

# 基于特征流的面部表情运动分析及应用\*

金辉<sup>+</sup>, 高文

(哈尔滨工业大学 计算机科学与工程系, 黑龙江 哈尔滨 150001)

## Analysis and Application of the Facial Expression Motions Based on Eigen-Flow

JIN Hui<sup>+</sup>, GAO Wen

(Department of Computer Science and Engineering, Harbin Institute of Technology, Harbin 150001, China)

+ Corresponding author: Phn: 86-451-86416485, E-mail: huijinbox@hotmail.com

<http://www.hit.edu.cn>

Received 1999-08-03; Accepted 1999-11-23

**Jin H, Gao W. Analysis and application of the facial expression motions based on eigen-flow. *Journal of Software*, 2003,14(12):2098-2105.**

<http://www.jos.org.cn/1000-9825/14/2098.htm>

**Abstract:** Analysis and recognition of the facial expressions play an important role in both the social society and the affective computing in the field of the computer science. There are three primary methods for the analysis of expression motive features: methods based on the facial geometrical structure features, on the definition of the expression space based on the eigen-face, and on the motion pattern matching. This paper extracts the feature regions of the expressions based on the facial physics-muscle model and evaluates the optical flow of the expression image sequences. The eigen-flow vectors can be calculated to constitute the eigen-sequences, and therefore, the expressions can be analyzed. The recognition system is implemented as an agent in the multi-perception machine and it is used as part of the video input for understanding the human body languages.

**Key words:** optical flow; eigen-sequence; combined expression analysis; multi-perception machine

**摘要:** 面部表情的分析与识别,不但在社会生活中具有普遍意义,而且在计算机的情感计算方面也起着有重要作用.关于表情运动特征的分析,有根据人脸面部几何结构特征的变化来分析的,有根据特征脸的概念定义的“表情空间”来分析的,也有从特征点跟踪的方法或运动模板的角度来分析的.基于人脸面部物理-几何结构模型,提取面部表情特征区域,通过动态图像序列中的光流估计,计算其运动场,进而计算特征流向量,把一组图像序列的运动向量组成运动特征序列,对表情的运动进行分析.该系统作为一个智能体应用到多功能感知机中,作为视频通道输入的一部分来理解人类的体势语言信息.

**关键词:** 光流;特征序列;混合表情分析;多功能感知机

中图法分类号: TP391 文献标识码: A

\* Supported by the National Natural Science Foundation of China under Grant No.69789301 (国家自然科学基金); the National High Technology Development 863 Program of China under Grant No.863-306-ZT03-01-2 (国家高技术研究发展计划(863)); the Hundred People Plan of the Chinese Academic of Sciences of China (中国科学院“百人计划”)

第一作者简介: 金辉(1972-),女,吉林长春人,博士,主要研究领域为计算机图像处理,模式识别.

面部表情的分析与识别,是实现智能化的人机接口必不可少的一个重要方面,是关于情感计算的一个重要组成部分,对促进计算机视觉系统、建模和数据库的发展都有重要作用.在律师、公安、医疗等需要解释脸部信号的特殊部门以及在商业应用方面,如可视电话和电视会议上,都有重要的应用价值.

表情是经面部区域的运动而产生的,表情的动作变化是表情聚类的实质和核心.对表情的分析,赵力庄<sup>[1]</sup>曾把特征脸(eigenface)的思想应用到表情识别中,用静止的单帧表情图像作为训练集,把表情投影到不同的空间区域,但存在的不足之处是:

- 用静止的单帧图像进行训练和识别,反映不出表情动作的变化,没有包含时间信息和运动信息;
- 提取的特征是局部的和线性的;
- 扩展性不够好,对混合表情难以识别.吕岩<sup>[2]</sup>等人曾对表情的运动图像序列作了分析识别,但所提取的特征只是帧内的静止灰度信息,而没有帧间的运动信息.在实验时,忽略了比较接近的、易误识的几种表情.

Mase<sup>[3]</sup>使用光流来跟踪运动单元.其主要的局限性是没有使用物理模型;脸部的运动是用静态的公式化表示的,没有用动态的优化模型;Irfan A.Essa<sup>[4]</sup>等人用图像序列作为输入的计算机视觉系统来观察脸部的运动单元.视觉观察与感知是通过优化估计光流方法与描绘脸部结构的几何、物理肌肉模型相结合得到的.这种建模方法产生了一个随时间变化的脸部形状的空间模板和一个独立的肌肉运动群的参数化表征.并从图像序列中选择出固定帧数的10帧图像来识别,在时序运动分析上受到限制.另外,James J.Lien<sup>[5]</sup>等人应用特征点跟踪、流跟踪和边缘检测这3种方法对表情进行分析.

本文基于人脸面部物理-几何结构模型,提取面部表情特征区域,通过动态图像序列中的光流估计,计算其运动场,进而计算特征流向量,把一组图像序列的运动向量组成运动特征序列,对表情的运动进行分析.该系统作为一个智能体应用到多功能感知机中,作为视频通道输入的一部分来理解人类的体势语言信息.

## 1 表情模型与编码

脸部是主要的感知输入和主要的信息输出地带.它是一个多信号(multi-signal)、多信息(multi-message)的响应系统,有很大的柔韧度和弹性<sup>[6]</sup>.这个系统通过4种普通的信号或信号载体来携带信息:(a) 静止的脸部符号表征相对永久的脸部特征,如构成个人形象的骨骼结构和软组织群;(b) 缓慢的脸部信号表征脸部表象的变化,如永久性皱纹的产生和皮肤纹理的变化;(c) 人工信号,如化妆等;(d) 快速的脸部信号,产生可见的能检测到的脸部表现的变化.

对于快速的脸部信号,Ekman和Friesen提出了人脸运动编码系统(facial action coding system,简称FACS)<sup>[7]</sup>,是人脸上所有导致脸部运动的运动单元(AU)的枚举.FACS有两个主要弱点:① 运动单元是纯粹的局部化的空间模板;② 没有时间描述信息,只是一个启发式信息.

FACS并不包含情绪信息,数据的分析只是单纯地描绘运动单元AU,或者把FACS通过字典规则转化成情绪.我们提出FACS'(即转换FACS),把运动单元的运动转化成基于物理和肌肉模型的运动特征流序列来对表情编码,相应的运动解释基于FACS的规则,同时克服了单纯FACS的弱点.

表情识别的一个难点就是建立表情模型.人脸是一个柔性体而不是一个刚体,很难把脸部的运动与表情联系起来.我们根据表情序列图像的特点建立了动态的表情模型,Onset→Application→Apex→Release→Offset.

## 2 运动场估计

### 2.1 表情特征区域的定位

人脸面部感知系统的已知条件是人脸图像.面部表情特征的定位也是在人脸检测的基础上的.本文采用了一个具有多级结构、由粗到精的人脸检测模型,测试图像通过各模块的逐级筛选,最后检测得到的图像可以确定为人脸.

人脸图像定位以后,根据脸部的结构信息,可以确定出各特征部件的大致区域:眼睛区域、鼻子区域、嘴部区域.它们是携带表情信息量最大的区域.虽然最终提取的特征不是基于精确的几何特征,不必精确描绘其形状

信息,但各特征区域必须分别归一、标准化.即先确定各个特征部件,根据其位置与比例,重新校正特征区域,同时根据 KL 变换调整主轴的思想去掉每组图像的头部刚体运动对表情运动的影响,然后对特征区域的大小进行归一,得到眼睛区域大小为  $20 \times 30$ ,嘴部区域大小为  $30 \times 40$  的图像.特征区域的定位只在第 1 帧中计算得到,其余帧与第 1 帧相同.

具体地,在眼睛区域中,最具明显特征的是虹膜,在边缘图像中,虹膜的边缘很强,并呈圆状.因此很容易检测到.进而可以得到眉毛特征,根据眼睛及眉毛的比例可以重新校正眼睛区域.嘴部区域也是先从大致确定的粗定位区域中根据灰度积分信息找到嘴部,然后重新校正嘴部特征区域.这样就保证了特征区域中特征部件的位置与比例的标准化.

## 2.2 光流计算

特征区域的运动场是基于序列图像中特征区域基础上计算的.用光流计算的方法提取其运动特征,一方面反映了表情的变化实质,另一方面特征受光照不均匀性的影响较小,只要保证在同一组图像序列中光照强度基本不变即可,光照条件对各组之间的影响较小.关于运动向量即光流的估计,主要有 3 种方法:基于时空梯度的方法(Horn 和 Schunck, Nagel, Uras 等,还有 Essa)、基于相关的方法(Anandan, Fan 等)以及频率域的方法(Heeger, Fleet 和 Jepson).此外,基于立体视的方法也逐渐受到重视.由传统的 Horn 和 Schunck 提出的基于梯度的方法比较适合于皮肤的变形计算,而且计算量比较简单,只是逐点估计位置的瞬时速度场.我们采用的就是这种方法<sup>[8]</sup>.

在视觉心理学中,当人与被观察物体发生相对运动时,被观察物体表面带光学特征部位的移动对观察者提供了运动以及结构的信息.这种带光学特征部位的移动投影到视网膜平面上形成了所谓的光流.运动产生出光流,因而有光流一定存在着运动.光流是一定灰度的点在图像平面上的运动速度,即

$$\frac{dl}{dt} = \left( \frac{dx}{dt}, \frac{dy}{dt} \right).$$

在基于梯度的方法中,时空梯度之间的关系是极其重要的,它构成了对光流计算的一个重要约束.设在时间  $t$  和  $t+dt$  时有两幅连续的投影,在图  $t$  上有灰度为  $f(x,y)$  的像素点,这里  $x,y$  为该点的坐标,该点在  $t+dt$  上移至  $f(x+dx,y+dy)$ ,  $f(x,y,t)=f(x+dx,y+dy,t+dt)$ .当变化是连续的,而且两帧间的时间间隔  $dt \rightarrow 0$  时,由泰勒级数展开可以得到:

$$\frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} + \frac{\partial f}{\partial t} = 0, u = \frac{dx}{dt}, v = \frac{dy}{dt}. \quad (1)$$

但在实际情况中,图像中的灰度变化以及速度场都可能不出现不连续性,即图像模式中的不连续性:一方面是表面模式的不连续性,另一方面是场景的可见表面中由一个面跳变到另外一个面的不连续性.光流的约束方程可以适用于以下情况:其一是图像中可以由图条形的不连续性存在,但在不连续性周围图像的变化应该是均匀的;其二是在图像平面中每个点上的灰度变化应该完全是由于图像中模式的运动而引起的,不应该加进某一点的反射性质变化带来的影响.也就是说,图像中的模式的位置发生变化,因而带来光流速度场,但该模式本身是没有变化的.

在求解光流时,Horn 和 Schunck 所采用方法的基本思想是光流本身尽可能平滑,即引入对光流的平滑性约束,设平滑性约束项为

$$E_s = \iint (u_x^2 + u_y^2 + v_x^2 + v_y^2) dx dy. \quad (2)$$

由基本等式(1),显然要求

$$E_c = \iint (f_x u + f_y v + f_t)^2 dx dy. \quad (3)$$

于是,由式(2)和式(3)可知,最后求得的光流应满足

$$\min \{ E_s = \iint [u_x^2 + u_y^2 + v_x^2 + v_y^2 + \lambda (f_x u + f_y v + f_t)^2] dx dy \}. \quad (4)$$

对形如式(5)的形式变分问题

$$\min \left\{ \iint F(u, v, u_x, u_y, v_x, v_y) dx dy \right\} \quad (5)$$

的解是对应的 Euler 方程

$$\begin{cases} F_u - \frac{\partial F_{u_x}}{\partial x} - \frac{\partial F_{u_y}}{\partial y} = 0 \\ F_v - \frac{\partial F_{v_x}}{\partial x} - \frac{\partial F_{v_y}}{\partial y} = 0 \end{cases}$$

的解.对于式(4),有

$$F = u_x^2 + u_y^2 + v_x^2 + v_y^2 + \lambda(f_x u + f_y v + f_t),$$

于是,对应的 Euler 方程为(其中 $\nabla^2$ 是 Laplace 算子)

$$\begin{cases} \nabla^2 u = \lambda f_x (f_x u + f_y v + f_t) \\ \nabla^2 v = \lambda f_y (f_x u + f_y v + f_t) \end{cases} \quad (6)$$

式(4)中的 $\lambda$ 反映对数据及约束的信度.当数据本身含有较多噪声时,原始数据的可信度较低,更多地依赖于对光滑性的约束, $\lambda$ 可以取较小的值,反之可以取较大的值.

在实际计算时,对象是离散化的图像,因此需要离散化处理.离散化后,光滑性约束变成了

$$s_{ij} = \frac{1}{4}[(u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2 + (v_{i+1,j} - v_{i,j})^2 + (v_{i,j+1} - v_{i,j})^2]. \quad (7)$$

而基本等式的约束变成了

$$c_{ij} = (f_x u_{ij} + f_y v_{ij} + f_t)^2. \quad (8)$$

于是,极小化目标函数为

$$\min \left\{ e = \sum_i \sum_j (s_{ij} + \lambda c_{ij}) \right\}. \quad (9)$$

对其求关于  $u_{kl}$  和  $v_{kl}$  的偏导,并令其为 0,有

$$\begin{cases} (1 + \lambda f_x^2) u_{kl} + \lambda f_x f_y v_{kl} = \bar{u}_{kl} - \lambda f_x f_t \\ \lambda f_y f_x u_{kl} + (1 + \lambda f_y^2) v_{kl} = \bar{v}_{kl} - \lambda f_y f_t \end{cases} \quad (10)$$

其中  $\bar{u}_{kl}$ ,  $\bar{v}_{kl}$  分别是  $u_{kl}$  和  $v_{kl}$  的四邻域平均.解得:

$$\begin{cases} u_{kl} = \bar{u}_{kl} - \frac{f_x \bar{u}_{kl} + f_y \bar{v}_{kl} + f_t}{1 + \lambda(f_x^2 + f_y^2)} f_x \\ v_{kl} = \bar{v}_{kl} - \frac{f_x \bar{u}_{kl} + f_y \bar{v}_{kl} + f_t}{1 + \lambda(f_x^2 + f_y^2)} f_y \end{cases} \quad (11)$$

于是得到自然的迭代过程

$$\begin{cases} u_{kl}^{n+1} = \bar{u}_{kl}^n - \frac{f_x \bar{u}_{kl}^n + f_y \bar{v}_{kl}^n + f_t}{1 + \lambda(f_x^2 + f_y^2)} f_x \\ v_{kl}^{n+1} = \bar{v}_{kl}^n - \frac{f_x \bar{u}_{kl}^n + f_y \bar{v}_{kl}^n + f_t}{1 + \lambda(f_x^2 + f_y^2)} f_y \end{cases} \quad (12)$$

的可求解光流.如图 1 所示是各种表情特征区域的光流场.

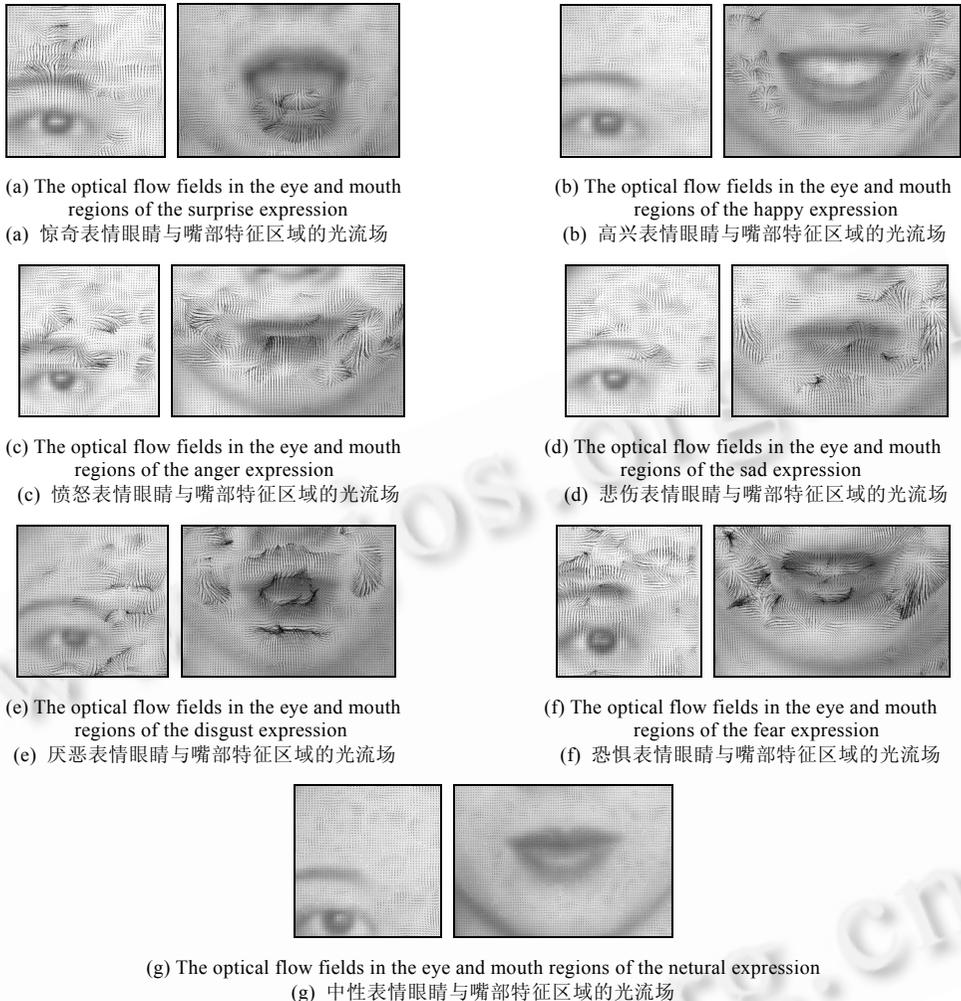


Fig.1 Velocity fields of the feature regions

图1 特征区域的速度场

### 3 特征序列的获取

对各个表情区域的运动场,我们根据 Karhunen-Loève 展开式<sup>[9]</sup>:

$$x(t) = \sum_{n=1}^{\infty} r_n x_n \phi_n(t), a \leq t \leq b,$$

采用较少数量的特征对样本进行描述,以降低特征空间维数,并且去掉了各特征分量间的相关性.用所有组图像序列的所有帧,共 1 341 帧(不包括中性表情的数据组,因为其他表情序列中都含有中性表情流数据)作为训练数据,求其协方差矩阵的特征值和特征向量,用前  $n$  个最大特征值对应的特征向量作基底,对于各特征区域的速度场的  $u$  分量和  $v$  分量分别求其基底.降维后,眼部区域的  $u$  分量和  $v$  分量分别降到 16 和 18 维,这样取其较大的 18 维作为两分量的维数.同理可得嘴部区域的两分量维数为 20 维.

在求特征序列时,把一组图像序列的每一帧中特征区域的运动场向其对应的基底作投影,把  $u$  分量和  $v$  分量串行合并作为一个特征流向量.然后把一组序列的特征流向量组成一个特征序列,如图 2 所示.其中 PCA 表示特征部件分析; $A, B, C$  分别表示各特征部件.

### 4 表情识别

#### 4.1 表情的时序分析方法

对一个实际的物理过程如动态的面部表情所产生的可观察序列,建立一个模型来描述这个序列的特征是非常重要的.在信号模型化问题上,隐马尔可夫模型既解决了用短时模型描述平稳段的信号,又解决了每个短时平稳段是如何转变到下一个短时平稳段的.即它既可以描述瞬态的随机过程,又可以描述动态的随机过程的转移特性.因而我们采用此模型<sup>[10]</sup>来描绘表情序列.HMM 在语音识别和手势识别中都得到了较成功的应用,同时,HMM 在分析时序过程中的良好特性也得到了证明.

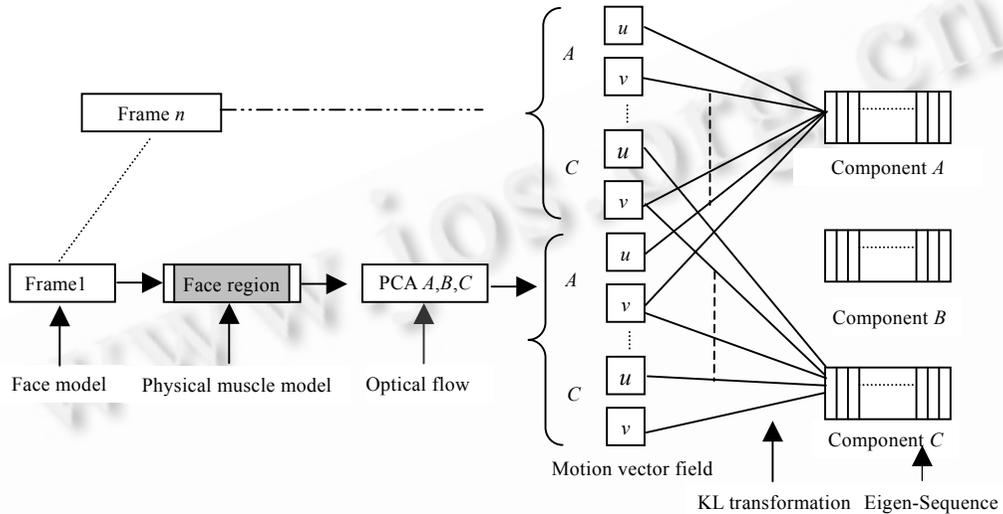


Fig.2 Acquisition of the eigen-sequences

图 2 特征序列的获取

人的面部表情图像序列与人的语音序列特点一样,是有时间顺序且不可逆的过程,因而采用无跨越从左向右的 HMM 模型.在此模型中限定一个起始状态,每个状态只能向右侧编号高一位的状态或本状态转移,大大减少了模型参数估值的计算量,参数的估计上采用的是半连续的,计算量较小,精度较高.为了得到可靠的模型估计,必须有多观察序列.本实验中每个 HMM 的训练数据大约有 20~30 组.在计算过程中,为防止溢出,在求解过程中加入比例因子,并且原始数据缩小到原来的 1/20.用 HMM 来识别表情序列,图像序列的长短不受限制.其中,惊奇表情的嘴部模型参数的实验结果如图 3 所示.

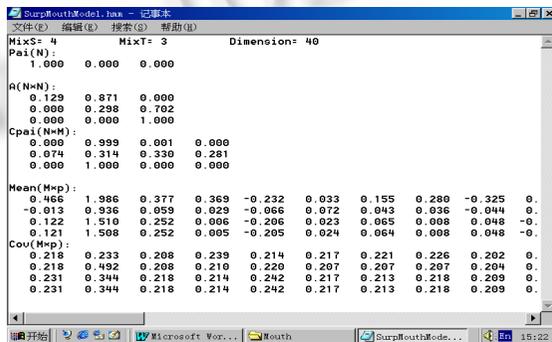


Fig.3 The trained results of the HMM parameters to the mouth region modal of the surprise expression

图 3 惊奇表情嘴部区域的 HMM 参数训练结果

## 4.2 实验

我们用心理学家 Ekman 所定义的最基本的 6 种表情加上中性表情,共 7 种最基本的表情做实验,即中性、高兴、惊奇、愤怒、悲伤、厌恶、恐惧.对每种表情采样 20~30 组图像序列,共 196 组,采样频率为每秒钟 7 帧.每组内表情图像的帧数不一样,大约有 10 多帧,大小为  $256 \times 256$ .在同一组序列中的第 1 帧,用人脸面部结构特征的先验信息和灰度积分特征确定表情特征区域,其余各帧与第 1 帧相同,并进行归一化和标准化;然后对每组序列的每帧图像进行时间-空间上的平滑,用基于梯度的光流方法提取特征部件区域的运动信息,得到水平方向  $u$  和垂直方向  $v$  的运动图像,再对运动场图像进行平滑,这样得到的光流数据噪声很小;在对运动特征进行降维时,把每组所有帧的运动场图像作为训练基底,分别对各特征部件的  $u$  和  $v$  方向进行降维,各部件的特征向量长度的确定是分别取  $u$  和  $v$  较大维数作为长度的一半,各部件的特征向量是  $u$  和  $v$  分量上分别投影后得到的系数串行连接起来的,把一组特征向量称为一个特征序列.眼睛部件向量长度为  $18 \times 2 = 36$ ,嘴部特征向量的长度为  $20 \times 2 = 40$ .系统的流程如图 4 所示.

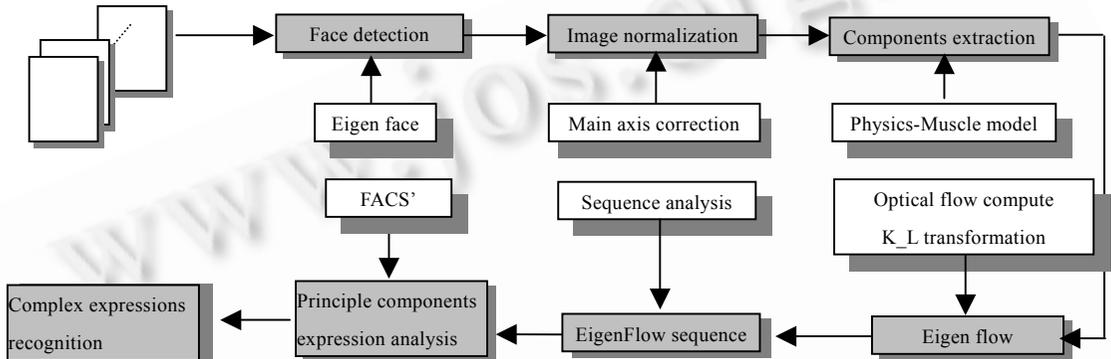


Fig.4 System flow

图 4 系统流程

在对 HMM 进行训练时,我们分 7 种表情的眼部加上嘴部共 14 种 HMM.把各种表情的眼部与嘴部分别训练、分析、识别,是因为在实际生活中人的面部表情并不是单一的某种表情,它与心情和情绪一样是混合复杂的,表现在面部表情上,不同的特征部件所包含的表情信息是不同的.对不同特征区域进行分析,进而分析混合表情的含义.HMM 的状态数与混合项数是由实验确定的,分别取  $N=3, M=4$ .各种表情的总体识别率达到了 96.9%,HMM 在分析时序过程中,速度和识别结果都非常理想.

本实验证明了用特征流作为表情的运动特征,稳定性高,可分性好,反映了各种表情变化的实质.

## 5 应用

人类的感知是一个多通道、多信息的综合决策过程.多功能感知机<sup>[11]</sup>是一个多输入流、多输出流的多模块系统.该系统包括视频、音频、文本等多个输入通道的并行输入,包括音频、视频和文本等多通道输出.各个模块之间相互协作,共享资源.如图 5 所示.各个感知通道并行工作,也就是说,各个感知通道是一个具有独立工作能力的 Agent.可以从 Agent 观点来看待多功能感知机系统.将完成不同功能的部件看作特定的 Agent,系统本身则是由许多合作 Agent 所组成,系统整体功能的实现依赖于各个 Agent 之间的合作.

显然,该模型接受 3 个通道的输入,即一个顺序的听觉通道、一个并行的阅读通道和一个并行的(除阅读外)视觉通道.这 3 个通道的数据经过相应的处理模块处理以后,所得到的信息送给融合模块加工成一个信息或命令序列.其中, S 表示声音信息处理, V 表示视觉信息处理, F 表示信息融合处理, Sd 表示声音驱动, Md 表示动作驱动, FE 表示面部表情识别模块.表情的识别作为一个独立的模块,在该系统中识别表情动作的含义和理解心理情绪,并且驱动表情的合成.自然化的人机交互只是其中的一个应用,随着面部表情的识别,系统功能的提高,以后会在其他领域得到更多的应用.

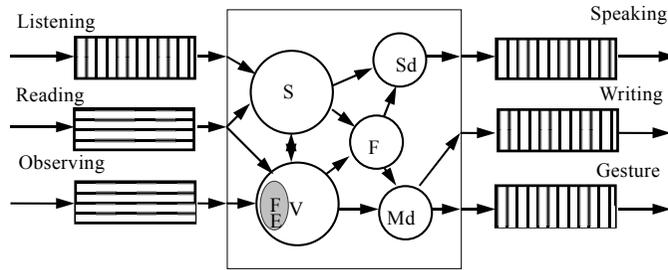


Fig.5 Expressions recognition in the multi-modal perception machine

图 5 表情识别在多功能感知机中

**References:**

- [1] Zhao LZ. Analysis and recognition of the facial expressions [MS. Thesis]. Harbin: Harbin Institute of Technology, 1997 (in Chinese with English abstract).
- [2] Lü Y. Research and realization of the facial expression recognition [MS. Thesis]. Harbin: Harbin Institute of Technology, 1999 (in Chinese with English abstract).
- [3] Mase K. Recognition of facial expressions for optical flow. IEICE Transactions on Special Issue on Computer Vision and Its Applications, 1991,E74(10):3474~3483.
- [4] Essa IA, Pentland AP. Coding, analysis, interpretation, and recognition of facial expressions. MIT Media Laboratory Perceptual Computing Section Technical Report, No.325, 1995. 19~20.
- [5] Lien JJ, Takeo Kanade. Automated facial expression recognition based on FACS action units. In: Proceedings of the 3rd International Conference on Automatic Face and Gesture Recognition. 1998. 390~395.
- [6] Ekman P, Huang TS, Report to NSF of the planning workshop on facial expression understanding. Technical Report, National Science Foundation, Human Interaction Lab., UCSF, 1993. 14~29.
- [7] Ekman P, Friesen WV. Facial Action Coding System. Palo Alto: Consulting Psychologists Press Inc., 1978.
- [8] Gao W, Chen XL. Computer Vision. Beijing: Tsinghua University Press, 1999 (in Chinese).
- [9] Bian ZQ. Pattern Recognition. Beijing: Tsinghua University Press, 1998 (in Chinese).
- [10] Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 1989,77(2):257~286.
- [11] Gao W. Research on multi-perception technology. Technical Report, Harbin: Harbin Institute of Technology, 1998 (in Chinese with English abstract).

**附中文参考文献:**

- [1] 赵力庄.面部表情的分析与识别[硕士学位论文].哈尔滨:哈尔滨工业大学,1997.
- [2] 吕岩.面部表情识别的研究与实现[硕士学位论文].哈尔滨:哈尔滨工业大学,1999.
- [8] 高文,陈熙霖.计算机视觉.北京:清华大学出版社,1999.
- [9] 边肇祺.模式识别.北京:清华大学出版社,1998.
- [11] 高文.多功能感知机技术研究.科技报告,哈尔滨:哈尔滨工业大学,1998.