

基于动态贝叶斯网络的手势识别*

侯亭亭 肖秦琨 杨永侠

(西安工业大学未央校区 西安 710021)

摘要: 研究了动态手势的识别技术,实现人机交互。采用 HSV 空间的肤色模型实现手势定位,光流场算法进行手势追踪,具有较好的快速性、准确性和鲁棒性。提出应用链码描述手势运动矢量,双手相对位置标量以及双手与脸的相对位置标量,将多条链码作为动态手势的特征进行提取。建立了一个动态手势识别的动态贝叶斯网络模型,将一部分手势链码作为训练样本,通过对 DBNs 的推理学习实现手势识别。实验构建动态手势识别系统,可应用于多媒体、智能电器或幻灯片控制之中。

关键词: 手势识别; HSV 空间; 光流场算法; 链码; 动态贝叶斯网络模型

中图分类号: TP39 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Gesture recognition based on dynamic bayesian networks

Hou Tingting Xiao Qinkun Yang Yongxia

(Xi'an University of Technology Electronic and Information Engineering, Xi'an 710021, China)

Abstract: The paper studies the dynamic gesture recognition technology which is used in human-computer interaction. It uses HSV color space model to achieve gestures positioning, optical flow algorithm to gesture tracking. It has well quickness, accuracy and robustness. The direction chain code is applied to describe the gesture motion vectors that is the dynamic gesture feature descriptor. In this paper, we establish an new Dynamic Bayesian Network model to finish the recognition of dynamic gesture. The chain codes are the training samples to learn the DBN model. The experiment constructs a demonstration system of dynamic gesture recognition which can be applied to multimedia, smart appliances or slide control.

Keywords: gesture recognition; HSV space; optical flow algorithm; chain code; Dynamic Bayesian Network model

1 引言

生活中,人们通过自然语言和肢体语言可以清晰表达自己的意图,自然语言会因为地域的不同受很大的影响,肢体语言是全世界通用的语言,受地域的影响比较少,其中,手势可以最自然、简单并且直观地实现人与外界一切的沟通^[1]。随着计算机技术的快速发展,人们更加希望计算机可以清晰识别自己表达的意思,作出相应的操作,利用手势进行人机交互成为研究的热点。手势识别技术已由最初的基于硬件设备输入发展为基于视觉的识别技术,其通过采集手势的视觉信息,从图像中提取手势,进行识别,实现与计算机的人机交互^[2]。目前,国内外已取得的研究成果^[3]包括:Lee、Kim 的文献中,其方法实现了识别率为 93.14%,美国密歇根州的 Cybernet Systems 技术研发公司

成功开发的一套名为 Gesture Storm 的软件系统;吴江琴、高文通过研究 ANN-HMM 混合方法,成功开发出由 18 个传感器构成的 Cyber-Glover 型号数据手套的中国手语识别系统,它采用基于视觉的技术实现单手或双手完成的动态手势识别。

2 基于 DBNs 的手势识别技术

2.1 手势定义

它定义了 10 个动态手势,其具有一定的控制性含义,其分别具有的含义如图 1 所示,其中黑点表示手势起点,有向矢量代表手势运动轨迹。它使用单目 USB 摄像头拍摄了 100 个 avi 格式的视频库,拍摄速度为 30 帧/s,像素为 640×480,其中每个含义的手势拍摄了 10 个视频。

收稿日期:2014-11

* 基金项目:国家自然科学基金(61271362)项目



图1 动态手势及其含义

2.2 基于肤色模型的手势定位

视频中,首先对手势的双手区域进行检测。手和脸部的肤色是其最明显的特征,可以较容易从周围的环境中提

$$\begin{cases} V = \max(R, G, B) \\ S = 1 - \min(R, G, B) / V \\ H = \begin{cases} \theta, G \geq B \\ 2\pi - \theta, G < B, \theta = \arccos^{-1} \left\{ \frac{[(R-G) + (R-B)]}{2\sqrt{(R-G)^2 + (R-B)(G-B)}} \right\} \end{cases} \end{cases} \quad (1)$$

2.3 基于光流法的手势跟踪

它在图像中定位出肤色位置之后,发现会有许多小面积的肤色区域出现,通过设置肤色区域面积的阈值进行去除小面积的肤色区域,获得准确的双手和脸的肤色区域,分别计算3个区域的重心,分别用 M_1, M_2, M_3 表示,即用3个重心来近似表示双手和脸的位置,如图2所示。



图2 肤色区域重心

对运动的手势进行跟踪,进而描述手势,一般实现方法有基于 Meanshift 算法^[6]、基于卡尔曼滤波模型算法、基于光流法等。它使用光流法,不仅可以定位手势和追踪手势,同时可以定量计算出手势运动矢量。光流是指空间运动物体在观测成像面上的像素运动的瞬时速度。

设 $I(x, y, t)$ 是图像序列中成像点 (x, y) 在时刻 t 的亮度,如果 $u(x, y)$ 和 $v(x, y)$ 是该点光流的 x 和 y 分量,假如间隔时间足够小,图像有恒定亮度,即运动物体的图像有相同的灰度,得到光流场约束方程:

$$I_x u + I_y v + I_t = 0 \quad (2)$$

目前计算光流矢量的方法^[7]主要有 HS 算法、Lucas-Kanade 算法和金字塔算法,其中 LK 算法相较于 HS 算法

取出来。它使用肤色模型将其与背景分离开来,其实现速度较快、效率较高。相关研究表明,不同人的肤色相差主要是饱和度与亮度的不同,而在色度上的差异并不大。在计算机视觉中,HSV 色彩空间采用色调 H 、饱和度 S 和亮度 V 3 个维度来表示颜色,可以成功将 3 个变量分别开来。

它首先将视频的每帧图像由 RGB 空间转换到 HSV 空间^[4],进而设定一个与肤色相近的 H 阈值区间,再对图像的像素点 H 值进行判断,如果输入像素在其区间内,则被认为是肤色像素。他根据式(1)^[5]实现由 RGB 空间转化为 HSV 空间。选择一个肤色区域,同样根据式(1)将其转化为 HSV 空间,根据其绘制出肤色的 H 分量分布图,得到肤色的 H 阈值,最终它设置肤色的 H 的阈值区间为: $0.01 \leq H \leq 0.1$,使用该阈值实现肤色定位。

更加准确,相较于金字塔算法原理简单、实现速度更快。它根据 LK 方法计算出相邻帧图像的光流矢量 $U = (u, v)$ 。其实现方法是假设在一个小的空间邻域 Ω 上运动矢量保持恒定,即满足光流误差为零,通过加权最小二乘法求解出 U ^[8]。

计算出光流矢量场后,对整个手势视频进行基于光流法对运动手势的跟踪,通过 $U_{k+1} = U_k + l$ 实现,其中, U_k 表示点 k 的光流矢量,表示光场矢量的平均值,其计算公式为:

$$l = \frac{1}{N} \sum_{i=1}^N f(i) \quad (3)$$

式中: $f(i) = [f_x(i), f_y(i)]^T$ 表示第 i 个光流矢量, N 表示与第 k 个点相联系的光流矢量的数量。它对 3 个重心,即 M_1, M_2, M_3 分别进行跟踪。

2.4 手势特征提取

2.3 节中的光流矢量已经定量的表示手势运动向量,但是单手和双手做的相近手势可能产生歧义,所以只用光流矢量并不能完全的表征一个手势。在 2.2 节中近似得到 3 个重心点,进而得到双手的相对位置以及左右手分别与脸的相对位置,它提出通过对矢量和标量进行量化和编码得到链码,将链码作为视频手势的特征,这样对运动手势的轨迹描述的更加准确,对手势含义表达更加清晰。

它定义相对位置矢量 $P_1 = M_1 - M_2, P_2 = M_1 - M_3, P_3 = M_2 - M_3$, P_1 表示脸与左手的相对位置, P_2 表示脸与右手的相对位置, P_3 表示左手与右手的相对位置。它定义 16 等级量化和编码规则如图 3 所示,根据该规则对每帧图像

的 M_2, M_3 的光流矢量 U_{M_2}, U_{M_3} 进行编码, 将其存入一个矩阵, 得到一个运动手势的两条特征链码 L_1 和 L_2 。

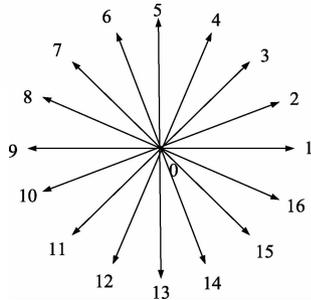


图3 光流矢量量化编码规则

它对视频中每帧图像的 P_1 和 P_2 进行8等级量化编码, 双手的相对位置标量 P_3 , 进行4等级量化编码, 其定义的规则为逆时针方向对 180° 分别进行8和4等分, 进而分别量化编码为1~8, 圆点编码为0。根据定义的该编码规则分别对每个视频的每帧图像进行量化编码, 依次分别保存到3个矩阵之中, 所以得到3条描述位置信息的链码 L_3, L_4 和 L_5 。

2.5 基于动态贝叶斯网络的手势识别

目前实现手势识别的方法有基于图像匹配算法^[9]、隐马尔科夫模型、Meanshift 算法等, 它们实现的理论较简单、系统结构简单、但是实现速度较慢、且识别效率不高, 它基于动态贝叶斯网络模型 (DBNs Model) 实现手势识别。首先建立适合系统的 DBNs, 通过对其学习, 得到相应最优参数, 进而对待识别手势特征输入系统, 进行推理, 最终识别出动态手势的含义。

假定使用一阶马尔可夫来简化运动动态^[10], 它提出了一个适合于所提取的手势特征的 DBNs 模型, 如图4所示。它有3个隐变量 (X^1, X^2, X^3) 和5个观察变量 (O^1, \dots, O^5)。2个隐变量 X^1 和 X^2 分别是左手和右手的运动模型, 每个隐变量与对应手的运动、手相对于脸部的位置特征这2个观察变量相联系。第3个隐变量 X^3 表示两只手当前的相对位置, 解决了相似手势之间的歧义问题, 与2个节点 X^1 和 X^2 有条件的相互依赖。

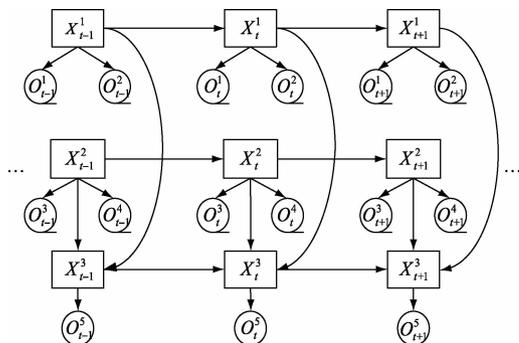


图4 构建的DBNs模型

图4中, 隐变量用方形节点表示, 观察变量用圆形节点表示。观察变量 O^1 和 O^3 表示左手和右手运动的链码, O^2 和 O^4 表示每只手和脸之间的空间关系, O^5 表示两只手之间的空间关系。在这个模型中, 隐变量 X^3 随时间变化, 即 $\dots \rightarrow X_{t-1}^3 \rightarrow X_t^3 \rightarrow X_{t+1}^3 \rightarrow \dots$, 被认为是耦合 HMM 中的一对 (X^1, X^2) 间接的捕获的两个隐变量 X^1 和 X^2 。

使用 DBNs 的一个优点是图形结构表示变量之间是条件独立的, 从而可以高效地计算变量子集的联合概率^[11]。它定义一定的初始状态概率 (π), 状态转换概率 (A), 和输出概率 (B), 将得到的链码数据 $L(L_1, L_2, L_3, L_4, L_5)$ 作为训练的样本, 通过 EM 算法^[12] 对其进行学习, 计算出最大似然估计, 即 $\hat{\theta} = \text{argmax}_{\theta} P(O_{1:T}^{1:5} | \theta)$ 从而得到该模型的最优参数 $\hat{\theta}$ 。对待识别的手势链码通过对该 DBN 的推理来实现, 即给定观察序列 (链码), 通过滤波, 预测, 滞后点平滑和脱机 (固定间隔平滑) 过程, 从而计算出隐变量 X_i 的边缘概率 $P(X_i | O_{1:t})$, 概率最大的为最终识别结果。

3 实验结果及分析

将视频库中的每个视频进行手势特征链码的提取, 将其保存到链码库之中。对每一种含义的手势分为2组, 1组为7个视频, 将它们的特征链码作为 DBNs 的训练样本, 经过学习得到最优参数, 再将另1组的3个视频作为待识别的手势, 分别将他们的特征链码分别作为 DBN 的输入序列, 进行 DBNs 推理得到边缘概率, 概率最大的为识别的手势, 经过统计得到该方法的识别率, 如表1所示。

表1 识别结果

手势含义	识别结果 正确的视 频个数	识别结果 错误的视 频个数	识别率 (%)	平均识别 时间/s
开启	2	1	66.67	2.2
关闭	3	0	100	2.0
播放	1	2	33.33	2.9
暂停	2	1	66.67	2.5
下一页	3	0	100	2.1
上一页	2	1	66.67	2.2
快进	2	1	66.67	2.6
快退	2	1	66.67	2.4
返回首页	3	0	100	2.1
进至末页	3	0	100	2.0

由表1中可以看出对于运动轨迹较简单的手势, 关闭、下一页、返回首页、快进至末页这4个手势, 该系统第识别率都可以达到100%, 但对于其他的手势识别率一般, 最终平均手势识别率达到76.67%。通过实验证明它提出的技术可以较准确的识别出来动态的手势视频, 但是它构建的视频库较小, 造成用于 DBNs 学习和推理的

特征链码较少,没有得到更精准的识别率,其完成识别的平均时间为 2.3 s,时间较短,有一定的时效性。

4 结 论

提出了一种对拍摄的动态手势视频实现手势识别的技术,构建了一种适合的 DBN 模型。在 MATLAB 编译环境中,编程实现读入视频,对帧图像进行手势和脸部的定位,计算出图像序列的光流场矢量,对运动矢量进行编码。每个运动手势使用 5 条链码作为特征表述符,将样本视频特征链码作为 DBN 模型的学习训练样本,对其进行学习得到其最优参数,待识别手势特征链码输入,通过推理实现手势识别,通过实验可以验证出该方法的鲁棒性较好,实现速度较快,可以被应用于一般的多媒体控制方面。

参 考 文 献

- [1] GROBEL K, ASSAN M. Isolated sign language recognition using hidden Markov models[J]. Computational Cybernetics and Simulation, 2008(1):162-167.
- [2] 王晓华,李才顺,胡敏,等. 服务机器人手势识别系统研究[J]. 电子测量与仪器学报,2013,27(4):305-311.
- [3] JUST A, MARCEL S. A comparative study of two state-of-the-art sequence processing techniques for hand gesture recognition[J]. Computer Vision and Image Understanding, 2009(113): 532-543.
- [4] BHUYAN M, GHOSH D, BORA P. Continuous hand gesture segmentation and co-articulation detection[A]. Lecture Notes in Computer Science: Computer Vision, Graphics and Image Processing, 2006(43):564-575.
- [5] SUK H I, SIN B K. HMM-based gait recognition with human profiles[C]. Proceedings of Joint IAPR International Workshops SSPR2006 and SPR2006 2006:596-603.
- [6] 李红岩,毛征,袁建建,等. 一种基于算法融合的运动

目标跟踪算法[J]. 国外电子测量技术, 2013, 32(12):36-40.

- [7] CHOI J, PARK H. Hand shape recognition using distance transform and shape decomposition [C]. IEEE International Conference on Image Processing, 2010:3605-3608.
- [8] AVILE 'S-ARRIAGA H, SUCAR L, MENDOZA C. Visual recognition of similar gestures[C]. Proceedings of IEEE International Conference on Pattern Recognition, 2006:1100-1103.
- [9] 刘佳,付伟平,王雯,等. 基于改进 SIFT 算法的图像匹配. 仪器仪表学报[J]. 2013,34(5):1107-1112.
- [10] MINERS B, BASIR O, KERNEL M. Understanding hand gestures using approximate graph matching[J]. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2011, 35 (2): 239-248.
- [11] SUK H I, SIN B K, LEE S W. Recognizing hand gestures using dynamic Bayesian network[C]. Proceedings of 8th IEEE International Conference on Automatic Face and Gesture Recognition,2008:1-4.
- [12] YANG H D, SCLAROFF S, LEE S W. Sign languagespotting with a threshold model based on conditional random fields[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2009,31(7): 1264-1277.

作 者 简 介

侯亭亭,1989 年出生,硕士,研究生,主要研究方向为通信与信息系统、信息检索、图像处理。
E-mail:b08030235houting@163.com

肖秦琨,1974 年出生,博士,教授,主要研究方向为图像检索、三维模型描述及学习,目标检测、先进控制理论及应用。