www.s2is.org
International Journal on Smart Sensing and Intelligent Systems
ISSN 1178-5608

AUTOMATIC RECOGNITION OF FACIAL EXPRESSION BASED ON COMPUTER VISION

Shaoping Zhu
Department of Information Management Hunan University of Finance and Economics, 410205
Changsha, China
Email: zhushaoping_cz@163.com

Submitted: Apr. 15, 2015 Accepted: Aug. 1, 2015 Published: Sep. 1, 2015

Abstract Automatic facial expression recognition from video sequence is an essential research area in the field of computer vision. In this paper, a novel method for recognition facial expressions is proposed, which includes two stages of facial expression feature extraction and facial expression recognition. Firstly, in order to exact robust facial expression features, we use Active Appearance Model (AAM) to extract the global texture feature and optical flow technique to characterize facial expression which is determined facial velocity information. Then, these two features are integrated and converted to visual words using "bag-of-words" models, and facial expression is represented by a number of visual words. Secondly, the Latent Dirichlet Allocation (LDA) model are utilized to classify different facial expressions such as "anger", "disgust", "fear", "happiness", "neutral", "sadness", and "surprise". The experimental results show that our proposed method not only performs stably and robustly and improves the recognition rate efficiently, but also needs the least dimension when achieves the highest recognition rate , which demonstrates that our proposed method is superior to others.

Index terms*:* Facial expression recognition, Active Appearance Model (AAM), Bag of Words model, LDA model, computer vision.

I. INTRODUCTION

Automatic facial expression recognition is an interesting and challenging problem, and impacts important applications in many areas such as human computer interaction. Human Computer Interaction moves forward in the field of computer vision. Computer vision has higher precision and processing speed than the human eye. It is an important research topic and has made great progress in the past decades. Archana S. Ghotkar and Dr. Gajanan K. Kharate [1] have discussed vision based hand gesture recognition system as hand plays vital communication mode. Yongqing Wang and Yanzhou Zhang [2] presented an improved SVDD algorithm to handle object tracking based on machine vision efficiently.

Computer vision technology has been widely used. For example, it has been successfully applied in the field of industrial detection and greatly improves the quality and reliability of the product. In the field of biomedicine, it is used to assist doctors in medical images analysis. Facial expressions recognition based on vision closely relates to the study of psychological phenomena and the development of human-computer interaction. Such a research has both significant theoretic values and wide potential applications. As a scientific issue, facial expressions recognition is a typical pattern analysis, understanding and classification problem, closely related to many disciplines such as Pattern Recognition, Computer Vision, Intelligent Human-Computer Interaction, Computer Graphics, and Cognitive Psychology etc. Its research achievements would greatly contribute to the development of these disciplines. While as one of the key technologies in Biometrics, facial expressions recognition techniques are believed having a great deal of potential applications in public security, law enforcement, information security, and financial security.

After more than 30 years' development, automatically recognizing facial expression has made great progress especially in the past ten years. There are many researches already carried out to recognize facial expressions from video sequence. F. Samaria and S. Young [3] proposed a new architecture of hidden Markov models (HMMs) for representing the statistics of facial images from video sequences. C. Shan and S. Gong et al. [4] used Boosted-LBP to extract the most discriminant LBP features for person-independent facial expression recognition in compressed low-resolution video sequences captured in real-world environments. X. Feng et al. [5] divided automatically the face area into small regions, extracted the local binary pattern (LBP) histograms, and concatenated into a single feature histogram for representing facial expression. Xiang et al. [6] employed Fourier transform to extract features for an expression representation and used the fuzzy C means computation to generate a spatiotemporal model for each expression type. Neggaz et al. [7] proposed the improved Active Appearance Model (AAM) to recognize facial expressions of an image frame sequence. Zhao and Pietikäinen [8] extended the LBP-TOP features to multi-resolution spatiotemporal space for describing facial expressions and used support vector machine (SVM) classifier to select features for facial expressions recognition. K. Yu and Z. Wang et al. [9] present a Support Vector Machine (SVM) based active learning approach for facial expressions recognition. Siyao Fu et al. [10] proposed a spiking neural network based cortex-like mechanism and application for facial expression recognition. Dailey et al. [11] used a simple biologically plausible neural network model to train and classify facial expressions.

Facial expression recognition based on vision is a challenging research problem. However, these approaches have been fraught with difficulty because they are often inconsistent with other evidence of facial expressions [12]. 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. L. Wang et al. [13] used Active Appearance Models (AAM) to decouple shape and appearance parameters from face images, and used SVM to classify facial expressions. In [14, 15], 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 [16], 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 [17] 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. A great number of challenges are to be solved before one can implement a robust practical application. The following key issues are especially pivotal: (1) the accurate facial feature location problem, which is the prerequisite for sequent feature exaction and classification; (2) efficient face representation and corresponding classifier with high accuracy; (3) how to improve the robustness of facial expression recognition to inevitable misalignment of the facial feature.

In this paper, we propose a method for automatically recognizing facial expressions from video sequences. This approach includes two steps. One step is facial expression features extraction; the other step is facial expressions classification. In the extracting feature, we use Active Shape Model (AAM) for facial global texture feature and motion descriptor based on optical flow for facial velocity features. Then, these two features are integrated and converted to visual words using "bag-of-words" models, and facial expression is represented by a number of visual words. Final, the LDA 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 LDA model.

Input unlabeled facial video sequence, our goal is to automatically learn different classes of facial expressions present in the data, and apply the LDA model for facial expressions categorization and recognition in the new video sequences. Our approach is illustrated in figure 1.

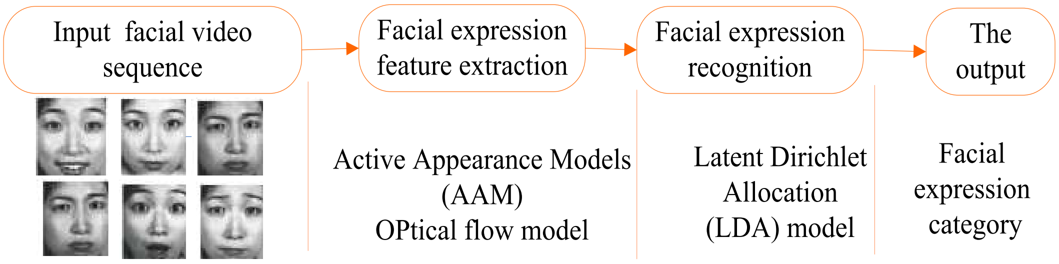

Input facial video sequence
Facial expression feature extraction
Facial expression recognition
The output
Active Appearance Models (AAM)
OPtical flow model
Latent Dirichlet Allocation (LDA) model
Facial expression category

FIGURE 1. Flowchart of facial expression recognition

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

II. FACIAL EXPRESSION FEATURE EXTRACTION AND REPRESENTATION

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. 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 AAM model for facial global texture feature extraction and optical flow model for facial velocity features, which can describe facial expression effectively.

a. Active Appearance Models for global texture feature extraction

Accurate facial feature alignment is the prerequisite. Active Shape Model (ASM) and Active Appearance Model (AAM) are the main methods for this problem. ASM is a parameterisation-based statistical model, which is mainly used in image feature points extraction and image segmentation . ASM [18] 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. AAM [19] contains a statistical model of the shape and grey-level appearance of the object of interest and can generalize to almost any valid example. These models represent objects as sets of labeled 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. AAM are parametric statistical models which describe the visual appearance of arbitrary object classes. The variation in object appearance is modeled by a shape component (represented by image landmarks) and a shape-free texture component. AAM are trained from sample images with annotated landmark positions. We use AAM to extract global texture feature. The steps can be briefly described as follows:

AAM training is based on annotated sample images. First, calibration of feature points. Given a training set of facial image $X = \cdots$, Where N is the number of images.

$L_n = \quad), \cdots \quad {}_{nM})\}$ is corresponding landmarks, where $n \in \quad \cdots$. The shape model is built by aligning the given shape samples $\{L_1, L_2 \cdots$, which includes translation, rotation, and scale via Procrustes analysis and results in shapes $\varsigma \; \varsigma \cdots$. The shape variations are then parameterized by applying principal component analysis (PCA) to obtain the matrix $\zeta = \quad - \quad \cdots$, where $\varsigma = \frac{1}{N} \sum$ is the mean shape. The result is a linear model which describes an arbitrary shape ς based on its shape parameters b_s and the shape eigenvectors p_s. The mean shape ς of all samples is calculated as follow:

$$\varsigma = \quad + \qquad (1)$$

The second step of AAM training is building a texture model. Firstly, each image $I_1, I_2 \cdots$ is warped into a common reference frame, namely the mean shape ς . The shape-normalized images are then vectorized and result in the texture vectors $\gamma \; \gamma \cdots$. Then, we employ the very same PCA-based procedure as for the shape model, which results in the linear texture model. It is expressed as follow:

$$\gamma = \quad + \qquad (2)$$

Where γ is an arbitrary shape-normalized texture with texture, parameters p_g and b_g are the texture eigenvectors, and $\gamma = \frac{1}{N} \sum$ is the mean texture of the samples.

The third step of AAM training is to merge shape and texture parameters into one parameter set, so we can obtain a combined representation of both shape and texture, which is achieved by concatenating the variance-normalized shape and texture parameter vectors for each training sample and again applying PCA. Thus we can obtain appearance parameter as follows:

$$\begin{pmatrix} \\ \\ \end{pmatrix} \qquad \begin{pmatrix} \\ \\ \end{pmatrix} \qquad (3)$$

$$\varsigma = \quad + \quad {}_{s}c \qquad (4)$$

$$\gamma = \quad + \qquad (5)$$

Where w_s is diagonal matrix and can adjust the weights between the shape model and texture model, p_c is feature vector, and c is comprehensive model parameters, it can be used to control the shape and texture of the model.

The last step is AAM search. We are synthetic appearance model according to the image and the model parameters. By adjusting the model parameters, model appearance and the actual image difference is minimal. Vector differential is expressed as follow:

$$\delta = \ - \qquad (6)$$

Where I_i is gray vector of image texture, I_m is gray vector of texture synthesis model. We can minimize the difference between the model and the image vector by changing the parameter c of the model. AAM algorithm has a good effect to the extraction of target and good robustness to the shade and degradation in real life problems.

b. Facial velocity feature extraction

Optical flow-based face representation has attracted much attention. B. Fasel and J. Luettin [20] used optical flow to estimate facial muscle actions for facial expressions recognition. According to the physiology, facial expressions are the result of facial muscle actions which are triggered by the nerve impulses generated by emotions. 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) $\begin{pmatrix} \\ \\ \end{pmatrix}$ at each frame using optical flow equation, which is expressed as:

$$\frac{\partial}{\partial}u+\frac{\partial}{\partial}+\frac{\partial}{\partial}- \ , \qquad (7)$$

where (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, $u = \frac{\partial}{\partial}$, $v = \frac{\partial}{\partial}$ is the horizontal and vertical velocities in pixel (x, y).

There are many methods to solve the optical flow equation. We use the Lucas and Kanade algorithm [21] to calculate the optical flow velocity. Lucas and Kanade method has better accuracy and stability.

We can obtain the facial velocity $\left(\quad \right)$ by using the least squares:

$$u\sum_{D} \qquad \sum_{D} \qquad \sum_{D} \tag{8}$$

$$u\sum_{D} \qquad \sum_{D} \qquad \sum_{D} \tag{9}$$

According to the equation (9) and equation (10), calculate optical flow vector $\left(\quad \right)$:

$$\left(\quad \right) \left(\begin{matrix} \sum & \sum & \sum \\ \sum & \sum & \sum \end{matrix} \right. \tag{10}$$

where φ is the window function. Lucas and Kanade method is fast to calculate and easy to implement. It meets the requirements of stable, reliable, high precision.

u and v are further half-wave rectified into four nonnegative channels u_x^{+}, u_x^{-}, v_y^{+}, v_y^{-} so that $u_x = \quad - \quad$ and $v_y = \quad - \quad$ These four nonnegative channels are then blurred with a Gaussian kernel and normalized to obtain the final four channels ub_x^{+}, ub_x^{-}, vb_y^{+}, vb_y^{-}.

Facial expression is represented by facial velocity features that are composed of the channels ub_x^{+}, ub_x^{-}, vb_y^{+}, vb_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.

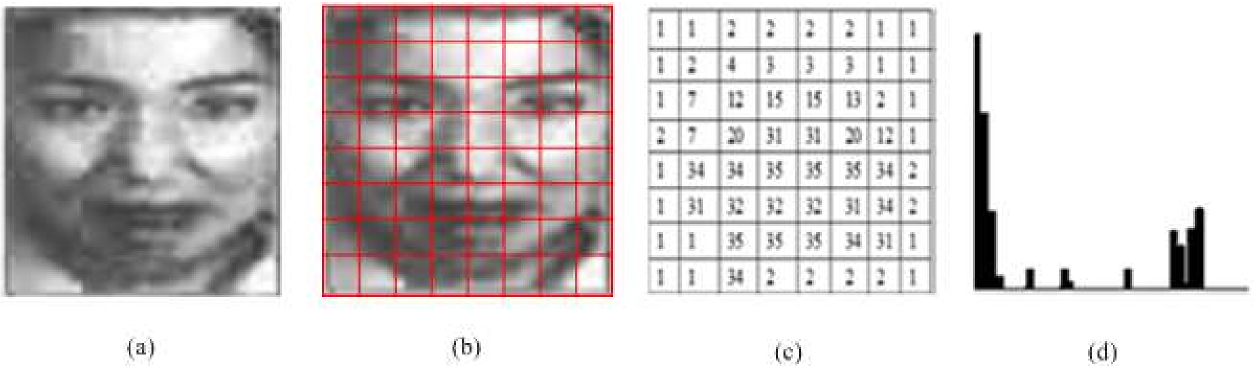
c. Visual words representation of facial expression

The human facial expression is a dynamic event; it must be represented by the motion information of the human face. To improve the accuracy of facial expression recognition, fusing the facial global texture feature vector and optical flow vector forms a hybrid feature vector by using the method of BoW (Bag of Words) [22], which can be better more effective for facial expression representation.

We divide each facial image into Lx L blocks , and each image block is represented by hybrid feature vector of all pixels in the block. On this basis , facial expressions are represented by visual words using the method of BoW. We randomly select a subset from all image blocks to construct the codebook. Then, we use k-means clustering algorithms to obtain clusters. Codewords are then defined as the centers of the obtained clusters, namely visual words. In the end, each face image is converted to the "bag-of-words" representation by appearance times of each codeword in the image is used to represent the image.

$$d \;=\; \qquad),\cdots \qquad \cdots \qquad , \tag{11}$$

where $n(I,w_i)$ is the number of visual word w_i included in image, M is the number of vision words in word sets. The processing pipeline of the "bag-of-words" representation is shown in figure 2.



1	1	2	2	2	2	1	1
1	2	4	3	3	3	1	1
1	7	12	15	15	13	2	1
2	7	20	31	31	20	12	1
1	34	34	35	35	35	34	2
1	31	32	32	32	31	34	2
1	1	35	35	35	34	31	1
1	1	34	2	2	2	2	1

FIGURE 2 (a) given a image,(b) divide into Lx L blocks (c) represent each block by a "visual word," and (d) ignore the ordering of words and represent the facial image as a histogram over "visual words."

III. LDA MODEL 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 classification problem, we use the Latent Dirichlet Allocation (LDA) model to learn and recognize human facial expression.

The LDA model has been applied to various computer vision applications, such as text classification, object recognition, action recognition, human detection, etc. The LDA framework is machine learning algorithm. Our approach is directly modeling topic by a body of work on using generative LDA models for visual recognition based on the "bag-of-words" paradigm. We propose a novel LDA framework, which can increase the recognition efficiency without compromising accuracy.

a. Analysis of LDA algorithm

Latent Dirichlet Allocation (LDA) was proposed by D. M. Blei [23] in 2003. LDA is a three-level hierarchical Bayesian model. The intuition behind LDA is the documents exhibit multiple topics. It is the key idea to model as a finite mixture over an underlying set of topics for each item of a collection. L. Shang and K. P. Chan. [24] extend the LDA topic model for modeling facial expression dynamics and proposed a very efficient facial expression recognition. A. Bansal and S. Chaudhary et al. [25] used a novel LDA and HMM-based technique for emotion recognition from facial expressions. LDA has widely used in image classification [26], human activities for classification [27, 28], etc.

LDA topic model has become popular tools for the unsupervised analysis of large document collections. The model posits a set of latent topics, multinomial distributions over words, and assumes that each document can be described as a mixture of these topics. With algorithms for fast approximate posterior inference, we can use the topic model to discover both the topics and an assignment of topics to documents from a collection of documents. The model of LDA is shown in figure 3.

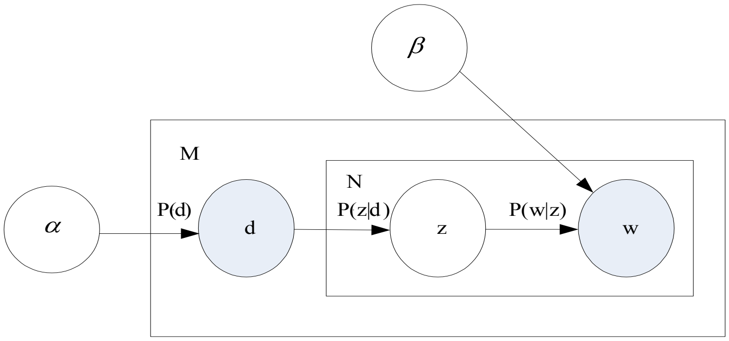


FIGURE 3 Graph model of LDA.

In figure 3, nodes represent random variables, shaded nodes are observed variables, and unshaded ones are unseen variables. LDA model is made up of three-level hierarchical Bayesian. Each item of a collection is modeled as a finite mixture over an underlying set of topics. Each topic is modeled as an infinite mixture over an underlying set of topic probabilities in turn.

Suppose document, word, topic is represented by d_j, w_i, z_k, respectively. Given (d_j, z_k, w_i), the joint probability of document d_j, topic z_k and word w_i can be expressed as:

$$p(d_j, z_k, w_i | \alpha \ \beta \ = \quad \prod_{k=}^{N} \quad \beta \qquad (12)$$

where $\alpha =$) is the probability of topic z_k occurring in image d_j, $\beta =$) is the probability of word ω occurring in topic z_k, $p(d_j | \alpha \propto \sum_{i}$ can be considered as the prior probability of d_j, and $p(w_i | z_k, \beta$ is the probability of word ω occurring in topic z_k, β. $p(w_i | z_k, \beta$ and $p(z_k | d_j)$ are Multinomial distributions. The conditional probability of $p(w_i | \alpha \ \beta$ can be obtained by marginalizing over all the topic variables z_k:

$$p(w_i | \alpha \ \beta \ = \int \quad {}_{K=}^{N} \quad \int \quad \beta \qquad (13)$$

To maximize the objective function is:

$$F = \prod_{i-}^{M} \prod_{-}^{N} \quad {}^{n(w_i, d_j)} \qquad (14)$$

The topic probabilities provide an explicit representation of a document. We present efficient approximate inference techniques based on variation methods and an EM algorithm for empirical Bayes parameter estimation. The EM algorithm consists of an expectation (E) step and a maximization (M) step: E-step computes the posterior probability of the latent variables, and M-step maximizes the completed data likelihood computed based on the posterior probabilities obtained from E-step. Both steps of the EM algorithm for LDA parameter estimate are listed below:

E-step:

Given α and β, estimate $p(z_k | d_j, w_i)$:

$$\alpha = \quad) \qquad (15)$$

$$\beta = \quad) \qquad (16)$$

$$p(z_k | d_j, w_i) \propto \quad) p(z_k | d_j) \qquad (17)$$

M-step:

Given the estimated $p(z_k | d_j, w_i)$ in E-step, and $n(w_i, d_j)$,estimate α and β:

$$\alpha = \quad \infty\sum_{i} \qquad {}_{k} \mid d_j, w_i) \qquad (18)$$

$$\beta = \quad \infty\sum_{i} \qquad {}_{k} \mid d_j, w_i) \qquad (19)$$

where $\sum_{i}$ is the length of document d_j.

Two step iteration until convergence.

Given a new document, the conditional probability distribution over aspect $p(z \mid d_{new})$ can be inferred by maximizing the objective function of d_{new} using a fixed word-aspect distribution $p(w_i \mid z_k)$ learned from the observed data. The iteration of inferring $p(z \mid d_{new})$ is the same as the learning process except that the word- topic distribution $p(w_i \mid z_k)$ in Eq.(20) is a fixed value that is learned from training data.

b. LDA-based facial expression recognition

LDA is the Latent Dirichlet Allocation model. It is an un-supervised learning method. To improve the recognition efficiency, we build LDA model. LDA is one of the most efficient machine learning algorithms. In LDA, learning must simultaneously learn. To obtain training samples, we treat each block in a image as a single word w_i, a image as a document d_j, and a facial expression category as a topic variable z_k. For the task of facial expression classification, our goal is to classify a new face image to a specific facial expression class. During the inference stage, given a testing face image, the document specific coefficients $p(z_k \mid d_{test})$, We can treat each aspect in the LDA model as one class of facial expressions So, the facial expression categorization is determined by the aspect corresponding to the highest $p(z_k \mid d_{test})$. The facial expression category ξ of d_{test} is determined as：

$$\xi = \qquad {}_{k}\, p(z_k \mid d_{test}) \qquad (20)$$

In order to shorten training time, we adopt a supervised algorithm to train LDA. Each image has a class label information in the training images. We make use of this class label information in the training images for the learning of the LDA model, so each image directly corresponds to a certain facial expression class on train sets, the image for training data become observable. The parameter $p(z_k \mid d_j)$ in the training step defines the probability of a word w_i drawing from a topic

z_k . Letting each topic in LDA corresponds to a facial expression category, the distribution $p(z_k | d_j)$ in the training can be simply estimated as:

$$\mathrm{p}(z_{\mathrm{k}} | d_j) = \frac{n}{n_j} \qquad (21)$$

where n_j is the number of the images corresponding to the j-th facial expression class and $n_{k,j}$ is the number of the k-th word (block) in the images corresponding to the j-th facial expression class. The LDA for facial expression recognition proceeds as follows:

Step 1: Local texture features are extracted by Active Shape Model (ASM).

Step2: Facial velocity features are extracted, each facial image is divide into *n*x *n* blocks, which is represented by optical flow vector of all pixels in the block.

Step3: Vision words are obtained using *k*-means clustering algorithms.

Step4: Facial expression is represented by fusion feature extraction of ASM and Optical flow, which use BoW histogram *d*.

Step5: For all *k* and *j*, calculate:

$$\mathrm{p}(z_{\mathrm{k}} | d_j) = \frac{n}{n_j} \qquad (22)$$

Step6: E-Step, for all (d_{test}, w_i) pairs, calculate:

$$p(z_k | d_{test}, w_i) = \frac{p(w_i | z_k) p(z_k | d_{test})}{\sum_{l=} \quad z_l | d_{test})} \qquad (23)$$

Step 7: Partial M-Step, fix $p(w_i | z_k)$ as calculated in step 5, for all *k*, calculate :

$$p(z_k | d_{test}) = \sum^{N} \frac{(z_k | d_{test}, w_i)}{n(d_{test})} \qquad (24)$$

Step 8: Repeat step 6 and step 7 until the convergence condition is met.

Step 9: Calculate facial expression class:

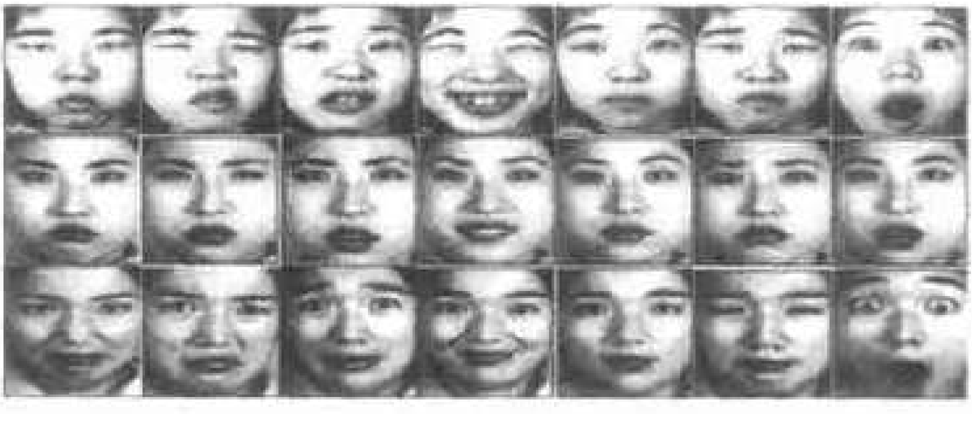
$$\xi = \quad {}_k\, p(z_k | d_{test}) \qquad (25)$$

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 LDA 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 LDA 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 LDA model, since each image directly corresponds to a certain facial expression class on train sets.

IV. 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 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 [29-30]. 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 4.



(a) (b) (c) (d) (e) (f) (g)

FIGURE 4 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 200 face images per class randomly for training, 120 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.

We randomly selected a subset from all image blocks to construct the codebook and used k-means clustering algorithms to obtain clusters. Visual words (Codewords) are then defined as the centers of the obtained clusters. To determine the value of *M* that is the number of the visual word set, the relation between *M* and recognition accuracy was observed, which is displayed in figure 5. It is revealed in figure 5 that with the increasing of *M* recognition accuracy is rise up at the beginning and if *M* is larger than or equal to 60, the recognition accuracy is stabled to 94.7%. As a result, *M* is set as 60.

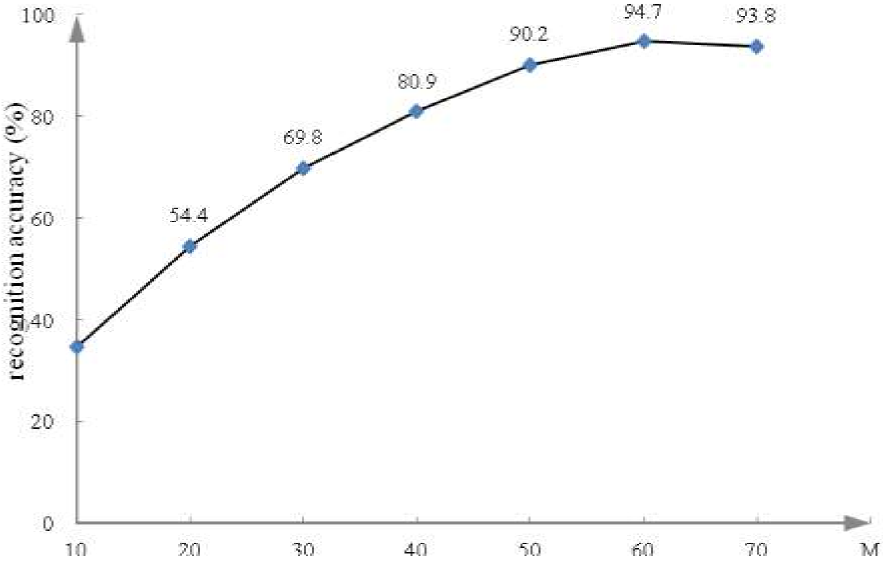


FIGURE 5 Relation curve between M and accuracy

In order to examine the accuracy of our proposed facial expressions recognition approach, we used 200 different face images for this experiment. Some images contained the same person but in different expressions. The recognition results are presented in the confusion matrices. The confusion matrix for per-video classification is shown in figure 6.

	“anger	disgust	fear	happiness	neutral	sadness	surprise
“anger	0.97	0	0	0	0	0.03	0
disgust	0.02	0.96	0.02	0	0	0	0
fear	0.04	0.03	0.93	0	0	0	0
happiness	0	0	0	0.95	0.03	0.02	0
neutral	0	0	0	0.09	0.90	0.01	0
sadness	0	0	0.03	0.02	0.01	0.94	0
surprise	0	0	0.02	0	0	0	0.98
The average recognition accuracy : 0.947							

FIGURE 6 Confusion matrix for facial expression recognition

Each cell in the confusion matrix is the average result of facial expression respectively. We can see that the algorithm correctly classifies most facial expressions. Average recognition rate gets to 94.7%. 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 “SLPP+ MKSVM”, which used SLPP to extract facial expression feature, and multiple kernels support vector machines (MKSVM) was used to recognize. 300 different expression images were used for this experiment, where some images contained the same person but in different facial expression. The results of recognition accuracy comparison are shown in figure 7.

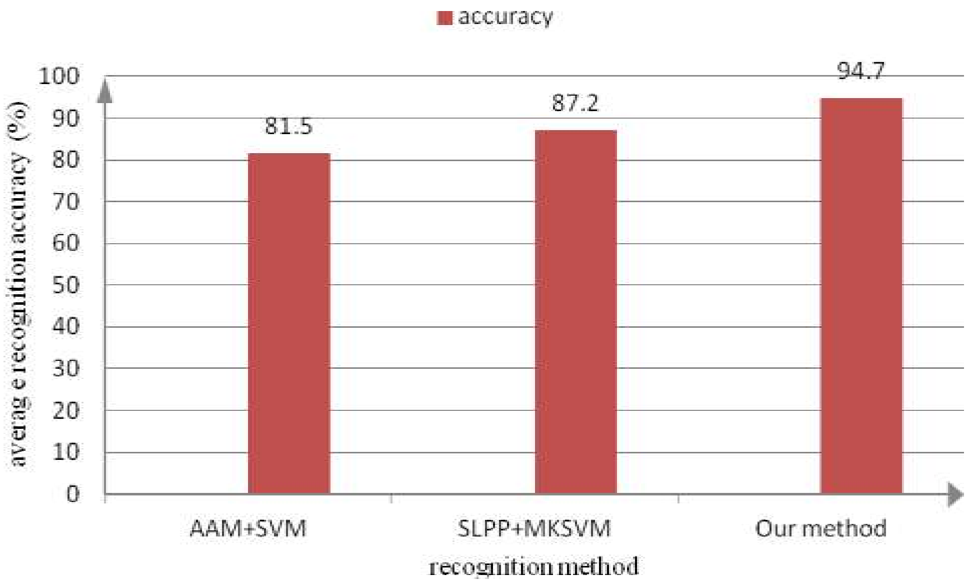


FIGURE 7 Recognition accuracy comparison of different method

In figure 7, we can see that AAM+SVM obtain average recognition accuracy of 81.5%. The average recognition accuracy of SLPP+ KSVM is to 87.2%. Our method is stabled to average recognition accuracy of 94.7%. Our method has higher recognition accuracy 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 global texture feature and motion features that were reliably with noisy image sequences and bag-of-words frame work to describe facial expression effectively. In the stage of expression recognition, we used LDA algorithm to classify facial expression images. Our method performs the best, its recognition accuracies and speeds are satisfactory.

V. 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) AAM model was used for global texture features extraction. Optical flow model was used to extract facial velocity features, then after fusing global texture features and facial velocity

features, we got hybrid features and converted them into visual words using “bag-of-words” models. Visual words were used for facial expression representation.

(2) LDA model was used for facial expression recognition, 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 LDA 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 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. XJK015BGD007), by the Teaching Reform Foundation of Hunan Province Ordinary College under Grant (No. 2012401544), by the Foundation for Key Constructive Discipline of Hunan Province and by the Foundation for Key Laboratory of Information Technology and Information Security in Hunan Province.

REFERENCES

[1] A. S. Ghotkar and G. K. Kharat, “Study of vision based hand gesture recognition using Indian sign language,” International Journal on Smart Sensing and Intelligent Systems, vol.7, no.1, 2014, pp. 96-114.

[2] Y. Wang and Y. Zhang, “Object tracking based on machine vision and Improved svdd algorithm,” International Journal on Smart Sensing and Intelligent Systems, vol.8, no.1, 2015, pp.677-696.

[3] F. Samaria and S. Young, “HMM-based architecture for face identification,” Image and Vision Computing, vol.12, No.8, 1994, pp. 537-543.

[4] C. Shan, S. Gong and P. W. McOwan, “Facial expression recognition based on local binary patterns: A comprehensive study,” Image and Vision Computing, vol.27, No.6, 2009, pp. 803-816.

[5] X. Feng, M. Pietikäinen and A. Hadid, “Facial expression recognition based on local binary patterns,” Pattern Recognition and Image Analysis, vol.17, No.4, 2007, pp.592-598.

[6] T. Xiang, M. K. H. Leung and S. Y. Cho, “Expression recognition using fuzzy spatio-temporal modeling,” Pattern Recognition, vol.41, No.1, 2008, pp.204-216.

[7] N. Neggaz, M. Besnassi and A. Benyettou, “Facial expression recognition,” Journal of Applied Sciences, vol.10, No.15, 2010, pp. 1572-79.

[8] G. Zhao and M. Pietikäinen, “Boosted multi-resolution spatiotemporal descriptors for facial expression recognition,” Pattern recognition letters, vol.30, No.12, 2009, pp. 1117-27.

[9] K. Yu, Z. Wang, L. Zhuo, et al, “Learning realistic facial expressions from web images,” Pattern Recognition, vol.46, No.8, 2013, pp.2144-2155.

[10] S. Y. Fu, G. S. Yang and X. K. Kuai, “A spiking neural network based cortex-like mechanism and application to facial expression recognition,” Computational Intelligence and Neuroscience, pp.1-13. Online publication date: 1-Jan-2012.

[11] M. N. Dailey, G. W. Cottrell, C. Padgett, et al, “EMPATH: A neural network that categorizes facial expressions,” Journal of Cognitive Neuroscience, vol.14, No.8, 2002, 1158-1173.

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

[13] L. Wang, R. F. Li, and K. Wang, “A novel automatic facial expression recognition method based on AAM,” Journal of Computers, vol.9, No.3, 2014, pp.608-617.

[14] K. M. Prkachin, “The consistency of facial expressions of pain: a comparison across modalities,” Pain, vol.3, No.5, 1992, pp. 297-306.

[15] K. M. Prkachin and P. E. Solomon, “The structure, reliability and validity of pain expression: Evidence from patients with shoulder pain,” Pain, vol.2, No.139, 2008, pp.267-274.

[16] S. Zhang, B. Jiang and T. Wang, “Facial expression recognition algorithm based on active shape model and gabor wavelet,” Journal of Henan University (Natural Science), vol.40, No.5, 2010, pp. 521-524.

[17] W. Zhang and L. M. Xia, “Pain expression recognition based on SLPP and MKSVM,” Int. J. Engineering and Manufacturing, No. 3, 2011, pp.69-74.

[18] K. W. Wan, K. M. Lam, and K. C. Ng, “An accurate active shape model for facial feature extraction,” Pattern Recognition Letters, vol.26, No. 15, 2005, pp.2409-23.

[19] T. F. Cootes, G. J. Edwards, C. J. Taylor, “Active appearance models,” Computer Vision—ECCV’98, Springer Berlin Heidelberg, 1998, pp. 484-498.

[20] B. Fasel and J. Luettin, “Automatic facial expression analysis: a survey,” Pattern recognition, vol.36, No.1, 2003, pp. 259-275.

[21] B. Lucas and T. Kanade, “An iterative image restoration technique with an application to sereo vision,” Proceedings of the DARPA IU Workshop, 1981, pp.121-130.

[22] G. J. Burghouts and K. Schutte, “Spatio-temporal layout of human actions for improved bag-of-words action detection,” Pattern Recognition Letters, vol.34, No. 15, 2013, pp.1861-1869.

[23] D. M. Blei, A. Y. Ng and M. I. Jordan, “Latent dirichlet allocation,” The Journal of Machine Learning Research, No. 3, 2003, pp.993-1022.

[24] L. Shang and K. P Chan, “A temporal latent topic model for facial expression recognition,” Computer Vision–ACCV 2010, Springer Berlin Heidelberg, 2011, pp. 51-63.

[25] A. Bansal, S. Chaudhary and S. D. Roy, “A novel LDA and HMM-Based technique for emotion recognition from facial expressions,” Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction, Springer Berlin Heidelberg, 2013, pp. 19-26.

[26] F. Monay and D. Gatica-Perez, “Modeling semantic aspects for cross-media image indexing,” 2007 IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.29, No. 10, 2007, pp.1802-1817.

[27] G. Sen Gupta, S.C. Mukhopadhyay and M Finnie, Wi-Fi Based Control of a Robotic Arm with Remote Vision, Proceedings of 2009 IEEE I2MTC Conference, Singapore, May 5-7, 2009, pp. 557-562.

[28] H. Zhang, Z. Liu and H. Zhao, “Human activities for classification via feature points,” Information Technology Journal, vol.10, No. 5, 2011, pp.974-982.

[29] J. C. Niebles, H. Wang and L. Fei-Fei, “Unsupervised learning of human action categories using spatial-temporal words,” International Journal of Computer Vision, vol.79, No3, 2008, pp.299-318.

[30] F. Cheng, J. Yu and H. Xiong, “Facial expression recognition in JAFFE dataset based on Gaussian process classification,” IEEE Transactions on Neural Networks, vol.21, No.10, 2010, pp.1685-1690.