

# 基于 D-S 证据理论的表情识别技术

王 嵘 马希荣

(天津师范大学计算机与信息工程学院 天津 300387)

**摘 要** 在情感计算理论基础上,提出了基于 D-S 理论的信息融合的表情识别技术,设计并实现了系统 IFFER。在表情识别模块中的分类器训练采用 JAFFE 表情库。识别中首先利用色度匹配及亮度匹配将人脸图像进行眼部及嘴部的分割,再分别用训练好的眼部 SVM 分类器及嘴部 SVM 分类器进行识别,将识别后的结果利用 D-S 证据理论进行融合。实验结果表明,对分割后的两部分图像进行识别,无论从训练上还是识别上,数据的维数都大大减少,提高了效率。在识别率上,融合后的结果相对于融合前的有显著的提高。

**关键词** 情感计算,信息融合,表情识别,D-S 理论

## Research on Facial Expression Recognition Technology Based on D-S Theory

WANG Rong MA Xi-rong

(College of Computer and Information Engineering, Tianjin Normal University, Tianjin 300387)

**Abstract** This paper proposed a kind of facial expression recognition technology based on the D-S theory of information fusion, in the background of affective computing theory, moreover designed and implemented the system IFFER. In the part of facial expression recognition, the training for the classification uses JAFFE expression library. Using color match and brightness match, this method divides face image into the eye and mouth images, then identifies facial expression respectively by trained eye SVM classifier and mouth SVM classifier, in the end, fuses the identification using D-S theory. The experimental results show that recognizing the two parts of image, in terms of training or recognition, the data dimension has greatly reduced and increased efficiency. In recognition rate, the fusion of the results has more improved than the pre-fusion.

**Keywords** Affective computing, Information fusion, Facial expression recognition, D-S theory

## 1 引言

随着科技的进步,人们越来越多地接触到计算机。在人机交互的过程中,人与机器间缺乏情感的交流成为当前急需解决的问题,于是便产生了情感计算理论。美国 MIT 大学媒体实验室的 Picard 教授于 1997 年出版的“*Affective Computing*(情感计算)”专著中给出如下定义:“情感计算是关于、产生于、或故意影响情感方面的计算”<sup>[1]</sup>。人们的情感可以通过多种方式表现出来,其中面部表情是最重要的途径。由此,本文是在情感计算的理论基础上,通过对面部表情的识别来对用户的情感进行识别,从而达到和谐的人机交互。

在人类的面部表情中,主要信息都集中在眼部和嘴部,其他部位的信息量相对较少。因此本文的表情识别技术将眼部与嘴部表情分别用两个支持向量机分类器进行识别,最后利用 D-S 证据理论的信息融合技术对识别结果进行融合。这样做不仅减少了训练和识别的维数,提高速度,而且最后将主要信息进行融合,增强了识别效率。同时,本文设计并实现了基于信息融合的表情识别系统 IFFER(Information Fusion-Facial Expression Recognition),在系统实现中,根据 Ekman 和 Friesen 的理论<sup>[2]</sup>将表情分为 6 种:高兴、悲伤、惊奇、恐惧、愤怒、厌恶。并用在日本女性表情库(JAFFE)<sup>[3]</sup>进行训练,取

得了良好的识别效果。

## 2 基于信息融合的表情识别技术

为了解决在表情识别中的识别速度问题,本文将信息融合技术应用到表情识别中去。将捕获到的人脸图像经过预处理后分割为脸部和嘴部两个子图像,再分别对这两个子图像进行特征提取、分类识别,最后将识别结果进行融合。表情识别模块的主要技术由下面这 3 部分构成:特征提取、支持向量机(SVM)基本理论、D-S 证据理论。

### 2.1 特征提取

有限 K-L 变换,又称霍特林(Hotelling)变换或主分量分解,它是一种基于目标统计特性的最佳正交变换。称其为最佳变换是因为它具有重要的优良性质:使变换后产生新的分量正交或不相关;以部分新的分量表示原向量均方误差最小;使变换向量更趋确定、能量更趋集中等。这使它在特征提取、数据压缩等方面都有着极为重要的应用。其主要原理如下:

设  $n$  维随机向量  $X = (x_1, x_2, \dots, x_n)^T$ , 其均值向量  $\bar{X} = E[X]$ , 相关阵  $R_X = E[XX^T]$ , 协方差  $C_X = E[(X - \bar{X})(X - \bar{X})^T]$ ,  $x$  经标准正交变换矩阵  $A^T$  正交变换后成为向量  $Y = (y_1, y_2, \dots, y_n)^T$ , 即

$$Y = A^T X \triangleq (A_1, A_2, \dots, A_n)^T X = \begin{bmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_n^T \end{bmatrix} X \quad (1)$$

从而  $Y$  的各分量

$$y_i = A_i^T X \quad i=1, 2, \dots, n \quad (2)$$

变换矩阵  $A^T$  在这里是由向量  $X$  的协方差矩阵的特征向量组成, 这里由于主要的信息都集中在特征值较大的特征向量上, 因此变换矩阵  $A^T$  由  $m$  个较大特征值对应的特征向量构造即可。

## 2.2 支持向量机(SVM)基本原理

支持向量机简称 SVM, 是统计学习理论中最年轻的内容, 也是最实用的部分<sup>[4,5]</sup>。SVM 是一种新的非常有发展前景的分类技术, 可以替代多层感知机、RBF 神经网络和多项式神经网络等已有的学习算法。其核心内容是在 1992 到 1995 年间提出的, 目前仍处在不断发展的阶段。

SVM 以结构风险最小化准则为理论基础, 通过适当的选择函数子集以及该函数子集中的判别函数, 使得学习机器的实际风险达到最小, 保证了通过有限训练样本的小误差分类器, 对独立测试集的测试误差仍然较小, 因而是一个具有最优分类能力和推广能力的学习机器。

SVM 是通过求解样本集的最优分类面来达到分类的目的。对于两类线性可分的情况, 如图 1 所示。最优分类面不但要求能将两类正确分开(训练错误率为 0), 而且使分类间隔最大。

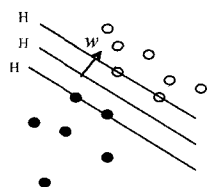


图1 最优超平面划分

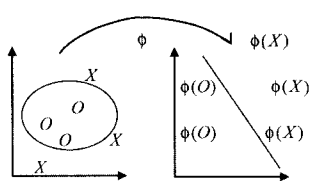


图2 非线性映射

对于线性不可分的情况, 支持向量机通过非线性变换  $\Phi$  把样本输入空间转化为某个高维空间中的线性问题, 在高维空间中求线性最优分类超平面, 如图 2 所示。变换  $\Phi$  可以利用核函数  $K$  来实现。因此, 两类问题的非线性 SVM 的求解过程为:

1. 设已知训练集  $T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (X \times Y)^l$ , 其中  $x_i \in X = R^n, y_i \in Y = \{1, -1\}, i=1, \dots, l$

2. 选择适当的核函数  $K(x, x')$ , 构造并求解最优化问题

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j K(x_i, x_j) - \sum_{j=1}^l \alpha_j \quad (3)$$

$$s. t. \sum_{i=1}^l y_i \alpha_i = 0 \quad (4)$$

$$\alpha_i \geq 0, i=1, \dots, l \quad (5)$$

得最优解  $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^T$

3. 选择  $\alpha^*$  的一个正分量  $\alpha_j^*$ , 并据此计算

$$b^* = y_i - \sum_{i=1}^l y_i \alpha_i^* K(x_j, x_i) \quad (6)$$

4. 构造决策函数

$$f(x) = \text{sgn}(\sum_{i=1}^l \alpha_i^* y_i K(x, x_i) + b^*) \quad (7)$$

利用 SVM 求解多类问题时, 可以将其转化成两类问题来解决。主要方法有一类对余类、成对分类等。一类对余类方法是构造一系列两类分类机, 其中的每一类分类机都把其中的一类同余下的各类分开, 然后据此推断某个输入  $x$  的

归属。这个算法的缺点是, 有可能  $x$  对应的多个类别的决策函数的值都很大, 以至于无法判断  $x$  属于哪一类。同时, 在一类对余类方法中, 两类问题是不对称的, 训练样本不平衡。另一种方法是成对分类法, 它也是基于两类问题的分类方法。它的主要思想是, 在所有类别中任意两类训练一个分类机, 那么对于第  $i$  类和第  $j$  类的分类机为,

$$f^{i-j}(x) = \begin{cases} i, & g^{i-j}(x) > 0 \\ j, & \text{其他} \end{cases} \quad (8)$$

其中  $g^{i-j}(x)$  为两类 SVM 的决策函数。对于一个输入  $x$ , 推断它属于第几类是考虑上述所有分类机对  $x$  所属类别的意见, 一个分类机判定  $x$  属于第  $i$  类就意味着第  $i$  类获得一票。得票数最多的类别就是最终判定  $x$  所属的类别。实际上对  $M$  类问题就有  $M(M-1)/2$  个两类分类机。这种方法的缺点是当类别很多时要训练多个两类分类机, 比较耗时。

## 2.3 D-S 证据理论

证据理论<sup>[6]</sup>是 Dempster 在 1967 年最先提出的, Shafer 进一步发展完善, 使 Dempster 合成法则推广到更加一般的情况, 为了纪念他们的重要贡献, 有人称证据理论为 D-S 理论。

D-S 理论用“识别框架  $\Theta$ ”表示所感兴趣的命题, 它定义了一个集函数  $m: 2^\Theta \rightarrow [0, 1]$ , 满足 1)  $m(\emptyset) = 0$ ; 2)  $\sum_{A \subset \Theta} m(A) = 1$  称  $m$  为识别框架  $\Theta$  上的基本可信度分配。  $\forall A \subset \Theta, m(A)$  成为  $A$  的基本可信数, 基本可信数反映了对  $A$  本身的信度大小。

Dempster 合成法则是对于多个信度的合成, 令  $m_1, \dots, m_n$  分别表示  $n$  个信息的信度分配, 如果它们是由独立的信息推得的, 则融合后的信度函数  $m = m_1 \oplus m_2 \oplus \dots \oplus m_n$  可表示为

$$m(A) = \frac{\sum_{\cap A_i = A} \prod m_i(A_i)}{1 - \sum_{\cap A_i = \emptyset} \prod m_i(A_i)} \quad (9)$$

通过合成法则就可以对不同的特征组合进行融合了。

## 3 IFFER 系统实现及应用

用户在交互过程中的情感是实时变化的, 为了提高人机交互过程中的表情识别速率, 同时又不影响识别率, 本文设计并实现的系统去掉了人脸图像中表情信息较少的部分, 只保留眼部和嘴部图像, 这样做可减小 SVM 分类器输入的维数, 提高速度, 最后将识别结果进行信息融合又不会降低识别率。

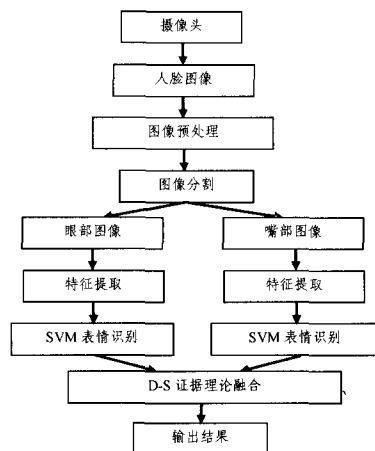


图3 表情识别流程

系统 IFFER 的主要流程, 如图 3。首先通过摄像头捕获到用户的图像, 进行图像的预处理, 光照补偿, 利用肤色模型

定位人脸,再利用色度匹配和亮度匹配定位出眼和嘴,最后将图像分割成眼和嘴两部分。对这两部分图像分别归一化,进行 K-L 变换特征提取。最后,输入到 SVM 中进行识别,将两个 SVM 的识别结果归一化,再进行 D-S 证据理论的融合,这样就可以根据面部表情得到用户的情感。

在本实验中的训练图像由两部分组成,一部分是在日本女性表情库 JAFFE,如图 4。JAFFE 表情图像来自 10 名日本女性的 7 种表情,其中同一人的每种表情有 3 张图像,共 210 张图像。本试验在 JAFFE 库中选取了 10 名日本女性的 6 种表情:高兴、悲伤、惊奇、恐惧、愤怒、厌恶,共 180 个图像。另一部分训练图像是从日常生活中提取出来的,共 7 个人的 126 个表情图像。这样训练图像一共为 306 幅。



图 4 表情库 JAFFE 样本



图 5 分割后的眼部及嘴部图像

将训练图像进行预处理,分别分割出眼部和嘴部,如图 5。将提取出的眼部和嘴部样本进行归一化,眼部图像按照 3:1 的比例进行处理,归一化为 21×7 的图像,嘴部图像按照 2:1 的比例进行处理,归一化为 14×7 的图像。在特征提取中,选取眼部图像和嘴部图像各 100 张图像计算 K-L 变换,对提取出的特征向量数再进行图像 K-L 逆变换,这样反复验证,确定眼部图像提取出 15 维特征,嘴部图像提取出 10 维特征。对眼部和嘴部图像的两个 SVM 分类器进行训练,本文利用 SVM 分类时采用成对分类法,这样每个分类器就产生 15 个两类的分类机,选取径向基核函数进行训练。根据成对分类法的原理,将图像识别后的各类票数进行归一化使其总合为 1。这样一幅待分类图像就会得到眼部和嘴部的两组分类数据,再根据 D-S 证据理论的合成法则进行融合。



图 6 待识别图像 P1,P2

图 6 是两张待识别的图像,编号为 P1,P2。E1 和 E2 分别为眼部和嘴部的两个特征子集,6 个状态,F0 为高兴,F1 为悲伤,F2 为惊奇,F3 为恐惧,F4 为愤怒,F5 为厌恶。经过两个 SVM 分类器分别对眼部和嘴部进行识别后得出票数归一化后数据如表 1,融合后的数据如表 2。

表 1 归一化后的分类数据

| 测试图像 | 特征子集 | m(F0) | m(F1) | m(F2) | m(F3) | m(F4) | m(F5) |
|------|------|-------|-------|-------|-------|-------|-------|
| P1   | E1   | 0.333 | 0.067 | 0.200 | 0.267 | 0     | 0.133 |
|      | E2   | 0.267 | 0.067 | 0.333 | 0.200 | 0.133 | 0.133 |
| P2   | E1   | 0.067 | 0.133 | 0.267 | 0.267 | 0.067 | 0.200 |
|      | E2   | 0     | 0.067 | 0.333 | 0.267 | 0.200 | 0.133 |

表 2 融合后的数据结果

| 测试图像 | m(F0) | m(F1) | m(F2) | m(F3) | m(F4) | m(F5) |
|------|-------|-------|-------|-------|-------|-------|
| P1   | 0.400 | 0.020 | 0.300 | 0.240 | 0     | 0.040 |
| P2   | 0     | 0.074 | 0.741 | 0.593 | 0.111 | 0.222 |

从数据中可以看出,在第一张图像的眼睛特征识别结果中,m(F0)值最大,而在嘴的特征识别结果中,m(F2)值最大。进行融合后 m(F0)的值最大,所以将其分类为高兴。在第二张图中,在眼睛的特征识别结果中 m(F2)和 m(F3)值一样,这说明眼睛同时具有惊奇和恐惧的特征。在嘴的特征识别结果中,m(F2)的值最大。融合后的结果 m(F2)的值最大,所以将其分类为惊奇。

表 3 是利用本文的方法对 6 种表情共 1000 个图像进行表情识别,得出了对 6 种表情的识别率。通过对比得出运用本文的方法大多数表情的识别率都有了提高。其中高兴和悲伤的识别率较高,这是因为在高兴和悲伤的面部表情中,眼部和嘴部的特征比较明显。而惊奇和恐惧的识别效果却不怎么明显,这是因为识别和恐惧表情的眼部及嘴部特征及其相似,所以容易造成错误识别。

表 3 表情识别率对比

| 表情种类 | 本文的方法  | 文献[9]中的 CKFD-神经网络方法 |
|------|--------|---------------------|
| 高兴   | 94.50% | 93.13%              |
| 悲伤   | 92.75% | 84.75%              |
| 惊奇   | 83.66% | 83.25%              |
| 恐惧   | 84.32% | 83.75%              |
| 愤怒   | 85.98% | 81.25%              |
| 厌恶   | 84.77% | 82.64%              |

**结束语** 本文将信息融合理论与表情识别技术相结合,提出了基于 D-S 理论融合技术的表情识别方法,设计并实现了系统 IFFER。此系统将待识别的表情图像进行分割,形成眼部与嘴部图像,同时训练两个分类器进行识别。这种方法不仅去掉了图像中情感信息较少的数据,提高了速率,而且最后的融合技术还提高了最终的识别率。

参考文献

[1] Picard R W. Affective Computing. London:MIT Press,1997

[2] Ekamn P, Friesen W V. Facial action coding system (FACS): manual[M]. Consulting Psychologists Press,1978

[3] Lyons M, Budynek J, Akamastu S. Automatic classification of signal facial images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1999,21(12):1357-1362

[4] Wang L P. Support Vector Machines: Theory and Application. Springer,Berlin Heidelberg New York,2005

[5] Platt J C. Fast training of support vector machines using sequential minimal optimization. Advance in Kernel method-support vector learning. Cambridge,MA,MIT Press,1999:185-208

[6] Shafer G. A mathematical theory of evidence. Princeton University Press,1976

[7] Yambor W S,Draper B A,Beveridge J R. Analyzing PCA-based Face Recognition Algorithms: Eigenvector Selection and Distance Measures // Second Workshop on Empirical Evaluation Methods in Computer Vision, 2000

[8] Zhang Y, Ji Q. Active dynamic information fusion for facial expression understanding from image sequences. IEEE Transactions on PAMI,2005,27:699-714

[9] 黄勇,应自炉. 基于双决策子空间和神经网络的人脸表情识别. 计算机工程与应用,2007,43(14):227-229

[10] 赵力庄,高文,等. Eigenface 的变维分类方法及其在表情识别中的应用. 计算机学报,1999,22(6)