

基于人机互动系统情感分析计算研究*

杨杰¹, 赵强²

(1. 湖南科技学院 计算机与通信工程系, 湖南 永州 425100; 2. 国防科学技术大学 计算机学院, 长沙 410073)

摘要: 由于人的情感包含大量的信息,特别是面部的表情与手势图像中存在大量干扰信息,造成了情感计算正确率下降,为此提出了一种从用户的面部表情与手势中系统地分析情感线索的方法。对面部表情与手势进行分析是感情丰富的人机交互系统的基本组成部分,采用非语言线索算法来判断用户的感情状态。从图像序列中提取与表情相关的特征,通过智能法则系统分析用户的情感状态,得到的感情信息最终与用户的真实反映相适应。最后采用基于主体接口技术来处理如沮丧或愤怒等特定的情绪状态。实验结果表明,提出的方法能够准确地分析计算出用户的情感信息。

关键词: 情感计算; 表情识别; 人机交互; 图像序列

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2012)07-2763-03

doi:10.3969/j.issn.1001-3695.2012.07.099

Analysis and calculation on human-computer interaction system of emotion

YANG Jie¹, ZHAO Qiang²

(1. Dept. of Computer & Communication Engineering, Hunan University of Science & Engineering, Yongzhou Hunan 425100, China; 2. School of Computer Science, National University of Defense Technology, Changsha 410073, China)

Abstract: Especially facial expression and gesture image existed in a large number of interference information disadvantages, resulting in affective computing accuracy down, this paper presented a from the user's facial expressions and gestures in a systematic analysis of the emotional cues of method. First analysis of facial expressions and gestures were the feelings of the rich interactive basic part of the system, and then the nonverbal cued algorithm to judge user emotional state. At the same time, these papers extracted from an image sequence and expression of related characteristics, through the intelligent rule system analysis of the user's emotional state, and reflected the user's suit. Finally based on the main interface technology, it treated such as frustration or anger and other specific emotional states. The experimental results show that, the proposed method can accurately analyze and calculate the user's emotional information.

Key words: affective computing; facial expression recognition; human-computer interaction; image sequence

当前的信息处理和可视化系统能够提供先进、直观的方式来接收输入的信息并将其输出给用户。因此人机交互系统主要利用用户情感状态的信息,目前该领域处于计算机视觉和人工智能的最前沿。这样的接口让那些不太精通计算机技术的人们能够更有效地使用计算机,克服了相关的恐惧与偏见。此外,大多数与情感有关的面部表情和体态信息都是具有一定的通用性,因为它们都能够被不同时代文化背景的人所认可。所以说引入情感词典可以增强人机互动系统应用程序的情感特征,这种词典可以详细描述各种面部表情与体态所表达出来的信息,从这些信息中就可以推断出用户所处的情感状态。

目前相关的研究已经取得了很大的进展。如果要弄清某个表情或情感的具体意义,首先要模拟人类思维运转的方式,努力识别计算这种感情。一般来说,语言可能会引起误解或歧义,但是一些特征传递信息的方式比语言更传神、更贴切。人们已经为核查人类表达的各个方面付出了很多努力。文献[1,2]中提出一种改进的算法,该算法表明,如果能考虑到多模式信息,这种方式将会取得更好的效果。本文提出了一种从用户的面部表情与手势中系统地分析计算情感线索的新方法。

1 人机互动系统中的情感分析

1.1 面部表情的情感分析

人们很早以前就希望能从面部表情中辨别出情感状态^[3],在过去的20年里,研究人员对人的面部表情作了大量的研究^[4,5]。从某些方面看,从面部表情辨别情感状态最突出的问题与从声音辨别情感状态很相似,但是从其他方面来看,两者却有很多不同点。如果是从面部表情中辨别情感状态,那么就需要将原型情感的例子进行分类,Sun等人^[6]认为,可以通过面部表情来表现人们原本的情感状态。他们将一种变性技术应用于探查原型表情之间的中间体上,可以在一定程度上计算出面部表情信息特征,但是该特征信息较弱。本文采用了一种改进的方法来给出与脸部参数在互动评价空间的映射关系。

1.2 对手势进行情感分析

用户的手势是一种能够有效传递人类感情的方法,近年来,观察并分析手势已经成为人机互动系统研究一个重要组成部分。一般来说,一个简单的手势就能准确地传递信息,如果

收稿日期: 2011-11-28; 修回日期: 2011-12-30 基金项目: 湖南省自然科学基金资助项目(11JJ6056)

作者简介: 杨杰(1976-),男(瑶族),湖南永州人,副教授,硕士,主要研究方向为情感技术、数据挖掘(hnyangjie1976@163.com);赵强(1975-),男,湖南永州人,副教授,博士研究生,主要研究方向为移动计算、网络协议。

一个人用手蒙住耳朵,表明他不想再听下去了,这个简单的手势比任何言语都更加清晰明了。通过分析手势进行情感识别是一项复杂的研究,其主要包含了手势建模、手势分析、模式识别、机器学习甚至心理语言学方面的研究。

对手势情感识别的第一阶段是要选择建立一个手势数学模型,该数学模型既要考虑手势的空间特征,也要考虑时间特征。建模方法对于手势分析的本质与性能起着十分关键的作用。决定了模型以后就进入了分析阶段,本文分析阶段是指从单个或多重的视屏输入流中提取图像特征,并对图像特征中的模型参数进行计算,这些参数取决于建模方式,对手势与手运动的路径进行了详细的描述。分析过程中最重要的问题包括手势的定位与追踪,以及选择合适的图像特征。根据已被接受的模型与某些强加的语法规则对模型参数进行分类与解释。语法不仅显示了手势指令的内部句法,也显示了手势与其他沟通模式(如语言、眼神、面部表情等)之间的相互作用的可能性。

2 对面部表情进行分析

2.1 提取面部表情

面部表情分析包括几个步骤:观察并追踪面部表情;找出脸部各个区域(如眼睛、嘴、鼻子)的特征;根据面部特征(如面部各区域的特征点)的运动规律或脸部的解剖信息对面部表情进行建模。尽管固定面部动作模式^[7]给出 MPEG-4 兼容动画所需的所有要素,但是由于其缺少一个清晰的定量定义框架,所以还是不能用它来分析视频中的表情。为了测量真实影像序列中的固定动作模式,必须要规定它们之间的影像以及特定的 FDP 特征点,这些 FDP 特征点即相当于面部折点。

本文提出的面部特征提取机制是一个有层次的、成熟的机制,以应对出现不同物体或者在真实影像序列中的同一个物体在不同的条件下引起的大变动。根据面部表情以及表情特征的大致定位作出了一个先验假定。多层次机制在优化的范围内都支持面部信息的逐渐式显现,归纳产生了一些相关的信息,并且使面部特征变得逐步视觉化。检测面部表情首先需要检测脸部的各个区域或信息点,根据它们处于同一个面部区域的可能性以及对最明显的点或区域的识别将它们合并起来。主要的面部特征如眼睛、鼻子、嘴在分区的、任意转动的脸上都是非连续的。

2.2 实验结果分析

本文采用梯度滤波器检测眼部以及眼睛与嘴之间的区域,可以有效地估算出眉毛与鼻子的位置。根据已经检测到的脸部特征的位置能够计算并估量出特征点。本文所有的实验数据来自 www.image.ntua.gr/ermis,实验数据采用的都是资料库中的人脸图像。

图 1 显示了影像序列中一个具有代表性的脸框,图 2 显示了经过皮肤检测与分割后的主要面部特征,图 3 显示了估量出的眼睛、嘴、眉毛和鼻子的位置,图 4 显示了用于计算 FP 距离的最初的温和表情。

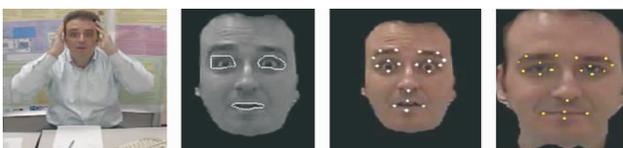


图 1 输入视频中 图 2 检测主要 图 3 情感特征 图 4 温和
原始图像帧 面部特征 点检测 情感

3 对手势进行分析

3.1 手势分析与追踪

本文从手势的运动中提取与情感相关的特征,在提出的人机交互系统中,首先建立一个移动皮肤掩码,即在相继的框架中可以跟踪到相关皮肤的区域,通过追踪这些皮肤掩码的重心,就能够估量出用户的运动状况。为了采用计算灯光系统,在构建系统结构时,本文同时考虑了与输入图像的期望特征相关的先验知识,将人的头部定位于框架上半部分的中间位置,而将手部定位于相对靠下的角落位置。同时,本文还集中给出了指腹的运动,它是手语胳膊链的末端效应器,因此其触觉也是比较灵敏的。

如同人的脸部检测一样,通过计算 C_r/C_b 图像值的联合概率率来计算肤色概率矩阵,如图 5 所示。图 6 显示的是从带阈值的皮肤概率矩阵中可以得出肤色掩码,根据当前框架与下一个框架之间像素的差异可以计算出可能会移动的区域,这些区域可能会导致运动掩码。如图 7 所示,这些掩码不能显示运动的方向与幅度,只能显示是否在运动,它们可以用来加速运算,颜色掩体与运动掩体中都混杂着那些类似于皮肤的有色物体的噪声与物体。为了解决这个问题,本文运用形态滤波来清除掩体中的物体,所有的形态操作都需使用半径为图像宽度 1% 的磁盘结构元素完成。颜色掩体的距离转换是事先计算好的,只有符合尺寸的对象才能得以保留。这些保留下来的物体则变成了初始颜色掩体进行形态重建的成分。

将处理后的皮肤与移动掩体融合起来创造移动的皮肤掩体。如图 8 所示,将头移除之后,移动的皮肤掩体由许多大块的部分构成,下一个框架里又产生了另一个移动皮肤掩体,也是一一对应的。如图 9 所示,图中主要分为两部分,一个为手部,另外一部分为脸部,其在颜色掩体上相对应,处于类似区域内的物体基于物体中心距离对应。在图 10 中,红色标记表示检测到的用户的右手中心运动的轨迹位置,绿色表示左手中心的位置(见电子版)。当双手合并或者分开时,如人们鼓掌时,对用户右手能得到的最左边的物体与用户左手能得到的最右边的物体进行了一项新的匹配。在下一个移动皮肤掩体中,物体匹配的是计算掩体流,也就是描述每个框架中运动方向与运动幅度的矢量。

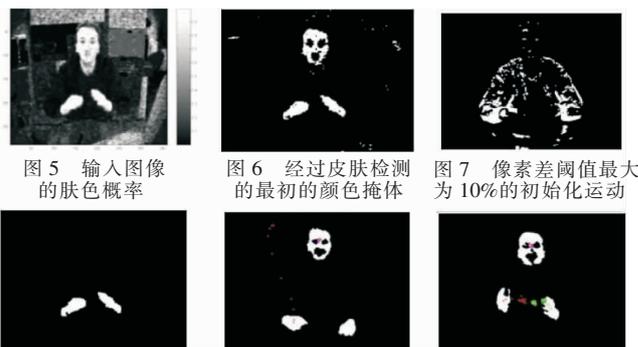


图 5 输入图像 图 6 经过皮肤检测 图 7 像素差阈值最大
的肤色概率 的最初的颜色掩体 为 10% 的初始化运动
图 8 形态重建后 图 9 抬手时对手中 图 10 拍手时的
检测到的手部 一个物体的追踪轨迹 跟踪轨迹

3.2 采用隐马尔可夫模型进行手势分类

本文采用隐马尔可夫模型来对手势进行分类。所提出的识别器是由与标准手势类型一致的 M 个不同隐马尔可夫模型组成。如表 1 所示,实验取 $M = 7$,采用第一个从左至右的模

型,该模型由不同数据组成(每个隐马尔可夫模型一个),通过算法的学习过程来确定模型内部状态 $G_{k,j}$ 。例如,第三个识别抬手的隐马尔可夫模型仅由 $G_{3,1}$ 、 $G_{3,2}$ 和 $G_{3,3}$ 三个状态组成的,而更复杂的手势类型,如拍手姿势,需要多达八种状态来充分展现其对应的隐马尔可夫模型。

表 1 手势、类型特征实验个数(隐马尔可夫模型输入/输出结果)

特征信息	$X_{lh} - X_{rh}, X_f - X_{rh}, Y_f - X_{lh}, Y_{lh} - Y_{rh}, Y_f - Y_{rh}, Y_f - Y_{lh}$ $C_f = (X_f, Y_f)$ 为面心坐标轴, $C_{rh} = (X_{rh}, Y_{rh})$ 为右面心坐标轴, $C_{lh} = (X_{lh}, Y_{lh})$ 为左面心坐标轴
手势类型	高频拍手, 低频拍手, 低速抬手, 快速抬手, 手靠头手势, 手靠头姿势

3.3 实验结果

为了确定提出的算法的有效性,本文进行了实验。实施测试和识别假定运算规则实验(三位男性手势顺序),用一台网络照相机以每秒 10 个镜头的速度捕捉三位男性的手势序列,每个手势的最长持续时间为 3 s。每种手势类别都需 15 个序列,3 个用于隐形马尔可夫模型初始化参数,7 个用于训练和参数的再评估,5 个用于测试。每个训练序列由大约 15 个画面组成,本文在脱机状态下选择画面以取得手势类型的典型例子。测试序列取样以每秒 5 个画面的速度进行,以便出现大幅的运动,这样可获得 94.3% 的总体识别率。

从获得的结果可以看出,意大利式手势和拍手—高频之间的分类错误主要是由意大利式手势在不同个体上的变化引起的,因此,对隐马尔可夫模型的分级器进行个体训练有利于区分两者。

4 多模态情感计算分析

4.1 面部情感分析子系统

面部情感分析子系统是本文提出的情感识别系统的主要部分,本文将分析手势以证明该子系统的有效性。首先将一个 15 元素长的矢量 f 作为情感分析次系统,将该矢量与 f_i [15] 相对应。如图 11 所示,使用矢量 G 作为输入与用 f 作为矢量产生的 FAP 值一样。 G 的成分显示了相应的 FAPs 的测定值。

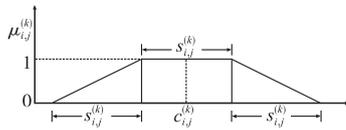


图 11 隶属函数形式

假定 $X_{i,j}^{(k)}$ 为包含在情感 k -th 剖面里的 FAPF_j 的变化范围。如果 $C_{i,j}^{(k)}$ 及 $S_{i,j}^{(k)}$ 分别为中点和间隔 $X_{i,j}^{(k)}$, 那么用图 11 中的从属函数 $\mu_{i,j}^{(k)}$ 描述一个模糊类 $A_{i,j}^{(k)}$, 假定 $A_{i,j}^{(k)}$ 对应剖面 $P_i^{(k)}$, 面部状态 $P_i^{(k)}$ 与 b_i 分别与剖面 $P_i^{(k)}$ 和情感 i 相对应,可以用方程式进行计算:

$$P_i^k = \prod_{A_{i,j}^{(k)} \in \Delta_{i,j}^{(k)}} r_{i,j}^{(k)} \quad b_i = \max_k (p_i^{(k)}) \quad (1)$$

其中: $r_{i,j}^{(k)} = \max \{g_i \cap A_{i,j}^{(k)}\}$ 表达出了输入特征矢量的第 i 个元素与类别 $A_{i,j}^{(k)}$ 之间的关联。实际上, $g = A'(G) = \{g_1, g_2, \dots\}$ 是一个因简单模糊程序引起的模糊输入矢量。如果需要做一个艰难的关于观测到的情感的决定,那么可使用以下方程:

$$q = \arg \max_i b_j \quad (2)$$

各种情感剖面与模糊交集相对应,并且通过 $t(a, b) = a \times b$ 形式的 τ -norm 来实施。类似的观点是一个观测到的特征向量与情感结果相对应,该情感源于一个模糊的集合。

4.2 情感分析子系统

通过分析手势来证实面部表情分析子系统的有效性。传统的算法中,由于过于抽象而难以表现一个具体表情。然而在特定互动情况下,一些手势明显与某个特定表情关联。快速地拍手表达快乐和满足,而其他手势则可体现出那些使用者传达出的情感。尤其是手迹中的量化特征(如动作的速度和幅度)可以让观测到的情感程度更明显。例如,随着拍手速度加快,满足转为快乐,甚至是狂喜。如第 3 章中所提到,头部中心和手的位置形成的特征是两序列,该序列是一种 HMM 分级器,分级器的输入信息与特定的手势类别相对应。表 2 显示的是检测到的六种典型情绪的手势。

表 2 手势与情感状态之间的联系

情感	手势类别
快乐	高频拍手
悲伤	手靠头姿势
愤怒	快速抬手,意大利化的手势
恐惧	手靠头手势,意大利化的手势
厌恶	慢速抬手,低频拍手
惊讶	手靠头手势

给定一个特定交际情景,将相对应的手势类别与情感组合成逻辑形式。表 2 显示的是特定手势可能与不止一个手势类别相对应。例如,拍手的手势,高频表示快乐,而低频拍手则表示讽刺并有可能表示一个厌恶的面部表情。

实际生活中,源于隐马尔可夫模型导出的手势类型通过使用表 2 中的信息转换成情感状态指示器。假定 EI_k 为情感状态 $k(k = \{1, 2, 3, 4, 5, 6\})$ 的情感指示器,与表 2 中的一个情感状态在出现顺序上相对应,即 $1 \rightarrow \text{Joy}$, $6 \rightarrow \text{Surprise}$ 。GCS = $\{gc_1, gc_2, \dots, gc_N\}$ 为能被隐马尔可夫模型分级器识别的手势类别集合,分类器 ($N = 7$), $GCS^k = GCS$ 是与情感状态 K 有关的表情类别的集合。其中, $p(gc_i)$ 表示隐马尔可夫模型分级器中手势类别 gc_i 的概率,使用以下公式对 EI_k 进行计算。

$$EI_k = \max_{gc_i \in GCS^k} (gc_i) \quad (3)$$

4.3 综合决策人机交互系统

本文提出的系统的最后一步是将面部表情分析子系统与情感手势分析子系统合并为一个可显示用户可能的情感系统。尽管面部是用户表情最主要的识别器,识别伴随的手势增加了面部表情子系统结果的可信度。在目前的实际运用中,将这两个子系统联合在一起形成加权总和。

假设 b_k 为一观测到的序列表现出 k -th 情感状态的程度,序列由面部表情分析次系统获得, EI_k 为对应的情感状态指示器, EI_k 从情感手势分析次系统获得,整个理念程度取决于

$$d_k = w_1 \times b_k + w_2 \times EI_k \quad (4)$$

由于涉及情感状态评估, w_1 及 w_2 要用于两个子系统可靠度测定。本文基于文献[8]中的实验结果,取 $w_1 = 0.75$, $w_2 = 0.25$, 这些值使得情感手势分析子系统在面部表情分析子系统中产生模糊结果,同时后一个子系统则变成整个决策系统的主要部分。对于图 1 中所示的输入序列,情感手势分析子系统总是出现让人意外的选择,用这来论证面部表情分析子系统的可信度,一般是 85%。

5 结束语

本文给出了一个整体的情感建模方法以及(下转第 2769 页)

(上接第 2765 页)其在 MMI 中的运用。结果表明,将视频序列的数量特征信息转换成估量用户的情感状态是可能有效的,其转换是基于一个模糊规则体系结构完成的。该体系结构将情感表现以及人类表情的内在特征考虑在内,这些特征可用于简单地表现目的。情感模型领域的未来研究包括丰富的带有更多感情手势的手势词汇,以及以特征为基础的描述。

参考文献:

- [1] WU Ying, HUANG T S. Hand modeling, analysis, and recognition for vision-based human computer interaction [J]. *IEEE Signal Processing Magazine*, 2010, 18(3): 51-60.
- [2] RAOUZAIYOU A, TSAPATSOUZIS N, KARPOUZIS K, *et al.* Parameterized facial expression synthesis based on MPEG-4 [J]. *EURASIP Journal on Applied Signal Processing*, 2008, 10(4): 1021-1038.
- [3] VOTSIS G, DROSOPOULOS A, KOLLIAS S. A modular approach to facial feature segmentation on real sequences [J]. *Signal Processing, Image Communication*, 2009, 18(3): 67-89.
- [4] 孙权森, 曾生根, 王平安, 等. 典型相关分析的理论及其在特征融合中的应用 [J]. *计算机学报*, 2005, 28(9): 1524-1533.
- [5] 刘永俊, 陈才扣. 基于差空间的散度差鉴别分析及人脸识别 [J]. *计算机应用*, 2008, 26(10): 2460-2462.
- [6] SUN Quan-sen, ZENG Sheng-gen, YANG Mao-long, *et al.* Combined feature extraction based in canonical correlation analysis and face recognition [J]. *Journal of Computer Research and Development*, 2005, 42(4): 614-621.
- [7] 杨静宇, 金忠, 郭跃飞. 人脸图像有效鉴别特征提取与识别 [J]. *南京理工大学学报*, 2008, 24(3): 193-198.
- [8] KRUEGER V, SOMMER G. Gabor wavelet networks for object representation [J]. *Journal of the Optical Society of America*, 2010, 19(6): 1112-1119.
- [9] SUN Quan-sen, ZENG Sheng-gen, YANG Mao-long, *et al.* The theory of canonical correlation analysis and its application to feature fusion [J]. *Chinese Journal of Computers*, 2005, 28(9): 1524-1533.
- [10] WONG Jia-jun, CHO S Y. Facial emotion recognition by adaptive processing of tree structures [C] // Proc of ACM Symposium on Applied Computing. New York: ACM Press, 2006: 23-30.
- [11] CHO S Y, WONG Jia-jun. Probabilistic based recursive model for face recognition [C] // Lecture Notes in Computer Science, vol 3641. [S. l.]: Springer-Verlag, 2005: 1245-1254.
- [12] SUNG J, KANADE T, KIM D. A unified gradient-based approach for combining ASM into AAM [J]. *International Journal of Computer Vision*, 2007, 75(2): 279-309.
- [13] KRUEGER V, SOMMER G. Gabor wavelet networks for object representation [J]. *Journal of the Optical Society of America*, 2009, 19(6): 1112-1119.
- [14] LIU C, WECHSLER H. Independent component analysis of Gabor features for face recognition [J]. *IEEE Trans on Neural Networks*, 2009, 14(4): 919-928.