视频序列中表情和姿态的双模态情感识别

姜明星[1,2]，胡敏[1]，王晓华[1]，任福继[1,3]，王浩文[1]

[1]合肥工业大学计算机与信息学院情感计算与先进智能机器安徽省重点实验室，安徽 合肥 230009；

[2]安徽国际商务职业学院信息服务系，安徽 合肥 231131；

[3]德岛大学先端技术科学教育部，日本 德岛 7708502

摘要 针对时空局部方向角模式应用到视频情感识别时，出现的特征稀疏、噪声敏感等问题，提出了一种新的特征提取算法——时空局部三值方向角模式(SLTOP)。考虑到表情和姿态特征的互补性，提出云加权决策融合的分类方法。对视频图像进行预处理，得到表情和姿态两种模态的序列；分别提取表情序列和姿态序列的 SLTOP 特征，并借鉴灰度矩阵思想解决特征直方图过于稀疏的问题；在决策分类阶段，引入云模型对表情和姿态两种模态进行云加权决策融合，实现双模态情感的最终识别。在 FABO 数据库中，表情和姿态单模态分别取得了 92.21%和 96.76%的平均识别率；与体积局部二值模式、三正交平面局部二值模式(LBP-TOP)、时空局部三值模式矩(TSLTPM)比较时，在表情模态上分别高约 18.42%、22.01%、9.15%，而在姿态模态上分别高约 26.59%、29.53%、1.98%。通过云加权融合得到平均识别率为 97.54%，均高于其他实验得到的数据。所提出的 SLTOP，对噪声和光照具有很好的稳健性。利用云模型的加权决策融合方法可以较好地发挥表情和姿态分类器的性能，得到较好的识别结果，与其他分类识别方法进行对比实验，结果同样表现出优越性。

关键词 图像处理；表情；姿态；时空局部三值方向角模式；云模型

中图分类号 TP301.6 **文献标识码** A **doi**：10.3788/LOP55.071004

Dual-Modal Emotion Recognition Based on Facial Expression and Body Posture in Video Sequences

Jiang Mingxing[1,2], Hu Min[1], Wang Xiaohua[1], Ren Fuji[1,3], Wang Haowen[1]

[1]*Anhui Province Key Laboratory of Affective Computing and Advanced Intelligent Machine, School of Computer and Information, Hefei University of Technology, Hefei, Anhui* 230009, *China*;

[2]*Information and Service Department, Anhui Institute of International Business, Hefei, Anhui* 231131, *China*;

[3]*Graduate School of Advanced Technology & Science, University of Tokushima, Tokushima* 7708502, *Japan*

Abstract Aiming at the problems of feature sparseness and noise sensitivity when the temporal-spatial local direction angle mode is applied to the video emotion recognition, we propose a new feature extraction algorithm, the spatiotemporal local ternary orientation pattern (SLTOP). Considering the complementarity of facial expression and posture characteristics in recognition, a classification method based on the cloud weighted decision fusion is proposed. The video image is preprocessed to obtain the sequence of the two modes of facial expression and gesture. For reducing the sparseness of the feature histogram, we extract the SLTOP feature of the sequences of expression and posture, learning from the idea of gray level co-occurrence matrix. In the stage of decision fusion, the cloud model is introduced to implement the cloud weighted decision fusion for the two modes of expression and posture making to realize the final recognition of dual-modal emotion. The average recognition rate of the single modal of facial expression and body posture in the FABO database is 92.21% and 96.76%, respectively. And they are

收稿日期：2017-11-20；**收到修改稿日期**：2017-12-31

基金项目：国家自然科学基金(61672202，61502141)、国家自然科学基金-深圳联合基金(U1613217)、高校优秀青年骨干人才国内外访学研修(gxfx2017189)、安徽高校自然科学研究(KJ2016A126)

作者简介：姜明星(1984—)，男，硕士，副教授，主要从事计算机视觉、图像处理、数据挖掘等方面的研究。

E-mail：mx0551@163.com

approximately 18.42%, 22.01% and 9.15% higher in expression, respectively, when compared with the volume local binary mode, local binary mode three orthogonal planes (LBP-TOP) and temporal-spatial local ternary pattern moment (TSLTPM). In the single-posture modal, they are 26.59%, 29.53%, 1.98% higher, respectively. The average recognition rate obtained by cloud-weighted fusion is 97.54%, which is higher than that of other experiments. The proposed SLTOP has good robustness to the noise and illumination. The weighted decision fusion method of cloud model is used to greatly express the performance of two classifiers with expression and posture. The superiority of the recognition results in this paper is shown comparing with other classification methods.

Key words image processing; facial expression; body posture; spatiotemporal local ternary orientation pattern; cloud model

OCIS codes 100.5010; 040.7290; 100.2960

1 引 言

近年来,随着计算机情感计算技术的发展,传统的以机器为中心的人机交互方式已无法满足人们的生活需求,人们开始尝试建立“情感模型”,使计算机像人类一样具有识别、理解、表达“情感”的能力,建立和谐、生动的人机交互环境。虽然面部表情[1]、身体姿态[2]、语音信号[3]等单模态方式均能展示一定的情感,但是人在交流时更多地依赖多重感官以获取和表达情感。因此建立多通道的人机交互环境,才能使交互更加自然。同时,由于情感行为是连续的,依赖单模态获取的情感不全面,因此需要其他模态方式提供补充信息。目前,多模态情感计算正在成为人机交互的研究热点,而多模态技术能有效提供连续情感行为的动态信息,有力提升情感计算的研究深度[4]。

迄今为止,单模态情感识别已经取得较好的识别效果,国内外学者也已逐渐转向双模态或多模态情感识别研究。文献[5]改进了体积局部二值模式(VLBP)[6],提出了三正交平面体积局部二值模式(VLBP-TOP)并成功提取了视频图像中面部表情的三维时空特征,但随着邻域半径的扩大,获得的特征向量长度急速增长,所得特征处于较高的维度,但该算法未考虑纹理变化的梯度信息。文献[7]在VLBP基础上,引入语义特征并实现了时空特征的有效降维,但缺少判别性的有效特征。王晓华等[8]对VLBP算子进行改进,并融合三维方向梯度直方图(3DHOG)特征,有效提取了局部纹理变化信息,但缺少脸部和姿态特征对情感识别贡献大小的分析。付晓峰等[9]基于多尺度时空局部方向角模式(LOP)直方图映射有效捕捉了三维空间的时空特征,具有较快的识别速度,但由于方向角差值采用二值化处理,对图像噪声及剧烈光照稳健性差,特征较为稀疏,特征空间分类性能受到限制。Gunes等[10-13]针对表情和姿态的双模态情感识别进行长期的研究,建立了表情和姿态的FABO数据库,并使用基于视频的隐马尔可夫模型和最大投票顶点帧的方法进行情感识别研究,但实时特征处理较复杂,同时时间复杂度相对较高。Chen等[14]提出采用运动历史图像定向梯度直方图(MHI-HOG)和Image-HOG特征,通过时间归一化和词袋模型来提取运动和外观特征,双模态识别率显著提高。闫静杰等[15]通过双边稀疏偏最小二乘(BSPLS)方法对表情和姿态两种模态提取的时空特征进行融合,用最近邻(NN)分类器和支持向量机(SVM)分别进行分类识别,计算复杂度降低,但识别准确率不高。

受文献[9]工作的启发,本文提出一种时空局部三值方向角模式(SLTOP)将表情和姿态的序列图像堆叠成多尺度的时空特征。采用三值编码的局部方向角差值,设定自适应阈值,并将特征拓展到三维空间,将时域信息融入特征分布。考虑到三进制编码转换为十进制后,得到的直方图过于稀疏,使用灰度共生矩阵[16]统计字符跳变次数,实现降维。在分类环节,为体现两种模态对情感识别结果的贡献大小不同,将训练集中通过交叉验证得到的不同情感类别的准确率作为云滴,利用云模型[17-19]求取两个模态的权值,并利用NN得到测试样本属于不同情感类别的后验概率,进行各个情感类别局部特征分类结果的加权融合,得出最终的识别结果。

2 SLTOP特征

2.1 LOP算子

LOP算子[9]在局部二值模式(LBP)特征基础上加入方向角信息(图1),通过设定4个方向,即$g_c \rightarrow g_0$、$g_c \rightarrow g_1$、$g_c \rightarrow g_2$、$g_c \rightarrow g_3$,得到的4个差值分别为

$$\begin{cases} D_0(g_c) = g_c - g_0 \\ D_1(g_c) = g_c - g_1 \\ D_2(g_c) = g_c - g_2 \\ D_3(g_c) = g_c - g_3 \end{cases}, \tag{1}$$

式中$D_i(g_j)$表示g_j为中心的4个方向的差值,而

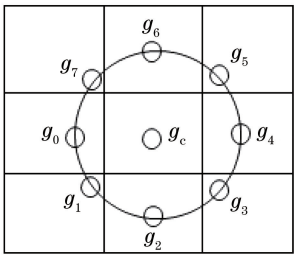

图 1 LOP 算子

Fig. 1 LOP operator

$g_c, g_1, g_2, \cdots, g_7$ 分别为各自位置的灰度值。得到 $g=g_c$ 处的 LOP 编码(LOPC)为

$$\mathrm{LOPC}(M,R)=\{\mathrm{LOPC}_i(M,R), i=0,1,2,3\}, \tag{2}$$

$$\mathrm{LOPC}_i(M,R)=\sum_{m=0}^{M-1}J[D_i(g_c)\cdot D_i(g_m)]2^m, \tag{3}$$

式中 M 为近邻点个数，R 为半径，$J[x]=1-D(x)$，$D(x)=\begin{cases}1, x\geqslant 0\\ 0, x<0\end{cases}$。通过比较近邻点的两个方向角差值来标注 g_j 像素点，针对比较差值变化得到的方向角模式特征进行编码，即得到 LOP 算子。LOP 算子不同于 LBP 近邻差值的二进制编码，通过邻域内方向角差值的变化进行编码，得到更准确的局部特征信息。

2.2 局部三值模式

为提升 LBP 算子在局部区域噪声和光照变化的稳健性，Tan 等[20]将 LBP 扩展到局部三值模式(LTP)，增加了−1 模式和自定义阈值 ε，增加了特征的信息量，同时提高特征空间分类性能。LTP 修改了 LBP 模型的阈值函数，具体定义 $g=g_c$ 处的差值如式(4)，其中 g_i 为邻域 P 内的其他像素点，如图 2 所示。

$$D(g_c)=\begin{cases}-1, & g_c-g_i\geqslant\varepsilon\\ 0, & |g_c-g_i|<\varepsilon\\ 1, & g_c-g_i<-\varepsilon\end{cases}, \tag{4}$$

式中 ε 为自定义的阈值，ε 的取值决定 LTP 抗噪能力，假定 ε 取值为 5。虽然 LTP 较 LBP 表现出较强的特征分类性能，极大地增强了对噪声的敏感性，但若阈值取固定值，无法区别样本间的差异，局部特征

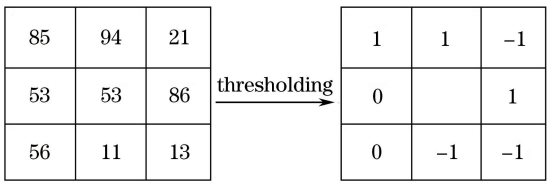

图 2 LTP 算子

Fig. 2 LTP operator

的稳健性不高，有待进一步优化阈值。

2.3 时空局部三值方向角模式(SLTOP)

鉴于上述 LOP 算子在描述方向角差值时，采用二值模式，对图像噪声及剧烈光照变化不够敏感，为了提升局部纹理特征的描述能力，SLTOP 通过使用自适应阈值 ε，将 LOP 算子拓展为三值编码模式。鉴于 LOP 得到的特征较为稀疏，融合 Zhao 等[5]提出的体积局部二值模式(VLBP)的思想，将视频序列图像堆叠成三维时空特征，使用灰度共生矩阵[14]表示字符跳变次数，从而有效解决直方图分布稀疏的问题。

自适应阈值 ε 的具体计算步骤如下：

1) 通过计算中心 g_j^c 在 4 个方向角的差值 $D^i(g_j^c)$，并统计求取所有差值的均值，使用差值次数 $q=p+1$。其中，$D^i(g_j^c)$ 表示 $g_j^c\rightarrow g_j^i$ 方向角差值。

$$D_j^i(g_j^c)=g_j^c-g_j^i,$$

$$(i=0,1,\cdots,p-1; j=t-1,t,t+1), \tag{5}$$

$$q=p+1, \tag{6}$$

$$\overline{\Delta D}=\Big[\sum_{i=0}^{p=1}\sum_{j=t-1}^{t+1}D_j^i(g_j^c)\Big]\Big/12q。 \tag{7}$$

2) 根据各差值和均值，计算方差 V 为

$$V=\sum_{i=0}^{p=1}\sum_{j=t-1}^{t+1}(D_j^i(g_j^c)-\overline{\Delta D})^2/12q。 \tag{8}$$

3) 由方差近似估计阈值 ε 为

$$\varepsilon=\sqrt{V}, \tag{9}$$

式中方差 V 随图像像素的变化而呈正比例变化。虽然 ε 是由方差 V 近似得到，对不同的样本具有一定的自适应性，但一般情况下，ε 变化范围有限，表现出较强的稳健性，有助于提升识别精度。将三值 LOP 拓展到选取视频帧图像 I_t 的像素值 g_t^c 作为阈值，根据公式对邻域 P 内的所有像素值进行三值处理，得到 $j=t$ 时像素三值向量 $\boldsymbol{U}$ 为

$$\begin{aligned}\boldsymbol{U}=\boldsymbol{U}\{&J[D_t^i(g_c)\cdot D_{t-1}^0(g_0)], J[D_t^i(g_c)\cdot D_{t-1}^1(g_1)],\cdots, J[D_t^i(g_c)\cdot D_{t-1}^{m-1}(g_{p-1})],\\ &J[D_t^i(g_c)\cdot D_t^0(g_0)], J[D_t^i(g_c)\cdot D_t^1(g_1)],\cdots, J[D_t^i(g_c)\cdot D_t^{m-1}(g_{p-1})],\\ &J[D_t^i(g_c)\cdot D_{t+1}^0(g_0)], J[D_t^i(g_c)\cdot D_{t+1}^1(g_1)],\cdots, J[D_t^i(g_c)\cdot D_{t+1}^{m-1}(g_{p-1})]\},\end{aligned} \tag{10}$$

式中 $i=0,1,2,3$。将像素三值向量 U 按照一定的顺序变换成 3 pixel×8 pixel 的矩阵 $\boldsymbol{M}$ 为

$$\boldsymbol{M}=\begin{bmatrix} J[D_t^i(g_c)\cdot D_{t-1}^0(g_0)] & J[D_t^i(g_c)\cdot D_{t-1}^1(g_1)] & \cdots & J[D_t^i(g_c)\cdot D_{t-1}^{m-1}(g_{p-1})] \\ J[D_t^i(g_c)\cdot D_t^0(g_0)] & J[D_t^i(g_c)\cdot D_t^1(g_1)] & \cdots & J[D_t^i(g_c)\cdot D_t^{m-1}(g_{p-1})] \\ J[D_t^i(g_c)\cdot D_{t+1}^0(g_0)] & J[D_t^i(g_c)\cdot D_{t+1}^1(g_1)] & \cdots & J[D_t^i(g_c)\cdot D_{t+1}^{m-1}(g_{p-1})] \end{bmatrix}。 \tag{11}$$

在矩阵 $\boldsymbol{M}$ 中，行方向上的值表示每一帧上像素点方向角差值进行编码得到的三进制序列，列方向上的值表示相邻帧间对应方向角差值变化的三进制序列。对于 $\boldsymbol{M}$ 纵向统计字符跳变次数，即 $-1\to-1$、$-1\to0$、$-1\to1$、$0\to-1$、$0\to0$、$0\to1$、$1\to-1$、$1\to0$、$1\to1$ 的跳变次数，借鉴灰度共生矩阵中符角二阶矩(ASM)思想，得到 $\boldsymbol{M}$ 的灰度共生矩阵 $\boldsymbol{G}$，并计算得到 ASM 能量值：

$$\boldsymbol{G}=\begin{Bmatrix} & -1 & 0 & 1 \\ -1 & m_{-1,-1} & m_{-1,0} & m_{-1,1} \\ 0 & m_{0,-1} & m_{0,0} & m_{0,1} \\ 1 & m_{1,-1} & m_{1,0} & m_{1,1} \end{Bmatrix}, \tag{12}$$

$$E_{\mathrm{ASM}}=\sum_{i=0}^{1}\sum_{j=0}^{1}\boldsymbol{G}(i,j)^2, \tag{13}$$

式中 $\boldsymbol{G}$ 为灰度共生矩阵三值表示形式；m_{ij} 为 $i'\to j'$ 的跳变次数，其大小反映了相邻两帧图像局部纹理信息的变化。如果矩阵 $\boldsymbol{G}$ 的值集中分布在对角线上，说明相邻三帧视频序列中纹理变化幅度较大或未产生变化，ASM 值较大；如果正负对角线上分布比较均匀，说明相邻三帧视频序列中纹理发生较小变化，ASM 值较小。

3 决策融合

3.1 云模型及分类器权重计算

为描述表情和姿态两种不同模态对不同类别的识别能力，对每种单模态构建一个局部分类模型，模型的样本数量和多样性皆由训练样本的划分和变化来决定。为了描述样本变化的模糊性和随机性，引入了云模型，并对两种模态局部特征分类器的可靠性进行定性定量分析。云模型是李德毅院士提出的一种定性定量不确定性转换模型[17-19]，解决了从定性概念与其定量数值表示之间的不确定转换问题，通过期望值 E_x、熵 E_n、超熵 H_e 这 3 个数字特征来整体表征一个概念。针对两种模态各个情感类别对识别结果的贡献权重不同，提出一种云加权决策融合算法，分别由两个分类器计算得出权值，实现两种模态逆向云发生器对三维时空特征从定性到定量的转换。

构建逆向正态发生器，期望值 E_x 通过交叉验证训练得到每个情感类别的准确率，将每种模态的所有情感类别得到的准确率作为云滴，通过求取期望值得到；熵 E_n 被用来综合度量每种模态样本的模糊度和概率，反映每种模态样本的离散程度和不确定程度，并通过求云滴的熵得到。超熵 H_e，即熵的熵，是熵的不确定度量，体现每种模态样本不确定度的集聚性。超熵的大小间接地反映了每种模态样本的离散程度。

1) 计算所有云滴的均值：

$$\bar{X}=\frac{1}{N}\sum_{i=1}^{N}x_i。 \tag{14}$$

2) 计算云滴的方差：

$$S^2=\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{X})^2。 \tag{15}$$

3) 计算云滴的期望值：

$$E_x=\bar{X}。 \tag{16}$$

4) 计算云滴的熵：

$$E_n=\sqrt{\frac{\pi}{2}}\times\frac{1}{N}\sum_{i=1}^{N}|x_i-E_x|。 \tag{17}$$

5) 计算超熵：

$$H_e=\sqrt{S^2-E_n}。 \tag{18}$$

为描述两种模态分类器在不同样本空间上的类准确率的变化，利用 Sigmoid 函数 $f(x)=1/[1+\exp(-x)]$，构造每种模态的权值 w_i 为

$$w_i=\frac{1}{1+\exp[-(E_{xi}-(E_{ni}+H_{ei})]}, \tag{19}$$

式中 $i=1,2$，将权值映射到 0,1 之间，形成期望值和熵、超熵的反比关系。

3.2 双模态情感识别方法

SLTOP 特征是从时空特征的角度对情感视频进行描述，通过交叉验证得到训练样本的每种模态不同情感类别的准确率，融合云模型思想，利用 Sigmoid 函数求取每种模态的权值；使用 NN 方法获取待测样本的后验概率，再与 Sigmoid 函数得出的权值相乘，合成得到待测样本所属的情感类别，如图 3 所示。

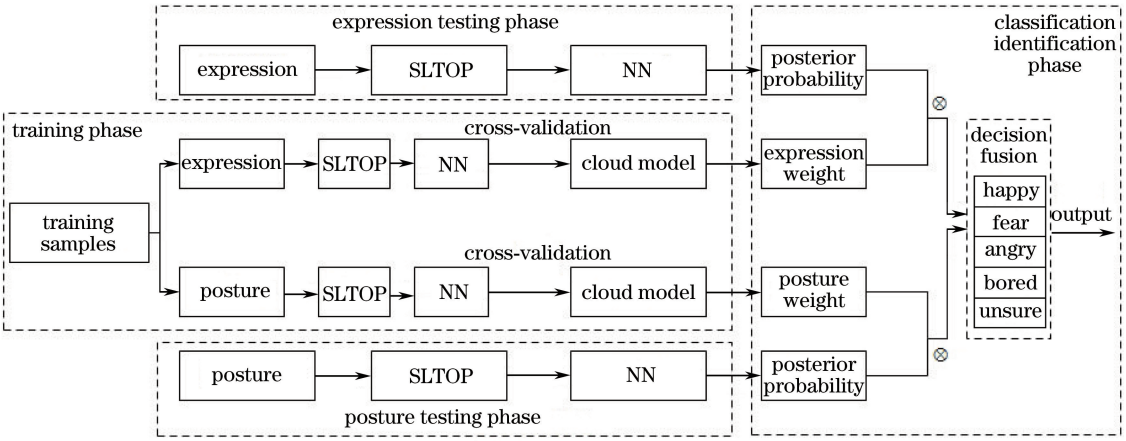

图 3 双模态情感识别系统流程图

Fig. 3 Flow chart of dual-modal emotion recognition

图 4 为面部表情 SLTOP 特征直方图谱，对表情和姿态关键帧图像序列，从前往后选取连续的 7 帧图像，对每一帧图像进行相同大小分块，然后提取中间 3 幅图像的 SLTOP 特征，然后将这 3 幅图像的 SLTOP 特征进行级联获取 SLTOP 总特征，得到表情的三维时空特征图谱。

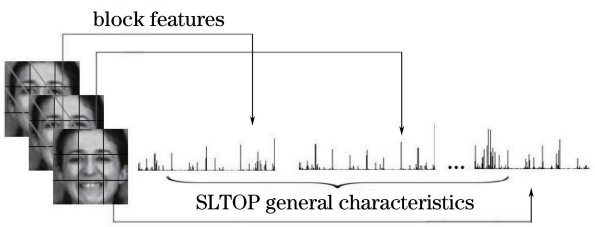

图 4 表情三维时空特征图谱

Fig. 4 Three-dimensional temporal and spatial features of expression

表情和姿态双模态情感识别的具体过程如下。

输入：待测视频样本 X。

输出：待测视频样本 X 的情感类别。

1）预处理就是首先将视频样本转化为序列图片，提取表情序列和姿态序列；然后采用 k 均值聚类（本文取 $k=7$）的办法分别对表情序列和姿态序列图像进行聚类，分别得到 k 帧的图像序列。

2）将聚类得到的图像序列 $T=[t_1,t_2,\cdots,t_k]$，对每一帧的图像进行分块处理，得到大小均匀的局部矩形子块。

3）以姿态图像序列为例，提取中间 3 幅图像每一个子块的 SLTOP 特征，然后将所有子块的 SLTOP 特征进行级联，得到待测样本 X 姿态的 SLTOP 总特征。待测样本 X 的表情图像序列按照和姿态特征提取方法相同的处理办法，得到样本 X 表情的 SLTOP 总特征。

4）按照步骤 1）～3），处理训练样本库的所有视频，分别得到训练样本库的表情和姿态的 SLTOP 特征集。

5）依据 NN 方法对训练样本的每个类别进行交叉验证，得到每个情感类别的准确率，将准确率作为云滴，导入到逆向云模型中，利用(19)式分别得到表情和姿态的各自的权值 w_1、w_2。

6）针对测试样本，同样依据 NN 方法求得待测样本 X 表情序列和姿态序列分别属于不同情感类别的后验概率 p_i^{ex}、p_i^{px}。

7）用表情和姿态各自的权值和不同情感类别的后验概率相乘并分别求和：

$$C_i=w_1p_i^{\mathrm{ex}}+w_2p_i^{\mathrm{px}},(i=1,2,\cdots,N),\quad(20)$$

式中 w_1 为表情模态的权值，w_2 为姿态模态的权值，N 为情感类别（本文取 $N=5$）。取 C_i 最大值，输出最终的情感识别结果。

4 实验与分析

4.1 FABO 双模态数据库

采用的数据库是目前唯一公开的表情和姿态双模态情感数据库——FABO 数据库，其使用摄像机记录多人脸部和姿态情感数据，如图 5 所示。由于现实世界应用中使用面部和身体相机都不实用，所以这里只选择了包含脸部和身体姿态信息的身体相机。

这些视频包括基本和非基本表达。其中，基本表达是“高兴”、“厌恶”、“害怕”、“惊喜”、“悲伤”和“愤怒”。而非基本表达是“焦虑”、“无聊”和“不确定性”等。每个视频包含 2～4 个表达周期。每个表达类别中的视频被随机分成 3 个子集。其中两个被选为训练数据，剩余的子集用作测试数据。对于训练和测试，不再出现相同的视频，但由于随机分离过

程,同一主题可能出现在训练和测试集中。由于该数据库视频未完全标注,实验从脸部和身体手势中各选择 246 个具有相同表情标签的视频,主要包含“高兴”、“害怕”、“生气”、“厌恶”和“不确定”5 类情感,并逐一进行标注。本文实验是在 Windows 7 系统下,基于 VS2015+OpenCV3.0 软件实现。实验中将表情图片和姿态图片归一化为 96 pixel×96 pixel和 128 pixel×96 pixel。

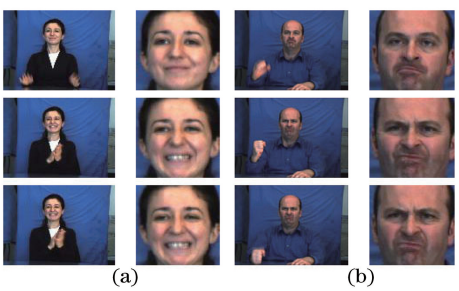

图 5 来自 FABO 数据库表情和姿态的样本。(a)高兴;(b)生气

Fig. 5 Facial expression and body gesture samples from the FABO database. (a) Happy; (b) angry

4.2 单模态情感识别实验

在对表情和姿态两种模态进行决策融合前,首先对表情和姿态两个单模态进行单独的实验。由于图像分块大小也会影响到识别结果,通过实验对比,得到图 6 结果,表情划分为 6 pixel×6 pixel 子块,姿态划分 4 pixel×4 pixel 子块,识别率较高。

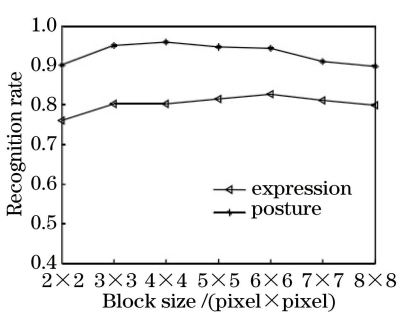

图 6 分块大小与识别率关系图

Fig. 6 Relationship between block size and recognition rate

对面部表情序列和上身姿态序列进行特征提取,得到 SLTOP 直方图特征,并利用得到的特征向量进行训练分类。分别使用 SVM、最小距离分类器和 NN 方法进行分类识别,其中 NN 分类器取得识别率高于其他两个分类器。表 1 和 2 分别给出 NN 方法分类识别时,表情和姿态的混淆矩阵,对角线上的值表示样本的正确识别概率。对于同一分类器,表情和姿态的识别率也有所不同,从识别结果来看,姿态识别率明显高于表情识别率。究其原因,姿态变化幅度大于表情变化幅度,且纹理特征也较为明显,更加有利于判断情感识别的结果。

表 1 表情单模态情感识别混淆矩阵

Table 1 Expression single modal emotion recognition confusion matrix

Emotion	Happy	Fear	Angry	Bored	Unsure
Happy	**0.91**	0.02	0.03	0	0.04
Fear	0.03	**0.92**	0.03	0	0.02
Angry	0.02	0	**0.94**	0.03	0.01
Bored	0	0	0.04	**0.92**	0.04
Unsure	0.04	0.01	0.03	0.01	**0.91**

表 2 姿态单模态情感识别混淆矩阵

Table 2 Posture single modal emotion recognition confusion matrix

Emotion	Happy	Fear	Angry	Bored	Unsure
Happy	**0.96**	0	0.01	0	0.03
Fear	0	**1**	0	0	0
Angry	0	0	**0.97**	0	0.03
Bored	0.05	0	0	**0.95**	0
Unsure	0.01	0	0.04	0	**0.95**

通过多次实验求取识别率的平均值,并作为最终的平均识别率,表 3 所示为本文算法与其他算法识别率的对比结果。由表 3 可知,本文 SLTOP 算法识别率明显优于其他算法识别率,更加适合视频情感识别。

表 4 所示为不同算法在 FABO 数据库的平均识别时间,在表情和姿态两种单模态下,本文 SLTOP 算法识别时间明显优于传统 VLBP 算法,但较三正交平面局部二值模式(LBP-TOP)、时空局部三值模式矩(TSLTPM)+3DHOG 和多尺度时空局部方向角模式(MSLOP)算法的识别时间略有增加。在适当的范围内增加时间,也是为了更加准确地描述局部纹理特征。

表 3 基于单模态的不同特征提取方法的平均识别率比较

Table 3 Comparison of the average recognition rates of different feature extraction methods based on single modal %

Algorithm	Average recognition rate	
	Expression	Posture
MHI-HOG[14]	66.50	66.70
Optical flow method[11]	76.40	89.90
VLBP[6]	73.79	70.17
LBP-TOP[5]	70.17	67.23
TSLTPM+3DHOG[8]	83.06	94.78
MSLOP[9]	87.68	90.81
SLTOP	92.21	96.76

表 4 基于单模态的不同特征提取方法的平均识别时间比较

Table 4 Comparison of the average recognition time for different feature extraction methods based on single modal ms

Algorithm	Average recognition time	
	Expression	Posture
VLBP[6]	423.15	467.85
LBP-TOP[5]	215.77	226.33
TSLTPM+3DHOG[8]	244.22	253.42
MSLOP[9]	281.46	296.23
SLTOP	298.57	308.16

4.3 双模态情感实验

为了论证云加权决策融合的有效性，增加手工设置权值的对比实验。手工设置表情权值(w_1)和姿态权值(w_2)，其中 $w_1+w_2=1$；w_1 从 0 逐渐增加到 1，以 0.1 为步长，图 7 给出了表情和姿态双模态情感平均识别率变化。通过云模型对相同的训练样本进行计算，得到表情权值 $w_1=0.431$ 和姿态权值 $w_2=0.569$，分别与待测样本 X 表情序列和姿态序列不同情感类别的后验概率的乘积求和，取不同情感类别的最大值，得出最终的平均识别率为 97.54%，略高于图 7 中平均识别率的峰值。由于表情和姿态两种模态在情感识别中贡献大小不同，表情变化主要集中在人的脸部，纹理变化范围较小，而姿态主要是上半身动作，纹理变化范围较大。两种情感表达相互补充，将两者进行融合，准确率和可靠性显著提升。

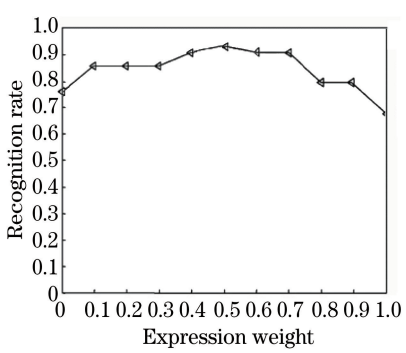

图 7 手工设置表情权值的双模态识别率

Fig. 7 Dual-modal recognition rate of manual set of expression weights

表 5 所示为不同融合方法的平均识别率对比，从表 5 可以明显看出，云加权融合之后的识别率为 97.54%，高于表情单模态的 92.21%和姿态单模态的 96.76%，说明云加权融合比表情单模态和姿态单模态的识别更加准确可靠。

表 5 不同融合方法的平均识别率对比

Table 5 Comparison of the average recognition rates for different fusion methods %

Algorithm	Average recognition rate
BSPLS[15]	65.22
DS evidence[8]	96.86
Proposed method	97.54

5 结 论

针对视频序列中的表情和姿态双模态的情感识别，提出了 SLTOP 特征描述方法和云加权决策融合方法。SLTOP 算子通过自适应阈值，获得三值模式矩阵的能量值，从局部纹理边缘、局部方向角对视频序列进行描述，较全面地表达局部特征，有较好的抗噪性；通过在单模态下与其他算子的对比实验表明：其具有较好的纹理描述性能。考虑表情和姿态两种模态对最终识别结果的作用不同，本文融合云模型的思想，提出融合 Sigmoid 函数的云加权融合方法，分别对两种模态各个情感类别的训练样本进行交叉实验得出识别准确率，通过云模型计算两种模态的权值，对测试样本进行双模态的加权决策融合，得到平均识别率为 97.54%。实验结果表明，云加权融合方法对不同模态的特征信息进行决策融合，获得的识别率高于单模态下的识别率，且与其他文献中的融合方法相比也表明了其可靠性。考虑进一步论述表情和姿态两种模态的纹理特征差异性，下一步工作将在多特征描述、特征融合环节做些研究，并在其他表情和姿态数据库中进行实验，提升视频图像情感识别系统的性能。

参 考 文 献

[1] Xia J, Pei D, Wang Q Z, *et al*. Face recognition based on local adaptive ternary derivative pattern coupled with Gabor feature [J]. Laser & Optoelectronics Progress, 2016, 53(11): 111004.
夏军, 裴东, 王全州, 等. 融合 Gabor 特征的局部自适应三值微分模式的人脸识别[J]. 激光与光电子学进展, 2016, 53(11): 111004.

[2] Hong Y, Sun X X, Wang D, *et al*. Fast pose estimation method for unmanned aerial vehicle based on rectangular geometry feature[J]. Chinese Journal of Lasers, 2016, 43(5): 0508006.
洪洋, 孙秀霞, 王栋, 等. 基于矩形几何特性的小型无人机快速位姿估计方法[J]. 中国激光, 2016, 43(5): 0508006.

[3] Huang Z W, Xue W T, Mao Q R, *et al*.

Unsupervised domain adaptation for speech emotion recognition using PCANet[J]. Multimedia Tools & Applications, 2017, 76(5): 6785-6799.

[4] Tkalčič M, Odić A, Košir A. The impact of weak ground truth and facial expressiveness on affect detection accuracy from time continuous videos of facial expressions[J]. Information Sciences, 2013, 249(16): 13-23.

[5] Zhao G, Pietikäinen M. Dynamic texture recognition using local binary patterns with an application to facial expressions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 915-928.

[6] Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971-987.

[7] Mu Y D, Yan S C, Liu Y, *et al*. Discriminative local binary patterns for human detection in personal album [C] // Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2008: 1-8.

[8] Wang X H, Hou D Y, Hu M, *et al*. Dual-modality emotion recognition based on composite spatio-temporal features [J]. Journal of Image and Graphics, 2017, 22(1): 39-48.
王晓华, 侯登永, 胡敏, 等. 复合时空特征的双模态情感识别[J]. 中国图象图形学报, 2017, 22(1): 39-48.

[9] Fu X F, Fu X J, Li J J, *et al*. Facial expression recognition using multiscale spatiotemporal local orientational pattern histogram projection in video sequences[J]. Journal of Computer-Aided Design & Computer Graphics, 2015, 27(6): 1060-1066.
付晓峰, 付晓鹃, 李建军, 等. 视频序列中基于多尺度时空局部方向角模式直方图映射的表情识别[J]. 计算机辅助设计与图形学学报, 2015, 27(6): 1060-1066.

[10] Gunes H, Piccardi M. A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior [C] // Proceedings of IEEE Internatioual Conference on Pattern Recognition, 2006, 4: 1148-1153.

[11] Gunes H, Piccardi M. Bi-modal emotion recognition from expressive face and body gestures[J]. Journal of Network and Computer Applications, 2007, 30(4): 1334-1345.

[12] Gunes H, Piccardi M. Fusing face and body gesture for machine recognition of emotions[C] // Proceedings of IEEE International Workshop on Robots and Human Interactive Communication, 2005: 306-311.

[13] Gunes H, Piccardi M. Affect recognition from face and body: early fusion vs. late fusion [C] // Proceedings of IEEE International Conference on Systems, Man and Cybernetics, 2005, 4: 3437-3443.

[14] Chen S Z, Tian Y L, Liu Q S, *et al*. Recognizing expressions from face and body gesture by temporal normalized motion and appearance features[J]. Image and Vision Computing, 2013, 31(2): 175-185.

[15] Yan J J, Zheng W M, Xin M H, *et al*. Bimodal emotion recognition based on body gesture and facial expression[J]. Journal of Image and Graphics, 2013, 18(9): 1101-1106.
闫静杰, 郑文明, 辛明海, 等. 表情和姿态的双模态情感识别[J]. 中国图象图形学报, 2013, 18(9): 1101-1106.

[16] Wang Y P, Hu Y H, Lei W H, *et al*. Aircraft target classification method based on texture feature of laser echo time-frequency image[J]. Acta Optica Sinica, 2017, 37(11): 1128004.
王云鹏, 胡以华, 雷武虎, 等. 基于激光回波时频图纹理特征的飞机目标分类方法[J]. 光学学报, 2017, 37(11): 1128004.

[17] Song Y J, Li D Y, Yang X Z, *et al*. Reliability evaluation of electronic products based on cloud models[J]. Acta Electronica Sinica, 2000, 28(12): 74-76.
宋远骏, 李德毅, 杨孝宗, 等. 电子产品可靠性的云模型评价方法[J]. 电子学报, 2000, 28(12): 74-76.

[18] Zeng F X, Li L, Diao X P. Iterative closest point algorithm registration based on curvature features [J]. Laser & Optoelectronics Progress, 2017, 54(1): 011003.
曾繁轩, 李亮, 刁鑫鹏. 基于曲率特征的迭代最近点算法配准研究[J]. 激光与光电子学进展, 2017, 54(1): 011003.

[19] Liu H J, Liu Z, Jiang W L, *et al*. Approach based on cloud model and vector neural network for emitter identification[J]. Acta Electronica Sinica, 2010, 38(12): 2797-2804.
刘海军, 柳征, 姜文利, 等. 基于云模型和矢量神经网络的辐射源识别方法[J]. 电子学报, 2010, 38(12): 2797-2804.

[20] Tan X Y, Triggs B. Enhanced local texture feature sets for face recognition under difficult lighting conditions [J]. IEEE Transactions on Image Processing, 2010, 19(6): 1635-1650.