A Hierarchical DGMM Recognizer for Chinese Sign Language Recognition*

WU Jiang-qin[1], GAO Wen[1,2], CHEN Xi-lin[1], MA Ji-yong[2]

[1](Department of Computer Science and Engineering, Harbin Institute of Technology, Harbin 150001, China)

[2](Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100080, China)

E-mail: jqwu@cti.com.cn

Received January 6, 1999; accepted July 7, 1999

Abstract: Sign language is the language used by the deaf, which is a comparatively steadier expressive system composed of signs corresponding to postures and motions assisted by facial expression. And it is a language communicated by motion/vision. The objective of sign language recognition research is to "see" the language of the deaf. The integration of sign language recognition and sign language synthesis jointly comprises a "human-computer sign language interpreter", which facilitates the interaction between deaf people and their surroundings. The issue of sign language recognition is to recognize dynamic gesture signal, that is, to recognize sign language signal. Considering the speed and performance of the recognition system, Cyberglove is selected as gesture input device in sign language recognition system, DGMM (dynamic Gaussian mixture model) is used as recognition technique, and hierarchical recognizer is used in recognizing module, which can recognize 274 sign language words coming from the dictionary of Chinese sign language with an accuracy of 97.4%, based on Chinese sign language's own characteristic. Compared with the recognition system based on single-DGMM recognizer, the recognition rate of hierarchical DGMM recognizer is nearly equal to that of single-DGMM recognizer, but its recognition speed is apparently much faster than that of single-DGMM recognizer.

Key words: sign language recognition; DGMM (dynamic Gaussian mixture model); hierarchical DGMM recognizer

Sign language is the language used by the deaf, which is a comparatively steadier expressive system composed of signs corresponding to postures and motions assisted by facial expression. And it is a language commu-

* Project is supported by the National Natural Science Foundation of China under Grant No. 69789301 (国家自然科学基金重点项目) and the National High Technology Development Program of China under Grant No. 863-306-ZT03-01-Z (国家863高科技项目基金). **WU Jiang-qin** was born in 1965. She is an associate professor and now is a Ph. D. student in Department of Computer Science and Engineering, Harbin Institute of Technology. Her current research areas include artificial intelligence, pattern recognition, and optimization. **GAO Wen** was born in 1956. He received his Ph. D. degree in computer application from Harbin Institute of Technology in 1988 and received his Ph. D. degree in electronics from Tokyo University in 1991. He is a professor now. His current research interests include computer vision, pattern recognition and image processing, multimedia data compressing, multimodal human computer interaction, virtual reality etc. **CHEN Xi-lin** was born in 1965. He received his Ph. D. degree in computer application from Harbin Institute of Technology in 1994. He is a professor now. His current research interests include image understanding, multimodal human computer interaction, data compressing, virtual reality etc. **MA Ji-yong** was born in 1963, received his Ph. D. degree in computer application technology from Harbin Institute of Technology. Now he is a postal doctor in the Institute of Computing Technology, the Chinese Academy of Sciences. His research interests include speech&speaker recognition, biometrics and multimodal human-computer interaction.

nicated by motion/vision[1].

The objective of sign language recognition research is to "see" the language of deaf. To "see" has two meanings: one is translating the language of deaf into corresponding written language by word to word; the other is making correct response to the requirement or query deaf language contains. The integration of sign language recognition and sign language synthesis jointly comprises a "human-computer sign language interpreter", which facilitates the interaction between deaf people and their surroundings.

According to the development history, Chinese sign language can be divided into two kinds: gesture language and finger-spelling language. Finger-spelling language is developed from alphabet language, in which a finger pattern stands for a bopomofo and is spelled into mandarin based on spelling scheme of Chinese. Gesture language is developed from glyph language, which makes full use of emotion and body motion to represent the most basic features of subject and action. As deaf people usually use gesture language to communicate in daily life, our research aims at gesture language recognition.

Some commonly used techniques in previous sign language recognition include: template matching, artificial neural network and hidden Markov model. Template matching is limited to a fixed pattern and inflexible, which is often used to recognize small-scale static set. Although artificial neural network has the strong property of discriminating and anti-jamming, its ability to deal with time sequences is weak, so now it is widely applied in recognizing postures. Fels's famous GloveTalk system[2] selects neural network as recognition technique. HMM[3] is a well known and widely used statistical method. HMM with standard topology has strong capability to describe temporal-spatial variety of sign language signal, which has been widely used in the domain of dynamic gesture recognition. Liang's[4] sign language recognition system uses a VPL's DataGlove as sign language input device, which can recognize 250 basic words from sign language textbook with an accuracy of 90.5% based on HMM. Starner and Pentland's[5] American sign language recognition system uses camera as sign language input device, which can recognize simple sentences randomly constructed from a 40-word lexicon using HMM, an accuracy of 99.2% is reached. Vogler and Metaxas's[6] American sign language recognition system uses a 3-D tracker and three orthogonal cameras as gesture input device, which implements recognition of 53 isolated words with an accuracy of 89.9% based on HMM. In addition, Kirsti Grobel and Marcell Assam[7] use HMM to recognize signals of 262 isolated language words inputted by the user wearing colored glove, with an accuracy of 91.3%. However, just the generalization of HMM topology leads to the model's difficulty in analyzing sign language signal and the computation complexity in HMM training and recognizing. Especially, for continuous HMM, the speed of training and recognition is comparatively slow, because of the requirement of computing large amount of emission probability and requirement of estimating too many model parameters. Thus, the HMM used by previous sign language recognition systems is usually discrete HMM. Aiming at the above problems, this paper avoids computing large amount of emission probabilities and estimating too many model parameters from the other point of view, that is, it uses DGMM (dynamic Gaussian mixture model) to model sign language signal. In order to reduce the computation complexity of multidimensional Gaussian density function, each multidimensional Gaussian density function is approximated by one-dimension equivalent probability density. And according to Chinese sign language's material characteristic, hierarchical DGMM recognizer is used in this paper to decrease model matching time. Compared with single DGMM recognizer, the recognition time of hierarchical DGMM recognizer almost reduces a half.

DGMM model is described in Section 1. Two basic problems for DGMM are proposed and their solutions are given in Section 2. In Section 3, the structure of a general sign language recognition system is presented briefly and the hierarchical DGMM recognizer applied in sign language recognition system is described too. Experimental results are given and analyzed in Section 4. The conclusion is drawn in the last section.

1 DGMM Model

In continuous DGMM, sign language signal $X(t)$ $(t=1,\ldots,T)$, the frame at moment t, is modeled by a time-varied mixture probability density function with M components, that is

$$P(X(t))=\sum_{j=1}^{M}\pi_j p_j(t)q_j(X(t)), \tag{1.1}$$

where t is time variable, π_j is Mixture Proportion, $\pi_j\geqslant 0$, $\sum_{j=1}^{M}\pi_j=1$, $q_j(X(t))$ is N-Gaussian mixture density, defined as

$$q_j(X(t))=\sum_{n=1}^{N}c_{jn}N_{\infty}(X(t),\mu_{jn},U_{jn}), \tag{1.2}$$

here c_{jn} is mixture weighted coefficient, N_{∞} is Gaussian density, μ_{jn} and U_{jn} are mean vector and covariance matrix related to component j and mixture n, $X(t)=(X_1(t),\ldots,X_p(t))$ is p dimension feature vector, i.e.,

$$N_{\infty}(X(t),\mu_{jn},U_{jn})=\frac{1}{(2\pi)^{p/2}\sqrt{|U_{jn}|}}\exp\left(-\frac{1}{2}(X(t)-\mu_{jn})^T U_{jn}^{-1}(X(t)-\mu_{jn})\right), \tag{1.3}$$

$p_j(t)$ is a probability density of time variable, and is selected as Gaussian distribution, that is

$$p_j(t)=\frac{1}{(2\pi)^{1/2}\sigma_j}\exp\left(-\frac{(t-\tau)^2}{2\sigma_j^2}\right), \tag{1.4}$$

here τ_j and σ_j are mean and deviation respectively.

To reduce the amount of data for model training and the computation complexity of model, $q_j(X(t))$ in Eq. (1.1) is selected as the following density function

$$q_j(X(t))=\sum_{n=1}^{N}c_{jn}N_{\infty}(X(t),\mu_n,U_n), \tag{1.5}$$

where μ_n, U_n is the mean vector and covariance matrix related to mixture n. This model is called SCDGMM (semi-continuous dynamic Gaussian mixture model).

For above SCDGMM, the probability density function with exponential function is needed to calculate, so the computation amount is comparatively large. In particular, if the dimension of feature vector used to describe sign language signal is too large, it will result in overflow of density function computing. In order to decrease the computation amount of probability density and avoid overflow, U_n is selected as diagonal matrix, i.e., $U_n=\mathrm{diag}(u_{n1}^2,u_{n2}^2,\ldots,u_{np}^2)$ and Gaussian density function $N_{\infty}(X(t),\mu_n,U_n)$ is approximated by one-dimensional equivalent probability function

$$f_m(X(t))=c_{nm}\frac{1}{\left(\prod_{i=1}^{p}u_m\right)^{p^{-1}}}\frac{1}{1+d_h(X(t))+\ldots\frac{1}{m!}d_h^m(X(t))}, \tag{1.6}$$

where $d_h(X(t))=\frac{1}{2}\|Y(t)\|^2$, $\dot{Y}(t)=\left(\frac{X_1(t)-\mu_{n1}}{u_{n1}},\ldots,\frac{X_p(t)-\mu_{np}}{u_{np}}\right)$.

2 Two Basic Problems for SCDGMM and Their Solutions

2.1 Two basic problems for SCDGMM

In order to make full use of SCDGMM in real-world application, two basic SCDGMM problems must be solved. With the notation $\lambda=(\pi,c,U,\mu,\tau,\sigma)$ to indicate the complete set of the model, where $\pi=(\pi_1,\pi_2,\ldots,\pi_M)$, $c=(c_{jn})_{M\times N}$, $U=(U_n)_{1\times N}$, $U_n=\mathrm{diag}(u_{n1}^2,u_{n2}^2,\ldots,u_{np}^2)$, $\mu=(\mu_n)_{1\times N}$, $\mu_{jn}=(\mu_{jn1},\mu_{jn2},\ldots,\mu_{jnp})$, $\tau=(\tau_1,\tau_2,\ldots,\tau_M)$, $\sigma=(\sigma_1,\sigma_2,\ldots,\sigma_M)$, the two basic problems are described as follows.

Problem 1. Given the observation sequence $X=X(1)X(2)\ldots X(T)$ and a model λ, how to compute $P(X/\lambda)$ efficiently, that is, the probability of the observation sequence, given the model?

Problem 2. How to adjust model parameter λ to maximize $P(X/\lambda)$?

Problem 1 is the evaluation problem, scoring the matching degree of the given model and the observation sequence. Problem 2 is the problem of estimating model parameters, adjusting model parameters to maximize the probability of the observation sequence, given the model. The observation sequences used to adjust model parameter are called training data.

2.2 Solutions to the two basic problems of SCDGMM

Solution to Problem 1. Assuming statistical independence of observations, the probability of the observation sequence, given model parameter λ, is

$$P(X/\lambda)=\prod_{t=1}^{T}P(X(t)/\lambda)=\prod_{t=1}^{T}\sum_{j=1}^{M}\pi_j p_j(t)q_j(X(t)), \tag{2.1}$$

$$q_j(X(t))=\sum_{n=1}^{N}c_{jn}N_{\infty}(X(t),\mu_n,U_n), \tag{2.2}$$

where $N_{\infty}(X(t),\mu_n,U_n)$ is approximated by one-order one-dimensional equivalent probability density, i.e.,

$$N_{\infty}(X(t),\mu_n,U_n)=V_n(X(t))(u_{n1}u_{n2}\cdots u_{np})^{-p^{-1}}, \tag{2.3a}$$

$$V_n(X(t))=(1+d_h(X(t))p^{-1})^{-1}, \tag{2.3b}$$

$p_j(t)$ is approximated by one-order one-dimensional equivalent probability density, i.e.,

$$p_j(t)=w_j(t)\sigma_j^{-1}, \tag{2.4a}$$

$$w_j(t)=\left[1+\frac{(t-\tau_j)^2}{2\sigma_j^2}\right]^{-1}. \tag{2.4b}$$

Solution to Problem 2. For the given K groups of training data $x_k(t)$, $t=1,\ldots,t(k)$, $k=1,\ldots,K$, where $t(k)$ is the total frame number of the kth group training data, the reestimation formulas of $\lambda=(\pi,c,U,\mu,\tau,\sigma)$ are obtained by likelihood estimate,

$$\bar{\pi}_j=\gamma_j\Big/\sum_{j=1}^{M}\gamma_j, \tag{2.5}$$

$$\bar{c}_{jn}=\gamma_{jn}/\gamma_j, \tag{2.6}$$

$$\bar{\tau}_j=\sum_{k=1}^{K}\sum_{t=1}^{t(k)}\gamma_j(t,k)w_j(t)t\Big/\sum_{k=1}^{K}\sum_{t=1}^{t(k)}\gamma_j(t,k)w_j(t), \tag{2.7}$$

$$\bar{\sigma}_j^2=\sum_{k=1}^{K}\sum_{t=1}^{t(k)}\gamma_j(t,k)w_j(t)t(t-\tau_j)^2\Big/\sum_{k=1}^{K}\sum_{t=1}^{t(k)}\gamma_j(t,k), \tag{2.8}$$

$$\bar{\mu}_n=\sum_{j=1}^{M}\sum_{k=1}^{K}\sum_{t=1}^{t(k)}\gamma_{jn}(t,k)v_n(x_k(t))x_k(t)\Big(\sum_{j=1}^{M}\sum_{k=1}^{K}\sum_{t=1}^{t(k)}\gamma_{jn}(t,k)v_n(x_k(t))\Big)^{-1}, \tag{2.9}$$

$$\bar{u}_{nl}=\sum_{j=1}^{M}\sum_{k=1}^{K}\sum_{t=1}^{t(k)}\gamma_{jn}(t,k)v_n(x_k(t))(x_{kl}(t)-\mu_{nl})^{\cdot p^{-1}}\Big(\sum_{j=1}^{M}\sum_{k=1}^{K}\sum_{t=1}^{t(k)}\gamma_{jn}(t,k)\Big)^{-1}, \tag{2.10}$$

here,

$$\gamma_j(t,k)=\pi_j p_j(t)q_j(x_k(t))/P(x_k(t)), \tag{2.11a}$$

$$\gamma_j=\sum_{k=1}^{K}\sum_{t=1}^{T(k)}\gamma_j(t,k), \tag{2.11b}$$

$$\gamma_{jn}(t,k)=\pi_j c_{jn}p_j(t)N(x_k(t),\mu_n,U_n)/P(x_k(t)), \tag{2.11c}$$

$$\gamma_{jn}=\sum_{k=1}^{K}\sum_{t=1}^{T(k)}\gamma_{jn}(t,k). \tag{2.11d}$$

3 Hierarchical DGMM Recognizer

As shown in Fig. 1, a current sign language recognition system includes front end processing module, training module (model parameter estimation module) and recognizing module. A hierarchical DGMM recognizer is

used in recognizing module.

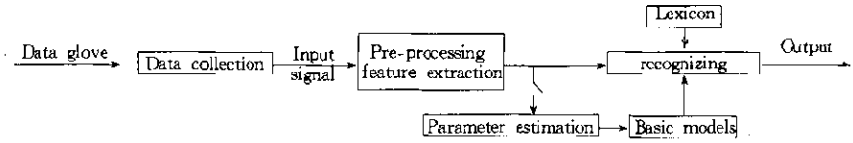


Fig. 1 System diagram

3.1 The front end processing module

The front end processing module includes data collection module, data preprocessing module and feature extraction module.

Data collection module: Two CyberGloves with 18 sensors are used as gesture input device with 38 400 baud rate of sampling. Using this module raw sensor data of each joint shown in Fig. 2 are obtained, they are, pinkie MPJ, pinkie PIJ, ring MPJ, ring PIJ, middle MPJ, middle PIJ, index MPJ, index PIJ, thumb MPJ, thumb IJ, pinkie-ring abduction, ring-middle abduction, middle-index abduction, thumb abduction, thumb rotation/TMJ, wrist yaw, wrist pitch and palm arch.

Data preprocessing: The gesture corresponding to a sign language word can stay for several seconds, during which all the data from dataglove are regarded as the signal representing this word, thus large amount of redundant information covers the essential property of the gesture. In addition, the physical property of dataglove leads to white noise for output data, while unconscious twitter of the subject can also result in the twitter of output. Moreover, gesticulating is a dynamic process, and considering the temporal characteristics of gestures may help in the temporal segmentation of gestures from other unintentional hand/arm movements. Three phases make a gesture: preparation, nucleus and retraction. Thus, the raw sensor data should be preprocessed. According to the stead characteristic and motion characteristic of the raw data, the following method is used for preprocessing raw sensor data:

Firstly, stead segment is determined by using one-dimensional window to filter each dimension data.

Secondly, common stead segment of each dimension is determined as the stead segment of raw data.

Lastly, according to each stead segment, preparation and retraction phases are deleted from raw sign language signal, and the observation sequence is obtained.

Feature extraction: Features such as posture, orientation and movement path are generally used to characterize sign language signal with time-space property (as Fig. 3 shows). As only dataglove is used as gesture input device in our system, it is apparent that extracting signal's orientation and movement path is impossible. Thus the left-hand and right-hand posture features extracted from dataglove's output parameters comprise the feature vector of training and testing samples.

3.2 Hierarchical DGMM Recognizer

In model parameter estimation module, a two-level DGMM recognizer is built: a first-level single DGMM threshold recognizer and two sub-recognizers. Assume that the word lexicon of system includes V words, each word in the lexicon is repeatedly gesticulated several times, and the collected data are stored in database. Firstly, word lexicon is divided into two sub-word lexicons, where single-hand gesture sub-word lexicon includes V_1 words, double-hand gesture sub-word lexicon includes V_2 words. Here assume all single-hand gesture words are gesticulated by the same hand (e.g. right hand). Secondly, using left-hand information of single-hand gesture to estimate single DGMM threshold model parameters λ_{th}, the first-level DGMM recognizer is built. Thirdly, using right-hand information of single-hand gesture to estimate V_1 DGMM model parameters $\lambda_v^{(1)}$ ($v=1\sim V_1$) corresponding to each single-hand gesture word, which is stored in basic model base 1, sub-recognizer 1 is

built; similarly using double-hand information of double-hand gesture to estimate V_2 DGMM model parameters $\lambda_v^{(2)}(v=1\sim V_2)$ corresponding to each double-hand gesture word, which is stored in basic model base 2, sub-recognizer 2 is built. Here, semi-continuos DGMM reestimation Eqs. (2.5)~(2.10) are used to estimate model parameters.

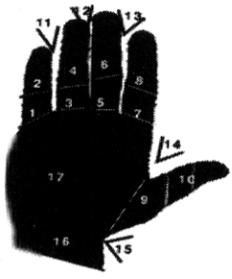
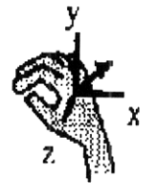
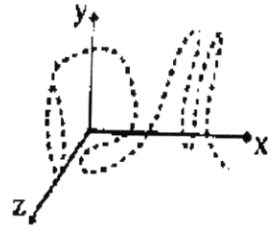


Fig. 2 The sketch map of Cyberglove measuring angle

(a) Posture



(b) Orientation



(c) Movement path

Fig. 3 Gesture features

According to the process shown in Fig. 4, for each sign language signal G from dataglove, an observation sequence $X=X(1)\ldots X(T)$, left-hand information sequence $X_l=X_l(1)\ldots X_l(T)$ and right-hand information sequence $X_r=X_r(1)\ldots X_r(T)$ are obtained through data preprocessing and feature extraction module. First, let threshold be Th and compute the probability of X_l corresponding to threshold model λ_{th}

$$P(X_l/\lambda_{th})=\prod_{t=1}^{T}P(X_l(t)/\lambda_{th}), \tag{3.1}$$

then the index of sub-recognizer is determined by the value $P(X_l/\lambda_{th})$:

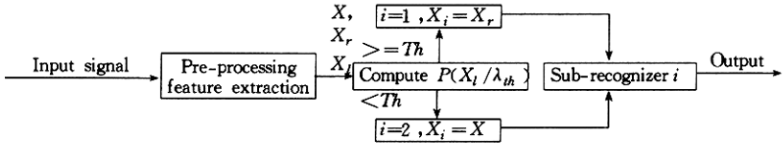


Fig. 4 The diagram of hierachical DGMM recognizer

If $P(X_l/\lambda_{th})>Th$, enter recognizer 1; otherwise enter recognizer 2.

Let $X_1=X_r$, $X_2=X$. According to the process shown in Fig. 5, the recognition process of sub-recognizer i ($i=1,2$) is as follows.

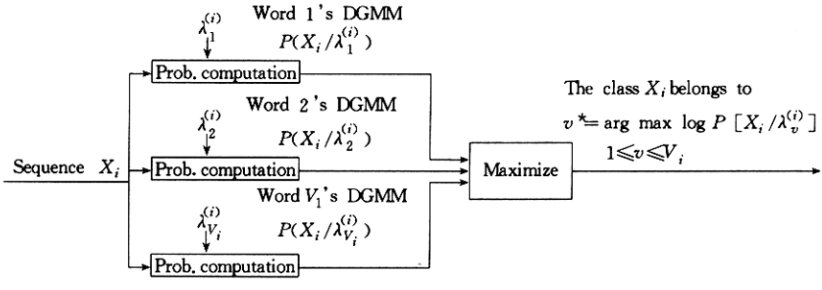


Fig. 5 The diagram of sub-recognizer i

Calculate the probability of x_i corresponding to each model $\lambda_v^{(i)}(1\leqslant v\leqslant V_i)$ in basic model base i

$$P(X_i/\lambda_v^{(i)})=\prod_{t=1}^{T}P(X_i(t)/\lambda_v^{(i)}). \tag{3.2}$$

The class X_i belongs to is determined by the following formula.

$$v^* = \arg\max_{1 \leqslant v \leqslant V_i} \frac{1}{T} \sum_{t=1}^{T} \log P(X_i(t)/\lambda_v^{(i)}). \tag{3.3}$$

4 Experimentation and Result

The experimentation aims at the word lexicon (as shown in Table 1) which includes 274 sign language words from Chinese sign language lexicon. The word lexicon is divided into two sub-word lexicons. They are single-hand gesture word lexicon including 115 words and double-hand gesture word lexicon including 159 words. Gesture input device is two CyberGloves with 18 sensors, whose sampling rate is 38 400 baud rate.

Every sign language word in each sub-word lexicon is gesticulated 10 times by the user who wears one data-glove with each hand. Through data collection, data preprocessing and feature extraction module, 10 observation sequences corresponding to every word are obtained and respectively stored in single-hand gesture database and double-hand gesture database, 8 of which are training sequences, the remaining of which are testing sequences. Considering the speed and performance of recognition, semi-continuous DGMM is used to model sign language signal and a hierarchical DGMM recognizer is used in the experimentation too. Firstly, using left-hand information of training data in single-hand gesture database to estimate single DGMM threshold model parameter λ_θ, the first-level DGMM recognizer is built. Secondly, using right-hand information of training data in single-hand gesture database to estimate V_1 DGMM model parameters $\lambda_v^{(1)}$ ($v=1\sim115$) corresponding to each single-hand gesture word, which is stored in basic model base 1, sub-recognizer 1 is built; similarly using double-hand information of training data in double-hand gesture database to estimate V_2 DGMM model parameter $\lambda_v^{(2)}$ ($v=1\sim159$) corresponding to each double-hand gesture word, which is stored in basic model base 2, sub-recognizer 2 is built. Thirdly, for each testing sample hierarchical DGMM recognizer (as Fig. 4 shows) is used to determine the index corresponding to the best matching model. Finally, in contrast to the lexicon shown in Table 1 and the basic model base, the corresponding meaning is determined.

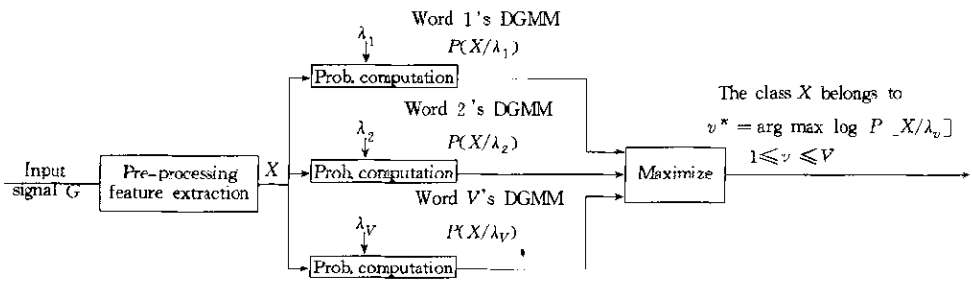


Fig. 6 The diagram of single DGMM recognizer

By adjusting time mixture M and Gaussian density mixture N of the model, different experimental results are obtained (as Fig. 7 shows), where average number of the wrongly recognized words refers to the mean of wrongly recognized words in 2 groups of testing samples.

According to the result shown in Fig. 7 and considering the speed and performance of recognition, select $M=2$, $N=1$, and an accuracy of 97.4% is reached.

From the result shown in Figs. 7 and 8, it is shown that the recognition performance of hierarchical DGMM recognizer is comparable to that of single DGMM recognizer, but its recognizing speed is much faster than single DGMM recognizer's, because the required time for recognizing is reduced almost a half.

In addition, single DGMM recognizer is used as the contrast of hierarchical DGMM recognizer, and the ex-

perimental result is shown in Fig. 8.

By calculating statistically the probability of wrongly recognized word in the experimentation and analyzing the reason of being wrongly recognized, a conclusion is drawn: the feature of double-hand posture is not enough for characterizing sign language signal. For example, in all testing experiments using hierarchical and single DGMM recognizers, "zuzhi" is wrongly recognized as "danwei" by 0.9 in hierarchical DGMM recognizer and by 0.95 in single DGMM recognizer. It is due to that "danwei" (as Fig. 9 shows) and "zuzhi" (as Fig. 10 showns) are almost the same on posture, which can be also shown from the variety on each dimension of the feature vector of "zuzhi" in Fig. 11 and the variety on each dimension of the feature vector of "danwei" in Fig. 12.

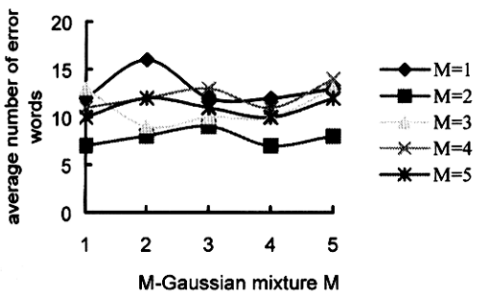


Fig. 7 The relation between error words number and N, M in hierarchical DGMM recognizer

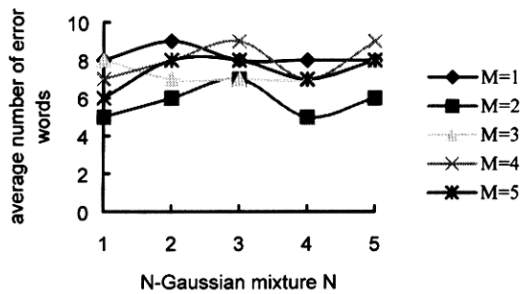


Fig. 8 The relation between error words number and N, M in single DGMM recognizer

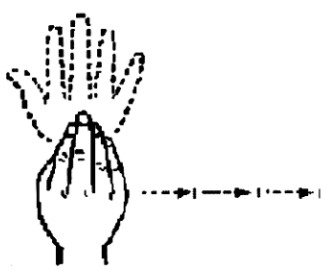


Fig. 9 The gesture graph of"danwei"

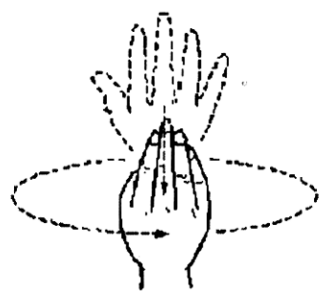


Fig. 10 The gesture graph of"zuzhi"

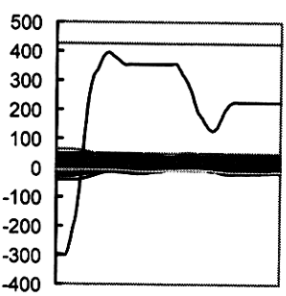


Fig. 11 The variety on each dimension of "zuzhi" feature vector

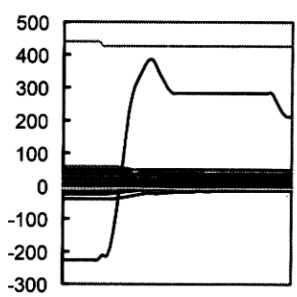


Fig. 12 The variety on each dimension of "danwei" feature vector

But their movement paths are comparatively different. So, orientation and movement path should be considered to recognize such similar words. Thus, to improve the recognition performance, it is necessary to select

3-D tracker as dataglove's assistant to obtain the feature of orientation and movement path.

Table 1 Lexicon

Single-Hand gesture word	Double-Hand gesture word
ayi;bashi;baoming;biyao;bocai;bu;buran;canguan;zangzu;chabuduo;chengji;chengxiao;chongzi;chouxiang;cizhi;dazao;danchun;danwei;danshi;daoli;daolu;dianbao;dianzi;dengji;duanwujie;dun;duibuqi;chua;ershi;fanying;fenbi;fouze;guwen;guilv;guoqi;he;houlai;hushi;jiating;jiao;jieli;jiushi;jiucai;kecue;kekao;keneng;keshi;kongqi;kongzhi;labi;litang;lizi;liushi;lvshi;manzu;mengguzu;mi;miantiao;miaozu;naxie;nainai;nver;nvshi;pizhun;qishi;qingcai;qingshi;sanshi;shan;shao;shengdong;shenghuo;shengren;shiyou;shiyong;shouqiang;shichang;shucai;shuxi;shuoshi;sishi;ti;tigao;tiankong;tian;touzi;warle;weiwuerzu;wushi;xibao;xijun;xiangxiang;xinyu;xing;xingming;xiongxin;yeye;youju;youpiao;you;yuan;yuanze;yueyue;zhexue;zhengce;zhibu;zhiyuan;zhiyuan;zhidu;zhihui;zhongwu;zhuyi;zhuangzu;zili;zizun;zuzhi;gangcai;biaoxian;	ai;aidai;aihu;aiqing;baicai;beizi;bijiao;biaozhi;bcwu;cailiao;chanlan;cha;chanpin;chi;chucai;chuang;ciyao;cong;danran;diren;dikang;difang;dianxing;dongwu;dongwuyuan;douzheng;fandui;fenqi;fendou;funvjie;gongcheng;gonghui;gongju;gongyi;gonggongqiche;gongmin;gongping;gongsi;guanjian;guanxin;guanli;guanchuan;guanghui;guangming;guiding;guohui;guoji;guomin;guoqingjie;guowuyuan;guoran;huanhe;huoche;jijinhui;jisuan;jijie;jiagong;jiaju;jianduan;jianrui;jianyi;jiankang;jiegou;jieguo;jiejue;jieshao;jindu;jinzhan;juece;jueding;kaoyan;kexi;kecheng;laibuji;kaolv;li;mishu;mianbao;minzhu;minzu;mo;neirong;nianling;paichusuo;peiyang;pinzhing;pohuai;qici;qiaqiao;qiangdu;qunzong;renmen;rongyu;ruozhuren;shehui;sheng;shengwu;shigong;shidai;shizhi;shiwu;shizi;shiru;shuiguo;taitai;taiyang;taozi;tixi;tianye;tiaojian;tongyong;tuanti;wazi;wancheng;wanju;weifeng;weiwang;weixie;weida;wuqi;wuli;xitong;xianjin;xingshi;xingxiang;xingming;xiongxin;yeyu;yihuier;yinxiang;yingxiong;youyu;yulun;yuzhou;yuanliao;yuesu;zanshi;zenme;zhangwo;zhengfu;zhichi;zhishifenzi;zhixing;zhiliang;zhonggong;zhongguo;zhongqiujie;zhongyang;zhongyu;zhudong;zhuzhang;zhuanjia;zhuanyong;zhuangzhi;ziben;zuqiu;zuihou;zunshou;

5 Conclusions

In this paper hierarchical DGMM recognizer applied in sign language recognition is expounded, dynamic Gaussian mixture model (DGMM) is used as recognition technique, and multidimensional Gaussian density function is approximated by one-order one-dimension equivalent density function. The reestimation formula estimating the parameters of SCDGMM (semi-continuous dynamic Gaussian mixture model) is also given in this paper, and by hierarchical DGMM recognizer, the sign language words in Table 1 are recognized using SCDGMM. Compared with the system using single DGMM recognizer, the recognition performance of hierarchical DGMM recognizer is equivalent to that of single DGMM recognizer, while its recognizing speed is much faster than that of single DGMM recognizer. By the experimental result, it is shown that if a 3-D tracker is used in the system, the recognition performance will be increased a lot. In addition, some words in Chinese sign language should be characterized with the assistant of face emotion and head gesture, so the research on fusion of multi-channel will further the development of Chinese sign language recognition.

References:

[1] China Deaf-Mute Association. Chinese Sign Language. Beijing: Huaxia Publishing Company, 1991. i~xi.

[2] Fels, S. S., Hinton, G. E. Glove-talk: a neural network interface between a dataglove and a speech synthesizer. IEEE Transactions on Neural Networks, 1993,4(1):2~8.

[3] Rabiner, L. R., Juang, B. H. An introduction to hidden Markov models. IEEE ASSP Magazine, 1986,3(1):4~16.

[4] Liang, R., Ouhyoung, M. A Sign Language Recognition System Using Hidden Markov Model and Context Sensitive Search. In: Mark Green ed. Proceedings of the ACM Symposium on VR Software and Technology. Hong Kong, 1996. 59~66.

[5] Starner, T., Pentland, A. Real-time American sign language recognition from video using hidden Markov models. In: MIT Media Laboratory Perceptual Computing Section, TR-375, 1996.

[6] Vogler, C., Metaxas, D.. ASL Recognition Based on a Coupling between HMMs and 3D Motion Analysis. In: Proceedings of the International Conference on Computer Vision. Bombay, 1998. 363~369.

[7] Grobel, K., Assam, M. Isolated sign language recognition using hidden Markov models. In: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics. Orlando, 1987. 162~167.

多层 DGMM 识别器在中国手语识别中的应用

吴江琴[1], 高 文[1,2], 陈熙霖[1], 马继涌[2]

[1](哈尔滨工业大学 计算机科学与工程系,黑龙江 哈尔滨 150001)

[2](中国科学院 计算技术研究所,北京 100080)

摘要: 手语是聋人使用的语言,是由手形动作辅之以表情姿势由符号构成的比较稳定的表达系统,是一种靠动作/视觉交际的语言.手语识别的研究目标是让机器"看懂"聋人的语言.手语识别和手语合成相结合,构成一个"人-机手语翻译系统",便于聋人与周围环境的交流.手语识别问题是动态手势信号即手语信号的识别问题.考虑到系统的实时性及识别效率,该系统选取 Cyberglove 型号数据手套作为手语输入设备,采用 DGMM(dynamic Gaussian mixture model)作为系统的识别技术,并根据中国手语的具体特点,在识别模块中选取了多层识别器,可识别中国手语字典中的 274 个词条,识别率为 97.4%.与基于单个 DGMM 的识别系统比较,这种模型的识别精度与单个 DGMM 模型的识别精度基本相同,但其识别速度比单个 DGMM 的识别速度有明显的提高.

关键词: 手语识别;动态高斯混合模型;多层 DGMM 识别器

中图法分类号: TP391 **文献标识码**: A