

基于分块局部纹理描述的手势识别

柴秀娟^{①*} 汪孔桥^{**}

(* 北京邮电大学信息工程学院 北京 100088)

(** 诺基亚研究中心系统研究中心 北京 100176)

摘 要 提出了一种用于解决手势识别问题的新的纹理描述,该描述基于一种有效的纹理编码模式——局部二值模式(LBP)。该方法对图像中的每一个像素进行 LBP 变换,提取其邻域内的亮度变化以作为特征来刻画像素点。为了更好地保持图像的空间结构信息,采用一种分块策略以获取对手势的更加鲁棒的描述。该策略将每一小块对应的局部直方图连接起来,形成一个完整的向量,即生成具有很强描述能力的纹理特征。从本质上讲,该方法有效地将全局结构信息和局部纹理信息结合起来,共同作用得到非常好的结果。在 Jochen Triesch 数据库的实验结果表明,与其他方法相比,这种基于分块局部纹理描述的手势识别方案具有显著的优势。

关键词 局部纹理描述,手势识别,局部二值模式(LBP),直方图

0 引 言

手势识别有很大的商业应用潜力,在日常生活中的作用越来越重要,尤其体现在人机交互方面^[1-3]。手势识别的主要应用有数字家电的命令控制,文件或网页的浏览导航,游戏控制以及线条输入等。作为最自然的交互方式之一,基于视觉的手势识别逐渐引起了广泛关注并取得了很大进展。目前,性能突出的手势识别策略可以在限制环境条件下取得很高的精度,然而当面对复杂的背景条件以及未限制的光照条件时,识别的精度就会急速下降。针对手势识别问题,研究人员提出了诸多方法,其中有被广泛研究的基于外观的方法。按照用于描述手势的特征,基于外观的建模方法可分为两个范畴:基于几何特征的方法和基于纹理的方法。几何特征包括轮廓特征、形状特征等。文献[4,5]通过轮廓特征的匹配得到了手势类别的分类结果,但其性能易受手分割及轮廓提取的精度限制。除轮廓外,还可从图像中提取圆或椭圆等形状特征对手势进行描述,同时尺度、位置、方向等信息也被用来进行分类决策^[6,7]。如在文献[7]中,作者使用了特征相似度图(feature likelyhood map, FLM)来近似多尺度空间

中的点的相似性,然后通过寻找大的 FLM 来检测得到圆形特征(blob)和山脊线特征(ridge),从而进行手势识别。基于纹理的方法也是一种主流方法,并且可以得到更好的识别精度。纹理描述需要对图像进行变换,生成描述子,如直方图变换^[8,9]、小波变换^[10]或一些其他的变换^[11]。相对来讲,基于整个手部区域的纹理描述比基于几何特征的描述更加鲁棒,识别结果也更理想。

近几年,有强大的纹理描述能力的局部二值模式(local binary pattern, LBP)在纹理分类、图像匹配及检索领域中得到了广泛应用^[12-14]。本文将 LBP 引入到手势描述中,并提出了一种基于分块的局部纹理描述来解决手势识别问题。尽管手势图像在颜色空间中看起来很平滑,但从灰度级上考虑,它实际上还是包含了足够多的纹理信息。LBP 通过比较像素点局部邻域的亮度变化来提取像素点的局部纹理特征。由于不同手势之间存在结构差异性,很明显,对 LBP 图像进行全局直方图统计会丢掉手势图像本身的空间结构特征,使得对手势的描述不够完备。因此,本文采用一种基于分块的策略来保持每一个手势的空间结构关系。最终,将每一个小块区域对应的直方图向量连接起来形成用于描述手势图像的完整的特征向量。

① 女,1978 年生,博士;研究方向:模式识别,计算机视觉,人机交互等;联系人,E-mail:ext-xiujian.chai@nokia.com
(收稿日期:2008-12-5)

1 局部纹理描述

对于图像匹配任务来讲,最重要的步骤就是寻找图像的不变特征描述。本文采用一种鲁棒的局部纹理描述作为图像特征来对不同的手势类别进行分类。

1.1 基本的 LBP 算子

在这一节,首先介绍一下基本的 LBP 算子^[12],它可以有效地衡量图像纹理的局部变化。

对于图像中的像素点 $c(x_c, y_c)$, 可用一个二进制串来表示,这个二进制串是通过将该像素与 8 邻域内的像素 $p(x_p, y_p)$ 灰度值进行比较得到的,如式

$$s(p, c) = \begin{cases} 1, & f_p \geq f_c \\ 0, & f_p < f_c \end{cases} \tag{1}$$

所示,这里, f 表示像素对应的灰度值。

将此二进制串与一个变换系数矩阵进行卷积运算(如图 1(c)所示)。最终对应于该像素的变换后的数值可通过等式

$$LBP(c) = \sum_{i=0}^7 s(p_i, c) \cdot 2^i \tag{2}$$

计算得到。

图 1 给出了一个 LBP 变换的示例。图 2 给出了对应不同手势图像的 LBP 滤波的图像示例,其第一

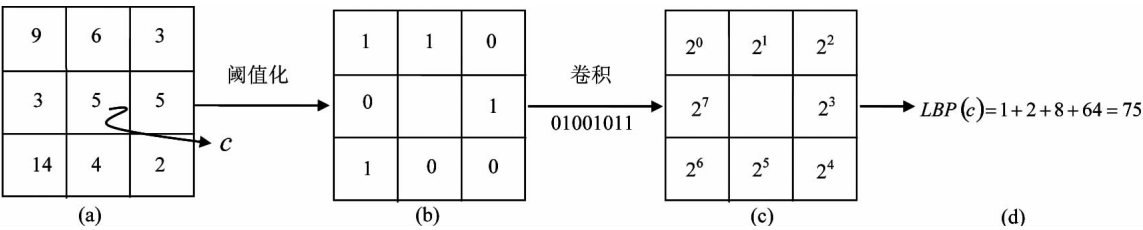


图 1 LBP 变换示例



图 2 对应不同手势的 LBP 滤波图像示例

行是 Jochen Triesch 数据库中的原始图像示例^[15],第二行是其对应的 LBP 滤波图像。从这个图中,我们可以看到 LBP 算子能够提取出很多图像细节。尽管复杂的背景会影响变换的结果,但同一手势类别内的不同样本的相似度仍然比不同类间样本的相似度要高得多。

1.2 基于分块的局部纹理描述

一旦得到灰度级的 LBP 纹理描述,就可以用直方图作为特征向量来描述该图像,即对每一个灰度级上出现的像素个数来进行计数,从而得到对其灰度信息分布的统计。然而直方图特征的一个缺点就在于它丢失了图像内容的空间结构关系,这无疑会导致一些错误的匹配,如图 3 所示。该图给出了对应于不同手势图像的三个直方图分布,上面的两个图像对应的全局直方图分布过于相似,导致手势难

以区分。为了保持直方图特征向量的空间结构关系,一种自然的方法就是采用分块策略。

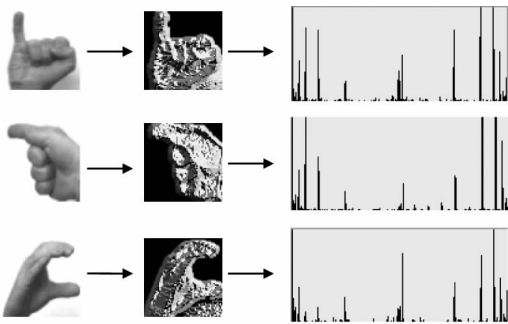


图 3 直方图描述造成误匹配的样例

对于 LBP 滤波图像,首先将它分割成很多小块的图像区域。在每一个局部的小块中统计直方图信息,将所有小块的直方图向量按序连接就形成了最终完整的特征向量,来刻画原始图像。图 4 给出了本文提出的基于分块的局部纹理特征描述方案。可见,由于采取分块的策略,每一种手势图像全局的空间结构位置关系都被有效的保留下来,通过与局部的纹理变化信息相结合,便可形成一种强大的手势纹理描述,用来实现准确的手势识别。

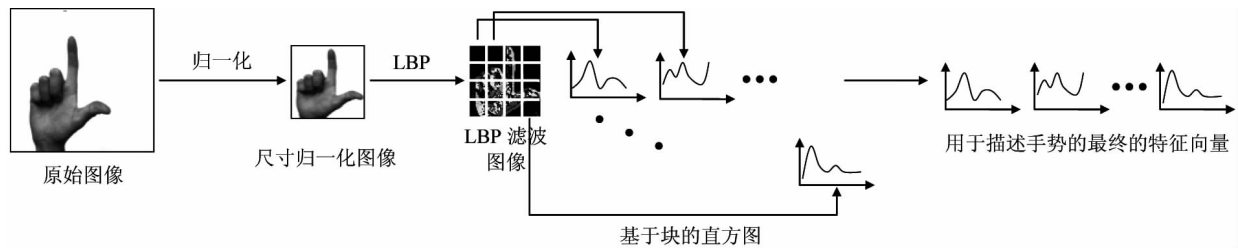


图 4 基于分块的局部纹理特征描述方案

2 识别方案

为了进行手势图像的识别,本文采用简单的直方图交作为相似性度量,并根据最近邻原则给出最终的识别结果。

两幅图像(I_1 和 I_2)的相似度由式

$$d(I_1, I_2) = \frac{\sum_{i=0}^{M \cdot Bin} \min(H_i^1, H_i^2)}{(W \times H)} \tag{3}$$

定义。这里, M 是块的总数, Bin 是 0 ~ 256 灰度分级数目, W 和 H 是经过尺寸归一化的图像的宽度和高度。计算测试图像和样本集合中每一个样本图像的相似度,最大相似度对应的图像所属的类别即为测试样本的识别结果:

$$Identity = \arg \max_j d(I_P, I_{G_j}) \tag{4}$$

此处, j 遍历整个样本数据库。

3 实验结果

3.1 数据库

对本文提出的方法,我们在两个数据库上进行了验证。一是可作为手势识别方法评测基准数据库的 Jochen Triesch 数据库,在此数据库上,我们对算法中各参数对性能的影响进行了分析,并将本文算法的结果与现有其他算法进行了比较。二是在地图导航中应用的真实场景中自采的手势图像序列数据库,它用于评测算法的性能。

Jochen Triesch 数据库包含 10 个不同的手势图像,如图 5 所示。每一个手势包含 24 个人在 3 种不同的背景(亮、暗、复杂背景)下的图像,除去两个丢失的图像,共有 718 幅图像。在此数据库上的实验严格地按照 Marcel 协议来进行,即 3 个人的所有图像作为样本数据库,其余 21 个人的图像形成测试集合。在识别之前,所有图像按照数据库提供的参数被剪切,并统一归一化到 64 × 64 大小。

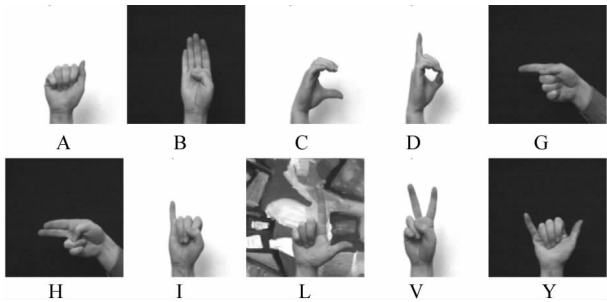


图 5 Jochen Triesch 数据库中的手势样本示例

在地图导航应用中,我们定义了 6 种基本的手势,如图 6 所示。分别代表上/下/左/右 4 个方向指令,以及启动(放大/缩小)、停止操作,其中放大/缩小指令可通过计算启动手势前后帧占有面积的变化来获得。此自采数据库共包含 460 帧图像,去除不同手势间的转移图像,有效手势为 441 帧。其测试结果将在性能分析部分中给出。

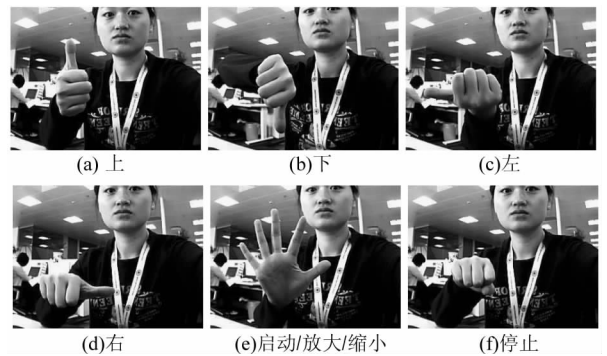


图 6 地图导航应用中手势类别定义

3.2 性能分析

在本文采取的识别方案中,性能与参数之间关系密切,尤其是直方图的 Bin 数以及小块的尺寸。直观来讲,直方图的 Bin 数如果太小(如 2,即 256 灰度级仅仅被分成 2 级),基于此得到的统计直方图毫无意义。随着直方图 Bin 数的增长,原始图像的更多的信息被保留下来,当然其计算代价也相对增长。

对于块的尺寸,过大的尺寸会损失空间结构信息,而过小的尺寸不仅会使最终特征向量的维数迅

速增长,而且,相同索引的块之间的对应也由于小块的过分琐碎而失去意义。对不同直方图 *Bin* 数以及不同小块尺寸的实验结果在图 7 中给出。如图所示,

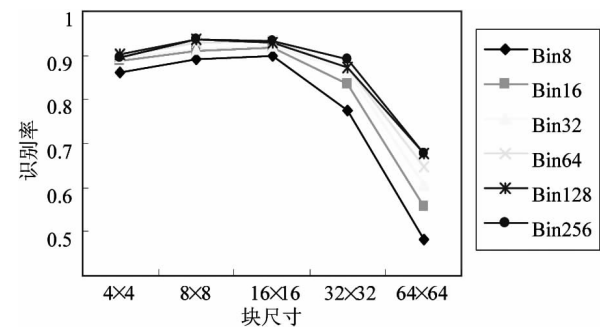


图 7 不同参数的识别结果比较

当块的尺寸固定不变时,不同的直方图 *Bin* 数对识别结果有轻微的影响。而对于固定的直方图 *Bin* 数来讲,块尺寸的变化会使得识别率随之发生较大的变化。为取得精度和计算效率的平衡,最终确定较为合适的 *Bin* 数是 128,小块的尺寸为 8×8 (相对于 64×64 的图像尺寸)。

这里,我们遵循 Jochen Triesch 数据库的 Marcel 协议,将本文方法与 MCT 方法^[11]的结果进行比较,该方法将 MCT 特征描述与 Adaboost 方法结合起来,实现了对手势的比较准确的识别,MCT 是一种强调图像局部空间结构的特征描述。在文献[11]中,4 个

人的图像用作训练样本,其他 4 个人的图像作为验证集合,剩余 16 个人的图像形成了测试集合。由于本文提出的识别方法不需要验证集合,因此,我们按照 Marcel 协议,将协议指定的 3 个人的图像作为训练样本,其余 21 人的图像形成测试集合,在相似的测试协议下,比较结果由表 1 给出,这里,MCT 方法

表 1 本文方法与 MCT 方法^[11]的识别性能比较

	识别率(%) (均衡背景)		识别率(%) (复杂背景)	
	MCT 方法	本文方法	MCT 方法	本文方法
A	100	97.5	100	100
B	93.75	97.5	93.75	95.0
C	93.75	92.5	93.75	100
D	84.38	95.0	81.25	100
G	100	97.5	68.75	100
H	90.63	80.0	87.5	80.0
I	90.63	97.5	62.5	80.0
L	96.88	100	75.0	95.0
V	96.77	89.7	87.5	78.9
Y	81.25	94.7	62.5	100
平均	92.79	94.2	81.25	92.9

的结果摘自文献[11]。由此可见,本文方法在均衡和复杂背景条件下,都取得了比 MCT 方法更好的识别结果。

为了更深刻地对实验结果进行分析,我们将本文方法的识别结果做了可视化呈现,如图 8 所示,均

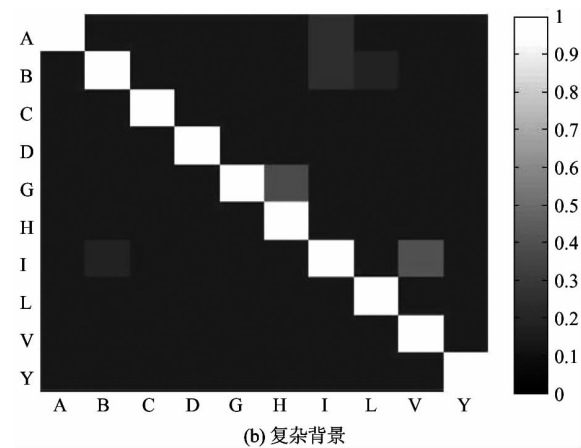
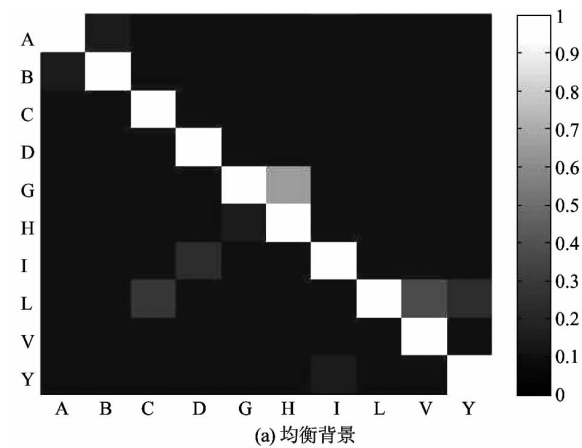


图 8 本文方法识别结果的可视化呈现

衡背景下和复杂背景下的识别结果分别在图(a)和(b)中给出。其中,横坐标为手势的真实类别,纵坐标为本算法识别得到的手势类别,小方格的亮度对应着识别率。很明显,对角线处基本显示着极高的亮度,即本文提出的方法得到了非常准确的识别结果。

除此以外,本文还对基于手势分析的地图导航应用中采集的手势视频数据进行了测试,以对本文提出方法进行客观的评价。对每一帧图像,本系统通过肤色检测和跟踪技术对手势进行了自动的分割,如图 9 所示,进而对其进行识别,识别结果由表 2 给出。从图 9 示例图像可见,手势所处背景相对复



图 9 自采手势图像对应的跟踪和分割结果以及不同手势之间的转移状态示例

表 2 自采数据手势识别结果

手势类别	左	右	上	下	启动	停止	总数(平均)
图像数目	64	57	50	55	177	38	441
正确识别图像数	62	55	49	51	176	37	430
识别率(%)	96.9	96.5	98	92.7	99.4	97.4	97.5

杂,并不限制为单一颜色的背景,而可以是正常的工作环境。由于识别是根据自动的分割结果进行的,因此,分割的精度对识别的结果也有一定的影响。除去不同手势之间变换时产生的差异较大的转移状态情况(如图 9 最后一列),对所有其他手势图像的识别结果表明,即使在复杂的背景下,本文提出的方法也可以取得很高的识别精度,实现自然的人机交互。

4 结论

本文提出了一种基于分块局部纹理描述的方案,用来解决手势识别问题。该方法使得图像的空间结构信息和局部的纹理信息都得到了有效的保留,形成了最终的图像特征描述。本文方法与 MCT 方法相比,识别准确度有很大提高,在复杂的背景条件下尤为显著。并且,在实际手势交互系统中的应用也表明了本方法的有效性。

参考文献

[1] Manresa C, Varona J, Mas R, et al. Hand tracking and gesture recognition for human-computer interaction. *Electronic Letters on Computer Vision and Image Analysis*, 2005, 5(3): 96-104

[2] Buades J M, Perales F J, Varona J. Real time segmentation and tracking of face and hands in VR application. In: *Proceedings of the 3rd International Workshop on Articulated Motion and Deformable Objects*, Palma de Mallorca, Spain, 2004. 259-268

[3] MacCormick J, Isard M. Partitioned sampling articulated ob-

jects, and interface-quality hand tracking. In: *Proceedings of the 6th European Conference on Computer Vision*, Dublin, Ireland, 2000. 3-19

[4] Nielsen E S, Canalis L A, Tejera M H. Hand gesture recognition for human-machine interaction. In: *Proceedings of the 12th International Conference in Central Europe on Computer Graphics, Visualization and computer Vision*, Plzeň, Czech, 2004. 395-402

[5] Lee J, Lee Y, Lee E, et al. Hand region extraction and gesture recognition from video stream with complex background through entropy analysis. In: *Proceedings of the 26th Annual International Conference of the IEEE EMBS*, San Francisco, USA, 2004. 1513-1516

[6] Bretzner L, Laptev I, Lindeberg T. Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering. In: *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition*, Washington, D.C., USA, 2002. 423-428

[7] Laptev I, Lindeberg T. A multi-scale feature likelihood map for direct evaluation of object hypotheses. In: *Proceedings of the IEEE Workshop on Scale-Space and Morphology*, Vancouver, Canada, 2001. 98-110

[8] Bradski G, Yeo B L, Yeung M M. Gesture for video content navigation. In: *Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Database*, San Jose, USA, 1999. 230-242

[9] Zhou H N, Lin D J, Huang T S. Static hand gesture recognition based on local orientation histogram feature distribution model. In: *Proceedings of the 2004 IEEE Computer Society Conference on Vision and Pattern Recognition Workshops*. Miami, FL, USA, 2004, 161-169

[10] Binh N D, Shuichi E, Ejima T. Real-time hand tracking and

- gesture recognition system. In: Proceedings of the International Conference on Graphics, Vision and Image Processing, Cairo, Egypt, 2005. 362-368
- [11] Just A, Rodriguez Y, Marcel S. Hand posture classification and recognition using the modified census transform. In: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, Southampton, UK, 2006. 351-356
- [12] Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on feature distribution. *Pattern Recognition*, 1996, 29(1): 51-59
- [13] Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans PAMI*, 2002, 24(7): 971-987
- [14] Zhang W C, Shan S G, Gao W, et al. Local gabor binary histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition. In: Proceedings of the 10th IEEE International Conference on Computer Vision, Beijing, China, 2005. 786-791
- [15] Triesch J, von der Malsburg C. Robust classification of hand postures against complex backgrounds. In: Proceedings of 7th International Conference on Automatic Face and Gesture Recognition, Killington, USA, 1996. 170-175

Patch-based local texture description for hand posture recognition

Chai Xiujuan^{* **}, Wang Kongqiao^{**}

(^{*} School of Information Engineering, Beijing University of Posts and Telecommunications, Beijing 100088)

(^{**} System Research Center, NOKIA Research Center, Beijing 100176)

Abstract

This paper presents a novel texture description scheme for accurate recognition of hand postures. The description is based on the local binary patterns (LBP), an effective texture coding pattern. This recognition method performs the LBP transformation for each pixel of an image, and extracts the brightness variations in the neighborhood of each pixel as the features to characterize the pixels. To well maintain the spatial structure information, a patch-based scheme is adopted to get more robust representation for each posture. By concatenating all the local histograms of each patch into a whole one, a powerful descriptive texture feature is generated. Intrinsically, the method effectively combines the global structure information and the local texture information together and gets ideal results. The experimental results on Jochen Triesch database convincingly show that the proposed patch-based local texture description scheme for hand posture recognition has the distinct advantage compared with other methods

Key words: local texture description, hand posture recognition, local binary pattern (LBP), histogram