

基于 C 均值 K 近邻算法的面部表情识别

张一鸣, 欧宗瑛, 王 虹

(大连理工大学 机械工程学院, 辽宁 大连 116023)

摘 要:随着人工智能与模式识别技术的不断发展,面部表情识别在智能人机交互中发挥着越来越重要的作用.通过对人的面部表情分类的研究,提出了一种使用 C 均值聚类、 K 近邻算法的面部表情分类方法.对参加训练的表情图像先进行 Gabor 小波变换,然后使用 Fisherface 判别分析方法进行变换,求得特征空间.再将已进行 Gabor 变换的标准表情图像投影到特征空间,进行 C 均值聚类得到子类表情模板.对于一幅待识别的表情图像,使用 K 近邻算法与子类表情模板比较,将该表情图像分类.使用该方法,在公开的日本女人表情人脸库上实测达到了 95.8% 的识别率.

关键词:面部表情识别; C 均值聚类; K 近邻; Gabor 小波; Fisherface 判别分析

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 1673-4785(2008)01-0057-05

Facial expression recognition based on C -means and K -nearest neighbor algorithms

ZHANG Yi-ming, OU Zong-ying, WANG Hong

(School of Mechanical Engineering, Dalian University of Technology, Dalian 116023, China)

Abstract: With the rapid development of artificial intelligence and pattern recognition, facial expression recognition plays an important role in intelligent human-machine interaction. In this paper, a facial expression classification method is presented which uses a C -means and K -nearest neighbor algorithm as the basis of analysis for the classification of facial expressions. First the images to be analyzed are transformed with Gabor wavelets, and then Fisherface discriminate analysis is performed to generate a feature space. Next, the images which were transformed with Gabor wavelets are projected into the feature space and C -means clustering performed on the projected images to generate sub-expression templates. Finally, the type of expression is identified by comparing the input expression images with the sub-expression templates by using a K -nearest neighbor algorithm. Experiments on the public Japanese female facial expression database show that the method proposed in this paper can achieve a 95.8% recognition rate.

Keywords: facial expression recognition; C -means clustering; K -nearest neighbor; Gabor wavelet; fisherface discriminant analysis

人的面部表情(高兴、厌恶、愤怒、惊讶等)是表达人的情绪的主要方法,是人的内心世界的外在表现形式.表情属于肢体语言,它和自然语言一起传达信息,因此面部表情的识别作为生物特征识别中的一种,是人机交互中不可缺少的一部分.

鉴于表情识别的重要性与复杂性,国内外很多

学者对表情识别的方法进行了研究.1978年 Ekman 和 Friesen^[1]开发了面部运动编码系统(facial action coding system),用以描述面部表情和人类的认知行为,被认为是表情识别与分析的里程碑.后来又出现了光流(optic flow)法^[2-3]、基于主成分分析法(principal component analysis)^[4-5,11]、局部加权的二维主成分分析算法(partially weighted two-dimensional, PCA)^[6]、线性判别分析(linear discriminant analysis)^[6]、Gabor 小波的方法^[7-9]、独立成份分析^[7,9]、局部特征分析^[10]、神经网络^[12-13]等多种方

收稿日期:2007-02-10.

基金项目:大连理工大学与中科院沈阳自动化研究所联合探索基金资助项目(DU T-SIA 2006).

通讯作者:欧宗瑛. E-mail: ouzyg@dlut.edu.cn.

法.综合考虑了以上算法,文中采用了 Gabor 小波变换和 Fisherface 判别分析的方法进行特征提取,采用了 C 均值聚类对标准的表情图像进行聚类,得到子类表情模板.对待识别的表情图像使用了 K 近邻判别方法进行了分类.该方法考虑到同一种表情可能有很多种不同的表现形式,比如高兴,有的人是张开嘴大笑,有的人是闭着嘴微笑,这就表明高兴这类表情还可以继续划分成多个子类,如果使用聚类的方法,就可以使每种表情聚成更多的子类,这样,所得到的表情模板才能更精确地代表该种表情的一种表现形式.使用 K 近邻算法将待识别的图像与各个聚类中心的表情模板进行比较,从而判别出该种表情与哪些表情的哪些表现形式更加接近,进而将该种表情进行分类.文中通过对公开的日本女人表情图像数据库 (JAFFE) 的测试表明,与其他方法相比,使用了 C 均值聚类与 K 近邻算法使识别率得到提高.

1 表情分类算法

文中所使用的表情分类方法分为以下几个步骤:首先对表情图像进行几何与灰度预处理.几何预处理是先找到图像中人眼的位置,经过旋转将两眼调整到同一水平位置.再将该图像按照一定的比例剪切下来,假设人两眼间的距离为 d ,剪切比例如图 1 所示.为排除人相貌对表情识别的影响,所以将人的额头与脸的外轮廓剪切掉.人表情的变化主要表现在眉毛、眼睛和嘴角的变化上,图示的剪切方法能很好地保留表情的变化的表现,同时削减了部分相貌差异对表情的影响.最后将剪切下来的图像放缩到统一大小,文中使用 100×128 像素大小的图像.对于灰度预处理,文中使用了直方图均衡化,使图像的对比度大大提高,灰度分布也趋于均匀.对图像进行预处理之后,使用 Gabor 小波提取图像的频率特征,然后使用 Fisherface 判别分析对图像进行训练,求得特征空间.文中使用模板匹配的方法分类表情,表情模板是通过在特征空间对一些标准的表情图像进行动态 C 均值聚类,将聚类中心作为表情模板而得到的.对于待识别的图像使用了 K 近邻算法,该算法与 C 均值聚类一起使用,达到了很好的识别效果.

1.1 表情图像的二维 Gabor 小波变换

文中使用 Gabor 小波来编码人脸图像.二维 Gabor 滤波器是带通滤波器,在空间域和频率域均有较好的分辨能力,它在空间域有良好的方向选择性,在频率域有良好的频率选择性. Gabor 小波变换是具有最好的时-频局部化描述能力的窗函数.二维 Gabor 小波变换是图像的多尺度表示和分析的有力工具,作为唯一能够取得空域和频域联合不确定关系下限的 Gabor 函数经常被用作小波基函数,

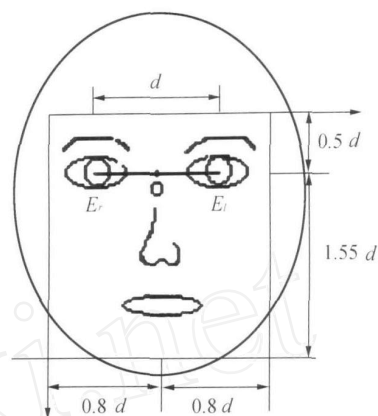


图 1 面部表情图像的剪切比例示意图

Fig. 1 Schematic of cut proportion on expression image

对图像进行各种分析.二维 Gabor 小波变换描述了图像 $I(x)$ 上给定一点 $x = (x, y)$ 附近区域的灰度特征,这可以用一个卷积来定义^[13]:

$$J_j(x) = I(x) \otimes j(x - x) d^2 x. \quad (1)$$

Gabor 核函数的定义为

$$j_j(x) = \frac{k_j^2}{2} \exp\left(-\frac{k_j^2}{2} \frac{x^2}{2}\right) \cdot \left[\exp(ik_j x) - \exp\left(-\frac{2}{2}\right)\right]. \quad (2)$$

式中:

$$k_j = \begin{bmatrix} k_{jx} \\ k_{jy} \end{bmatrix} = \begin{bmatrix} k_v \cos \mu \\ k_v \sin \mu \end{bmatrix}, k_v = 2 \cdot \frac{\mu^2}{2}, \mu = \mu \frac{\pi}{8}, \mu = 2. \quad (3)$$

文中使用了 5 个频率,8 个方向的 Gabor 核函数,即 $v = 0, 1, \dots, 4, \mu = 1, 2, \dots, 8$.对图像中的一个点进行 Gabor 滤波,便得到 40 个 Gabor 系数.文中对预处理后的图像使用网格结构均匀取点,每隔 20 个像素点取一个点(共 30 个点)进行 Gabor 滤波,得到长度为 1 200 Gabor 系数的向量 J .即经过该滤波之后,将大小为 128×100 的灰度图像变换为 1 200 维 Gabor 向量.

1.2 Fisherface 判别分析

灰度图像经过二维的 Gabor 小波变换得到了 Gabor 向量,接下来使用主成分分析 (PCA) 与线性判别分析 (LDA) 相结合的方法,即 Fisherface 法对 Gabor 向量进行训练,得到投影空间.该方法先将 Gabor 向量通过 PCA 提取主成分信息并降维,然后对降维后的图像应用 LDA 训练一个最佳分类器. Fisherface 方法既保留了原始向量的主成分信息,又克服了直接使用 LDA 带来的大矩阵和类内散布矩阵奇异的问题,且该方法的识别率比只使用 PCA

的方法高.

对于已经过 Gabor 小波变换的表情图像训练集 X , 先将图像集进行分类, 将同一种表情的图像归为一类, 记为 X_i . 文中要识别 7 种表情, 所以 $i = 1, 2, \dots, 7$. Fisherface 的计算过程如下.

首先, 计算每一类的类内平均表情图像 m_i 和所有训练图像的总体平均表情图像 m . 然后, 将每个类内的表情图像减去自己类内的平均表情图像, 得到每个表情的差值表情; 最后每一个类内平均表情图像减去总体平均表情图像.

$$\begin{aligned} \forall x \in X_i, X_i \in X, \\ \hat{x} = x - m_i, \hat{m}_i = m_i - m. \end{aligned} \quad (4)$$

再创建一个数据矩阵, 即把所有中心化后的表情图像按次序排列成一个表情数据矩阵. 接下来求解表情数据矩阵的正交向量: 可以通过奇异值分解或者求解表情数据矩阵的协方差矩阵的方法得到正交基, 也就是采用 PCA 方法得到一个表情子空间, 记为 W_{PCA} . 投影所有的中心化后的表情图像、中心化后的类内平均表情图像及总体表情平均图像到表情子空间.

$$\begin{aligned} x = W_{PCA}^T \hat{x}, m_i = W_{PCA}^T \hat{m}_i, \\ m = W_{PCA}^T m. \end{aligned} \quad (5)$$

计算第 i 类的类内散布矩阵 S_i 和总的类内散布矩阵 S_w . C 为表情分类的类数, 文中 $C = 7$.

$$S_i = \sum_{x \in X_i} (x - m_i)(x - m_i)^T, S_w = \sum_{i=1}^C S_i. \quad (6)$$

类间散布矩阵 S_B 是所有类内表情平均图像在表情子空间的投影的加权协方差矩阵之和, 权值 N_i 是该类表情的图像数.

$$S_B = \sum_{i=1}^C N_i (m_i - m)(m_i - m)^T. \quad (7)$$

求解类内散布矩阵 S_w 和类间散布矩阵 S_B 的广义特征值 V 和特征向量:

$$S_B V = S_w V. \quad (8)$$

按照特征值的大小对特征向量从大到小排序, 保留最大的前 $C - 1$ 个特征向量, 组合成最佳分类空间 W_{FLD} . 最后, 组合 PCA/LDA 方法, 得到最优的表情投影子空间 $W_{OPT} = W_{FLD} W_{PCA}^T$. 对于后文表情模板的计算和待识别的表情图像, 都要在这个空间进行计算, 即要投影到这个空间. 对于任意一个向量 x , 在该空间的投影变换公式为

$$z = W_{OPT}^T x = W_{FLD}^T W_{PCA}^T (x - m). \quad (9)$$

式中: z 是一个 $C - 1$ 维的向量, 文中 $C = 7$, 所以是 6 维的向量. 可见经过该变换后数据量大大减小, 这样才能做到表情的实时识别.

1.3 C 均值聚类

因为表情因人而异, 所以同一种表情往往会有

很多种表现方式, 因此很难只使用一个表情模板来代表一种表情, 也就是说每种表情还应再划分成多个子类. 聚类是在样本类别未知的情况下进行的无监督的学习方法. 通过聚类分析, 可以把相似的样本聚成一类, 不相似的样本分别聚在不同的类里. 文中对表情模板使用了动态 C 均值聚类方法, 将每种表情聚成多个子表情类, 取每类的质心作为标准的表情模板. C 均值聚类算法过程如下:

首先, 确定聚类的类数 C 并选择 C 个代表点作为最初的质心. 然后计算代表点之外的所有点到各个代表点的距离, 按照最近邻的原则, 将这些点进行分类. 至此完成第 1 次迭代. 接下来再分别计算各个类的质心, 然后计算所有点到各个质心的距离, 按照最近邻法将所有点再一次进行分类. 反复迭代, 直到算法收敛或达到预定义的最大迭代次数为止. 该过程如图 2 所示.

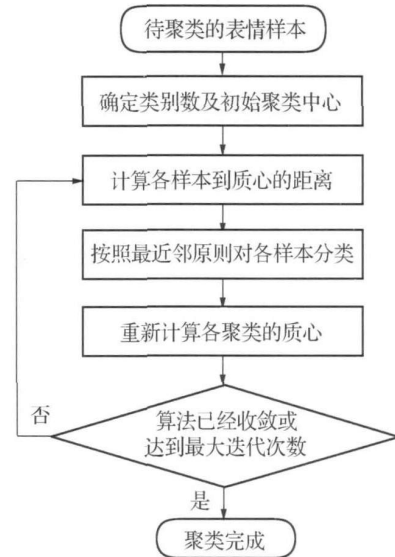


图 2 C 均值聚类算法过程图

Fig. 2 Process of C-means clustering algorithm

1.4 K 近邻分类

该方法简单地说就是取未知样本 x 的 k 个近邻, 看这 k 个近邻中多数属于哪一类, 就把 x 归为哪一类^[15]. 具体说就是在 N 个已知样本中, 找出 x 的 k 个近邻. 设这 N 个样本中, 来自 1 类的样本有 N_1 个, 来自 2 类的样本有 N_2 个, ..., 来自 c 类的样本有 N_c 个, 若 k_1, k_2, \dots, k_c 分别是 k 个近邻中属于 1, 2, ..., c 类的样本数, 则定义判别函数为

$$g_i(x) = k_i, i = 1, 2, \dots, C. \quad (10)$$

决策规则为

$$\begin{aligned} \text{若 } g_j(x) = \max_i k_i, \\ \text{则决策 } x \text{ 为 } j. \end{aligned} \quad (11)$$

这就是 K 近邻法的基本原则. 对一幅待识别的表情图像, 文中先将该图像进行 Gabor 小波变换,

再利用式(9)投影到特征空间上,最后使用 K 近邻法与 1.3 中使用 C 均值聚类得到的表情模板进行比较,以达到对该表情分类的目的. K 近邻法与 C 均值聚类方法联合使用,考虑了同种表情的表现多样性,能够有效地提高表情的识别率.

1.5 表情识别的整体流程

表情识别的全部过程见图 3. 所有的图像首先都要进行 Gabor 小波变换,然后对训练图像进行训练得到特征空间,标准的表情图像在该特征空间中进行 C 均值聚类得到表情模板,待识别的表情图像在该特征空间中与表情模板匹配从而实现该幅表情图像的分类.

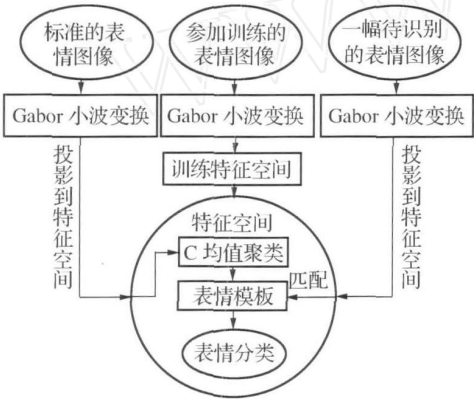


图 3 表情识别的全部过程

Fig. 3 Process of facial expression recognition

2 测试实验

测试实验使用了日本女人表情数据库 (Japanese female facial expression JAFFE). 该数据库包含了 10 个日本女人的 213 幅表情图像,每个人 7 种表情,包括高兴、生气、恐惧、厌恶、害怕、悲伤和中性表情,每个人的同一种表情图像有 3~4 幅. JAFFE 的部分表情图像如图 4 所示,其表情依次为中性、生气、厌恶、恐惧、高兴、悲伤、惊讶.

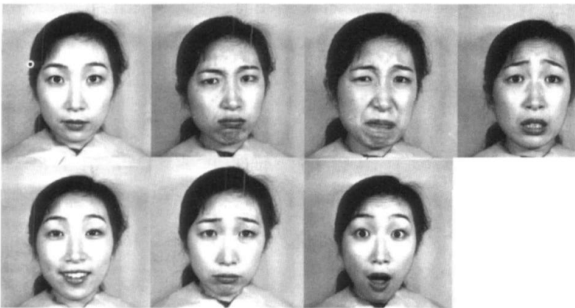


图 4 JAFFE 表情图像数据库中的表情样本

Fig. 4 Expression samples in JAFFE expression database

本测试实验与 Michael 等^[16]的测试方法相同,把数据库中的 213 张图像分成 10 部分,每次使用其中的 9 部分训练,剩下的 1 部分测试. 循环测试 10 次,然后对这 10 次取平均值,得出正确识别率. 由于 C 均值算法的 C 取值和 K 近邻的 K 取值不同,识别率有所不同,实验结果见表 1.

表 1 C 均值 K 近邻算法的不同取值的实验结果

Table 1 Performance of C-means cluster and Knearest neighbor algorithm

C	K					
	1	2	3	4	5	6
1	91.0	91.0	91.0	91.0	91.0	91.0
2	91.3	91.3	91.7	91.9	91.9	91.9
3	92.8	92.8	93.2	93.1	93.1	93.0
4	95.2	95.2	95.5	95.8	95.3	95.3
5	94.2	94.2	94.3	94.4	94.4	94.3
6	93.1	93.1	93.3	93.5	93.5	93.4

由表 1 的实验数据可以看出,采用把每个表情类分成多个子表情类的方法,由于充分利用了每种表情类之间的差异信息,因此能够提高识别率. K 的取值对识别率影响不大,识别率主要取决于 C 的取值,即每种表情聚成的子类个数. 随着 C 的取值从 1 逐渐增大,识别率也随着提高,但是当 C 的取值超过某一个数值之后识别率又开始下降. 这是因为当聚类的数目很多时,每类的图像数目变少,聚类所形成的表情模板受到了人相貌的影响,影响了表情识别的准确率. 从表 1 看出当 C=4, K=4 时,达到最高识别率 95.8%. 该识别率高于文献[16]92% 的识别率. 此时输入的表情与被识别出的表情的混淆矩阵(confusion matrix)见表 2.

表 2 C=4, K=4 时的混淆矩阵

Table 2 Confusion matrix when C=4 and K=4

O	I						
	高兴	厌恶	生气	恐惧	惊讶	悲伤	中性
高兴	31	0	0	0	0	0	0
厌恶	0	29	1	0	0	1	0
生气	0	0	28	0	0	1	0
恐惧	0	0	0	30	2	0	0
惊讶	0	0	0	2	28	0	0
悲伤	0	0	1	0	0	29	1
中性	0	0	0	0	0	0	29

I: 识别输入 (recognition input); O: 识别输出 (recognition output)

从表 2 可以看出,惊讶和恐惧不易区分,生气、厌恶、悲伤 3 种表情也不易区分,高兴的识别率最高.这是由于不同表情之间的差异不同造成的.

3 结束语

文中通过 Gabor 小波变换提取面部表情信息,使用 Fisherface 判别分析的方法训练投影空间,对于表情模板使用了 C 均值动态聚类的方法,将每种表情聚成多个子表情类,最后使用 K 近邻法进行表情识别.从实验结果来看,文中的方法由于考虑了每种表情表现形式的多样性,将每种表情划分成多个子类,识别率得到一定的提高.由于表情的多样性与复杂性,表情识别的研究,仍然是模式识别领域的一项重要重要的课题.

参考文献:

- [1] EKMAN P, FRIESEN W V. Facial action coding system: a technique for the measurement of facial movement [M]. Palo Alto, CA: Consulting Psychologists Press, 1978.
- [2] MASE K. Recognition of facial expression from optical flow [J]. IEICE Trans E, 1991, 74(10): 3474-3483.
- [3] YACOOB Y, DAVIS L. Recognizing human facial expressions from long image sequences using optical flow [J]. IEEE Trans on PAMI, 1996, 18(6): 636-642.
- [4] COTTRELL G, METCALFE J. Face, gender and emotion recognition using Holons [C]// Advances in Neural Information Processing Systems. Denver, USA, 1990, 3: 564-571.
- [5] PADGETT C, COTTRELL G. Representing face images for emotion classification [C]// Advances in Neural Information Processing Systems. Cambridge: MIT Press, 1997.
- [6] 金一, 阮秋琦. 一种局部加权的二维主成分分析算法及其在人脸识别中的应用 [J]. 智能系统学报, 2007, 2(3): 25-29.
JIN Yi, RUAN Qiuqi. A part weighted two-dimensional PCA for face recognition [J]. CAAI Transactions on Intelligent Systems, 2007, 2(3): 25-29.
- [7] DONATO G, STEWART M B, HAGER J C, et al. Classifying facial actions [J]. IEEE Trans on PAMI, 1999, 21(10): 974-989.
- [8] DAUGMAN J G. Complete discrete 2D Gabor transform by neural networks for image analysis and compression [J]. IEEE Trans on ASSP, 1998, 36(7): 1169-1179.
- [9] BUCIU I, KOTROPOULOS C. ICA and Gabor representation for facial expression recognition [C]// Proceedings of IEEE ICIP. Barcelona, Spain, 2003.

- [10] PENEV P S, ATICK J J. Local feature analysis: a general statistical theory for object representation [J]. Network: Computation in Neural Systems, 1996, 7(3): 477-500.
- [11] CALDER A J, BURTON A M, MILLER P, et al. A principal component analysis of facial expressions [J]. Vision Research, 2001, 41(9): 1179-1208.
- [12] SEYEDARABI H, AGHA GOLZADEH A, KHAN-MOHAMMADI S. Recognition of six basic facial expressions by feature-points tracking using RBF neural network and fuzzy inference system [C]// Proceedings of 2004 IEEE International Conference on Multimedia and Expo. Taipei, China, 2004.
- [13] MA L, KHORASANI K. Facial expression recognition using constructive feedforward neural networks [J]. IEEE Trans on SMC-Part B, 2004, 34(3): 1588-1595.
- [14] LADES M, JAN C. Distortion invariant object recognition in the dynamic link architecture [J]. IEEE Trans on Computer, 1993, 42(3): 300-311.
- [15] 边肇祺, 张学工. 模式识别 [M]. 北京: 清华大学出版社, 2000.
- [16] LYONS M J, BUDYNEK J, KAMATSU S. Automatic classification of single facial images [J]. IEEE Trans on PAMI, 1999, 21(12): 1357-1362.

作者简介:



张一鸣,女,1981年生,硕士研究生,主要研究方向为模式识别.



欧宗瑛,男,1936年生,教授,博士生导师,主要研究方向为计算机辅助设计、计算机图形学和图像处理.获教委和机械部科技三等奖各一项,辽宁省科技一等奖一项.参与主编的机械设计手册和机电类规划教材分别获1995全国科技图书二等奖和2002全国优秀教材二等奖.发表的学术论文被SCI检索12篇,被EI检索95篇.



王虹,男,1964年生,博士研究生,主要研究方向为模式识别、图像处理.