Facial expression recognition based on ASM and Multi-Instance Boosting

Shaoping Zhu[*]

Department of Information Management, Hunan University of Finance and Economics, 410205, China

Received 1 August 2014, www.cmnt.lv

Abstract

In this paper, a novel method for facial expression recognition in dynamic facial images is proposed, which includes two stages of feature extraction and facial expression recognition. Firstly, Active Shape Model (ASM) is used to extract the local texture feature, and optical flow technique is determined facial velocity information, which is used to charaterize facial expression. Then, fusing the local texture feature and facial velocity information get the hybrid characteristics. Finally, Multi-Instance Boosting model is used to recognize facial expression from video sequences. In order to be learned quickly and complete the recognition, the class label information was used for the learning of the Multi-Instance Boosting model. Experiments were performed in the JAFFE database to evaluate the proposed method. The proposed method shows substantially higher accuracy at facial expression recognition than has been previously achieved and gets a recognition accuracy of 95.3%, which validates its effectiveness and meets the requirements of stable, reliable, high precision and anti-interference ability etc.

Keywords: Facial expression recognition; Active Shape Model; Multi-Instance boosting

1 Introduction

Facial expression is the most expressive way humans display emotions. It delivers rich information about human emotion and provides an important behavioral measure for studies of emotion and social interaction et al. Human facial expression plays an important role in human communications. Facial expressions recognition based on vision closely relates to the study of psychological phenomena and the development of human-computer interaction. It is an important addition to computer vision research at present, and has numerous significant theoretic and practical value. It has been widely applied in human-computer interfaces, human emotion analysis, medical care and cure, public security, and financial security, such as real-time video surveillance, bank cryptography and so on.

Automatically recognizing facial expression has recently become a promising research area. There are many researches already carried out to recognize facial expressions from video sequence. Yeasin, M. et al. [1] trained discrete hidden Markov models (HMMs) to learn the underlying model for each facial expression from video sequences. Morishima and Harashima [2] used emotion space to recognize facial expression. Jabid, T. et al. [3] put forward robust facial expression recognition based on local directional pattern. Aleksic et al. [4] used facial animation parameters and multistream HMMs for automatic facial expression recognition. Edwards G J et al. [5] used the Active Appearance Model (AAM) as a basis for face recognition and obtained good results for difficult images. Zhao and Pietikäinen [6] extended the LBP-TOP features to multi-resolution spatio-temporal space for describing facial expressions and used support vector machine (SVM) classifier to select features for facial expressions recognition. Shih et al. [7] combined 2D-LDA and SVM to recognize facial expressions. Matsugu, M. et al. [8] proposed a rule-based algorithm for robust facial expression recognition combined with robust face detection using a convolutional neural network. Shan et al. [9] empirically evaluated facial representation based on local binary patterns (LBP) for facial expression recognition in 2009.

Facial expression recognition based on vision is an challenging research problem. However, these approaches have been fraught with difficulty because they are often inconsistent with other evidence of facial expressions [10]. It is essential for intelligent and natural human computer interaction to recognize facial expression automatically. In the past several years, significant efforts have been made to identify reliable and valid facial indicators of expressions. Wang L et al. [11] used Active Appearance Models (AAM) to decouple shape and appearance parameters from face images, and used SVM to classify facial expressions. In [12, 13], Prkachin and Solomon validated a Facial Action Coding System (FACS) based measure of pain that could be applied on a frame-by-frame basis. Most must be performed offline, which is both timely and costly, and makes them ill-suited for real-time applications. In [14], Zhang et al. combined the advantages of Active Shape Model (ASM) with Gabor Wavelet to extract efficient facial expressions feature and proposed ASM+GW model for facial expression recognition. Zhang [15] used supervised locality preserving projections (SLPP) to extract facial expression features, and multiple kernels support vector machines (MKSVM) is used for facial expression recognition. Methods described above use static features to characterize facial expression, but these static features cannot fully represent facial expressions.

However, evaluation results and practical experience have shown that facial expression automatically technologies are currently far from mature. Many challenges are to be solved before it can implement a robust practical

[*] ***Corresponding author's*** e-mail: zhushaoping_cz@163.com

application. In this paper, we propose a method for automatically recognizing facial expressions from video sequences. This approach includes extracting facial expression features and classifying facial expressions. In the extracting feature, we use Active Shape Model (ASM) for facial local texture feature and motion descriptor based on optical flow for facial velocity features. Then, the two features are integrated, and facial expression is represented by hybrid feature. Final, the Multi-Instance boosting model is used for facial expression recognition. In addition, in order to improve the recognition accuracy, the class label information is used for the learning of the Multi-Instance boosting model.

Given unlabeled facial video sequence, our goal is to automatically learn different classes of facial expressions present in the data, and apply the Multi-Instance boosting model for facial expressions categorization and recognition in the new video sequences.

The paper is structured as follows. After reviewing related work in this section, we describe the facial expression feature representation base on ASM model, optical flow technique in section 2. Section 3 gives details of Multi-Instance boosting model for recognize facial expression. Section 4 shows experiment result, also comparing our approach with two state-of-the-art methods, and the conclusions are given in the final section.

2 Facial expression representation

Automatic facial expression recognition has many potential applications in different areas of human computer interaction. However, they are not yet fully realized due to the lack of an effective facial feature descriptor. Deriving an effective facial representation from original face images is a vital step for successful facial expression recognition. Due to great changes in the dynamic characteristics and lack of constraints, this paper proposes ASM model for facial local texture feature extraction and optical flow model for facial velocity features, which can describe facial expression effectively.

2.1 LOCAL TEXTURE FEATURE EXTRACTION

Active Shape Model (ASM) [16] can iteratively adapt to refine estimates of the pose, scale and shape of models of image objects, and is a powerful method for refining estimates of object shape and location. González-Jiménez, D., & Alba-Castro, J. L. [17] built Point Distribution Models (PDM) to derive from sets of training examples and identified the parameters. This model represents objects as sets of labelled points. An initial estimate of the location of the model points in an image is improved by attempting to move each point to a better position nearby. Adjustments to the pose variables and shape parameters are calculated. Limits are placed on the shape parameters ensuring that the example can only deform into shapes conforming to global constraints imposed by the training set. An iterative procedure deforms the model example to find the best fit to the image object. The steps can be briefly described as follows:

First, given a calibration facial key feature point training set X .

$$X = \{(I_i, s_i) \,|\, i = 1, 2, \cdots, \gamma; s_i = (x_i^1, y_i^1, \cdots, x_i^\beta, y_i^\beta)^T\}, \qquad (1)$$

where γ is the number of training sample, β is the number of predefined key feature points, s_i is Shape vector in training set X , which is concatenated by predefined and manual calibration β key feature points of horizontal ordinate on the training images I_i .

Then, all shapes in training set are aligned to the same coordinate system by shape alignment algorithm, and get feature set X' after alignment.

$$X' = \{s_i' \,|\, i = 1, 2, \cdots, \gamma\}, \qquad (2)$$

these alignment shape are analysized by PCA, get the active shape model as follow:

$$X = \bar{X} + p_s b_s, \qquad (3)$$

where $\bar{X}$ is the average shape, b_s for the shape parameter, p_s is eigenvector of main component of transformation matrix, which is obtained by the training set of the eigenvalues of the covariance matrix decomposition. Eigenvector of main component reflects the main mode of shape change. Any shape can be approximately expressed by the average shape deformation, which is modeled by the shape parameter b_s to sum several model weighted. $-3\sqrt{\lambda_i} < b_i < 3\sqrt{\lambda_i}$, $i = 1, 2, \cdots, \gamma$, where λ_k is the eigenvalues of the covariance matrix, $\lambda_k \ge \lambda_{k+1}, \lambda_k \ne 0, k = 1, 2, \cdots, 2l$.

Finally, we calculate markov distance to determine the best matching position by analyzing the gray information of neighborhood.

Given local texture model:

$$\bar{G}_{ij} = \frac{1}{\gamma} \sum_{i=1}^{\gamma} G_{ij}, \qquad (4)$$

$$T_{G_{ij}} = \frac{1}{\gamma - 1} \sum_{i=1}^{\gamma} (G_{ij} - \bar{G}_{ij})(G_{ij} - \bar{G}_{ij})^T, \qquad (5)$$

where $\bar{G}_{ij}$ is the average texture, G_{ij} is the texture vector after the gray level information of j-th fixed point normalizes in the i-th training image.

$$G_{ij} = \frac{1}{\sum_{j=1}^{2k+1} \left| d_{g_{ij}} \right|} d_{g_{ij}}, \qquad (6)$$

where $d_{g_{ij}} = [g_{ij,2} - g_{ij,1}, \cdots, g_{ij,2k+1} - g_{ij,2k}]$, g_{ij} is the gray information of the i-th feature points, which is the gray level of k points from each up and down along the normal direction with feature points as the center, $T_{G_{ij}}$ is the covariance matrix.

Calculate markov distance as follow:

$$d(G_{ij}') = (G_{ij}' - \bar{G}_{ij})^T (T_{G_{ij}})^{-1} (G_{ij}' - \bar{G}_{ij}), \qquad (7)$$

where G_{ij}' is the normalized vector texture by sampling near the j-th point of target search images.When $d(G_{ij}')$ takes minimum value, the corresponding point is the best candidate.

2.2 FACIAL VELOCITY FEATURE EXTRACTION

Emotion expression is far more varied. Optical flow-based facial expression representation has attracted much attention

[18]. According to the physiology, the expression is a dynamic event, it must be represented by the motion information of a face. So we use facial velocity features to characterize facial expression. The facial velocity features (optical flow vector) are estimated by optical flow model, and each facial expression is coded on a seven level intensity dimension (A–G): “anger”, “disgust”, “fear”, “happiness”, “neutral”, “sadness” and “surprise”.

Given a stabilized video sequence in which the human face appears in the center of the field of view, we compute the facial velocity (optical flow vector) $\upsilon = (\upsilon_x, \upsilon_y)$ at each frame using optical flow equation, which is expressed as:

$$I_x\upsilon_x + I_y\upsilon_y + I_t = 0, \tag{8}$$

where $I_x = \frac{\partial I}{\partial x}, I_y = \frac{\partial I}{\partial y}, I_t = \frac{\partial I}{\partial t}$, $\upsilon_x = \frac{dx}{dt}, \upsilon_y = \frac{dy}{dt}$, (x, y, t) is the image in pixel (x, y) at time *t*, where $I(x, y, t)$ is the intensity at pixel (x, y) at time t , υ_x , υ_y is the horizontal and vertical velocities in pixel (x, y) . We can obtain $\upsilon = (\upsilon_x, \upsilon_y)$ by minimizing the objective function:

$$C = \int_D \left[\lambda^2 \|\nabla\upsilon\|^2 + (\nabla I \cdot \upsilon + I_t)^2\right] dxdy, \tag{9}$$

where there are many methods to solve the optical flow equation. We use the iterative algorithm [19] to compute the optical flow velocity：

$$\begin{aligned} \upsilon_x^{k+1} &= \bar{\upsilon}_x^k - \frac{I_x\left[I_x\bar{\upsilon}_x^k + I_y\bar{\upsilon}_y^k + I_t\right]}{\lambda + I_x^2 + I_y^2} \\ \upsilon_y^{k+1} &= \bar{\upsilon}_y^k - \frac{I_y\left[I_x\bar{\upsilon}_x^k + I_y\bar{\upsilon}_y^k + I_t\right]}{\lambda + I_x^2 + I_y^2} \end{aligned}, \tag{10}$$

where k is the number of iterations, initial value of velocity $\upsilon_x^0 = \upsilon_y^0 = 0$, $\bar{\upsilon}_x^k, \bar{\upsilon}_y^k$ is the average velocity of the neighborhood of point (x, y) .

The optical flow vector field υ is then split into two scalar fields υ_x and υ_y corresponding to the x and y components of υ . υ_x and υ_y are further half-wave rectified into four none-gative channels ${\upsilon_x}^+$, ${\upsilon_x}^-$, ${\upsilon_y}^+$, ${\upsilon_y}^-$ so that $\upsilon_x = {\upsilon_x}^+ - {\upsilon_x}^-$ and $\upsilon_y = {\upsilon_y}^+ - {\upsilon_y}^-$ These four nonnegative channels are then blurred with a Gaussian kernel and normalized to obtain the final four channels $\upsilon {b_x}^+$, $\upsilon {b_x}^-$, $\upsilon {b_y}^+$, $\upsilon {b_y}^-$.

Facial expression is represented by facial velocity features that are composed of the channels $\upsilon {b_x}^+$, $\upsilon {b_x}^-$, $\upsilon {b_y}^+, \upsilon {b_y}^-$ of all pixels in facial image. Facial expression can be regard as facial motion which are important characteristic features of facial expression, in addition to, the velocity features have been shown to perform reliably with noisy image sequences, and has been applied in various tasks, such as action classification, motion synthesis, etc.

2.3 FACIAL EXPRESSION HYBRID FEATURE REPRESENTATION

In order to improve the accuracy of facial expression recognition, fusing local texture feature vector and optical flow feature vector become a hybrid feature vector, which is expressed as:

$$X_{ij} = [d, \upsilon], \tag{11}$$

where X_{ij} is the hybrid feature vector of each frame image, d is local optical flow vector, υ is contour vector.

Facial expressions are represented by the hybrid feature vectors of local texture feature vector and optical flow vector. Because facial expressions can be regard as motion, the local optical flow features can describe facial expression effectively, in addition to, local texture features can describe the shape of facial movement information simply and visually. Thus, the hybrid features have been shown to perform reliably with noisy image sequences, and have been applied in various tasks, such as expressions classification, face recognition, etc.

3 Multi-Instance boosting for facial expression recognition

After characterizing human facial expression, there are many methods to recognize human facial expression. Because human facial expression recognition can regard as a Multiple Instance problem, we use the Multi-Instance boosting algorithm to learn and recognize human facial expression. The Multi-Instance boosting model has been applied to various computer vision applications, such as object recognition, action recognition, human detection, etc. The Multi-Instance boosting framework is used to learn a unified classifier instead of individual classifiers for all classes in order to increase recognition efficiency without compromising accuracy. Our approach is directly inspired by a body of work on using generative Multi-Instance boosting models for visual recognition based on the “bag-of-words” paradigm. We propose a novel Multi-Instance boosting framework, which learns a unified classifier instead of individual classifiers for all classes, so that the recognition efficiency can be increased without compromising accuracy.

3.1 DEFINITION OF MULTI-INSTANCE PROBLEM

Keeler, et. al [20] proposed originally the idea for the multi-instance learning for handwritten digit recognition in 1990. It was called Integrated Segmentation and Recognition (ISR), and it is the key idea to provide a different way in constituting training samples. Dietterich et al.[21, 22] proposed Multiple-Instance framework for the prediction of drug molecule activity. Multiple-Instance Learning has widely used in image classification [23, 24], human face detection [25], etc. Definition of multi-instance is as follows:

Assume the set of class labels $C_i \in \{0,1\}$, χ is the instance space, multi-instance data set $D = \{X_i, C_i\}_{i=1}^N$. The instances are defined as $\{x_k \,|\, k = 1, 2, \cdots, \tau\}$ in multi-instance data set D . All the instances in the positive bags and negative bags is defined as D^+ and D^- respectively, where $D^+ = \{{x_i}^+ \,|\, i = 1, 2, \cdots, \tau\}$, $D^- = \{{x_j}^- \,|\, j = 1, 2, \cdots, \varsigma\}$.

Given a bag X_i , X_i is a positive bag if at least one of its instances is positive; otherwise, X_i is a negative bag.

The multi-instance problem is a function. Its goal of multi-instance is to learn a classifier based on instance.

$f(x_i): x_i \to h$, $x_i \in \chi$ or a classifier based on bags.

$F(X_i): X_i \to h$, that correctly predicts the unlabeled bag.

Given two bags X_i and X_k $(i \neq k)$, $X_i \bigcap X_k \neq \phi$.

In the multi-instance problem, each instance only belongs to one specific bag. Namely, two different bags cannot share the same instance.

3.2 BOOSTING ALGORITHM ANALYSIZE

Boosting algorithm is a mathematical model based on a fuzzy rule system. A fuzzy rule system is defined as follows by using the classical case at the beginning. In the classical case, a rule is a function formulated with arguments coupled by logical operators, yielding a logical expression and a corresponding response. It is a semi-supervised learning method [26]. The steps of boosting algorithm can be briefly described as follows.

Assuming the training sample set $\{(x^1, c_1), (x^2, c_2), \cdots, (x^n, c_n)\}$, $c_n \in \{c_1, \cdots, c_l\}$

Give equal initial weights of each sample: $\omega^i = 1/n$, the training sample set trains for κ rounds of training and obtains κ fuzzy classification rules.

For $t = 1, 2, \cdots, \kappa$ Do

Under the current sample distribution, calculated the corresponding weights of fuzzy rules as follows:

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1 - E(R_t)}{E(R_{t)}}\right), \tag{12}$$

Update the sample weight as follows:

$$\omega^i(t+1) = \frac{\omega^i(t)}{z_t} \times \begin{cases} e^{-\alpha_t \mu_{rt}(x^i)} & c_i = c_t \\ e^{\alpha_t \mu_{R_t}(x^i)} & c_i \neq c_j \end{cases}, \tag{13}$$

The category is obtained by the fuzzy classifier as follows:

$$C_{\max}(x^k) = \arg\max_{C_m} \sum_t \alpha_t \sum_{R_i / c_t = C_k} \mu_{R_t}(x^k), \tag{14}$$

where x^k is unknown sample, $x^k = \left\{x_1^k, x_2^k, \cdots, x_N^k\right\}$.

3.3 MULTI-INSTANCE BOOSTING FOR EXPRESSION RECOGNITION

To improve the recognition efficiency, we combine Multiple-Instance and boosting to build Multi-Instance boosting model. Multi-Instance boosting is one of the most efficient machine learning algorithms. In Multi-Instance boosting, training samples are not singletons, at the same time they are in "bags", where all of the samples in a bag share a label [27]; Samples are organized into positive bags of instances and negative bags of instances, where each bag may contain a number of instances [28]. At least one instance is positive (i.e. object) in a positive bag, while all instances are negative (i.e. non-object) in a negative bag. In Multi-Instance boosting [29], learning must simultaneously learn that samples in the positive bags are positive along with the parameters of the classifier. To obtain training samples, each image is divided into $L \times L$ blocks. We treat each block in a image as a single word w_j and a image as a bag. Each block is used as an example for the purposes of training. It is suitable to represent the object by a bag of multiple instances (non-aligned human face images). Then, Multi-Instance boosting can learn that instances in the positive bags are positive, along with a binary classifier [30]. In this paper, Multi-Instance boosting is used for facial expression with non-aligned training samples. The Multi-Instance boosting for facial expression recognition proceeds as follows:

Input: Given dataset $\{X_i, C_i\}_{i=1}^N$, X_i is training bags, where $X_i = \{x_{i1}, x_{i2}, \cdots, x_{ij}, \cdots, x_{iN}\}$, C_i is the score of the sample, and $C_i \in \{0,1\}$. n is the number of all weak classifiers. A positive bag contains at least one positive sample, and $C_i = \max(C_{ij})$.

Pick out K weak classifiers and consist of strong classifier.

Step 1: Update all weak classifiers in the pool with data $\{x_{ij}, C_i\}$.

Step 2: Initialize all strong classifier: $H_{ij} = 0$ for all i, j.

Step 3: Calculate the probability that the *j*-th sample is positive in the *i*-th bag as follows:

For $k = 1, 2, \cdots, K$ do

For $m = 1, 2, \cdots, N$ do

$$P_{ij}^m = \sigma(H_{ij} + h_m(x_{ij})), \tag{15}$$

where $P_{ij}^m = p(C_i | x_{ij}) = \dfrac{1}{1 + \exp(-c_{ij})}$.

We calculate the probability that the bag is positive as follow:

$$P_i^m = 1 - \prod\nolimits_j (1 - p_{ij}^m), \tag{16}$$

where $P_i^m = \mathrm{p}(C_i \mid X_i)$.

The likelihood assigned to a set of training bags is:

$$C^m = \sum\nolimits_i (C_i \log p_i^m + (1 - C_i)\log(1 - p_i^m)). \tag{17}$$

End for

Finding the maximum *m** from *N* as the current optimal weak classifier as follow:

$$m^* = \arg\min_m C^m, \tag{18}$$

The *m** come into the strong classifier:

$$h_k(x) \leftarrow h_{m*}(x), \tag{19}$$

$$H_{ij} = H_{ij} + h_k(x), \tag{20}$$

End for

Step 4: Output: Strong classifier which consist of weak classifiers as follow:

$$H(x) = \sum\nolimits_k h_k(x), \tag{21}$$

where h_k is a weak classifier and can make binary predictions using $sign(H_K(x))$.

In Multi-Instance boosting, samples come into positive bags of instances and negative bags of instances. Each instance x_{ij} is indexed with two indices, where *i* for the bag and *j* for the instance within the bag. All instances in a bag share a bag label C_i. Weight of each sample composes of the weight of the bag and the weight of the sample in the bag.

The quantity of the samples can be interpreted as a likelihood ratio, where some (at least one) instance is positive in a bag. P_{ij}^{m} is the probability that some instance is positive. So the weight of samples in the bags is P_{ij}^{m}. We calculate: $w_{ij} = \frac{\partial \log C^{m}}{\partial y_{ij}}$, and get weight of the bags w_{ij}.

Training in the initial stages is the key to a fast and effective classifier. The result of the Multi-Instance boosting learning process is not only a sample classifier but also weights of the samples. The samples have high score in positive bags which are assigned high weight. The final classifier labels these samples to be positive. The remaining samples have a low score in the positive bags, which are assigned a low weight. The final classifier classifies these samples as negative samples as they should be. We train a complete Multi-Instance boosting classifier to achieve the desired false positive rates and false negative rates. Retrain the initial weak classifier so that a zero false negative rate is obtained on the samples, which label positive by the full classifier. This results in a significant increase in many samples to be pruned by the classifier. Repeating the process so that the second classifier is trained to yield a zero false negative rate on the remaining samples.

For the task of facial expression recognition, our goal is to classify a new face image to a specific facial expression class. During the inference stage, given a testing face image, we can treat each aspect in the Multi-Instance boosting model as one class of facial expression. For facial expression recognition with large amount of training data, this will result in long training time. In this paper, we adopt a supervised Algorithm to train Multi-Instance boosting model. The supervised training algorithm not only makes the training more efficient, but also improves the overall recognition accuracy significantly. Each image has a class labeled information in the training images, which is important for the classification task. Here, we make use of this class label information in the training images for the learning of the Multi-Instance boosting model, since each image directly corresponds to a certain facial expression class on train sets.

4 Experimental results and analysis

We studied facial expression feature representation and facial expression classification schemes to recognize seven different facial expressions, such as “anger”, “disgust”, “fear”, “happiness”, “neutral”, “sadness” and “surprise” in the JAFFE database. We verified the effectiveness of our proposed algorithm using C++ and Matlab7.0 hybrid implementation on a PC with Intel CORE i5 3.2 GHz processor and 4G RAM.

JAFFE data set [31] is the most available video sequence dataset of human facial expression. In this database, there are seven groups of images by 10 Japanese women and a total of 213 images, which are “anger”, “disgust”, “fear”, “happiness”, “neutral”, “sadness” and “surprise” respectively. The size of each image is 256×256 pixels in the JAFFE database. Each face image was normalized to a size of 8×8. Some sample images are shown in figure 1.

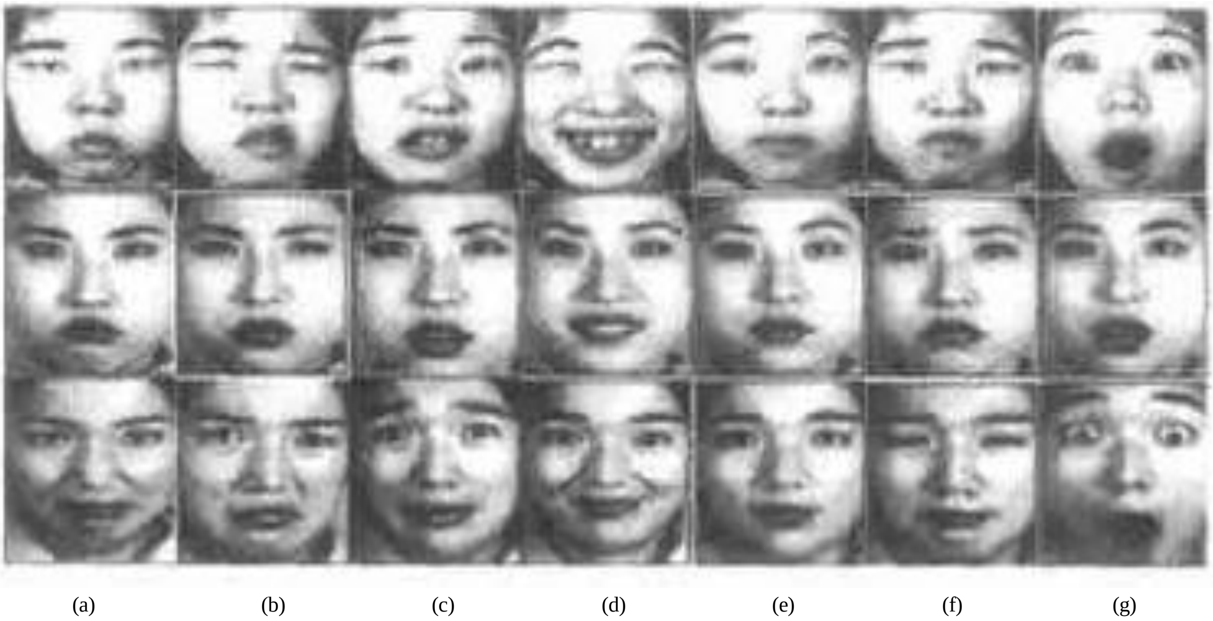


FIGURE 1 Example of seven facial expression images in JAFFE. (a) “anger”, (b) “disgust”, (c) “fear”, (d) “happiness”, (e) “neutral”, (f) “sadness”, (g) “surprise”

In experiments, we chose 30 face images per class randomly for training, 20 face images for testing in JAFFE. These images were pre-processed by aligning and scaling, thus the distances between the eyes were the same for all images, and ensured that the eyes occurred in the same coordinates of the image. The system was run seven times, and we obtained seven different training and testing sample sets. The recognition rates were obtained by average recognition rate of each run.

In order to examine the effectiveness of our proposed facial expressions recognition approach, we used 150 different face images for this experiment. Some images contained the same person but in different expressions. There are seven facial expressions. They are “anger”, “disgust”, “fear”, “happiness”, “neutral”, “sadness” and “surprise” respectively. In experiments, we chose 150 samples per facial expression. The recognition results for per-video classification are presented in the table 1.

TABLE 1 Facial expression recognition results with different samples

facial expressions	number of samples	correct recognition number	correct recognition rate(%)
anger	150	143	95.3
disgust	150	146	97.3
fear	150	142	94.6
happiness	150	145	96.7
neutral	150	136	90.7
sadness	150	140	93.3
surprise	150	148	98.7

We can see that the algorithm correctly classifies most facial expressions, where “surprise” obtains recognition accuracy of 98.7%, the recognition rate of “disgust” is to 97.3%. The recognition accuracy of “happiness” is to 96.7%. Average recognition rate gets to 95.3%. Face image for some expression changes are minor, some image expressions are compound expressions, So it is difficult to identify accurately. Most of the mistakes are confusions between “anger” and “sadness”, between “happiness” and “neutral”, between “fear” and “surprise”. It is intuitively reasonable that they are similar facial expressions.

To examine the accuracy of our proposed facial expression recognition approach, we compared our method with two state-of-the-art approaches for facial expression recognition using the same data. The first method is “AAM+SVM”, which used Active Appearance Models (AAM) to extract face features, and SVM to classify facial expression. The second method is “ASM+GW”, which used Active Shape Model and Gabor Wavelet (ASM+GW) for facial expression recognition. The results of recognition accuracy comparison are shown in figure 2.

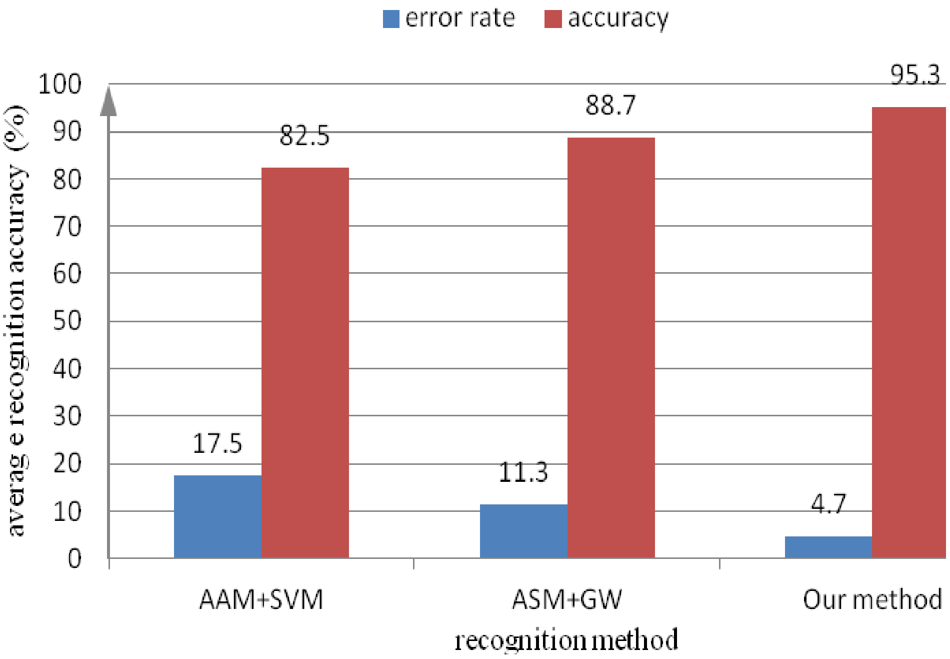


FIGURE 2 Recognition accuracy comparison of different method

In figure 2, we can see that AAM+SVM obtains average recognition accuracy of 82.5%. The average recognition rate of ASM+GW is to 88.7%. Our method is stabled to average recognition accuracy of 95.3%. Our method has higher recognition accuracy, lower error rate and performs significantly better than the above two state-of-the-art approaches for facial expression recognition. Because we improved the recognition accuracy in the two stages of facial expression features extraction and facial expression recognition. In the stage of facial expression feature extraction, we used local texture feature and motion features that were reliably with noisy image sequences and hybrid feature to describe facial expression effectively. In the stage of expression recognition, we used Multi-Instance boosting algorithm to classify facial expression images. Our method performs the best, its recognition accuracies are satisfactory.

To examine the speed of our proposed facial expression recognition approach, we compared our method with three state-of-the-art approaches for facial expression recognition using the same data. 200 different expression images were used for this experiment, where some images contained the same person but in different facial expression. The results of detection speed comparison are shown in Table 2.

TABLE 2 Recognition accuracy comparison of different method

method	Number of samples	Recognition time(s)
AAM+SVM	200	2.87
ASM+GW	200	2.62
SLPP+ MKSVM	200	2.18
Our method	200	1.65

In table 2, we can see that AAM+SVM takes 2.87s, The time of ASM+GW is to 2.62s, SLPP+ MKSVM need 2.18s. Our method is only 1.65s. Our method has faster recognition speed than the above three state-of-the-art approaches for facial expression recognition. Because we improved facial expression feature extraction, reduces the dimensions of the feature, and improve the speed of recognition.

5 Conclusion

Facial expression recognition can provide significant advantage in public security, financial security, drug-activity prediction, image retrieval, face detection, etc. In this paper, we have presented a novel method to recognize the facial expression and given the seven facial expression levels at the same time. The main contribution can be concluded as follows:

(1) ASM model was used for local texture features extraction. Optical flow model was used to extract facial velocity features, then after fusing local texture features and facial velocity features, we got hybrid features and to be used for facial expression representation.

(2) Multi-Instance boosting model was used for facial expression recognition. In our models, Multi-Instance and boosting were used to create Multi-Instance boosting. We proposed a new Multi-Instance boosting framework, which recognized different facial expression categories. In addition, in order to improve the recognition accuracy, the class label information was used for the learning of the Multi-Instance boosting model.

(3) Experiments were performed on a facial expression dataset in JAFFE and evaluated the proposed method. Experimental results reveal that the proposed method significantly improves the recognition accuracy, speed and performs better than previous ones.

Acknowledgments

This work was supported by Research Foundation for Science & Technology Office of Hunan Province under Grant (No.2014FJ3057), by Hunan Provincial Education Science and “Twelve Five” planning issues (No. XJK012CGD022), by the Teaching Reform Foundation of Hunan Province Ordinary College under Grant (No. 2012401544) and by the Foundation for Key Constructive Discipline of Hunan Province.

References

[1] Yeasin M, Bullot B and Sharma R. 2006, Recognition of facial expressions and measurement of levels of interest from video, *IEEE Transactions on Multimedia*, **8**(3): 500-508.

[2] Morishima S and Harashima H. 1993, Emotion space for analysis and synthesis of facial expression, *Proc. 2nd IEEE Int. Workshop on Robot and Human Communication*, pp. 188-193.

[3] Jabid T, Kabir M H, Chae O 2010 Robust facial expression recognition based on local directional pattern, *ETRI Journal*, 32(5): 784-794

[4] Aleksic P S and Katsaggelos A K. 2006, Automatic facial expression recognition using facial animation parameters and multistream HMMs. *IEEE Transactions on Information Forensics and Security*, 1(1): 3-11.

[5] Edwards G J, Cootes T F and Taylor C J. 1998, Face recognition using active appearance models, *Proc. Computer Vision—ECCV'98*, Springer Berlin Heidelberg, pp. 581-595.

[6] Zhao G and Pietikäinen M. 2009, Boosted multi-resolution spatiotemporal descriptors for facial expression recognition, *Pattern recognition letters*, 30(12): 1117-1127.

[7] Shih F Y, Chuang C F and Wang P S P. 2008, Performance comparisons of facial expression recognition in JAFFE database, *Int. J. Pattern Recognition and Artificial Intelligence*, 22(03): 445-459.

[8] Matsugu M, Mori K, Mitari Y and Kaneda Y. 2003, Subject independent facial expression recognition with robust face detection using a convolutional neural network, *Neural Networks* **16**(5) 555-559

[9] Shan C, Gong S and McOwan P W. 2009, Facial expression recognition based on local binary patterns: A comprehensive study, *Image and Vision Computing*, 27(6): 803-816.

[10] Turk D C, Dennis C and Melzack R. 2001, The measurement of pain and the assessment of people experiencing pain, *Handbook of Pain Assessment*, ed Turk D C and Melzack R, New York: Guilford, 2nd edition: pp. 1-11.

[11] Wang L, Li R F, and Wang K. 2014, A novel automatic facial expression recognition method based on AAM, *Journal of Computers*, 9(3): 608-617.

[12] Prkachin K M. 1992, The consistency of facial expressions of pain: a comparison across modalities, *Pain*, **3**(5): 297-306.

[13] Prkachin K M and Solomon P E. 2008, The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain, *Pain*, **2**(139): 267-274.

[14] Zhang S J, Jiang B and Wang T. 2010, Facial expression recognition algorithm based on active shape model and gabor wavelet, *Journal of Henan University (Natural Science)*, **40**(5): 521-524.

[15] Zhang W and Xia L M. 2011, Pain expression recognition based on SLPP and MKSVM, *Int. J. Engineering and Manufacturing*, **3**: 69-74.

[16] Cootes T F, Taylor C J, Cooper D H and Graham J. 1995, Active shape models-their training and application, *Computer vision and image understanding*, **61**(1): 38-59.

[17] González-Jiménez D and Alba-Castro J L. 2007, Toward pose-invariant 2-d face recognition through point distribution models and facial symmetry, *IEEE Transactions on Information Forensics and Security*, **2**(3): 413-429.

[18] Cohn J F, Zlochower A J, Lien J J and Kanade T. 1998, April, Feature-point tracking by optical flow discriminates subtle differences in facial expression. *Proc. Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 396-401.

[19] Bouguet J Y. 2001, Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm, *Intel Corporation*, **5**: 1-10.

[20] Keeler J D, Rumelhart D E and Leow W K. 1990, Integrated segmentation and recognition of hand-printed numerals, *1990 NIPS-3: Proc. Conf. on Advances in neural information processing systems 3*, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc 557–563

[21] Dietterich T G, Lathrop R H and Lozano-Perez T. 1997, Solving the multiple instance problem with axis-parallel rectangles, *Artificial Intelligence*, **89**: 31-71.

[22] Zafra A, Pechenizkiy M, and Ventura S. 2012, Relief-MI: an extension of relief to multiple instance learning, *Neurocomputing*, **75**: 210-218.

[23] Zha Z J, Hua X S, Mei T, Wang J, Qi G J and Wang Z. 2008, June, Joint multi-label multi-instance learning for image classification, *Proc. 2008 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR 2008, pp. 1-8.

[24] Song X F, Jiao L C, Yang S Y, Zhang X R, and Shang F H. 2013, Sparse coding and classifier ensemble based multi-instance learning for image categorization, *Signal Processing*, **93**, 1-11.

[25] Zhang C, Platt J C and Viola P A. 2005, Multiple instance boosting for object detection, *In Advances in neural information processing systems*, pp. 1417-1424.

[26] Zhu X and Goldberg A B. 2009, Introduction to semi-supervised learning, *Synthesis lectures on artificial intelligence and machine*

learning, **3**(1): 1-130.

[27] Kumar J, Pillai J and Doermann D. 2011, Document Image Classification and Labeling using Multiple Instance Learning, Proc. *2011 International Conference on Document Analysis and Recognition (ICDAR)*, IEEE, pp. 1059–1063.

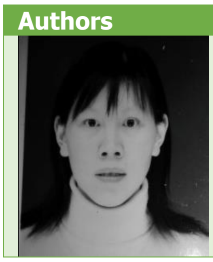
[28] Andrews S and Hofmann T. 2004, Multiple instance learning via disjunctive programming boosting, *Advances in Neural Information Processing Systems*, **16**: 65-72.

[29] Song X, Jiao L C, Yang S, Zhang X and Shang F. 2013, Sparse coding and classifier ensemble based multi-instance learning for image categorization, *Signal Processing*, **93**(1): 1-11.

[30] Yakhnenko O, Honavar V. 2011, Multi-Instance multi-label learning for image classification with large vocabularies, *BMVC*, pp.1-12

[31] Cheng F, Yu J, Xiong H. 2010, Facial expression recognition in JAFFE dataset based on Gaussian process classification, *IEEE Transactions on Neural Networks*, **21**(10): 1685-1690.

Authors



Shaoping Zhu, P.R.China

Current position, grade: Associate Professor of the Department of Information Management, Hunan University of Finance and Economics, in China.

Scientific interest: image and signal processing, computer vision, artificial intelligence and pattern recognition, in particular, high-level recognition problems in computer vision, human facial expression recognition, human activity recognition, object and scene recognition.