaBoosted Gabor features are not only low dimensional but also discriminant, 900 Gabor features are enough to achieve high recognition accuracy. Significant reduction in computation and memory cost has been achieved since the number of convolution operations has been reduced from 412160 to 900 for the image of 112×92 pixels. A series of experiments compared the recognition rates of SVMs classifying the following category pairs: pain versus non-pain, pain versus calm, pain versus cry. We concluded that SVM with a polynomial kernel of degree 3 provided the best recognition rates of pain versus non-pain (85.29%), pain versus calm (94.24%), pain versus cry (78.24%). The result of this study indicates a potential research area for developing a neonatal pain assessment system based on facial expressions.

Our future work would focus on the further investigation of the AdaBoosted Gabor features. Also, we are trying other feature selection tools and comparing their performance with AdaBoost-based method. Another research direction would be to address multi-class decisions for evaluating pain intensity by combining clinical pain assessment measures based on physiological indices. In addition, automatic face detection and expression feature extraction from video are also our future works.

5. ACKNOWLEDGEMENTS

This research was Sponsored by Swedish Research Council and the Swedish International Development Co-operation Agency (Grant No. 348-2005-6434), the Scientific Research Foundation for the Returned Overseas Chinese Scholars, Nanjing Municipal Personnel Bureau (Grant No. TJ206015) and Nanjing health Bureau (No. VKX07020).

6. REFERENCES

[1] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, Elsevier, Amsterdam, Netherlands, pp. 259-275, 2003.

[2] N. Sebe, M.S. Lew, Y. Sun, I. Cohen, T. Gevers, and T.S. Huang, "Authentic facial expression analysis," *Image and Vision Computing*, Elsevier Ltd, United Kingdom, pp. 1856-1863, 2007.

[3] Gwen Littlewort, Marian Stewart Bartlett, Ian Fasel, Joshua Susskind, and Javier Movellan, "Dynamics of facial expression extracted automatically from video," *Image and Vision Computing*, Elsevier Ltd, United Kingdom, pp.615-625, 2006.

[4] Bouchra Abboud, Franck Davoine, and Mô Dang, "Facial expression recognition and synthesis based on an appearance model," *Signal Processing: Image Communication*, Elsevier, Amsterdam, Netherlands, pp.723-740, 2004.

[5] I. Kotsia and I. Pitas, " Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE Transactions on Image Processing*, Institute of Electrical and Electronics Engineers Inc., Piscataway, United States , pp. 172-187, 2007.

[6] K.J.S. Anand, "Clinical importance of pain and stress in preterm neonates," *Biology of the Neonate*, S. Karger AG, Switzerland, pp. 1-9, 1998.

[7] M. S. Lidow, "Long-term effects of neonatal pain on nociceptive systems," *Pain*, International Association for the Study of Pain (IASP), pp. 377–383, 2002.

[8] K.J.S. Anand, The International Evidence-Based Group for Neonatal Pain, "Consensus statement for the prevention and management of pain in the newborn," *Archives of Pediatric and Adolescent Medicine*, pp. 173-180, 2001.

[9] Pat Hummel and Monique van Dijk, "Pain assessment: Current status and challenges," *Seminars in Fetal & Neonatal Medicine*, Elsevier Science , pp. 237-245, 2006.

[10] Huda Huijer Abu-Saad, Gerrie J.J.w. Bours, Bonnie Stevens, and Jan P.H. Hamers, "Assessment of pain in the neonate," *Seminars in Perinatology*, pp. 402-416, 1998.

[11] K.D. Craig, M.F. Whitfield, R.V.E. Grunau, "Pain in the preterm neonate: behavioral and physiological indices," *Pain*, International Association for the Study of Pain (IASP) , pp. 287-99, 1993.

[12] R. V. E. Grunau and K. D. Craig, "Pain expression in neonates: facial action and cry," *Pain*, International Association for the Study of Pain (IASP) , pp. 395-410, 1987.

[13] K.M. Prkachin, P. Solomon, T. Hwang, and S.R.

Mercer, “Does experience influence judgments of pain behavior? Evidence from relatives of pain patients and therapists,” *Pain Res Manage*, International Association for the Study of Pain (IASP), pp. 105-112, 2001.
[14] T. Mori, Y. Okazaki, T. Kawai, and T. Sato, “Objective estimation of pain based on facial expression, vital signs and body movements,” *Advanced Robotics*, VSP BV, Zeist, Netherlands, pp. 413-414, 2000.
[15] M. Pantic and L.J.M Rothkrantz, “Machine understanding of facial expressions of pain,” *Behavioral and Brain Sciences*, pp. 469-470, 2002.
[16] J. G. Daugman, “Two-dimensional spectral analysis of cortical receptive field profiles,” *Vision Research*, pp.847–856, 1980.
[17] L. Shen and L. Bai, “MutualBoost learning for selecting Gabor features for face recognition,” *Pattern Recognition Letters*, Elsevier, Amsterdam, Netherlands, pp.1758-1767, 2006.