

基于 Haar 小波分解的实时手势识别

安 涛, 彭进业, 吴 静

(西北大学信息科学与技术学院, 西安 710127)

摘 要: 传统基于样例的识别方法由于数据量大而难以应用于实时手势识别。为此, 利用 Haar 小波可用于分解图像, 保留细节部分而丢弃高频部分的特点, 将视频序列中每一帧的图像尺寸降到不影响系统准确率的程度, 同时降低识别过程的计算量。采用单摄像头在一个 36 个元素的美国手语手势集上进行实验, 结果证明系统的有效识别速率可以提高到 30 f/s 且准确率几乎未变, 可以满足实时性需求。

关键词: 手势识别; Haar 小波; 小波分解; 基于样例的识别; 最近领域匹配

Real-time Gesture Recognition Based on Haar Wavelet Decomposition

AN Tao, PENG Jin-ye, WU Jing

(School of Information Science and Technology, Northwest University, Xi'an 710127, China)

【Abstract】 Toward the basic exemplar-based recognition method is hard to apply in real-time gesture recognition for the huge amount of data, the Haar wavelet can be used to decompose images to keep the detail parts and discard the high-frequency parts, where each frame image in the video sequence can be downsized to the minimal level such that no obvious effect on the system's accuracy and the computational cost in the recognition stage can be saved. Comparing by the experiment with a single Web camera on a 36 elements of American sign language gesture dataset, the results show that the effective system recognition can be sped up to 30 f/s with the same accuracy, which meets real-time requirement.

【Key words】 gesture recognition; Haar wavelet; wavelet decomposition; exemplar-based recognition; nearest-neighbor matching

DOI: 10.3969/j.issn.1000-3428.2011.24.046

1 概述

手势识别在当前计算机视觉研究中是一个非常活跃的领域, 把自然手势加强到传统的人机交互(HCI)中, 将最终比仅使用键盘鼠标的交互方式获得更高的输入带宽。研究进展见综述^[1]。多数研究针对多摄像头输入或 3D 手势跟踪。尽管在快速检索和连续手语识别上得到了很好的结果, 但包含大词汇量和手势复杂且快速变换时, 实时分类和识别的问题仍然存在。

一个流行的基于视觉的手势识别方法是通过时间 Markov 模型^[2]表示手势动作, 但是限制了能被正确识别的手势范围。实际中希望系统在移出摄像头的视野范围或当手势被部分遮盖时, 识别准确率不损失或者系统重新初始化时没停顿, 因此, 基于时空相干的方法不满足实时性。

基于手势外观的方法常使用较小词汇量数据库(20 个~100 个手语)而且有时会依赖于彩色标记, 与本文背景相似的是文献[3]方法。其中手势词汇量为 31 个手形, 使用手指定位技术最高达到 99%识别率, 然而需要复杂的初始化过程, 且跟踪和识别分开进行, 运算量非常大^[3]。文献[4]采用了多模式的手势识别方法, 采用基于视线跟踪和手势识别的交互方式通过一个摄像头能远程有效地操控计算机。实验中系统的视觉处理部分可以达到 15 f/s 的速度^[4]。

本文提出一个完全基于单手识别的方案, 使用流行的基于样例的识别方法, 仅用单摄像头作为输入传感器, 可以在 2.2 GHz 主频的笔记本电脑上以每秒 30 帧的系统识别速率可靠地解释一个由 36 个元素组成的手势集。

2 手势模型

手势集如图 1 所示。

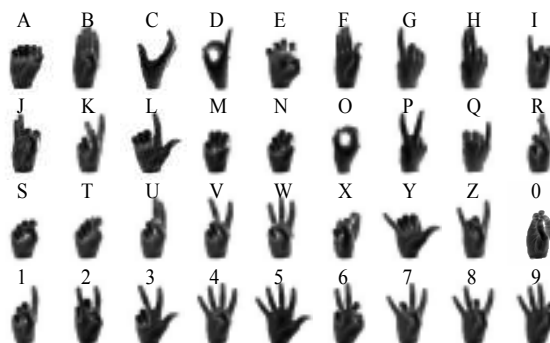


图 1 实验用的手势集

对 ASL 作如下修改: 动态手势“J”和“Z”用图中的静态模式代替以便能在单帧中被识别; 和字母手形一样的数字手形修改为不同手形。

本文的创新点在于将 Haar 小波分解能降低图像尺寸的功能扩展到基于样例的识别方法中, 节省了每一帧在识别阶段的计算量, 而使实时性成为可能。基于样例的方法能够很容易地应用于手势类有少量样本的情况。以 3 阶 Haar 小波分解为例, 图像的尺寸降 1/8, 输入视频由每帧 640×480 减少到 80×60, 则识别阶段的计算量从 $O(10^6)$ 降到大约 $O(10^4)$, 速度会明显提高。显然这是对帧图像的细节部分有可忽略的损失才达到的, 但是系统保留了几乎和全模板匹配一样的准确率。

为了突显所要研究的问题, 简化手势模型在三维空间变

作者简介: 安 涛(1984—), 男, 硕士研究生, 主研方向: 图像处理, 手势识别; 彭进业, 教授、博士生导师; 吴 静, 硕士研究生

收稿日期: 2011-06-16 **E-mail:** fateflame@163.com

化而造成的复杂性, 对系统的一般性做如下限制: 只输入右手手势; 为避免远近景对手势采集后尺度变化的干扰, 固定摄像头跟手之间的距离; 为避免平移变化的干扰, 确保手势质心与背景中心重合; 视频采集在室内进行, 光照条件要确保简单图像处理能检测到大多数肤色像素点; 为避免手势分割涉及对复杂背景的处理, 要求背景单一且与肤色有较明显的对比。

3 Haar 小波分解对帧图像的预处理

由视频输入的 RGB 帧图像灰度化后, 通过设定一个全局阈值(每一帧图像强度的平均值)或者局部阈值(包含手势的矩形框中的平均值)把手势和背景分割。这对帧图像中大部分的手势像素是可靠的, 但在手势边缘, 高亮和内部阴影区域会丢失像素。经过 Haar 小波分解对每一帧图像降尺寸后, 由于部分肤色像素误分类产生的可变因素将进一步减少, 变成附加于分类过程中的次要因素, 在随后的基于样例的识别阶段被忽略。图 2 显示了对一幅帧图像的处理步骤。

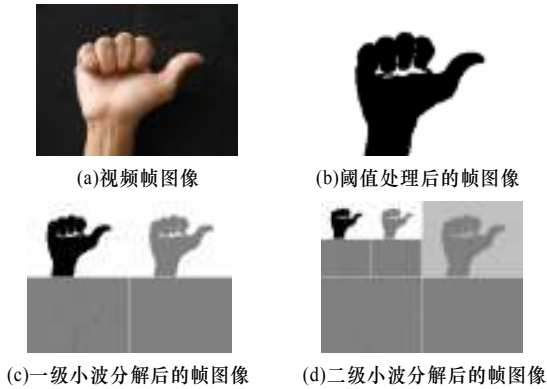


图 2 预处理阶段

基于 Haar 小波的图像处理基本思想见文献[5]。对帧图像的 Haar 小波分解是行列独立进行的, 以 Haar 小波对帧图像 f 的行分解为例: 假设 f 的每一行有 N 个像素, 则原始帧图像可以表示为 $\{f_r^0\}_{r=0}^{N-1}$, r 表示像素位置, 则 $\{f_r^1\}_{r=0}^{N_1-1}$ 和 $\{d_r^1\}_{r=0}^{M_1-1}$ 分别表示一级 Haar 小波分解后 f 的细节部分和低频部分, N 为偶数时 $N_1 = N/2$, N 为奇数时 $N_1 = (N+1)/2$, $M_1 = N - N_1$ 。对帧图像而言, 像素个数是 2 的幂次方的倍数, 所以实际上可以用下面的算法[5]:

$$\begin{cases} d_r^1 = f_{2r}^0 - f_{2r+1}^0 & r = 0, 1, \dots, N/2 - 1 \\ f_r^1 = \text{int}(d_r^1/2) + f_{2r+1}^0 & r = 0, 1, \dots, N/2 - 1 \end{cases}$$

$\text{int}(x)$ 是取整函数, 即 $x-1 \leq \text{int}(x) \leq x$ 。 f_r^1 为新的手势帧图像继续进行下一级分解, 最终得到一组保留图像大部分信息的细节小波系数和若干组高频小波系数。细节系数用于识别阶段, 而高频细节将被去除, 包含这些冗余信息的样本在识别过程中价值不大, 反而增加了系统的计算量。

从图 2(c)和图 2(d)可以看到, 每一级分解后, 尺寸 r 降为原来的一半, 但是手势的大部分信息仍然保留。若进行多级分解, 图像尺寸降更低。但在提高识别速率的同时, 将丢失手势帧图像的细节信息而导致误识别率增加为代价。

4 基于样例的识别方法

由前文得到的手势帧图像被表示成一维列向量 $x \in X^N$, 所有帧图像的列向量级联成集合 X^N , N 是经过分解后的帧图像像素个数。所有的帧图像都是标准化的二值图像, 如图 2(b)所示。系统对测试手势集的识别是基于一个事先获得

的训练手势集。假设训练集中有 H 类手语, 第 H 类手语已经得到 M_h 个训练手势样本。以下假设所有手语类都有同样数量的训练样本数 M , 这样, $M_h = M \forall H$ 。

训练样本表示为 x_{mh} , $m=1, 2, \dots, M$ 且 $h=1, 2, \dots, H$ 。规定样本 x_{mh} 的正确标识是 h 。分类器就是一个函数 $c(x)=l$, 输入测试图像 x 并返回一个标识 $l \in \{1, 2, \dots, H\}$ 。理想分类器将遵循: $c(x_{mh}) = h \forall m, h$ 。实际中使用的是软分类器, 不返回标识而是对给定的图像标识组合 (x, l) 赋值:

$$c(x, l): X^N \times \{1, 2, \dots, H\} \rightarrow [0, 1]$$

其值接近于 1 则说明图像可能就是这个标识, 接近 0 则说明不可能。理想的软分类器对训练集中的标识将赋值为 1, 反之为 0, 即: $c(x_{mh}, l) = \delta_{hl} \forall l, m, h$, δ 是 Kronecker 符号。注意这里不需要软分类器返回概率值, 也不要求 $\sum_{l=1}^H c(x, l) = 1$ 。

比较成熟的技术是最近邻域匹配分类器, 一个二值图像 x 与所有的训练集样本进行比较, 与之最接近的样本标识将被分类器识别出来。2 幅二值图像 x 和 y 的相似性度量可以是一个中心相关:

$$k(x, y) = \frac{4}{N} \sum_{n=1}^N \left(x_n - \frac{1}{2} \right) \left(y_n - \frac{1}{2} \right)$$

当 x 和 y 完全相同其值为 +1, 当 $y=1-x$ 时其值达到最小值 -1。

定义一个距离量为 $d(x, y) = 1 - k(x, y)$, 则其最近邻域分类器 c_{NN} 为:

$$c_{NN}(x) = \min_{h \in H} \left(\min_{m \in M} d(x, x_{mh}) \right)$$

其软分类器形式是:

$$c_{NN} = \max_{m \in M} \frac{1}{2} k(x, x_{ml}) + \frac{1}{2}$$

最近邻域匹配的计算复杂度是 $O(MHN)$, 同图像数、手势数和像素数是线性关系。

5 实验及比较结果

训练集中有 2 655 个手势帧图像训练样本(20 个手语如“M”、“N”的样本数为 85~90, 其余为 50~55), 原始分辨率为 640×480 像素。为了对比, 使用 Haar 小波对每个样本进行 4 级分解, 分辨率分别为 320×240 像素、 160×120 像素、 80×60 像素和 40×30 像素, 最终的训练集样本数扩展为 10 620。

测试集直接从输入视频中抽取, 然后设定不同的分解级数。系统对每一帧中的手势进行匹配并给出与训练集中相似度匹配最大的标识, 这样确保测试视频每一帧都有一个与之配对的训练集手势。如果帧速率为 30 f/s, 则对一个 5 s 的视频, 产生一个 150 个视频帧图像的测试集, 图 3(a)和图 3(b)演示了不同帧速率下从摄像头输入的手语字母“V”的手势, 经过 3 级 Haar 小波分解, 然后系统给出与之最为匹配的测试集手势帧图像。



(a) 帧速率=15 f/s

(b) 帧速率=30 f/s

图 3 手语字母“V”在不同帧速率下的识别过程

为了便于比较, 首先设定固定的帧速率, 视频序列的长度不定, 但要知道每个序列的确切长度, 这样可以知道测试

图像帧数,然后统计系统给出的匹配标识数和正确的匹配数,就可以计算出识别过程的有效帧速率和正确率。以手语字母“S”的识别为例,当输入视频设定为 15 f/s,若视频长度为 20 s,那么一共有 300 个测试帧图像,系统给出 300 个匹配标识,其中只有一个是错误的匹配,说明在此情况下,正确率为 99.67%,有效帧速率为 15 f/s;而当输入视频设定为 30 f/s,系统只给出了 217 个匹配,说明有效帧速率小于 30 f/s,则系统会忽略其正确率。

表 1 是从实验结果中截取的部分数据,通过表中的数据

表 1 帧速率为 15 f/s 和 30 f/s 时部分实验结果

图像	帧速率/(f·s ⁻¹)	手语字母识别正确率/(%)							
		3	0	C	J	M	O	S	W
原始帧图像 (640×480 像素)	15	100.00	—	100.00	100.00	—	100.00	99.67	100.00
	30	—	—	—	—	—	—	—	—
1 级小波分解 (320×240 像素)	15	100.00	100.00	100.00	100.00	100.00	100.00	99.52	100.00
	30	100.00	—	100.00	100.00	100.00	—	99.21	100.00
2 级小波分解 (160×120 像素)	15	100.00	99.54	100.00	100.00	100.00	100.00	99.53	100.00
	30	100.00	99.50	99.88	100.00	100.00	100.00	99.28	100.00
3 级小波分解 (80×60 像素)	15	100.00	99.81	99.69	100.00	100.00	100.00	99.13	100.00
	30	100.00	99.59	99.37	100.00	100.00	100.00	98.77	100.00
4 级小波分解 (40×30 像素)	15	87.99	79.28	85.59	90.69	94.55	93.41	75.62	91.15
	30	85.44	78.63	84.71	90.19	93.19	93.15	73.26	91.02

6 结束语

本文所采用的方法尽管在降低手势识别阶段的计算量上取得了很好的效果,能够满足系统的实时性,但是距离实用性还是很远。系统对手势视频采集的背景和输入方式做了比较严格的要求,但是这个要求比使用手套或者色彩标记更方便。此外,系统所采用的基于样例的识别方法,计算量仍然会随着手势训练集的扩充而急剧增加,所以不适合较大词汇量规模的手语识别数据库,如何使用更高级的识别方法将可能减少样本数量并且更进一步加快速率,这将是下一步的研究重点。

参考文献

- [1] Ong S C W. Automatic Sign Language Analysis: A Survey and the

可以看出,利用 Haar 小波分解的确可以提高系统的识别速率,同时可以保证正确率在一个可靠的范围内。尤其当输入视频帧分辨率较高时,在保证系统的正确率没有较大损失的情况下,可利用 Haar 小波分解显著地降低数据量。实验中最好的情况是利用 Haar 小波对原始帧图像进行 3 级分解,可以达到 50 f/s 的系统识别速率,而仅仅有个别手势识别的正确率略有降低。但是利用 Haar 小波进行 4 级分解后,由于信息损失的过于严重而引入较大的误分类,以至于系统识别的正确率很低。

- Future Beyond Lexical Meaning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(6): 873-891.
- [2] Moni M A. HMM Based Hand Gesture Recognition: A Review on Techniques and Approaches[C]//Proc. of the 2nd International Conference on Computer Science and Information Technology. [S. l.]: IEEE Press, 2009.
- [3] 周航,阮秋琦. 基于 ROI 分割和相干映射的裸手字母手势识别[J]. 通信学报, 2007, 28(5): 94-101.
- [4] 肖考勇,秦华标. 基于视线跟踪和手势识别的人机交互[J]. 计算机工程, 2009, 35(15): 198-200.
- [5] 吴昊. 一种基于 Haar 小波的图像处理技术[C]//2006 年全国数学技术应用科学学术论坛论文集. 昆明: [出版者不详], 2006: 159-162.

编辑 顾逸斐

(上接第 137 页)

上进行仿真测试,结果表明,可以对芯片中的 SHA-1 算法 IP 核、COS 集成的 MD5 算法进行成功的调用。以 MD5 算法调用为例,测试结构如下:

```

Main(){
//初始化
Intrfc_Hash EEPROM intrfc_MD5=**;
Class_Hash class_MD5;
intrfc_MD5->byAttribute=**;
intrfc_MD5->pIntrfcHash(HashMode,**);
//设置算法接口
class_MD5->OprState=**;
class_MD5->HshBlkLen= 0x10;
class_MD5->MsgBlkLen= 0x40;
class_MD5->dwByteCounter= 0x00ul;
class_MD5->pHash= Md5;
//数据操作
class_MD5->pHash(**,**,**);}

```

5 结束语

本文针对当前 COS 密码服务系统存在的函数使用不灵活、系统扩展性差等问题,以 Hash 算法应用为例,设计一种可重构的 Hash 算法接口,可以比较有效地解决上述问题。今

后将完善所设计的算法结构,使之操作更方便、占用存储空间更少、兼容性更好,并探索分组密码、序列密码、公钥密码等密码算法的可重构设计方法,完善 COS 密码服务系统。

参考文献

- [1] ISO/IEC. 7816-4-1996 Identification Cards Integrated Circuit(s) Cards with Contacts Part4: Interindustry Commands for Interchange[S]. 1996.
- [2] ISO/IEC. 7816-8-1999 Identification Cards Integrated Circuit(s) Cards with Contacts Part8: Security Related Interindustry Commands[S]. 1999.
- [3] Mihir B, Ran C, Hugo K. Keying Hash Function for Message Authentication[C]//Proc. of the 16th Annual International Cryptology Conference. [S. l.]: Springer-Verlag, 1996.
- [4] 陈士伟. 基于杂凑函数的消息认证码的进一步分析[J]. 密码与信息安全学报, 2010, 22(8): 19-21.
- [5] 史肖燕,熊璋,蒲菊华. 智能卡操作系统-BHCOS 的设计和实现[J]. 计算机工程, 2003, 29(2): 207-209.
- [6] 金晨辉,郑浩然,张少武,等. 密码学[M]. 北京: 高等教育出版社, 2009.

编辑 陈文

