BHU人脸表情数据库的设计与实现

薛雨丽 毛 峡 张 帆

(北京航空航天大学 电子信息工程学院, 北京 100083)

摘 要: 提供了一个较为全面的人脸表情数据库,它包含 3类人脸表情:单一表情、混合表情和复杂表情. 本数据库具有在其它人脸数据库中未曾出现的情感表情,对世界人脸表情数据库进行了有益的补充. 对数据库单一表情的部分样本进行了问卷调查,给出了对每个样本的评价,并通过统计分析给出了对数据库中单一表情可识别性的评价. 通过对 9种单一表情的识别实验,说明了多类表情识别的可行性. 本表情数据库对人脸表情识别和人机交互领域的研究具有积极的促进作用.

关 键 词: 情感表情; 人脸表情数据库; 表情识别

中图分类号: TP 391. 41

文献标识码: A **文 章 编 号**: 1001-5965(2007)02-0224-05

Design and realization of BHU facial expression database

Xue Yuli Mao Xia Zhang Fan

(School of Electronics and Information Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: A comprehensive facial expression database was provided, which includes three kinds of facial expressions: pure facial expressions, mixed facial expressions and complex facial expressions. The database contains some unique emotional facial expressions not appeared in other facial databases, and is benefit to the world facial expression databases. Questionnaire investigation for some videos of pure facial expressions was done to give the evaluation of every video. The evaluation of each pure facial expression data was obtained through statistical manner. Facial expression recognition experiment on nine pure facial expressions demonstrated the feasibility of recognizing multiple facial expressions. The database is helpful to the development of facial expression recognition technologies and emotional human-machine interaction systems.

Key words: emotional facial expressions; facial expression database; facial expression recognition

人脸表情含有丰富的人体行为信息,在人们的日常生活中起着重要的作用,对它研究可以进一步了解人类对应的心理状态,人脸表情识别的研究在情感计算和智能交互领域有重要意义. 由于人脸表情非常复杂,而且受种族、文化等差异的影响,目前大多数人脸表情识别的研究仅限于少数典型的表情[1]. 然而情感化的智能交互要求机器具有接近人类的表情识别能力,为了实现该目标,就需要有丰富、有效的人脸表情数据,所以收集一个高质量的人脸表情数据库就成为一项至关重要的任务.

目前的人脸表情数据库大致分为 2类[2]. 一类为基于美国心理学家 Ekman提出的 6种基本情感[3] (分别为:高兴、悲伤、惊讶、生气、嫌恶和害怕)的数据库. 例如,马里兰大学人脸表情数据库[4]和日本女性人脸表情 (JAFFE, Japanese Female Facial Expression)数据库[5]均包含了上述 6种基本表情,后者还包含了中性表情. 而另一类人脸表情数据库集中于提取人脸更加细微的表情. 例如,基于 Ekman提出的面部动作编码系统 (FACS, Facial Action Coding System)的 AU-Coded人脸表情数据库[6]中包含了单一运动单元和复合运动单元共 23种表情. 此外,还有一些数据库包含了部分非基本表情,例如, Yale表情数据

收稿日期: 2006-03-08

基金项目: 国家自然科学基金资助项目 (60572044); 国家高技术研究发展计划资助项目 (2006AA01Z135)

作者简介: 薛雨丽 (1980 -), 女, 江苏南通人, 博士生, YuLi_Xue@ee.buaa.edu.cn.

库[7]包含了困乏、眨眼等常见表情,PIE (Pose, Illumination, and Expression)人脸数据库[8]包含眨眼、谈话等表情,AR人脸数据库[9]包含尖叫等表情,清华人脸表情数据库[10]包含大笑、眨眼等表情.综上所述,人脸表情数据库不仅数量不足,而且包含的表情种类有限,尤其是具有情感意义的表情不多,极大地限制了人脸表情分析及情感识别研究的发展.

本文针对国内国外人脸表情数据匮乏的现状,设计并实现了较全面的视频人脸表情数据库,涉及人类主要的情感表情,包含单一表情、混合表情和复杂表情等共 25种,并对表情数据进行了评价及表情识别的实验.

1　表情数据库的设计

BHU (Beihang University)人脸表情数据库包含了正面和侧面 2类人脸表情彩色视频数据,侧面表情是以正面为基准旋转约 30 拍摄得到的.鉴于人脸具有对称性,这里仅选取一侧的表情.

为了使机器能够通过表情识别出人类的情感,BHU人脸表情数据库在 6种基本表情的基础上进行了扩展,使其具有人类表达的主要表情.

人类具有十分丰富的面部表情,社会学家将人类的表情大致分为 18种:失望、愁、忧、惊愕、仓皇、恐怖、嫌恶、暴怒、惧怕、怀疑、焦急、厌恶、鄙夷、藐视、嘲笑、微笑、恳求、大笑[11].本文在上述表情的基础上进行筛选并加以补充构成了 18种单一表情;参考文献[12]列举的 15种混合表情,选择了 3种以惊讶为结合点的混合表情:惊讶—喜悦、惊讶—悲伤和惊讶—生气等;设计了有代表性的 4种复杂表情:人的面部被遮挡时微笑和生气,人在谈话时微笑和生气等.

2　表情实验的设计

为了获得各种表情数据,本文对实验场景、数据的采集过程以及对被试者的诱导方式进行了设计.

2.1　实验场景的设计

场景的复杂度,如杂乱的背景或其它人的突然出现,会潜在地影响到人脸检测、特征跟踪和表情识别的精度.在大多数的人脸表情数据库中场景单一且仅有一个人出现,但在自然环境中很可能有多人存在.因此在训练数据中引入复杂的场景,将有利于提高算法对于这些变化的鲁棒性[6].BHU人脸表情数据库是针对人机交互的情感需求设计的,所以实验场景被设计为背景比较复杂的实验室,光照条件为头顶日光灯照且受自然光影响,并允许有其它人的出现,但不能主动影响被试者的情感表达.

被试者被要求坐在一台计算机前依次表达指定的表情.被试者的正前方和以正前方为基准的约 30 的右侧方处固定 2个连接到计算机的普通 PC摄像头 (ANC - 网乐宝贝 II, 38万像素)以摄录被试者的表情,如图 1所示.

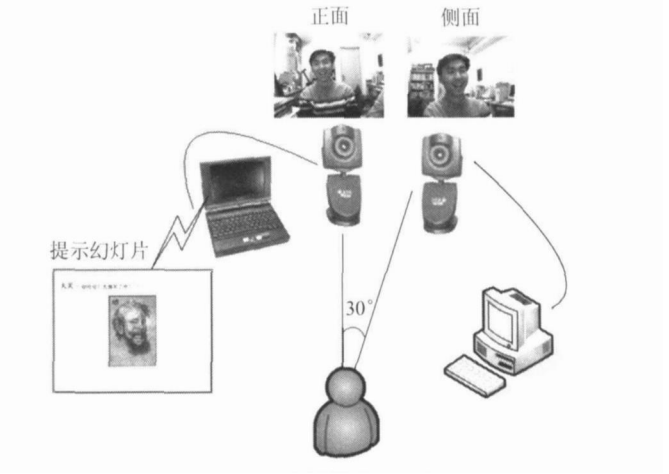


图 1　人脸表情拍摄示意图

从表情识别鲁棒性研究考虑,本文尽量减少对被试者的约束.被试者在表达面部表情时可以轻微地移动头部,但需基本保持正面面对正前方的摄像头.被试者的穿着和发型也不受任何约束.

2.2　数据采集过程的设计

为了获得被试者的表情数据,实验者需诱导被试者依次表达单一表情、混合表情和复杂表情,并和助手进行数据的采集:①被试者被要求正坐于一个摄像头的正前方,而另一个摄像头位于被试者的右侧 30 左右.摄像头与被试者的距离约为 0.6m;②被试者被要求自然地表达指定的表情;③如果被试者难于表达指定的表情,实验者应用各种提示对其进行诱导;④实验者发出口令指示被试者表达指定的表情,并控制 2个摄像头同步进行摄录,直到被试者表达完毕以口令结束录像;⑤当被试者的表达不够理想时,重复 ③和 ④,但重复的次数不宜过多;⑥对下一个指定的表情进行 ②～⑥的操作,直到被试者将 25种表情表达完毕.

2.3　提示的设计

为了使被试者尽量自然地表达指定的表情,本文设计了各种提示.通常,实验者会向被试者展示摘自文献[11]的表情素描、JAFFE数据库中的典型表情图片、虚拟场景描述、潜台词或独白构成

的幻灯片,以启发被试者更好地表达指定的表情.实验者还会根据情况演示前期录制的较成功的表情视频(如图 2所示)或现场示范部分表情,使被试者对指定的表情有更深刻的理解.

3 BHU人脸表情数据库的组成

3.1 介 绍

BHU人脸表情数据库的数据采集工作自2005年 4月在中国北京航空航天大学的某个实验室中进行,历时两个月. 32位 21至 25岁的大学生(男性: 14人,女性: 18人)参加了数据采集实验.每位被试者表达了指定的 25种表情,每种表情均采集了约 6 s的视频数据,数据存储为 AVI格式, 640像素 ×480像素, 24位彩色,帧率为 30帧/s.最初获得的每位被试者的视频存储空间大约为 3 GByte,并且包含很多冗余信息.为了精减数据库,本文用视频剪辑工具 Adobe Premiere 6.5版将每个视频剪辑成约 2 s的视频.

3.2 数据库的组成

BHU人脸表情数据库由 3部分组成: 18种单一表情、3种混合表情和 4种复杂表情.单一表情包括:微笑、大笑、嘲笑、忧愁、生气、暴怒、惊讶、害怕、嫌恶、怀疑、焦急、鄙夷、恳求、失望、悲伤、好奇、害羞和打哈欠等;混合表情包括:惊讶—喜悦、惊讶—悲伤、惊讶—生气等;复杂表情包括:手遮挡脸时微笑、手遮挡脸时生气、谈话时微笑和谈话时生气等.图 2~图 4所示为某大学生的正面单一表情、混合表情及复杂表情.



图 2 正面单一表情



图 3 正面混合表情



图 4 正面复杂表情

4 对表情数据库的评价

4.1 问卷调查

由于 BHU人脸表情数据库的被试者不是演员,缺乏通过面部表情充分表达情感的经验,且由于个性的因素,表达的表情不一定与指定的表情相吻合.为了给研究者提供质量较高的表情数据,本文采用问卷调查的方式,对采集的数据进行了客观评价.具体方法是让被调查者观看某一段表情视频(不给出提示),请其在 18种单一表情中选择该视频可能表现的表情或给出 18种以外的其它表情,并做出 1分、2分或 3分(得分越高则越相似)的评价,未做出评价的一律计为 0分.每位被调查者需要对同一位被试者的 18段单一表情视频做出评价.在这里,由 20位被调查者对 16位被试者的每个单一表情数据进行评价.

令 $x_{ij}^{(kl)}$为第 j位被调查者对第 i位被试者表达的第 k种表情给出的关于第 l种表情的评分($x_{ij}^{(kl)}=0, 1, 2, 3$),则由式(1)计算出每位被试者的每个单一表情数据关于每种表情的平均得分:

$$x_i^{(kl)}=\frac{1}{N}\sum_{j=1}^{N}x_{ij}^{(kl)}$$

$$i=1, 2, \cdots, 16;\ k, l=1, 2, \cdots, 18 \qquad (1)$$

式中, $x_i^{(kl)}$为第 i位被试者表达的第 k种表情获得的关于第 l种表情的平均得分; $N=20$表示对 20个样本求均值.

4.2 均值统计分析

为了得到对表情数据的总体描述,本文进一步由式(2)计算得到每一单一表情的平均得分,结果如表 1所示.

$$x^{(kk)}=\frac{1}{M}\sum_{i=1}^{M}x_i^{(kk)} \qquad k=1, 2, \cdots, 18 \qquad (2)$$

式中, $x^{(kk)}$为第 k种单一表情获得的关于自身的平均得分; $M=16$表示对 16个样本求均值.

表 1 单一表情数据可识别性平均得分

表情	得分	表情	得分	表情	得分
打哈欠	2.45	好奇	1.56	忧愁	1.29
大笑	2.01	鄙夷	1.53	悲伤	1.25
惊讶	1.94	嫌恶	1.47	怀疑	1.23
害羞	1.89	害怕	1.45	暴怒	1.18
微笑	1.85	嘲笑	1.44	焦急	1.13
失望	1.65	生气	1.40	恳求	1.12

表 1说明如打哈欠、大笑、惊讶和害羞等表情是易于表达或识别的,而恳求、焦急和怀疑等表情则表现得不明显或由于缺乏语境而难于识别. 以上评价不仅有助于有效的表情识别,也对人机交互中的情感推理具有参考价值.

4.3　分层聚类分析

为了说明各种表情之间的关系,本文进一步对单一表情进行了聚类分析. 聚类分析是根据事物性质的相近程度将其分类.

本文以 18种单一表情作为 18个变量,一位被试者表达的某一表情得到的关于 18种表情的一组平均得分 $(x_i^{(kl)},\ l=1, 2, \cdots, 18)$作为一个样本,以 Pearson相关系数为距离,采用类间平均链锁法对 16位被试者 $(i=1, 2, \cdots, 16)$的 18个表情数据 $(k=1, 2, \cdots, 18)$构成的 288个样本进行了分层变量聚类分析.

如图 5所示,暴怒和生气表情有较大的相似性,悲伤和忧愁表情也较为相似,可见某些相近的情感会产生相似的表情. 此外,某些差异较大的情感 (如害怕与嫌恶)也会产生相似的表情,需要结合多模信息 (如语音、姿势和生理信号等)才能准确识别情感.

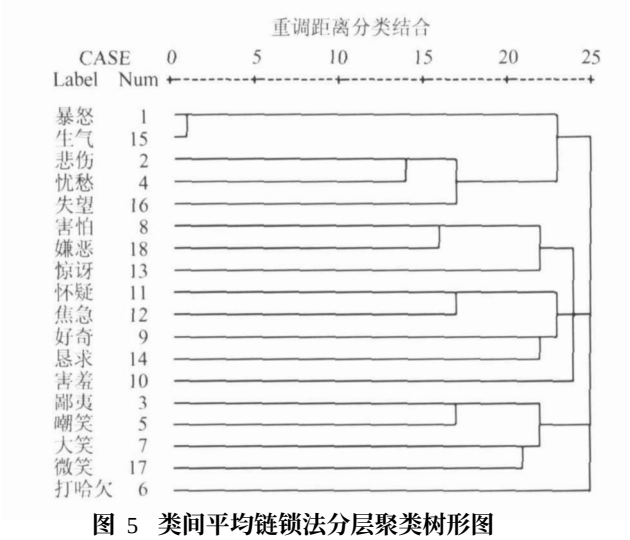
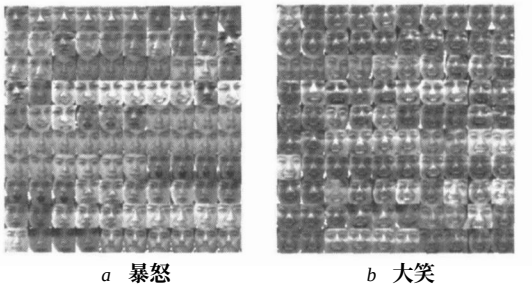


图 5　类间平均链锁法分层聚类树形图

5　表情识别实验

人脸表情识别主要包括人脸检测和表情特征提取与识别 2个技术环节,本文使用自行开发的实时人脸表情识别系统的人脸检测器从 BHU人脸表情数据库中所有单一表情视频中全自动检测出人脸并进行筛选作为表情分类器的训练样本,以确保表情分类环节在训练与识别上的一致性,减少人为干预的影响.

在选择训练样本时,本文依据表情类别对训练样本进行分类,未对表情强度进行细分,图 6所示尺寸归一化为 24 ×24点像的部分灰度训练样本. 本文针对每种单一表情,以该类表情样本为正例样本,其它各类表情样本及平静样本为反例样本,经过灰度方差归一化后训练专门表情分类器.



a 暴怒　　b 大笑

图 6　部分表情训练样本

本表情识别系统采用 Viola提出的积分图像和基于 Adaboost方法[13]训练人脸检测分类器和表情分类器. Adaboost是一种迭代方法,它是通过从大量的弱分类器中选取最具有分类意义的组合成一个强分类器,其定义如式 (3)所示.

$$h(x)=\begin{cases}1 & \sum_{t=1}^{T}\alpha_t h_t(x)\geq\frac{1}{2}\sum_{t=1}^{T}\alpha_t\\ 0 & \text{其它}\end{cases}\qquad(3)$$

式中, T为循环次数; h_t为每轮选取误差最小的弱分类器; $\alpha_t=\log\frac{1-\varepsilon_t}{\varepsilon_t}$,其中 ε_t为 h_t的判别误差.

专门表情分类器将 Adaboost迭代后的强分类器的定义作如式 (4)的修改,输出为 [0, 1]区间的数,将“硬 判决转换为“软 判决.

$$h_e(x)=\frac{\sum_{t=1}^{T}\alpha_t h_t(x)}{\sum_{t=1}^{T}\alpha_t}\qquad(4)$$

本文从 18种单一表情中选取 9种表情 (分别为:暴怒、悲伤、打哈欠、大笑、害怕、好奇、惊讶、微笑、嫌恶)进行表情识别实验. 最终表情分类器对各专门表情分类器的得分进行判决得出表情识别结果,其规则如图 7所示.

本文选取了共 814个样本对各专门表情分类器进行了训练,并对表情分类器进行了测试. 表 2给出了阈值为 0.4时表情分类器的识别结果,可

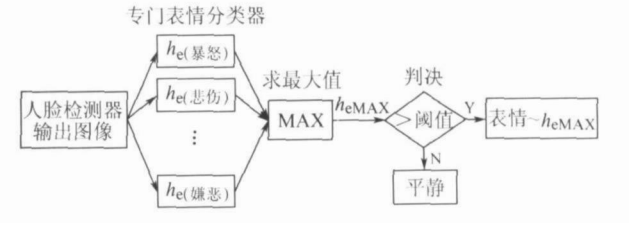


图 7　表情分类器判决规则

见表情分类器对大多数表情的正确识别率较高,误识别为其它表情的概率较小,打哈欠、大笑、暴怒等表情有较高的识别率,而惊讶、害怕、好奇等表情的识别率不够理想. 可根据各专门表情分类器的可靠性进行加权数据融合,以达到更好的表情判决效果[14].

表情识别结果表明,对多类表情进行识别是可行的. 多类表情识别对于人机交互中的多种情感识别是有利的. 本文的表情分类器还未达到很高的识别率,可以通过对姿势进行归一化,选择更好的分类器以及结合面部运动编码、模糊理论、信息融合等,以达到更好的识别效果.

表 2 表情分类器识别结果

测试样本		表情分类器判决结果(概率) 阈值 =0.4									
表情	样本数	暴怒	悲伤	打哈欠	大笑	害怕	好奇	惊讶	微笑	嫌恶	平静
暴怒	100	0.83	0.01	0.01	0	0.02	0.02	0.04	0.02	0.05	0
悲伤	100	0.10	0.71	0	0	0.04	0.04	0.03	0.01	0.03	0.04
打哈欠	82	0.02	0.01	0.94	0	0.01	0.01	0	0	0	0
大笑	100	0	0	0	0.85	0	0.02	0	0.09	0.04	0
害怕	100	0.06	0.05	0	0.01	0.67	0.04	0.05	0.05	0.06	0.01
好奇	98	0.10	0.03	0	0	0.05	0.69	0.02	0.06	0.03	0.01
惊讶	73	0.05	0.01	0.04	0	0.05	0.16	0.62	0.01	0.04	0
微笑	79	0	0.03	0	0.04	0.01	0.10	0	0.77	0.03	0.03
嫌恶	82	0.06	0.10	0	0.01	0.04	0.05	0.01	0.01	0.72	0

6 结束语

人脸表情识别是情感计算和人机交互的重要研究领域. 而高质量的表情数据库对该领域的发展有着至关重要的作用. 本文针对国内外情感表情数据匮乏的现状,设计并实现了较全面的视频人脸表情数据库,涉及人类主要的情感表情,包含单一表情、混合表情和复杂表情等共 25种,并对表情数据进行了评价及表情识别实验. BHU人脸表情数据库对促进人脸表情识别技术及情感化人机交互系统的发展具有积极的意义.

致谢 感谢日本 ATR研究所的 Michael Lyons提供了 JAFFE数据库. 感谢众多的被试者,以及对该人脸表情数据库给予过帮助的人们.

参考文献 (References)

[1] Picard R W. Affective computing [M]. Cambridge, MA: MIT Press, 1997

[2] Ralph Gross Face databases [C]// Li S, Jain A. Handbook of face recognition. New York: Springer-Verlag, 2005

[3] Ekman P, Friesen W. Constants across cultures in the face and emotion [J]. Journal of Personality and Social Psychology, 1971, 17(2): 124 - 129

[4] Black M J, Yacoob Y. Recognizing facial expressions in image sequences using local parameterized models of image motion [J]. International Journal of Computer Vision, 1997, 25(1): 23 - 48

[5] Lyons Michael J, Shigeru Akamatsu, et al. Coding facial expressions with gabor wavelets [C]// Third IEEE International Conference on Automatic Face and Gesture Recognition. Nara: IEEE Computer Society, 1998: 200 - 205

[6] Kanade T, Cohn J, Tian Y L. Comprehensive database for facial expression analysis [C]// Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition. Grenoble: IEEE Computer Society, 2000: 46 - 53

[7] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces: vs fisherfaces: recognition using class specific linear projection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 711 - 720

[8] Sim Terence, Baker Simon, Bsat Maan. The CMU pose, illumination, and expression (PIE) database [C]// Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition. Washington: IEEE Computer Society, 2002: 46 - 51

[9] Martinez A R, Benavente R. The AR face database [R]. Computer Vision Center (CVC) Technical Report #24, 1998

[10] 吴丹,林学[illegible].人脸表情视频数据库的设计与实现[J]. 计算机工程与应用, 2004, 40(5): 177 - 180
Wu Dan, Lin Xueyin. The design and realization of a video database for facial expression analysis [J]. Computer Engineering and Applications, 2004, 40(5): 177 - 180 (in Chinese)

[11] 周昌乐. 心脑计算举要[M]. 北京:清华大学出版社. 2003: 155 - 157
Zhou Changle. The reviews of mental computation [M]. Beijing: Tsinghua University Press, 2003: 155 - 157 (in Chinese)

[12] Hara F, Kobayashi H. Computer graphics for expressing robot - artificial emotions [C]// Proceedings of Robot and Human Communication. Tokyo: IEEE International Workshop, 1992: 155 - 160

[13] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Cambridge: IEEE Computer Society, 2001: 511 - 518

[14] Chair Z, Varshney P K. Optimal data fusion in multiple sensor detection systems [J]. IEEE Trans, AES, 1986, 22(1): 98 - 101

(责任编辑:娄 嘉)