Posed and Spontaneous Expression Recognition Through Restricted Boltzmann Machine

Chongliang Wu and Shangfei Wang(✉)

School of Computer Science and Technology,
University of Science and Technology of China, Hefei, China
clwzkd@mail.ustc.edu.cn, sfwang@ustc.edu.cn

Abstract. This paper presents a new method to recognize posed and spontaneous expression through modeling their global spatial patterns in Restricted Boltzmann Machine (RBM). First, the displacements of facial feature points between apex and onset facial images are extracted as features, which capture spatial variations of facial points. Second, the point displacement related facial events are extracted from its displacements. Third, two RBM models are trained to capture spatial patterns embedded in posed and spontaneous expressions respectively. The recognition results on both USTC-NVIE and SPOS databases demonstrate the effectiveness of the proposed RBM approach in modeling complex spatial patterns embodied in posed and spontaneous expressions, and good performance on posed and spontaneous expression distinction.

Keywords: Posed and spontaneous · Expression recognition · Restricted Boltzmann Machine · Spatial pattern

1 Introduction

As we all know, spontaneous expressions convey one's true feelings, while posed expressions disguise one's real emotions. A method to distinguish posed expressions from spontaneous expressions can be used in many areas, including real-life human-robot communications, healthcare, and security. For example, robots can be more perceptual by analyzing users' true emotions through posed and spontaneous expressions differentiating technology. Doctors can make a more precise diagnosis by detecting patients' genuine pain. Deceptive facial expression recognition can be used by the police for lie detection.

Many nonverbal behavior researches presented the differences between posed and spontaneous expressions in both spatial and temporal patterns [4–7]. Spatial patterns mainly involve the movements of facial muscles. For example, when one smiles spontaneously, both the zygomatic major and the orbicularis oculi should be contracted. However, if a smile is posed, the contraction will only appear on zygomatic major, but not on orbicularis oculi [5]. The contraction of zygomatic major is more likely to occur asymmetrically in posed smiles than in spontaneous ones [6]. Ekman *et al.* [5,7] claimed that a good way to recognize

Q. Tian et al. (Eds.): MMM 2016, Part I, LNCS 9516, pp. 127–137, 2016.
DOI: 10.1007/978-3-319-27671-7_11

a posed expression from a spontaneous one is to analyze the absence of muscles movements, since some movements is difficult to make voluntarily [5,7]. Temporal patterns include the trajectory, speed, amplitude and total duration of onset and offset. Such as, for posed expressions, the total duration is usually longer and the onset is more abrupt than spontaneous expressions in most cases [4,5]. For spontaneous expressions, the trajectory appears often smoother than posed expressions [5].

Motivated by the properties revealed by behavior researches, researchers in computer vision have begun to pay attention to posed and spontaneous expressions recognition. The first research on posed and spontaneous expression recognition using machine learning method is presented by Cohn and Schmidt [1], which extracted temporal features, i.e. amplitude, duration, and the ratio of amplitude to duration, and applied a linear discriminant as classifier for posed and spontaneous smile recognition. Valstar [18] proposed a posed and spontaneous smile recognition method. They studied posed and spontaneous brow actions using velocity, duration, and the order of occurrence and fused head, face, and shoulder modalities for the recognition. Littlewort *et al.* [10] proposed a real and faked pain expression classification method by feeding the detected 20 facial action units into a classifier. Dibeklioglu *et al.* [2] used the dynamics of eyelid, check and lip corner movement to distinguish posed and spontaneous smile.

However, most computer vision researches only focus on one specific expression. To the best of our knowledge, only two works [15,20] considered all six basic expressions (i.e. happiness, disgust, fear, surprise, sadness and anger) for posed and spontaneous expressions recognition. Zhang *et al.* [20] used SIFT [3] and FAP [12] features to investigate the performance of a machine vision system for posed and spontaneous expressions recognition of six basic expression on USTC-NVIE database. Pfister *et al.* [15] proposed a spatiotemporal local texture descriptor (CLBP-TOP) and 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. Few works captured the spatial patterns embedded in posed and spontaneous expressions explicitly. Thus, we proposed to use Restricted Boltzmann Machine (RBM) to explicitly model spatial patterns embedded in both posed and spontaneous expressions respectively. As a graphical model, RBM can model higher-order dependencies among random variables by introducing a layer of latent units [11]. It has been widely used to model complex joint distributions over structured variables such as image pixels.

In this paper, we first extract the facial point displacements between the apex and the onset facial images. Second, these displacements are discretized to extract facial point displacement related facial events which are used as inputs to RBMs. Third, two RBM models are trained from posed and spontaneous expression samples to capture the spatial patterns embedded in posed and spontaneous expressions explicitly. During testing, the label of an unknown sample

is given by selecting the RBM model that is more likely to have generated its set of displacements. The recognition results on both USTC-NVIE and SPOS databases demonstrate the effectiveness of the proposed RBM approach in modeling complex spatial patterns embodied in posed and spontaneous expressions.

2 The Proposed Method

The procedure of our proposed approach is shown in Fig. 1, includes features extraction, and posed and spontaneous expression modeling by RBMs. The details are described as follows.

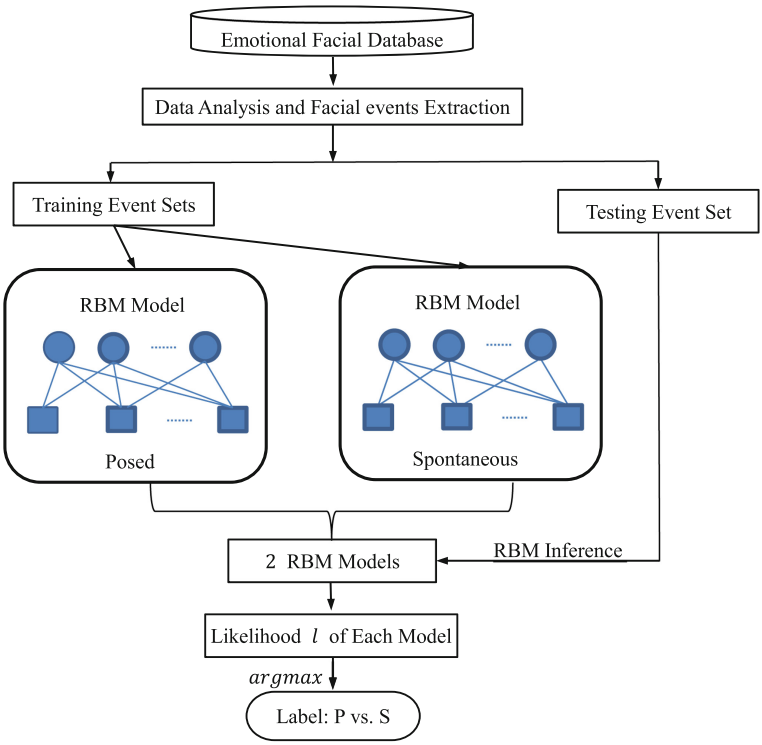

Fig. 1. The framework of our proposed method

2.1 Feature and Facial Events Extraction

First, 27 facial feature points, as shown in the bottom part of Fig. 2, are automatically detected on both the onset and apex expression frames using the algorithm introduced in [17]. The onset frame is the beginning of the onset phase, which is similar to the neutral frame here, and the apex frame is the most exaggerated expression frame during the apex phase. Both onset and apex frames are

provided by the databases. Second, the facial region around each facial feature point is extracted and normalized to 100×100, in which the locations of the eyes and the tip of the nose are fixed. Through the face alignment and normalization, the facial feature points are robust to different subjects and to moderate face pose variation.

Then, the displacements of the feature points between the onset frame and the apex frame are calculated as features. After that, for each facial feature point, the displacements are discretized with unequal intervals. Each displacement interval represents a certain movement of facial feature points, called as a facial event here.

Last, facial events are represented as binary codes and used as inputs of RBMs. The coding rules are as follows: for every facial point, the displacements of x and y coordinate values of all facial images are computed, respectively. We assume the movement of facial point i along x coordinate ranging from x^i_{min} to x^i_{max}, and that along y ranging from y^i_{min} to y^i_{max}. Let $[x^i_{min}, x^i_{max}]$ be discretized into A intervals, and $[y^i_{min}, y^i_{max}]$ be discretized into B intervals, then we have $A \times B$ different facial events on point i. Accordingly, we use $t_i = \lceil \log_2(A \times B) \rceil$ bit binary code to describe facial events on point i.

2.2 Posed and Spontaneous Expression Modeling Using RBM

In order to model the spatial patterns embodied in posed and spontaneous expressions, RBM which consists of two layers is used. The first layer of RBM, $v \in \{0,1\}^n$, is visible and represents facial point displacement events. The second layer is a latent layer, $h \in \{0,1\}^m$, applied to capture facial spatial patterns. Latent layer of RBM is capable of modeling complex joint distribution over visible layer, i.e. feature point displacement events. In this way, spatial patterns embodied in posed or spontaneous expressions can be captured in the model. Figure 2 shows a RBM structure and the facial point displacements events combination patterns captured by an hidden node, i.e. h_1.

Since RBM is an undirected graphical model, the total energy of RBM is defined in Eq. 1.

$$E < v, h; \theta >= -\sum_i \sum_j v_i W_{ij} h_j - \sum_i b_i v_i - \sum_j c_j h_j \tag{1}$$

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 distribution over visible units of RBM is calculated by marginalizing over all hidden units with Eq. 2, where $Z(\theta)$ is the partition function and $P(h, v; \theta)$ is joint distribution over hidden nodes and visible nodes. This allows RBM to capture global dependencies among the visible variables.

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

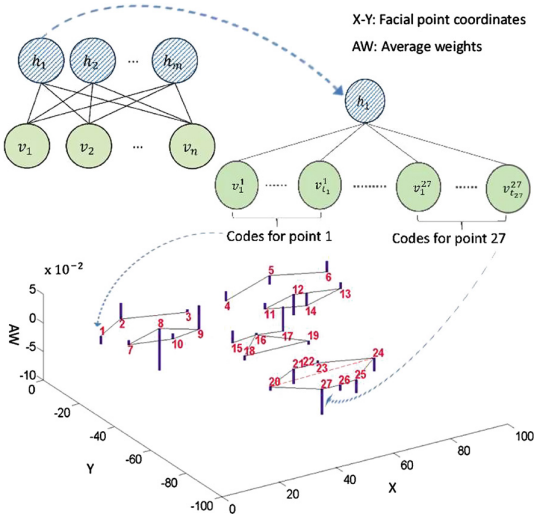

Fig. 2. RBM structure and the combination patterns of facial point events captured by h_1. t_i represents the length of codes for facial point i. At the bottom part, we drew a facial points distribution map at $x - y$ plane. z coordinate is the average weights (AW) of visible nodes for each facial point

According to Bayesian theorem, conditional probability of hidden nodes given visible nodes and of visible nodes given hidden nodes can be estimated as follow:

$$\begin{aligned} P(v|h,\theta) &= \prod_i \delta(\sum_j w_{ij}h_j + b_i); \\ P(h|v,\theta) &= \prod_j \delta(\sum_i v_i w_{ij} + c_j) \end{aligned} \tag{3}$$

Given the training data $\{v_i\}_{i=1}^N$, where N indicates the number of the training samples, the parameters are learned by maximizing the log likelihood with Eq. 4.

$$\theta^* = argmax_\theta L(\theta); L(\theta) = \frac{1}{N}\sum_{i=1}^{N} logP(v;\theta) \tag{4}$$

The gradient with respect to θ can calculated with Eq. 5,

$$\frac{\partial logP(v;\theta)}{\partial\theta} = \langle\frac{\partial E}{\partial\theta}\rangle_{P(h|v,\theta)} - \langle\frac{\partial E}{\partial\theta}\rangle_{P(h,v|\theta)} \tag{5}$$

where $\langle\cdot\rangle_P$ represents the expectation over distribution P. In Eq. 5, inferring partition function $Z(\theta)$ in $P(h,v)$ analytically is intractable. However, Hinton *et al.* [9] proposed an very efficient way to estimate it approximately, namely contrastive divergence (CD). The basic idea of CD algorithm is to approximate $P(h,v)$ with one step sampling from the data. Gradient calculation is basic procedure of RBM training. With the gradients of all parameters, stochastic gradient

descent is applied for RBM training. Meta-parameters such as the learning rate, the momentum, are decided by following the instruction of [8].

We train two RBMs for posed and spontaneous expression respectively. After training, for a test sample t, the log probability that RBM trained on class c assign to the test sample is as follows:

$$\log P(t;\theta_c) = \log\left(\sum_h exp(-E(t;\theta_c))\right) - \log Z(\theta_c) \tag{6}$$

This is the logarithm of Eq. 2. The partition function $Z(\theta_c)$ can be approximately estimated through Annealed Importance Sampling (AIS) approach [16]. With these log probabilities, the label of the test sample is the class with greater value.

3 Experiments and Analysis

3.1 Experimental Conditions

As far as we know, there are several databases available for posed and spontaneous expressions recognition, such as, BBC Smile Dataset [13], MAHNOB-Laughter database [14], UvA-NEMO smile database [2], SPOS database [15], and USTC-NVIE database [19]. Among them, the BBC, MAHNOB-Laughter and UvA-NEMO databases only contain posed and spontaneous expressions for smiles, while the USTC-NVIE and SPOS databases consist of posed and spontaneous expressions for six basic expression categories. Thus, these two databases are adopted in our experiments.

The USTC-NVIE database [19] is a natural visible and thermal infrared facial expression database, which contains both spontaneous and posed expressions with six basic categories (i.e. happiness, disgust, fear, surprise, anger and sadness) of more than 100 subjects. The onset and apex frames are provided for both posed and spontaneous subsets. The SPOS database [15] is a visible and near infrared expression database, including both posed and spontaneous expressions with six basic categories from seven subjects (four males and three females). The image sequences in this database start from onset frame and end with apex frame.

For USTC-NVIE database, both the apex and onset frames of all posed and spontaneous expression samples, which come in pairs from the same subject, are selected. In this procedure, we discarded spontaneous samples which have no expressions, and finally select 1028 samples, including 514 posed and 514 spontaneous expression samples from 55 male and 25 female subjects. The distribution of posed and spontaneous expression samples is shown in Table 1. Our experimental results on the database are obtained by applying a 10-fold cross validation method on all samples according to the subjects.

For SPOS database, the first and the last frames of all the posed and spontaneous samples are selected, including 84 posed expression samples and 150

Table 1. The distribution of samples on USTC-NVIE database

	Happiness	Disgust	Fear	Surprise	Anger	Sadness
Posed	104	93	68	78	91	80
Spontaneous	104	93	68	78	91	80

spontaneous expression samples, as shown in Table 2. Since SPOS database consists of images from only seven subjects (4 males and 3 females), and it does not include all six expression images for certain subjects, we did not select samples in pairs from SPOS database as we did on USTC-NVIE database. In order to compare with [15], leave-one-subject-out cross validation is used.

Table 2. The distribution of samples on SPOS database

	Happiness	Disgust	Fear	Surprise	Anger	Sadness
Posed	14	14	14	14	14	14
Spontaneous	66	23	32	11	13	5

3.2 Experimental Results and Analysis

Figure 3 shows the histogram of the feature point displacements along x and y axis, From Fig. 3, we find that most displacements are at the middle, which means that the movements of feature points are small in most cases.

In order to make each facial event cover similar number of samples, the displacements are discretized into multiple intervals with unequal length. In our experiments, we let the number of samples fall into every interval as close as possible to 100 but no more than 100. After extracting point related facial events, these events are used to form binary codes which are the inputs of RBMs. According to the binary coding rule described in Sect. 2.1, we are able to generate 108-bit binary codes for samples in SPOS database and 216-bit for samples in USTC-NVIE database.

We trained two RBM models from posed and spontaneous samples using the generated binary codes, respectively. Then, the RBMs are used for distinguishing posed vs. spontaneous expression.

In order to analyze the ability of our proposed RBM for modeling spacial patterns in posed and spontaneous expressions, the global spacial pattern captured by both RBMs are showed in Fig. 4. As described in Sect. 2.2, 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 facial events affect the captured spacial pattern. We first summated W over all hidden units for every visible unit, and then computed the average W for visible units represent facial events of the same facial point. Due to the unbalanced data in

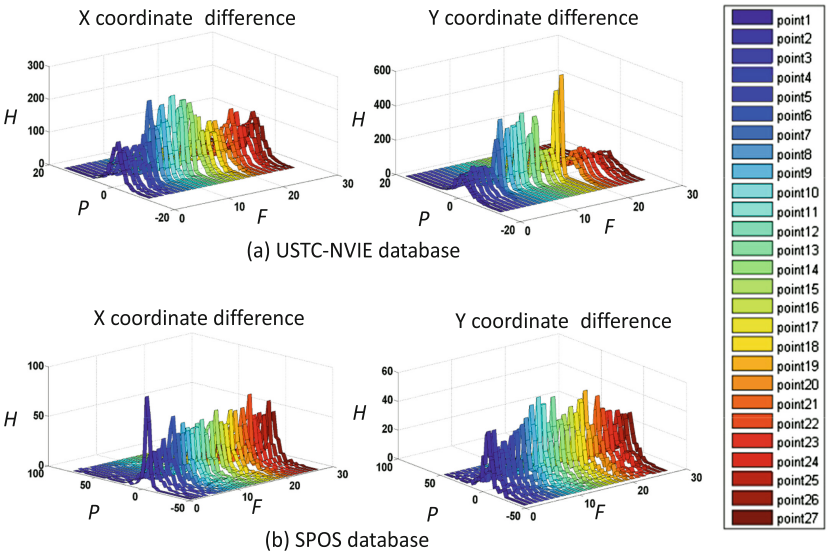

Fig. 3. The histogram of x and y coordinate value differences between apex and onset facial points for all samples on two database. The number of bins is 50 for all facial points. P: Position of bins; F: Facial points number; H: Height of bins

SPOS database, the weights are normalized by the number of samples for each RBM. From Fig. 4, we can obtain the following observations: first, the W values of RBM for posed expressions (red bars) and those for spontaneous expressions (blue bars) are different, proving that the spacial pattern for posed expressions and that for spontaneous expressions are different. This is consistent with current behavior research. Second, in most cases, the weights of posed RBM are larger than those of spontaneous RBM. It further confirms that posed expressions are more exaggerated than spontaneous one.

The recognition results on both databases are shown in Table 3. The recognition accuracy for all expressions reaches 81.23 % and the F1-score is 0.8012, on USTC-NVIE database. On SPOS database, the accuracy achieves 74.36 %, however, due to the unbalanced data the F1-score is only 0.5231.

3.3 Comparison with Other Methods

In order to further demonstrate the effectiveness of our proposed method, we compared our work with Zhang's [20] and Pfister's [15] works. In additional, as a baseline, we conducted experiments using support vector machine (SVM) with linear kernel under the same experimental conditions with our work.

Zhang *et al.* selected 3572 posed and 1472 spontaneous images. Since they did not explicitly state which images were selected, it is hard for us to select the same images as theirs. We can only compare the experimental results as a reference. The results of our work and the best results of [20] are shown in

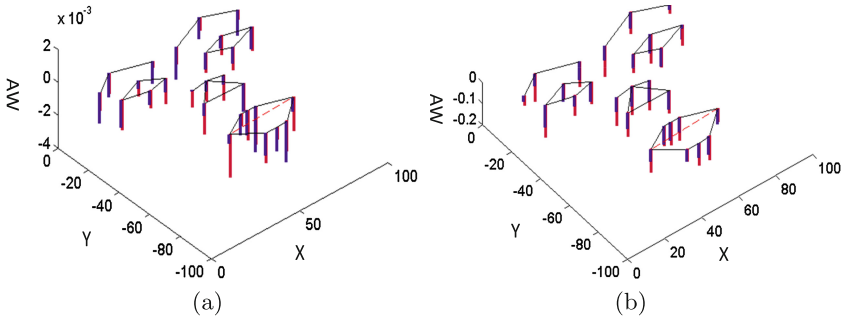

Fig. 4. Average W at every facial points from the trained RBMs, (a) and (b) are from the RBMs trained on USTC-NVIE database and SPOS database, respectively. We drew a facial points distribution map at $x-y$ plane. z coordinate is the average weights (AW). The red bars represent average W value from RBM for posed expressions, and the blue bars represent that for spontaneous ones (Color figure online)

Table 3. Experiment result on USTC-NVIE database and SPOS database

	USTC-NVIE database			SPOS database	
Confusion matrix		Posed	Spont.	Posed	Spont.
	Posed	389	125	35	49
	Spont.	68	446	11	139
Accuracy	81.23 %			74.36 %	
F1-score	0.8012			0.5385	

"Posed" represents posed expression.
"Spont." represents spontaneous expression.

Table 4. From Table 4, we can find that although the number of samples are smaller than Zhang *et al.*'s, our proposed models outperform their.

Pfister *et al.* [15] distinguished posed and spontaneous expression from both visible and near-infrared image sequences on SPOS database. Here, we only compare our work with their work on visible images, as shown in Table 4. From this table, we can find that the accuracy of our method is 2.36 % higher than Pfister *et al.*'s. Furthermore, they extracted CLBP-TOP texture features from facial expression sequence, while our features are extracted from the apex frames and onset frames. It indicates that, with less information, our proposed method achieve better results.

Table 4 also shows the results achieved by using SVM with linear kernel as classifier. The classification accuracy reaches 78.02 % on USTC-NVIE database, and 69.66 % on SPOS database. Our approach outperforms the baseline.

The above comparison demonstrates the performance of our approach is better than the state of the art. The most important reason that contributes the performance of our method is the use of the RBM to globally capture the relationships among the spatial movements of facial landmark points. This explains

Table 4. Comparison between our method and related work

USTC-NVIE database			
Method	ours	L. Zhang *et al.* [20]	SVM
Accuracy (%)	81.23	79.43	78.02
SPOS database			
Method	ours	T. Pfister *et al.* [15]	SVM
Accuracy (%)	74.36	72.0	69.66

why our method outperforms [15,20] in spite of their use of more powerful features and classifiers.

4 Conclusion and Future Works

In this paper, we proposed a new method to recognize posed and spontaneous expressions by capturing the global facial spatial patterns of posed and spontaneous expressions using RBM models. First, the displacements of facial feature points between apex and onset facial images are recorded in the form of coordinates value variation. We analyzed distributions of these displacements of all facial points by computing their histograms. Second, the coordinates value variation of all facial points are discretized as facial point displacement related facial events which are used to form binary codes as inputs of RBMs. Third, two different RBM models are trained to capture spatial patterns embedded in posed and spontaneous expressions respectively. The recognition results on both USTC-NVIE and SPOS databases demonstrated the effectiveness of our method in modeling spacial patterns of posed and spontaneous expressions, good performance on posed and spontaneous expression recognition.

In the future works, we will consider modeling the temporal patterns of posed and spontaneous expressions to improve recognition performance.

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

References

1. Cohn, J., Schmidt, K.: The timing of facial motion in posed and spontaneous smiles. Int. J. Wavelets Multiresolut. Inf. Process. **2**(02), 121–132 (2004)
2. Dibeklioğlu, H., Salah, A.A., Gevers, T.: Are you really smiling at me? spontaneous versus posed enjoyment smiles. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 525–538. Springer, Heidelberg (2012)
3. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**, 91–110 (2004)

4. Ekman, P.: Darwin, deception, and facial expression. Ann. NY Acad. Sci. **1000**(1), 205–221 (2003)
5. Ekman, P., Friesen, W.: Felt, false, and miserable smiles. J. Nonverbal Behav. **6**(4), 238–252 (1982)
6. Ekman, P., Hager, J., Friesen, W.V.: The symmetry of emotional and deliberate facial actions. Psychophysiology **18**(2), 101–106 (1981)
7. Ekman, P., Rosenberg, E.: What the Face Reveals: Basic and Applied Studies of Spontaneous Expression using the Facial Action Coding System (FACS). Oxford University Press, New York (1997)
8. Hinton, G.: A practical guide to training restricted boltzmann machines. Momentum **9**(1), 926 (2010)
9. Hinton, G.E.: Training products of experts by minimizing contrastive divergence. Neural Comput. **14**(8), 1771–1800 (2002)
10. Littlewort, G., Bartlett, M., Lee, K.: Automatic coding of facial expressions displayed during posed and genuine pain. Image Vis. Comput. **27**(12), 1797–1803 (2009)
11. Nie, S., Ji, Q.: Capturing global and local dynamics for human action recognition. In: ICPR (2014)
12. Pandzic, I.S.: Mpeg-4 Facial Animation: The Standard, Implementation and Applications. John Wiley & Sons Inc., New York (2003)
13. Paul, E.: Bbc-dataset. http://www.bbc.co.uk/science/humanbody/mind/surveys/smiles/
14. Petridis, S., Martinez, B., Pantic, M.: The mahnob laughter database. Image Vis. Comput. **31**, 186–202 (2013)
15. Pfister, T., Li, X., Zhao, G., Pietikainen, M.: Differentiating spontaneous from posed facial expressions within a generic facial expression recognition framework. In: ICCV Workshops, pp. 868–875. IEEE (2011)
16. Salakhutdinov, R., Murray, I.: On the quantitative analysis of deep belief networks. In: ICML, pp. 872–879. ACM (2008)
17. Tong, Y., Wang, Y., Zhu, Z., Ji, Q.: Robust facial feature tracking under varying face pose and facial expression. Pattern Recogn. **40**(11), 3195–3208 (2007)
18. Valstar, M., Gunes, H., Pantic, M.: How to distinguish posed from spontaneous smiles using geometric features. In: Proceedings of the 9th International Conference on Multimodal Interfaces, pp. 38–45. ACM (2007)
19. Wang, S., Liu, Z., Lv, S., Lv, Y., Wu, G., Peng, P., Chen, F., Wang, X.: A natural visible and infrared facial expression database for expression recognition and emotion inference. IEEE Trans. Multimedia **12**(7), 682–691 (2010)
20. Zhang, L., Tjondronegoro, D., Chandran, V.: Geometry vs. appearance for discriminating between posed and spontaneous emotions. In: Lu, B.-L., Zhang, L., Kwok, J. (eds.) ICONIP 2011, Part III. LNCS, vol. 7064, pp. 431–440. Springer, Heidelberg (2011)