

基于动态时间规整和主动外观模型的动态表情识别

许良凤^{*①} 王家勇^① 崔婧楠^② 胡 敏^① 张柯柯^① 滕文娣^①

^①(合肥工业大学计算机与信息学院 合肥 230009)

^②(华南理工大学计算机与信息学院 广州 510006)

摘 要: 针对静态表情特征缺乏时间信息, 不能充分体现表情的细微变化, 该文提出一种针对非特定人的动态表情识别方法: 基于动态时间规整(Dynamic Time Warping, DTW)和主动外观模型(Active Appearance Model, AAM)的动态表情识别。首先采用基于局部梯度 DT-CWT(Dual-Tree Complex Wavelet Transform)主方向模式(Dominant Direction Pattern, DDP)特征的 DTW 对表情序列进行规整。然后采用 AAM 定位出表情图像的 66 个特征点并进行跟踪, 利用中性脸的特征点构建人脸几何模型, 通过人脸几何模型的匹配克服不同人呈现表情的差异, 并通过计算表情序列中相邻两帧图像对应特征点的位移获得表情的变化特征。最后采用最近邻分类器进行分类识别。在 CK+ 库和实验室自建库 HFUT-FE(HeFei University of Technology-Face Emotion)上的实验结果表明, 所提算法具有较高的准确性。

关键词: 动态表情识别; 动态时间规整; 主动外观模型; 双树复小波变换; 主方向模式

中图分类号: TP391.41

文献标识码: A

文章编号: 1009-5896(2018)02-0338-08

DOI: 10.11999/JEIT170416

Dynamic Expression Recognition Based on Dynamic Time Warping and Active Appearance Model

XU Liangfeng^① WANG Jiayong^① CUI Jingnan^②
HU Min^① ZHANG Keke^① TENG Wendi^①

^①(School of Computer and Information, Hefei University of Technology, Hefei 230009, China)

^②(School of Computer and Information, South China University of Technology, Guangzhou 510006, China)

Abstract: To overcome the deficiency of static expression feature, which lacks time information and can not reflect the subtle changes of expression adequately, a dynamic expression recognition method is proposed for non-specific face: the dynamic expression recognition based on Dynamic Time Warping (DTW) and Active Appearance Model (AAM). Firstly, the method of DTW based on local gradient Dual Tree-Complex Wavelet Transform (DT-CWT) dominant direction pattern is used to warp expression sequence. Secondly, using AAM to locate 66 feature points of face image and track them. The changing feature of expression can be obtained by calculating the displacement of corresponding feature points in two adjacent expression sequences image. And using the feature points of neutral face to build the facial geometry model. The matching of facial geometry model can overcome the expression differences between various people. Finally, the nearest neighbor classifier is used for classification and recognition. The experimental results on CK+ database and HeFei University of Technology-Face Emotion (HFUT-FE) database show that the proposed algorithm has a high degree of accuracy.

Key words: Dynamic expression recognition; Dynamic Time Warping (DTW); Active Appearance Model (AAM); Dual Tree-Complex Wavelet Transform (DT-CWT); Dominant direction pattern

1 引言

人脸表情是人们内心活动的一种外在表现, 甚

至包含了语言中无法体现的重要信息。早在 20 世纪 70 年代, Mehrabian^[1]就展开了一系列的研究, 结果表明在人们交流的过程中, 语句本身所展现出来的情感信息仅仅只占 7%, 而人脸表情却包含了 55% 的情感信息。人的情感在很大程度上由面部表情来呈现, 在日常交际中, 可以通过表情的变化来判断人们的内心想法。近年来, 随着人工智能和模式识别的迅速发展, 人脸表情识别逐渐成为一个研究热点^[2-5]。

收稿日期: 2017-05-05; 改回日期: 2017-11-08; 网络出版: 2017-12-04

*通信作者: 许良凤 xulfejn@hfut.edu.cn

基金项目: 国家自然科学基金(61300119, 61432004), 安徽省自然科学基金(1408085MKL16)

Foundation Items: The National Natural Science Foundation of China (61300119, 61432004), The National Natural Science Foundation of Anhui Province (1408085MKL16)

人脸表情识别包括 4 个步骤：人脸检测、图像预处理、特征提取和分类识别。其中，特征提取是表情识别的重要环节，表情特征是影响识别结果的重要因素。根据表情库类型的不同，表情特征可以分为静态表情特征和动态表情特征。静态表情特征是从静态表情图像中提取的，特征提取速度快而且简单，对于特定人的表情识别具有较好的识别效果。但是静态表情特征包含的仅仅是表情的空间信息，缺乏表情的时间信息，并且容易受到人脸形状、肤色、光照变化等因素的影响，对非特定人的表情识别效果较差。动态表情特征体现的是表情变化的过程，从本质上展现了表情的特性，可以很好地刻画表情的细微变化，对非特定人的表情识别效果较好。因此，近年来许多研究者把研究重心转移到动态表情识别上。动态表情识别以动态表情库为基础，通过分析表情帧之间的相关性，提取表情的变化特征。动态表情识别结合了表情图像的时间和空间信息，提取了更加完整的表情特征。

文献 [6] 提出了一种基于 LBP-TOP(Local Binary Patterns on Three Orthogonal Planes)+SRC(Sparse Representation-based Classification)的动态人脸表情识别方法，该方法将 LBP 推广到 3 维空间，通过提取 3 个正交平面 XY、XT、YT 的 LBP 来获得动态表情特征，并用 SRC(Sparse Representation-based Classifier)进行分类识别。文献 [7] 利用 LPQ-TOP(Local Phase Quantization from Three Orthogonal Planes)特征描述人脸的运动单元，并用 SRC 进行分类识别。文献 [8] 提出了一种基于字典的人脸表情分析方法，该方法首先通过 FACS 建立 AU(Action Unit)字典，然后又提出了一种新的算法来学习得到 SP(Structure-Preserving)字典，对于表情图像，将其视为字典的成分并纳入其中，从而进行表情分析。文献 [9] 提出了一种基于 PHOG-TOP(Pyramid Histogram of Oriented Gradients on Three Orthogonal Planes)和光流法的动态表情识别方法，分别提取表情序列的 PHOG-TOP 特征和光流特征，然后将两种特征结合起来，并用 SVM 分类器进行分类。文献 [10] 提出了一种基于 AAM(Active Appearance Model)的动态人脸表情识别方法，首先利用 AAM 定位人脸表情图像的 68 个特征点，然后将表情序列中表情帧与中性帧相对应的特征点的坐标相减，用坐标差作为表情序列的特征。文献 [11] 利用差分 AAM 来提取动态表情特征，首先通过表情图像和中性图像的参数差分来计算差分 AAM，然后运用流形学习对差分 AAM 进行平滑和降维。

上述提出的动态表情识别方法虽然取得了不错的识别效果，但是它们并没有考虑到人脸信息对表情识别的影响，即所做的研究是针对特定人的表情识别。为了增加表情识别的实用性，针对非特定人的动态表情识别也一直是研究的焦点，针对非特定人的表情识别主要突出不同人脸呈现同一表情的共性，而忽略其个性，即排除人脸因素对表情识别影响。

本文在文献 [10,12,13] 的启发下，提出了一种针对非特定人的动态表情识别方法：基于 DTW(Dynamic Time Warping)和 AAM 的动态表情识别。该方法的主要步骤如下：(1)采用基于局部梯度 DT-CWT(Dual Tree-Complex Wavelet Transform)主方向模式的 DTW 对表情序列进行规整。(2)采用 AAM 定位出表情图像的 66 个特征点并进行跟踪。对于序列中的中性表情图像，利用特征点之间的距离来刻画人脸的几何模型。对于整个表情序列，计算相邻两帧图像对应特征点之间的位移形成表情特征。(3)对于一个测试表情序列，首先将中性人脸表情的几何模型与训练样本的中性人脸模型进行匹配，寻找 K 个近邻样本。然后，将 K 个近邻样本所对应的表情序列作为分类识别的依据，而其他的表情序列不参与该测试样本的识别。最后，采用最近邻分类器进行分类识别。实验部分将分别在 CK+库和实验室自建库 HFUT-FE(HeFei University of Technology-Face Emotion)上进行验证，实验结果证明了本文方法的优越性。本文方法的优点：(1)利用 DTW 对表情序列进行规整，使表情序列更加规范化。(2)与大多数 AAM 定位 68 个人脸特征点相比，本文定位了 66 个特征点，去掉了鼻子附近的两个特征点，因为人脸发生表情变化时，鼻子附近基本没有变化。所以在保证特征质量的同时降低了特征维数。(3)通过中性人脸几何模型的建立和匹配，有效排除了人脸信息对表情识别的影响，针对非特定人的表情识别具有较好的效果。

2 表情序列的动态时间规整

由于每个人对表情呈现的快慢不一样，所以每个表情序列的长度不一样。在表情识别时，需要对表情序列进行预处理使其长度相同。目前大多数表情序列的规整都是手动的，具有一定的人为性和不规范性，对识别效果造成影响。本文采用基于局部梯度 DT-CWT 主方向模式特征的 DTW 对表情序列进行规整，与手动规整相比，规整后的表情序列更加规范化。

2.1 动态时间规整概述

DTW 的优势在于能够解决时间序列的扭曲，

实现两个时间序列的对齐,并且可以根据对齐的时间序列计算相似度^[14]。图1展示了两个时间序列的欧氏距离和DTW距离,由图可以看出,欧氏距离只是简单地计算两个序列对应时间点之间的距离和,而DTW距离计算的是序列对齐后的距离和。由于DTW性能的优越性,其在模板匹配、序列对齐、相似度估计及分类等方面的应用相当广泛。

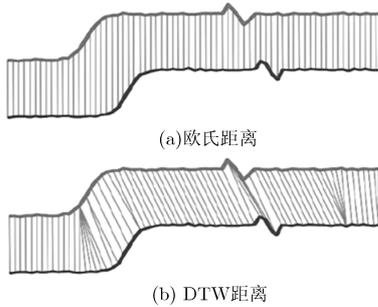


图1 欧氏距离与DTW距离的对比

DTW基本思想是:尽量使两个时间序列之间的距离最小,实现一个序列到另一个序列的非线性映射。对于两个时间序列 \mathbf{A} 和 \mathbf{B} ,它们的特征用向量可表示为 $\mathbf{A} = \{a_1, a_2, \dots, a_i\}$ 和 $\mathbf{B} = \{b_1, b_2, \dots, b_j\}$,其中 $i \neq j$,若用DTW将序列 \mathbf{B} 对齐到序列 \mathbf{A} ,需要寻求一个时间规整函数 \mathbf{W} ,使序列 \mathbf{B} 的时间轴 j 能非线性的映射到序列 \mathbf{A} 的时间轴 i 上。假设 $\mathbf{W} = \{w(1), w(2), \dots, w(N)\}$ (N 是规整路径长度), $w(n) = (i(n), j(n))$ 表示两个序列第 n 个相对应的点,把这两个对应的点之间的距离记为 $d(i(n), j(n))$ 。将 $1 \sim n$ 所有相对对应点的距离加权后累加,寻找出一条距离最短的路径:

$$D = \min \sum_{n=1}^N [d(i(n), j(n))] \quad (1)$$

2.2 基于DT-CWT的动态时间规整

DTW是通过计算图像特征的相似度来实现图像的对齐,因此,高区分性的特征是DTW的基础。DT-CWT具有近似的平移不变性和方向选择性,可以从多层次多方向上提取图像的特征^[15]。本文提取表情图像的局部梯度DT-CWT主方向模式特征,首先采用DT-CWT提取表情图像的特征,为了进一步增加特征的区分性,构建一种新的主方向模式(Improved Dominant Direction Pattern, IDDP)对DT-CWT特征进行编码,并对编码后的特征进行梯度方向融合^[16]。

2.2.1 IDDP编码 IDDP算子是对DDP算子^[17]的改进,DDP算子的邻域大小为 3×3 ,每一个像素点

的周围8邻域分别对应由3 bit二进制数构成的8个方向数,如图2(a)所示。DDP编码的基本思想为:比较中心像素点的8邻域像素值,选出最大值和次大值作为主要像素进行编码,如果2个或2个以上的邻域像素拥有相同的最大值或次大值,选择较小方向数上的像素作为最大值或次大值。DDP编码规则为 $\min(D_1 \times 8 + D_2, D_2 \times 8 + D_1)$,其中 D_1 和 D_2 分别是最大像素值和次大像素值所对应的方向数。DDP编码后灰度值范围为 $1 \sim 55$,相对于LBP, LGC和LDC等一些经典算子,DDP算子在提取丰富纹理信息的基础上有效地降低了特征数据间的冗余。由于DDP在处理拥有相同最大值和次大值的邻域像素时,只选择较小方向数上的像素,而舍弃其他方向数上的像素,造成图像特征的丢失。故本文对DDP进行了改进,把那些拥有相同最大值和次大值的邻域像素都作为主要像素进行编码。编码规则为

$$\min \left\{ \left[8 \times \sum_{i=1}^N D_i / N \right] + \left[\sum_{j=N+1}^{N+M} D_j / M \right], \right. \\ \left. \left[8 \times \sum_{j=1}^M D_j / M \right] + \left[\sum_{i=M+1}^{N+M} D_i / N \right] \right\}$$

其中, D_i, D_j 分别表示邻域像素最大值和次大值所对应的方向数, N, M 分别表示拥有相同最大值和次大值的邻域像素个数, $N \in [1, 8], M \in [0, 7], M + N \in [1, 8]$, $\lceil \bullet \rceil$ 表示向上取整。

对于同一幅 3×3 图像,DDP, IDDP编码示例分别如图2(b),图2(c)所示。在 3×3 图像中,最大值91所对应的方向数分别为010和111;次大值80对应的方向数分别为101和110。根据DDP的编码规则,选择较小方向数上的最大值和次大值进行编码,最大值和次大值较小方向分别为010和101,并且 $(010)_2 = (2)_{10}, (101)_2 = (5)_{10}, \min(2 \times 8 + 5, 5 \times 8 + 2) = 2 \times 8 + 5 = 21$ 。根据IDDP的编码规则,对所有方向上的最大值和次大值进行编码,最大值方向数为010和111,次大值方向数为101和110, $\min\left(\frac{(7 \times 8 + 2 \times 8)}{2} + \left\lceil \frac{(6 + 5)}{2} \right\rceil, \frac{(6 \times 8 + 5 \times 8)}{2} + \left\lceil \frac{(7 + 2)}{2} \right\rceil\right) = \frac{(7 \times 8 + 2 \times 8)}{2} + \left\lceil \frac{(6 + 5)}{2} \right\rceil = 42$ 。

与DDP相比, IDDP结合了更多邻域的方向信息,从而使得编码后的特征具有更高的区分性。如图3所示,用DDP和IDDP分别对4幅灰度分布不同的 3×3 图像进行编码,4幅图像的DDP编码值相同,均为10。而IDDP编码值却各不相同,因此证明IDDP特征具有更高的鉴别能力。

2.2.2 表情序列的规整 对于待规整表情序列,将其与标准序列对齐的步骤如下:

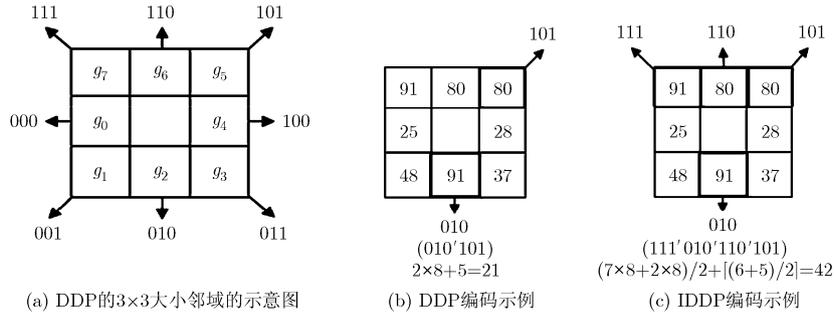


图 2 DDP 和 IDDP 编码示例

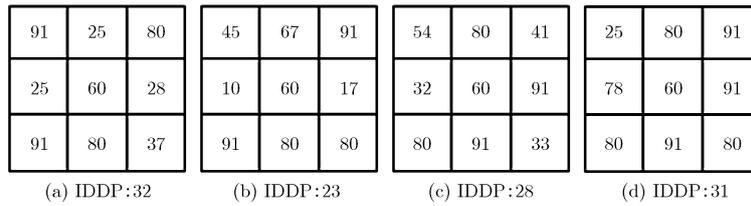


图 3 不同灰度分布示意图

步骤 1 提取两个表情序列中每张图像的局部梯度 DT-CWT 主方向模式特征。待规整表情序列和标准序列的特征矩阵分别为 P, Q ，其中 P, Q 的每一列代表序列中一张表情图像的特征；

步骤 2 对矩阵 P, Q 的列向量两两之间求距离，得到一个距离矩阵 W ，称之为代价矩阵；

步骤 3 在代价矩阵中，通过式(1)搜索出一条最短路径，在这条路径下序列之间最为相似，即对齐效果最好。

图 4 展示了基于局部梯度 DT-CWT 主方向模式对一段待规整表情序列进行动态时间规整的效果图。

3 基于 AAM 的特征提取

AAM 集纹理信息与形状信息于一体，建立的模

型与原图像契合度极高，特征点的定位和跟踪较准确。本文利用 AAM 定位了人脸表情图像的 66 个特征点并跟踪，以一种新的思路来提取面部表情特征。首先通过中性脸几何模型的建立和匹配来克服不同人呈现表情的差异，有效降低了人脸特征对表情识别的影响；然后通过计算相邻两帧图像对应特征点之间的位移来获得脸部表情的运动特征。

(1)几何特征：如式(2)所示，通过计算中性脸两两特征点之间的相对距离来建立人脸几何模型。

$$d_{ij} = \|p_i - p_j\| = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (2)$$

其中， $i, j \in [1, 66], i \neq j, (x_i, y_i)$ 和 (x_j, y_j) 表示特征点的坐标， d_{ij} 表示特征点间的距离。

(2)动态表情特征：假设一个表情序列 X 共有 n 帧图像，表情序列可用向量表示为

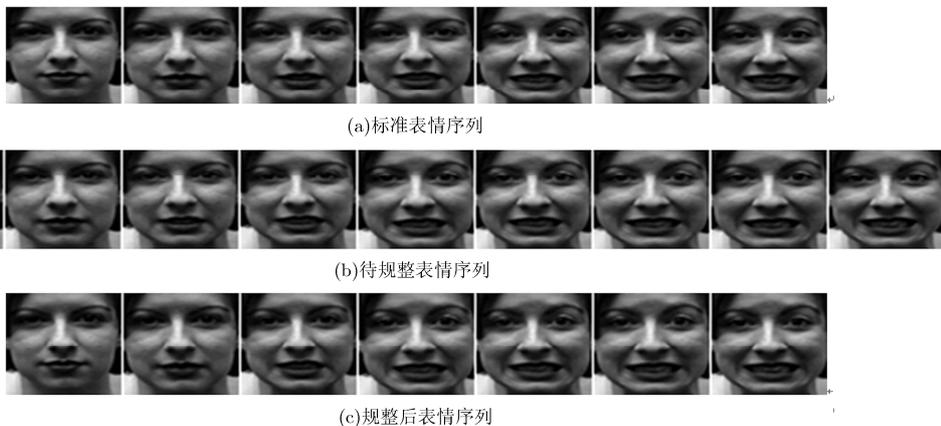


图 4 表情序列规整的效果图

$$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_i, \dots, \mathbf{X}_n), \quad 1 \leq i \leq n \quad (3)$$

式(3)中, \mathbf{X}_i 表示第 i 帧图像, 对于表情序列中第 i 帧图像的第 k 个特征点, 其坐标可表示为 (x_i^k, y_i^k) 。

因此, 在表情序列中第 k 个特征点的坐标可分别表示为 $(x_1^k, y_1^k), (x_2^k, y_2^k), \dots, (x_n^k, y_n^k)$, 通过计算相邻两坐标点的位移获得第 k 个特征点在表情序列中的运动特征。位移的计算公式如式(4)、式(5):

$$D_{(i,i-1)}^k = \sqrt{(x_i^k - x_{i-1}^k)^2 + (y_i^k - y_{i-1}^k)^2} \quad (4)$$

$$a_{(i,i-1)}^k = \begin{cases} \frac{y_i^k - y_{i-1}^k}{x_i^k - x_{i-1}^k}, & x_i^k \neq x_{i-1}^k \\ \frac{\pi}{2}, & x_i^k = x_{i-1}^k \end{cases} \quad (5)$$

式(4)和式(5)中, $2 \leq i \leq n$, $D_{(i,i-1)}^k$ 和 $a_{(i,i-1)}^k$ 分别表示第 k 个特征点从第 $i-1$ 帧图像到第 i 帧图像的位移大小和方向。

4 实验结果及分析

为了检验和分析基于 DTW 和 AAM 动态表情识别算法的性能, 本文实验基于 Windows XP 系统上(单核 2.8 GHz CPU, 2.0 GB 内存), 利用 VS2010 和 MATLAB2010a 在 CK+库和实验室自建表情库 HFUT-FE 上进行了一系列实验。

HFUT-FE 是动态表情库, 拍摄于合肥工业大学情感计算研究所, 由情感所的所有研究人员共同完成拍摄, 对每位参与拍摄的人员均采集 6 种表情, 每个人的每种表情由一个表情序列来体现。为了使表情库接近现实环境, 在拍摄过程中对光照和姿态进行了改变。HFUT-FE 库的部分表情序列如图 5 所示。

4.1 实验步骤

本文算法的步骤如下:

(1)训练阶段:

步骤 1 提取表情图像的局部梯度 DT-CWT 主方向模式特征, 然后利用 DTW 对表情序列进行规整;

步骤 2 利用 AAM 定位表情特征点并跟踪;

步骤 3 对于序列中的中性表情图像, 通过计算关键点之间的欧氏距离建立人脸几何模型, 对于整个表情序列, 通过计算特征点在相邻两帧图像的位移大小和方向得到动态表情特征。

(2)测试阶段: 对于一个测试表情序列, 其识别过程为:

步骤 1 提取表情图像的局部梯度 DT-CWT 主方向模式特征, 然后利用 DTW 对表情序列进行规整;

步骤 2 利用 AAM 定位表情特征点并跟踪, 并建立中性表情图像的人脸几何模型和表情序列的动态特征;

步骤 3 将人脸模型与训练库中的人脸模型进行匹配, 寻找 K 个最近的人脸模型;

步骤 4 将找到的 K 个人脸模型所对应的表情序列的动态特征作为识别的依据, 利用最近邻分类器分类识别。

4.2 K 值的选取

本文针对非特定人进行表情识别, 排除人脸对表情识别的影响, 所以同一个人的 6 种表情只出现在训练库中或测试库中。对于 CK+库, 选取 41 个人的表情共 246 个序列作为实验的样本, 随机选取 31 个人的表情序列作为训练库, 利用剩下的 10 个人的表情序列进行测试; 对于 HFUT-FE 库, 选取 20 个人的表情共 120 个序列作为实验的样本, 随机选取 12 个人的表情序列作为训练库, 利用剩下的 8 个人的表情序列进行测试。在每个库上均做 3 次实验, 用平均识别率作为识别的结果。

由于不同人的面部轮廓、五官存在一定的差异, 在表达相同的情感时也会产生差异, 所以在进行非特定人的表情识别时, 可以先匹配出与测试序列所对应的人脸模型最接近的人脸, 人脸模型越相似, 所呈现的表情特征也就越接近, 利用相似人脸模型对应的表情序列进行识别, 可以改善识别效果。由于客观因素的影响, 最近的人脸呈现出来的表情特征不一定是明显的, 因此本文寻找测试人脸模型的 K 个近邻样本。



图 5 HFUT-FE 库部分表情序列

为了选取 K 的最优值，在 CK+库和 HFUT-FE 库上做了一系列实验，实验结果如表 1、表 2 所示。由表 1、表 2 可以看出，CK+库和 HFUT-FE 库的实验结果呈相同的趋势，当 K 的值为 3 时，识别率最高。经过分析，其原因可归纳成两个方面：一方面，当 K 值小于 3 时，虽然可以匹配到相似的人脸模型，但是样本量不够充分，不能有效地诠释一类表情的特征；另一方面，当 K 值大于 3 时，虽然样本量充足，但是匹配的近邻人脸模型过多，导致某些人脸模型与测试的人脸模型相似度并不高，这类人脸模型对应的表情特征会对识别的准确性造成干扰。所以，选取 K 的值为 3。

表 1 CK+库上不同 K 值的识别率比较

K	第 1 组	第 2 组	第 3 组	平均识别率(%)
1	83.33	83.33	88.33	84.44
2	88.33	85.00	90.00	87.78
3	91.66	86.67	91.66	90.00
4	88.33	86.67	90.00	88.33
5	85.00	85.00	88.33	85.56

表 2 HFUT-FE 库上不同 K 值的识别率比较

K	第 1 组	第 2 组	第 3 组	平均识别率(%)
1	68.75	64.58	64.58	64.19
2	72.92	70.83	68.75	70.83
3	75.00	75.00	72.92	74.31
4	72.92	72.92	70.83	72.22
5	68.75	66.67	66.67	67.36

对于参数 K ，当训练样本较少时， K 的取值可能略有不同，但是当训练样本足够多时，训练样本包含的人脸类型足够丰富时，对于不同的测试集， K 的值将基本一致。

4.3 实验结果分析

4.3.1 本文算法实验结果分析 表 3、表 4 分别列出了本文算法在 CK+库和 HFUT-FE 库上的实验结果，由表 3、表 4 可以看出 CK+库上的平均识别率为 90%，在 HFUT-FE 库上的平均识别率为 74.31%。本文算法在两个数据库上的识别率存在一定的差距，经过对数据库的分析可知 CK+库的拍摄环境比较单一，图片的质量比较高，没有光照和姿态的变化，所以识别率较高。而 HFUT-FE 库更加接近现实环境，拍摄时人脸的姿态和光照会产生一定的变化，因此相比 CK+库识别率较低。

表 3 基于 CK+人脸表情库的实验结果

表情	测试样本数	正确识别数			平均识别率(%)
		第 1 组	第 2 组	第 3 组	
高兴	10	9	9	10	93.33
悲伤	10	9	8	9	86.67
生气	10	10	9	8	90.00
惊讶	10	9	9	9	90.00
厌恶	10	10	9	9	93.33
恐惧	10	8	8	10	86.77
总计	60	55	52	55	90.00

表 4 基于 HFUT-FE 人脸表情库的实验结果

表情	测试样本数	正确识别数			平均识别率(%)
		第 1 组	第 2 组	第 3 组	
高兴	8	6	6	6	75.00
悲伤	8	5	6	6	70.83
生气	8	6	5	6	70.83
惊讶	8	7	6	6	79.17
厌恶	8	6	6	6	75.00
恐惧	8	6	7	5	75.00
总计	48	36	36	35	74.31

4.3.2 不同算法的识别率对比 为了验证提出算法的有效性，本文做了大量对比实验。由表 5 可以看出，在 JAFFE 库和 CK 库上 DT-CWT+IDDP+融合算法的识别率最高，由表 6 可以看出，基于 DT-CWT+IDDP+融合特征的 DTW 规整后识别率最高，表 5、表 6 的实验结果表明了局部梯度 DT-CWT 特征具有较高的区分性，基于此特征的动态时间规整具有较好的效果。

同时本文在分别 CK+和 HFUT-FE 库上比较了几种表情识别算法的识别率，实验结果如表 7、表 8 所示。由表 7、表 8 可以看出，本文算法的识别率均高于其他算法，表明了该算法的有效性。

表 5 JAFFE 库和 CK 库上不同算法识别率比较

算法	识别率(%)	
	JAFFE	CK
Gabor+Radial Encoding	89.67	91.51
GLDPE	83.67	97.08
多特征多分类器	94.77	97.21
Shape+Texture	93.25	95.94
DT-CWT+DDP+融合	93.24	95.24
DT-CWT+LBP+融合	95.94	97.14
DT-CWT+IDDP+融合	96.85	98.41

表6 CK+库上基于不同特征的规整算法识别率比较

算法	识别率(%)
手动规整	86.11
DTW(LBP)	86.98
DTW(HOG)	87.22
DTW(DT-CWT+IDDP+融合)	90.00

表7 CK+库上不同算法的识别率比较

算法	识别率(%)
LBP +SRC ^[18]	79.97
LPQ +SRC ^[19]	80.78
Gabor+SRC ^[20]	82.82
LBP_TOP+SRC ^[6]	83.06
LPQ_TOP+SRC ^[7]	86.75
SPSD ^[8]	88.50
本文算法	90.00

表8 HFUT-FE 库上不同算法的识别率比较

算法	识别率(%)
LBP +SRC ^[18]	68.75
LPQ +SRC ^[19]	70.83
Gabor+SRC ^[20]	72.22
LBP_TOP+SRC ^[6]	71.53
LPQ_TOP+SRC ^[7]	73.61
SPSD ^[8]	73.61
本文算法	74.31

5 结束语

针对静态表情特征的不足, 本文提出了一种针对非特定人的动态表情识别方法: 基于 DT-CWT 和 AAM 的动态表情识别。大量实验表明, 该方法有一定的优越性:

(1) 目前大多数算法在进行动态表情识别时, 表情序列的规整都是手动的, 具有一定的人为性和不规范性, 影响表情识别效果。本文采用基于局部梯度 DT-CWT 主方向模式特征的 DTW 对表情序列进行规整, 使表情序列更加规范化。与基于其他特征的规整方法相比, 本文的规整方法取得了更好的效果;

(2) 采用 AAM 定位图像的特征点, AAM 集纹理信息和形状信息为一体, 可以准确的定位出图像的特征点并进行跟踪, 通过计算相邻两帧表情图像的关键点在时间轴上的位移, 获得了更加全面的动

态表情特征;

(3) 通过对中性脸几何模型的建立和匹配, 排除人脸模型与测试样本相差甚多的表情序列, 利用与测试样本相近的人脸模型所对应的表情序列进行分类识别, 有效克服了不同人脸模型呈现表情的差异性, 降低了人脸特征对表情识别的影响。

随着人脸表情识别技术的逐渐成熟, 针对非特定人的表情识别日益重要, 目前针对非特定人的表情识别方法的准确性虽然有所提高, 但若应用到实际产品中还有欠缺, 所以进一步提高识别的准确性仍然有待研究。

参考文献

- [1] MEHRABIAN A. Communication without words[J]. *Psychological Today*, 1968, 2: 53-55.
- [2] 於俊, 汪增福, 李睿. 一种同步人脸运动跟踪与表情识别算法[J]. *电子学报*, 2015, 43(2): 371-376. doi: 10.3969/j.issn.0372-2112.2015.02.025.
- [3] YU Jun, WANG Zengfu, and LI Rui. A simultaneous facial motion tracking and expression recognition algorithm[J]. *Acta Electronica Sinica*, 2015, 43(2): 371-376. doi: 10.3969/j.issn.0372-2112.2015.02.025.
- [4] JIA J, XU Y, ZHANG S, et al. The facial expression recognition method of random forest based on improved PCA extracting feature[C]. *IEEE International Conference on Signal Processing, Communications and Computing*, Hong Kong, China. 2016: 1-5. doi: 10.1109/ICSPCC.2016.7753643.
- [5] REN Fuji and HUANG Zhong. Facial expression recognition based on AAM-SIFT and adaptive regional weighting[J]. *IEEE Transactions on Electrical & Electronic Engineering*, 2015, 10(6): 713-722. doi: 10.1002/tee.22151.
- [6] LI Y, REN F, and HU M. Facial expression recognition based on DT-CWT and weighted inverse cloud model[C]. *Modelling, Simulation and Identification 841: Intelligent Systems and Control*, Campinas, Brazil. 2016: 380-388. doi: 10.2316/P.2016.841-010.
- [7] ZHAO Guoying and PIETIKAINEN Matti. Dynamic texture recognition using local binary patterns with an application to facial expressions[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2007, 29(6): 915-928.
- [8] JIANG Bihan, VALSTAR Member, MARTINEZ Brais, et al. A dynamic appearance descriptor approach to facial actions temporal modeling[J]. *IEEE Transactions on Cybernetics*, 2014, 44(2): 161-174. doi: 10.1109/tcyb.2013.2249063.
- [9] TAHERI Sima, QIU Qiang, and CHELLAPPA Rama. Structure-preserving sparse decomposition for facial expression analysis[J]. *IEEE Transactions on Image Processing*, 2014, 23(8): 3590-3603. doi: 10.1109/tip.2014.2331141.

- [9] FAN Xijian and TJAHHADI Tardi. A spatial-temporal framework based on histogram of gradients and optical flow for facial expression recognition in video sequences[J]. *Pattern Recognition*, 2015, 48(11): 3407–3416. doi: 10.1016/j.patcog.2015.04.025.
- [10] 施毅. 基于主动外观模型的人脸表情识别研究[D]. [硕士学位论文], 上海交通大学, 2012.
SHI Yi. Facial expression recognition based on active appearance models[D]. [Master dissertation], Shanghai Jiao Tong University, 2012.
- [11] CHEON Yeongjae and KIM Daijin. Natural facial expression recognition using differential-AAM and manifold learning[J]. *Pattern Recognition*, 2009, 42(7): 1340–1350. doi: 10.1016/j.patcog.2008.10.010.
- [12] LEE Yonghwan, HAN Wuri, KIM Youngseop, *et al.* Robust emotion recognition algorithm for ambiguous facial expression using optimized AAM and k-NN[J]. *International Journal of Security & Its Applications*, 2014, 8(5): 203–212. doi: 10.14257/ijisa.2014.8.5.19.
- [13] HIYADI H, ABABSA F, MONTAGNE C, *et al.* Combination of HMM and DTW for 3D Dynamic gesture recognition using depth only[C]. *Informatics in Control, Automation and Robotics 12th International Conference, ICINCO 2015*, Colmar, France, 2016: 229–245. doi: 10.1007/978-3-319-31898-1_13.
- [14] 傅颖, 郭晶云. 基于动态时间规整的人体动作识别方法[J]. *电子测量技术*, 2014, 37(3): 69–72. doi: 10.3969/j.issn.1002-7300.2014.03.018.
FU Ying and GUO Jingyun. Dynamic time warping-based human action recognition[J]. *Electronic Measurement Technology*, 2014, 37(3): 69–72. doi: 10.3969/j.issn.1002-7300.2014.03.018.
- [15] 宋怡焕, 饶秀勤, 应义斌. 基于 DT-CWT 和 LS-SVM 的苹果果梗/花萼和缺陷识别[J]. *农业工程学报*, 2012, 28(9): 114–118. doi: 10.3969/j.issn.1002-6819.2012.09.019.
SONG Yihuan, RAO Xiuqin, and YING Yibing. Apple stem/
- calyx and defect discrimination using DT-CWT and LS-SVM [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2012, 28(9): 114–118. doi: 10.3969/j.issn.1002-6819.2012.09.019.
- [16] 胡敏, 朱弘, 王晓华, 等. 基于梯度 Gabor 直方图特征的表情识别方法[J]. *计算机辅助设计与图形学学报*, 2013, 25(12): 1856–1861.
HU Min, ZHU Hong, WANG Xiaohua, *et al.* Expression recognition method based on gradient Gabor histogram features[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2013, 25(12): 1856–1861.
- [17] LIU Jun, JING Xiaojun, SUN Songlin, *et al.* Local gabor dominant direction pattern for face recognition[J]. *Chinese Journal of Electronics*, 2015, 24(2): 245–250. doi: 10.1049/cje.2015.04.004.
- [18] HUANG M W, WANG Z W, and YING Z L. A new method for facial expression recognition based on sparse representation plus LBP[C]. *IEEE International Congress on Image and Signal Processing*, Yantai, China, 2010: 1750–1754. doi: 10.1109/cisp.2010.5647898.
- [19] WANG Zhen and YING Zihu. Facial expression recognition based on local phase quantization and sparse representation [J]. *IEEE International Conference on Natural Computation*, 2012, 2: 222–225. doi: 10.1109/icnc.2012.6234551.
- [20] ZHANG Shiqing, ZHAO Xiaoming, and LEI Bicheng. Robust facial expression recognition via compressive sensing[J]. *Sensors*, 2012, 12(3): 3747–3761. doi: 10.3390/s120303747.
- 许良凤：女，1970 年生，副教授，研究方向为多媒体通信、图像处理等。
- 王家勇：男，1992 年生，硕士生，研究方向为模式识别、图像处理等。
- 崔婧楠：女，1996 年生，本科生，研究方向为高性能计算和大数据、图像处理等。
- 胡敏：女，1967 年生，教授，研究方向为计算机视觉、数字图像处理等。