



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2013
09
VOL.18

ISSN1006-8961
CN11-3758/TB



城市交通规划 P1197



第18卷第9期（总第209期）

2013年9月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

2013 all rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院

主办单位 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 李小文

编辑出版《中国图象图形学报》编辑出版委员会

邮政信箱 北京9718信箱

邮 编 100101

电子信箱 jig@irsa.ac.cn

电 话 010-64807995 010-82614429

网 址 www.cjig.cn

广告经营许可证 京朝工商广字第0361号

总 发 行 北京报刊发行局

订 购 全国各地邮局

海外发行 中国国际图书贸易集团有限公司

（邮政信箱：北京399信箱 邮编：100048）

印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian

Monthly, Started in 1996

Superintended by Chinese Academy of Sciences

Sponsored by Institute of Remote Sensing and Digital Earth, CAS
China Society of Image and Graphics

Institute of Applied Physics and Computational Mathematics

Editor-in-Chief LI Xiaowen

Editor, Publisher Editorial and Publishing Board of Journal of
Image and Graphics

P.O.Box 9718, Beijing, P.R.China

Zip code 100101

E-mail jig@irsa.ac.cn

Telephone 010-64807995 010-82614429

Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals

Domestic All Local Post Offices in China

Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China))

Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB

ISSN 1006-8961

CODEN ZTTXFZ

国外发行代号 M1406

国内邮发代号 82-831

国内定价 50.00元

综述

高分辨率遥感影像空间结构特征建模方法综述

秦昆, 陈一祥, 甘顺子, 冯霞, 任文力

1055

简缩极化SAR数据信息提取与应用

张红, 谢镭, 王超, 张波, 吴樊, 汤益先

1065

图像处理和编码

融合CCCTI码和谱系聚类的仿射配准

张桂梅, 江少波, 刘丕玉, 张松

1074

面向虚拟视点绘制的深度图滤波及上采样方法

高凯, 陈贺新, 赵岩, 耿英楠, 王钢

1085

图像分析和识别

包空间多示例图像自动分类

王科平, 杨艺, 王新良

1093

表情和姿态的双模态情感识别

闫静杰, 郑文明, 辛明海, 邱伟

1101

等级关联的粒子群迭代多目标检测跟踪

程旭, 李拟珺, 张索非, 吴镇扬

1107

改进后的TLD视频目标跟踪方法

周鑫, 钱秋朦, 叶永强, 王从庆

1115

多特征稳健主成分分析的视频运动目标分割

甘超, 王莹, 王向阳

1124

融合全局最好和声搜索算法的模糊C均值聚类图像分割

崔兆华, 高立群, 欧阳海滨, 李文娜

1133

从运动中恢复目标结构的改进因子分解法

邱少华, 文贡坚, 回丙伟, 张鹏

1142

计算机图形学

并行扫描转换结构中的状态管理

殷诚信, 韩俊刚, 黄虎才

1149

医学图像处理

3维颅骨表面模型的复杂孔洞修补

包佳蕊, 梁荣华, 吴福理, 黄鲜萍, 蒋莉, 陈矛

1156

归一化Radon变换和统计测度的超声图像肌束方向自动跟踪

李乔亮, 李振宇, 张会生, 张小飞, 陈昕, 陈思平, 汪天富

1164

遥感图像处理

改进的多光谱双边滤波图像融合

吴川, 杨冬

1170

长度归一化扫描的合成孔径雷达图像船舶尾迹恒虚警率检测

南杰, 张红, 王超, 吴樊, 张波

1176

地理信息技术

3D可视化输电线路中杆塔空间挂点计算

何国辉, 周聪

1184

地图符号服务与地图服务的耦合

焦东来, 张海涛, 闫国年, 姜杰

1190

不同表达粒度对城市路网结构健壮性评价的影响

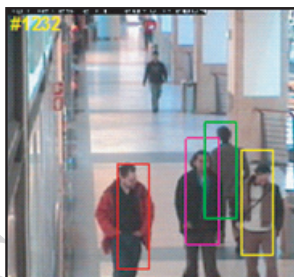
段滢滢, 陆锋

1197

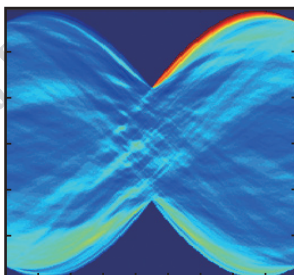
DEM保真精度量化描述模型

张锦明, 游雄

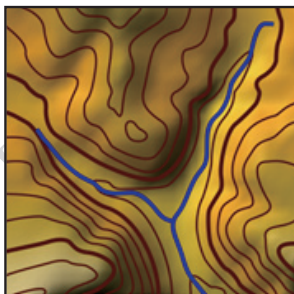
1206



等级关联的粒子群迭代多目标检测跟踪(第1107页)



归一化Radon变换和统计测度的超声图像肌束方向自动跟踪(第1164页)



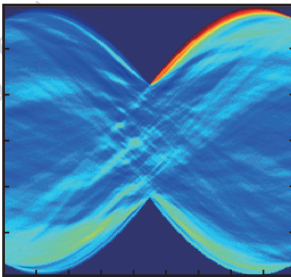
DEM保真精度量化描述模型(第1206页)

CONTENTS

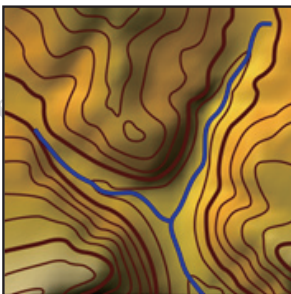
JOURNAL OF IMAGE AND GRAPHICS



Hierarchical associative multi-target detection and tracking with particle swarm optimization(P1107)



Automatic tracking of muscle fascicle orientation in ultrasound images: a approach based on a normalized Radon transform and statistical measure (P1164)



Quantitative descriptive model of DEM fidelity accuracy(P1206)

Review

Review on methods of spatial structural feature modeling of high resolution remote sensing images

Qin Kun, Chen Yixiang, Gan Shunzi, Feng Xia, Ren Wenli 1055

Information extraction and application of compact polarimetric SAR data

Zhang Hong, Xie Lei, Wang Chao, Zhang Bo, Wu Fan, Tang Yixian 1065

Image Processing and Coding

Affine registration based on CCCTI and hierarchical clustering

Zhang Guimei, Jiang Shaobo, Liu Piyu, Zhang Song 1074

Depth map filtering and upsampling method for virtual view rendering

Gao Kai, Chen Hexin, Zhao Yan, Geng Yingnan, Wang Gang 1085

Image Analysis and Recognition

Automatic classification of multiple-instance image based on the bag space

Wang Keping, Yang Yi, Wang Xinliang 1093

Bimodal emotion recognition based on body gesture and facial expression

Yan Jingjie, Zheng Wenming, Xin Minghai, Qiu Wei 1101

Hierarchical associative multi-target detection and tracking with particle swarm optimization

Cheng Xu, Li Nijun, Zhang Suofei, Wu Zhenyang 1107

Improved TLD visual target tracking algorithm

Zhou Xin, Qian Qiumeng, Ye Yongqiang, Wang Congqing 1115

Multi-feature robust principal component analysis for video moving object segmentation

Gan Chao, Wang Ying, Wang Xiangyang 1124

Improved fuzzy C-means clustering combined with the global best harmony search algorithm for image segmentation

Cui Zhaozhuo, Gao Liqun, Ouyang Haibin, Li Wenna 1133

Improved factorization method for structure from motion

Qiu Shaohua, Wen Gongjian, Hui Bingwei, Zhang Peng 1142

Computer Graphics

Efficient state management for parallel rasterization

Yin Chengxin, Han Jungang, Huang Hucai 1149

Medical Image Processing

Various hole repairing algorithm for 3D mesh surface models of human skull

Bao Jiarui, Liang Ronghua, Wu Fuli, Huang Xianping, Jiang Li, Chen Mao 1156

Automatic tracking of muscle fascicle orientation in ultrasound images: a approach based on a normalized Radon transform and statistical measure

Li Qiaoliang, Li Zhenyu, Zhang Huisheng, Zhang Xiaofei, Chen Xin, Chen Siping, Wang Tianfu 1164

Remote Sensing Image Processing

Improved multispectral bilateral filter video fusion algorithm

Wu Chuan, Yang Dong 1170

Ship wake constant false alarm rate detection algorithm in synthetic aperture radar images based on length normalized scanning

Nan Jie, Zhang Hong, Wang Chao, Wu Fan, Zhang Bo 1176

Geoinformatics

Tower hanging point calculation for transmission line on three-dimensional visualization

He Guohui, Zhou Cong 1184

Coupling of map symbol web service and web map service

Jiao Donglai, Zhang Haitao, Lv Guonian, Jiang Jie 1190

The impact of different granularity representations on robustness evaluation of city road network

Duan Yingying, Lu Feng 1197

Quantitative descriptive model of DEM fidelity accuracy

Zhang Jinming, You Xiong 1206

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2013)09-1101-06

论文引用格式: 闫静杰, 郑文明, 辛明海, 邱伟. 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. [DOI: 10.11834/jig.20130906]

表情和姿态的双模态情感识别

闫静杰¹, 郑文明², 辛明海², 邱伟²

1. 东南大学信息科学与工程学院, 南京 210096; 2. 东南大学学习科学研究中心, 南京 210096

摘要: 多模态情感识别是当前情感计算研究领域的重要内容, 针对人脸表情和动作姿态开展双模态情感识别研究, 提出一种基于双边稀疏偏最小二乘的表情和姿态的双模态情感识别方法。首先, 从视频图像系列中分别提取表情和姿态两种模态的空时特征作为情感特征矢量。然后, 通过双边稀疏偏最小二乘(BSPLS)的数据降维方法来进一步提取两组模态中的情感特征, 并组合成新的情感特征向量。最后, 采用了两种分类器来进行情感的分类识别。以国际上广泛采用的 FABO 表情和姿态的双模态情感数据库为实验数据, 并与多种子空间方法(主成分分析、典型相关分析、偏最小二乘回归)进行对比实验来评估本文方法的识别性能。实验结果表明, 两种模态融合后相比单模态更加有效, 双边稀疏偏最小二乘(BSPLS)算法在几种方法中得到最高的情感识别率。

关键词: 表情; 姿态; 双模态情感识别; 空时特征; 双边稀疏偏最小二乘(BSPLS)

Bimodal emotion recognition based on body gesture and facial expression

Yan Jingjie¹, Zheng Wenming², Xin Minghai², Qiu Wei²

1. School of Information Science and Engineering, Southeast University, Nanjing 210096, China

2. Research Center for Learning Science, Southeast University, Nanjing 210096, China

Abstract: Multimodal emotion recognition has been a very important research topic in affect computing. This paper mainly focuses on the methods of bimodal emotion recognition based on body gesture and facial expression and presents a new bimodal emotion recognition method based on bilateral sparse partial least squares (BSPLS). First, the spatio-temporal feature is extracted as the emotion feature vector for video-based body gesture and facial expression respectively. Then we propose a new bilateral sparse partial least squares (BSPLS) method to extract emotion feature and fuse facial expression and body gestures as new emotion feature. Finally, we use two classifiers in emotional classification. We compared the BSPLS method with some subspace methods including PCA, CCA and PLSR based on the data from the FABO database. The experimental results show that the fusion feature methods are all better than the monomodal emotion recognition and our BSPLS feature fusion provides the best recognition performance.

Key words: body gesture; facial expression; bimodal emotion recognition; spatio-temporal feature; bilateral sparse partial least squares (BSPLS)

收稿日期: 2012-12-18; 修回日期: 2013-02-27

基金项目: 国家自然科学基金项目(61231002, 61073137); 教育部博士点基金项目(20120092110054)

第一作者简介: 闫静杰(1983—), 男, 东南大学信息科学与工程学院博士研究生, 主要研究方向为模式识别、计算机视觉等。E-mail: yanjingjie1212@163.com

通讯作者: 郑文明, E-mail: wenmin_zheng@seu.edu.cn

0 引言

近年来,随着信息等技术的快速发展,如何实现自然、智能、和谐的人机交互引起很多研究者的兴趣和重视。以往的人际交互模式,大多是采用键盘和鼠标等中介方式,没有具备情感能力,也很难实现和谐与自然的人机交互。智能情感识别是完善自然、和谐人机交互的核心技术^[1]。以往的大部分情感识别研究工作主要集中在人脸表情识别、情感语音识别和姿态情感识别等单模态情感识别^[2-5]。但是单模态情感识别有一定的局限性,因为人类的情感表达并不仅仅通过单种模态,而是多种模态情感信息融合的方式。研究表明,多模态情感识别比单模态情感识别更加准确可靠^[6]。

国内外对单模态情感识别已经做了很多的研究工作。相比表情识别和语音情感识别,姿态动作的情感识别还比较少。但是研究结果显示,姿态动作也是一种重要的表达人脸情感的方式^[7]。Coulson 等人^[8]最先研究了基于姿态动作的情感识别,结果表明基于姿态的情感识别和基于语音的情感识别在人类表达情感上的作用相差不大,某些姿态动作的作用跟人脸表情的作用相当。Camurri 等人^[9]研究了基于舞蹈动作的情感识别。Ravindra 等人^[10]设计了基于姿态的自动情感识别系统用来识别游戏中的儿童们的各种情感状态。Ambady 和 Rosenthal^[11]同时研究了人脸表情和身体姿态,当同时观察人脸表情和身体姿态时,能更加准确可靠地判定人的行为状态。

在表情识别和身体姿态情感识别等单模态的情感识别基础上,已经有一些研究者进行了表情和姿态的双模态情感识别研究。Kapoor 等人^[12]通过融合视频中人脸表情信息和身体姿态动作信息等来识别游戏中的儿童们的各种情感状态。Shan 等人^[13]进行了基于空时特征的双模态情感识别研究,他们通过提取视频数据中的姿态和表情的空时特征情感特征向量并使用典型相关分析方法融合。Chen 等人^[14]针对目前实时自动情感识别中表情和姿态特征提取问题,提出了联合 MHI-HOG 和 Image-HOG 的时间归一化方法来有效描述动态表情和姿态情感识别的过程。Gunes 和 Piccard 在基于姿态和表情的多模态情感识别研究领域做了一系列工作^[15-18]。他们建立了基于表情和姿态的 FABO 双模态数据

库^[15],分别用 Ekman 的人脸面部运动单元^[19]和身体运动单元(BAU)的来表示人脸表情和身体姿态变化与情感的关系。在 FABO 数据库上,他们比较了人脸表情和身体姿态两种模态分别采用特征层融合和决策层融合时候的效果^[16-18]。实验结果表明,两种融合方式都能得到比单模态识别时候更好的实验结果,而且基于特征层融合方式的实验结果要比基于决策层融合方式更加有效。

本文研究了基于视频的表情和姿态双模态情感识别,提出了一种基于双边稀疏偏最小二乘的表情和姿态的双模态情感识别方法。首先,分别对姿态和表情两种单模态的视频数据提取空时特征,完成两种单模态的情感识别实验。然后针对视频数据中可能存在冗余信息和噪声,通过双边稀疏偏最小二乘(BSPLS)算法对降维后的两组单模态的特征向量进行特征选择并融合,并与主成分融合、典型相关分析融合、偏最小二乘融合等方法进行了比较。最后,使用最近邻分类器和支持向量机作为分类器进行分类识别。

1 双模态情感特征提取

双模态情感特征提取包括表情特征提取和姿态特征提取两部分。进行双模态融合前,先分别提取人脸表情和身体姿态的情感特征。特征提取方法采用空时特征^[20]。空时特征通过计算视频数据中兴趣点的响应函数局部最大值得到。

空时特征的提取首先通过原始视频数据与预先定义的响应函数进行卷积,响应函数一般定义为^[20]

$$L = (H * g * h_{ev})^2 + (H * g * h_{od})^2 \quad (1)$$

式中, H 是需要提取空时特征的原始视频数据, g 是高斯包络函数 $g(x, y; \sigma)$,其定义为

$$g(x, y; \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (2)$$

h_{ev} 和 h_{od} 是 Gabor 滤波器的一组正交函数,它们的具体形式为

$$h_{ev}(t; \tau, \omega) = -\cos(2\pi t\omega) e^{-t^2/\tau^2} \quad (3)$$

$$h_{od}(t; \tau, \omega) = -\sin(2\pi t\omega) e^{-t^2/\tau^2} \quad (4)$$

通过响应函数,对相应的原始视频数据进行卷积可以求得视频中每个像素点的响应函数值,并通过比较响应函数值的大小选取一定数量的兴趣点。求得兴趣点后提取兴趣点周围的立方体并进一步地处理就可以得到最后的特征向量。

空时特征的提取流程主要包括以下 4 个步骤^[20]:检测视频数据中的兴趣点并提取兴趣点周围的立方体、计算立方体的描述子、建立立方体原型类别库、计算行为描述子。

经过上述步骤,可以得到一个低维的空时特征向量来代表原有的视频数据,其中每个视频数据对应一个低维空时特征向量,这个空时特征向量是一个直方图向量,直方图向量的维数是第 3 步中建立的立方体的原型类别库的个数。

尽管每个视频数据中提取的立方体数量很大而且每个视频所对应的立方体数量不同,但立方体们所属的类别数却是有限的。因此为了识别视频中的表情和姿态情感类别,可以通过计算每个视频中大量立方体在各个类别中的数量来区分不同的情感状态。计算出来的立方体在各个类别中的数量就是最后的空时特征直方图向量。

空时特征直方图向量中包含了视频中的表情或者姿态信息,空时特征直方图向量是对视频中具体表情或者姿态行为的一种低维描述,不同情感状态的原始视频数据,其差异会体现到这个低维空时特征直方图向量里。图 1 与图 2 分别为 FABO 数据库中的表情和姿态数据提取空时特征后的效果图。

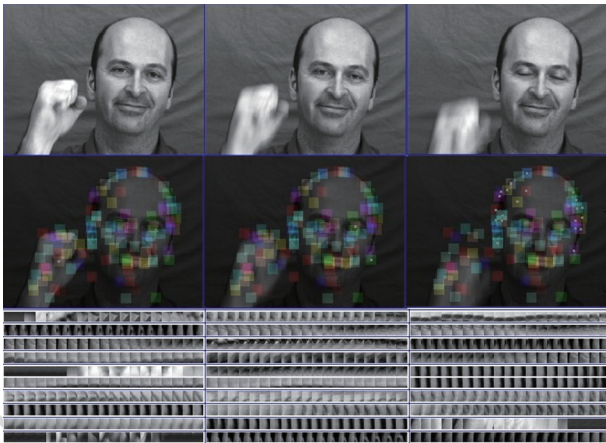


图 1 姿态样本空时特征提取效果图

Fig. 1 Examples of spatio-temporal features extracted from the body gesture samples

图 1 中第 1 行是原始表情视频的部分帧图像。第 2 行是表情视频中经过兴趣点检测后得到的立方体示意图。立方体是以检测得到的兴趣点为中心,在空间和时间维度上提取预设长度的立方体。提取的立方体能够包括绝大多数对响应函数局部最大值有贡献的像素区域。第 3 行是提取的部分立方

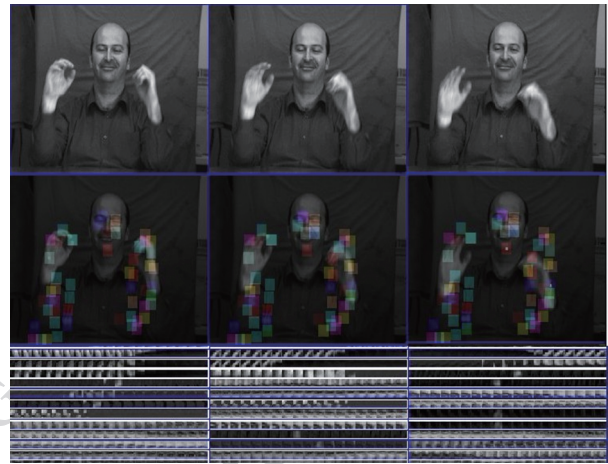


图 2 表情样本空时特征提取效果图

Fig. 2 Examples of spatio-temporal features extracted from the facial expression samples

体数据按时间展开的示意图。

图 2 跟图 1 类似,从图 1 和 2 中可以看到,对于表情和身体姿态视频数据,提取的立方体绝大部分都集中在表情和动作发生的区域。通过上述两图,可以观察到空时特征提取的兴趣点的位置与人眼观察到的表情和动作发生区域非常接近,这也证明了空时特征方法在提取视频数据中表情和动作信息的有效性和合理性。因此本文采用空时特征方法来提取原始视频数据中的人脸表情和身体姿态情感信息。

2 基于双边稀疏偏最小二乘的表情和姿态双模态情感融合方法

2.1 偏最小二乘 (PLS)

偏最小二乘回归是一种新型的多元统计数据分析方法,它将回归建模(多元线性回归)、数据结构简化(主成分分析)和两组变量之间的相关性分析(典型相关分析)的基本功能结合起来,得到了较好的建模效果,其广泛应用在化学、计量学、教育评测学、生物信息学、统计学、医学、模式识别等领域^[21]。

设 $X_{N \times m}$ 、 $Y_{N \times n}$ 为两组零均值的随机变量矩阵,PLS 的目的就是找到一组投影方向 u 和 v ,使得 Xu 和 Yv 之间具有最大的协方差。

假设共提取 T 个潜变量,PLS 可以归结为解决最优化问题

$$\begin{aligned} \max \operatorname{cov}(X_{r-1}u_r, Y_{r-1}v_r) &= \max u_r^T X_{r-1}^T Y_{r-1} v_r \\ \text{s. t. } \|u_r\| &= 1, \|v_r\| = 1, r = 1, 2, \dots, t \end{aligned} \quad (5)$$

通过拉格朗日乘子法求解上述优化问题,最后

可得

$$X_{r-1}^T Y_{r-1} Y_{r-1}^T X_{r-1} u_r = \lambda^2 u_r \quad (6)$$

$$v_r = \frac{1}{\lambda} Y_{r-1}^T X_{r-1} u_r \quad (7)$$

u_r 就是矩阵 $X_{r-1}^T Y_{r-1} Y_{r-1}^T X_{r-1}$ 的最大特征值所对应的特征向量。求得 u_r 后可以通过式(7)求得 v_r 。

2.2 双边稀疏偏最小二乘 (BSPLS)

通过上述 PLS 方法求得的投影向量 u_r 和 v_r 是不稀疏的,而视频中自带的冗余信息和噪声会影响识别效果,因此下面通过奇异值分解的方法引入双边稀疏偏最小二乘方法来求得稀疏的投影向量^[22]。

设 $M_r = X_{r-1}^T Y_{r-1}$, $\text{rank}(M_r) = h$, 对 M_r 进行奇异值分解有

$$M_r = A_r D_r B_r^T \quad (8)$$

式中, A_r 和 B_r 第 1 列向量分别为待求的 u_r 和 v_r 。

定义并改写矩阵 M_r 的秩 -1 近似矩阵

$$\|M - d u v^T\|_2^2 = \|M - \tilde{u} \tilde{v}^T\|_2^2 \quad (9)$$

式中, d 为 M_r 最大的奇异值。

为了使 u 和 v 都能达到稀疏的要求,可以通过式(9)中附加两个 LASSO 惩罚函数。因此,最优化问题就变为使式(10)取得最小值^[22]。

$$\|M - \tilde{u} \tilde{v}^T\|_2^2 + \lambda_1 \sum_{i=1}^m |\tilde{u}_i| + \lambda_2 \sum_{j=1}^n |\tilde{v}_j| \quad (10)$$

通过以下的迭代更新算法,可以求得稀疏投影向量 u 和 v 。

1) 令 $\tilde{v} = B_1$, 然后展开式(10)得

$$\sum_j (m_{ij} - \tilde{u}_i \tilde{v}_j)^2 = \sum_j m_{ij}^2 - 2(M\tilde{v})_i \tilde{u}_i + \tilde{u}_i^2 \quad (11)$$

可求得 $\tilde{u}$ 的值为

$$\tilde{u} = \text{sgn}(M\tilde{v}) \left(|M\tilde{v}| - \frac{\lambda_1}{2} \right)_+ \quad (12)$$

2) 令 $\tilde{u} = A_1$, 然后展开式(10)得

$$\sum_j (m_{ij} - \tilde{u}_i \tilde{v}_j)^2 = \sum_j m_{ij}^2 - 2(M^T \tilde{u})_i \tilde{v}_i + \tilde{v}_i^2 \quad (13)$$

可求得 $\tilde{v}$ 的值为

$$\tilde{v} = \text{sgn}(M^T \tilde{u}) \left(|M^T \tilde{u}| - \frac{\lambda_2}{2} \right)_+ \quad (14)$$

上述求解过程通过迭代更新最后收敛求得稀疏投影向量 u 和 v 。

2.3 表情和姿态特征融合

假如求得两组稀疏投影向量 u_i 和 v_i , 因此可以得到 $T = X[u_1, \dots, u_d]$ 和 $C = Y[v_1, \dots, v_d]$, 其中 X 和 Y 分别代表表情和姿态向量。因此可以得到融

合后的表情和特征向量

$$Z = \begin{pmatrix} T \\ C \end{pmatrix} = \begin{pmatrix} X[u_1, \dots, u_d] \\ Y[v_1, \dots, v_d] \end{pmatrix} \quad (15)$$

最后把融合后的特征作为下一步识别时的特征向量。

3 实验与分析

3.1 FABO 双模态情感数据库

采用 FABO 身体姿态和人脸表情的双模态情感数据库进行实验^[15], FABO 数据库是目前唯一公开的表情和姿态双模态情感数据库。由于 FABO 数据库只是部分标注,并且各个人的样本和情感类别数目不一样,所以选择其中样本数目较多且样本情感类别比较均匀的 10 个人的数据作为实验样本,并选择厌烦,厌恶,高兴和不确定 4 类情感,其中姿态和表情各取 89 个样本。本文的实验使用 10 折交叉验证。最后的识别率为 10 次实验的平均值。

3.2 单模态情感识别实验

单模态情感识别实验,分别针对姿态和表情两个单模态的数据进行。首先分别对姿态和表情的视频样本提取其空时特征,然后对得到的空时特征向量分别进行分类。表 1 显示了在 FABO 数据库上采用 SVM 分类器和 KNN 分类器(采用欧氏距离,文献[23]采用的是卡方距离)时表情和姿态两种单模态情感识别的平均识别率。图 3 和图 4 分别为采用 SVM 分类器时表情和姿态对应的混淆矩阵。

表 1 采用 SVM 分类器和最近邻分类器时
表情和姿态两种单模态情感识别的平均识别率

Table 1 The average recognition rate of facial expression and body gesture using the SVM classifier and the KNN classifier

模态	平均识别率/%
表情 + SVM	67.42
姿态 + SVM	71.90
表情 + KNN	52.80
姿态 + KNN	51.69

从姿态和表情两种单模态情感识别的实验结果来看,使用 SVM 分类器,姿态单模态的识别效果要好于表情单模态的识别效果。其原因可能是,对于 FABO 的数据样本,姿态动作变化较大,空时特征更

容易提取有用的动作判别信息, 而人脸表情的变化比较微弱, 相比姿态动作较难区分。

厌烦	0.82	0.18	0.00	0.00
厌恶	0.24	0.60	0.12	0.04
高兴	0.00	0.28	0.50	0.22
不确定	0.00	0.22	0.06	0.72
	厌烦	厌恶	高兴	不确定

图 3 表情单模式识别的混淆矩阵

Fig. 3 Confusion matrices of body gesture recognition

厌烦	0.75	0.21	0.04	0.00
厌恶	0.20	0.68	0.04	0.08
高兴	0.06	0.00	0.83	0.11
不确定	0.11	0.22	0.06	0.61
	厌烦	厌恶	高兴	不确定

图 4 姿态单模式识别的混淆矩阵

Fig. 4 Confusion matrices of facial expression recognition

空时特征相比于其他算法, 在特征提取时候不需要人工预先标定特征点, 不需要其他分割定位或者跟踪等算法, 而且对使用的视频数据不需要背景和遮挡等先验的假设, 而且原始视频数据空时特征经过提取后得到直方图向量的维数远远低于原始视频数据的维数, 所以复杂度相对于其他需要定位和跟踪的算法较小, 例如 Gunes^[15]的方法就需要预先标定特征点, 并需要分割和跟踪, 所以复杂度很高。空时特征在计算复杂度方面有优势, 而且简单易用。

3.3 双模态情感融合识别实验

3.2 节实验已经通过表情和姿态单模态情感识别获得了单模态的识别结果和各单模态情感特征向

量, 这里将进行姿态和表情的情感特征融合实验, 分别使用 PCA 降维融合方法、CCA 降维融合方法, 偏最小二乘 (PLS) 融合和双边稀疏偏最小二乘 (BSPLS) 融合 4 种方法进行融合。

由于表情和姿态两种模态对情感影响不同会对融合的结果造成不同的影响, 因此在融合前我们先观察姿态和表情在不同权重情况下对最后融合识别率的影响。假设姿态的权重为 r , 则表情的权重为 $1-r$, 取 r 为 0.2、0.4、0.5、0.6 和 0.8, 采用 PCA 降维后融合方法, 在使用 SVM 分类器时得到融合后的识别率分别为 69.56、70.26、71.05、72.91 和 70.81。在 r 为 0.6 时识别率最高, 因此在下面的融合实验中令权重 r 为 0.6。

表 2 显示了在 FABO 数据库上采用 SVM 分类器和 KNN 分类器时 4 种融合方法的平均识别率。

表 2 采用 SVM 分类器时 4 种融合方法的平均识别率

Table 2 The average recognition rate of four fusion methods using the SVM classifier and the KNN classifier

分类器	融合方法	平均识别率/%
SVM	PCA	72.91
	CCA	74.06
	PLS	73.03
	BSPLS	76.40
KNN	PCA	59.55
	CCA	63.67
	PLS	61.80
	BSPLS	65.22

实验结果表明, 在使用 SVM 分类器和最近邻分类器的情况下, 各种融合方法的识别率都超过了对应的两个单模态的识别率, 说明了融合姿态和表情后进行情感识别的有效性。比较几种融合方法的识别效果可以看出, 双边稀疏偏最小二乘 (BSPLS) 算法得到最高的情感识别率, 说明通过稀疏视频数据中可能存在的冗余信息和噪声, 更好地提取视频数据中的判别信息, 取得更好的分类效果。此外, 双边稀疏偏最小二乘方法融合身体姿态与人脸表情进行情感识别, 比单独使用人脸表情进行情感识别时提高了十几个百分点。

4 结 论

本文研究了基于姿态和表情的双模态情感识别

并提出了一种基于双边稀疏偏最小二乘的表情和姿态的双模态情感融合识别方法。首先采用空时特征提取视频中的表情和姿态两种模态的情感特征向量。然后针对视频数据中可能存在冗余信息和噪声,通过 BSPLS 算法对降维后的两组单模态的特征向量进行特征选择并融合。实验表明,BSPLS 算法在几种融合方法中得到最高的情感识别率,BSPLS 融合方法能更好地提取视频数据中的判别信息,取得更好的分类效果。但本文方法并没有利用两种模态的类别信息,如何发挥双模态情感识别中的类别信息需要更一步的研究。

参考文献 (References)

- [1] Zeng Z, Pantic M, Roism G, et al. A survey of affect recognition methods: audio, visual and spontaneous expressions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(1): 39-58.
- [2] Valstar ME, Jiang BH, Mehu M, et al. The first facial expression recognition and analysis challenge[C]//Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition. Washington: IEEE Computer Society, 2011: 314-321.
- [3] Bai X, Pei Y, Ma L, et al. Automatic facial expression recognition using gabor filter and expression analysis[C]//Proceedings of International Conference on Computer Modeling and Simulation. Washington: IEEE Computer Society, 2010, 2: 215-218.
- [4] Chena S, Tiana Y, Liu Q, et al. Recognizing expressions from face and body gesture by temporal normalized motion and appearance features[J]. Image and Vision Computing, 2012, 31(2): 175-185.
- [5] Ayadi M, Kamel M, Karray F. Survey on speech emotion recognition: features, classification schemes, and databases[J]. Pattern Recognition, 2011, 44(3): 572-587.
- [6] Hudlicka E. To feel or not to feel: the role of affect in human-computer interaction[J]. Int. J. Hum.-Comput. Stud., 2003, 59(1-2): 1-32.
- [7] Meeren H, Heijnsbergen C, Gelder B. Rapid perceptual integration of facial expression and emotional body language[C]//Proceedings of the National Academy of Sciences of USA. Washington: United States National Academy of Sciences, 2005, 102(45): 16518-16523.
- [8] Coulson M. Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence[J]. Journal of Nonverbal Behavior, 2004, 28(2): 117-139.
- [9] Camurri A, Lagerlöf I, Volpe G. Recognizing emotion from dance movement: comparison of spectator recognition and automated techniques[J]. International Journal of Human-Computer Studies, 2003, 59(1-2): 213-225.
- [10] Silva P, Osano M, Marasinghe A. Towards recognizing emotion with affective dimensions through body gestures[C]//Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition. Washington: IEEE Computer Society, 2006, 269-274.
- [11] Ambady N, Rosenthal R. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis[J]. Psychological Bulletin, 1992, 111(2): 256-274.
- [12] Kapoor A, Picard R. Multimodal affect recognition in learning environments[C]//Proceedings of the 13th Annual ACM International Conference on Multimedia. New York: ACM 2005. 677-682.
- [13] Shan C, Gong S, G MOWan P W. Beyond facial expressions: Learning human emotion from body gestures[C]//Proceedings of Brit. Mach. Vis. Conf.: Manchester British: BMVA Press, 2007: 43(1-10).
- [14] Chen S, Tian Y, Liu Q, et al. Recognizing expressions from face and body gesture by temporal normalized motion and appearance features[C]//Proceedings of 24th IEEE Conference on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 2011. 7-12.
- [15] Gunes H, Piccardi M. A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior [C]//Proceedings of the 18th International conference on pattern Recognition. Washington: IEEE Computer Society, 2006, 1: 1148-1153.
- [16] 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.
- [17] Gunes H, Piccardi M. Fusing face and body gesture for machine recognition of emotions [C]//Proceedings of IEEE International Workshop Robots and Human Interactive Communication. Washington: IEEE Computer Society, 2005: 306-311.
- [18] Gunes H, Piccardi M. Affect recognition from face and body: early fusion vs. late fusion [C]//Proceedings of IEEE SMC. Washington: IEEE Computer Society, 2005, 4: 3437-3443.
- [19] Ekman P, Friesen W. Facial Action Coding System: A Technique for the Measurement of Facial Movement[M]. Palo Alto: Consulting Psychologists Press, 1978.
- [20] Ryo M, Aggarwal J. Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities [C]//Proceedings of International Conference on Computer Vision. Washington: IEEE Computer Society, 2009(5): 1593-1600.
- [21] Krishnan A, Williams L J, McIntosh AR, et al. Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review[J]. Neuroimage, 2011, 23(2): 302-332.
- [22] Cao K, Rossouw D. A sparse PLS for variable selection when integrating omics data[J]. Statistical Applications in Genetics and Molecular Biology, 2008, 5(7): 389-422.
- [23] Yan J, Xin M. Facial expression recognition based on fused spatio-temporal features [C]//Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering. Paris, France: Atlantis Press, 2013: 2113-2116.