

基于压缩感知与 SURF 特征的手语关键帧提取算法

王民, 李泽洋, 王纯, 石新源

西安建筑科技大学信息与控制工程学院, 陕西 西安 710055

摘要 针对实时、大词汇集、连续的手语视频高效准确地识别, 提出了一种基于压缩感知与加速稳健特征(SURF)的手语关键帧提取算法。利用压缩感知将手语视频降维成低维多尺度帧图像特征, 通过自适应阈值完成子镜头分割, 以处理大量的手语帧数据; 运用 SURF 特征点完成特征匹配, 绘制其间的相似度曲线进而提取关键帧。在前期预处理阶段, 采用基于 HSV 空间自适应颜色检测提取手势区域。实验验证, 由本文算法提取到的关键帧具有较高的准确性, 且算法具备处理大量复杂数据的能力。

关键词 图像处理; 图像特征提取; 压缩感知; 加速稳健特征; 关键帧; 手势检测

中图分类号 TP391.411

文献标识码 A

doi: 10.3788/LOP55.051013

Key Frame Extraction Algorithm of Sign Language Based on Compressed Sensing and SURF Features

Wang Min, Li Zeyang, Wang Chun, Shi Xinyuan

School of Information and Control Engineering, Xi'an University of Architecture and Technology,
Xi'an, Shaanxi 710055, China

Abstract A key frame extraction algorithm of sign language based on compressed sensing and speed up robust features(SURF) feature is proposed to recognize the real-time, large vocabulary sets and continuous sign language videos efficiently and accurately. The sign language videos are reduced to the image features of low dimensional and multi-scale frame with compressed sensing. The segmentation of sub lens is completed by a adaptive threshold value, and a large number of sign language frame data are processed. We use SURF feature points to complete the feature matching, and the SURF frame similarity curve is drawn for extracting the key frames. In the pre-processing stage, we use the HSV space adaptive color detection to abstract the sign language area. Experimental results show that the key frames extracted by the proposed algorithm have high accuracy, and the proposed algorithm has the ability to process large amounts of complex data.

Key words image processing; image feature extraction; compression sensing; speed up robust features; key frame; sign language detection

OCIS codes 100.2960; 110.2960; 100.2000

1 引言

手语是聋哑人以手型信息取代有声语言进行交谈的一种方式。目前的手语识别方法主要分为数据采集设备和计算机视觉^[1]。相比基于数据采集设

备, 手语识别具有价格昂贵、使用复杂且不适合推广的特点, 基于计算机视觉的手语识别采用普通摄像设备获取手语数据, 再运用图像处理、模式识别等领域的概念和方法进行手语识别, 因此价格低廉, 易于实现推广^[2-3]。

收稿日期: 2017-12-19; 收到修改稿日期: 2018-02-07

基金项目: 住房和城乡建设部科学技术项目计划(2016-R2-045)、陕西省自然科学基金基础研究资金(2014JM8343)、陕西省自然科学基金青年基金(2013JQ8003)

作者简介: 王民(1959—), 男, 本科, 教授, 硕士生导师, 主要从事数字语音处理、多媒体通信技术等方面的研究。

E-mail: wangmin1329@163.com

* 通信联系人。E-mail: zenolzy@163.com

当前,基于计算机视觉的手语识别算法主要是逐帧对手语视频进行手势检测与跟踪^[4-5],再根据手部运动规律进行手语识别。但由于手语帧数据量庞大,手语识别算法多采用人工神经网络(ANN)和隐马尔科夫模型(HMM)等复杂算法,难以实现手语识别的实时性需求^[6]。

通过研究《中国手语》^[7]能够发现手语词汇语义的表述方式为几个关键手势动作的组合。在手语识别中,一个动态手语可由1~5个关键静止手势进行完整地表述,但通过视频表现就需要30~250帧的图像。将其中大多数的过渡帧剔除,筛选出1~5个关键帧,就能表达一个手语词汇。因此,将手语识别中研究的动态视频转化为对一系列帧图像的处理研究,之后在帧图像中选取具有代表性的手势作为关键帧^[8],从而实现手语识别的目的。

目前,对于关键帧的检测识别算法研究主要在基于内容的视频检索与分类方面。其中,最基本的方法就是对初始数据进行等间隔采样,但由于方法过于简单,容易对不同情况下视频内容变化造成过采样或欠采样。运用聚类算法提取关键帧,是最为常用的方法。由于聚类数目无法提前得出,且其结果过于依赖阈值,故通常采用运算复杂的自适应算法,但效率较低难以满足实时性需求^[9-10]。基于运动分析的关键帧提取算法,也较为常用,其选取帧数据中相对运动量最多或最少的帧作为关键帧,但这种算法运算量较大且关键帧检测结果易受外部环境影响^[11]。针对本文实时、大词汇集、连续的手语识别需求,综合比较以上算法的优缺点,并基于计算机视觉的动态手语研究,设计了本文关键帧提取算法。

2 应用原理

2.1 压缩感知

使用单一特征无法完整描述手语图像整体信息,而多个特征则会明显提高计算复杂度和存储量,压缩感知原理可以有效解决上述问题。对于高维多尺度图像,通过稀疏随机投影矩阵进行稀疏表示,降维而得到低维特征,其通过重构算法依旧可以保留图像原有信息^[12]。

设 m 维向量 \mathbf{y} , 可由 n 维信号 \mathbf{x} 与 m 行 n 列矩阵 \mathbf{R} 表示为

$$\mathbf{y} = \mathbf{R}\mathbf{x}, \quad (1)$$

式中 $n \gg m$, 根据压缩感知原理,当原始信号 \mathbf{x} 是 K 稀疏的,同时当稀疏随机投影矩阵 \mathbf{R} 满足约束等距性条件(RIP),则

$$\mathbf{R} = (r_{ij}) = \sqrt{\rho} \times \begin{cases} 1, & P(1) = \frac{1}{2\rho} \\ 0, & P(0) = 1 - \frac{1}{\rho} \\ -1, & P(-1) = \frac{1}{2\rho} \end{cases} \quad (2)$$

且当 $\rho = 1$ 或 3 时,原始信号 \mathbf{x} 通过对 m 维向量 \mathbf{y} 的最优 l_1 范数问题的求解来实现重构。

2.2 SURF 特征提取

加速稳健特征(SURF)是一种尺度不变与旋转不变的特征点检测和描述方法^[13],具有处理速度快、稳健性好等特点,能够很好地描述帧间手势运动的变化情况。SURF对于处理手语这类尺度特征数据具有良好的效果^[14-15]。

SURF运用近似 Hessian 矩阵检测特征点,并通过积分图像^[16]大幅度降低运算复杂度。其算法主要由特征点检测、特征点表述与特征匹配组成。

在 SURF 特征提取算法中,首先运用快速 Hessian 矩阵遍历图像上的极值点。Hessian 矩阵定义为

$$\mathbf{H}[f(x, y)] = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \quad (3)$$

基于尺度空间,其在尺度 σ 上的定义为

$$\mathbf{H}(\bar{x}, \sigma) = \begin{bmatrix} L_{xx}(\bar{x}, \sigma) & L_{xy}(\bar{x}, \sigma) \\ L_{xy}(\bar{x}, \sigma) & L_{yy}(\bar{x}, \sigma) \end{bmatrix}, \quad (4)$$

式中 L_{xx} 、 L_{xy} 和 L_{yy} 为在点 $\bar{x} = (x, y)$ 处图像与相应二阶高斯模板相卷积的结果。

SURF 特征提取算法通过方框滤波近似表述为二阶高斯滤波,运用积分图像提高卷积计算速度;在尺度金字塔的构造中,SURF 算法通过扩大方框滤波的尺寸以近似不同尺度的图像金字塔模型。设方框滤波后, L_{xx} 、 L_{yy} 和 L_{xy} 分别由 D_{xx} 、 D_{yy} 和 D_{xy} 表示,则在尺度 σ 上,判别图像某点是否为特征点,可通过(5)式近似计算该点 Hessian 矩阵行列式值,若满足 $\det \mathbf{H} > 0$,则点 (x, y) 是特征点^[17]。

$$\det \mathbf{H} = D_{xx}D_{yy} - (0.9D_{xy})^2 \quad (5)$$

对于多尺度上的特征点检测,SURF 算法将通过 Hessian 矩阵求得极值后的点,与其所在的多尺度空间立体邻域内的点进行非极大值抑制;与 26 个邻域值相比,只有当该点值都大于(或小于)所有相邻点时,才将该点作为候选特征点。之后在尺度空间与图像空间中,对每个候选特征点,进行函数插值运算法以得到稳定的特征点位置和尺度值。

为保证旋转不变性,选取以特征点为中心,计算半径为 $6l$ (l 为特征点所在的尺度值) 的邻域内的点在 x 和 y 轴方向的 Haar 小波(边长为 $4l$) 的响应值(分别记为 dx 和 dy),得到该特征点的主方向。最终以特征点为中心,选取特征点的主方向作为 x 轴方向,边长为 $20l$ 的正方形区域,并划分成 4×4 的子区域,在每一个子区域计算 Haar 小波响应值,生成四维分量的矢量 $\mathbf{v} = [\sum dx, \sum dy, \sum |dx|, \sum |dy|]$,则每个特征点形成由一个 $4 \times 4 \times 4 = 64$ 维的特征向量来描述。

3 算法设计

研究动态手语识别算法,通过剔除手语视频中的过渡帧,筛选出具有代表性的手势作为关键帧,从而实现手语语义的快速识别。

3.1 预处理

提出一种基于压缩感知与 SURF 特征的手语关键帧提取算法,首先需要对手语帧序列进行预处理,主要是完成手势区域的检测与提取。当前阶段,对于手势区域的检测算法,仅用一台普通摄像机,难以简单地解决手部与面部重合的问题^[18]。因此,要解决此问题,同时还要满足实时、大词汇集、连续的手语识别的需求,采取手语者佩戴纯色普通手套再利用颜色检测算法提取手部区域的方法,消除了做手语的过程中身体其他部位带来的影响。由于手势区域检测只是预处理过程,其与后续关键帧提取算法相对独立,所以对预处理算法的更改,对关键帧提取算法没有太大影响,有益于算法的改进。

由于 HSV 空间对颜色敏感而对光照等外部因素不敏感^[19],因此采取基于 HSV 空间自适应颜色检测提取手势区域,而在做手语的过程中,目标颜色手势区域位置偏移最大,因此通过帧间差分运动检测,能够自适应地确定手套颜色。将手语帧图像转换成 HSV 颜色空间,之后逐一判别每个像素值是否属于手套颜色范围内,并标识出来,再运用轮廓查找的方式,获取手势区域。

3.2 压缩感知子镜头分割

对于手语图像,其多尺度低维特征是由多尺度高维特征通过稀疏随机投影矩阵 \mathbf{R} 投影得到的^[20]。对手语帧进行多尺度滤波,定义双向同性均值滤波器为

$$F = \left\{ F_{i,i}(x, y) = \frac{1}{i \times i}, i = 1, 3, \dots, 2N + 1 \right\}, \quad (6)$$

式中 i 为滤波器的尺度, N 为滤波器的个数。

利用 F 对手语图像滤波,将滤波后得到的图像序列以列向量的形式表达,即图像高维多尺度特征 \mathbf{f} 。设原始图像的宽、高分别为 w 和 h ,则 \mathbf{f} 维度为 Nwh ,根据(1)式得到的低维多尺度特征为

$$\mathbf{g} = \mathbf{R}\mathbf{f}. \quad (7)$$

根据 2.1 节可知, \mathbf{f} 可由 \mathbf{g} 表达,由此实现减少计算复杂度和空间存储量的多尺度图像降维。

构建手语帧图像低维多尺度特征时,每个元素 \mathbf{g}_i 对应于稀疏随机投影矩阵 \mathbf{R} 与高维多尺度特征投影的一个采样向量。为表现对不同特征元素的信任度,设随机权值向量 \mathbf{W} 有

$$\mathbf{W} = [\omega_1 \ \omega_2 \ \dots \ \omega_i \ \dots \ \omega_k], \quad (8)$$

式中 k 为图像低维多尺度特征的维数, ω_i 为第 i 个元素对应的随机权值。为避免权值过小,造成某些采样值过度抑制, ω_i 满足约束性条件为

$$\sum_{i=1}^p \omega_i = 1, 0 < \omega_i < 1, \frac{\max\{\omega_i\}}{\min\{\omega_i\}} \leq 2. \quad (9)$$

在一段完整手语视频序列中,其包含的各个图像低维多尺度特征,按顺序组成低维多尺度特征列向量集合 \mathbf{G} 为

$$\mathbf{G} = [\mathbf{g}_1 \ \mathbf{g}_2 \ \dots \ \mathbf{g}_i \ \dots \ \mathbf{g}_k], \quad (10)$$

式中 k 为镜头中帧的数量。将随机权值向量 \mathbf{W} 中的 ω_i 与低维特征列向量集合 \mathbf{G} 中第 i 个特征 \mathbf{g}_i 对应相乘,即有运算

$$\mathbf{G}' = [\mathbf{g}'] = [\mathbf{W}^T \cdot * \mathbf{g}_i], \quad (11)$$

式中 $*$ 为向量的阿达马乘积。此时,将得到 \mathbf{G}' 的作为镜头低维多尺度特征列向量组,则 \mathbf{G}' 中每一列 \mathbf{g}' 代表一幅帧图像的匹配特征。

在上述运算中,得到图像低维多尺度特征 \mathbf{g}' 后,根据各特征之间的相似性度量进行子镜头的分割,其中选取欧氏距离作为相似性度量。对拥有 k 个帧的镜头序列,能够计算出 $k-1$ 个帧间相似性值 (sm_1, \dots, sm_{k-1}) , 得到数组 $D = \{sm_1, \dots, sm_{k-1}\}$; 以数组 D 中元素作为数据空间的样本,进行子镜头分割。对 D 中的元素由大到小进行排序,假设排序后有 $sm_1 \geq sm_2 \geq \dots \geq sm_{k-1}$, 令 T 为

$$T = \operatorname{argmin} \delta_W^2, \quad (12)$$

式中 $\delta_W^2 = q_H \delta_H^2 + q_L \delta_L^2$, $q_H = T$, $q_L = k - T - 1$, $\delta_H^2 = \frac{1}{q_H} \sum_{i=1}^T [sm_i - \mu_H]^2$, $\delta_L^2 = \frac{1}{q_L} \sum_{i=T+1}^{k-1} [sm_i - \mu_L]^2$, $\mu_H = \frac{1}{q_H} \sum_{i=1}^T sm_i$ 和 $\mu_L = \frac{1}{q_L} \sum_{i=T+1}^{k-1} sm_i$, 则 sm_T 为所求阈值,其

值自适应得出。从第一帧开始,遍历所有帧数据,若相邻两帧帧差 $\geq sm_T$,则包括后续帧都划分为新的子镜头。其中, g' 具有尺度不变性。

3.3 基于 SURF 特征的关键帧提取

将 SURF 算法应用于经过预处理后的手语帧序列,用像素值 $I(x, y)$ 取代函数值 $f(x, y)$,采用尺度归一化的二阶高斯滤波计算帧序列的二阶偏导。于是,定义图像 $I(x, y)$ 在尺度 σ 的 Hessian 矩阵^[21-22]为

$$\mathbf{H}(x, y, \sigma) = \begin{bmatrix} L_{xx}(x, y, \sigma) & L_{xy}(x, y, \sigma) \\ L_{xy}(x, y, \sigma) & L_{yy}(x, y, \sigma) \end{bmatrix}, \quad (13)$$

式中 $L_{xx}(x, y, \sigma)$ 为图像在点 (x, y) 处与高斯函数二阶偏导 $\partial^2[g(x, y, \sigma)]/\partial x^2$ 相卷积的结果。类似地,可定义 $L_{xy}(x, y, \sigma)$ 和 $L_{yy}(x, y, \sigma)$ 。运用 Hessian 矩阵行列式计算帧图像上特征点位置和尺度信息。

手语视频完成子镜头分割后,再从各子镜头中提取出关键帧。对于手语帧数据,利用 SURF 特征点提取算法提取各帧中的特征点,并匹配相邻帧的 SURF 特征点,之后根据本文关键帧判别准则,判别、筛选关键帧,最终输出所提取到的关键帧。

对于相邻帧的 SURF 特征点匹配方法,当两幅手语帧图像的 SURF 特征矢量生成后,运用特征点匹配算法寻找帧图像中某一特征点对应的匹配点,设图 1(a)中某一特征点 A ,根据欧氏距离度量,在图 1(b)中寻找与其最近的两个特征点,令其中最近的特征点 B 的欧氏距离为 l_1 ,特征点 C 的欧氏距离为 l_2 ,设定判别阈值 β ,其中常取判别阈值 $\beta = 0.5$,若下式成立:

$$\frac{l_1}{l_2} \leq \beta, \quad (14)$$

则得到相应的特征点 B 与原特征点 A 互相匹配,即点 A 与 B 为一对特征点,反之,则认为图 1(b)中不存在与特征点 A 相匹配的特征点,选取两帧图像的匹配效果如图 1 所示。

对于关键帧判别准则,当手语视频内容发生较大变化时,帧间相似度会发生显著的变化,视频帧间相似性曲线出现拐点,因此寻找帧间相似性曲线拐点位置的帧图像作为该子镜头手语视频中的关键帧。在得到相邻两帧间的特征点匹配数之后,进行该帧图像的相似性度量,其不但需要计算相匹配特征点数,而且要考虑帧图像整体信息,即两帧中的特征点总数,帧图像的相似度 τ 为



图 1 两手语帧匹配效果示例图

Fig. 1 Matching image of two sign language frame

$$\tau = \frac{2m}{n_1 + n_2}, \quad (15)$$

式中 m 为两帧相匹配特征点对数, n_1 和 n_2 分别为这两帧的特征点总数。

3.4 基于压缩感知与 SURF 特征的手语关键帧提取算法

对于实时、大词汇集、连续的手语视频,需要完成从动态手语转换为静态手势的过程,并且要对手语帧图像进行粗划分,将具有大致相似性的手语动作快速地分割到一起。由于手语动作存在一定的细节变化,需要对划分后的各集合再进行精细筛选的关键帧提取,因此提出了一种基于压缩感知与 SURF 特征的手语关键帧提取算法。其算法框图如图 2 所示,具体步骤如下:

- 1) 手语视频帧图像预处理,利用基于 HSV 空间自适应颜色检测提取手势区域。
- 2) 根据压缩感知原理,构造稀疏随机投影矩阵,将高维多尺度帧图像特征转换为低维多尺度帧图像特征。
- 3) 构建随机权值向量与低维多尺度特征向量的阿达马乘积运算,生成各帧的匹配特征,并根据匹配特征的相似性度量以及自适应的判别阈值,进行帧图像序列的子镜头分割。
- 4) 在各子镜头中,提取各帧 SURF 特征点,再对相邻帧的 SURF 特征点进行匹配。
- 5) 绘制 SURF 帧间相似度曲线,寻找曲线拐点,再进行局部极值筛选,得到局部最优手势位置,其所在的帧即作为关键帧。

4 实验分析

通过具体实验,验证本文算法的稳定性与准确性。选用 4 段利用普通摄像头录制、时长约为 5 min 的手势视频,以及 4 段相同时长、相同拍摄条件、不同目标人与背景的手势视频,实验平台为 Matlab R2015b。图 3(a)为实验室环境下的两段视频,分别

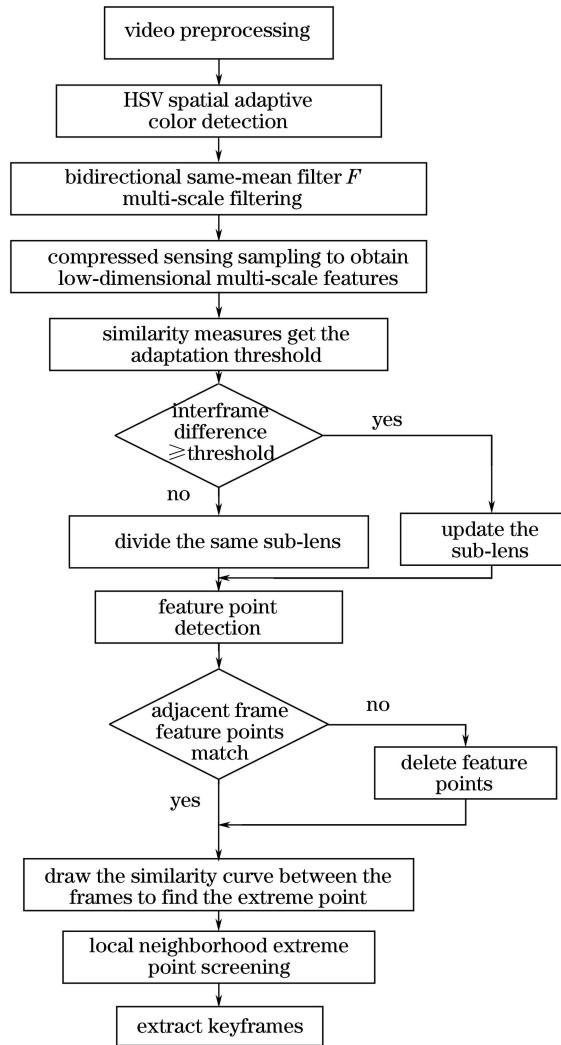


图 2 基于压缩感知与 SURF 特征的手语关键帧提取算法框图

Fig. 2 Key frame extraction algorithm diagram of sign language based on compressed sensing and SURF feature

不断重复数字 0~5 的手势;图 3(b)为不同环境与光照条件下的两段视频,分别不断重复数字 0~5 手

势及复位 0 手势作为本次实验样本。其中,摄像头采样频率为 $30 \text{ frame} \cdot \text{s}^{-1}$ 。



图 3 部分测试视频示例。(a) 实验室环境;(b) 不同环境与光照条件

Fig. 3 Partial test videos. (a) Laboratory environment; (b) different environment and light conditions

基于压缩感知与 SURF 特征的关键帧提取算法运用基于压缩感知的子镜头分割算法对以上视频数据进行处理,并按时间顺序将相关性较大的帧数据划分到一个子镜头当中,最后得到若干子镜头视频片段,如表 1 所示,其中视频 Test1 与 Test2 共得

到 27 段子镜头片段,而 Test3 与 Test4 得到 24 段子镜头片段。

完成分割后得到的各子镜头片段,并运用基于 SURF 特征的关键帧提取算法,得到关键帧,部分提取结果如表 1 所示。

表 1 子镜头分割及检测关键帧表

Table 1 Subshot segmentation and key frame detection

Video sample	Number of subshots	Average key frame number /part position
Test1	27	6/[2, 12, 40, 86, 113, 149]
Test2	27	6/[7, 26, 51, 76, 100, 132]
Test3	24	7/[33, 49, 67, 101, 129, 140, 167]
Test4	24	7/[37, 55, 85, 115, 134, 153, 178]

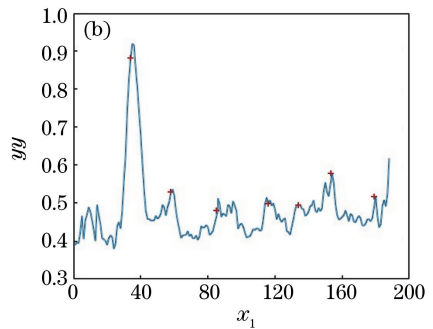
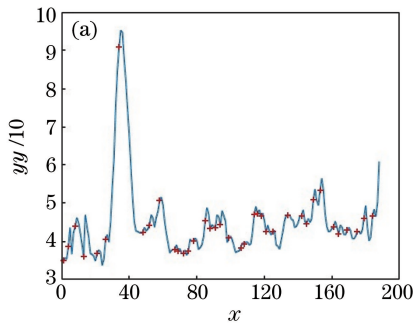


图 4 帧间相似性曲线极值点时间序列图。(a)极点;(b)关键帧

Fig. 4 Time sequence diagram of interframe similarity curve extreme point. (a) Extreme point; (b) key frame

图 5 为其他算法与本文算法的部分输出结果对比示例。图 5(a)为本文算法结果,图 5(b)为非特定人动态手语视频关键帧检测算法结果^[6],图 5(c)为经典的

无监督聚类自适应关键帧提取算法结果^[23],其中无监督聚类算法运用计算各帧的特征向量与聚类中心的距离,来确定各子聚类及聚类中心,得到关键帧。



图 5 3 种算法输出关键帧示例。(a)本文算法;(b)非特定人动态手语视频关键帧检测算法;(c)无监督聚类算法

Fig. 5 Output key frames of three algorithms. (a) Proposed algorithm;

(b) non-specific dynamic sign language video key frame detection algorithm; (c) unsupervised clustering algorithm

为验证本文算法对动态手语的识别有效性,进行对比实验。运用非特定人动态手语视频关键帧检测算法^[6]、无监督聚类自适应关键帧提取算法^[23]对实验中相同样本完成提取;比对三种算法实验效果,结果如表 2 所示。以视频 Test4 为例,对分割到的各子镜头分别运用三种算法进行关键帧提取。由于视频在不断重复手势动作,可以计算得到子镜头中平均检测关键帧的数量,通过对比,本文算法提取到的结果相对准确;由关键帧漏检情况可知,本文算法存在关键帧漏检,但概率较小;由关键帧稳定性结果可知,虽然算法存在漏检或多检的情况,但对比其他两种算法,本文算法提取到的结果稳定性更高。本文算法实现过程比较简单,运算效率相对较高。同

时针对解决实时、大词汇集、连续动态视频的特点,其算法结果更加稳定、准确率相对较高,且运算时间与其他算法比较没有太大差别。

两种比对算法中,无监督聚类自适应关键帧提取算法^[23]采用复杂的自适应算法确定聚类数目,需要计算的数据量较大、时间复杂度较高,而其算法运行效率不高,虽然提取到的关键帧数量较多,但较多无意义的同类帧作为关键帧会使个别子镜头中需要提取的关键帧出现缺失。比对算法中,非特定人动态手语视频关键帧检测算法^[6]的关键帧提取算法简单、时间复杂度不高,但其运算效果过于依赖前期的预处理过程,造成算法结果波动性较大,并且易出现关键帧漏检的情况。

表 2 实验数据比较表

Table 2 Comparison of experimental data

Item	Proposed algorithm	Non-specific dynamic sign language video key frame detection algorithm	Unsupervised clustering algorithm
Average number of key frames	7.125	9.375	12.958
Key frame leak detection rate /%	0.75	1.75	3.25
Average running time /s	13.551	11.236	18.912
Stability of key frame detection /%	98.25	93.75	82.50

由于本文算法是面向非特定人的手语识别,对于不同背景和目标人的手势视频,依旧能够准确高效地完成关键帧的提取,其部分结果如图 6 所示。本文算法在前期预处理阶段基于手套颜色对手部区

域进行检测提取,而对于身体其他部位及背景信息不考量。在关键帧提取阶段,算法作用于已预处理后的图像,不匹配某一特定人的具体信息。因此,算法能够完成针对不同背景和特定人的手语识别。



图 6 非特定人部分关键帧示例

Fig. 6 Non-specific part of the key frames

5 结 论

相比于传统的手语识别算法,本文算法能够快速、准确且易于实现地处理大量视频数据,实现实时手语识别的目的;与其他的关键帧提取算法相比,本文算法高效地针对不同的场景环境和目标人的大量帧数据完成子镜头分割与关键帧提取。并且,沿用通过一组静态手势组合表征一个动态手语词汇的观点,保持与《中国手语》中通过一组关键动作分解表现手语语义方式的一致性。利用识别手势关键帧的方法完成手语识别的效果,并对传统关键帧提取算法加以改进,提出基于压缩感知与 SURF 特征的手语关键帧提取算法。高效地完成了手语识别,实现了人机和谐交互。在今后的学习研究中,应继续对手势区域检测算法与关键帧提取算法加以改进,不断提升运算效率。

参 考 文 献

- [1] Maraqa M, Al-Zboun F, Dhyabat M, *et al.* Recognition of arabic sign language (ArSL) using recurrent neural networks[J]. Journal of Intelligent Learning Systems and Applications, 2012, 4: 41-52.
- [2] Sarkar A R, Sanyal G, Majumder S. Hand gesture

recognition systems: a survey [J]. International Journal of Computer Applications, 2013, 71(15): 26-37.

- [3] Samantaray A, Nayak S K, Mishra A K. Hand gesture recognition using computer vision [J]. International Journal of Scientific & Engineering Research, 2013, 4(6): 1602-1608.
- [4] Liu W, Zhao W J, Li C. Long-term visual tracking based on spatio-temporal context [J]. Acta Optica Sinica, 2016, 36(1): 0115001.
刘威, 赵文杰, 李成. 时空上下文学习长时目标跟踪 [J]. 光学学报, 2016, 36(1): 0115001.
- [5] Shen Q, Yan X L, Liu L F, *et al.* A scale adaptive correlation filtering tracker based on adaptive feature selection [J]. Acta Optica Sinica, 2017, 37(5): 0515001.
沈秋, 严小乐, 刘霖枫, 等. 基于自适应特征选择的多尺度相关滤波跟踪 [J]. 光学学报, 2017, 37(5): 0515001.
- [6] Li S R, Huang Y Y, Hu Z J, *et al.* Effective algorithm of key frame detection from dynamic sign language video [J]. Journal of Chinese Computer Systems, 2015, 36(7): 1597-1601.
李姝蓉, 黄元元, 胡作进, 等. 非特定人群动态手语视频中的关键帧检测算法 [J]. 小型微型计算机系

- 统, 2015, 36(7): 1597-1601.
- [7] The Department for Education and Employment of China Disabled Persons' Federation, China Association of the Deaf and Hard of Hearing. Chinese sign language[M]. Beijing: Huaxia Publishing House, 2003.
中国残疾人联合会教育就业部, 中国聋人协会. 中国手语[M]. 北京: 华夏出版社, 2003.
- [8] Shih H C. A novel attention-based key-frame determination method [J]. IEEE Transactions on Broadcasting, 2013, 59(3): 556-562.
- [9] Hao H F. Key technology research of video shot segmentation and key frame extraction[D]. Wuhan: Central China Normal University, 2015.
郝会芬. 视频镜头分割和关键帧提取关键技术研究[D]. 武汉: 华中师范大学, 2015.
- [10] Zhou W Z, Chen G L, Du S S, *et al.* An improved iterative closest point algorithm using clustering[J]. Laser & Optoelectronics Progress, 2016, 53(5): 051202.
周文振, 陈国良, 杜珊珊, 等. 一种聚类改进的迭代最近点配准算法[J]. 激光与光电子学进展, 2016, 53(5): 051202.
- [11] Zhong X, Yang G, Lu Y S. Method of key frames extraction based on double-threshold values sliding window sub-shot segmentation and fully connected graph[J]. Computer Science, 2016, 43(6): 289-293.
钟欣, 杨光, 卢炎生. 基于双阈值滑动窗口子镜头分割和完全连通图的关键帧提取方法[J]. 计算机科学, 2016, 43(6): 289-293.
- [12] Nagesh P, Li B X. A compressive sensing approach for expression-invariant face recognition [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2009: 1518-1525.
- [13] Bay H, Ess A, Ess A, Tuytelaars T, *et al.* Speeded-up robust features (SURF)[J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.
- [14] Yang Q, Peng J Y. Chinese sign language recognition method based on depth image information and SURF-BoW [J]. Pattern Recognition and Artificial Intelligence, 2014, 27(8): 741-749.
杨全, 彭进业. 基于深度信息和 SURF-BoW 的中国手语识别算法[J]. 模式识别与人工智能, 2014, 27(8): 741-749.
- [15] Bao J T, Song A G, Guo Y, *et al.* Dynamic hand gesture recognition based on SURF tracking [J]. Robot, 2011, 33(4): 482-489.
包加桐, 宋爱国, 郭晏, 等. 基于 SURF 特征跟踪的动态手势识别算法[J]. 机器人, 2011, 33(4): 482-489.
- [16] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features [C]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001: 511-518.
- [17] Zhang Z L, Xue J, Qiao H H, *et al.* The vehicle retrieval method of traffic video based on improved SURF algorithm [J]. Journal of Northwestern Polytechnical University, 2014, 32(2): 297-302.
张子龙, 薛静, 乔鸿海, 等. 基于改进 SURF 算法的交通视频车辆检索方法研究[J]. 西北工业大学学报, 2014, 32(2): 297-302.
- [18] Liu B, Zhao X, Hu C H, *et al.* Gesture recognition method of hand over face occlusion in color and depth images[J]. Laser & Optoelectronics Progress, 2016, 53(6): 061001.
刘斌, 赵兴, 胡春海, 等. 面向颜色深度图像手脸近距遮挡的手势识别[J]. 激光与光电子学进展, 2016, 53(6): 061001.
- [19] Huang S Y, Liu Z Y, Ruan T Y. Algorithm of license extraction based on HSV color space and SVM [J]. Computer Systems & Applications, 2014, 23(8): 150-153.
黄社阳, 刘智勇, 阮太元. 基于 HSV 颜色空间和 SVM 的车牌提取算法[J]. 计算机系统应用, 2014, 23(8): 150-153.
- [20] Pan L, Shu X, Cheng K, *et al.* A key frame extraction algorithm based on compressive sensing and entropy computing [J]. Journal of Optoelectronics • Laser, 2014, 25(10): 1977-1982.
潘磊, 束鑫, 程科, 等. 基于压缩感知和熵计算的关键帧提取算法[J]. 光电子 • 激光, 2014, 25(10): 1977-1982.
- [21] Athani S, Tejeshwar C H. Performance analysis of key frame extraction using SIFT and SURF algorithms[J]. International Journal of Computer Science and Information Technologies, 2016, 7(4): 2136-2139.
- [22] Grycuk R, Knop M, Mandal S. Video key frame detection based on SURF algorithm [C]. International Conference on Artificial Intelligence and Soft Computing, 2015: 566-576.
- [23] Zhuang Y T, Rui Y, Huang T S, *et al.* Adaptive key frame extraction using unsupervised clustering [C] // Proceedings International Conference on Image Processing, 1998, 2: 866-870.