Capturing global spatial patterns for distinguishing posed and spontaneous expressions

Shangfei Wang[a,*], Chongliang Wu[a], Qiang Ji[b]

[a]*Key Lab of Computing and Communication Software of Anhui Province*
School of Computer Science and Technology, University of Science and Technology of China
Hefei, Anhui, P.R.China, 230027
[b]*Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute*
Troy, NY, USA, 12180

Abstract

In this paper, we introduce methods to differentiate posed expressions from spontaneous ones by capturing global spatial patterns embedded in posed and spontaneous expressions, and by incorporating gender and expression categories as privileged information during spatial pattern modeling. Specifically, we construct multiple Restricted Boltzmann Machines (RBMs) with continuous visible units to model spatial patterns from facial geometric features given expression-related factors, i.e. gender and expression categories. During testing, only facial geometric features are provided, and the samples are classified into posed or spontaneous expressions according to the RBM with the largest likelihood. Furthermore, we propose efficient inference algorithm by extending annealing importance sampling to RBM with continuous visible units for calculating partition function of RBMs. Experimental results on benchmark databases demonstrate the effectiveness of the proposed approach in modelling global spatial patterns as well as its superior posed and spontaneous expression distinction performance over existing approaches.

Keywords: global spatial patterns, posed and spontaneous expression distinction, privileged information,

1. Introduction

Spontaneous expressions reveal one's real emotions, while posed expressions may disguise one's inner feelings. Automatically distinguishing between spontaneous and posed expressions can benefit many real life scenes. For example, service robots can make human-robot interaction more realistic by perceiving users' true feelings. Doctors can be more certain during diagnosis by knowing patients' genuine feelings. Detectives may detect a lie by differentiating posed expression from spontaneous ones.

Behavior research indicates that posed and spontaneous expressions are different from each other in both temporal and spatial patterns. Temporal patterns involve the speed, amplitude, trajectory and total duration of onset and offset. For example, Ekman *et al* [1, 2] revealed that the trajectory appears often smoother for spontaneous expressions than for posed ones, and the total duration is usually longer, and onset is more abrupt for posed expressions than spontaneous expressions in most cases. Spatial patterns mainly consists of the movement of facial muscles. Ekman *et al* [1] found that both zygomatic major and orbicularis oculi are contracted during spontaneous smiles, while only zygomatic major is contracted for posed smiles, as shown in Figure 1. Furthermore, the contraction of zygomatic major is more likely to occur asymmetrically for posed smiles than spontaneous ones [3]. **Recently, some works reveal contradictory findings in spatial patterns. For example, Krumhuber *et al* [4] questioned the differences of orbicularis oculi muscle movements between posed and spontaneous smile. Schmidt *et al* [5] suggested that asymmetry of facial movements may play a much smaller role in distinguishing posed and spontaneous smile. But they observed other differences between posed and spontaneous smile, such as smile intensity [4], amplitude, maximum speed, and duration [5]. Despite lack of a consensus on the differences between posed and spontaneous expression, we believe there indeed exist differences in spatial and temporal facial patterns between posed and spontaneous facial expressions as demonstrated by existing research. And, the goal of this research is to automatically capture the differences and to leverage them for distinguishing posed and spontaneous facial expressions.**

Inspired by the observations from nonverbal behavior research, researchers have begun to pay attention to posed and spontaneous expression distinction. The main com-

*This is the corresponding author (Telephone: +86-551-3602824).
Email addresses: sfwang@ustc.edu.cn (Shangfei Wang), clwzkd@mail.ustc.edu.cn (Chongliang Wu), qji@ecse.rpi.edu (Qiang Ji)

(a) Posed smile (b) Spontaneous smile

Figure 1: Posed and spontaneous smile, frames in (a) are from posed smile, and frames in (b) are from spontaneous smile. Both the zygomatic major (facial mouth area) and the orbicularis oculi (eyes area) are contracted during spontaneous smiles (the first frame of (b)), while only the zygomatic major is contracted for posed smiles (the first frame of (a)). The contraction of zygomatic major is more likely occur asymmetrically for posed smiles than spontaneous ones (second frame in (a) and (b))

ponents of posed and spontaneous expression distinction consists of feature extraction and classification. Although various features are proposed to describe temporal patterns embedded in spontaneous and posed expressions, and many classifiers are adopted, most studies only focus on one kind of expression, such as smile or pain, and little research explicitly models spatial patterns embedded in posed and spontaneous expression respectively. Furthermore, little research incorporates expression-related factors, such as gender, age, for posed and spontaneous expression classification. Thus, in this paper, we propose Restricted Boltzmann Machine (RBM) to explicitly capture the high-order spatial patterns embedded in posed and spontaneous expressions from facial geometric features, and incorporate gender and expression categories as privileged information during spatial pattern modeling. Specifically, we construct multiple RBMs with continuous visible units to model high order spatial patterns embedded in posed and spontaneous expressions given expression-related factors. During training, contrastive divergence (CD) [6] is adopted to learn the parameters of RBMs. During testing, only facial geometric features are provided, and the samples are classified into posed or spontaneous expressions according to the RBM with the largest likelihood. Furthermore, to calculate the partition function of RBMs, we extended Annealing Importance Sampling (AIS) [7] to RBM with continuous visible units case.

The rest of the paper is organized as follows: Section 2 presents an overview of the related works on posed and spontaneous distinction. The detailed introduction of our method is given in Section 3. Section 4 discusses the experimental results. Finally, the paper is concluded in Section 5.

2. Related Work

Current research of posed and spontaneous expression differentiation mainly consists of two steps: feature extraction and classification. For feature extraction, most research proposes features specially designed for differentiating posed expressions from spontaneous ones. For example, Cohn and Schmidt [8] proposed temporal features, i.e. duration, amplitude, and the ratio of amplitude to duration. Valstar [9] defined several mid-level feature, including intensity, speed, duration, symmetry, trajectory and the occurrence order of brow actions, from the displacements of facial fiducial points. Dibeklioglu et al. [10] extracted distance and angular features to discriminate the movements of eyelids. They [11] further extracted amplitude, duration, speed, and acceleration to describe dynamics of eyelid, cheek, and lip corner movements. Seckington [12] defined six features including morphology, apex overlap, symmetry, total duration, speed of onset and speed of offset, to represent temporal dynamics, which is essential for distinguishing between posed and spontaneous smiles. In addition to defining posed vs spontaneous expression specified features, some research adopts commonly used features for expression recognition. For example, Littlewort *et al* [13] fed the extracted Gabor wavelet features into SVM to recognize 20 facial action units as the middle-level features for posed and spontaneous pain classification. Pfister et al. [14] proposed a spatio-temporal local texture features, CLBP-TOP. Zhang et al. [15] used Scale-invariant feature transform (SIFT) appearance features and facial animation parameters (FAP) geometric features.

After feature extraction, classifiers should be trained. Cohn and Schmidt [8] adopted a linear discriminant classifier for posed and spontaneous smile recognition. Littlewort *et al* [13] employed SVM, Adaboost, and linear discriminant analysis to classify posed and spontaneous pain from recognized 20 facial action units. Valstar [9] adopted gentle Boost and relevance vector machines to distinguish posed vs. spontaneous brow actions. Dibeklioglu et al. [10] used continuous HMM, k-NN and naive Bayes classifiers to differentiate spontaneous smiles from posed ones. They [11] also employed individual SVM classifiers for different facial regions, and fuse them to classify genuine and posed smiles. Seckington [12] proposed to use dynamic Bayesian networks to model the temporal dynamics to distinguish between posed and spontaneous expressions. Zhang et al. [15] adopted minimal redundancy maximal relevance for feature selection, and support vector machine (SVM) as classifier for discrimination between posed and spontaneous versions of six basic emotions. Although various approaches have been developed for posed and spontaneous expression differentiation, there still exist several limitations. First, most computer vision works only focus on one specific expression, such as smile. To the best of our knowledge, only two works [15][14] considered all six basic expressions (i.e. happiness, disgust, fear, surprise, sadness and anger) for posed and spontaneous expressions recognition. Zhang *et al* [15] investigated the performance of a machine vision system for posed and spontaneous expressions recognition of six basic expression on USTC-NVIE

database. Pfister *et al* [14] proposed a generic facial expression recognition framework to differentiate posed from spontaneous expressions from both visible and infrared images on SPOS database.

Furthermore, most current works applied different classifiers for posed and spontaneous expression recognition, without capturing the spatial patterns embedded in posed and spontaneous expressions explicitly. We call them feature-driven method. Only recently, Wang *et al* [16] proposed multiple Bayesian networks (BN) to capture posed and spontaneous spatial facial patterns respectively given gender and expression categories. We call it a model-based method. Their recognition results on the USTC-NVIE and SPOS databases outperform those of the state of the art. However, due to the first-order Markov assumption of BN, their model can only capture the local dependencies among geometric features instead of the global and high-order relations among them. Furthermore, finding the optimal structure of a large geometric feature network for posed and spontaneous expression recognition is difficult. Compared with BN, restricted Boltzmann machine can model higher-order dependencies among random variables by introducing a layer of latent units [17]. It has been widely used to model complex joint distributions over structured variables such as image pixels. Thus, in this paper, we propose to use RBM to explicitly model complex joint distributions over feature points , i.e. spatial patterns, embedded in posed and spontaneous expressions respectively.

In addition, little work incorporates expression-related factors, such as gender, age and expression categories, for posed and spontaneous expression distinction, although researches indicate that different gender have different facial expression manifestation, face structures develop with ages, expression manifestation varies with ages, and different expressions usually evokes different spatial patterns [18, 19]. Recently, Dibeklioglu et al. [11] analyzed effect of age and gender on posed and spontaneous expression distinguishing by using age or gender as one feature. Wang *et al* [16] employed gender and expression categories as privileged information to help classify posed and spontaneous expressions. Compared with these two works, the former requires expression-related factors during both training and testing, while the later requires expression-related factors only for training. It means expression-related factors should be predicted during testing in the former. Such sequential approach may propagate the error of expression-related factor recognition to the subsequent expression recognition. Therefore, we prefer to incorporate expression-related factor as privileged information in this paper. Specifically, we construct multiple RBMs with continuous visible units to model spatial patterns in posed and spontaneous expressions given expression-related factors. During training, contrastive divergence (CD) [6] is adopted to learning the parameters of RBMs. During testing, the samples are classified into posed or spontaneous expressions according to the RBM with the largest likelihood. In addition, to solve the partition function of RBMs, we extended Annealing Importance Sampling (AIS) to RBM with continuous visible units.

Compared with related works, our contributions are as follows:

1. **We are the first to use** RBM to explicitly model the high-order spatial patterns embedded in posed and spontaneous expressions.

2. We further propose a partition function estimation method by extending AIS to RBM with continuous visible units.

3. We incorporate gender and expression category as privileged information into posed and spontaneous expression distinction.

3. Proposed method

The framework of our proposed method is shown in Fig.2, including feature extraction, spatial pattern modeling using RBM, and posed and spontaneous expression distinction. The details are described in the following subsections.

3.1. Feature extraction

In this paper, we extract the displacements of facial points between apex and onset expression frames as features. The apex frame is the frame with the most exaggerated expression during apex phase, and the onset frame is the first frame of the onset phase. The first step of feature extraction is to locate several facial points which shift obviously when a facial expression occurs and these points which will be used for geometric normalization. There are 29 feature points, the $1^{st} \dots 27^{th}$ facial point are showed in Figure 2, automatically detected on apex and onset frames using the algorithm introduced in [20]. We assigned the center of two eyes as the 28^{th} and 29^{th} point. Next step is to use face alignment and normalization to make the facial features robust to different subjects and different face pose variation. We rotate every facial image to make the inter-ocular line horizontal and with fixed length, and change the position of other facial points accordingly. In the meantime, **the facial region is cropped according to the width between center of two eyes and the height between eyes center with nose tip. Then, we resize facial region** to 100×100 by applying bicubic interpolation [21] and Anti-aliasing filter [22]. After that, we extract spatial movements of facial points by calculating the difference of facial points coordinates value between apex and onset frames. Since we located 27 facial points which moves significantly when expressions appear, a 54 dimensional feature vector is generated.

3.2. Spatial pattern modeling using RBM

A RBM for modeling spatial pattern embedded in posed or spontaneous expressions consists of two layers as shown in Figure. 3, one layer with n visible variables $\mathbf{v}$, representing the feature point displacements, and one layer with m

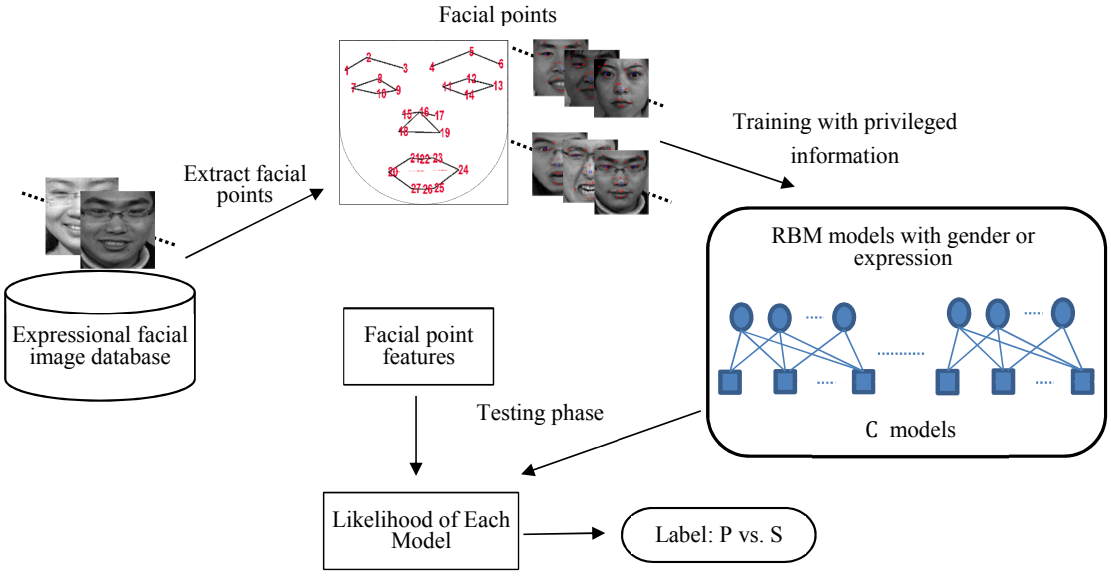


Figure 2: The framework of our proposed method, where "P" and "S" represent posed and spontaneous expression respectively.

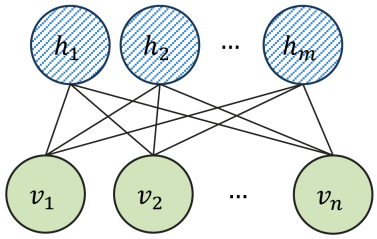


Figure 3: Restricted Boltzmann Machine

hidden variables $\mathbf{h} \in \{0,1\}^m$. Since feature point displacements are continuous, the visible units we used are continuous and have Gaussian marginal distributions. By introducing the latent layer, the RBM can model complex joint distributions over structured visible variables, i.e. feature point displacements. Thus, it can capture the spatial pattern embedded in posed or spontaneous expressions.

The total energy of RBM with continuous visible variables is defined in Eq. 1.

$$E(\mathbf{v},\mathbf{h};\theta) = -\sum_i \sum_j v_i W_{ij} h_j - \sum_j c_j h_j + \frac{1}{2}\sum_i (v_i - b_i)^2 \tag{1}$$

where, $\theta = \{\mathbf{W}, \mathbf{b}, \mathbf{c}\}$ are the parameters. W_{ij} are the weight of the connection between visible node v_i and hidden node h_i, which measures the compatibility between v_i and h_j. $\{b_i\}$ and $\{c_j\}$ are the biases of v_i and h_i respectively. The joint distribution over visible and hidden variables is described as Eq. 2

$$P(\mathbf{v},\mathbf{h}|\theta) = \frac{1}{Z(\theta)} \exp(-E(\mathbf{v},\mathbf{h};\theta)) \tag{2}$$

where $Z(\theta)$ is the partition function. The distribution over visible units of RBM is calculated by marginalizing over all hidden units with Eq. 2, as shown in Eq. 3. This allows RBM to capture global dependencies among the visible variables.

$$P(\mathbf{v}|\theta) = \sum_{\mathbf{h}} P(\mathbf{v},\mathbf{h}|\theta) = \frac{\sum_{\mathbf{h}} \exp(-E(\mathbf{v},\mathbf{h};\theta))}{Z(\theta)} \tag{3}$$

Given the training data $\{v_i\}_{i=1}^N$, where N indicates the number of training samples, the goal of RBM training is to maximize the joint distribution over visible units, as follows:

$$\theta^* = argmax_\theta L(\theta) = argmax_\theta \frac{1}{N}\sum_{i=1}^{N} \log P(\mathbf{v}|\theta) \tag{4}$$

The gradient with respect to θ can be calculated as Eq.5

$$\frac{\partial \log P(\mathbf{v}|\theta)}{\partial \theta} = \left\langle \frac{\partial E}{\partial \theta} \right\rangle_{p(\mathbf{h}|\mathbf{v},\theta)} - \left\langle \frac{\partial E}{\partial \theta} \right\rangle_{p(\mathbf{h},\mathbf{v}|\theta)} \tag{5}$$

where $\langle\cdot\rangle_p$ represents the expectation over distribution p. Calculating the gradient involves inferring $P(\mathbf{h},\mathbf{v}|\theta)$, which is intractable. However there is an efficient way to estimate its approximation called contrastive divergence (CD) [6]. The basic idea is to approximate $P(\mathbf{h},\mathbf{v}|\theta)$ with an one step sampling from the data. In the case of continuous visible nodes, during sampling, the probability distributions of $P(\mathbf{v}|\mathbf{h},\theta)$ and $P(\mathbf{h}|\mathbf{v},\theta)$ can be calculated as:

$$\begin{aligned} P(\mathbf{v}|\mathbf{h},\theta) &= \prod_i N(b_i + \sum_j w_{ij} h_j, 1); \\ P(\mathbf{h}|\mathbf{v},\theta) &= \prod_j \delta(\sum_i v_i w_{ij} + c_j) \end{aligned} \tag{6}$$

In this work, expression categories and gender are used as privileged information, which is only available during

training. It means we construct one RBM for each expression or gender. For example, if we consider gender as privileged information, 2×2 RBMs ($\theta_l, l = 1, ..., 2 \times 2$) are trained for modeling spatial patterns embedded in male posed expression, female posed expression, male spontaneous expression and female spontaneous expression respectively. Similarly, 2×6 RBMs ($\theta_l, l = 1, ..., 2 \times 6$) are trained, when expression categories are used as privileged information.

3.3. Posed and spontaneous expression distinction

After training, we obtain multiple RBM given expression related factors. During testing, only geometric features are provided, without expression related factors. For a test sample t, the log likelihood that RBM trained on class l assign to t is as follows:

$$\log P(t|\theta_l) = \log\left(\sum_h exp(-E(h, t; \theta_l))\right) - \log Z(\theta_l) \quad (7)$$

Then, the label of the test sample is the class with greatest log likelihood value according to Eq. 8:

$$l^\star = \max_{l \in [1,C]} \{\log P(t|\theta_l)\} \quad (8)$$

where $l^\star$ represents the predicted label, C is the number of RBMs.

It is intractable to compute the partition function $Z(\theta)$ of RBM directly. Salakhutdinov and Murry [7] proposed to use Annealed Importance Sampling (AIS) [23] to estimate the partition function of a RBM with discrete visible units. In this work, we extend the AIS method to calculate the partition function of a RBM with continuous visible units. The algorithm of distinguishing posed and spontaneous expression is shown in Algorithm 1.

Algorithm 1 Posed and Spontaneous expression recognition

Input: Training samples: S_{tr},
Label of training samples: L_{tr},
Gender (expression) categories of S_{tr}: L_g (L_e),
Test samples: S_{te}.
Output: Label of test samples: L_{te}.
Training phase
Divide S_{tr} into C classes, according to L_{tr} and L_g (L_e)
for $l = 1 : C$ **do**
Train a RBM (θ_c) for $\{t|t \in S_{tr} \cap t \in c\}$ using CD algorithm with Eq. 5 and Eq. 6;
end for
Testing phase
for $l = 1 : C$ **do**
Estimate $Z(\theta_c)$ using the Algorithm 2;
Estimate $\log P(t|\theta_c)$ ($t \in S_{te}$) with θ_c and Eq. 7;
end for
Predict L_{te} for S_{te} using Eq. 8.

AIS estimates the ratio of partition function of the object RBM to a "base-rate" RBM. In order to evaluate the partition function of a RBM, we suppose p_0 and p_K are two probability distributions over V which represents the visible units of two RBMs (i.e. "base-rate" RBM and the object RBM). Here, elements in V comply with Gaussian marginal distribution. Parameters of two RBMs are represented as $\theta_0 = \{W^0, b^0, c^0\}$ and $\theta_K = \{W^K, b^K, c^K\}$. RBMs can have different number of hidden units $\mathbf{h}^0 \in \{0,1\}^{m_0}$ and $\mathbf{h}^K \in \{0,1\}^{m_K}$.

First, a sequence of intermediate distributions for $k = 0, \ldots, K$ are defined as:

$$p_k(\mathbf{v}) = \frac{p_k^*(\mathbf{v})}{Z_k} = \frac{1}{Z_k}\sum_h \exp(-E_k(\mathbf{v}, \mathbf{h})) \quad (9)$$

where the energy function is given by:

$$E_k(\mathbf{v}, \mathbf{h}) = (1 - \beta_k)E(\mathbf{v}, \mathbf{h}^0; \theta_0) + \beta_k E(\mathbf{v}, \mathbf{h}^K, \theta_K) \quad (10)$$

where $0 = \beta_0 < \beta_1 < \ldots < \beta_K = 1$. Different with [7], here, $E(\mathbf{v}, \mathbf{h}; \theta)$ is from Eq. 1, in which the visible units are continuous.

Then, we define a Markov chain transition operator $T_k(v'; v)$ that leaves $p_k(\mathbf{v})$ invariant. With Eq.9 and Eq.10 to derive a block Gibbs sampler, the conditional distribution over RBM's hidden or visible units can be defined as follow:

$$p(h_j^0 = 1|\mathbf{v}) = \delta\left((1 - \beta_k)(\sum_i W_{ij}^0 v_i + c_j^0)\right) \quad (11)$$

$$p(h_j^K = 1|\mathbf{v}) = \delta\left(\beta_k(\sum_i W_{ij}^K v_i + c_j^K)\right) \quad (12)$$

$$\begin{aligned} p(v_i'|\mathbf{h}) = N\Big(&(1 - \beta_k)(\sum_j W_{ij}^0 h_j^0 + b_i^0) \\ &+ \beta_k(\sum_j W_{ij}^K h_j^K + b_i^K),\ 1\Big) \end{aligned} \quad (13)$$

where Eq. 13 is a Gaussian marginal distribution. Given a sample $\mathbf{v}$, Eq. 11 and Eq. 12 are used to draw samples of hidden units within two RBMs. Then, with Eq. 13, we can draw a new sample v'. The unnormalized probability over visible units can be estimated as:

$$\begin{aligned} p_k^*(\mathbf{v}) &= \sum_{\mathbf{h}^0, \mathbf{h}^K} e^{(1-\beta_k)E(\mathbf{v}, \mathbf{h}^0; \theta_0) + \beta_k E(\mathbf{v}, \mathbf{h}^K; \theta_K)} \\ &= e^{-\frac{1-\beta_k}{2}\sum_i (v_i - b_i)^2} \cdot \prod_{j=1}^{m_0}(1 + e^{(1-\beta_k)(\sum_i W_{ij}^0 v_i + c_j^0)}) \\ &\cdot e^{-\frac{\beta_k}{2}\sum_i (v_i - b_i)^2} \cdot \prod_{j=1}^{m_K}(1 + e^{\beta_k(\sum_i W_{ij}^K v_i + c_j^K)}) \end{aligned} \quad (14)$$

With equations 11, 12, 13, and 14, we can perform AIS starting by running a blocked Gibbs sampler (Eq. 6) to

Algorithm 2 Annealed Importance Sampling

Input: Parameters of object RBM: θ_K
Parameters of "base-rate" RBM: θ_0
β_k, $k = 0, \ldots, K$
Output: Partition function of object RBM: Z_K.
for $i = 1 : M$ **do**
Sample $\mathbf{v}_1$ from p_0, by running a blocked Gibbs sampler with Eq. 6;
for $k = 2 : K$ **do**
Sample $\mathbf{v}_k$ given $\mathbf{v}_{k-1}$ using T_{k-1} with Eq. 11, 12, and 13;
end for
Compute the importance weight w^i with Eq. 14:

$$w^i = \frac{p_1^*(\mathbf{v}_1)}{p_0^*(\mathbf{v}_0)} \frac{p_2^*(\mathbf{v}_2)}{p_1^*(\mathbf{v}_1)} \cdots \frac{p_K^*(\mathbf{v}_K)}{p_{K-1}^*(\mathbf{v}_{K-1})}; \quad (15)$$

end for
The ratio of partition function:

$$\frac{Z_K}{Z_0} \approx \frac{1}{M} \sum_{i=1}^{M} \frac{p_K^*(\mathbf{v}^i)}{p_0^*(\mathbf{v}^i)} = \frac{1}{M} \sum_{i=1}^{M} w^i = \hat{r}_{IS}; \quad (16)$$

Compute Z_K with Eq. 16 and Eq. 17

generate samples from p_0. We gradually change β_k from 0 to 1. The procedure of AIS algorithm displayed in Algorithm 2.

The partition function of the object RBM (Z_K) can be estimated by finding the ratio to the normalizer for p_0 with $\theta_0 = \{0, b^0, 0\}$ where the weight matrix is zero. The partition function Z_0 is computed as follow:

$$\begin{aligned} Z_0 &= \sum_{\mathbf{v}} \sum_{\mathbf{h}} \exp(E_0(\mathbf{v}, \mathbf{h})) \\ &= (\sqrt{2\pi})^n \cdot 2^{(m_0)} \end{aligned} \quad (17)$$

In this case, we can draw exactly independent samples from p_0, since the weights between visible and hidden nodes are zero. By annealing from this simple model to the final model, we can estimate the partition function through AIS.

4. Experiments and Analysis

4.1. Experimental conditions

For experiments, we evaluate our methods on the SPOS [14], the USTC-NVIE [24], and the MMI [25] databases. The SPOS database and the USTC-NVIE database contain posed and spontaneous expression for six basic expression categories (i.e. happiness, disgust, fear, surprise, anger and sadness). The MMI database contains happiness and disgust spontaneous expressions and six basic expression categories for posed expressions.

The USTC-NVIE database is a natural visible and thermal infrared facial expression database. The onset and apex frames are provided for both posed and spontaneous subsets. Both apex and onset frames from all posed and spontaneous expression samples, which come in pairs from the same subject are selected. During this procedure, we discarded spontaneous samples whose maximum evaluation value on six expression categories are zero, since these samples have no expression. Finally 1028 samples, including 514 posed and 514 spontaneous expression samples from 55 male and 25 female subjects, are selected. Our experimental results on the database are obtained by applying a 10-fold cross validation to all samples according to the subjects. Given the databases, facial feature point displacements between apex and onset frames are used for modeling spatial patterns embedded in posed and spontaneous expressions from three aspects: all samples without gender and expression information, with gender information and with expression information, respectively. We first build 2 RBM models, which denote as "PS model", using posed and spontaneous samples without the gender and expression labels respectively. Then, 4 RBM models are built, which denote as "PS_gender model", using male posed, male spontaneous, female posed, and female spontaneous samples respectively. Last, we build 12 RBM models, which denote as "PS_exp model", using posed and spontaneous samples for each expression respectively.

The SPOS database is a visible and near infrared expression database. The image sequences in this database start from onset frame and end with apex frame. Therefore, the first and last frames of all posed and spontaneous samples are selected, including 84 posed and 150 spontaneous expression samples. Since only seven subjects (4 males and 3 females) are in SPOS database and it does not include all six expression images for a certain subject, we can not select samples in pairs as we did on USTC-NVIE database. In order to compare with [14], leave-one-subject-out cross validation is used during our experiments.

For the MMI database, according to the description in [25], spontaneous expression contains two subsets. The first subset of spontaneous expression, part IV described in [25], includes 383 manually segmented sequences that contain happiness and disgust expressions. The second subset of spontaneous expression, part V, contains nine unsegmented visual and audio recordings. We selected 318 segments which are onset-apex-offset segments from the first subset of spontaneous expression. For posed expression, we selected sessions with expression labels from part I, II, and III which consist of posed expressions as described in [25]. Since there are only happiness and disgust expression categories for spontaneous expressions, only happiness and disgust expression categories from the selected posed segments are used in our experiments. Finally, we obtained 64 posed segments (35 happiness and 29 disgust)

from 24 subjects and 318 spontaneous segments (258 happiness and 60 disgust) from 15 subjects, on MMI database. We manually extract onset and apex frames from the selected segments. After that, following the same procedures as USTC-NVIE database, we extract facial feature point displacements between apex and onset frames as input to RBM, and build RBMs from three aspects: all samples without gender and expression information (i.e. "PS model"), with gender information (i.e. "PS_gender model") and with expression information (i.e. "PS_exp model"), respectively. 10-fold cross validation to all samples according to the subjects were then performed in the experiments on the MMI database.

4.2. Experimental results of posed and spontaneous expression recognition

Experimental results on the USTC-NVIE database are shown in Table 1. Comparing the results of PS model with those of the remaining models, we can find that using gender and expression information during training can **help model the variation of muscle** in posed and spontaneous expression, since both accuracy and F1 score of P-S_gender and PS_exp are higher than that of PS model. All experimental accuracies are greater than 90%, which demonstrated that our proposed method can capture the spatial patterns effectively by using multiple RBM models.

Table 1: P vs. S recognition results on USTC-NVIE database

Model	PS		PS_gender		PS_exp	
	P	S	P	S	P	S
P	482	32	467	47	478	36
S	53	461	29	485	48	466
Accuracy(%)	91.73		92.61		91.83	
F1-score	0.9190		0.9248		0.9192	

"P" represents posed expression.
"S" represents spontaneous expression.

Table 2: P vs. S recognition results on SPOS database

Model	without any information	
	P	S
P	51	33
S	23	127
Accuracy(%)	76.07	
F1-score	0.6456	

"P" represents posed expression.
"S" represents spontaneous expression.

Experimental results on the MMI database are shown in Table 3. Our model achieved 89.01% by using PS model alone. By using gender and expression as privileged information, recognition accuracy reached 89.27% and 89.79% respectively, which are higher than that using PS model. F1 scores of using privileged information are also higher than that of PS model. The results of experiments on MMI database once again demonstrated the effectiveness of RBM models capturing facial spatial patterns and the effectiveness of using gender and expression as privileged information to help the modeling task.

Table 3: **P vs. S recognition results on MMI database**

Model	PS		PS_gender		PS_exp	
	P	S	P	S	P	S
P	33	31	36	28	33	31
S	11	307	13	305	8	310
Accuracy(%)	89.01		89.27		89.79	
F1-score	0.6111		0.6372		0.6286	

"P" represents posed expression.
"S" represents spontaneous expression.

Experimental results on the SPOS database are shown in Table 2. The accuracy and F1-score are achieved 76.07% and 0.6456, respectively. **The results are acceptable, but not as good as those on the USTC-NIVE database and the MMI database. Since the number of samples from USTC-NVIE database and the MMI database vastly exceed that from SPOS database and RBM models require much data to train, it is reasonable that the accuracy rate and F1-score obtained on SPOS Database are a little lower than those on the NVIE database and the MMI database.**

To show the differences of spatial patterns between posed and spontaneous expressions captured by RBM intuitively, **we analyzed the weights of the RBMs trained on the USTC-NVIE, SPOS and MMI databases.** In Figure 4, the global spatial pattern captured by both RBMs are demonstrated. As described in Section 3.2, parameters W_{ij} measures the compatibility between visible node v_i and latent node h_j. The greater the absolute value of W_{ij}, the more the point displacement affect the captured spatial pattern. Figure 4 (a) and (b) are the weights of two different hidden nodes of "PS_happiness" model built on USTC-NVIE database. **Figure 4 (d) and (e) are of "PS_happiness" model built on MMI database.** Figure 4 (a) and **(d)** show the pattern that the displacement of points around mouth area are more likely to occur asymmetrically for posed happiness than spontaneous ones. Figure 4 (b) and **(e)** show the pattern that the displacement of points around eyes area are greater for spontaneous happiness than posed ones. Figure 4 (a), **(d)** and (b), **(e)** conform to the statements in section 1 that the contraction of zygomatic major (mouth area) is more likely to occur asymmetrically for posed smiles than spontaneous ones and that the orbicularis oculi muscle (eyes area) is contracted only during spontaneous smiles. **These empirical findings are**

consistent with those findings in [1] and [3].

We sum W over all hidden units in "PS" model trained on the SPOS database **and the MMI database**, and demonstrate them in Figure. 4 (c) **and (f)**, respectively. We can find that the W values of RBM for posed expressions and those for spontaneous expressions are very different, proving that the spatial pattern for posed expressions and that for spontaneous expressions are different. This is consistent with behavior researches. Besides, in most cases, the weights of posed RBM are greater than those of spontaneous RBM. It further confirms that posed expressions are more exaggerated than spontaneous ones.

4.3. Comparison with other methods

We compare our work with related works for distinguishing posed vs. spontaneous expressions using feature driven methods [15, 14, 26] and model based methods capturing local facial spatial patterns through BN [16]. In addition, we conduct posed and spontaneous expression distinction experiment using a linear kernel support vector machine (SVM) with the same features and the same experimental conditions with ours as a baseline.

In order to better compare with other methods, we conducted our experiments five times on each database. Mean and variance of accuracy of experiments on each database are listed in Table 4-6. Since we do not have code for these methods, we cannot perform a statistical test to evaluate the statistical significance of our method.

Table 4 shows the comparison results on the USTC-NVIE database. Zhang *et al* recognized posed and spontaneous expression for six basic expressions on the USTC-NIVE database by extracting both geometric and appearance features. They removed the images whose face or facial point cannot be detected correctly, and finally selected 3572 posed and 1472 spontaneous images. Since they did not explicitly state which images were selected, we cannot select the same images as theirs. Therefore, we compare the experimental results as a reference. Although Zhang *et al* extracted more complex features and applied more samples for training, we achieve a better result. This demonstrates that the trained RBMs capture spatial patterns for posed and spontaneous expressions successfully. The experimental conditions of our work are close to that of Wang *et al*' work [16]. From Table 4, we can find that our approach slightly outperforms theirs for "PS model" and "PS_gender model", demonstrating the superiority of high-order dependence modeled by RBM. For "PS_exp model" , Wang *et al*' is slightly better than ours. It may be due to the small size of training set. Under the same experimental conditions, our approach significantly outperforms the baseline, i.e. SVM. It further demonstrates the effectiveness of the proposed method.

Pfister *et al* [14] distinguished posed and spontaneous expression from both visible and near-infrared images on the SPOS database. We only compare our work with their work on visible images. Wang *et al* [16] conducted experiments using visible images on the SPOS database. The comparisons are shown in Table 5. From Table 5, we can find that our approach outperforms Pfister *et al*'s. **Although the average recognition accuracy of our method is slightly lower than Wang *et al*'s, the best recognition accuracy, As shown in Table 2, of our method is 76.06% which is better than Wang *et al*'s.** Besides, the features used in Pfister *et al*'s, Wang *et al*'s and ours are texture features, action unit related geometric features, and feature point displacement respectively. It means we use simpler features, but achieve better results. In addition, our method outperforms the baseline significantly.

Table 4: Comparison with other methods on USTC-NVIE database

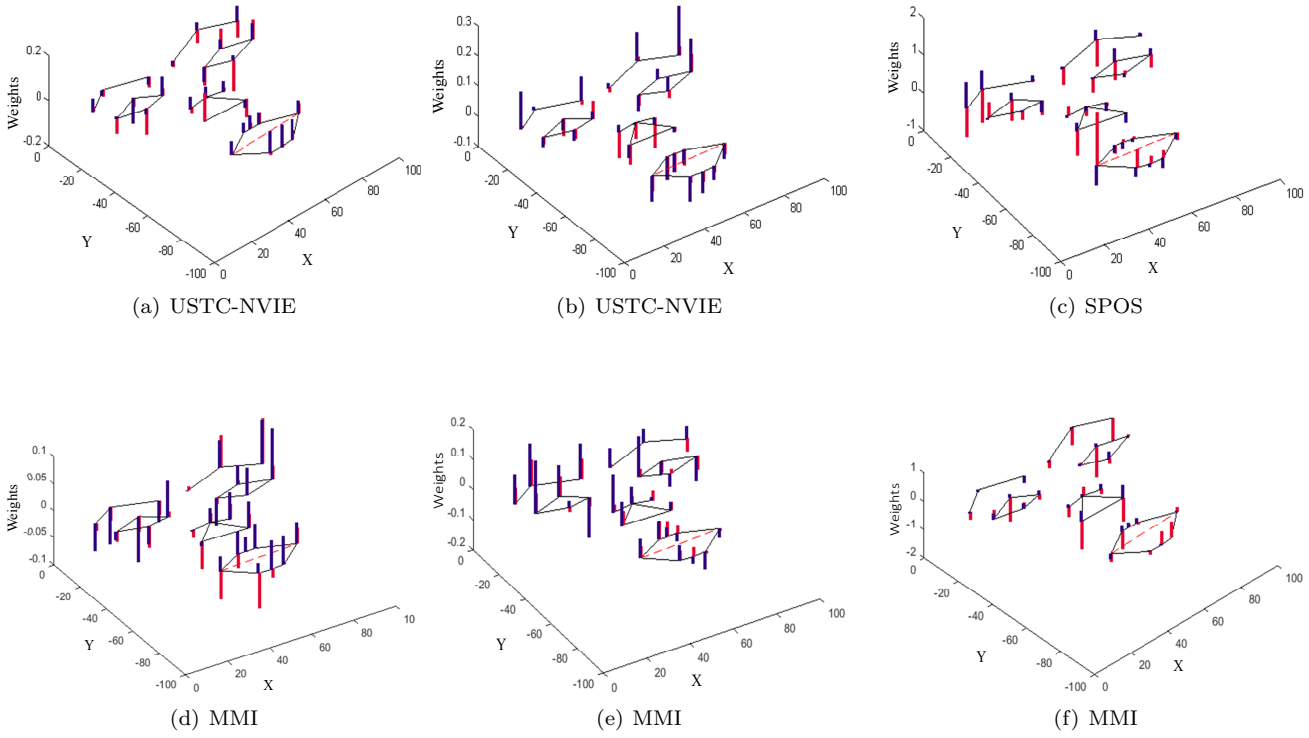
Comparison with methods based on capturing spatial patterns			
Accuracy(%)	PS	PS_gender	PS_exp
Our method (Mean/Variance)	91.71/ 1.73e-5	92.24/ 7.43e-6	91.46/ 6.36e-6
S. Wang *et al* [16]	91.63	92.22	92.90
Comparison with feature driven methods			
Accuracy(%)	PS	PS_gender	PS_exp
SVM	81.52	/	/
L. Zhang *et al* [15]	79.43	/	/

Table 5: Comparison with other methods on SPOS database

Comparison with methods based on capturing spatial patterns		
Method	Ours (Mean/Variance)	Wang *et al* [16]
Accuracy (%)	74.10/1.33e-4	74.79
Comparison with feature driven methods		
Method	T. Pfister *et al* [14]	SVM
Accuracy (%)	72.0	63.25

Table 6 shows the comparison results on the MMI database. Dibeklioglu *et al* [26] extracted features to describe the dynamics of eyelid, cheek, and lip corner movements, and fused them over different regions and over different temporal phases for posed and spontaneous smile recognition. They selected 74 posed smiles from 30 subjects and 120 spontaneous smiles from 15 subjects. Since we do not know which samples they used in MMI database for posed and spontaneous smile recognition, we can not compare our works with theirs under exactly same experimental conditions. We can only compare with their published results. They also investigate gender effects on their system. Different from using gender as privileged information, they applied gender information during both training and testing phase. We can find that although with more unbalancing data and more simple features, our method still outperforms Dibeklioglu's. It further demonstrates the effectiveness of the proposed method.

From the above comparison, we can find that our method significantly outperforms current feature-driven



(a) USTC-NVIE (b) USTC-NVIE (c) SPOS

(d) MMI (e) MMI (f) MMI

Figure 4: Weights at every facial points from the trained RBMs, (a) and (b) are from two different hidden nodes of "PS_happiness" model trained on USTC-NVIE database, and (c) is from the sum of all hidden nodes of RBMs trained on SPOS database. **(d) and (e) are from two different hidden node of "PS_happiness" model trained on MMI database, and (f) is from the sum of all hidden nodes of "PS" model.** We drew a facial points distribution map same as that in Figure 2 at $x-y$ plane. z coordinate is the weight values, i.e. W. The red bars represent weight values from RBM for posed expressions, and the blue bars represent weight values from RBM for spontaneous expressions

Table 6: Comparison with other methods on MMI database

Comparison with methods based on capturing spatial patterns			
Accuracy(%)	PS	PS_gender	PS_exp
Our method (Mean/Variance)	88.33 /1.59e-5	88.64 /1.58e-5	89.63 /3.58e-5
Comparison with feature driven methods			
Accuracy(%)	PS	PS_gender	PS_exp
SVM	84.55	/	/
H. Dibeklioglu *et al* [26]	88.14	87.63	/

methods on both databases, and is superior to current model-based methods in most cases. It demonstrates that our proposed RBM spatial models successfully capture the global spatial patterns, which are crucial for distinguishing posed and spontaneous expressions.

5. Conclusions

In this paper, we propose to use RBM to explicitly model complex joint distributions over feature points, i.e. spatial patterns, embedded in posed and spontaneous expressions respectively, and incorporate expression-related factor as privileged information, which is only available during training. Specifically, we construct multiple RBMs with continuous visible units to model spatial patterns embedded in posed and spontaneous expressions given expression-related factors. During training, contrastive divergence is adopted to learn the parameters of RBMs. During testing, the samples are classified into posed or spontaneous expressions according to the RBM with the largest likelihood. In addition, to solve the partition function of RBMs, we extended annealing importance sampling to RBM with continuous visible units case. Experimental results on three benchmark databases demonstrate the power of the proposed model in capturing spatial patterns as well as its advantage over existing methodologies for posed and spontaneous expression distinction.

The proposed model has two major advantages over existing methods. 1) Unlike methods that can only capture local spastical patterns, our model is developed upon the restricted Boltzmann machine, and therefore can exploit the global relations among geometric features. 2) Although expression-related factors can influence patterns embedded in posed and spontaneous expressions, these factors are generally ignored by the current methods. Our approach, however, can successfully capture them to help more accurately characterize facial spatial patterns.

6. Acknowledgments

This work has been supported by the National Science Foundation of China (Grant No. 61175037, 61228304, 61473270), and the project from Anhui Science and Technology Agency (1106c0805008).

[1] P. Ekman, W. Friesen, Felt, false, and miserable smiles, Journal of nonverbal behavior 6 (4) (1982) 238–252.

[2] P. Ekman, Darwin, deception, and facial expression, Annals of the New York Academy of Sciences 1000 (1) (2003) 205–221.

[3] P. Ekman, J. Hager, W. F., The symmetry of emotional and deliberate facial actions, Psychophysiology 18 (2) (1981) 101–106.

[4] E. G. Krumhuber, A. S. Manstead, Can duchenne smiles be feigned? new evidence on felt and false smiles., Emotion 9 (6) (2009) 807.

[5] K. L. Schmidt, S. Bhattacharya, R. Denlinger, Comparison of deliberate and spontaneous facial movement in smiles and eyebrow raises, Journal of nonverbal behavior 33 (1) (2009) 35–45.

[6] G. E. Hinton, Training products of experts by minimizing contrastive divergence, Neural computation 14 (8) (2002) 1771–1800.

[7] R. Salakhutdinov, I. Murray, On the quantitative analysis of deep belief networks, in: ICML, ACM, 2008, pp. 872–879.

[8] J. Cohn, K. Schmidt, The timing of facial motion in posed and spontaneous smiles, International Journal of Wavelets, Multiresolution and Information Processing 2 (02) (2004) 121–132.

[9] M. Valstar, M. Pantic, Z. Ambadar, J. Cohn, Spontaneous vs. posed facial behavior: automatic analysis of brow actions, in: Proceedings of the 8th international conference on Multimodal interfaces, ACM, 2006, pp. 162–170.

[10] H. Dibeklioglu, R. Valenti, A. A. Salah, T. Gevers, Eyes do not lie: spontaneous versus posed smiles, in: Proceedings of the international conference on Multimedia, ACM, 2010, pp. 703–706.

[11] H. Dibeklioglu, A. Salah, T. Gevers, Recognition of genuine smiles, Multimedia, IEEE Transactions on PP (99) (2015) 1–1. doi:10.1109/TMM.2015.2394777.

[12] M. Seckington, Using dynamic bayesian networks for posed versus spontaneous facial expression recognition, Mater Thesis, Department of Computer Science, Delft University of Technology.

[13] G. Littlewort, M. Bartlett, K. Lee, Automatic coding of facial expressions displayed during posed and genuine pain, Image and Vision Computing 27 (12) (2009) 1797–1803.

[14] T. Pfister, X. Li, G. Zhao, M. Pietikainen, Differentiating spontaneous from posed facial expressions within a generic facial expression recognition framework, in: ICCV Workshops, IEEE, 2011, pp. 868–875.

[15] L. Zhang, D. Tjondronegoro, V. Chandran, Geometry vs. appearance for discriminating between posed and spontaneous emotions, in: Neural Information Processing, Springer, 2011, pp. 431–440.

[16] S. Wang, C. Wu, M. He, J. Wang, Q. Ji, Posed and spontaneous expression recognition through modeling their spatial patterns, Machine Vision and Applications (2015) 1–13.

[17] S. Nie, Q. Ji, Capturing global and local dynamics for human action recognition, in: ICPR, 2014.

[18] C. Lithari, C. Frantzidis, C. Papadelis, A. Vivas, M. Klados, C. Kourtidou-Papadeli, C. Pappas, A. Ioannides, P. Bamidis, Are females more responsive to emotional stimuli? a neurophysiological study across arousal and valence dimensions, Brain topography 23 (1) (2010) 27–40.

[19] S. Yunus, T. Christopher, Cascaded classification of gender and facial expression using active appearance models, Automatic Face and Gesture Recognition, IEEE International Conference on 0 (2006) 393–400.

[20] Robust facial feature tracking under varying face pose and facial expression, Pattern Recognition 40 (11) (2007) 3195 – 3208.

[21] R. Keys, Cubic convolution interpolation for digital image processing, Acoustics, Speech and Signal Processing, IEEE Transactions on 29 (6) (1981) 1153–1160.

[22] D. P. Mitchell, A. N. Netravali, Reconstruction filters in computer-graphics, in: ACM Siggraph Computer Graphics, Vol. 22, ACM, 1988, pp. 221–228.

[23] R. M. Neal, Annealed importance sampling, Statistics and Computing 11 (2) (2001) 125–139.

[24] S. Wang, Z. Liu, S. Lv, Y. Lv, G. Wu, P. Peng, F. Chen, X. Wang, A natural visible and infrared facial expression database for expression recognition and emotion inference, Multimedia, IEEE Transactions on 12 (7) (2010) 682–691.

[25] M. Valstar, M. Pantic, Induced disgust, happiness and surprise: an addition to the mmi facial expression database, in: Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC): Corpora for Research on Emotion and Affect, 2010, p. 65.

[26] H. Dibeklioglu, A. A. Salah, T. Gevers, Recognition of genuine smiles, Multimedia, IEEE Transactions on 17 (3) (2015) 279–294.