Emotion Recognition Using Hidden Markov Models from Facial Temperature Sequence

Zhilei Liu and Shangfei Wang*

Key Lab of Computing and Communicating Software of AnHui Province
School of Computer Science and Technology
University of Science and Technology of China
HeFei, AnHui, P.R. China
leivo@mail.ustc.edu.cn, sfwang@ustc.edu.cn

Abstract. In this paper, an emotion recognition from facial temporal sequence has been proposed. Firstly, the temperature difference histogram features and five statistical features are extracted from the facial temperature difference matrix of each difference frame in the data sequences. Then the discrete Hidden Markov Models are used as the classifier for each feature. In which, a feature selection strategy based on the recognition results in the training set is introduced. Finally, the results of the experiments on the samples of the USTC-NVIE database demonstrate the effectiveness of our method. Besides, the experiment results also demonstrate that the temperature information of the forehead is more useful than that of the other regions in emotion recognition and understanding, which is consistent with some related research results.

Keywords: emotion recognition, facial temporal sequence, Hidden Markov Models.

1 Introduction

As the development of Human-Computer Interaction (HCI) in the domain of health care, service robotic, security industry, gaming and so on, emotional HCI has attracted more and more attentions in the past few years, in which, proper understanding of human emotions is a key problem to be solved first.

Human's emotions could be manifested in various ways, including both the external signals, such as facial expressions, body gestures, speech and so on, and some internal signals, such as blood flow, heart rate, EEG, body temperature and so on [7]. Compared to the emotion recognition using other signals, the emotion recognition based on the temperature information reflected through the infrared thermal images may be more practical because of its non-invasive and non-verbal characteristics [5], [7], [8], [13]. Generally speaking, two kinds of features extracted from the infrared thermal data are considered in most of existing researches, the first one is the imaging features extracted from the infrared thermal images, for instance: Benjamín Hernández, Gustavo Olague et al. selected

* This author is the corresponding author.

S. D´Mello et al. (Eds.): ACII 2011, Part II, LNCS 6975, pp. 240–247, 2011.

the Gray Level Co-occurrence Matrix (GLCM) to compute region descriptors of the infrared images and used them to distinguish the expressions of surprise, happiness and anger [14], Guotai Jiang et al. conducted the facial expression recognition through drawing and analyzing the whole geometry characteristics and some geometry characteristics of the ROI in infrared images by using mathematics morphology [15], Yasunari Yoshitomi et al. extracted the features by using a two-dimensional discrete cosine transformation (2D-DCT) to transform the gray scale values of each block in the face portion of an infrared image into frequency components, and these features were used in their expression recognition systems [9]; the other feature is the temperature feature recorded by infrared cameras, such as: Masood Mehmood Khan et al. have tried to use the variances in the thermal intensity values recorded at thermally significant locations on human faces as the features to discern some pretended expressions as well as the pretended and the evoked emotional expressions [10], [11], Brain R. Nhan and Tom Chau have extracted the time, frequency and time-frequency features derived from 12 adults' thermal infrared data to classify the natural responses of subject-indicated levels of arousal and valence stimulated by the International Affective Picture System [6], A. Merla and G. L. Romani have studied the facial thermal signatures of 10 healthy volunteers' three fundamental emotional conditions: stress, fear and pleasure arousal [12]. All these works have shown that the human's emotion states or expressions are relevant to the properties of the facial temperature. However, to our best knowledge, most researches extracted the features from a single apex or onset and apex infrared thermal data or images, only a few features are extracted from the emotional data sequences, which may lose some useful information contained in the sequences [16].

In this paper, an emotion recognition method using the temporal information of the facial temperature data is provided. Firstly, the temperature data of the facial region are extracted and segmented into some facial sub regions. Secondly, the temperature difference histogram features (TDHFs) and five statistic features (StaFs) are extracted from each facial sub region's temperature difference matrix obtained from two consecutive frames in the sequences. Thirdly, a feature selection strategy based on each feature's recognition ability on the training data is used, and these selected features are used in some discrete Hidden Markov Models (HMMs) to recognize the emotion states. Experiments performed on the samples selected from the USTC-NVIE database [1] are implemented to verify the effectiveness of this method. Compared with other researches, the contribution of this paper is that the temporal information of the human facial temperature data in different emotion states is fully considered, and our research is one of the first concerted attempts at emotion recognition using the temporal information of human's facial temperature data sequences.

The remainder of the paper is organized as follows. The details of our approach are explained in Section 2, experiments and results conducted on the USTC-NVIE database are given in Section 3, finally, some conclusions and future works are described in Section 4.

2 Methodology

The schematic representation of our approach is shown as Fig. 1, which could be divided into four modules, named data preprocessing and feature extraction, feature selection and classification. The details are provided as follows.

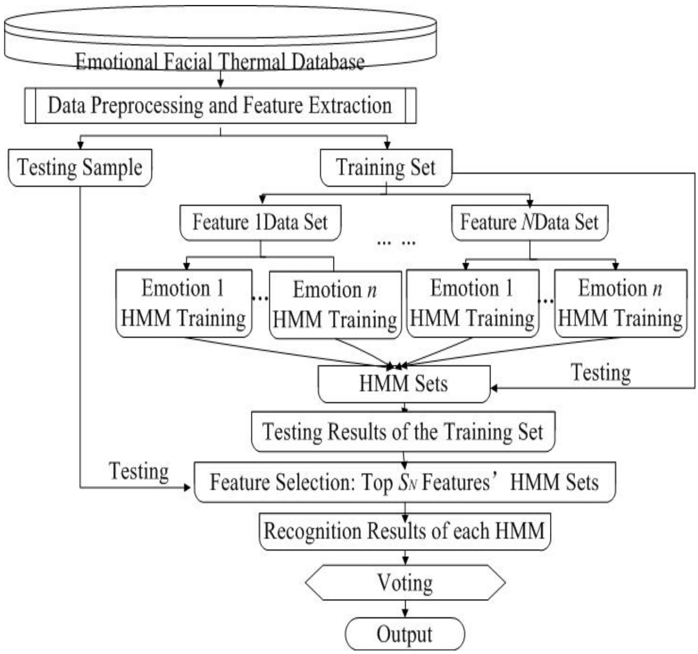

Fig. 1. The framework of the emotion recognition method

2.1 Data Preprocessing and Feature Extraction

The temperature variations of the objects could be detected by an infrared thermal camera based on the objects' infrared radiations and the black body radiation law. Firstly, the temperature data of each frame is converted into a gray-scale image, which is used to obtain the facial region's temperature data. Secondly, it is difficult to achieve automatic location of the feature points in the infrared image due to its low resolution and low contrast, for this reason, three points P_l, P_r and P_n are marked manually in our approach, which are the centers of the both eyes and the nose tip. After that, the facial region's temperature

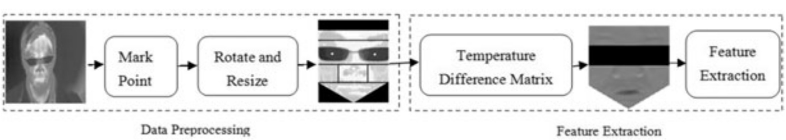

Fig. 2. The procedure of data preprocessing and feature extraction

is determined and marked out based on these three points. Thirdly, the facial region's temperature data matrix is rotated based on the angle of the P_l and P_r in the horizontal direction and resized based on the distance of P_l and P_r and the distance between the P_n and the connection line of P_l and P_r. Finally, this facial mask's temperature matrix is obtained as shown in Fig. 2, and this facial mask is divided into five regions, named forehead, eye, nose, mouth and cheek.

For each facial sub-region, the temperature difference matrix (DMat) between two consecutive frames could be obtained, from which some features are extracted. In this paper, two kinds of features are extracted, named temperature difference histogram features (TDHFs) and statistical features (StaFs). TDHF describes the data distribution of the DMat, which is similar to the gray level histogram of a gray image. Suppose the size of DMat is $M \times N$, the lower limit and the upper limit of the DMat data are L and U separately, and the dimension of the TDHF is D, then the TDHF of DMat could be defined as formula (1).

$$TDHF(i) = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N}[T(i-1)) < DMat(m,n) \leq T(i)]}{M \times N} \tag{1}$$

In which, $i = 1, 2, ..., D$, $T(i) = L + \frac{(U-L)\times i}{D}$ is the endpoint of the i-th interval and $TDHF(i)$ is the frequency of data points between the interval of $(T(i-1), T(i)]$ in DMat.

Besides, five StaFs are also extracted from each DMat, which are: VAR, which is the variance of the DMat; MEAN, which is the mean of DMat; ADDP and ADDN, which represent the mean of positive and negative values of DMat; ABS which is the mean of absolute values of DMat.

Thus, D-dimensional TDHFs and 5-dimensional StaFs are extracted from each DMat of each facial sub region in the recording temperature sequence.

2.2 Feature Selection Strategy

In our method, a feature selection strategy based on each feature's recognition results in the training set is considered. Suppose N-dimensional features have been extracted and the classifier of each feature could be trained by using the training set at first. Next, for each feature, the recognition results of the samples in the training set could be obtained through these well trained classifiers. After that, these features are sorted based on their average recognition rates in the training set. Finally, the classifiers of S_N best selected features with the highest S_N average recognition rates are selected and used in the testing phase.

2.3 Emotion Recognition Using HMMs

HMM is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved states, which could be well applied in the classification problem based on sequence features, especially known for its application in temporal pattern recognition such as speech, handwriting, gesture recognition and so on [2], [4]. In this paper, the discrete HMM is adopted as the classifier for each feature by using the HMM toolbox for Matlab [3].

Before the classification, the feature data quantization of each dimension is performed. Take the i-th dimension feature data $F(i)$ of all the samples as the example, these data are normalized into the interval of $[0, 1]$ and the probability distribution between 0 and 1 is calculated at first. Suppose the feature data are quantified to 1,2,...,N, which is the same as the number of the state variable in the HMMs, then the data within the probability distribution interval of $[\frac{(m-1)}{N}, \frac{(m)}{N})$ are quantified as m, in which $m = 1, 2, ..., N$.

Next, for the feature of each dimension in each facial sub region, n different HMMs are established for n-different specific emotion states respectively in the model training phase. Thus, $(D + 5) \times 5 \times n$ HMMs are constructed in our method. In the model testing phase, a test sample's recognition result of i-th feature is determined based on the category of the HMM with the maximum log-likelihood value of these n specific HMMs, and this sample's final emotion state is determined by the voting strategy based on these recognition results of $(D + 5) \times 5$ features.

3 Experiments and Results

3.1 Experimental Conditions

The samples in our experiment are selected from the USTC-NVIE database [1], which contains both spontaneous and posed expressional images of more than 100 subjects, recorded simultaneously by a visible and an infrared thermal camera, with illumination provided from three different directions. In each experiment, subject's emotions are elicited by watching some emotional videos and reported in the self-reported data including the evaluation value of the six basic emotions, named happiness, disgust, fear, sad, surprise and anger, on 5-point scale. This self-reported data are used to determine the emotion label of this subject when watching this emotional video. Based on the analysis results in [1], three emotions are considered in this paper, that is happiness, disgust and fear, which have the greatest impact to the facial temperature among these six emotions, and 176 samples' temperature data sequences are selected and used in the following experiments, including 69 happiness, 42 disgust and 65 fear. All these samples' self-report data about the primary emotion category are larger than 1.

The size of the facial mask is 80×84. As most of the subjects wore glasses, thereby masking the thermal features of the eye region, then eye regions is not taken into account in our experiment. The lower and upper limits of the DMat are -10 and 10, and the dimension of the TDHF is 20. Thus, for each facial sub region, 20 TDHFs and 5 StaFs are extracted.

In these discrete HMMs, the feature data are quantified into 1,2,...,12, which is the same as the number of state variable in our paper, the number of the observed variable is 12. In the training phase, the Baum-Welch estimation method is used, and the initial probability vector P, state transition matrix A and observation

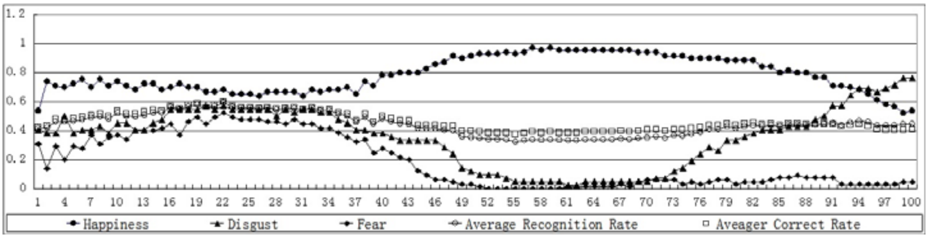

Fig. 3. The recognition rates with different number of selected features

probabilities matrix B are all random initialized. The leave-one-sample-out cross-validation is adopted in the following experiments.

3.2 Results and Analysis

Experiments with different number of the selected features are conducted, the results are shown as Fig. 3, in which, the recognition rate of each emotion state, the average recognition rate and average correct rate are described.

From Fig. 3, we could find that when the number of the selected feature is 22 the best overall recognition results are achieved, that is, the recognition rates of happiness, disgust and fear are 0.68116, 0.57143 and 0.52308 respectively, the average recognition rate and average correct rate are 0.59189 and 0.59660 respectively. When the classification is conducted without feature selection, in other words, the selected feature number is 100, the recognition rates of happiness, disgust and fear are 0.53623, 0.76191 and 0.04615 respectively, the average recognition rate and average correct rate are 0.44810 and 0.40910, which verifies the effectiveness of the feature selection strategy in our method.

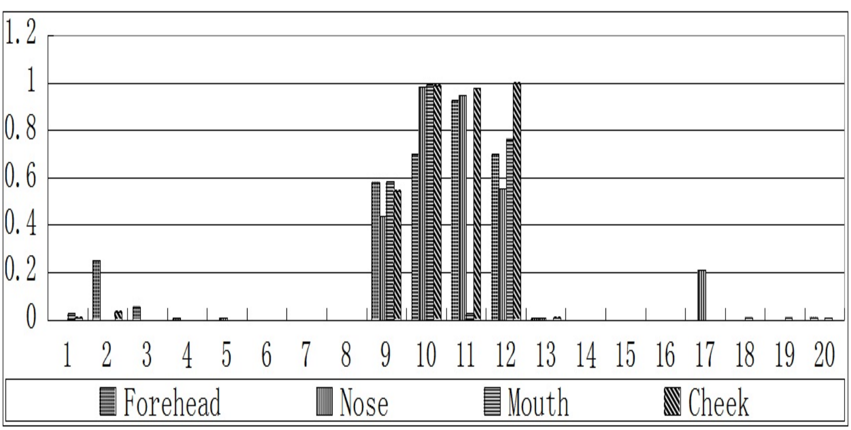

Fig. 4. The probability distribution of the TDHF features in each facial sub-region

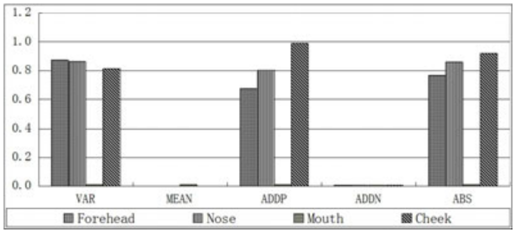

Fig. 5. The probability distribution of the StaFs features in each facial sub-region

Table 1. The distribution the best selected features

	Forehead	Nose	Mouth	Cheek
TDHF	9,10,11,12	10,11,12	9,10,12	9,10,11,12
StaF	VAR,ADDP,MEAN	VAR,ADDP,MEAN	NULL	VAR,ADDP,MEAN

Next, the distributions of all the features in each facial sub region of all the samples are analyzed as follows when the number of the selected features is set to 22. The probability distribution of the 20-TDHFs and five StaFs in four facial sub region are described in Fig. 4 and Fig. 5, and the distributions of these 22 best selected features with the highest recognition rates in each facial region are shown in Table 1, from which we conclude that:

1) From Fig. 4 and Fig. 5, we could find that the most used HDTFs are in 9-12 dimensions, and the most used StaFs are VAR, ADDP and ABS;

2) Most of the best selected features are belong to the forehead region, which indicates that forehead's temperature features are more useful than the other regions', however, the mouth region's features are the least, especially for the StaFs, no one is included, which means that these StaFs of temperature data in the mouth are inadequate to represent these emotion states, all these results are consistent with the analysis in [1], [17].

4 Conclusion and Future Work

In this paper, an emotion recognition method based on the temporal information of the human facial temperature data and discrete HMMs is introduced, in which, a feature selection strategy based on the recognition result of the training set is adopted to improve the overall recognition rate. Finally, experiments on the samples selected from the USTC-NVIE database are implemented to verify the effectiveness of our method.

Some additional works are necessary to improve the accuracy of the recognition before the practical applications can be realized, for example, the automatic facial location and feature extraction, parameter optimization in the classifiers, the decision strategy of the final results and so on, all these works will be completed in the future.

Acknowledgments. The authors would like to thank all the subjects who participated in the experiments. This work is supported by National Program 863 (2008AA01Z122), Youth Creative Project of USTC and SRF for ROCS, SEM.

References

1. 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 Transactions on Multimedia, 682–691 (2010)
2. Cohen, I., Garg, A., Huang, T.S.: Emotion recognition from facial expressions using multilevel HMM. In: NIPS (2000)
3. Hidden Markov Model Toolbox for Matlab, `http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html`
4. Nefian, A.V., Liang, L., Pi, X., Xiaoxiang, L., Mao, C., Murphy, K.: A coupled HMM for audio-visual speech recognition. In: International Conference on Acoustics, Speech and Signal Processing CASSP 2002, pp. 2013–2016 (2002)
5. Puri, C., Olson, L., Pavlidis, I., Levine, J., Starren, J.: StressCam: non-contact measurement of users' emotional states through thermal imaging. In: CHI Extended Abstracts, pp. 1725–1728 (2005)
6. Nhan, B.R., Chau, T.: Classifying affective states using thermal infrared imaging of the human face. IEEE Transaction on Biomedical Engineering 57(4), 979–987 (2010)
7. Gunes, H., Pantic, M.: Automatic, Dimensional and Continuous Emotion Recognition. International Journal of Synthetic Emotions 1(1), 68–99 (2010)
8. Pavlidis, I., Levine, J.: Thermal image analysis for polygraph testing. IEEE Engineering in Medicine and Biology Magazine 21(6), 56–64 (2002)
9. Yoshitomi, Y.: Facial Expression Recognition for Speaker Using Thermal Image Processing and Speech Recognition System. In: Proc. of 10th WSEAS International Conference on Applied Computer Science, pp. 182–186 (2010)
10. Khan, M.M.: Cluster-analytic classification of facial expressions using infrared measurements of facial thermal features. Ph.D. Thesis, Department of Computing and Engineering, University of Huddersfield, Huddersfield, UK (2008)
11. Khan, M.M., Ward, R.D., Ingleby, M.: Classifying pretended and evoked facial expressions of positive and negative affective states using infrared measurement of skin temperature. Trans. Appl. Percept. 6, 1 (2009)
12. Merla, A., Romani, G.L.: Thermal signatures of emotional arousal: A functional infrared imaging study. In: IEEE 29th Annu. Int. Conf., pp. 247–249 (2007)
13. Calvo, R.A., D'Mello, S.: Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. IEEE Transactions on Affective Computing 1(1), 18–37 (2010)
14. Hernández, B., Olague, G., Hammoud, R., Trujillo, L., Romero, E.: Visual learning of texture descriptors for facial expression recognition in thermal imagery. Computer Vision and Image Understanding 106(2-3), 258–269 (2007)
15. Jiang, G., Song, X., Zheng, F., Wang, P., Omer, A.M.: Facial Expression Recognition Using Thermal Image. In: 27th Annual International Conference of the Engineering in Medicine and Biology Society, IEEE-EMBS 2005, pp. 631–633, 17–18 (2006)
16. Bassili, J.N.: Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. J. Personality Social Psychology 37, 2049–2058 (1979)
17. Merla, A., Romani, G.L.: Thermal signatures of emotional arousal: A functional infrared imaging study. In: Proc. IEEE 29th Annu. Int. Conf., pp. 247–249 (2007)