

基于二级匹配策略的实时动态手语识别

梁文乐¹ 黄元元¹ 胡作进²

(南京航空航天大学计算机科学与技术学院 南京 210016)¹

(南京特殊教育师范学院数学与信息科学学院 南京 210038)²

摘要 动态手语可以利用其轨迹与关键手型加以描述。大量的统计实验数据表明,大多数的常用手语通过轨迹曲线的匹配即可实现识别,因此,提出一种针对动态手语的分级匹配识别算法。首先利用体感设备获取手势轨迹,并根据轨迹的点密度分布设计了一种关键帧检测算法以提取手势的关键手型,结合轨迹的曲线特征,实现对动态手语的精确描述。然后利用优化的动态时间规整(DTW)算法完成对手语的一级匹配,即轨迹匹配。若此时可以得到识别结果,那么识别过程可以结束,否则进入二级匹配,即针对关键手型再做匹配识别,从而得到最终的识别结果。实验证明,所提算法不仅实时性好,识别的准确率也较高。

关键词 动态手语识别,手语轨迹,关键帧,动态时间规整

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.07.054

Real-time Dynamic Sign Language Recognition Based on Hierarchical Matching Strategy

LIANG Wen-le¹ HUANG Yuan-yuan¹ HU Zuo-jin²

(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)¹

(School of Mathematics and Information Science, Nanjing Normal University of Special Education, Nanjing 210038, China)²

Abstract Dynamic sign language can be described by its trajectory and the key hand-action. However, a large number of statistical data show that most of the commonly used sign languages can be recognized by its trajectory curve. Therefore, a hierarchical matching recognition strategy for dynamic sign language was proposed in this paper. First, the gesture trajectory can be obtained by the somatosensory equipment like Kinect. According to its point density, an algorithm of key frame detection is designed and is used to extract the key gestures. Thus, we can achieve a precise description of dynamic sign language through trajectory curve and key frames. Then the dynamic time warping (DTW) algorithm is optimized and used to do the first-level matching, i. e. trajectory matching. If the recognition results can be get currently, the recognition process can be finished, otherwise the process should go into the second-level, i. e. key frame matching, and then get the final recognition results. Experiments show that this algorithm not only has better real-time performance, but also has higher recognition accuracy.

Keywords Dynamic sign language recognition, Gesture trajectory, Key frame, Dynamic time warping

1 引言

手势识别是人机交互技术中的一项关键技术,国内外已有众多学者对此课题从不同的角度、不同的层次进行了一定的研究。同时,手语也是聋人与正常人、聋人与聋人之间进行正常交流的主要方法。目前我国聋人数量达到2057万,约占全国人口总数的1.67%,但手语翻译人员的数量却远远难以满足市场需求,因此研究手语识别技术不仅可以让聋人与健听人之间进行无障碍交流,而且可以提高计算机的感知能力,增加人机交互的渠道与方法。

基于视觉的手势识别是一个富有挑战性的、多学科交叉的研究课题,是人机交互领域的一个前沿性课题和研究热点。与数据手套相比,计算机视觉具有交互方式更符合自然习惯、设备低廉易推广等优点。早期的单目摄像头由于无法对手部进行精确定位,因此只能采取对手部做标记,或者仅用于识别相对简单的静态手势^[1-2]。为了克服单目摄像头难以捕捉手在三维空间中的变化的缺点,研究者们尝试采用多摄像头采集不同维度的图像^[3-4]。但使用这种方法时,每次都需要对多个摄像头进行校准,操作非常不便。近年来,深度摄像机的出现使基于三维数据的手势识别研究得到了长足的发展。2010

到稿日期:2016-05-09 返修日期:2016-09-08 本文受国家自然科学基金项目(61375021),江苏省“双创”项目,扬州市“绿杨金凤”优秀博士项目资助。

梁文乐(1989—),男,硕士生,主要研究方向为模式识别、图像处理;黄元元(1975—),女,博士后,副教授,主要研究方向为多媒体技术、图像处理、模式识别,E-mail:805861040@qq.com;胡作进(1965—),男,博士,教授,主要研究方向为数据处理、机器学习。

年微软公司推出体感摄像头 Kinect,使得利用深度信息识别手语成为趋势。Jang 等提出了基于 Kinect 获取深度信息来识别手语的系统,使用连续自适应均值平移算法(Continuously Adaptive Mean Shift, CAMSHIFT)将深度概率和更新深度直方图用于手部跟踪^[6]。Chai 等利用 Kinect 获取手势三维特征,通过手的三维轨迹匹配实现识别,平均识别率达到 83.51%^[6]。Marin 使用 Kinect 定位手部区域,再使用 Leap Motion 得到手的精细信息,再将支持向量机(Support Vector Machine, SVM)作为分类器来识别手势,识别率达到 91.28%^[7]。目前通过使用 Kinect 获取深度信息来识别手语已经成为主流。

2 存在的问题

1)手语的描述。目前对动态手语的描述大多是通过手势的轨迹进行的。由于动态手语通过若干变化手势的组合进行表示,且人手是柔性物体,自由度大且手势变化多样,因此手势轨迹的不确定性与时空变化性是利用轨迹描述动态手语的难点问题。此外,除了手势的运动轨迹,手型的变化也包含了手语的重要语义信息。目前对手型的描述基本上是对手语中的每一帧图像检测其手型特征。但事实上并非所有的手型都对手语的语义有贡献,只有那些表示关键动作的关键手型才是构成手语语义的要素。因此,对动态手语更为精确的描述方法应该是手势轨迹加关键手型。

2)关键帧提取。关键帧是指手语中的关键动作所处的那一帧,而所谓的关键动作是组成手语语义的基本要素^[8]。由于长期以来习惯于以轨迹以及所有帧的手型来描述动态手语,因此如何在人机自然交互的情况下检测并提取关键帧的相关研究较少。

3)手语的匹配识别。目前对动态手语的识别都是通过轨迹曲线或者轨迹加所有手型进行一次性的分类完成的。首先,当前对曲线的相似性度量主要是解决曲线的平移、旋转以及等比例缩放问题。但手语轨迹所呈现出的曲线的情况要复杂得多,因为不同的手语者有不同的手语习惯,即使是同一个手语者,在不同时间做出的手语动作的轨迹也会不同,而这些不同并不是用平移或者等比例缩放就可以解决的。其次,通过大量的统计实验发现,大多数的手语轨迹曲线差别很大,而对于那些轨迹相似的手语,它们的关键手型一定会存在较大差异。因此,为了提高动态手语识别的效率与精度,可以考虑将轨迹与手型分开进行识别。

由上文可知,利用轨迹与关键手型可以更加精确而稳定地描述动态手语,在此基础上,设计合理的相似性度量算法可以实现对动态手语高效而准确的识别。

3 算法设计

本文提出了一种专门针对动态手语的分级匹配策略,即首先仅对手势的轨迹进行匹配,若可以得到结果,则识别过程结束;否则再对其关键手型进行匹配,从而得到识别结果。这种匹配策略能够实现的前提是要获取手势的轨迹曲线及其关键手型信息。因此本文提出了一种基于轨迹曲线点密度特征

的关键帧检测算法,并对传统的 DTW 算法做了优化,使之更加适用于手势轨迹的相似性度量。

3.1 手势轨迹获取

当前一般是通过采集手心的位置获取手势的轨迹,例如,文献[9]提取每一帧图像的人手中心点作为手势轨迹。Kinect 的帧率是 30 帧/秒。若手语动作“毕业证书”持续 2.13s,则会得到 64 帧手语图像以及相应的 64 个手部位置点,将这些点在时间轴上连起来形成手语的轨迹曲线,如图 1 所示(为了显示方便,将 x, y, z 三维数据分别显示),曲线上的每一个点(即轨迹的特征点)都具有 x, y, z 三维坐标。手语可以分为单手手语和双手手语,也可以认为单手手语是双手共同完成的(另一只手的位置始终保持不变即可),因此在手势的运动轨迹中,每一个特征点都可以看作是一个六维的特征向量,这六维的特征包含了两只手手心的 x, y, z 三维坐标。

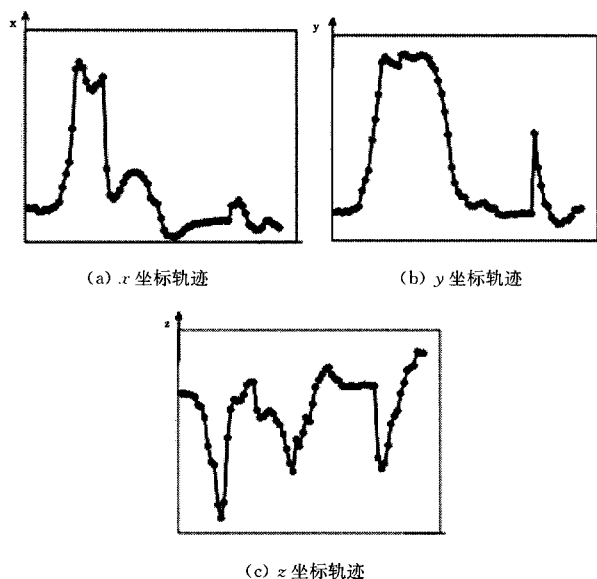


图1 手语“毕业证书”轨迹曲线(左手)

3.2 关键帧检测

对于手语的手型而言,对每一帧都提取手型特征是没有必要的。虽然一个手语单词的视频中会包含几十甚至一百多幅图像,但并不是每一帧手型都包含语义信息,只有那些为数不多的关键帧对手语语义才是有贡献的,而其余的绝大多数都是没有任何意义的过渡帧。这里定义关键帧中的手型为关键手型。从稳定性方面而言,不同的手语者对于相同手语的关键手型相对较稳定,而过渡帧中的手型则差别很大。因此,利用关键手型描述手语不仅可以减少数据量及提高手语识别效率,而且还可以提高对手语描述的稳定性与准确性。

通过大量的实验发现以下规律:一般在手语表达的过程中,为了明确手语语义,手语者会自然地强调手语中的关键动作,表现的方式为在关键动作处有一个时间的停留。这种停留只是相对过渡动作而言,并非刻意去做,是手语者做表达时的一个本能反应,这种本能反应体现在手语的轨迹曲线上,即在该关键动作时间附近的点会非常密集。基于该统计规律,本文提出了一种基于曲线点密度的关键帧检测算法。

假定有一个手语轨迹曲线 P ,若要定义其上某一点 X 的

密集度,则须首先获取在 P 上与 X 相邻的点的集合 T_X ,定义:

$$T_X = \{Y | \forall Y \in P, \delta(X, Y) \leq \Delta\} \quad (1)$$

其中, $\delta(X, Y)$ 表示曲线 P 上两个特征点 X 与 Y 的距离, Δ 为阈值。对于 T_X 而言,其中包含的点不仅在空间位置上相近,同时在时间上也必须是连续出现的。而式(1)的定义由于忽略了时间的连续性,在手语表达过程中,手部重复出现 T_X 的点密集度就会非常大,这与我们定义密集度的初衷不符,因此还需加入对时间的限制。假定由式(1)得到的 T_X 中包含 n 个点,按照时间顺序这 n 个点为 $T_X = \{Y_{t_1}, Y_{t_2}, \dots, Y_{t_m}\}$,其中点 Y 的下标 t_i 表示 Y 在时间序列中的标号,那么点 X 的密集度则定义为包含最多的时间标号连续的 Y_{t_i} 的 T_X 的子集:

$$\text{crow}(X) = \max\{|Y_{t_i}| | t_i \text{ 是连续的}, |Y_{t_i}| \subseteq T_X\} \quad (2)$$

对于一个特定手语而言,由于它是由左右双手共同完成的,因此需要分别计算左手和右手运动轨迹中所有点的密集度,并将它们相加得到该手语的总的点密集度。图 2(a) 示出手语“毕业证书”的点密集度曲线,其中横坐标表示采集到的帧数,为 64 帧;纵坐标表示每一帧中的手心位置点的密集度。理论上,在该曲线上密集度大于一定阈值的点应该对应关键帧。阈值的选取可以有以下几种方式:1)所有点密集度的平均值;2)点密集度的中位数;3)点密集度最大值与最小值的均值。由图 2(a) 可以发现,很多点的密集度都超过阈值,但实际上这些拥有较大密集度的点并非全部对应关键帧。因此可以考虑将图 2(a) 所示的密集度曲线划分为若干区间,然后在每个区间中寻找大于阈值的最大密集度点,以这些点对应的帧作为关键帧的候选帧。那么如何对密集度曲线进行划分呢?由于每一幅关键帧对手语的语义均有贡献,为了保证不漏检关键帧,选择将密集度曲线划分成的区间数为关键帧可能的最大数量的 1.5~2 倍,且对于绝大多数的手语,其关键帧数量不超过 5。为了显示方便,如图 2(b) 所示,将密集度曲线均匀划分为 7 个区间,得到了 6 个密集度极值点,从而得到对应的 6 幅候选关键帧,如图 3 所示。

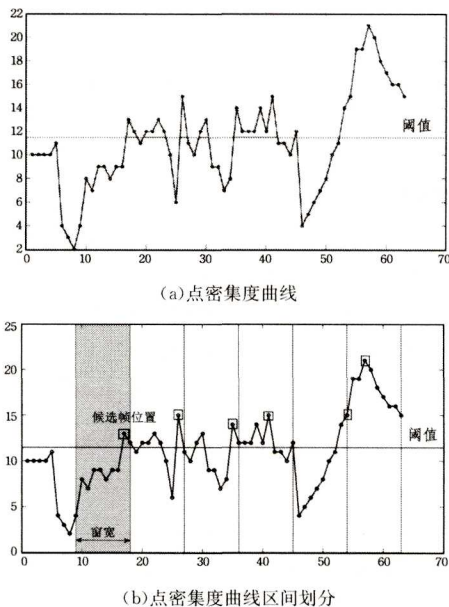


图2 “毕业证书”点密集度曲线



图3 “毕业证书”候选关键帧

由于划分的区间个数超过了实际的关键帧数量,因此得到的候选关键帧数量比实际数量更多。而连续的关键帧在手型上肯定是不相似的,根据这一特点,利用帧消减法去除多余的候选帧,得到最终的关键帧为图 3 中的候选关键帧 1、候选关键帧 2、候选关键帧 4,这与实际情况也正好相符。图 3 中的最后两帧属于终止手势,并不包含语义,但是由于拍摄的原因,终止处的点密集度也会比较大,因此它会入选候选关键帧的序列中。但由于一般的手语动作均在人腰部以上的位置完成,因此可以根据这个先验知识将终止手势排除在外。

3.3 轨迹曲线的归一化处理

不同的手语者有着不同的手语习惯,相同手语被不同手语者表达出来时其持续时间、手势快慢以及动作幅度等都存在显著差异,这就使得不同手语者的轨迹曲线在大体趋势上相似,但在局部细节上差别明显,从而给轨迹的相似性度量带来诸多问题。因此有必要对轨迹曲线进行归一化处理,尽量消除这些差异。假定有两条轨迹曲线 P 和 Q ,以 P 为模板,若要计算 Q 与 P 的相似度,则首先将 Q 按照 P 来做归一化。若 P 中包含 m 个特征点, Q 中包含 n 个特征点,即 $P = \{p_1, p_2, \dots, p_m\}$, $Q = \{q_1, q_2, \dots, q_n\}$,归一化的步骤如下。

1) 首先计算尺度缩放因子 ρ :

$$\rho = \frac{\left\| \sum_{i=1}^m \frac{p_i}{m} \right\|}{\left\| \sum_{j=1}^n \frac{q_j}{n} \right\|} = \frac{n \left\| \sum_{i=1}^m p_i \right\|}{m \left\| \sum_{j=1}^n q_j \right\|} \quad (3)$$

2) 曲线 Q 中的点的下标索引 i ($1 \leq i \leq n$) 经过归一化后应改为 $j = i * \frac{m}{n}$ ($1 \leq j \leq m$)。若 q_j 已经存在,则将 q_j 按照式(4)进行赋值;若 q_j 并不存在,则按照式(5)进行赋值。

$$q_j = \rho * \frac{(q_i - q_1) + p_1 + q_i}{2} \quad (4)$$

$$q_j = \rho * (q_i - q_1) + p_1 \quad (5)$$

3)归一化处理后的索引 j 由于四舍五入取整的关系,可能并不连续,因此对经过步骤2)还未赋值的 Q 集合中的点按照式(6)进行插值运算处理。

$$q_t = \frac{q_{t-1} + q_{t+1}}{2}, 1 < t < m \quad (6)$$

由此可以得到针对模板 P 归一化后的曲线 Q 。此时,它们起点相同,包含的特征点个数也相同,以便进行相似性度量。

3.4 基于轨迹的一级匹配

手语轨迹是一个时间序列,目前对时间序列的分类识别的常用方法是隐马尔科夫模型(Hidden Markov Model, HMM)和动态时间规整算法(Dynamic Time Warping, DTW)。文献[10-11]对DTW和HMM算法在手势识别方面的效果进行了对比,进而推荐在动态手语识别中使用DTW算法。但传统的DTW算法并没有考虑到关键帧对手语语义的重要贡献,而是同等对待所有的帧,这显然并不符合手语的特点。因此,本文提出了一种基于关键帧加权的DTW算法,用以提高动态手语识别的准确性。

假设有两条轨迹曲线 P 和 Q ,经过归一化处理后均包含 M 个特征点, P 中有 m 个关键帧,其序列号集合为 $KP = \{kp_1, kp_2, \dots, kp_m\}$; Q 中有 n 个关键帧,序列号集合为 $KQ = \{kq_1, kq_2, \dots, kq_n\}$ 。在曲线 P 上,点 p_i 到某一幅关键帧对应点的距离为 $\delta_{KP}(i)$,且 $\delta_{KP}(i) = \min\{|i-e|, e \in KP\}$;同样地, Q 中的点 q_j 到其某一关键帧对应点的距离为 $\delta_{KQ}(j) = \min\{|j-e|, e \in KQ\}$,那么使用式(7)作为累计代价矩阵的递推公式。

$$D(i, j) = (|\delta_{KP}(i) - \delta_{KQ}(j)| + 1) \times \delta(i, j) + \min \begin{cases} D(i-1, j) \\ D(i-1, j-1) \\ D(i, j-1) \end{cases} \quad (7)$$

其中, $\delta(i, j)$ 表示 p_i 与 q_j 之间的距离,一般可以用欧氏距离来表示。对于一条匹配路径 $W = \{w_1, w_2, \dots, w_l\}$ 而言,若 $w_k = (i, j)$, $1 \leq k \leq l$,则意味着在这条匹配路径上 p_i 与 q_j 相匹配,比较它们到各自曲线上关键帧点的距离,若 $\delta_{KP}(i) = \delta_{KQ}(j)$,则分两种情况:1) p_i 与 q_j 从相同方向接近各自曲线的关键帧,它们的匹配权值更高,对应的系数会较小;2)这两个点分别从不同方向接近各自曲线的关键帧,虽然它们的系数小,但在这条匹配路径上其他匹配点的系数会增大。由此可见,两种情况均可体现出关键帧的约束作用。由于采用归一化算法处理轨迹曲线能保证相互匹配的两条曲线长度一致,因此相同手语的曲线中关键帧所在位置是比较接近或者相同的,对于这样的两条曲线来说,它们之间的匹配距离更短,更容易被识别,反之亦然。

根据上述关键帧加权的优化DTW算法,可以将待识别的手语轨迹曲线与模板曲线一一匹配,计算距离,找到最小距离 $d_{\min 1}$ 与次小距离 $d_{\min 2}$,若 $d_{\min 2}/d_{\min 1} > \alpha$,则可以得到分类结果,即最小距离对应的类别;否则选择距离最小的 k 个类别,进入基于关键手型的二级匹配。参数 α 的取值对分类进程有

很大影响,一般在 $[1.2, 2]$ 之间取值。 α 取值越小,说明一级匹配的结果的可信度越高。

3.5 基于关键手型的二级匹配

有些手语的轨迹非常相似,例如“妻子”与“丈夫”、“儿子”与“女儿”等,对于这些手语,通过轨迹很难做出准确识别,因此有必要对其关键手型做进一步的分析比对。在关键帧中,借助深度与肤色信息,可以将图3中的手部区域分割出来,再对其做特征提取。假定对每一帧关键手型提取 N 维特征,那么关键手型可以表示成一个 N 维的特征向量 H 。

假设现有一个待识别的手语 S ,其包含 m 个关键手型,对应 m 个 N 维的特征向量,即 $S = \{H_{S1}, H_{S2}, \dots, H_{Sm}\}$ 。经过一级轨迹匹配,得到与 S 在轨迹上最相似的 k 个手语类别 $C_1, C_2, \dots, C_k$ 。以 C_1 为例,如何计算 S 与 C_1 在关键手型上的相似度呢?假定 C_1 包含 n 个关键帧,即 $C_1 = \{H_{C1}, H_{C2}, \dots, H_{Cn}\}$ 。理论上,若 S 与 C_1 相匹配,那么 m 应该与 n 相等。但在实际计算中,难以保证 m 等于 n 。根据前文所述的提取关键帧的算法,一般 $m > n$ 。构造一个 $m \times n$ 的矩阵 D ,其中第 i 行第 j 列的元素 d_{ij} 为 H_{Si} 与 H_{Cj} 之间的距离。可以将 S 与 C_1 看作是由关键帧构成的时间序列,那么同样利用DTW算法可以在矩阵 D 中找到一条匹配路径 $L = \{(x_1, 1), (x_2, 2), \dots, (x_n, n)\}$,其中 $1 \leq x_1 < x_2 < \dots < x_n \leq m$,使得这条路径中的 $d_{ij} ((i, j) \in L)$ 之和最小。那么 S 与 C_1 之间的距离定义为:

$$Dis(S, C_1) = \frac{\min\{\sum_{(i,j) \in L} d_{ij}\}}{\min(m, n)} \quad (8)$$

该值越小表明 S 与 C_1 越相似,反之亦然。用同样的方法可以分别计算 S 与其他几个类别的距离,若最小距离小于事先设定的拒识阈值,那么识别结果即为最小距离对应的类别,否则拒绝识别。

4 实验分析

实验分为两个部分:对关键帧检测算法的验证和对动态手语进行识别的验证。我们对5名手语者做了简单培训,教会他们60个常用手语词汇,并以其中一名男生和一名女生的手语轨迹和关键手型作为模板,然后利用其他3人的手语数据作为测试对象。

4.1 关键帧检测实验

为了验证本文提出的基于轨迹曲线点密度特征的关键帧检测算法的有效性,利用文献[12]中的聚类算法进行关键帧提取,用以比较。表1的实验数据为两种算法漏检率为0的情况下检测到的关键帧数量与运行时间。

表1 关键帧检测实验数据

	基于轨迹曲线点密度特征的关键帧检测算法	无监督聚类算法
检测到的关键帧平均数量	3.4	26
平均运行时间/s	0.054	12.978

从检测到的关键帧数量以及运行时间来看,本文算法无疑具有很大优势。文献[12]中的聚类算法并没有考虑到动态手语的特点,仅仅是利用每一帧手型的特征来进行聚类,计算

量很大;为了保证不漏帧,聚类的类别数也相对较多。而本文算法从动态手语的动作特点与手语者的心理意识出发,利用轨迹曲线上点的密集度大小来区分关键帧与非关键帧,计算量很小,并且准确率高。此外,由于关键帧的提取依赖于轨迹点的密集度,因此当手语时间变长、轨迹点数量增加时,本文方法检测关键帧的效率并不会下降;而随着数据量的增加,无监督的聚类算法将会耗费更多的时间,效率会明显下降。

4.2 手语识别实验

为了验证分级的匹配策略的有效性以及基于关键帧加权的 DTW 优化算法的有效性,做了 3 组对比实验:

第一组 用轨迹以及全部帧的手型描述动态手语,并采用传统的 DTW 算法进行分类识别;

第二组 用轨迹和关键手型描述动态手语,采用本文提出的分级的匹配策略,并利用传统的 DTW 算法做轨迹的匹配;

第三组 用轨迹和关键手型描述动态手语,利用基于关键帧加权的优化 DTW 算法做轨迹的匹配。

实验中,若一级匹配无法得到识别结果,则返回与待识手语轨迹最相似的前 5 个手语进入二级匹配。实验结果如表 2 所列,由实验数据可以看出,对于第一组实验,无论是在识别的效率还是识别的准确性上,其都是最低的。虽然只用了一级匹配,但是由于采用了全部帧的手型特征,数据量大,而且未区分关键帧与过渡帧对语义贡献的不同,因此识别效果并不理想。第二组与第三组实验虽然采用二级匹配,但是每一级的匹配数据量都不大,例如一级匹配仅对轨迹进行识别,特征的维数低;二级匹配中,关键帧的数量又不多,因此后两组实验的匹配效率是很高的,实时性好。此外,第二组与第三组实验的区别在于用于轨迹匹配的 DTW 算法与优化的 DTW 算法。由于充分考虑了动态手语的特点,突出了关键帧的作用,因此第三组实验在识别效率与识别准确率上均有所提高。此外,由于关键帧的稳定性,本文算法除了能够对上述经过培训的手语者的手语作出准确识别外,对一些现场学习手语的非特定人群也可以达到 80% 以上的识别率。需要指出的是,对于非特定人群,识别准确率并不稳定,动作的规范性在很大程度上决定了识别的准确率。

表 2 手语识别对比实验

	识别时间/s	识别准确率
第一组	4.200	0.7833
第二组	0.503	0.8667
第三组	0.396	0.9672

结束语 本文提出了一种对动态手语进行识别的新思路,即利用轨迹和关键手型描述动态手语,识别时将轨迹与手型分开,采用一种分级的匹配策略。由于充分考虑了动态手语自身的特点,无论是关键帧检测算法还是基于关键帧加权的优化 DTW 算法都取得了较好的实验效果。在对手语词汇

进行识别的基础上,今后可以进一步对更为复杂的手语语句进行识别。

参 考 文 献

[1] STARNER T,PENTLAND A. Real-time american sign language recognition from video using hidden markov models[M]// Motion-Based Recognition. Springer Netherlands, 1997: 227-243.

[2] MARAQA M,AL-ZBOUN F,DHYABAT M,et al. Recognition of Arabic Sign Language(ArSL) Using Recurrent Neural Networks[C]// Applications of Digital Information and Web Technologies,2008(ICADIWT 2008). IEEE,2008:41-52.

[3] ARGYROS A,LOURAKIS M I A. Binocular hand tracking and reconstruction based on 2D shape matching[C]//18th International Conference on Pattern Recognition,2006(ICPR 2006). IEEE,2006:207-210.

[4] VOGLER C,METAXAS D. Parallel hidden markov models for american sign language recognition [C] // IEEE International Conference on Computer Vision,1999. IEEE,1999:116-122.

[5] JANG Y. Gesture recognition using depth-based hand tracking for contactless controller application[C]//2012 IEEE International Conference on Consumer Electronics (ICCE). 2012:297-298.

[6] CHAI X,LI G,LIN Y,et al. Sign language recognition and translation with kinect[C]//IEEE Conf. on AFGR. 2013:1-2.

[7] MARIN G,DOMINIO F,ZANUTTIGH P. Hand gesture recognition with jointly calibrated Leap Motion and depth sensor[J]. Multimedia Tools and Applications,2016,72(22):14991-15015.

[8] SHEN Y L. Analysis of Chinese sign language morpheme [J]. Chinese Journal of Special Education,1993(1):1-13. (in Chinese)

沈玉林. 中国手语语素分析[J]. 特殊教育研究,1993(1):1-13

[9] DOLIOTIS P,STEFAN A,MCMURROUGH C,et al. Comparing Gesture Recognition Accuracy Using Color and Depth Information[C]//Proceedings of the Fourth International Conference on Pervasive Technologies Related to Assistive Environments(PETR) Crete, Greece,2011:1-7.

[10] CARMONA J,CLIMENT J. A performance evaluation of hmm and dtw for gesture recognition[C]//Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. 2012:236-243.

[11] RAHEJA J L,MINHAS M,PRASHANTH D,et al. Robust gesture recognition using Kinect: A comparison between DTW and HMM[J]. Optik-International Journal for Light and Electron Optics,2015,126(11):1098-1104.

[12] ZHUANG Y T,RUI Y,HUANG T S,et al. Adaptive Key Frame Extraction Using Unsupervised Clustering[C]//Proc. of IEEE Int. Conf. on Image Processing. 1998:866-870.