

基于结构分析的手势识别

朱继玉^{1),2)} 王西颖²⁾ 王威信²⁾ 戴国忠²⁾

¹⁾ (中国科学技术大学计算机科学与技术系 合肥 230027)

²⁾ (中国科学院软件研究所人机交互技术与智能信息处理实验室 北京 100083)

摘 要 该文从分割和表示(建模)两方面着手,提出了一种新颖的手势分割和整体及局部手势特征提取算法.用模糊集合来描述视频流中空域和时域上的背景、颜色、运动等信息,通过对它们执行模糊运算,分割出手;使用结构分析的方法来表示手势,根据人手不同部分在几何尺寸上的变化,从低到高逐次分析图像金字塔中各种分辨率的图像,以获取手势的整体和局部结构特征;将人手划分成手掌和手指几个部分,使用手掌和各手指的中心点的坐标和从手掌中心到所有手指的中心的方(作为手势方向)来表示一个 2D 手势.实验结果证明,该文算法具有很好的鲁棒性,对手势分割中间结果的精确性要求不高,因此能适应环境的变化.

关键词 手势识别;手势分割;模糊集合论;数学形态学;图像金字塔
中图法分类号 TP391

Hand Gesture Recognition Based on Structure Analysis

ZHU Ji-Yu^{1),2)} WANG Xi-Ying²⁾ WANG Wei-Xin²⁾ DAI Guo-Zhong²⁾

¹⁾ (Department of Computer Science & Technology, University of Science & Technology of China, Hefei 230027)

²⁾ (Intelligence Engineering Laboratory, Institute of Software, Chinese Academy of Sciences, Beijing 100083)

Abstract Camera-based user interface is one of the main style in multimodal user interface, which includes gesture interaction, pose interaction, gaze interaction, and so on. Camera-based gesture interaction is a natural, convenient way to interact with computer. Gesture recognition is the basis of gesture interaction. This paper considers both hand segmentation and gesture modeling, and presents a method of gesture segmentation and globe and local gesture features extraction. The background, motion, and color information of the video is described by fuzzy set, and is processed with fuzzy operation, and then the hand is segmented from the environment. Because different parts of the hand have different size, so Image Pyramid is used to analyze them to obtain the globe and local characteristics of the gesture. Then the hand is simply divided into palm and five fingers. The coordinates of the center of palm and every finger and the orientation from the center of the palm to the center of all the five fingers (as the orientation of the gesture) are used to denote a 2D gesture. This algorithm can be used to recognize 2D gesture. In experiment, the algorithm is applied to real-time gesture segmentation and recognition successfully, and robust enough against environment noise.

Key words gesture recognition; gesture segmentation; fuzzy set theory; mathematical morphology; graph pyramid

收稿日期:2006-03-07;修改稿收到日期:2006-10-12. 本课题得到国家“九七三”重点基础研究发展规划项目基金(2002CB312103)和国家自然科学基金(60303019)资助. 朱继玉,男,1981年生,硕士研究生,主要研究方向为人机交互技术、计算机视觉. E-mail: zhujiyu@ustc.edu; zhujiyuhappy@hotmail.com. 王西颖,男,1974年生,博士研究生,主要研究方向为图像处理、人机交互技术. 王威信,男,1970年生,博士,助理研究员,主要研究方向为信息可视化、人机交互技术. 戴国忠,男,1944年生,研究员,博士生导师,主要研究领域为人机交互技术、计算机图形学.

1 引言

在日常生活中,手势是一种常用的自然直观的交流方式. 手势识别研究的主要目的就是把手势引入到人机接口(Human Computer Interface, HCI)中,实现更加符合人类行为习惯的人机接口. 手势识别可用于虚拟现实、三维设计、临场感、可视化、医学研究、手语理解等领域中,它的解决方法对于表情识别、唇读、步态识别、时空纹理分类、视觉导航、图像拼接和基于内容的视频检索等研究都有直接推广的意义. 最近几年,基于计算机视觉的手势识别^[1]逐渐引起研究人员的关注和兴趣. 同采用数据手套(data glove)或电磁波等方式的手势识别相比,基于视觉的手势识别更加自然舒适. 数据手套和电磁波发射器等接触性传感器(contact type sensors)会削弱人机交互的自然性和灵活性.

基于视觉的手势识别过程可以分为分割、表示、识别 3 个阶段. 分割阶段需要解决的主要问题是环境因素,特别是背景的影响. 最简单而常用的方法是附加某种限制,如限制使用黑色或白色的墙壁、深色的服装等来简化背景,或要求人手戴上特殊颜色的手套或者加上标记物(marker)等来强调前景,这样就可以直接从背景中分割出人手^[2,3]. 由于人体皮肤颜色在 HSV 空间具有聚集的特性,有人应用这种特性将手势从复杂彩色背景中分割出来^[4]. 差值图像法^[5,6]首先录制背景图像,将含有手势的图像和背景图像进行比较,从而分割出手势. 手势的表示,也就是人手的模型,是指使用某种方法来表示手势. 常见的表示法是使用某个维度的特征向量,最简单的形式是用手心的三维坐标,但信息量太小. 目前出现了很多复杂的手势表示方法,如任海兵等^[7]提取出能够反应人手结构特性的特征线条,并将特征线条分割成小的曲线段,跟踪这些曲线段的运动,采用平面模型对手势图像表现的变化建模,根据各个曲线段的运动,融合成手的整体运动分量. 直板模型(cardboard)^[8]是手势识别中常用的一种表示模型,它用 3 个相连的直板来表示一根手指,是人手实际结构的一种简化形式,但它有 20 个自由度,使得算法复杂、计算量非常大.

本文从分割和建模表示两方面入手,提出了一种新颖的手势特征提取算法. 我们用模糊集合来描述视频流中空域和时域上的背景、运动、肤色等信息,通过对它们执行模糊运算,分割出精确的人手;

根据人手的结构特征——人手的不同部分在几何尺寸上变化较大,适合使用不同分辨率的图像进行分析,我们使用图像金字塔来提取手势的整体和局部特征. 其特点是:手势分割的鲁棒性较好,能较为精确地分割出人手;能提取出整体、局部、细节等不同级别的手势特征数据,方便建立各种手势模型. 本文使用整体和局部的特征,建立了一种简化的直板模型,但与直板模型相比,算法简单、运算量小,且对手势分割精度的要求较低.

本文第 2 节论述手势识别中背景、运动、颜色等信息的模糊表示法以及对它们进行的模糊处理;第 3 节论述使用图像金字塔提取手势特征的算法,使用这些特征来表示手势;第 4 节介绍了一个利用静态手势识别进行交互的演示系统;最后两节分别为实验结果和结论.

2 模糊手势分割

2.1 信息的模糊表示

我们将二维图像空间作为讨论问题的论域 E , 在该论域上定义非背景集、皮肤集和模糊手势集. 其中,非背景集描述视频序列中的背景、运动信息;皮肤集描述肤色信息;模糊手势集是由上述两个模糊集经过求交、代数积等运算得到,它综合描述视频序列中的背景、运动、肤色信息.

时空图是视频序列在空间的某一截面上因时间变化而形成的图像,即顺序取视频序列的各帧图像中的某一空间截面(一行或一列)组成的图像. 设视频序列为 $I(x, y, t)$, 其中 x 和 y 为空间坐标; t 为时间坐标. 取 $I(x, y, t)$ 的第 i 行像素构成的时空图可表示为 $\{I(x, t) \mid x = i, t = 0, 1, 2 \dots\}$. 时空图中梯度为竖直方向的点属于静止背景;梯度不是竖直方向的点为运动物体上的点. 用该性质可以分割出运动的人手.

定义 1. 背景图是指从视频序列的各帧图像中除去要分割的人手后剩下的背景相互叠加而成的图像.

从采集到的第一帧图像开始,在上面逐步叠加当前图像中去掉人手的部分,就得到一幅背景图. 叠加法则为:当时空图中某点的梯度不是竖直方向时,如果该点是人手上的一点,则背景图中该点对应的像素点仍保持原来的颜色值,如果该点不属于人手,则用该点的颜色值代替背景图中相应像素点的颜色值.

视频流中的每一帧图像都与上一帧图像的背景

图相比较,从中分割出人手,并在时空图中标记出属于人手的点,然后根据叠加法则更新时空图.为了增强系统的稳定性,增加抗干扰的能力,对叠加法则做如下修正:当时空图中的某个点连续若干次(实验中发现 3~5 次比较合适)的梯度都不是竖直方向且都不属于人手,而且这几次中该点的颜色值变化不大时,才用该点的颜色值代替背景图中相应像素点的颜色值.这样做可使背景图适应环境和光源的变化,对噪声不敏感,保持背景图的稳定性.

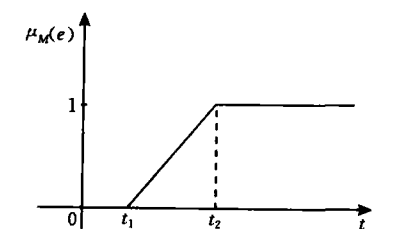
这里的背景图和差值图像法中使用的背景图不同,它是根据时空图的性质,通过统计分析得到的,只是实际背景的近似.

定义 2. 非背景集是二维图像论域 E 上的模糊集合,它描述当前帧中的各像素点不属于背景的程度.由于非背景集通常在视频中的对象移动一下时才能捕捉到,所以也叫运动对象集(moving object set),简记为 M .

当前图像中某个像素点与背景图中对应像素点的颜色值差别越大,该像素点属于 M 的隶属度就越大.设当前图为 CUR ,背景图为 BG , $CUR(x, y)$ 和 $BG(x, y)$ 分别表示当前图和背景图中的 (x, y) 点. E 中的某点 (x, y) 属于 M 的隶属度是当前图中点 $CUR(x, y)$ 与背景图中对应点 $BG(x, y)$ 的函数.当前图的非背景集 M 可用函数表示法表示如下:

$$\mu_M(x, y) = f(CUR(x, y), BG(x, y)) \quad (1)$$

通常可以用它们的 RGB 颜色值的差的平方和来定义隶属度.如



(a) 非背景集隶属度与当前帧上的点和背景图上对应点的颜色值的差的平方和的关系

$$\mu_M(x, y) = \begin{cases} 0, & t(x, y) < t_1 \\ [t(x, y) - t_1] / (t_2 - t_1), & t_1 \leq t(x, y) < t_2 \\ 1, & t(x, y) \geq t_2 \end{cases} \quad (2)$$

这里 t 为点 $CUR(x, y)$ 和点 $BG(x, y)$ 的 RGB 颜色值的差的平方和, t_1 和 t_2 是给定的两个阈值.

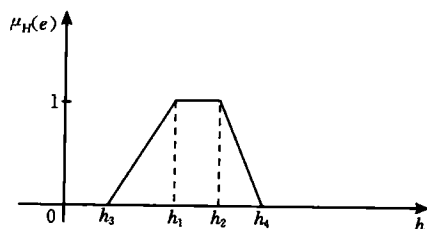
由背景图的性质可知,这里并不要求手必须运动才会包含到 M 里,只要求在算法开始时摇动一下手,使系统捕捉到手,然后即使手静止下来, M 也可以跟踪到它而不会丢失.

定义 3. 皮肤集(hue-set)是二维图像论域 E 上的模糊集合,它描述当前帧中的各像素点的颜色值接近肤色的程度.由于人手的色调非常稳定,这里使用色调值来刻画皮肤集,皮肤集简记为 H .

E 中点 (x, y) 属于 H 的隶属度是点 $CUR(x, y)$ 的颜色值的函数,该点颜色越接近手的颜色,隶属度就越大.例如,可以使用 HSV 颜色模型,定义集合 H 为

$$\mu_H(x, y) = \begin{cases} 1, & h_1 < h(CUR(x, y)) < h_2 \\ f(h(CUR(x, y))), & h_3 \leq h(CUR(x, y)) \leq h_1 \text{ 或 } h_2 \leq h(CUR(x, y)) \leq h_4 \\ 0, & h(CUR(x, y)) < h_3, \text{ 或 } h(CUR(x, y)) > h_4 \end{cases} \quad (3)$$

式中 $h(CUR(x, y))$ 表示当前图上点 (x, y) 处像素颜色的色调值, h_1, h_2, h_3, h_4 为设定的阈值.



(b) 皮肤集隶属度与当前帧中的像素点的色调值的关系

图 1 本文采用的 M 集和 H 集的隶属度函数草图

定义 4. 模糊手势集(fuzzy gesture set)是二维图像论域 E 上的模糊集,它描述当前帧中各像素点属于人手的程度,简记为 G .

G 不是直接定义的模糊集合,它是由 M 和 H 经过模糊运算得到的.计算 G 的方式有多种,如 G 可等于 M 和 H 的 F 交集、代数积或者加权和.对于不同的环境,它们具有不同的性能.因此,计算 G 的通用公式如下:

$$G = M + H + (M \cdot H) + (M \cdot H) + \dots + \dots + 1 \quad (4)$$

式中 $\alpha, \beta, \gamma, \dots$ 为正系数. $M \cdot H$ 表示 M 和 H 的代数积, $M + H$ 表示 M 和 H 的 F 交集,“+”号表示模糊集合的加权和运算.

2.2 模糊处理

由式(4)生成的模糊手势集含有很多噪声(如图 2(d)所示),为了得到更加清晰的人手,需要进行

一些后期处理. 数学形态学中的运算和变换都是以集合论为基础的, 我们把它推广到模糊集合上. 下面

的讨论中, 我们只考虑 E 和 S 是有限论域的情况.

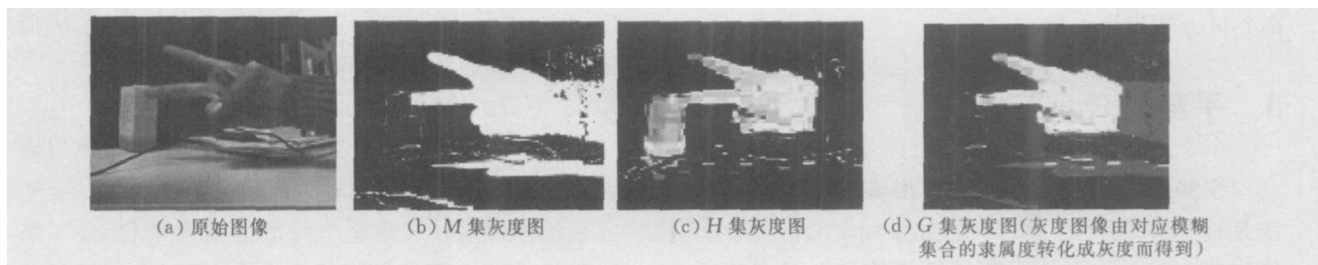


图 2

定义 5. 论域 S 平移到点 z , 记为 $(S)_z$, 如果

$$(S)_z = \{c \mid c = a + z, a \in S\} \quad (5)$$

$(S)_z$ 是由 S 经过平移得到的新论域, $(S)_z$ 中的元素和 S 中一一对应. 如果将 z 看成是一个向量, 则它表示由 S 中的每一点平移向量 z 后得到的空间.

定义 6. 设 B 为论域 S 上的一个模糊集, 定义 $(S)_z$ 上 B 的 z 平移模糊集 $(B)_z$ 为

$$\mu_{(B)_z}(e) = \mu_B(e - z), \quad e \in (S)_z \quad (6)$$

$(B)_z$ 是定义在 $(S)_z$ 上的模糊集, $(S)_z$ 的每一点在 $(B)_z$ 中的隶属度等于 S 上的对应点在 B 中的隶属度, 即 $(B)_z$ 和 B 是同构的.

在模糊数学中, 设 S 是 E 上的子论域, A 为论域 E 上的模糊集, A 在 S 上的投影 A_S 可表示为

$$\mu_{A_S}(e) = \begin{cases} \mu_A(e), & e \in S \\ 0, & e \notin S \end{cases} \quad (7)$$

定义 7. 设 A 为论域 E 上的模糊集, S 是某个包含原点的论域, B 为 S 上的模糊集, 则 A 被 B 膨胀记为 $A \oplus B$, 如果

$$\mu_{A \oplus B}(e) = \begin{cases} \mu_A(e) \vee \mu_{(B)_e}(e), & e \in \{x \mid ((B)_x)_E \subseteq (S)_x\} \\ \mu_A(e), & e \in \{x \mid ((B)_x)_E \subseteq A_{E \setminus (S)_x}, x \in E\} \\ 0, & e \in E - \{x \mid ((B)_x)_E \subseteq A_{E \setminus (S)_x}, x \in E\} \end{cases} \quad (8)$$

A 被 B 腐蚀记为 $A \div B$, 如果

$$\mu_{A \div B}(e) = \begin{cases} \mu_A(e), & e \in \{x \mid ((B)_x)_E \subseteq A_{E \setminus (S)_x}, x \in E\} \\ 0, & e \in E - \{x \mid ((B)_x)_E \subseteq A_{E \setminus (S)_x}, x \in E\} \end{cases} \quad (9)$$

这里 $A_{E \setminus (S)_x}$ 表示 A 在 $E \setminus (S)_x$ 上的投影. 运算中使用的 B 称为模板.

模糊集合的开、闭运算分别称为模糊开运算和模糊闭运算, 它们的定义同数学形态学中开、闭运算

的基本定义, 即模糊开运算是先对模糊集先腐蚀后膨胀, 模糊闭运算是先对模糊集先膨胀后腐蚀. 模糊开运算和模糊闭运算组成了一个对模糊集合进行去噪求精的处理器.

适当的选取模板可以提高处理结果的质量. 图 3 是一个 3×3 的模板和一个 5×5 的模板实例. 一般可以选用中心小的模板做模糊开运算, 中心较大的模板做模糊闭运算. 当使用中心较小的模板进行模糊开运算时, 如图 2(d) 所示的模糊集中暗处的“亮斑”由于中心较大而周围较小会在腐蚀操作中消失, 而亮处的“暗斑”很多都可以保留下来. 使用中心较大的模板进行模糊闭运算, 一方面可以使手势区域在模糊开运算中可能出现的“裂痕”得到弥合, 另一方面使手势区域变“亮”, 增加了模糊手势集的清晰度.

0.8	0.8	0.8	0.8	0.8
0.8	0.8	0.8	0.8	0.8
0.8	0.8	1.0	0.8	0.8
0.8	0.8	0.8	0.8	0.8
0.8	0.8	0.8	0.8	0.8

图 3 3×3 和 5×5 的模板

模糊手势集经过模糊开运算和模糊闭运算处理后, 取该集合内隶属度大于某个阈值的点, 就得到了精确的人手, 然后就可以使用各种对二值图像提取边界的算法提取手势轮廓; 也可以通过形态学的运算求得. 求清晰手势集和手势轮廓的完整过程用公式表示如下:

$$\begin{cases} G = ((M \oplus H) \oplus (M \cdot H)) \circ B_1 \cdot B_2, \\ \quad \quad \quad + + + = 1 \\ G_1 = \{x \mid \mu_G(x) > \} \\ L = \{x \mid x \in (G \oplus B_3) - G\} \end{cases} \quad (10)$$

这里 G 表示模糊处理后的模糊手势集; G_1 为清晰手势集; L 表示手势轮廓; B_1, B_2 分别为模糊开运算和模糊闭运算使用的模板; B_3 为 3×3 的处理二值图

像的模板;式(10)中的 \oplus 表示二值图像的膨胀运算符; \circ 和 \cdot 分别表示模糊开和模糊闭运算符;式中 $\alpha, \beta, \gamma, \delta$ 均为正系数.

3 手势的表示与描述

手势的表示(建模)方法有很多种,常见的表示法是将手势表示成一个向量,如使用手心的坐标、轮廓中曲率较大的点等手势特征组成的向量来表示.在手势识别中,以轮廓中曲率较大的点作为特征点的鲁棒性较低,一方面是由于手在运动过程中存在变形,同一部分在相邻两帧图像中的曲率往往存在较大差异;另一方面是因为这种方法对手势分割的精度要求较高,很难适用于复杂背景、复杂光照的条件.另一种方法是使用直板模型,它是人手物理结构的一种简化形式,能有效地描述人手的 3D 特征,但同时也有算法复杂、对手势分割的精度要求高等缺点.

本文将人手划分成手掌、手指和轮廓三部分,使用结构分析的方法来表示手势.这三部分描述手势

的角度和层次不同,是一种从整体到局部逐渐细化的表示法:手掌可以作为一点(手心)或一块来表示整个手;而手指描述了较为详细的人手信息,轮廓描述的是手势的更加精确的局部信息.

3.1 手势图像金字塔

图像金字塔是多分辨率分析中的一种有效但概念简单的结构.一幅图像的金字塔是一系列以金字塔形状排列的分辨率逐步降低的图像集合.金字塔的底部是待处理图像的高分辨率表示,塔顶是低分辨率的近似.一幅手势图像的金字塔,从塔顶到塔底移动时,相当于我们由远而近地观察人手,开始只看到模糊的一点或一块,然后可以分辨出手指,最后可以分辨出清晰的轮廓.图 5 是图 4(a)所示的由模糊手势集转化成的手势图的金字塔中相邻的三级.分辨率最小的一幅图中,整个人手只能看作一个小矩形.从它的放大图中可以看出,手指的宽度只有两个像素,它适合于确定出人手在图像中所处的基本范围和手掌的大体位置;分辨率次小的图可以区分出手指的骨架;分辨率最大的图像已经可以分辨出较为清晰的手势轮廓了.



图 4

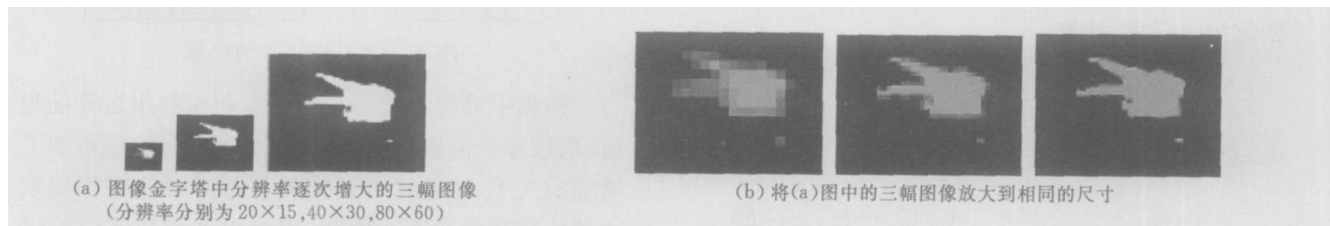


图 5

本文采用如下的方法生成手势图像的金字塔.金字塔的基础级图像,是将模糊手势分割法求出的模糊手势集中每点的隶属度乘以 255 转换而成的灰度图,在基础级图像的基础上,生成其它各级近似图像.设手势图像金字塔的第 i 级近似的分辨率为 $M \times N$,则它的上一级(第 $i-1$ 级,分辨率更低的一级)的分辨率为 $\lceil M/2 \rceil \times \lceil N/2 \rceil$.第 i 级中点 (x, y) 的灰度值取第 $i+1$ 级中点 $(2x, 2y), (2x+1, 2y), (2x, 2y+1), (2x+1, 2y+1)$ (如果它们存在的话)

的平均灰度值.将第 i 级中的点 (x, y) 称为第 $i+1$ 级中点 $(2x, 2y), (2x+1, 2y), (2x, 2y+1), (2x+1, 2y+1)$ 的父结点;将第 $i+1$ 级中点 $(2x, 2y), (2x+1, 2y), (2x, 2y+1), (2x+1, 2y+1)$ 称为第 i 级中点 (x, y) 的子结点;将子结点的子结点称为间接子结点,反之为间接父结点.这样第 i 级近似中点 (x, y) 的子结点或间接子结点在以下各级近似中形成一个矩形,点 (x, y) 的性质代表了这个矩形的整体性质.

3.2 金字塔中点的分类和判定

对手势图像金字塔中每级近似图像中的点, 根据其本身和它邻域的性质划分为三类: 内点、边界点、外点。内点是指属于人手的点; 外点是属于背景的点; 边界点是即不能判定为内点也不能判定为外点的点。

内点的充分条件: 如果某点周围八邻域的点都不满足外点的必要条件, 则该点一定是内点。

内点的必要条件: 该点的灰度值大于某个给定的阈值。

外点的充分条件: 如果某点周围八邻域的点都不满足内点的必要条件, 则该点一定是外点。

外点的必要条件: 该点的灰度值小于某个给定的阈值。

若第 i 级近似图像中的某个点已经判定为内点 (或外点), 则该点的子结点可以直接判定为内点 (或外点), 称为继承; 如果该点为边界点, 则它的子结点需要重新进行判断。首先将金字塔的塔顶判定为一

个边界点, 然后向下逐级判断, 内点或外点向下继承, 边界点的子结点则做进一步判断。这样手势图像金字塔就形成了一个手势边缘金字塔, 特征提取的过程也就变成了边缘求精的过程。

3.3 手指的判定

在手势边缘金字塔求精的过程中, 人手和背景逐渐分割开来, 最终在这两者之间形成了一条闭合连通的轮廓线。但由于手掌和手指上的点都是内点, 所以需要对内点做进一步的分类才能将手掌和手指分开。我们把判定为边界的点涂成蓝色, 判定为手掌的点涂成绿色, 判定为手指的点涂成红色, 将整个分类过程形象化为一个对手势图像金字塔进行染色的过程。

设在第 i 级近似中首次判定出内点, 则将这些内点都涂成绿色, 在第 i 级以下的各级近似中, 绿色结点直接向下继承。设 p 点是在第 j ($j > i$) 级判定出的内点, 它的父结点为 q , 若 q 的八邻域中有绿色的点, 则 p 涂成绿色; 否则 p 涂成红色。

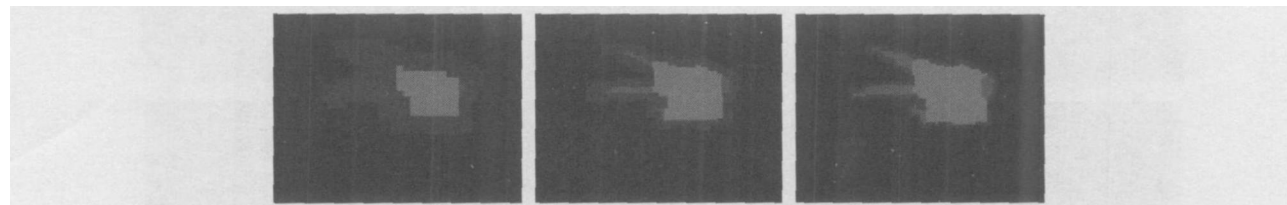


图 6 图 4(a) 的染色过程的示意图 (这里是金字塔中相邻的三级, 已将它们放大成同尺寸的图像)

染色结果中有时会在手掌棱角处出现红色区域, 即出现虚假手指。一种解决方法是: 在染色结束后, 统计所有红色区域的周长、面积、致密性 (周长的平方/面积) 等描绘子, 增加相互位置关系、形状等约束条件, 将不满足条件的红色区域涂成绿色。

成功分割出手掌、手指和轮廓三部分后, 统计各部分的各种性质, 如手掌的中心、各手指的中心和指向、轮廓的各种特征值等, 使用它们就可以准确地描述手势了。

4 手势交互演示系统

为了检验本文算法的性能, 我们完成了一个利

用静态手势识别进行交互的演示系统。该系统利用手心坐标 (x, y) 、手指个数 n 、手指指向组成的特征向量 (x, y, n, \cdot) 来识别手势, 实现了 2D 平面里的抓放、平移、旋转、状态转换等操作功能。手心设为手掌的中心, 定义手势的指向为从手掌中心指向所有手指整体中心的方向。

该系统分析摄像机实时捕获到的每一帧图像, 对手势进行实时识别。我们根据手指的个数定义操作种类: 伸开手为空闲状态或放开操作, 握上拳头为抓取操作。处于抓取状态的对象将随手一起移动, 伸出两根手指, 被抓住的对象将执行其默认的操作 (类似于用鼠标左键双击), 伸出一根手指, 被抓住的对象将随手势的指向变化而旋转。演示系统的效果如图 7 所示。

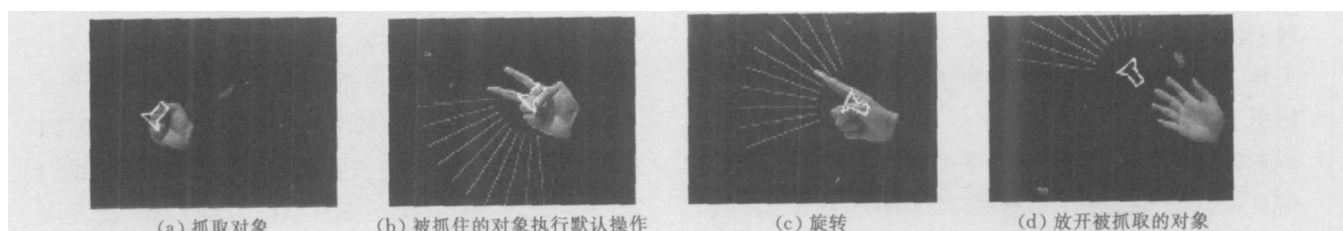


图 7 利用静态手势识别进行交互的演示系统

5 实验结果

为了检验本文算法的准确性和实时性,我们选取了几组不同背景和光照条件下的手势视频序列进行测试. 测试过程中,式(10)中的参数分别取 $\mu = 0.44$, $\sigma = 0$, $\gamma = 0$, $\alpha = 0.56$, $\beta = 0.8$, $\delta = 0.8$, 非背景集和皮肤集的隶属度函数的各个阈值分别取 $t_1 = 0$, $t_2 = 1000$, $h_1 = h_2 = 33$, $h_3 = 13$, $h_4 = 53$.

图8是一段视频序列的模糊分割和结构分析的结果. 图中黄色的小盒和人手的色调值非常接近,由图2(c)也可以看出单靠肤色是很难将它和人手分开的,而由于它一直保持不动,本文算法可以非常精确地排除它的干扰. 后面4幅图为前面对应图像的手势结构分析的结果.

对图8所示的视频序列中连续55帧含有手势的图像进行模糊分割的准确率达到96.4%. 该算法对环境的变化具有很好的抗干扰性. 图9是背景中有人走动时实时完成的手势分割和结构分析的结果,从中可以看出,由于人的走动造成环境和光线的剧烈变化,从而影响了分割的准确性,但使用手势图像金字塔进行结构分析的过程对分割结果具有一定的抗干扰性,仍然取得了较好的效果.

在实时性方面,我们的实验结果显示,当实时采集 320×240 的真彩色图像时,本文算法在800MHz、384MB的PC机上,处理一帧图像用时不超过0.42s. 当已捕捉到手势再处理下一帧图像时,可以引入预测机制来改进本文算法,从而可以大大提高运算速度.

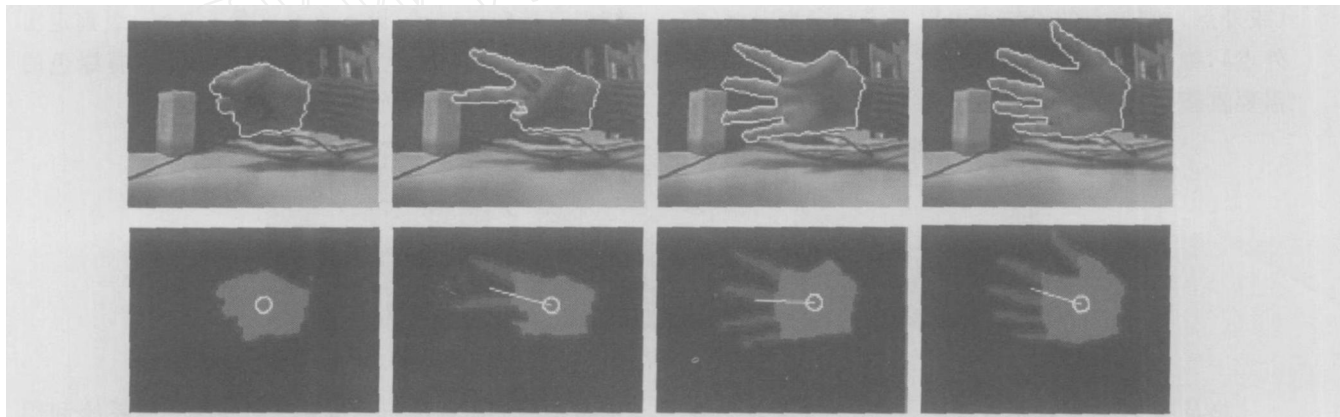


图8 图2所示视频流的分割和结构分析结果(由左向右分别为视频流的第20,30,35,39帧 其中白色的圆圈中心是手心,白色短线是手势指向,下图同)

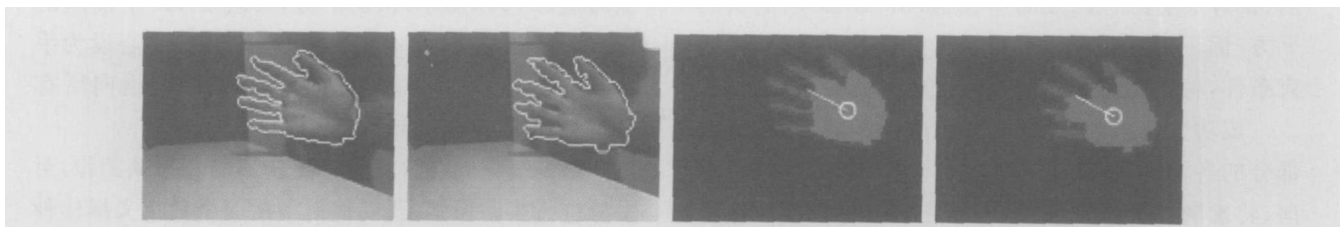


图9 背景发生变化时,使用本文算法实时完成的手势分割和结构分析的结果

6 结 论

本文使用模糊集合的概念表示视频流时域和空间上的不同信息,以模糊运算作为信息加工处理的工具,使手势分割算法建立在坚实的数学基础之上. 使用手势图像金字塔进行手势结构分析是一种多分辨率分析的方法,能够取得手势的总体和局部变化信息,增加了可识别的手势类型的种类. 实验结果表明,该算法具有很好的鲁棒性,对手势分割中间结果

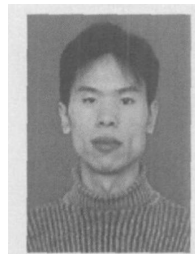
的精确性要求不高,因此能适应环境的变化.

本文算法适用于静态手势识别,且本文算法以及其中的处理思想,尤其是信息的模糊表示和处理、多分辨率分析等,对于动态手势识别和多目的3D手势识别,都具有极重要的推广意义.

该算法还有许多需要继续改进之处. 算法中有一些系数和常量,如基本肤色的色调值等,对不同的环境,适当调整这些参数是有益的,因此,我们下一步的工作是实现这些参数的自动选择,使算法能够自适应环境的变化.

参 考 文 献

- 1 Wu Ying, Huang Thomas S.. Vision-based gesture recognition: A review. In: Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction, Gif-sur-Yvette, France, 1999, 103 ~ 115
- 2 Volkert B., Stephen V., Mark B., Andy C.. FingARtips: Gesture based direct manipulation in augmented reality. In: Proceedings of the 2nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia, 2004, 212 ~ 220
- 3 Shahzad Malik, Joe Laszlo. Visual touchpad: A two-handed gestural input device. In: Proceedings of the ACM International Conference on Multimodal Interfaces (ICMI '04), State College, PA, USA, 2004, 289 ~ 296
- 4 Jiang Wei, Chen Yuan-Fei, Kong Yong, Li Wen-Ming. A new method for dividing hand image for complex background. Journal of Shandong University (Engineering Science), 2003, 33(4): 410 ~ 412(in Chinese)
- 5 Julien L., Francois B.. Visual tracking of bare fingers for interactive surface. In: Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology, Santa Fe, NM, USA, 2004, 119 ~ 122
- 6 Christian von Hardenberg, Francois Berard. Bare-hand human-computer interaction. In: Proceedings of the ACM Workshop on Perceptive User Interfaces, Orlando, Florida, USA, 2001, 1 ~ 8
- 7 Ren Hai-Bing, Xu Guang-You, Lin Xue-Yin. Hand gesture recognition based on characteristic curves. Journal of Software, 2002, 13(5): 987 ~ 993(in Chinese)
(任海兵, 徐光祐, 林学闻. 基于特征线条的手势识别. 软件学报, 2002, 13(5): 987 ~ 993)
- 8 Lin John, Wu Ying, Huang Thomas S.. Capturing human hand motion in image sequences. In: Proceedings of the Workshop on Motion and Video Computing (MOTION '02), Orlando, Florida, 2002, 99 ~ 104



ZHU Ji-Yu, born in 1981, M. S. candidate. His major research interests include human-computer interaction, computer vision.

WANG Xi-Ying, born in 1974, Ph. D. candidate. His

major research interests include computer graphics, human-computer interaction.

WANG Wei-Xin, born in 1970, Ph.D., professor. His major research interests include information visualization, human-computer interaction.

DAI Guo-Zhong, born in 1944, professor and Ph.D. supervisor. His major research interests include human-computer interaction, computer graphics.

Background

This research is supported by the National Basic Research Program (973 Program) of China under grant No. 2002CB312103 and the Key National Natural Science Foundation of China under grant No. 60033020.

This research is part of the project named Theory and Method of Natural and Harmonious Human-Computer Interaction supported by the National Basic Research Program (973 Program) of China under grant No. 2002CB312103. This project is focused on the important theories and methods of next generation user interface, especially related to virtual reality. The contents of this project mainly include pen-based

user interface and camera-based user interface, which are all important UI style in multimodal user interface. Currently some applications have been built based on this project.

The research group's main interests include hand gesture image understanding and analysis, and gesture interaction technology based on camera. This paper introduces Fuzzy Set Theory and Multi-resolution Analysis into gesture recognition, and proposes a novel algorithm about gesture segmentation and recognition. The research result of camera-based gesture recognition can be applied for future research on camera-based 3D interaction, etc.