

一种新的手势建模方法及其在手势识别中的应用

杨筱林 姚鸿勋

{yangxl, yhx}@vilab.hit.edu.cn

哈尔滨工业大学 计算机科学与工程系, 哈尔滨, 150001

摘要: 随着计算机性能的提高和人机交互技术的发展, 手势识别越来越受到人们的重视。但由于人手是复杂形体, 现有的方法对运动过程中手的形变的描述不够充分。本文从视觉角度提出了一种新的手势建模的方法—多尺度形状描述子。从分析手的基本形状入手, 利用了圆形的轴对称和中心对称的几何特点, 从多个尺度对手势进行形状描述, 在一定程度上解决了手势的精细区分问题。

关键字: 手势建模, 手势识别, 手语, 形状描述子

1 引言

近年来计算机的迅速普及, 使用户对计算机使用的方便程度要求越来越高, 随着计算机性能的逐步提高和各领域对计算机使用的不断深入, 计算机在现代社会中的影响迅速扩大, 高性能、高智能和高可靠性被普遍认为是当前计算科学发展的主要方向。从而, 人与计算机的交互活动越来越成为人们日常生活中的一个重要组成部分[1]。手势是一种自然、直观、易于学习的人机交互手段。以人手直接作为计算机的输入设备, 人机间的通讯将不再需要中间的媒体, 用户可以简单地定义一种适当的手势来对周围的机器进行控制。手语作为一种高度结构化的手势语言, 对于基于手势的人机交互技术的研究有着非凡的意义。手语(势)识别的研究涉及机器人学、空间几何学、心理学、生理学、人工智能、模式识别、概率统计学、计算语言学、计算机视觉、图像处理和分析等多个学科, 更是模式识别、人工智能和计算机视觉学科的经典案例之一。由于手语(势)识别研究具有的跨学科性, 加上手语(势)本身具有的多样性、多义性、以及与个体有关的时间和空间上的差异性, 并且人手是复杂的变形体的特点, 使得该项课题具有很大的挑战性。

为简化前端的特征提取过程, 很多研究人员采用数据手套(dataglove)或者电磁波等方法。然而手上带有数据手套或者电磁波发射器等接触性传感器(contact type sensors), 但这些方式不利于自然灵活的人机交互[2]。

基于视觉的手势识别, 从视觉通道获取信号, 是一种更为自然的人机交互方式。但由于人手是复杂的变形体, 无法用简单的模型加以表示。目前的解决办法通常有以下几种方式, 一种是要求人手佩戴特殊的颜色手套, 或标记物来强调前景, 可以简单地将人手分割出来, 然后利用 PCA 等技术分析手的局部特征[3]; 另一种是佩戴多种颜色标注的手套来确定人手的各个部位, 利用颜色特征来对手势进行建模[4]。在徒手方面, MIT 的 Starner 等[5]以椭圆为人手建模, 以椭圆中心位置、椭圆长短半轴长以及椭圆长半轴与坐标系 x 轴夹角这些参数的变化轨迹作为手势特征, 但这种方法没有描述手形的精确信息; 还有研究者利用手的边缘轮廓信息, 采用模板匹配的方法对每种手势进行识别, 该方法的缺点是不具备物体的旋转不变性和空间尺度的不变性描述。尤其在手语识别中, 很多词汇具有相同的运动特征, 它们的区别仅在于手的形状。因此, 一种既不需要对手加标志物限制, 又能够比较精确的描述人手特征的方法对于基于视觉的手势(语)识别有非凡的意义。

本文的主要贡献在于提出了一种对手的形状进行描述的方法—多尺度形状描述子, 它综合考虑了人手的特点, 并具有较好的尺度和旋转不变性。

2 多尺度形状描述子

人的手是由手指和手掌两部分构成，五个手指有弯曲和伸展状态，手又有不同的朝向信息。多尺度的形状描述子就是针对人手的上述形变提出来的。

圆形是具有轴对称和中心对称的几何形状，在描述物体方面也具有上述优点。该多尺度形状描述子将圆形所特有的性质应用到手形的描述上，通过由外到内多尺度的计算，可以得到手的精确形状，同时减少对手势描述的冗余信息。

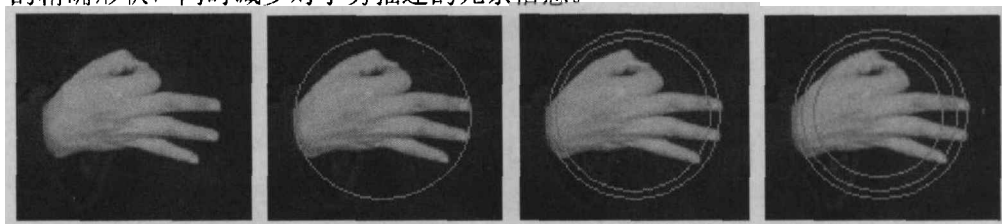


图 1 多尺度形状描述子示意图

(a 原图片 b 第一层次描述子 c 二个层次描述子 d 三个层次描述子)

每一尺度的描述子如下表示：

$$width_i = \sum_{\theta \in R} \#\{p: p[r, \theta] > Y\} \quad (2-1)$$

$$\theta_k = \{\theta: p[r, \theta] > Y \ \&\& \ p[r, \theta + 1] < Y\} \quad (2-2)$$

$$\theta_{k'} = \{\theta: p[r, \theta] < Y \ \&\& \ p[r, \theta + 1] > Y\} \quad (2-3)$$

这里 $width_i$ 表示每一肤色区域在该描述子上的分布大小， p 为 R 区域内的像素， r 为描述子在该尺度上的半径， θ_k 和 $\theta_{k'}$ 分别为肤色和非肤色区域的交界点在极坐标系下的极角， $p[r, \theta]$ 为在某颜色空间内的亮度值， Y 为亮度阈值， R 为 θ 的连续区域。

有了上面的两个最基本的表示，就可以根据需要来提取每一手势在该层次上的其它特征。这样随着尺度的变化，每一个手势的形状特征就变得精确了[6]。

3 多尺度形状描述子在手势识别中的应用

为简化分割过程，我们采用了与肤色区别较大的简单背景，目的在于验证该多尺度形状描述子对形状描述的有效性。

整个系统的工作过程如下：

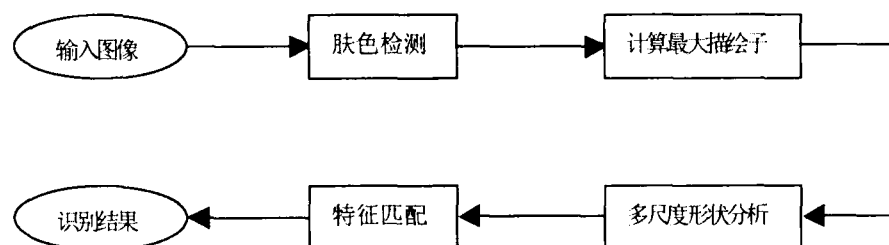


图 2. 手势识别系统工作流程

3.1 肤色检测

肤色检测实际上是根据皮肤的颜色,亮度特征,从整个图像的视野中找出属于肤色的区域。由于本实验所采的图像是24位真彩图,所以可以利用R,G,B信息进行肤色提取。同种颜色在不同分布的光照下,亮度是不同的,但是色的感觉基本上是保持恒定的倾向。所以,我们将它变换到另一个色彩系统Y,U,V空间。其中的Y指的是明视度,U和V指的是色调,R,G,B与Y,U,V的对应关系如下:

$$\begin{pmatrix} Y \\ U \\ V \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (3-1)$$

$$Ch = \sqrt{|U|^2 + |V|^2} \quad (3-2)$$

$$\theta = \tan^{-1}(|V|/|U|) \quad (3-3)$$

把U和V看作平面上的两个互相正交的矢量,色度信号(即U和V之和)是一个二维矢量。公式(3-1)为空间变换公式,公式(3-2)为模值公式,公式(3-3)为相位角公式。根据色度恒定这一特点,我们可以由 θ 的范围来确定肤色的分布。

3.2 最大形状描绘子的计算

由上面所得的肤色区域,可以计算肤色区域的中心,由此可以得到最大的描绘子的中心。

$$m_{00} = \int \int_{-\infty}^{\infty} f(x,y) dx dy \quad (3-4)$$

$$m_{10} = \int \int_{-\infty}^{\infty} xf(x,y) dx dy \quad (3-5)$$

$$m_{01} = \int \int_{-\infty}^{\infty} yf(x,y) dx dy \quad (3-6)$$

$$\bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}} \quad (3-7)$$

给定所给图像 $f(x,y)$ 在每一点 (x,y) 处的灰度级是在点 (x,y) 的“质量”, (\bar{x}, \bar{y}) 是 $f(x,y)$ 的重心点。

求得中心以后,在根据对肤色像素点的统计来估计 r 值,使以 (\bar{x}, \bar{y}) 为圆心,以 r 为半径的圆能够包括所有的肤色区域,这样就奠定了多尺度形状分析的基础。

3.3 多尺度形状描述子的应用

多尺度体现在,沿着最大形状描绘子的半径 r 仍然可以得到多个 $r' < r$ 的描绘子,在本

文中，我们采用了 4 个尺度，由公式(2-1)(2-2)(2-3)得到以下对于每个尺度的特征组合：

$$\frac{width_i}{360}, i, \frac{\theta_k - \theta_{k'}}{2}, \text{每个 } \frac{\theta_k - \theta_{k'}}{2} \text{ 之间的角度 (其中 } \theta_k, \theta_{k'} \text{ 经过了排序)}$$

由此，每一尺度上的肤色与非肤色区域的分布特征，包括手指与手掌的位置均非常明显的表现出来。

当然，由于存在 i 的数目不同，而导致每个 $\frac{\theta_k - \theta_{k'}}{2}$ 之间的角度数目也不同，所以，可以将特征分为两个层次，第一个层次为 $\frac{width_i}{360}, i, \frac{\theta_k - \theta_{k'}}{2}$ ，第二个层次为 i ，每个 $\frac{\theta_k - \theta_{k'}}{2}$ 之间的角度。

3.4 识别

由于我们的目标是寻找一个有效的特征提取方法，在识别阶段，我们仅采用简单的特征匹配方法。

在训练阶段，取得个参数的平均值和方差，来代表该手势的最一般特征。测试阶段，我们选用最近邻分类方法，求得各测试样本与训练样本的距离，距离最小者则判别为哪一类。

$$d_i = \|X - \omega_i\| = \sum_{j=1}^{M_i} (x_{ij} - \omega_{ij})^2 \quad (3-8)$$

$$d_k(X) = \min_i (d_i(X)) \quad (3-9)$$

满足 (3-9) 式，则 $X \in k$

在本文中，由于采用了两个特征层次，所以对两个层次进行了匹配。

4 实验结果及分析

我们以中国手语中的手指语字母作为训练和测试集。用 Sony Digital Camera 对每个手势采集十个，其中的 5 个作为训练样本，5 个作为测试样本。

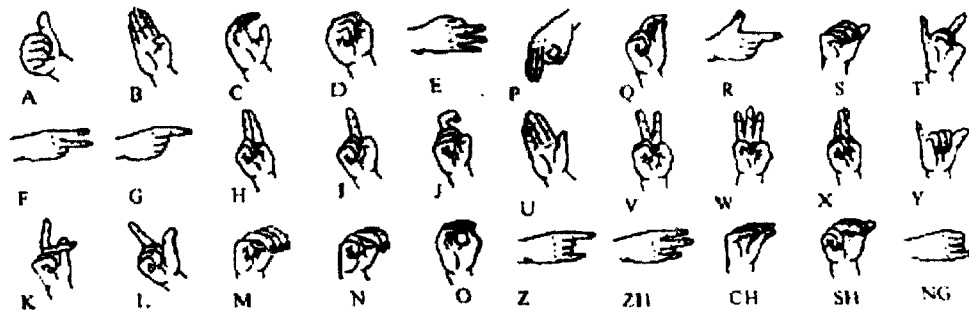


图 3. 训练测试集—中国手指语字母

由表 1 的识别结果我们可以看出，对其中的 26 个字母识别都有很好的效果，总体的识别率可以达到 94.6%。但对于 D,M,N,由于其形状极为相似，所以在识别过程中没有考虑这 3 类，但今后可以将它们作为一类进行识别。而 X 同 H 在 4 个尺度上的形状描述子中也具有相似的特征，目前没有把它加入到实验当中。

传统的基于徒手的手势识别方法，识别的手势集合比较小，而且没有对手势进行精确的建模，因此识别率都比较低，如朴素的模板匹配识别率仅为 81.3%，该方法无论在识别手势集合大小上还是识别率上，都较以前的方法有较大的进步。

手势	A	B	C	E	F	G	H	I	J	K
识别率	100	100	100	100	100	100	100	100	80	100
手势	L	O	P	Q	R	S	T	U	V	W
识别率	100	40	100	60	100	100	80	100	100	100
手势	Y	Z	ZH	CH	SH	NG	D	M	N	X
识别率	100	100	100	100	100	100	X	X	X	X

表 1. 识别结果

5 结论及展望

本文提出了一种新的手势建模方法—多尺度形状描述子。并在中国手语中 30 个手指语字母集合上进行了实验，实验结果说明：

1. 该方法能够较好的进行手势的建模。
2. 由该描述子自身的特性可以看出，该方法对在平面内旋转的物体的特征提取也有较为鲁棒的特性，这一性质可以在今后的实验中得到验证。
3. 由于该方法特征描述的精确性，可以同隐马尔可夫模型（Hidden Markov Model）相结合，进行动态的手势（语）识别。
4. 由特征选取的特点，可以进一步考虑非特定人的特征提取工作，并将其应用于非特定人的手势（语）识别。

不足之处：

该方法目前依赖于肤色提取的准确性，对背景限制比较严格，今后要在分割上做工作，以获取较好的手形分割，可以考虑同边缘信息相结合的办法。

在今后的工作中，打算进一步将该方法应用到手语识别当中，期待能够实现中等词汇规模以上手语识别。

参考文献

- [1] V. Pavlovic, R. Sharma, T. Huang. "Visual Interpretation of Hand Gestures for Human Computer Interaction: A Review". IEEE PAMI, 19(7), pp.677-695, 1997.
- [2] 任海兵, 徐光祐, 林学闯. "基于特征线条的手势识别". 软件学报, Vol. 13, No. 5, pp.987-993, 2002
- [3] Jiangwen Deng and H.T. Tsui. "A PCA-MDA Scheme for Hand Posture Recognition". Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002.
- [4] K. Grobel and M. Assan. "Isolated Sign Language Recognition Using Hidden Markov Models". Proceedings of the International Conference of System, Man and Cybernetics, pp.162-167, 1996.
- [5] T. Starner, J. Weaver and A. Pentland. "Real-time American Sign Language Recognition Using Desk and Wearable Computer Based Video". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.20, Dec, pp.1371-1375, 1998.
- [6] Serge Belongie, Jitendra Malik and Jan Puzicha. "Shape Matching and Object Recognition Using Shape Contexts". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24.No.24, pp.509-522, April 2002.