

Kappa 加权的子空间融合表情识别方法

贾熹滨¹, 张艳华², 包锡元³

(北京工业大学 计算机学院 多媒体与智能软件技术北京市重点实验室, 北京 100124)

摘要: 为提高面部表情识别效果, 提出基于 Kappa 计算面部表情图像子区域对表情的贡献程度, 并线性加权子空间预测结果. 将标准化后的人脸表情图像上下平均分割成 2 个子区域, 确定上半脸和下半脸及全脸 3 个表情子空间, 采用 Gabor 小波特征, 分别利用 SMO、MLP 和 KNN 三种分类器, 统计并计算基于 Kappa 的子空间表情信息. 在 Cohn-Kanade 和 JAFFE 两个表情图像库进行测试, 实验结果表明: 基于 Kappa 加权融合的表情识别方法识别率更高.

关键词: Gabor 小波; 面部分块; 加权融合; 特征提取

中图分类号: TP 391

文献标志码: A

文章编号: 0254 - 0037(2014)06 - 0900 - 08

Kappa Weighted Subspace Fusion Approach of Facial Expression Recognition

JIA Xi-bin¹, ZHANG Yan-hua², BAO Xi-yuan³

(Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology, College of Computer Science, Beijing University of Technology, Beijing 100124, China)

Abstract: To improve expression recognition rates, Kappa-based contribution degree computing approach of face sub-area in recognizing expression is proposed. It is utilized as the basis of deriving weights to fuse the subspace prediction results. The normalized face emotion images are partitioned averagely into two sub-regions to obtain the three expression subspaces containing the upper and lower half face parts and the whole face. Each part is represented with Gabor feature. Then, three classifiers: SMO, MLP and KNN are separately used. The expression prediction results are counted to obtain the Kappa. Experiments are done on the CMU and JAFFE two expression image databases and results show that Kappa weighted fusion expression recognition approach has higher recognition accuracy.

Key words: Gabor wavelet; facial block; weighted fusion; feature extraction

在人际交往中, 人脸表情是一种非常重要的信息传递方式, 非语言交流研究开创者 Mehrabian 的研究结果表明: 听众对语言的理解只有 7% 来自于声音, 而 93% 源于肢体语言、声调等信息. 通过对人脸的数据采集、分析和表情计算, 实现如虚拟教师等虚拟代理, 建立智能人机接口是目前计算

机视觉的重要研究方向之一^[1]. 由于人脸面部表情的复杂性, 目前表情建模仍是表情认知研究的难点. P. Ekman 等^[2]最早从心理学角度提出准确分析人脸表情的人脸动作编码系统 FACS, 根据所定义的人脸最小运动单元 (action unit, AU) 建立检测人脸信号架构, 为人脸表情识别提供依据^[3].

收稿日期: 2013-06-24

基金项目: 国家自然科学基金资助项目(61070117); 北京市自然科学基金资助项目(4122004)

作者简介: 贾熹滨(1969—), 女, 副教授, 主要从事视觉图像理解方面的研究, E-mail: jiaxibin@bjut.edu.cn

H. Ahn 等^[4]为避开解剖学定义单元不易提取的困难,提出组合 AU 构造对应几何特征,建立关联规则,利用主动外观模型 (active appearance models, AAM) 匹配算法提取特征点,计算 AU 动作识别表情. 然而 AAM 算法仍需要手工标定训练数据,受个人判断影响,有一定随意性,缺乏一致性, G. Donato 等^[5]采用主元分析 (principal component analysis, PCA)、线性判别分析 (linear discriminant analysis, LDA)、局部特征分析 (local feature analysis, LFA)、独立主元分析 (independent component analysis, ICA) 等全局特征表示方法,以及局部滤波分析如 Gabor 小波、局部主元分析对面部动作分类能力作了广泛的比较,通过比较 FACS 定义的上半部分和下半部分面部的 12 种人脸动作,发现利用 Gabor 小波,分类正确率可达 96%,高于其他全局特征,为局部空域特征在人脸细节动作表述的能力提供了证据.

D. W. Cunningham 等^[6]的工作进一步从人认知的角度证明了面部不同区域对不同表情作用不同. 其从心理学和神经学理论研究人脸表情,将人脸进行部分遮挡后,将部分面部区域 (包括眼睛、眼睛和眉毛、唇部区域) 呈现给受试者进行表情识别,测试了受试者依据局部区域对高兴、惊讶、厌恶等表情的识别,结果表明:自然人认知表情时不同区域对不同表情判断的贡献具有差异,如高兴唇部区域已含有足够的决策信息,而只提供眼睛区域不足以判断高兴^[6].

本文针对 Gabor 小波具有揭示人脸局部特性的优势,应用 Gabor 小波表述人脸表情图像,同时考虑到人脸不同区域对表情的贡献程度不同,对人脸进行上下平均分割,将上下 2 个子区域连同全脸分别计算 Gabor 特征,学习各部分对表情识别的贡献,并用于子空间表情融合算法中,以提高表情识别效果.

1 基于决策层融合的表情识别系统概述

本系统强调不同区域在不同表情识别中的作用,采用人脸分块策略,考虑到表情描述贡献的差异主要在分别包含眼睛和唇部的上下半脸,采取将预处理后的面部图像平均分为上下 2 个部分的策略,如图 1 所示,根据训练数据学习基本表情识别中各分块及全脸所起的作用,识别阶段通过投票选择具有最大贡献的块,计算表情状态. 将每部分进行单独特征提取和分类识别,得到的预测率矩阵应用融合模型进行加权融合.

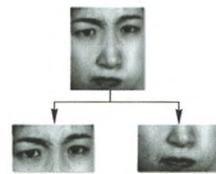


图 1 面部分块示意

Fig. 1 Diagram of facial block

具体算法流程如图 2 所示,对正则化的表情图像依据分块策略沿中心分为 2 个部分,本文使用整体、上半部、下半部分别单独进行特征提取. 提取表情图像特征后,分别利用 Weka 提供的分类器进行分类识别,将预测率矩阵导出到 Excel 表格中,如图 3 所示. 第 1 列代表实例编号,每种表情包含 10 个样本实例;第 2 列代表该样本实际的所属表情,如图 3 为 angry 表情的样本;第 3 列是该实例的预测结果,此结果是根据第 4 ~ 9 列中的最大预测率确定的,比如第 2 行中的第 4 ~ 9 列中最大的预测率是 57%,它所对应的标签是 3: fear,所以该实例最后就被判定为 fear;最后一列“Result”代表预测是否正确,即预测类与实际所属类别是否相同,比如实例 1 实际为 angry,但预测为 fear,所以该实例被分类错误,因此为 FALSE. 最后的识别率等于 TRUE 的数

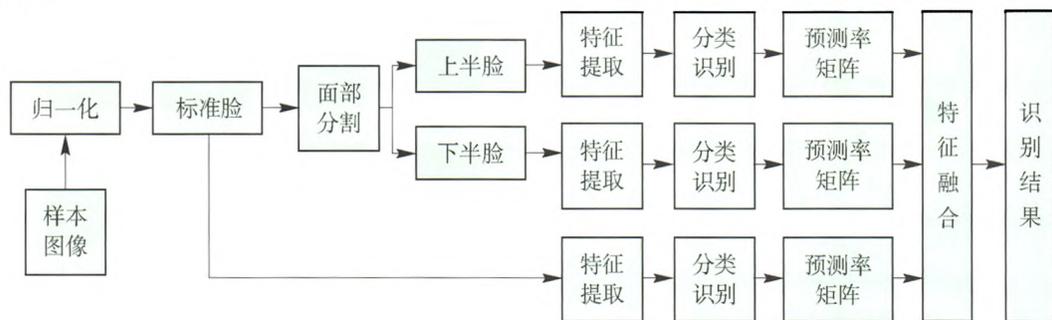


图 2 系统框图

Fig. 2 System diagram

量除以整个实例样本数;然后利用每部分的识别率计算每部分的 Kappa 值^[7];再求取 Kappa 的信息熵作为加权系数,对预测率矩阵进行加权.

| instance | actual | WHOLE | 1:angry | 2:disgust | 3:fear | 4:smile | 5:sadness | 6:surprise | Result |
|----------|---------|------------|---------|-----------|--------|---------|-----------|------------|--------|
| 1 | 1:angry | 3:fear | 1% | 42% | 57% | 0% | 0% | 0% | FALSE |
| 2 | 1:angry | 1:angry | 60% | 0% | 0% | 0% | 40% | 0% | TRUE |
| 3 | 1:angry | 3:fear | 11% | 0% | 88% | 0% | 1% | 0% | FALSE |
| 4 | 1:angry | 6:surprise | 17% | 2% | 34% | 0% | 5% | 42% | FALSE |
| 5 | 1:angry | 1:angry | 67% | 0% | 9% | 20% | 4% | 0% | TRUE |
| 6 | 1:angry | 1:angry | 96% | 1% | 0% | 0% | 1% | 2% | TRUE |
| 7 | 1:angry | 1:angry | 99% | 0% | 0% | 1% | 0% | 0% | TRUE |
| 8 | 1:angry | 1:angry | 74% | 25% | 0% | 1% | 0% | 0% | TRUE |
| 9 | 1:angry | 1:angry | 56% | 0% | 0% | 43% | 0% | 0% | TRUE |
| 10 | 1:angry | 1:angry | 99% | 0% | 1% | 0% | 0% | 0% | TRUE |

图3 预测率矩阵

Fig.3 Prediction rate matrix

2 表情图像预处理

由于表情信息主要集中在面部,为加强图像中的有用信息、减少头发等无用信息的影响,本文对标准人脸库中的表情图像进行尺度正则化.由于人脸数据库中不同人的脸型等差异,人脸在原始表情图像中所占尺度会有所不同,如图4所示.为规范面部图像大小,在一定程度上降低人脸脸型差异对表情识别带来的影响,根据面部区域各器官之间的位置比例关系^[8](如图5所示)对数据库中的表情图像进行预处理,在采用人工交互的方式标定出两眼中心的基础上,以眼中心的距离为基准,按照图5所示相对眼中心的位置和比例关系切割出矩形框所示的感兴趣区域(region of interest, ROI),同时以眼中心位置为基准,将图像缩放旋转到统一尺度上.这里原始图像的大小是 256 × 256,正则化后的图像尺寸是 90 × 100.切割后尺度正则化表情图像如图6所示.



图4 原始表情图像示例

Fig.4 Example of source image

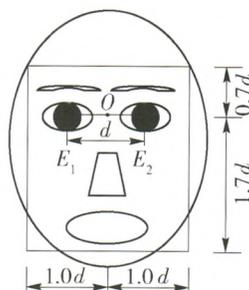


图5 人脸表情图像切割比例示意

Fig.5 Illustration of facial image cutting proportion



图6 正则化后的表情图像

Fig.6 Expression image after normalization

3 Gabor 小波特征尺度和方向的选择

二维 Gabor 滤波器具有在空间域和频率域同时取得最优局部化的特性,与人类生物视觉特性很相似,因此,能很好描述对应于空间频率(尺度)、空间位置及方向选择性的局部结构信息^[9].在提取人脸表情特征时,一般采用在不同尺度和方向上的 Gabor 滤波器族,并根据表情的特点和神经生理学的结论来选择参数.通常大尺度可描述全局性较强的信息,同时可掩盖掉图像中噪声的影响,而小尺度可描述比较精细的局部结构,但同时受噪声的影响也大^[10].根据对表情表达主要体现在局部细节,如边缘纹理等高频信息,同时考虑到人脸表情变化过程中垂直方向包含更多的表情信息,本文通过遍历不同方向和尺度的 Gabor 滤波器的设置,选取了 3 个尺度 0、1、2, 5 个方向 30°、60°、90°、120°、150° 共 15 种组合进行实验,将对图7所示的正则化后的表情图像进行变换,效果如图8所示.图片排列中从上到下尺度增加(0、1、2),从左到右是方向依次增加(30°、60°、90°、120°、150°).



图7 预处理后的图像

Fig.7 Image after pre-treatment

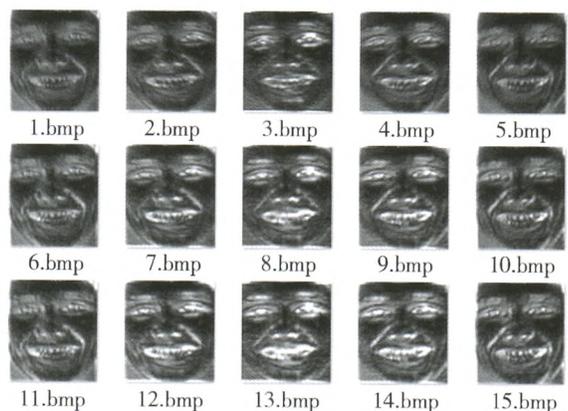


图8 不同尺度和方向 Gabor 变换结果

Fig.8 Gabor translation at different orients and scales

从图 8 中的实验结果可看出,小尺度(第 1 排图片相比下面 2 排)更能体现面部表情的细节特征,描述细节更加精确,而方向接近垂直的(第 3 列相比其他列)更能表现表情的变化方向,使特征更加明显. 所以本文选取尺度为 0、1,方向为 60°、90°、120°作为 Gabor 滤波器的参数设置. 这里 Gabor 特征由于维数过高,本文采用离散余弦变换(discrete cosine transform, DCT)进行降维处理,特征维数取为 300 维.

4 基于决策级的表情识别融合方案

4.1 基于概率模型的加权融合实验

为了更好地利用面部各部分对表情的贡献度,更加准确地识别出样本所属的表情类,本文将决策级融合理论应用到表情识别中,如图 9 所示,通过融合各个部分的识别结果,实现对面脸各部分信息的充分利用. 本文利用人脸上下部分对表情识别贡献的互补关系,将概率论中的加法定理和乘法定理应用于本节的决策级融合模型中,将 3 个情形(全脸时被识别为某种表情,上半脸时被识别为某种表情,下半脸时被识别为某种表情)进行融合计算,共同决策样本的所属类别.

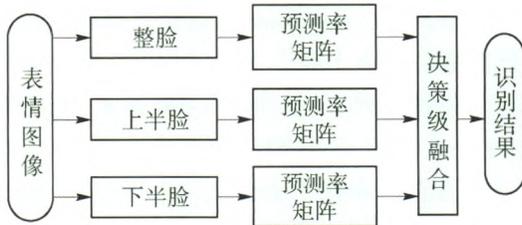


图 9 基于决策级融合的表情识别

Fig. 9 Facial expression recognition based on fusion of decision level

本文将人脸图像分成上下 2 个部分,结合原始图像,分别对这 3 个源信息进行特征提取,分类识别,得到各自的识别率矩阵,图 10~12 分别是全脸、上半脸和下半脸识别率矩阵的实验结果截图.

以上 3 个实验截图中的概率都是预测率,即对每个实例预测为每种表情的概率,本文结合概率论的知识,设全脸、上半脸、下半脸分别是 3 个条件,那么以上的预测率都是条件概率,即在全脸时预测为某种表情的概率记为 $P(T|W)$,在上半脸时预测为某种表情的概率记为 $P(T|U)$,在下半脸时预测为某种表情的概率记为 $P(T|L)$, W 代表应用全脸图像数据得到的预测率矩阵, U 代表应用上半脸图像数据得到的预测率矩阵, L 代表应用下半脸图像数

| | A | B | C | D | E | F | G | H | I | K | |
|----|----------|------------|------------|---------|-----------|--------|-----------|---------|------------|---|--------|
| 1 | instance | actual | WHOLE | 1:angry | 2:disgust | 3:fear | 4:sadness | 5:smile | 6:surprise | | Result |
| 2 | 1 | 1:angry | 1:angry | 51% | 3% | 7% | 37% | 1% | 1% | | TRUE |
| 3 | 2 | 1:angry | 5:smile | 35% | 1% | 8% | 7% | 50% | 0% | | FALSE |
| 4 | 3 | 1:angry | 3:fear | 5% | 16% | 70% | 7% | 1% | 1% | | FALSE |
| 5 | 4 | 2:disgust | 5:smile | 4% | 10% | 36% | 0% | 40% | 10% | | FALSE |
| 6 | 5 | 2:disgust | 2:disgust | 8% | 78% | 9% | 1% | 2% | 3% | | TRUE |
| 7 | 6 | 2:disgust | 3:fear | 2% | 2% | 64% | 4% | 17% | 12% | | FALSE |
| 8 | 7 | 3:fear | 3:fear | 3% | 13% | 46% | 35% | 2% | 1% | | TRUE |
| 9 | 8 | 3:fear | 3:fear | 17% | 1% | 56% | 14% | 11% | 2% | | TRUE |
| 10 | 9 | 3:fear | 5:smile | 9% | 9% | 29% | 0% | 45% | 8% | | FALSE |
| 11 | 10 | 4:sadness | 1:angry | 49% | 23% | 4% | 1% | 10% | 14% | | FALSE |
| 12 | 11 | 4:sadness | 2:disgust | 4% | 66% | 21% | 8% | 1% | 0% | | FALSE |
| 13 | 12 | 4:sadness | 3:fear | 11% | 25% | 48% | 13% | 0% | 2% | | FALSE |
| 14 | 13 | 5:smile | 5:smile | 12% | 13% | 6% | 1% | 66% | 3% | | TRUE |
| 15 | 14 | 5:smile | 5:smile | 4% | 14% | 1% | 1% | 73% | 7% | | TRUE |
| 16 | 15 | 5:smile | 5:smile | 13% | 5% | 35% | 0% | 43% | 4% | | TRUE |
| 17 | 16 | 6:surprise | 6:surprise | 2% | 21% | 1% | 0% | 2% | 74% | | TRUE |
| 18 | 17 | 6:surprise | 6:surprise | 2% | 0% | 36% | 1% | 1% | 60% | | TRUE |
| 19 | 18 | 6:surprise | 6:surprise | 2% | 20% | 7% | 0% | 3% | 67% | | TRUE |

图 10 基于全脸的预测率矩阵

Fig. 10 Prediction matrix based on whole face

| | A | B | L | M | N | O | P | Q | R | T | |
|----|----------|------------|------------|---------|-----------|--------|-----------|---------|------------|---|--------|
| 1 | instance | actual | UPPER | 1:angry | 2:disgust | 3:fear | 4:sadness | 5:smile | 6:surprise | | Result |
| 2 | 1 | 1:angry | 4:sadness | 24% | 16% | 5% | 53% | 2% | 1% | | FALSE |
| 3 | 2 | 1:angry | 1:angry | 61% | 1% | 0% | 3% | 18% | 2% | | TRUE |
| 4 | 3 | 1:angry | 3:fear | 6% | 1% | 81% | 2% | 2% | 8% | | FALSE |
| 5 | 4 | 2:disgust | 3:fear | 38% | 7% | 42% | 2% | 2% | 9% | | FALSE |
| 6 | 5 | 2:disgust | 2:disgust | 12% | 71% | 12% | 2% | 1% | 3% | | TRUE |
| 7 | 6 | 2:disgust | 5:smile | 0% | 21% | 21% | 13% | 32% | 13% | | FALSE |
| 8 | 7 | 3:fear | 3:fear | 0% | 36% | 53% | 2% | 2% | 7% | | TRUE |
| 9 | 8 | 3:fear | 4:sadness | 7% | 1% | 29% | 58% | 1% | 4% | | FALSE |
| 10 | 9 | 3:fear | 2:disgust | 5% | 74% | 14% | 1% | 2% | 4% | | FALSE |
| 11 | 10 | 4:sadness | 5:smile | 22% | 22% | 1% | 1% | 48% | 6% | | FALSE |
| 12 | 11 | 4:sadness | 2:disgust | 24% | 44% | 1% | 4% | 22% | 5% | | FALSE |
| 13 | 12 | 4:sadness | 4:sadness | 11% | 27% | 17% | 40% | 0% | 6% | | TRUE |
| 14 | 13 | 5:smile | 5:smile | 5% | 6% | 1% | 9% | 77% | 3% | | TRUE |
| 15 | 14 | 5:smile | 2:disgust | 1% | 76% | 6% | 1% | 12% | 4% | | FALSE |
| 16 | 15 | 5:smile | 5:smile | 33% | 12% | 0% | 1% | 46% | 9% | | TRUE |
| 17 | 16 | 6:surprise | 6:surprise | 2% | 4% | 2% | 3% | 5% | 84% | | TRUE |
| 18 | 17 | 6:surprise | 3:fear | 10% | 0% | 51% | 3% | 5% | 31% | | FALSE |
| 19 | 18 | 6:surprise | 6:surprise | 2% | 37% | 5% | 0% | 6% | 50% | | TRUE |

图 11 基于上半脸的预测率矩阵

Fig. 11 Prediction matrix based on upper face

| | A | B | U | V | W | X | Y | Z | AA | AC | |
|----|----------|------------|------------|---------|-----------|--------|-----------|---------|------------|----|--------|
| 1 | instance | actual | LOWER | 1:angry | 2:disgust | 3:fear | 4:sadness | 5:smile | 6:surprise | | Result |
| 2 | 1 | 1:angry | 1:angry | 65% | 1% | 3% | 30% | 2% | 0% | | TRUE |
| 3 | 2 | 1:angry | 1:angry | 44% | 1% | 28% | 16% | 12% | 0% | | TRUE |
| 4 | 3 | 1:angry | 2:disgust | 11% | 42% | 40% | 5% | 1% | 1% | | FALSE |
| 5 | 4 | 2:disgust | 3:fear | 6% | 2% | 64% | 0% | 19% | 10% | | FALSE |
| 6 | 5 | 2:disgust | 3:fear | 10% | 11% | 69% | 8% | 2% | 0% | | FALSE |
| 7 | 6 | 2:disgust | 5:smile | 2% | 5% | 16% | 2% | 53% | 23% | | FALSE |
| 8 | 7 | 3:fear | 3:fear | 6% | 7% | 54% | 32% | 1% | 0% | | TRUE |
| 9 | 8 | 3:fear | 3:fear | 0% | 1% | 79% | 7% | 12% | 1% | | TRUE |
| 10 | 9 | 3:fear | 5:smile | 16% | 11% | 11% | 15% | 46% | 1% | | FALSE |
| 11 | 10 | 4:sadness | 1:angry | 58% | 33% | 2% | 1% | 1% | 5% | | FALSE |
| 12 | 11 | 4:sadness | 4:sadness | 14% | 28% | 20% | 37% | 0% | 1% | | TRUE |
| 13 | 12 | 4:sadness | 2:disgust | 10% | 80% | 5% | 4% | 0% | 2% | | FALSE |
| 14 | 13 | 5:smile | 3:fear | 17% | 4% | 62% | 0% | 11% | 6% | | FALSE |
| 15 | 14 | 5:smile | 5:smile | 7% | 2% | 1% | 2% | 77% | 11% | | TRUE |
| 16 | 15 | 5:smile | 5:smile | 24% | 0% | 18% | 1% | 56% | 1% | | TRUE |
| 17 | 16 | 6:surprise | 6:surprise | 11% | 28% | 1% | 0% | 0% | 60% | | TRUE |
| 18 | 17 | 6:surprise | 6:surprise | 4% | 4% | 8% | 0% | 4% | 81% | | TRUE |
| 19 | 18 | 6:surprise | 3:fear | 11% | 4% | 71% | 0% | 6% | 8% | | FALSE |

图 12 基于下半脸的预测率矩阵

Fig. 12 Prediction matrix based on lower face

据得到的预测率矩阵, T 表示对表情预测正确.

而对于一种表情的决策,基于上半脸的决策结果和基于下半脸的决策结果可能起到互补的正向作用,于是本文采用概率论中的加法定理和乘法定理进行融合表达式的设计,同时也可能起到互扰的反向作用,于是本文同时利用求对立事件的方式进行融合. 对于某些表情,需要上下半脸的协同作用,即全脸的决策信息,所以也将全脸的决策数据加入到融合中.

基于上半脸和下半脸决策的互补性,本文采用常用的融合方法,利用加法定理和乘法定理进行融合,如式(1)(2). 有些表情的识别需要全脸的协同作用,于是加入全脸信息,将全脸、上半脸和下半脸 3 部分融合,如式(4)(5). 利用求对立事件的方式

进行融合,如式(3)(6).

于是,融合表达式选取以下几种:

$U+L$ 模型

$$P = P(T|U) + P(T|L) \quad (1)$$

UL 模型:

$$P = P(T|U)P(T|L) \quad (2)$$

$\overline{(UL)}$ 模型

$$P = 1 - (1 - P(T|U))(1 - P(T|L)) \quad (3)$$

$W+U+L$ 模型

$$P = P(T|W) + P(T|U) + P(T|L) \quad (4)$$

WUL 模型

$$P = P(T|W)P(T|U)P(T|L) \quad (5)$$

$\overline{(WUL)}$ 模型

$$P = 1 - (1 - P(T|W))(1 - P(T|U))(1 - P(T|L)) \quad (6)$$

分别利用上述几种公式,对预测率矩阵进行融合计算,得出融合后的预测率矩阵,通过计算每行中最大的预测率所对应的表情标签作为当前表情的识别结果,最后统计识别正确的个数比总个数的百分数作为该次实验的识别率.

4.2 基于 Kappa 的加权融合实验

人脸每部分(全脸、上半脸、下半脸)在表情预测中是一种协同作用的结果,利用加、乘等方式进行线性融合有效描述了表情各子空间共同作用的决策模式,但考虑到两者作用的差异,本文利用线性加权方案将子空间在各表情识别过程中不同作用加以体现.本文采用样本学习获取具体权值,即根据样本预测的正确率生成权值.

基于样本正确率进行加权的方法存在一个问题就是正确率包含猜测因素,从而不能获得每一部分的有效信息.为去除猜测因素影响所导致预测正确率对信息描述的偏差,本文提出利用 Kappa 作为权值的线性加权融合模式,即采用 Kappa 统计量对每部分得出的预测率矩阵进行计算求解出 Kappa 系数,作为每部分的加权系数进行决策级融合计算.

Kappa 统计量的计算公式为

$$Kappa = \frac{Acc - E(Acc)}{1 - E(Acc)} \quad (7)$$

式中:Acc 为正确率; $E(Acc)$ 为正确率的期望值. Kappa 统计量的优势在于,它可去除计算中对表情样本的猜测成分,更加客观地反映每部分的有用信息.从式(8)可看出,使用正确率减掉的期望值也就是猜测因子,例如本文中讨论的人脸表情,对于具有 7 种表情的表情库中的任意一个表情图像,都有

1/7 的概率猜中它属于哪种表情.这个 1/7 就是猜测成分,而 Kappa 巧妙地去除了这种猜测成分,更加客观地反映特征本身携带的表情信息.

$$H(1 - Kappa) = -\ln(1 - Kappa) \quad (8)$$

因为 Kappa 的取值范围是 0 ~ 1,为计算方便,本文不直接使用 Kappa 值,而是对 Kappa 取对数 $H(1 - Kappa)$,融合后预测率的表达式为

$$P = H_w(1 - Kappa_w)P(T|W) + H_u(1 - Kappa_u)P(T|U) + H_l(1 - Kappa_l)P(T|L) \quad (9)$$

即将 3 部分利用每部分得到的 $H(1 - Kappa)$ 值,对每部分进行加权融合,得到最后的识别率矩阵.然后,选取每行的最大预测率作为该表情的最终识别结果.

5 实验及结果分析

5.1 表情库

本文在实验中采用 2 个公共人脸数据库:CMU 的 Cohn-Kanade (Action Unit, AU)-coded 人脸表情库和日本的 JAFFE 人脸表情库.

表情数据库 Cohn-Kanade 由美国 CMU 机器人研究所和心理学系共同建立,使用基于 AU 编码的数据库.图 13 为 Cohn-Kanade 表情数据库的样本示例.该数据库含有 100 个 18 ~ 30 岁的大学学生的面部表情图像序列,尺寸:640 × 490 像素,包括单个 AU 或者 AU 组合,其中 65% 为女性,15% 为非裔美洲人,3% 为亚洲和拉丁美洲人.由于 Cohn-Kanade 数据库中并未对每个采集对象采集所有的表情状态,采集数据的中性表情较少,因而在该数据库中,本文主要针对不包括中性表情的 6 种基本表情(惊奇、恐惧、厌恶、愤怒、高兴、悲伤)进行识别,验证本文所提出的 Kappa 加权融合表情识别算法.



图 13 Cohn-Kanade 表情数据库中的部分样本图像
Fig. 13 Part sample images in the Cohn-Kanade database

表情数据库 JAFFE 是日本 ART 建立的 10 个日本女人的 7 种基本表情——惊奇、恐惧、厌恶、愤怒、高兴、悲伤及中性.库中所有图像均存在姿势端正、光照充分、成像背景单一、无眼镜胡须等遮挡物等特点,共有 210 幅 256 × 256 像素大小的静态表情图像.在此数据库上针对 7 种基本表情进行实验,验

证本文所提出的表情识别方法的有效性. 图 14 是 JAFFE 表情数据库的样本示例.



图 14 JAFFE 表情数据库中的部分样本图像

Fig. 14 Part sample images in the JAFFE database

5.2 分类器及其参数设置

利用 JAFFE 和 Cohn-Kanade 两种表情库, 本文分别对上半脸、下半脸和全脸进行 Gabor 小波特征提取, 如表 1 所示, 并采用 3 种分类器: k -近邻 (k -nearest neighbor)、多层感知网络 (multi-layer perception) 和最小序列优化算法 (sequence minimum optimization, SMO)^[12-13] 分别进行分类测试.

表 1 JAFFE 库下的概率加权融合实验结果

Table 1 Experiment fusion results of probability weighting in JAFFE database %

| Gabor | SMO | k -NN | MLP |
|--------------------|-------|---------|-------|
| W | 77.14 | 84.29 | 74.29 |
| U | 61.43 | 61.43 | 58.57 |
| L | 54.29 | 62.86 | 55.71 |
| UL | 64.29 | 61.43 | 74.29 |
| $U+L$ | 64.29 | 61.43 | 65.71 |
| $\overline{(UL)}$ | 65.71 | 61.43 | 64.29 |
| WUL | 71.43 | 80 | 75.71 |
| $W+U+L$ | 71.43 | 80 | 75.71 |
| $\overline{(WUL)}$ | 71.43 | 80 | 75.71 |
| $H(1-Kappa)$ | 78.29 | 85.43 | 78.57 |

对于 3 种分类器, 本文引用开源数据挖掘软件 Weka 中提供的算法, 对参数进行了调优设置, 其中 k -NN 搜集算法使用默认参数 Linear NN Search. 对于 MLP, 本文对比 Weka 文档中提供的参考经验值, Hider Layer: 8、10、12 和 Training Time: 50、100、150、200, 共计 12 种参数组合, 最后根据识别率选取网络层数为 10, 训练次数为 100. SMO 是支持向量机 (support vector machine, SVM) 的一种变型算法, 本文根据表情数据的特征, 选取多项式函数作为核函数, 实现特征空间到高维的映射, 其中一个重要的参数就是核函数的次数 Degree, 经验范围为 0~6, 经测试, 根据正确率最终将 Degree 选定为 3.

5.3 实验及结果分析

本文分别在 Cohn-Kanade 数据库和 JAFFE 数据库上对所提出的 Kappa 融合表情识别算法进行训练和测试. 训练阶段包括 2 个部分: 第 1 部分分类器训练, 在训练集 1 上利用上半脸、下半脸和全脸数据进行相应分类器训练; 第 2 部分融合模型 Kappa 权值训练, 利用训练集 2 的数据进行表情识别, 通过统计表情分类结果计算各面部分块的 Kappa 值. 测试阶段, 对测试集表情利用 Kappa 融合模型进行表情识别, 评测其表情识别效果. 测试阶段采用了 2 种方案, 第 1 种采用训练数据进行测试, 第 2 种采用非训练数据, 即测试集数据进行表情识别效果实验分析. 实验分别采用 3 种分类器, 分别对基于全脸、上半脸和下半脸和 Kappa 加权融合的表情识别效果进行了比较和分析.

5.3.1 采用训练集进行测试

考虑到样本数量的限制, 本文首先采用更多样本用于分类器训练和 Kappa 融合系数统计, 同时利用 Kappa 系数统计阶段的训练数据集 2 样本进行测试. 具体训练和测试样本选取方案如下: 在 JAFFE 数据库上, 140 幅作为训练集 1 (每种表情 20 幅), 剩余的 70 幅 (每种表情 10 幅) 作为训练集 2, 并作为测试数据. Cohn-Kanade 数据库中的 30 幅为训练集, 每种表情 5 幅, 剩余 18 幅 (每种表情 3 幅) 同时作为训练集 2 和测试集. 实验结果如表 1、2 所示. 同时, 本文对 4.1 所提出的几种可行的融合方案进行了实验, 如表 1、2 分别为在 JAFFE 和 Cohn-Kanade 数据库利用各融合方案的表情识别结果.

5.3.2 采用测试集进行测试

在 5.3.1 的实验中, 采用 Kappa 统计的训练数据作为测试数据, 存在训练结果对样本的拟合性好的局限性, 不具有广泛意义. 为了进一步验证融合表情识别算法的有效性, 本文对 2 个数据库重新进行训练样本和测试样本数据分配, 定义了 3 个数据集, 即训练集 1、2 和测试集, 利用非训练集的测试样本对融合表情识别算法进行验证. 本文从 JAFFE 数据库中选取 119 幅作为训练集 1 用于分类器训练, 每种表情选取 17 幅; 56 幅作为训练集 2, 用于 Kappa 权值系数统计, 每种表情 8 幅; 另取 35 幅作为测试集. 这里需要说明, 为了在一定程度上改善训练样本少的问题, 采用对重复复制每个样本 1 遍的策略增加训练样本, 即 119×2 幅图像作为训练集 1, 56×2 幅作为训练集 2. 在 Cohn-Kanade 数据库中选取其中 24 幅样本图像作为训练集 1, 每种表情 4

幅;12幅作为训练集2,这里用于统计计算 Kappa 值,每种表情2幅;另取12幅作为测试集,每种表情2幅,对所提出的基于 kappa 参数加权融合方案进行验证. 同样,采用每个样本复制1遍的方法增加样本,即以 24×2 幅表情图片作为训练集1, 12×2 幅作为训练集2. 其实验结果如表3、4所示.

表2 Cohn-Kanade 库下的概率加权融合实验结果

Table 2 Experiment fusion results of probability weighting in Cohn-Kanade database %

| Gabor | SMO | KNN | MLP |
|-----------------------|-------|-------|-------|
| <i>W</i> | 38.89 | 33.33 | 55.56 |
| <i>U</i> | 38.89 | 22.22 | 44.44 |
| <i>L</i> | 50.00 | 27.78 | 50.00 |
| <i>UL</i> | 44.44 | 16.67 | 50.00 |
| <i>U + L</i> | 44.44 | 16.67 | 55.56 |
| $\overline{(UL)}$ | 44.44 | 16.67 | 55.56 |
| <i>WUL</i> | 44.44 | 22.22 | 61.11 |
| <i>W + U + L</i> | 44.44 | 22.22 | 61.11 |
| $\overline{(WUL)}$ | 44.44 | 22.22 | 61.11 |
| $H(1 - \text{Kappa})$ | 64.44 | 37.78 | 61.11 |

表3 JAFFE 库下测试集的概率加权融合实验结果

Table 3 Experiment fusion results of probability weighting in the test set of JAFFE database

| Gabor/% | SMO | KNN | MLP |
|-----------------------|-------|-------|-------|
| <i>W</i> | 57.14 | 75.51 | 55.10 |
| <i>U</i> | 46.94 | 53.06 | 48.98 |
| <i>L</i> | 48.98 | 46.94 | 42.88 |
| <i>UL</i> | 59.18 | 38.58 | 57.14 |
| <i>U + L</i> | 63.27 | 53.06 | 59.18 |
| $\overline{(UL)}$ | 61.22 | 53.06 | 59.18 |
| <i>WUL</i> | 61.22 | 38.78 | 55.10 |
| <i>W + U + L</i> | 59.18 | 75.51 | 61.22 |
| $\overline{(WUL)}$ | 59.18 | 55.10 | 55.10 |
| $H(1 - \text{Kappa})$ | 63.27 | 75.51 | 59.18 |

5.3.3 实验结果分析

从表1~4的整体分析来看:1)部分融合后的结果高于融合前,但也有融合后识别率变低的,这是因为上半脸、下半脸对某些表情的预测率并非独立;2)加入全脸后,融合结果好于单独上下半脸融合的,因为有些表情依赖于全脸的协同作用,并不是单

独上下半脸能展现的,即使重新将上下半脸进行融合也无法实现全脸的整体效果. 与其中最好的融合识别结果相比,基于 Kappa 加权融合表情识别结果绝大部分都高于对应的最好识别结果. 该实验结果充分证明了 Kappa 统计量在表情分块对表情信息的捕捉性能方面具有更好的度量,因而利用该加权各表情分块的作用融合模型,能进一步提高表情识别结果.

表4 Cohn-Kanade 库下测试集的概率加权融合实验结果

Table 4 Experiment fusion results of probability weighting in the test set of Cohn-Kanade database %

| Gabor | SMO | KNN | MLP |
|-----------------------|-------|-------|-------|
| <i>W</i> | 41.67 | 37.50 | 45.83 |
| <i>U</i> | 37.50 | 37.50 | 33.33 |
| <i>L</i> | 33.33 | 33.33 | 37.50 |
| <i>UL</i> | 37.50 | 29.17 | 37.50 |
| <i>U + L</i> | 37.50 | 33.33 | 37.50 |
| $\overline{(UL)}$ | 41.67 | 33.33 | 37.50 |
| <i>WUL</i> | 37.50 | 29.17 | 45.83 |
| <i>W + U + L</i> | 45.83 | 33.33 | 50.00 |
| $\overline{(WUL)}$ | 41.67 | 33.33 | 50.00 |
| $H(1 - \text{Kappa})$ | 41.67 | 37.50 | 54.17 |

不难发现,5.3.1小节中基于 Kappa 参数加权融合的实验结果优于5.3.2小节的实验结果. 这是由于5.3.1中训练集2既作为训练集又作为测试集,得到的 Kappa 参数自然依赖于测试集,而在5.3.2中,Kappa 参数对测试集没有这么强的依赖性. 尽管如此,5.3.2小节的实验结果也足以证明该融合方法的有效性.

6 结论

1) 基于人脸不同部位对表情识别的贡献不同,本文对人脸进行上下分块处理,并对每部分提取 Gabor 小波特征,采用2个表情库作为训练和测试的原料,采用 SMO、MLP、KNN 三种分类器分别测试. 实验结果表明,人脸上下半脸对表情识别的贡献是不同的,经过融合提高了表情识别的正确率.

2) 通过使用常用的融合方法,对上半脸和下半脸进行加法和乘法方式的融合,表明单独使用上半脸、下半脸不能很好描述所有表情,有些表情需要全脸的协同作用.

3) 利用每一部分的预测率矩阵计算出相应的 Kappa 值,利用 Kappa 值对每部分进行加权融合,加强了每部分有用信息的融合,实验结果显示,基于 Kappa 的决策级融合方法显著提高了识别率.

参考文献:

- [1] FINCHAM F D. Communication in marriage [J]. Handbook of Family Communication, 2004, 84: 83.
- [2] EKMAN P, FRIESEN W V. Facial action coding system: a technique for the measurement of facial movement [M]. Palo Alt: Consulting Psychologists Press, 1978.
- [3] PICARD R W. Affective computing [M]. Cambridge: MIT press, 2000.
- [4] AHN H, PICARD R W. Affective cognitive learning and decision making: the role of emotions [C] // The 18th European Meeting on Cybernetics and Systems Research (EMCSR 2006). Vienna: Osterreichische Studiengesellschaft fur Kybernetik (ASCS), 2006: 1-6.
- [5] DONATO G, BARTLETT M S, HAGER J C, et al. Classifying facial actions [J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1999, 21: 974-989.
- [6] CUNNINGHAM D W, KLEINER M, WALLRAVEN C, et al. Manipulating video sequences to determine the components of conversational facial expressions [J]. ACM Transactions on Applied Perception (TAP), 2005 (2): 251-269.
- [7] COHEN J A. A coefficient of agreement for nominal scales [J]. Educational and Psychological Measurement, 1960, 20: 37-46.
- [8] 李倩. 基于 Gabor 小波变换的人脸表情识别技术研究 [D]. 南昌: 华东交通大学信息工程学院, 2011.
- LI Qian. The research of facial expression recognition technology based on Gabor wavelet transformation [D]. Nanchang: College of Infrmation Engineering, East China Jiaotong University, 2011. (in Chinese)
- [9] 杜智斌. 表情识别系统的设计与实现 [D]. 武汉: 武汉理工大学计算机学院, 2009.
- DU Zhi-bin. Design and implementation of facial expression recognition system [D]. Wuhan: College of Computer Science, Wuhan University of Technology, 2009. (in Chinese)
- [10] 姚伟, 孙正兴, 张岩. 面向脸部表情识别的 Gabor 特征选择方法 [J]. 计算机辅助设计与图形学学报, 2008, 20(1): 79-84.
- YAO Wei, SUN Zheng-xing, ZHANG Yan. Optimal gabor feature for facial expression recognition [J]. Journal of Computer-aided Design & Computer Graphics, 2008, 20(1): 79-84. (in Chinese)
- [11] AHMED N, NATARAJAN T, RAO K R. Discrete cosine transform [J]. IEEE Transaction on Computer, 1974, 100: 90-93.
- [12] PLATT J. Sequential minimal optimization: a fast algorithm for training support vector mechines [M] // Advance in Kernel Methods & Mdash, Support Vector Learning. Cambridge: MIT Press, 1999: 185-208.
- [13] 李航. 统计学习方法 [M]. 北京: 清华大学出版社, 2012.

(责任编辑 吕小红)