



文章编号:1001-9081(2018)06-1771-06

DOI:10.11772/j.issn.1001-9081.2017122876

飞机牵引车语音识别的动态时间规整优化算法

解本铭¹, 韩明², 张攀¹, 张威^{1,3*}

(1. 中国民航大学 航空工程学院, 天津 300300; 2. 中国民航大学 电子信息与自动化学院, 天津 300300;

3. 中国民航航空地面特种设备研究基地, 天津 300300)

(* 通信作者电子邮箱 drwadecheung@gmail.com)

摘要:为研究飞机牵引车智能语音控制, 实现机场环境下牵引车对飞行员语音命令的精确、高效识别, 同时针对传统动态时间规整(DTW)算法计算量大、时间复杂度高、算法识别效率低的问题, 提出了一种车辆语音识别的六边形弯曲窗口约束DTW优化算法。首先, 从DTW算法原理、牵引车指令的语音特性和机场环境三方面, 分析了弯曲窗口对DTW算法识别精度、效率的影响; 然后, 在Itakura Parallelogram菱形弯曲窗口约束DTW优化算法的基础上, 进一步提出了六边形弯曲窗口约束的DTW全局优化算法; 最后, 通过改变优化系数, 实现了最优六边形弯曲窗口约束的DTW算法方案。基于孤立词识别的实验结果表明, 所提最优算法与传统DTW算法、菱形弯曲窗口约束的DTW算法相比, 识别错误率分别降低77.14%和69.27%, 识别效率分别提高48.92%和27.90%。该最优算法更具鲁棒性、时效性, 可以作为飞机牵引车智能控制的理想指令输入端口。

关键词:飞机牵引车; 语音识别; 动态时间归整; 弯曲窗口; 全局优化; 孤立词

中图分类号: TP391.42 文献标志码:A

Optimization algorithm of dynamic time warping for speech recognition of aircraft towing vehicle

XIE Benming¹, HAN Mingming², ZHANG Pan¹, ZHANG Wei^{1,3*}

(1. College of Aeronautical Engineering, Civil Aviation University of China, Tianjin 300300, China;

2. College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China;

3. Aviation Ground Special Equipment Research Base of Civil Aviation Administration of China, Tianjin 300300, China)

Abstract: In order to study the intelligent voice control of aircraft towing vehicle, realize accurate and efficient recognition of the voice command of pilot in the airport environment, and solve the problems of large computation, high time complexity and low recognition efficiency of the traditional Dynamic Time Warping (DTW) algorithm, a new optimization algorithm of DTW with constraint of hexagonal warping window for vehicle speech recognition was proposed. Firstly, the influence of warping window on the accuracy and efficiency of DTW algorithm was analyzed from three aspects such as the principles of DTW algorithm, the speech characteristics of towing vehicle instruction and the airport environment. Then, on the basis of DTW optimization algorithm with constraint of Itakura Parallelogram rhombic warping window, a DTW global optimization algorithm with the constraint of hexagonal warping window was further proposed. Finally, by varying the optimization coefficient, the optimal DTW algorithm with the constraint of hexagonal warping window was realized. The experimental results based on isolated-word recognition show that, compared with the traditional DTW algorithm and the DTW algorithm with rhombic warping window constraint, the recognition error rate of the proposed optimal algorithm is reduced by 77.14% and 69.27% respectively, and its recognition efficiency is increased by 48.92% and 27.90% respectively. The proposed optimal algorithm is more robust and timeliness, and can be used as an ideal instruction input port for intelligent control of aircraft towing vehicle.

Key words: aircraft towing vehicle; speech recognition; Dynamic Time Warping (DTW); warping window; global optimization; isolated-word

0 引言

目前, 语音识别技术已广泛应用于计算机、办公自动化、机器人、车辆等在内的各个领域, 这使得人机交互变得更加智

能有效。然而, 牵引车作为机场的特种车辆, 其牵引飞机的工作并没有实现语音控制, 而是由地面机组人员通过耳机接收飞行员传达的塔台管制员命令, 然后和牵引车驾驶员沟通、协调进行控制^[1]。

收稿日期:2017-12-08;修回日期:2018-02-06;录用日期:2018-02-08。

基金项目:国家自然科学基金委员会与中国民航局联合资助项目(U1533103);中央高校基本科研业务费资助项目(3122017025)。

作者简介:解本铭(1956—),男,辽宁彰武人,教授,硕士,主要研究方向:机电液一体化; 韩明(1989—),女,山东德州人,硕士研究生,主要研究方向:语音处理、模式识别; 张攀(1984—),男,湖北随州人,讲师,博士,主要研究方向:智能诊断、动态监测; 张威(1979—),男,湖南衡阳人,教授,博士,主要研究方向:机器学、机构学。



近年来,在民用航空器地面安全事故中,牵引过程发生的安全事故所占比例不容小觑,事故造成巨大的经济损失,其原因主要包括:牵引车驾驶员违反牵引车使用规则、反应不及时,以及驾驶员与地面工作人员联络不畅或者配合失误^[1]。

为了减小牵引车驾驶员人为失误造成的民用航空器地面事故率,同时参考语音识别技术在车辆自动驾驶领域的应用研究,本文拟在牵引车语音控制方面作些探究性工作。

本文就如何实现飞机牵引车智能语音控制,使牵引车及时、精确地识别飞行员语音指令,安全高效地完成飞机牵引工作进行了研究。牵引车指令属于驾驶舱指令,是参照陆空通话标准扩充的机载短语语音指令^[2],具有语速平稳、发音力度强化变更的特点,并且驾驶舱对机场噪声分离性较好,因此指令识别相对简单并且具有一定的抗噪性。尽管对于短语语音识别,目前已有许多高效算法,如隐马尔可夫模型(Hidden Markov Model, HMM)、矢量量化(Vector Quantization, VQ)、人工神经网络(Artificial Neural Network, ANN)技术等,但是以上算法需要对大量发音样本进行训练,相对复杂,计算量大,更适合用于大词汇量连续语音识别^[3]。相反,动态时间规整(Dynamic Time Warping, DTW)技术可以在样本数据较少的情况下进行学习训练,并能够获得较好的训练模型,更适合训练样本缺乏情况下的简单、孤立、短语语音识别^[4]。因此,本研究选用DTW算法对牵引车指令语音进行识别。但是,由于DTW算法是通过在整个矩形弯曲窗口内展开若干条路径搜索,寻找最佳匹配路径来识别语音的,时间复杂度和空间复杂度都相对较高,所以算法识别效率较低^[5]。

针对上述DTW算法识别效率低的问题,目前主要有两种改进方法。第一种是提前终止技术^[6]:该方法在累积弯曲代价超过一定门限时停止搜索,认为两个语音序列不匹配,该方法能有效降低算法的计算复杂度。第二种是弹性粗粒度动态弯曲时序相似性算法^[7]:其思想是通过数据处理,降低时序数据的维度,用低维特征代替原时间序列,从而大幅提高DTW计算效率。上述两种改进方法的缺点都是以损失算法的识别精度为代价来提高算法运行效率^[6-7]。

由于机场对安全性的高标准要求,上述现有DTW改进方法很难满足本研究对算法的高识别性能要求。为了提高算法识别性能,本文结合对DTW算法原理、牵引车指令声学特性和机场环境的分析,在Itakura Parallelogram菱形弯曲窗口DTW算法^[8]的基础上,进一步提出了六边形弯曲窗口DTW全局优化算法,然后通过孤立词算法识别性能对比试验,找到所提六边形弯曲窗口DTW算法的最优方案,并验证了该最优算法良好的时效性和鲁棒性。

1 系统总体设计及实现过程

本文所涉及的飞机牵引车智能语音识别系统是通过飞行员语音直接控制牵引车的一种智能系统^[9],如图1所示。该系统的工作过程为:飞行员的语音信号通过飞机耳机接口传输给牵引车上的语音接收设备,经过语音处理后直接控制牵引车的运动,同时通过外放设备可以实现与地面机组人员的通信。

语音识别就本质而言是一种模式匹配^[10],它是飞机牵引车智能语音识别系统中重要的环节之一。该语音识别流程如图2所示。其中,端点检测选用的是双门限检测法^[11],特征

提取则选用的是基于人耳听觉机理的梅尔频率倒谱系数(Mel Frequency Cepstrum Coefficients, MFCC)。飞行员发出的语音指令经该流程识别后输出给牵引车,从而实现飞行员对牵引车的语音控制。

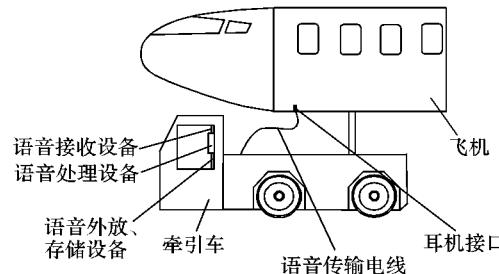


图1 飞机牵引车智能语音识别系统示意图

Fig. 1 Schematic diagram of intelligent speech recognition system of aircraft towing vehicle

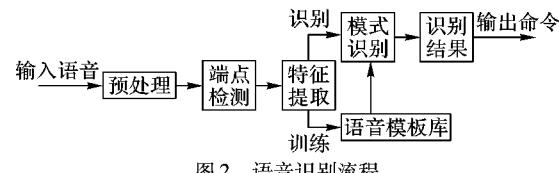


图2 语音识别流程

Fig. 2 Flow chart of speech recognition

2 DTW 识别算法及其优化

2.1 传统DTW算法的基本算法原理

DTW算法是基于动态规划思想,把距离测度计算和时间规整结合起来的一种非线性归整技术^[5]。

在语音识别过程中,语音经过训练生成参考模板,待测语音则经过相同的窗函数、相同的帧长和相同的帧移得到相同类型的语音特征参数,生成测试模板。测试模板与参考模板的匹配,就是寻求两者之间的声学相似度,相似度又可以通过两者之间的失真度来描述,失真度越小,相似度越高^[12-14]。

DTW算法通过寻找一个最优时间规整函数 $M = \phi(N)$,将语音测试模板的时间轴非线性映射到参考模板的时间轴上^[5]。假设参考模板共M帧语音特征矢量,记作 $\{\mathbf{R}(1), \mathbf{R}(2), \dots, \mathbf{R}(m), \dots, \mathbf{R}(M)\}$,其中m记作语音帧的时序标号,则 $\mathbf{R}(m)$ 表示参考模板的第m帧语音特征矢量;相应的假设测试模板有N帧特征矢量,记作 $\{\mathbf{T}(1), \mathbf{T}(2), \dots, \mathbf{T}(n), \dots, \mathbf{T}(N)\}$,其中n记作语音帧的时序标号,则 $\mathbf{T}(n)$ 表示测试模板的第n帧的特征矢量。

参考模板与测试模板匹配时,每帧特征矢量之间都有相应的失真度,假设m,n分别记作参考模板和测试模板的任意帧号,则两帧之间的失真度可表示为 $d\{\mathbf{T}(n), \mathbf{R}[\phi(n)]\}$,用D表示处于最优时间归整情况下测试模板和参考模板每一帧特征矢量匹配累积距离最小的距离矩阵,应满足式(1)^[12]:

$$D = \min \sum_{n=1}^N d\{\mathbf{T}(n), \mathbf{R}[\phi(n)]\} \quad (1)$$

此时测试模板与参考模板匹配度最高,语音失真度最小,该参考模板语音即为DTW最终识别结果^[14]。

2.2 DTW 算法优化

2.2.1 传统DTW算法存在的问题

DTW算法搜索路径^[15]如图3所示,横轴为测试模板时间轴的各个帧号 $n = 0 \sim N$,纵轴为参考模板时间轴的各个帧号 $m = 0 \sim M$,连接横、纵各个坐标轴就可以形成一个矩形网



络(即矩形弯曲窗口)。对于两个等长的语音时间序列,即 $M = N$,最佳搜索路径应该是弯曲窗口的对角线(如图4中细线条所示),此时两个时间序列是一一对应的,偏差为0,而实际搜索路径(如图4中粗线所示)和对角线有一定的偏离,对于不等长的时间序列,即 $M \neq N$,亦是如此,但是偏离不会太大;否则,当弯曲路径偏离对角线很远时,即使两个偏差很大的并不属于同一类的序列,也可能在整个传统矩形弯曲窗口内经DTW计算,得到很小的累积距离值,这将造成病态匹配^[8],从而产生误识别,降低算法的识别精度。

传统DTW算法是从 $(0,0)$ 出发,到 (N,M) 结束,在整个矩形弯曲窗口内展开若干条路径搜索,计算每条路径从 $(0,0)$ 到 (N,M) 时对应帧匹配的积累距离,累积距离最小者即为最佳匹配路径。该算法计算量大,时间复杂度较高,为 $O(MN)^{[16]}$,算法识别效率低。

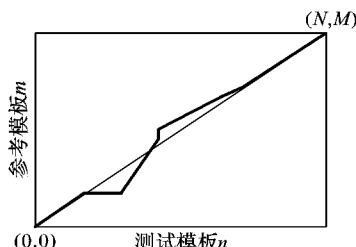


图3 DTW算法搜索路径

Fig. 3 Searching path of DTW algorithm

2.2.2 牵引车指令的语音特性分析

本研究飞行员对牵引车的语音指令属于驾驶舱指令,是参照陆空通话标准扩充的机载语音指令^[2],其基本要求包括用词简洁、准确、专业以及发音力度强化变更等,以此保证通话的时效性和准确度,从而提高工作效率,保证航行安全。其语音特性的具体要求将严格遵循民航陆空通话标准^[17]:

- 1) 机载指令语气易采用祈使句,通过省略主语使指令更加简练。
- 2) 指令应语义干脆、明确,指令词汇应避免使用单独的汉字即英文字母,尽量采用平均长度保持在2~4个字符的简单孤立词汇。
- 3) 指令语速平稳,应保持平均每秒2个字,以保证语音信息量和指令效率之间的平衡,从而保证指令的识别效率和识别准确率,使机载语音指令的控制效能得以较好发挥。
- 4) 指令发音力度应作强化变更,这在提高语音抗噪性的同时,保证了不同指令间较强的差异化程度。使得语音识别时指令间不易发生误识别,从而保证识别的准确率^[8]。

2.2.3 六边形弯曲窗口约束的DTW算法

由牵引车指令语速平稳特性可知,同一条指令的参考模板语音和测试模板语音发音时间轴长度相差不大,即两者时间轴分帧数 M,N 相差不大,根据DTW算法识别原理^[15],本研究语音指令的算法匹配路径沿算法弯曲窗口对角线偏差不大,所以适当减小算法弯曲窗口偏离对角线的区域,理论上将在保证识别准确率的同时减少算法计算量从而提高算法识别效率。

另外,用词简单、语义明确以及发音力度强化变更等语音特性,使得牵引车指令抗噪性强,同时不同指令间的差异化程度加强,根据DTW算法识别原理^[15],合理减小算法弯曲窗口偏离对角线的区域,理论上将有效避免指令间病态匹配现象,

从而提高算法识别正确率。

针对传统DTW算法存在的问题,并结合以上对牵引车语音指令声学特性与DTW算法弯曲窗口关系的分析,为了防止病态匹配、提高识别精度,同时为了减小时间复杂度、提高识别效率,本文对DTW算法的弯曲窗口进行了优化,即全局优化^[16];DTW路径的局部最优性搜索只在限制的弯曲窗口中进行。

目前,较为常用的全局优化方法是Itakura Parallelogram约束^[16,18],该方法是将DTW算法的矩形弯曲窗口优化为图4所示的菱形弯曲窗口。

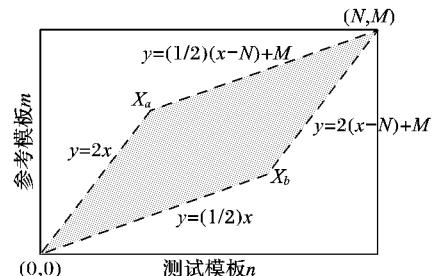


图4 Itakura Parallelogram 约束

Fig. 4 Constraint of Itakura Parallelogram

其中,菱形弯曲窗口斜率为 $1/2 \sim 2$,且两顶点分别为 $(0,0)$ 、 (N,M) ,由此可求得菱形弯曲窗口四条边的表达式分别为: $y = (1/2)x$, $y = 2x$, $y = (1/2)(x - N) + M$, $y = 2(x - N) + M$ 。由 $y = 2x$ 与 $y = (1/2)(x - N) + M$ 相交, $y = (1/2)x$ 与 $y = 2(x - N) + M$ 相交,分别求得菱形窗口另外两个顶点 $X_a = (2M - N)/3$, $X_b = (2/3)(2N - M)$ 。

在寻找DTW路径时,由于菱形弯曲窗口斜率的限制,语音识别只需要计算和存储菱形弯曲窗口内格点对应帧的匹配距离和累积距离^[18],相对于传统矩形弯曲窗口DTW,减少了算法的计算量。尽管如此,就本研究而言,由上文对牵引车语音指令声学特性与DTW算法弯曲窗口关系的分析可知,偏离菱形弯曲窗口对角线的两个顶角($\angle X_a$ 和 $\angle X_b$)范围内的帧匹配距离和累积距离,同菱形弯曲窗口范围外帧的匹配距离和累积距离,可能也没有必要计算和存储^[18]。基于此,本文在菱形弯曲窗口约束的DTW算法的基础上,进一步提出六边形弯曲窗口约束的DTW算法。

通过两条与菱形弯曲窗口对角线等斜率的直线 $y = (M/N)x + kb$ 和 $y = (M/N)x - kb$,分别去掉菱形弯曲窗口偏离对角线的两个顶角 $\angle X_a$ 和 $\angle X_b$,将其进一步优化为六边形弯曲窗口,如图5阴影部分所示。这时,待测模板的每一帧只需和参考模板在六边形弯曲窗口内的有限帧匹配,时间复杂度降低,同时有效避免病态匹配现象,提高算法识别精度^[8]。

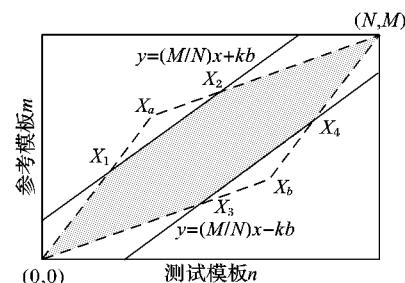


图5 六边形弯曲窗口

Fig. 5 Hexagonal warping window



设 k 为优化系数, $b = -(2M^2)/(3N) + (5/3)M - (2/3)N$ 为定值, 取 $k = 1$ 时, $y = (M/N)x - kb$ 过菱形窗口顶点 X_a , 在纵坐标轴上的截距。因为本文优化后的 DTW 弯曲窗口为六边形, 所以 $k \in (0, 1)$ 。 $X_1 = k(2M - N)/3, X_2 = k(2M - 4N)/3 + N, X_3 = 2k(2N - M)/3, X_4 = k(N - 2M)/3 + N$ 分别是 $y = (M/N)x + kb$ 和 $y = (M/N)x - kb$ 两条直线与菱形窗口的交点。

2.2.4 六边形弯曲窗口

随参考模板与测试模板时间序列长度 M, N 的不同、优化系数 k 的取值的不同, 优化后的 DTW 六边形弯曲窗口大小不同, 共有以下 7 种情况, 如图 6 所示。

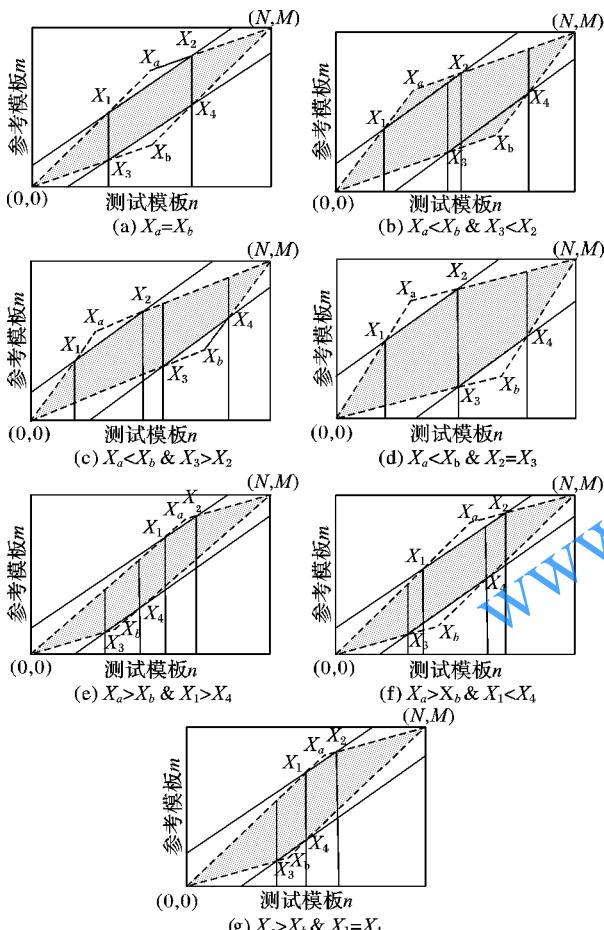


图 6 优化后的 DTW 六边形弯曲窗口

Fig. 6 Optimized DTW hexagonal warping window

因为以上 7 种六边形弯曲窗口的上下限各不相同, 所以待测模板每一帧与参考模板每一帧匹配时, DTW 算法的搜索范围和匹配路径各不相同。把每种情况下的六边形弯曲窗口按照各自上下限对应分成不同的小区域, 这样, DTW 匹配路径就能分段进行, 具体情况分析如下:

1) 当 $X_a = X_b$ 时, 计算可得 $X_1 = X_3, X_2 = X_4$, 匹配分为三段进行:

$$\begin{cases} (1/2)x \sim 2x, & 0 \leq x < X_1 \\ (M/N)x - kb \sim (M/N)x + kb, & X_1 \leq x < X_2 \\ 2(x - N) + M \sim (1/2)(x - N) + M, & X_2 \leq x < N \end{cases} \quad (2)$$

2) 当 $X_a < X_b$ 时, 计算可得 $X_1 < X_3, X_2 < X_4$, 当 $X_3 < X_2$

时, 匹配分为五段进行:

$$\begin{cases} (1/2)x \sim 2x, & 0 \leq x < X_1 \\ (1/2)x \sim (M/N)x + kb, & X_1 \leq x < X_3 \\ (M/N)x - kb \sim (M/N)x + kb, & X_3 \leq x < X_2 \\ (M/N)x - kb \sim (1/2)(x - N) + M, & X_2 \leq x < X_4 \\ 2(x - N) + M \sim (1/2)(x - N) + M, & X_4 \leq x < N \end{cases} \quad (3)$$

同理, 其他情况分析同上。

以上分段分析计算的方式让 7 种不同六边形弯曲窗口应用于电脑程序成为可能, 并为后续 Matlab 仿真实验中用六边形弯曲窗口 DTW 算法写语音识别程序建立了基础, 方便了后续对不同算法识别性能的研究。

3 实验结果与分析

本文共设计三组实验: 实验一通过改变优化系数, 改变六边形弯曲窗口约束的 DTW 算法, 并将其与传统 DTW 算法、菱形弯曲窗口约束的 DTW 算法进行识别精度对比, 找到识别精度最高的算法优化系数范围; 然后, 在该优化系数范围内, 对比三种算法的识别效率, 找到识别效率最高的优化算法, 该优化算法即为最优六边形弯曲窗口约束的 DTW 算法。实验二将实验一得到的最优算法, 与其他两种算法进行识别精度、效率对比, 进一步验证其优化性能。实验三则是在改变测试集样本大小的情况下, 进一步将最优算法与其他两种算法的识别精度、效率进行对比, 验证其鲁棒性和时效性。

3.1 语音库和测试集

根据本文 2.2.2 节对牵引车语音指令的语音特性分析, 定义了满足工况要求的牵引车语音基本指令集。

本实验语音信号处理方面主要考虑噪声影响。机场噪声主要来源于飞机发动机, 由于该机载语音指令控制的使用环境是飞机驾驶舱, 驾驶舱对机场环境噪声可分离性好^[14], 所以其噪声主要来源于本机发动机, 其他飞机发动机噪声对其影响不大。然而, 又因为在牵引车推拖飞机过程中, 飞机发动机并不工作, 所以, 本研究中本机发动机也不会对驾驶舱机载语音识别产生噪声影响。另外, 由于本研究中飞行员对牵引车的语音指令是通过插在飞机腹部的耳机传输线传到牵引车的(如图 1), 属于内话系统传输^[14], 所以, 尽管机场环境嘈杂, 在信号传输过程中, 机场噪声对语音识别的影响并不大。

以上分析可知, 本研究牵引车指令信号受环境噪声影响较小, 并且具有良好抗噪性, 鉴于此, 以下仿真实验在实验室低噪声环境下进行。

利用 Matlab 开发一个基于本研究内容的语音信号数据检测系统, 系统在内存为 4 GB 和 CPU 为 3 GHz 的 PC 上运行。利用 PC 的 Windows 录音设备, 其中, 采样率为 16 kHz, 采样位数为 16 bit, 录制了 10 个人(5 男 5 女)的语音, 内容包括所建牵引车语音基本指令集中“前进、后退、左转、右转、停止”这五条最常用指令, 并且严格按照牵引车指令词汇库设计原则, 要求所有人语速平稳, 保持每秒 2 字, 即每条语音指令用时 1 s, 并且发音力度相对于日常要作强化变更。每人每条语音均被采集 30 次, 即每条指令均被采集 300 次。以 12 维梅尔频率倒谱系数 MFCC 及 1 维标准化能量, 以及它们的一阶及二阶差分, 为语音特征参数^[19], 分别对每条语音的 200



次采样进行特征提取,建立该语音的参考模板。五条语音共1500条参考模板,组成本研究的孤立词指令语音库。

测试模板仍然由上述10个人录制,每人分别采集各条孤立词指令语音20次,即每条语音各采集200次,参照参考模板的建立,同样要求所有人语速平稳,保持每秒2字,同时发音作同等力度的强化变更。然后,分别提取与参考模板相同的特征参数建立测试模板,五条语音指令共1000条测试模板,组成本研究的测试集。

3.2 最优六边形弯曲窗口约束的DTW算法

将测试集里的每条测试模板与所建语音库里的参考模板进行匹配,分别统计传统DTW算法和菱形弯曲窗口约束的DTW算法的语音识别结果。不断改变六边形弯曲窗口约束的DTW算法的优化系数 k ,进行同样的实验,其中, k 从0到1每间隔0.05取值一次。由仿真实验结果可知,对于同一种DTW算法,各条语音指令的识别精度相差不大,所以,本文将同一种DTW算法计算下的各条语音指令的识别错误率进行了均值处理,最后将三种不同DTW算法均值处理后的识别错误率进行对比分析,如图7所示。

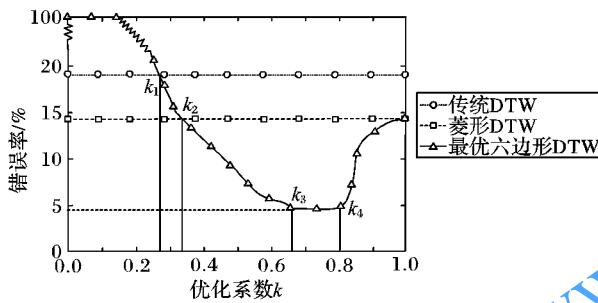


图7 不同优化系数时不同算法的识别错误率对比

Fig. 7 Comparison of recognition error rates of different algorithms under different values of optimization coefficient

由图7分析可知: $k \in (k_1, k_2)$ 时,六边形弯曲窗口约束的DTW算法,相对于传统DTW算法,识别精度较低; $k \in (k_1, k_2)$ 时,六边形弯曲窗口约束的DTW算法虽然相对于传统DTW算法识别精度有所提高,但是,相对于菱形弯曲窗口约束的DTW算法识别精度并无提高; $k \in (k_2, 1)$ 时,六边形弯曲窗口约束的DTW算法相对于传统DTW算法和菱形弯曲窗口约束的DTW算法,识别精度均提高; $k \in (k_3, k_4)$ 时,六边形弯曲窗口约束的DTW算法识别精度最高(其中 $k_1 \approx 0.25$, $k_2 \approx 0.3$, $k_3 \approx 0.65$, $k_4 \approx 0.81$)。

本研究重在提高算法识别精度,所以根据图7算法精度对比实验结论,不同DTW算法识别效率的对比实验是在 $k \in (k_3, k_4)$ 保证优化算法识别精度最高的前提下进一步进行的,并用Matlab中的tic和toc函数,分别对三种DTW算法的识别时间进行了统计,同样对其结果进行均值处理,如图8所示。

由图8分析可知,在优化算法识别精度最高的 $k \in (k_3, k_4)$ 取值范围内,六边形弯曲窗口约束的DTW算法,相对于其他两种算法,其识别效率均有提高;并且, k 值越小,算法识别效率越高, $k = 0.65$ 时,算法识别效率最高,此时六边形弯曲窗口约束的DTW算法即为本研究的最优算法。

3.3 最优算法与其他算法的识别错误率、效率对比

在实验一的基础上,进一步将 $k = 0.65$ 时最优六边形弯曲窗口约束的DTW算法与其他两种DTW算法对各条孤立

词指令语音的识别错误率、识别效率进行对比,结果如表1~表2所示。

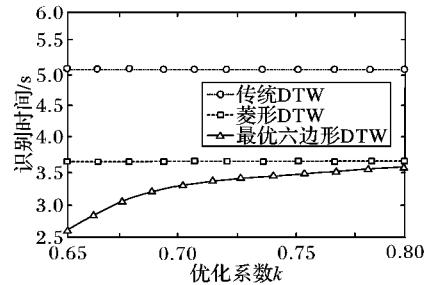


图8 不同优化系数时不同算法的识别时间对比

Fig. 8 Comparison of recognition time of different algorithms under different values of optimization coefficient

表1 $k = 0.65$ 时不同算法识别错误率对比

识别方法	孤立词					平均
	前进	后退	左转	右转	停止	
传统DTW算法	16.87	17.24	20.54	21.36	16.09	18.42
菱形弯曲窗口约束的DTW算法	12.16	11.98	15.26	16.09	13.01	13.70
最优六边形弯曲窗口约束的DTW算法	3.56	3.17	5.21	5.34	3.77	4.21

表2 $k = 0.65$ 时不同算法识别时间对比

Tab. 2 Comparison of recognition time of different algorithms with $k = 0.65$

识别方法	孤立词					平均
	前进	后退	左转	右转	停止	
传统DTW算法	5.01	4.93	5.75	5.24	4.62	5.11
菱形弯曲窗口约束的DTW算法	3.31	3.12	3.92	4.10	3.65	3.62
最优六边形弯曲窗口约束的DTW算法	2.83	2.50	2.91	2.73	2.08	2.61

由表1、表2分析可知,所提最优算法相比传统DTW算法在识别错误率减小了77.14%的同时识别效率提高了48.92%;相比菱形弯曲窗口约束的DTW算法在算法识别错误率减小了69.27%的同时识别效率提高了27.90%。

3.4 不同测试集下三种算法的识别性能对比

考虑到测试集大小可能会影响算法的识别性能,为了更好地考察不同算法的识别性能,本实验在不同测试集下进一步对比了三种算法在识别效率和识别正确率上的结果。本实验建立了测试模板数量为100条至1000条,以100条为间隔的大小不等的十种测试集。其中100条测试模板的测试集是由上述实验中的10个人分别录制前进、后退、左转、右转、停止语音指令各2次,200条测试模板的测试集则是由上述试验中的10个人分别录制各条语音指令4次,以此类推建立其他测试集。

分别将不同测试集里的测试模板与语音库里的参考模板进行匹配,对不同算法的识别性能进行对比,结果如图9所示。

由图9(a)可以看出,随着测试集增大,传统DTW算法的识别效率略有降低,而其他两种算法的识别效率均有提高,但最优六边形弯曲窗口约束的DTW算法相对于菱形弯曲窗口



约束的 DTW 算法,识别效率提高 20%~30%,并且该最优算法的识别效率呈线性增长。

与此同时,由图 9(b)可以看出,随着测试集的增大,传统 DTW 算法的识别错误率逐渐增大,对于大数据时代,这种错误率的增大是不允许的,相比之下,菱形弯曲窗口约束的 DTW 算法较稳定,其识别错误率随测试集增大没有明显变化。与以上两种算法不同,随测试集的增大,最优六边形弯曲窗口约束 DTW 算法的识别错误率明显降低,并且几乎成线性降低,这也体现了该最优算法对于大测试集具有较好的适应性。

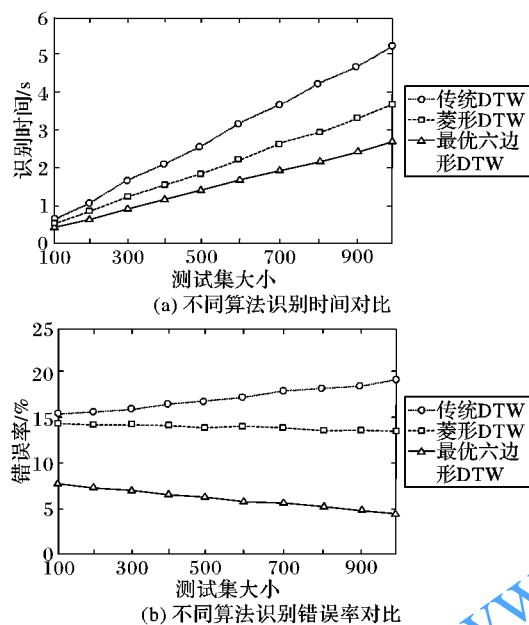


Fig. 9 Comparison of recognition performance of different algorithms under different sizes of test datasets

上述不同测试集下三种算法的性能对比实验进一步证明,本文所提最优六边形弯曲窗口约束的 DTW 算法相对于其他两种算法,尤其在大数据测试集情况下,其识别效率和识别正确率均有较好的表现。

4 结语

DTW 是近年来针对简单、孤立词汇识别的重要方法。识别过程中,DTW 会在整个算法弯曲窗口内展开若干条路径匹配,计算量大,然而,其最佳匹配路径是沿 DTW 弯曲窗口对角线进行,且偏差不大,否则容易出现病态匹配,造成误识别。由此可见,弯曲窗口的大小可能会影响 DTW 的识别性能。

由于机场对安全性的高标准要求,本研究对牵引车识别飞行员指令的准确性和时效性都提出了较高的要求。本文通过研究弯曲窗口对算法识别性能的影响,提出六边形弯曲窗口 DTW 优化算法,并通过孤立词算法性能对比实验,进一步找到最优六边形弯曲窗口 DTW 算法。实验结果表明,所提最优算法相比传统 DTW 算法和菱形弯曲窗口 DTW 算法,时间效率和识别正确率均有提高,更具时效性和鲁棒性。

由于本文算法只测试了牵引车基本指令的中文语音版本,下一步将针对牵引工况下更多功能及英语语音的应用进行研究。

参考文献 (References)

- [1] AIRPLANES B C. Statistical summary of commercial jet airplane accidents [EB/OL]. [2017-10-26]. http://www.boeing.com/resources/boeingdotcom/company/about_bca/pdf/statsum.pdf#search=%22Biman%20Bangladesh%20Airlines%20DC-10%20crash%22.
- [2] 须明,王博,王凯.驾驶舱语音控制指令的设计问题讨论[J].航空电子技术,2012,43(3):39~43.(XU M, WANG B, WANG K. Discussion of cockpit voice instruction design [J]. Avionics Technology, 2012, 43(3): 39~43.)
- [3] 朱国腾,孙伟.基于模板匹配的快速语音关键词检出方法[J].计算机应用,2013,33(11):3138~3140.(ZHU G T, SUN W. Rapid speech keyword spotting method based on template matching [J]. Journal of Computer Applications, 2013, 33(11): 3138~3140.)
- [4] GAGLIO S, RE G L, MORANA M. Human activity recognition process using 3-D posture data [J]. IEEE Transactions on Human-Machine Systems, 2015, 45(5): 586~597.
- [5] LI S C, REN H P. An isolated word recognition system based on DSP and improved dynamic time warping algorithm [C]// Proceedings of the 2010 IEEE International Conference on Progress in Informatics and Computing. Piscataway, NJ: IEEE, 2010: 136~139.
- [6] 陈胜利,李俊奎,刘小东.基于提前终止的加速时间序列弯曲算法[J].计算机应用,2010,30(4):1068~1071.(CHEN S L, LI J K, LIU X D. Early abandon to accelerating exactly warping matching of time series [J]. Journal of Computer Applications, 2010, 30(4): 1068~1071.)
- [7] CAI Q L, CHEN L, SUN J L. Piecewise statistic approximation based similarity measure for time series [J]. Knowledge-Based Systems, 2015, 85: 181~195.
- [8] 陈乾,胡谷雨.一种新的 DTW 最佳弯曲窗口学习方法[J].计算机科学,2012,39(8):191~195.(CHEN Q, HU G Y. New learning method for optimal warping window of DTW [J]. Computer Science, 2012, 39(8): 191~195.)
- [9] 张威,吴志恒,王立文.一种飞机牵引车智能语音识别系统:中国, CN104803007A [P]. 2015-07-29. (ZHANG W, WU Z H, WANG L W. An intelligent speech recognition system for aircraft tractor: China, CN104803007A [P]. 2015-07-29.)
- [10] 李瑞峰,王亮亮,王珂.人体动作行为识别研究综述[J].模式识别与人工智能,2014,27(1):35~48.(LI R L, WANG L L, WANG K. A survey of human body action recognition [J]. Pattern Recognition and Artificial Intelligence, 2014, 27(1): 35~48.)
- [11] 张备伟,吴琦,刘光徽.基于 DTW 的交警指挥手势识别方法[J].计算机应用研究,2017,34(11):3494~3499.(ZHANG B W, WU Q, LIU G H. Method for recognizing gesture of traffic police based on DTW algorithm [J]. Application Research of Computers, 2017, 34(11): 3494~3499.)
- [12] HAN S, CHEN D J, XIONG M, et al. Schedulability analysis of deferrable scheduling algorithms for maintaining real-time data freshness [J]. IEEE Transactions on Computers, 2014, 63(4): 979~994.
- [13] 陈明威,孙丽华,徐健锋.弹性粗粒度动态弯曲时序相似性算法[J].计算机应用,2016,36(6):1639~1644.(CHEN M W, SUN L H, XU J F. Temporal similarity algorithm of coarse-granularity based dynamic time warping [J]. Journal of Computer Applications, 2016, 36(6): 1639~1644.)

(下转第 1789 页)



参考文献 (References)

- [1] BERTALMIO M, SAPIRO G, CASELLES V, et al. Image inpainting [C] // Proceedings of the 2000 27th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM, 2000: 417 – 424.
- [2] RUDIN L I, OSHER S. Total variation based image restoration with free local constraints [C] // ICIP1994: Proceedings of the 1994 IEEE International Conference on Image Processing. Piscataway, NJ: IEEE, 1994: 31 – 35.
- [3] CHAN T F, SHEN J H. Non-texture inpainting by Curvature-Driven Diffusions (CDD) [J]. Journal of Visual Communication and Image Representation, 2001, 12(4): 436 – 449.
- [4] BAI J, FENG X C. Fractional-order anisotropic diffusion for image denoising [J]. IEEE Transactions on Image Processing, 2007, 16 (10): 2492 – 2502.
- [5] YOU Y L, KAVEH M. Four-order partial differential equations for noise removal [J]. IEEE Transactions on Image Processing, 2000, 9(10): 1723 – 1730.
- [6] GILBOA G, OSHER S. Nonlocal operators with applications to image processing [J]. SIAM Journal on Multiscale Modeling and Simulation, 2008, 7(3): 1005 – 1028.
- [7] 杨文波, 马天玮, 刘剑. 非局部变分修复法去除高密度椒盐噪声 [J]. 中国光学, 2013, 6(6): 876 – 884. (YANG W B, MA T W, LIU J. Elimination of impulse noise by non-local variation inpainting method [J]. Chinese Optics, 2013, 6(6): 876 – 884.)
- [8] ZHANG X D, FENG X C, WANG W W. Two-direction nonlocal model for image denoising [J]. IEEE Transactions on Image Processing, 2013, 22(1): 408 – 412.
- [9] CRIMINISI A, PEREZ P, TOYAMA K. Region filling and object removal by exemplar-based image inpainting [J]. IEEE Transactions on Image Processing, 2004, 13(9): 1200 – 1212.
- [10] 刘华明, 毕学慧, 叶中付, 等. 样本块搜索与优先权填充的弧形推进图像修复 [J]. 中国图象图形学报, 2016, 21(8): 993 – 1003. (LIU H M, BI X H, YE Z F, et al. Arc promoting inpainting using exemplar searching and priority filling [J]. Journal of Image and Graphics, 2016, 21(8): 993 – 1003.)
- [11] 林金勇, 邓德祥, 颜佳, 等. 基于自适应相似组稀疏表示的图像修复算法 [J]. 计算机应用, 2017, 37(4): 1168 – 1173. (LIN J Y, DENG D X, YAN J, et al. Self-adaptive group based sparse representation for image inpainting [J]. Journal of Computer Applications, 2017, 37(4): 1169 – 1173.)
- [12] 邹伟刚, 周志辉, 王洋. 基于非降采样轮廓波变换的图像修复算法 [J]. 计算机应用, 2017, 37(2): 553 – 558. (ZOU W G, ZHOU Z H, WANG Y. Image inpainting algorithm based on non-subsampled contourlet transform [J]. Journal of Computer Applications, 2017, 37(2): 553 – 558.)
- [13] PERONA P. Orientation diffusions [J]. IEEE Transactions on Image Processing, 1998, 7(3): 457 – 467.
- [14] CHAMBOLLE A. An algorithm for total variation minimization and applications [J]. Journal of Mathematical imaging and vision, 2004, 20(1/2): 89 – 97.
- [15] WANG Z , BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600 – 612.

This work is partially supported by the National Natural Science Foundation of China (61573012).

YANG Wenxia, born in 1978, Ph. D., associate professor. Her research interests include digital image processing, pattern recognition.

ZHANG Liang, born in 1977, Ph. D., professor. His research interests include controllability and stability of partial differential equation.

(上接第 1776 页)

- [14] GUERRA-CASANOVA J, SÁNCHEZ-ÁVILA C, de SANTOS SIERRA A, et al. Score optimization and template updating in a biometric technique for authentication in mobiles based on gestures [J]. Journal of Systems & Software, 2011, 84(11): 2013 – 2021.
- [15] BAILADOR G, SANCHEZ-AVILA C, GUERRA-CASANOVA J, et al. Analysis of pattern recognition techniques for in-air signature biometrics [J]. Pattern Recognition, 2011, 44(10/11): 2468 – 2478.
- [16] 刘贤梅, 赵丹, 郝爱民. 基于优化的 DTW 算法的人体运动数据检索 [J]. 模式识别与人工智能, 2012, 25(2): 352 – 360. (LIU X M, ZHAO D, HAO A M. Human motion data retrieval based on dynamic time warping optimization algorithm [J]. Pattern Recognition and Artificial Intelligence, 2012, 25(2): 352 – 360.)
- [17] 刘万凤, 胡军, 袁伟伟. 陆空通话标准用语(英语)的语音指令识别技术研究 [J]. 计算机科学, 2013, 40(7): 131 – 137. (LIU W F, HU J, YUAN W W. Research on technology of voice instruction recognition for air traffic control communication [J]. Computer Science, 2013, 40(7): 131 – 137.)
- [18] DUMPALA S H, ALLURI K N R K R, GANGASHETTY S V, et al. Analysis of constraints on segmental DTW for the task of query-by-example spoken term detection [C] // Proceedings of the 2015

Annual IEEE India Conference. Piscataway, NJ: IEEE, 2015: 1 – 6.

- [19] 尹浩, 宋成, 彭维平. 基于发音特征的声效相关鲁棒语音识别算法 [J]. 计算机应用, 2015, 35(1): 257 – 261. (CHAO H, SONG C, PENG W P. Robust speech recognition algorithm based on articulatory features for vocal effort variability [J]. Journal of Computer Applications, 2015, 35(1): 257 – 261.)

This work is partially supported by the National Natural Science Foundation of China and Civil Aviation Administration of China Jointly Funded Project (U1533103), the Fundamental Research Funding for the Central Universities (3122017025).

XIE Benming, born in 1956, M. S., professor. His research interest includes mechatronics and hydraulics integration.

HAN Mingming, born in 1989, M. S. candidate. Her research interests include speech processing, pattern recognition.

ZHANG Pan, born in 1984, Ph. D., lecturer. His research interests include intelligent diagnosis, dynamic monitoring.

ZHANG Wei, born in 1979, Ph. D., professor. His research interests include robotics, mechanisms.