

三角坐标系下人脸表情表示方法

肖 潇 孔凡芝

浙江传媒学院电子信息学院 杭州 310018

摘 要 对三角形坐标系作了推广,给出了广义三角坐标,使之使用于人脸表情特征表示,结合高斯核 SVM 分类器,采用留一主体交叉验证技术。针对 CK+人脸表情数据库,得到了人脸表情正确识别率为 98.2%,相比于其基准算法和 M-CRT 算法,正确率有较大提高。这表明所提出的人脸表情特征表示方法的有效性。

关键词 人脸表情;特征表示;三角坐标系;位移矢量

中图法分类号 TP391.41

New Representation of Facial Affect Based on Triangular Coordinate System

XIAO Xiao and KONG Fan-zhi

College of Electronics Information, Zhejiang University of Media and Communication, Hangzhou 310018, China

Abstract Based on the triangular coordinate system, the generalized triangular coordinates are given to be used in facial expression feature representation. Combined with the Gaussian kernel SVM classifier, the left-face cross-validation technique is used to obtain the correct facial expression. For the CK+ facial expression database, the recognition rate is 98.2%, which is greatly improved compared with benchmark algorithm and M-CRT algorithm, indicating the effectiveness of the proposed facial expression feature representation method.

Keywords Facial Expression, Feature representation, Triangle coordinate system, Transform vector

1 引言

人脸表情是人们情感交流的重要方式,人脸表情的自动识别对诸多领域,如人机交互、实时表情视频生成^[1]、动画等都有重要应用。如何快速准确地判别人脸表情,一直吸引众多研究者,依然是人们关注的研究热点^[2-4]。

如何提取与光照、姿势、身份无关的人脸表情特征,无疑是人脸表情识别中的关键。基于局部统计特征的 LBP^[5]已应用于人脸表情识别。Shan 等^[6]基于 Boosted-LBP 提取最有判别性的 LBP 特征,利用 SVM 得到最佳的识别性能,并将 LBP 用于低分辨率人脸图像的表情识别,得到可靠稳定的识别性能。Du 等^[7]给出了改进的分类及回归树(M-CRT)用于人脸表情识别,通过计算表情图像与中性表情图像之差,并用 LBP 等方法得到全局和局部特征。

通过测量和分析脸部肌肉运动,Ekman 等提出了脸部活动编码系统(FACS)^[8],以及相应的人脸活动单元(AU),通过多个 AU 的组合判别人脸表情。Eleftheriadis 等^[9]提出了新颖的高斯过程自编码建模方案,将多个观测特征编码投影到隐空间,而解码器负责重构原始特征,对人脸活动单元的预测有较好结果。文献^[10]基于部分的、稀疏表示测量 AU 强度的连续变化,检测离散的人脸活动单元强度,用于自动人脸表情分析。文献^[11]研究不同表情中的共同及特定信息,而呈现表情时仅有少数人脸部分活跃,从而可揭示对表情识别具

有重要影响的特定部分,并采用两阶段多任务稀疏学习框架,同时进行人脸表情识别和人脸识别。

利用人脸表情随时间变化的动态信息,Long 等^[12]对视频利用独立元分析(ICA)学习空间-时间滤波器,基于所学习的滤波器对输入视频构造特征表示,并基于无标记数据的特征学习识别人脸表情。Zhi 等^[13]利用图保持稀疏非负矩阵分解算法研究人脸表情识别,在有遮挡情形下具有更好的鲁棒性。

随着深度学习在图像分类等应用中的突出表现^[14],研究者提出了将深度学习应用于人脸表情识别的算法^[3-4]。Hasani 等^[15]结合深度神经网络(DNN)和条件随机场(CRF)对视频人脸表情建模,利用卷积层描述人脸图像中的空间关系,再用条件随机场刻画视频帧间的时间关系,对 3 个公开数据(CK+, MMI, 和 FERA)库做了比较分析。Rifai 等^[16]提出了半监督方法解决数据中所呈现的各种变化因素,利用深度学习所提供的层次式学习方法,将表情相关因素与其他无关特征分离。Li 等^[17]提出了深度卷积融合网络,给出了 2D+3D 的人脸表情识别方案。Usman 等^[18]利用深度自编码器对人脸表情识别做特征选择和维数归约,取得了较好的结果。

综上所述,研究者在人脸表情识别中取得了令人瞩目的成就,但依然存在着诸多不足,如特征表示维数过高,有些特征的训练提取过程时间过长,影响计算效率和实时性;也有些方法要求有三维或视频数据,这在某些条件下并不能得到满足,如对于简易摄像头获取的图像,难以精确得到三维图像,

基金项目:浙江省公益项目(LGG19E050002, LGG18F010001);院级教改项目《视频监控技术》课程讲义建设

This work was supported by Public Welfare Project of Zhejiang Province (LGG19E050002, LGG18F010001) and College Educational Reform Project 'Video Monitoring Technology' Course Construction.

通信作者:肖潇(xiaoxiao800412@163.com)

而视频存储过大,也会影响实际使用。随着机器视觉研究的深入,人脸特征点检测技术在精确性和效率两方面都有了显著的提高^[19],为此我们仅基于人脸的中性表情图像和峰值表情图像,利用所检测的人脸特征点,基于三角形坐标系^[20],给出了人脸表情位移矢量的概念,基于 SVM 分类算法^[21]对典型人脸表情数据进行了测试,结果表明了所提方法的有效性。

2 三角坐标系下人脸表情表示

文献[20]提出了应用于图像变形处理的三角形坐标系,并将其基本思想推广到三维变形,给出了四面体坐标系^[22],得到诸多有意义的结果。本文在此基础上为适用于人脸表情识别,将文献[20]中仅考虑三角形内及边上点的三角形坐标表示推广到整个平面上点,并给出基于齐次坐标的简洁表示。

2.1 广义三角形坐标

为简洁起见,用大写字母表示平面点的笛卡尔坐标,而用小写字母表示其对应的三角形坐标,文献[20]的三角形坐标系的定义如下。

如有一点 P 位于 $\triangle ABC$ 的内部或者边上,已知 A, B 及 C 各自的笛卡尔坐标为 $(A_x, A_y), (B_x, B_y)$ 和 (C_x, C_y) , P 点为 (P_x, P_y) , 则 P 点的三角坐标 (p_1, p_2) 定义如下: $p_1 = p_2 = 0$, 若 P 点与 C 点重合; 否则 $p_1 = k_3 / (1 - k_1), p_2 = (1 - k_3) / (1 - k_1)$ 。其中:

$$\begin{cases} k_2 = (A_x - B_x)(C_y - P_y) - (A_y - B_y)(C_x - P_x) \\ k_1 = [(A_x - B_x)(B_y - P_y) - (A_y - B_y)(B_x - P_x)] / k_2 \\ k_3 = [(P_x - B_x)(C_y - P_y) - (P_y - B_y)(C_x - P_x)] / k_2 \end{cases} \quad (1)$$

由相应的逆变换可知,若已知点 P 在 $\triangle ABC$ 中的三角形坐标为 (p_1, p_2) , 那么点 P 的笛卡尔坐标计算如下:

$$\begin{cases} P_x = (1 - p_1 - p_2)C_x + p_1A_x + p_2B_x \\ P_y = (1 - p_1 - p_2)C_y + p_1A_y + p_2B_y \end{cases} \quad (2)$$

其几何意义参看文献[20]。不同于文献[20]所描述的图像变形,在人脸表情的表示中,需要考虑在三角形内外点的三角坐标表示,如图 1 所示。



图 1 三角形内外点

Fig. 1 Interior and exterior points of a triangle

但当点 P 位于三角形外部,且 $CP \parallel AB$ 时,如图 2 所示,式(1)中 $k_2 = 0$,使得三角坐标没有定义,为此以齐次坐标形式将原始三角坐标进行推广,得到在整个平面上点的笛卡尔坐标与三角坐标的变换形式。

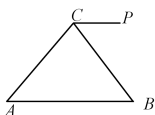


图 2 CP 平行于 AB

Fig. 2 CP is parallel to AB

定义 1(广义三角坐标) 记平面齐次笛卡尔坐标 $\mathbf{X}_P = [P_x, P_y, 1]^T$, 齐次三角形坐标 $\mathbf{W}_P = [p_1, p_2, 1 - p_1 - p_2]^T$, $\triangle ABC$ 所组成的齐次笛卡尔坐标矩阵 $\mathbf{O} = \begin{bmatrix} A_x & B_x & C_x \\ A_y & B_y & C_y \\ 1 & 1 & 1 \end{bmatrix}$,

则有变换关系:

$$\mathbf{X}_P = \mathbf{O} \times \mathbf{W}_P = [\mathbf{X}_A \ \mathbf{X}_B \ \mathbf{X}_C] \times \mathbf{W}_P \quad (3)$$

且有逆变换:

$$\mathbf{W}_P = \mathbf{O}^{-1} \times \mathbf{X}_P \quad (4)$$

可简单验证,广义三角坐标与文献[20]一致,且由于 \mathbf{O}^{-1} 的存在性,使得任意平面点的三角坐标都有定义。

2.2 广义三角坐标系下的人脸表示

人脸表情与人脸活动单元(AU)密切相关,但将精确测量 AU 的强度以及利用 AU 的组合用于人脸表情识别仍是个困难问题^[8]。为此,利用人脸特征点的位置变化表示不同表情,不失为一种有效的表示方法,随着人脸特征点检测技术的成熟,可较为精确且高效地获取人脸特征点^[19]。但使用一般的人脸特征点变化表示人脸表情都需要人脸对齐步骤^[2],而特征点相对于三角形的广义三角坐标具有较好的旋转、缩放、平移不变性^[20],从而可避免人脸对齐过程。

记人脸特征点序号为 $i \in \{1, \dots, t\}$, t 为特征点数,设中性表情人脸图像所提取的特征点的笛卡尔坐标为 $\{V_i | V_i = (V_{ix}, V_{iy}), i \in \{1, \dots, t\}\}$, 选择其中 3 个特征点,记为 $\{i_1, i_2, i_3\}$, 可组成 $\triangle i_1 i_2 i_3$, 由式(4)可计算相对于 $\triangle i_1 i_2 i_3$ 的三角坐标 $\{W_i | W_i = (p_{i1}, p_{i2}), i \in \{1, \dots, t\}\}$; 类似地对于表情图像所提取的特征点的笛卡尔坐标为 $\{V'_i | V'_i = (V'_{ix}, V'_{iy}), i \in \{1, \dots, t\}\}$, 选择与 $\{i_1, i_2, i_3\}$ 相同序号的特征点,组成 $\triangle i'_1 i'_2 i'_3$, 同样可由式(4)计算相对于 $\triangle i'_1 i'_2 i'_3$ 的三角坐标 $\{W'_i | W'_i = (p'_{i1}, p'_{i2}), i \in \{1, \dots, t\}\}$, 每个特征点的位移记为 $\Delta W_i = (p'_{i1} - p_{i1}, p'_{i2} - p_{i2})$, 串接所有特征点位移,记为表情位移矢量 $\Delta W = [\Delta W_1, \dots, \Delta W_t]^T$, 作为该种表情的矢量表示。这里需要注意的是,由于 $\triangle i_1 i_2 i_3$ 与 $\triangle i'_1 i'_2 i'_3$ 的对应顶点的三角坐标一致,没有判别作用。采用文献[19]的特征点检测方法,得到 68 个特征点,去除 3 个基准点(组成三角形),最后所组成的表情位移矢量为 130 维;另外考虑到计算误差,三角形的 3 个顶点尽量选得远些,如图 3 所示。



(a) 中性表情图像及特征点 (b) 表情图像及特征点 (c) 表情图像及位移矢量

图 3 特征点及表情位移矢量

Fig. 3 Feature points and expression displacement vectors

3 人脸表情识别实验结果

3.1 数据库

我们采用著名的 CK+ 数据库^[23], 其中年龄范围从 18 岁到 50 岁, 其中 69% 为女性, 81% 为欧美裔, 13% 为非美裔, 6% 为其他族群; 所有图像为正面视角或 30° 视角。本文将其数字化为 640×490 , 8-bit 灰度或 24-bit 彩色图像, 从 123 人中得到 593 个序列, 图像序列时长为 10~60 帧。表情分为

7 种类别:生气、蔑视、嫌恶、害怕、高兴、悲伤、惊讶,按照脸部活动编码系统的多步骤选择过程,其中 327 个序列满足这 7 种分类标准,并给出对应的类别标记。为此,我们以该 327 个

表情序列作为实验测试数据,表情示例见图 4。第 1 行为中性表情图像,第 2 行为对应的 7 种表情。各个类别的样本数如表 1 所列。



图 4 CK+中 7 种表情示例
Fig. 4 seven expressions in CK+

表 1 CK+中各类别样本数

Table 1 Sample size of each type in CK+

表情类别	样本数
愤怒	45
蔑视	18
嫌恶	59
害怕	25
高兴	69
悲伤	28
惊讶	83

3.2 识别方法

由于人脸表情识别并没有统一的协议,使得算法性能评估没有标准,有些算法采用多个峰值表情图像用于训练和识别,也有算法随机选取一定比率样本用于训练,剩下的用于识别,并重复多次得到平均识别率。文献[23]采用留一主体交叉验证法评估,即每次取一个人的表情作为测试集,其余作为训练集,最后计算总的识别正确率。我们采用[23]的评估方法,用留一主体交叉验证法评估识别性能。使用 SVM 分类器,其中核函数取高斯核,对得到的人脸表情位移矢量,对各分量做归一化处理,即均值为 0、标准方差为 1。由表 1 可看出,各类别样本数不平衡,容易导致样本数少的类别识别率较低,为此对样本数少的类别,通过添加很低的随机噪声的形式增加训练样本数,使样本类别平衡。

3.3 识别结果

未做样本类别平衡时,本文与文献[23]的评估协议一致,得到的混淆矩阵见图 5,其正确识别率为 95.1%。

	愤怒	蔑视	厌恶	恐惧	快乐	悲伤	惊讶	全体
愤怒	42 12.8%	0 0.0%	3 0.9%	0 0.0%	0 0.0%	2 0.6%	0 0.0%	89.4 10.6%
蔑视	0 0.0%	17 5.2%	1 0.3%	0 0.0%	0 0.0%	0 0.0%	1 0.3%	89.5 10.5%
厌恶	2 0.6%	0 0.0%	55 16.8%	0 0.0%	0 0.0%	0 0.0%	1 0.3%	94.8 5.2%
恐惧	0 0.0%	0 0.0%	0 0.0%	22 6.7%	0 0.0%	0 0.0%	0 0.0%	100 0.0%
快乐	0 0.0%	0 0.0%	0 0.0%	2 0.6%	68 20.8%	0 0.0%	0 0.0%	97.1 2.9%
悲伤	1 0.3%	1 0.3%	0 0.0%	1 0.3%	0 0.0%	25 7.6%	0 0.0%	89.3 10.7%
惊讶	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.3%	0 0.0%	82 25.1%	98.8 1.2%
全体	93.3 6.7%	94.4 5.6%	93.2 6.8%	88.0 12.0%	98.6 1.4%	89.3 10.7%	98.8 1.2%	95.1 4.9%

图 5 未做类别样本数平衡

Fig. 5 Not do category sample number balance

当对样本补充使得各类别样本数平衡时,依然采用留一主体交叉验证方法,得到的混淆矩阵见图 6。其正确识别率为 98.2%。

	愤怒	蔑视	厌恶	恐惧	快乐	悲伤	惊讶	全体
愤怒	43 13.1%	0 0.0%	2 0.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	95.6 4.4%
蔑视	0 0.0%	18 5.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.3%	94.7 5.3%
厌恶	2 0.6%	0 0.0%	57 17.4%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	96.6 3.4%
恐惧	0 0.0%	0 0.0%	0 0.0%	24 7.3%	0 0.0%	0 0.0%	0 0.0%	100 0.0%
快乐	0 0.0%	0 0.0%	0 0.0%	1 0.3%	69 21.1%	0 0.0%	0 0.0%	98.6 1.4%
悲伤	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	28 8.6%	0 0.0%	100 0.0%
惊讶	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	82 25.1%	100 0.0%
全体	95.6 4.4%	100 0.0%	96.6 3.4%	96.0 4.0%	100 0.0%	100 0.0%	98.8 1.2%	98.2 1.8%

图 6 类别样本数平衡的正确率

Fig. 6 Category samples is balanced

本文方法与文献[23]方法的各类别正确识别率如图 7 所示。

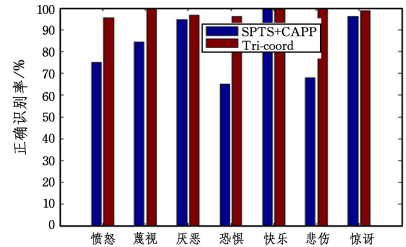


图 7 各类别正确识别率

Fig. 7 Correct recognition rate of each category

由图 7 可看出,采用广义三角坐标的人脸表情位移矢量作为特征(Tri_coord),使用高斯核的 SVM 分类器,相比文献[23],本文融合了相似归一化形状(SPTS)和正则观表(CAPP)特征,其识别性能有了较大的提高,其中愤怒类别的识别率从 74% 提高到 95.6%,蔑视类别从 84.4% 提高到 100%,恐惧类别从 65.2% 提高到 96%,悲伤类别从 68% 提高到 100%。

表 2 CK+的识别率

Table 2 Recognition rate on CK+

表情	M-CRT	Tri_coord(ours)
Anger	74.0	95.6
Contempt	—	94.7
Disgust	89.0	96.6
Fear	84.0	100
Happy	92.0	98.6
Sadness	85.3	100
Suprise	98.0	100
Neutral	94.3	—
Total accuracy	90.72	98.2

本文根据文献[23]所提供的表情类型将表情划分成7类,而文献[7]只划分了6类表情。但由表2仍可看出,我们所提出的识别方法在各种类别表情上都有提高,且总体识别准确率提高了7.48%。

结束语 本文对三角形坐标系进行了推广,给出了广义三角坐标用于人脸表情特征表示,结合高斯核 SVM 分类器,采用留一主体交叉验证技术,针对 CK+人脸表情数据库得到了人脸表情正确识别率为98.2%,相比于其基准^[23]和 M-CRT^[7]两种算法有了较大提高,表明了所提出的人脸表情特征表示方法的有效性。结合3维人脸特征,考虑人脸姿势改变对表情识别的影响,将是我们进一步研究的方向。

参 考 文 献

- [1] THIES J, ZOLLHÖFER M, NIEßNER M, et al. Real-time expression transfer for facial reenactment[J]. *ACM Transactions on Graphics (TOG)*, 2015, 34(6).
- [2] SARIYANIDI E, GUNES H, CAVALLARO A. Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(6): 1113-1133.
- [3] PRAMERDORFER C, KAMPEL M. Facial Expression Recognition using Convolutional Neural Networks: State of the Art [J]. arXiv:1612.02903.
- [4] LI S, DENG W. Deep Facial Expression Recognition: A Survey [J]. arXiv:1804.08348, 2018.
- [5] OJALA T, PIETIKÄINEN M, HARWOOD D. A Comparative Study of Texture Measures with Classification Based on Feature Distributions[J]. *Pattern Recognition*, 1996, 29: 51-59.
- [6] SHAN C, GONG S, MCOWAN P W. Facial expression recognition based on local binary patterns: A comprehensive study, *Image Vis*[J]. *Comput.*, 2009, 27(6): 803-816.
- [7] DU L, HU H. Modified classification and regression tree for facial expression recognition with using difference expression images[J]. *Electronics Letters*, 2017, 53(9): 590-592, 274.
- [8] EKMAN P, FRIESEN P. Facial action coding system: a technique for the measurement of facial movement[M]. Palo Alto: Consulting Psychologists Press, 1978.
- [9] ELEFThERIADIS S, et al. Variational Gaussian Process Auto-Encoder for Ordinal Prediction of Facial Action Units[C]// *Asian Conference on Computer Vision Springer, Cham*, 2016.
- [10] JENI L A, GIRARD J, COHN J, et al. Continuous A Uintensity estimation using localized, sparse facial feature space[C]// *Proc. IEEE Int. Conf. Autom. Face and Gesture Recognit. Workshops*, 2013: 1-7.
- [11] ZHONG L, LIU Q, YANG P, et al. Learning active facial patches for expression analysis[C]// *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012: 2562-2569.
- [12] LONG F, WU T, MOVELLAN J R, et al. Learning spatiotemporal features by using independent component analysis with application to facial expression recognition[J]. *Neurocomputing*, 2012, 93: 126-132.
- [13] ZHI R, FLIERL M, RUAN Q, et al. Graph-preserving sparse nonnegative matrix factorization with application to facial expression recognition[J]. *IEEE Trans. Systems, Man, Cybern. B, Cybern.*, 2011, 41(1): 38-52.
- [14] KRIZHEVSKY A, et al. Imagenet classification with deep convolutional neural networks[C]// *Advances in Neural Information Processing Systems*, 2012.
- [15] HASANI B, MAHOOR M H. Spatio-Temporal Facial Expression Recognition Using Convolutional Neural Networks and Conditional Random Fields[C]// *Proceedings 12th IEEE International Conference on Automatic Face and Gesture Recognition. FG 2017*.
- [16] RIFAI S, BENGIO Y, COURVILLE A, et al. Disentangling factors of variation for facial expression recognition[C]// *European Conference on Computer Vision*, 2012: 808-822.
- [17] LI H B, SUN J, XU Z B, et al. Multimodal 2D+3D Facial Expression Recognition with Deep Fusion Convolutional Neural Network[J]. *IEEE Transactions on Multimedia*, 2017, 19(12): 2816-2831.
- [18] USMAN M, et al. Using Deep Autoencoders for Facial Expression Recognition[J]. arXiv:1801.08329, 2018.
- [19] KAZEMI V, JOSEPHINE S. One millisecond face alignment with an ensemble of regression trees[C]// *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2014: 1867-1874.
- [20] 王守觉, 梁先扬. 图像变形计算方法及其应用[J]. *计算机辅助设计与图形学学报*, 2011, 23(8): 1304-1310.
- [21] CHANG C C, LIN C J. LIBSVM: a library for support vector machines[J]. *ACM Transactions on Intelligent Systems and Technology*, 2011, 2: 27: 1-27: 27.
- [22] 王守觉, 张晓盟, 肖泉, 等. 基于四面体坐标系的三维变形方法及应用[J]. *计算机辅助设计与图形学学报*, 2015, 27(3): 469-477.
- [23] LUCEY P, COHN J, KANADE T, et al. The extended Cohn-Kanade dataset (CK+): A completed dataset for action unit and emotion-specified expression[C]// *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2010: 94-101.



XIAO Xiao, born in 1980, Ph.D, lecturer. Her research interests include pattern recognition and intelligence computation.