Person independent facial expression analysis using Gabor features and genetic algorithm

Shojaeilangari, Seyedehsamaneh; Yau, Wei-Yun; Teoh, Eam Khwang

2011

Shojaeilangari, S., Yau, W. Y., & Teoh, E. K. (2011). Person independent facial expression analysis using Gabor features and genetic algorithm. 8th International Conference on Information, Communications & Signal Processing, 1-5.

<https://hdl.handle.net/10356/101109>

<https://doi.org/10.1109/ICICS.2011.6173537>

© 2011 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at: [<http://dx.doi.org/10.1109/ICICS.2011.6173537>].

Downloaded on 26 Apr 2023 22:02:56 SGT

Person independent facial expression analysis using Gabor features and genetic algorithm

Seyedehsamaneh Shojaeilangari
Division of Control, EEE
Nanyang Tech. Univ.
Singapore

Yau Wei Yun
Institute for Infocomm Research
A-star
Singapore

Teoh Eam Khwang
Division of Control, EEE
Nanyang Tech. Univ.
Singapore

Abstract—Over the last decade, automated analysis of human affective behavior has become an active research area in computer science, psychology, neuroscience, and related fields. This study investigates the application of Gabor filter based features in combination of Genetic Algorithm (GA) and Support Vector Machine (SVM) for dynamic analysis of six basic facial expressions from video sequences. Traditionally, a set of Gabor filters is used for feature extraction from static images of face. However, we employed Sum of Difference (SOD) approach to analysis the dynamics of facial expression from a video sequence. We also used GA to overcome the problem of high dimensional feature vectors and computation cost. A local Gabor filter bank with selected frequencies and orientations is produced by GA. The experimental results show that the proposed method is effective for temporal analysis of affective states. The detection rate of six basic emotions has been reached to 92.97% for Cohn-Kanade (CK⁺) database.

Keywords—*affect analysis; Gabor filter; emotion recognition; SVM; genetic algorithm*

I. INTRODUCTION

Affects or the conscious subjective aspect of emotions are considered as essential factors in communication and social interaction. The changes in a person's affective state usually complement the transmission of a message in human-human interaction. People are able to sense these changes and exploit it to enhance their communication. This fact has motivated a huge number of researches to enable machines to recognize humane emotion. The ability to understand human affective states is beneficial to facilitate human-compute interaction, enable the intelligent machines to respond to users based on their preference.

Essentially, facial expression as a visual information is considered the most important cue for assessment of human affective behavior. In spite of the various techniques reported in this field, there are still challenges remaining. The majority of previous works tried to recognize facial expressions via static images and ignore the temporal information of an emotion in time domain. However, temporal analysis of a dynamic process can strongly improve the performance of an affect analysis system in real world application.

In this paper, we focus on facial-based emotion recognition from video sequences considering the temporal information of each expression. In the proposed work, the 2D Gabor filter is applied to each frame of a video sequence to extract the relevant features. Then we compute the between-frame differences to track the facial feature changes for expressing a certain emotion. Since huge dimensional Gabor features require computation and memory cost, a feature reduction technique is applied.

Past literatures [1, 2] applied statistics of Gabor features (such as mean and standard deviation) to reduce the number of features. In our case, we applied Genetic Algorithm (GA) to select several filters which are more effective to discriminate various emotions. In other word, GA selects a number of frequency and orientation set of Gabor filters from the global filter bank to reduce the length of the feature vector. Using mean and standard deviation of Gabor feature which is usually used to deal with high dimensional feature vector will not be able to pick up the key features while using GA to reduce the feature length preserves the unique information.

The novelty of our proposed method are: (1) using Gabor features in a dynamic approach to interpret the temporal information of an affective face from a video sequence and (2) selecting the most relevant Gabor filters by GA instead of using the entire filter bank.

The paper is organized as follows: Section II summarizes the literature review. Section III describes the database used. The proposed method including pre-processing, feature extraction, classification, and feature selection is explained in section IV. The experimental results are discussed in section V and finally, section VI concludes the paper.

II. RELATED WORKS

Facial expression recognition systems are usually classified into two categories: target- and gesture-oriented [3]. The target-oriented systems select a single key frame involving the peak facial expression from a given sequence of images. Gesture-oriented approaches use temporal information in the sequence by tracking some specific facial features over all frames of a video sequence.

The fundamental step for pattern recognition is the feature extraction process. In target-oriented approaches, the process of extracting the facial expression information is addressed as localization of face and its features. In Gesture-oriented approaches, the key challenge is tracking the face and its features [4]. The widely used facial features are geometric and appearance-based features. Geometric features considered are either the shape of facial components (eyes, eyebrows, nose and mouth) or the location of facial landmarks (corner of eyes, brows, lip, and etc.). Appearance features used are the facial intensity distribution and texture (wrinkles, furrows and bulges) [5].

The work of Pantic et al. [6] is an example of geometric-feature-based methods. They tracked 15 facial points around the eyes, brows, nose, mouth and chin of an input face video to observe the expression.

Gabor wavelets, Haar like wavelets and Local Binary Pattern are three examples of appearance-feature-based methods. The major appearance-based researches have focused on Gabor-wavelet representation. Filtering face images with a bank of Gabor filters is compute and memory intensive. However, it is still widely used because of its effective performance on affect analysis. The study of Bartlett et al. explored the fully automatic spontaneous facial expression detection via Gabor features [7].

Whitehill and Olmin used some Haar-like features selected via boosting algorithm which was proposed by Viola and Jones [8, 9]. They evaluated this approach (Haar+Adaboost) both in classification accuracy and computational time and then compared it to the standard Gabor+SVM approach and showed that both methods have the same recognition rate, but the first approach is superior in term of lower computational time. Kim et al. extended the Haar features for emotion recognition purpose, because the Viola-Jones rectangle features were initially suggested for face detection and are more able to classify the face images of different subjects [10]. In this work, novel rectangle features were proposed to distinguish between different expressions even they are from one subject. In other word, they defined some expression specific features based on characteristics of human facial expressions.

Local Binary Pattern (LBP) is another example of appearance-based features which was used by Shan et al. [11]. In this research, they illustrated that LBP features can be useful for low resolution facial expression analysis which is a critical issue in real world applications. In addition, LBP features are computationally simple and robust against illumination changes [11].

Typical examples of gesture-oriented systems which are based on feature tracking can be found in [12] and [13]. Automatic lip tracking for emotion estimation was investigated in [12]. The strategy applied in this paper is tracking the feature points selected around the lip by dynamic programming and Viterbi algorithm.

Up to now, the most comprehensive emotional facial systems is the Facial Action Coding System (FACS) proposed by Ekman et al. [14]. In this system, Coders manually coded approximately all facial expressions by viewing a video sequence of the facial behavior. Then, the facial expressions were decomposed to some smallest

visually discriminable facial movements called action units (AUs) where each AU is related to one or more facial muscles. The database is rich and the proposed methodology is able to code all facial movements. However, it requires frame by frame manual coding of the facial features which is highly manual intensive [15].

Tian et al. developed an automatic face analysis system which can categorize facial expression changes into AUs, instead of predefined basic emotions [16]. Their approach have been proposed for tracking and modeling the various facial features including both permanent facial features (eyes, brows, mouth and cheeks) and transient facial feature (deepening of facial furrows).

III. DATASET DESCRIPTION

The extended Cohn-Kanade dataset (CK⁺) [17] is used in this work. The database includes approximately 593 video sequences recorded from 123 university students ranging from 18 to 30 years old. The subjects were asked to express a series of 23 facial displays including single or combined action units. Six of the displays were based on descriptions of prototype basic emotions (joy, surprise, anger, fear, disgust, and sadness). Image sequences from neutral to target display were digitized into 640 by 480 or 490 pixel arrays with 8-bit precision for grayscale values. For our work, 321 sequences from the dataset which have been labeled as one of six basic emotions are used.

IV. PROPOSED APPROACH

The key components of the proposed facial-based affect analysis system are described below.

A. Pre-processing and feature extraction

The preprocessing stage normalizes the face and then segments the face region from the background. Since the facial landmarks' locations are given in the dataset, we manually performed the face alignment and face cropping. The images were aligned to have a constant distance between two eyes, and then rotated to line up the eye coordinates. Finally, the faces were cropped using a rectangle of size 60*70.

To recognize facial expressions from video sequences, a set of features that best describe facial muscle changes during an expression is needed. Gabor filters are considered as a powerful method in image processing and computer vision because of their ability to describe the image properties in both spatial and frequency domains. The frequency and orientation representations of Gabor filters are found to be similar to those of human visual system.

In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave and can be formulated as follows:

$$G(x, y, \omega, \theta) = \frac{1}{2\pi\sigma^2} e^{-\left(\frac{x'^2 + y'^2}{2\sigma^2}\right)} \left[e^{-i\omega x'} - e^{-\left(\frac{\omega^2 \sigma^2}{2}\right)} \right],$$

$$x' = x \cos \theta + y \sin \theta, y' = -x \sin \theta + y \cos \theta \quad (1)$$

Where (x, y) is the pixel position, ω and θ are the radial center frequency and filter orientation respectively and σ is the standard deviation of Gaussian function.

The Gabor feature of an image from sequenced images is the convolution of the image with the Gabor filter as follow:

$$O_{m,n}(x, y, t) = |I(x, y, t) * G(x, y, \omega_m, \theta_n)| \quad (2)$$

Indeed, $O_{m,n}(x, y, t)$ is the magnitude of the filtered image at time t , and $G(x, y, \omega_m, \theta_n)$ is a Gabor filter with frequency m and orientation n . Usually five frequencies and eight orientations are used to define a Gabor filter bank.

Finally, Sum of Difference (SOD) technique is applied for all filtered frames of a video sequence to estimate the facial changes during an expression. The feature is described by:

$$F_{m,n}(x, y) = \sum_{t=1}^{k-1} O_{m,n}(x, y, t+1) - (x, y, t) \quad (3)$$

Where $F_{m,n}$ is a feature matrix with the same size with the face images and K is the total number of frames for each video sequence.

B. Filter Selection

In practice, the dimension of the extracted feature vector for a Gabor filter bank is huge and thus would require a very long processing time. It is found that extracted feature vectors for some Gabor filters in a filter bank are correlated and contains redundancy. Indeed, a few Gabor filters can be selected from a filter bank to reduce the number of features and the computation time.

Genetic Algorithm (GA) is a bio-inspired search heuristic which can generate useful solutions to an optimization problem. For feature selection purpose, the algorithm requires a fitness function to provide a measure to evaluate each individual population.

In this study, each individual which is a binary string represents a subset of Gabor filters. Each binary digit (gene) represents the presence (1) or absence (0) of predefined filters. The length of individuals should be set based on the number of filters. Since we have defined 40 Gabor filters (five frequencies and eight orientations) in our experiments, 40 genes was considered for each individual. Also the detection rate of the classifier was used as the fitness function.

C. Classification

Support Vector Machine (SVM) has been successfully applied to various pattern recognition tasks such as classification [18]. SVM classifier is based on statistical learning theory which is able to handle both linear and nonlinear classification problems. SVM maps the input data set into a higher dimensional feature space through several kernel functions [19]. It then attempts to find the optimal separation hyperplane with the maximum margin in the feature space. A hyperplane $w^T x + w_0 = 0$ (w is the vector of coefficients and w_0 is a bias term) is the optimal separating hyperplane which has maximum margin to the nearest sample. After finding the optimal hyperplane, the decision function is obtained as follow:

$$f(y) = \text{sgn} \left[\sum_{i=1}^N \alpha_i y_i K_{\theta}(x_i, y) + b \right] \quad (4)$$

Where N is the sample size, $K_{\theta}(x_i, y)$ represents defined kernel function, θ is a set of parameters and b is the bias. α_i are non-negative Lagrange multipliers that satisfy $\sum_{i=1}^N \alpha_i y_i = 0$ and will be obtained by solving the following quadratic optimization problem:

$$w(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K_{\theta}(x_i, y) \quad (5)$$

The popular kernel functions used in SVM are sigmoid, polynomial, and radial basis function. In this study, polynomial kernel function is applied.

V. RESULTS AND DISCUSSION

To evaluate our proposed approach, we used Leave-One subject-Out (LOO) cross validation. For our database with N subjects, we performed N experiments. For each experiment, the samples of $N-1$ subjects are used for training and the remaining samples for testing. Finally, the true detection rate is estimated as the average classification accuracy on test samples.

SVM has been originally proposed as a binary classifier, and then extended for multi-class problems [20]. For classification of six basic expressions, we have used one-against-all technique that considers 6 binary SVM classifiers to categorize each emotion against all remains. The final recognition will be accomplished by majority voting.

The first experiment was conducted on CK⁺ database using all 40 Gabor filters (5 frequencies and 8 orientations) to test the effect of pre-processing including face alignment. Table1 shows the results. The detection rate of SVM classifier has been improved from 84.43% to 88.99% after face alignment. Based on the results, it can be concluded that Gabor features are not robust to displacement and rotation.

The next experiment was designed to select a subset of Gabor filters which produced more discriminative features. Table 1 gives the GA parameters which have been set in our experiment. The number of selected Gabor filters, length of extracted feature vector, and detection rate of SVM classifier are given in Table 2. We have also included the results without filter selection in this table. The results show that GA is effective for both feature reduction and recognition performance improvement compared to using the entire Gabor filter bank. The classification accuracy increased to 92.97% by using only 8 filters. Indeed, a large number of features does not necessarily result in better performance.

Fig.1 displays the real part of our Gabor bank with 5 frequencies and 8 orientations. The filters indicated by arrows are the selected filters by GA. It is notable that GA has selected most of filters from the first set of filter bank which have the same frequency but different orientations. It can be concluded that the features related to various frequencies are probably correlated with each other and considered as redundant features.

We have also presented the results of LOO cross validation as a confusion matrix in table 4 to analyze the performance of our proposed method on each expression. It can be seen that the detection rate of two expressions - “sadness” and “anger” (68% and 66.66% respectively) are not as high as the other expressions (93.55%, 96.49%, 95.52%, and 96.38% for surprise, happiness, fear, and disgust respectively). This implies that the Gabor feature alone is not able to detect all six basic emotions successfully.

Table 5 shows a comparison of our method with some reported methods (either static or dynamic analysis approaches) on the same database. A brief information of each method including the number of subjects and video sequences, number of emotions categories (include or exclude neutral expression), dynamic or static process, and evaluation measurement are mentioned in this table. It is noted that direct comparison of the results is not fair since there are still differences in the experimental setup, pre-processing approaches, number of sequences used and evaluation methods among the reported results in the various works. However, it can still be useful to compare them in relative sense to understand the strengths and weaknesses of each approach. We can see that our experiment has been applied to the greatest number of subjects and our person-independent results are comparable to the others.

TABLE I. COMPARISON THE RESULTS BEFORE AND AFTER PRE-PROCESSING

method	Detectionrate
Without pre-processing	84.43
After pre-processing	88.99

TABLE II. GA PARAMETERS FOR FILTER SELECTION

Used parameters	value
Population size	10
No. of genes	40
Crossover rate	0.5
Mutation rate	0.01

TABLE III. COMPARISON THE NUMBER OF EXTRACTED FEATURES AND DETECTION RATE BY DIFFERENT GABOR FILTER BANK

method	# filters	# features	Detection rate
Entire filter bank	80	165600	88.99
Selected filter bank	8	33120	92.97

TABLE IV. COFUSION MATRIX FOR SIX BASIC EMOTIONS

Predict label \	Surprise	Happiness	Sadness	Fear	Anger	Disgust
Surprise	93.55	3.22	0	1.61	1.61	0
Happiness	1.75	96.49	0	1.75	0	0
Sadness	8	0	68	8	8	8
Fear	1.49	0	1.49	95.52	0	1.49
Anger	18.52	3.70	3.70	3.70	66.66	3.70
Disgust	1.20	0	1.20	0	1.20	96.38

Figure 1. The real part of a Gabor filter bank. The indicated filters by arrows are selected by GA.

TABLE V. COMPARISON OF DIFFERENT APPROACHES

Method	# Subject	# Sequence	Dynamic	Measure	Recognition Rate (%)
LBP+SVM [21]	96	320	N	10-fold	88.4
Gabor+ AdaSVM [22]	90	313	N	10-fold	86.9
Gabor+ adaboost+ SVM[23]	90	313	N	LOO	93.8
Optical flow+ HMM[24]	97	-	Y	5-fold	90.9
Multistream NN[25]	90	284	Y	-	93.66
Geometrical features+NN [26]	97	375	N	-	93.8
Spatio-temporal LBP[18]	97	375	Y	10-fold	96.26
Proposed (Gabor+GA)	115	321	Y	LOO	92.97

VI. CUNCLUSSION AND FUTURE WORK

In this study, we introduced a method for dynamic analysis of facial expressions from video sequences using Gabor filters. To consider temporal information of a facial emotion, dissimilarity of each sequence was measured and then summed for each video clip. Experimental results show that the approach provided an effective feature compression for each video sequence. To reduce the dimension of the Gabor filters, GA is used instead of computing the statistics of Gabor features. The experimental results showed that the selected Gabor filters outperformed the entire filter bank in the classification detection rate and yet required less processing time for feature extraction. However, our proposed method is not robust to face displacement and rotation. To design an automatic facial expression recognition system, we need to align the faces correctly. However, spontaneous facial expression in natural situation may contain considerable head movement and out-of-plane rotation. Solving this will be our future work.

REFERENCES

- [1] Wang, Y. and L. Guan, *Recognizing Human Emotional State From Audiovisual Signals**. Multimedia, IEEE Transactions on, 2008. **10**(5): p. 936-946.
- [2] Meghiani, M., F. Ferrie, and G. Dudek. *Bimodal information analysis for emotion recognition. in Applications of Computer Vision (WACV), 2009 Workshop on*. 2009.
- [3] Cowie, R., et al., *Emotion recognition in human-computer interaction*. Signal Processing Magazine, IEEE, 2001. **18**(1): p. 32-80.
- [4] Pantic, M. and L.J.M. Rothkrantz, *Automatic analysis of facial expressions: the state of the art*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2000. **22**(12): p. 1424-1445.
- [5] Zhihong, Z., M. Pantic, G.I. Roisman, and T.S. Huang, *A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2009. **31**(1): p. 39-58.
- [6] Pantic, M. and I. Patras, *Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences*. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 2006. **36**(2): p. 433-449.
- [7] Bartlett, M.S., G. Littlewort, M. Frank, C. Lainscsek, I.Fasel, and J. Mvellan, *Recognizing facial expression: machine learning and application to spontaneous behavior. in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. 2005.
- [8] Whitehill, J. and C.W. Omlin. *Haar features for FACS AU recognition. in Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*. 2006.
- [9] Viola, P. and M. Jones. *Rapid object detection using a boosted cascade of simple features. in Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. 2001.
- [10] Kun, L. and Z. Xin. *Facial Expression Recognition from Image Sequences Based on Feature Points and Canonical Correlations. in Artificial Intelligence and Computational Intelligence (AICI), 2010 International Conference on*. 2010.
- [11] Shan, C., S. Gong, and P.W. McOwan, *Facial expression recognition based on Local Binary Patterns: A comprehensive study*. Image and Vision Computing, 2009. **27**(6): p. 803-816.
- [12] ur Rehman, S., L. Li, and L. Haibo. *Lipless Tracking and Emotion Estimation. in Signal-Image Technologies and Internet-Based System, 2007. SITIS '07. Third International IEEE Conference on*. 2007.
- [13] Park, S. and D. Kim, *Subtle facial expression recognition using motion magnification*. Pattern Recognition Letters, 2009. **30**(7): p. 708-716.
- [14] Cohn, J.F., Z. Ambadar, and P. Ekman, *Observer-Based Measurement of Facial Expression with the Facial Action Coding System*. The handbook of emotion elicitation and assessment, 2005. **Oxford University Press Series in Affective Science**.
- [15] Bailenson, J.N., et al., *Real-time classification of evoked emotions using facial feature tracking and physiological responses*. International Journal of Human-Computer Studies, 2008. **66**(5): p. 303-317.
- [16] Tian, Y.I., T. Kanade, and J.F. Cohn, *Recognizing action units for facial expression analysis*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2001. **23**(2): p. 97-115.
- [17] Lucey, P., et al. *The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. in Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. 2010.
- [18] Guoying, Z. and M. Pietikainen, *Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2007. **29**(6): p. 915-928.
- [19] Çomak, E. and A. Arslan, *A new training method for support vector machines: Clustering k-NN support vector machines*. Expert Systems with Applications, 2008. **35**(3): p. 564-568.
- [20] Übeyli, E.D., *Multiclass support vector machines for diagnosis of erythematous-squamous diseases*. Expert Systems with Applications, 2008. **35**(4): p. 1733-1740.
- [21] Caifeng, S., G. Shaogang, and P.W. McOwan. *Robust facial expression recognition using local binary patterns. in Image Processing, 2005. ICIP 2005. IEEE International Conference on*. 2005.
- [22] Bartlett, M.S., et al. *Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction. in Computer Vision and Pattern Recognition Workshop, 2003. CVPRW '03. Conference on*. 2003.
- [23] Littlewort, G., et al. *Dynamics of Facial Expression Extracted Automatically from Video. in Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on*. 2004.
- [24] Yeasin, M., B. Bullot, and R. Sharma. *From facial expression to level of interest: a spatio-temporal approach. in Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. 2004.
- [25] Asteriadis, S., P. Tzouveil, K. Karpouzis, S. Kollias, *Estimation of behavioral user state based on eye gaze and head pose—application in an e-learning environment*. Multimedia Tools and Applications, 2009. **41**(3): p. 469-493.
- [26] Ying-li, T. *Evaluation of Face Resolution for Expression Analysis. in Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on*. 2004.