

基于独立子空间分析特征学习的表情识别

詹永杰^{1,2}, 龙飞¹, 卜轶坤²

(1. 数字媒体计算研究中心厦门大学软件学院, 厦门 361005; 2. 厦门大学信息科学与技术学院, 厦门 361005)

摘要: 手工设计的特征(如 Gabor、LBP 等)在表情识别中得到了广泛的应用。独立子空间分析是一种无监督特征学习方法, 可从图像中学习出具有相位不变的特征。在表情识别应用中, 由于复杂背景的影响以及人脸对齐方法的局限性, 很难得到精确对齐的人脸图像序列。研究了在非精确对齐情况下, 基于独立子空间分析的表情识别问题。通过分析不同子空间尺寸下的表情识别效果发现, 在非精确对齐情况下, 选择合适的子空间尺寸能提升学到的特征对表情识别的鲁棒性。

关键词: 表情识别; 独立子空间分析; 无监督特征学习; 时空特征学习

中图分类号: TP 391.41

文献标识码: A

文章编号: 1004-731X (2015) 10-2316-05

Facial Expression Recognition with Independent Subspace Analysis Based Feature Learning

Zhan Yongjie^{1,2}, Long Fei^{1*}, Bu Yikun²

(1. Center for Digital Media Computing, Software School, Xiamen University, Xiamen 361005, China;

2. School of Information Science and Engineering, Xiamen University, Xiamen 361005, China)

Abstract: Hand-designed features (such as Gabor, LBP) has been widely employed in facial expression recognition. In the real-world applications of facial expression recognition, it is very difficult to achieve perfect face alignment because of the impact of complex background and the limitations of face alignment approaches. Independent Subspace Analysis (ISA) is an unsupervised feature learning method, which can be used to learn phase-invariant visual features from images. The problem of facial expression recognition based on ISA in the situation of not precise face alignment was investigated. *Through analyzing the facial expression recognition performances with different subspace size, it was turned out that choosing an appropriate subspace size is important to improve the robustness of learned features for facial expression recognition in the situation of not precise alignment.*

Keywords: facial expression recognition; independent subspace analysis; unsupervised feature learning; spatial-temporal feature learning

引言

人脸表情对识别人的情绪起到非常重要的作用。心理学家认为人的面部表情具有固定而有意义的结构, 这些结构能够反向推导出人的内在心理状

态。由于脸部信息可以通过非接触的方式(如摄像头)取得, 所以人脸图像分析的研究具有非常深远的意义。近年来, 表情识别作为人机交互学科的一部分, 受到了越来越多的关注。

根据是否使用表情的动态信息作为输入, 可以将表情识别方法分为两类。一类是基于单帧图像的表情识别, 另一类是基于图像序列的表情识别。基于图像序列的方法可以利用表情的动态变化特性从而进一步提高识别效果。



收稿日期: 2015-06-14 修回日期: 2015-07-30;
基金项目: 福建省自然科学基金项目(2014J01246);
虚拟现实技术与系统国家重点实验室开放基金
(BUAA-VR-14KF-01); 2014 年安徽省科技厅重大
科技专项项目(1301021018);
作者简介: 詹永杰(1990-), 男, 福建三明, 硕士生,
研究方向为计算机视觉、机器学习。

由于样本采集环境的复杂性、眼睛定位及人脸对齐方法的局限性,输入图像序列难以精确对齐是一个必须考虑的问题。一些表情识别方法是在单帧图像特征点定位的基础上实现的^[1]。逐帧处理时,非精确的特征点定位会对识别结果产生不利影响。另外,也有一些表情识别方法是在跟踪人脸运动的基础上设计的,其中包括基于外观的跟踪识别方法^[2]和基于特征点的跟踪识别方法^[3-4]。然而在不受控的环境下,跟踪效果很难保证不受到噪声或图像采集质量的影响,进而影响表情识别效果。文献[5]提出了从 LBP (Local Binary Patterns)发展出来的时空特征描述子,VLBP (Volume Local Binary Patterns)和 LBP-TOP (Local Binary Patterns From Three Orthogonal Planes),它们的提出推动了时空特征在表情识别中的发展和应用。文献[6]提出了利用 GME (Gabor Motion Energy)滤波器从视频中提取时空信息,实验证明这种方法比单纯用空间 Gabor 滤波器的方法在表情识别中效果更好。使用精心设计的手工特征,表情识别技术已经有了很大发展。然而,这些方法的识别效果在很大程度上受限于特征提取之前的人脸对齐精度。

ISA (Independent Subspace Analysis)是一种在 ICA (Independent Component Analysis)基础上发展出来的机器学习方法,它通过非线性模型找到高阶的统计独立性,可以从自然图像中学习出具有相位不变性的视觉特征^[7]。文献[8]提出了一种两层 ISA 模型,并将其应用于视频分类任务,证明了 ISA 用于时空特征学习的有效性。本文探讨了在非精确人脸对齐情况下,基于视频的表情识别鲁棒性问题。和传统方法如何提高样本对齐精度所不同的是,本文是从特征学习模型本身出发,探究 ISA 模型的子空间尺寸对非精确对齐情况下表情识别的影响。

1 ISA 时空特征学习

本节我们先简单介绍 ISA 算法,以及将其用于

时空特征学习的具体方法。之后,我们介绍怎样从一个图像序列构造特征表达,并用于表情识别任务。

ICA 提取的是线性特征,线性特征面临的一个问题就是缺乏不变性。从视神经科学的角度来讲,ICA 学到的特征检测器可以被看作是简单细胞^[7]。而一个复杂细胞对不同相位可以给出相同的响应,这就是所谓的相位不变性。线性系统的响应依赖于输入信号相位和检测器相位的一致性。ISA 通过对不同线性特征检测器的结果进行池化的方式实现了一个对相位不敏感的特征表达,其学习产生的特征检测器可以看作复杂细胞。

ISA 将观察数据的向量分解成相互独立的子空间来表示。从结构上来看,ISA 可以被看作是一个两层网络,如图 1 所示。

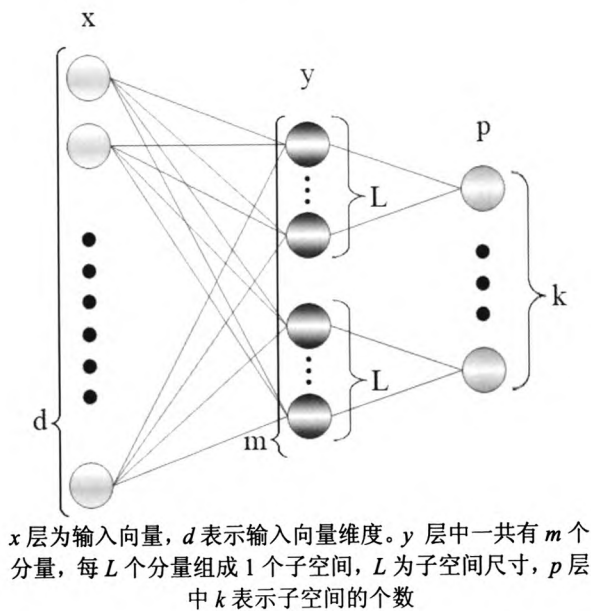


图 1 ISA 模型

第一层学到的是线性特征检测器,第二层对每个子空间内的 L 个线性特征检测器的结果进行池化。获取线性特征 y_i 和非线性特征 p_i 的公式如下:

$$y = W \cdot x \tag{1}$$

$$p_i = \sqrt{\sum_{j \in S_i} (y_j)^2} \quad i = 1, 2, \dots, k \tag{2}$$

其中 $W \in \mathbb{R}^{m \times d}$ 是线性特征检测器的权值, 将

白化后的输入 $x \in \mathfrak{R}^d$ 变换为特征 $y \in \mathfrak{R}^m$ 。能量检测器 p_i 代表第 i 个子空间 S_i 的特征。 L 和 k 作为参数输入。权值 W 则通过解下列优化问题学习得到：

$$\begin{aligned} &\underset{W}{\text{minimize}} \quad \sum_{n=1}^N \sum_{i=1}^k p_i(x_n) \\ &\text{subject to} \quad WW^T = I \end{aligned} \tag{3}$$

其中 N 为输入样本个数。

下面介绍将 ISA 用于表情识别中的时空特征学习方法。首先，从表情视频中抽取大量的 3D 时空图像块，构成训练数据。将训练数据经过 PCA (Principal Component Analysis)降维和白化预处理后，输入到 ISA 网络进行时空滤波器的学习。在 ISA 学习产生时空滤波器之后，将其和输入视频进行卷积提取时空特征。直接卷积的结果是产生数量巨大的特征。为了降低特征维数，我们对卷积产生的特征序列进行下采样处理，最后采用 SVM (Support Vector Machine)方法进行分类。

2 实验

MMI 数据库是世界著名的表情识别数据库。数据库包括 30 个学生和研究人员，年龄从 19 岁到

62 岁不等。和 Cohn-Kanade 数据库相比，MMI 是一个更有挑战性的数据库。眼镜、头巾、胡须等因素加大了识别难度。

MMI 数据库的六种表情数量统计如表 1 所示。

表 1 MMI 数据库的六种表情数量统计

Expression	Count
Anger	33
Disgust	32
Fear	29
Smile	42
Sadness	32
Surprise	41

本文实验是用 MMI 数据库中的表情数据进行的，按照能否分为六种表情的原则，我们从中提取了 209 段视频。对于每一帧图像，通过手工标定眼睛位置对人脸进行标定。根据眼睛的位置将图像旋转矫正，提取矩形区域。为了探究非精确对齐情况下的识别效果，对每帧图像得到的矩形区域的位置、大小和旋转参量加入零均值的高斯噪声。然后用最近邻插值将图像缩小到 96×96。图 2 对比了不加噪声和加高斯噪声的情况下剪裁得到的图像序列。



图 2 MMI 数据库中人脸表情图像序列样例

实验探究了 ISA 方法中不同子空间尺寸对非精确对齐表情识别效果的影响。同时与时空 Gabor 滤波器进行了比较。实验采用 hold-out 交叉验证方法，每次实验从每种表情的图像序列样本中随机选取 75%作为训练集，剩下 25%作为测试集。使用 AUC (area under ROC curve)来评估结果的好坏。重

复 20 次实验，以平均值作为最终结果。
实验中，我们采样 16×16×10 的时空图像块作为输入训练数据，学习 120 个线性特征检测器，改变子空间尺寸，观察识别效果的变化。ISA 学习产生的时空特征如图 3 所示。

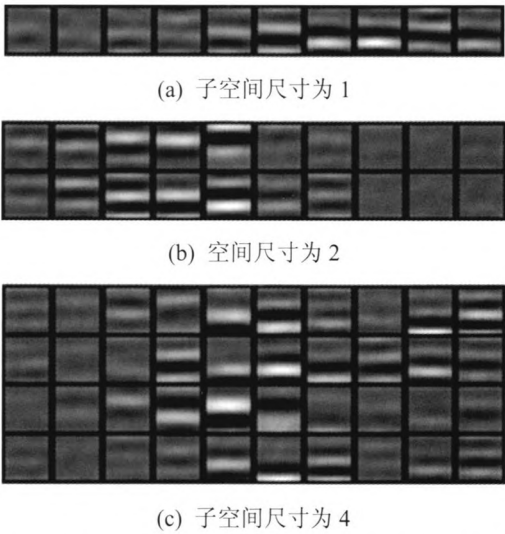


图 3 ISA 从 MMI 数据库学到的 16×16×10 的时空特征

表 2 不同子空间尺寸 ISA 模型以及时空 Gabor 滤波器的识别结果(AUC).

Method	Anger	Disgust	Fear	Smile	Sadness	Surprise	Mean
Sub1	0.8343	0.8225	0.5703	0.9060	0.8059	0.9221	0.8102
Sub2	0.8507	0.8333	0.6898	0.9303	0.8071	0.9250	0.8394
Sub3	0.8588	0.8438	0.6396	0.9308	0.8102	0.9341	0.8362
Sub4	0.8610	0.8467	0.6609	0.9425	0.7888	0.9105	0.8351
Sub5	0.8740	0.8482	0.6736	0.9555	0.8155	0.9323	0.8498
Sub6	0.8705	0.8666	0.6581	0.9556	0.8232	0.9267	0.8501
Sub8	0.8741	0.8577	0.6519	0.9555	0.8371	0.9204	0.8494
Sub10	0.8911	0.8650	0.6507	0.9531	0.8100	0.9271	0.8495
Sub12	0.8839	0.8372	0.6323	0.9609	0.8384	0.9182	0.8452
Sub15	0.8922	0.8695	0.6384	0.9542	0.8176	0.9208	0.8488
Sub20	0.8870	0.8676	0.6502	0.9629	0.8342	0.9189	0.8535
Sub24	0.8940	0.8636	0.6456	0.9563	0.8291	0.9071	0.8493
Sub30	0.9013	0.8664	0.6842	0.9574	0.8176	0.9216	0.8581
Sub40	0.8943	0.8650	0.6331	0.9529	0.8434	0.9096	0.8497
Sub60	0.8835	0.8783	0.6503	0.9510	0.8270	0.8973	0.8479
Gabor	0.7972	0.7633	0.6908	0.9398	0.8228	0.8958	0.8183

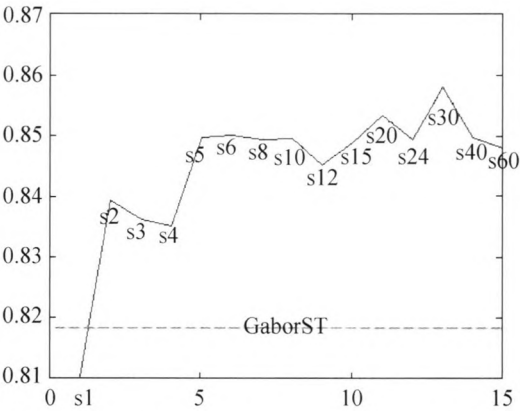


图 4 ISA 识别效果(AUC)随着子空间尺寸的变化情况

使用 ISA 方法学习出时空滤波器后,用前面介绍的方法得到表情样本的特征表达,然后用线性 SVM 进行分类。实验结果的 AUC 比较如表 2 所示。

从表中可以看出, ISA 在子空间尺寸 ≥ 2 的情况下,识别效果优于时空 Gabor 滤波器。在子空间尺寸为 30 的时候效果最好。识别效果随着子空间尺寸变化情况如图 4 所示。从图中可以看出,随着子空间尺寸增加,识别率呈上升趋势。在子空间尺寸较小的时候,这种上升趋势更加明显。

3 结论

本文探讨了在非精确对齐情况下,基于 ISA 特征学习的表情识别问题。我们分析了在不同子空间尺寸情况下,基于 ISA 特征学习方法的表情识别效果。实验结果显示,识别效果随着子空间尺寸变化总体上呈上升趋势。在子空间尺寸较小的区段,这种上升趋势更加明显。可见,选择合适的 ISA 子空间尺寸能够提升学到的特征对非精确对齐情况下表情识别的鲁棒性。

5 结论

本文针对虚拟装配操作仿真问题, 从人机工程理论出发, 基于多目标遗传算法结合人机工程学理论构建了虚拟人智能操作仿真模型, 以发动机轴承盖紧固螺栓装配为例进行了仿真验证。结果表明, 该方法能较真实的仿真出实际操作的结果, 相对原有装配仿真方法, 优化模型的引入减少了仿真中手动调节的工作, 仿真人员主观因素的影响被极大削弱, 仿真结果的可信度得到有效提升, 对于产品可装配性的仿真分析具有较大的实际应用意义。

参考文献:

- [1] Zhu H, Wu D, Fan X. Interactive assembly tool planning based on assembly semantics in virtual environment[J]. *The International Journal of Advanced Manufacturing Technology* (S0268-3768), 2010, 51(5): 739-755.
 - [2] 王朝晖, 陈昱, 朱心雄. 一种虚拟人作业行为的自主优化模型 [J]. *软件学报*, 2012, 23(9): 2358-2373.
 - [3] 李石磊, 梁加红, 吴冰, 等. 虚拟人运动生成与控制技术综述 [J]. *系统仿真学报*, 2011, 23(9): 1758-1771.
 - [4] 杨锋, 袁修干. 基于舒适度最大化的人体运动控制 [J]. *计算机辅助设计与图形学学报*, 2005, 17(2): 267-272.
 - [5] 邱世广, 武殿梁, 范秀敏, 等. 基于船舶人机工程的虚拟人操作驱动建模仿真技术 [J]. *上海交通大学学报*, 2012, 46(9): 1366-1370.
 - [6] 袁修干, 庄达民, 张兴娟. *人机工程计算仿真* [M]. 北京: 北京航空航天大学出版社, 2005.
 - [7] Edward Y L Gu. *A Journey from Robot to Digital Human: Mathematical Principles and Applications with MATLAB Programming* [M]. Germany: Springer-Verlag Berlin and Heidelberg GmbH & Co. K, 2013: 422-463.
 - [8] 霍伟. *机器人动力学与控制* [M]. 北京: 高等教育出版社, 2005.
 - [9] 王斌锐, 李璟, 金英连, 等. *运动建模与控制系统设计* [M]. 北京: 清华大学出版社, 2014.
 - [10] D B Chaffin, G B J Andersson. *Occupational Biomechanics* [M]. USA: John Wiley & Sons, 1991.
 - [11] Qiu S, He Q, Fan X, Wu D. Virtual human hybrid control in virtual assembly and maintenance simulation [J]. *International Journal of Production Research* (S0020-7543), 2014, 52(3): 867-887.
 - [12] Tolani D, Goswami A, Badler N I. Real-time inverse kinematics techniques for anthropomorphic limbs [J]. *Graphical Models* (S1524-0703), 2000, 62(5): 353-388.
-
- (上接第 2319 页)
- ### 参考文献:
- [1] Jaewon Sung, Daijin Kim. Real-time facial expression recognition using STAAM and layered GDA classifier [J]. *Image and Vision Computing* (S0262-8856), 2009, 27: 1313-1325.
 - [2] Fadi Dornaika, Abdelmalik Moujahid, Bogdan Raducanu. Facial expression recognition using tracked facial actions: Classifier performance analysis [J]. *Engineering Applications of Artificial Intelligence* (S0952-1976), 2013, 26: 467-477.
 - [3] Yongqiang Li, Shangfei Wang, Yongping Zhao, *et al.* Simultaneous Facial Feature Tracking and Facial Expression Recognition [J]. *IEEE Transactions on Image Processing* (S1057-7149), 2013, 22: 2559-2573.
 - [4] Hui Fanga, Neil Mac Parthaláin, Andrew J Aubrey, *et al.* Facial expression recognition in dynamic sequences: An integrated approach [J]. *Pattern Recognition* (S1369-703X), 2014, 47: 1271-1281.
 - [5] Guoying Zhao, Matti Pietikainen. Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (S0162-8828), 2007, 29: 915-928.
 - [6] Wu Tingfan, Bartlett Marian S, Movellan Javier R. Facial Expression Recognition Using Gabor Motion Energy Filters [C]// *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition - Workshops (CVPRW)*. USA: IEEE, 2010: 42-47.
 - [7] A Hyvarinen, J Hurri, P Hoyer. *Natural Image Statistics* [M]. Germany: Springer, 2009.
 - [8] Quoc V Le, Will Y Zou, Serena Y Yeung, *et al.* Learning hierarchical invariant spatiotemporal features for action recognition with independent subspace analysis [C]// *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. USA: IEEE, 2011: 3361-3368.