

中图法分类号: TP391.4 文献标志码: A 文章编号: 1006-8961(2011)11-2041-06

论文索引信息: 冯文刚, 高隽, Bill P. Buckles, 吴克伟. 无监督学习的无线胶囊内窥镜视频分类方法 [J]. 中国图象图形学报, 2011, 16(11): 2041-2046

# 无监督学习的无线胶囊内窥镜视频分类方法

冯文刚<sup>1)</sup>, 高隽<sup>1)</sup>, Bill P. Buckles<sup>2)</sup>, 吴克伟<sup>1)</sup>

<sup>1)</sup>(合肥工业大学计算机与信息学院, 合肥 230009)

<sup>2)</sup>(北德克萨斯大学计算机科学与技术系, 丹顿 76203)

**摘要:**近年来在对病人的消化道系统检查中,无线胶囊内窥镜(WCE)是一种最新技术,可以让医生直接观察到病人的病灶所在,但是对于消化道系统中的口腔、胃、小肠和大肠的WCE视频分类却是难点所在。相关研究中均采用通过人工标记的训练库的有监督学习方法。为了在WCE训练数据中获得高识别率,提出一种无监督学习方法,它利用融合颜色信息的尺度不变特征转换(SIFT)获取局部特征,再利用概率隐语义分析模型(pLSA)数据训练中进行语义内容分析。实验结果表明,在WCE图像分类中本方法与当前最新的监督分类方法一样可以获得高准确率。

**关键词:** 视频分类; 无监督学习; 概率隐语义模型; 尺度不变特征变换

## Wireless capsule endoscopy video classification using an unsupervised learning approach

Feng Wengang<sup>1)</sup>, Gao Jun<sup>1)</sup>, Bill P. Buckles<sup>2)</sup>, Wu Kewei<sup>1)</sup>

<sup>1)</sup>(School of Computer and Information, Hefei University of Technology, Hefei 230009 China)

<sup>2)</sup>(Department of Computer and Science, University of North Texas, Denton 76203 USA)

**Abstract:** Since Wireless Capsule Endoscopy (WCE) is a novel technology for recording the videos of the digestive tract of a patient, the problem of segmenting the WCE videos of the digestive tract into sub-images corresponding to the mouth, stomach, small intestine and large intestine regions is not well addressed in the literature. A few papers addressing this problem use a supervised learning approach that presumes availability of a large database of correctly labeled training samples. Considering the difficulties in procuring sizable WCE training data sets needed for achieving high classification accuracy, we introduce an unsupervised learning approach that employs Scale invariant feature transform (SIFT) with color information for extraction of local features and uses probabilistic latent semantic analysis (pLSA) model for data semantic analysis. Our results indicate that this method compares well in classification accuracy with the state-of-the-art supervised classification approach to WCE image classification.

**Keywords:** video classification; unsupervised learning approach; probabilistic latent semantic analysis (pLSA); Scale invariant feature transform (SIFT)

## 0 引言

无线胶囊内窥镜(WCE)是一种新颖的技术,用

来记录病人的消化道系统中各部分的视频,而这些视频恰恰是其他类型的内窥镜(例如结肠镜检查)所不能可视化观测到的。在WCE检测中,病人吞咽下一个带有微摄像机的胶囊,利用消化道系统的蠕动贯穿

收稿日期:2010-09-25;修回日期:2010-11-24

基金项目:国家自然科学基金项目(60875012,60905005);美国国家防御研究规划中心项目(DARPA-BAA-10-53)。

第一作者简介:冯文刚(1982—),男。合肥工业大学智能信息处理研究室计算机应用技术专业博士研究生,主要研究方向为模式识别、计算机视觉、图像处理等。E-mail:wengang.feng@gmail.com。

整个消化道系统拍摄视频。通过在胶囊中的微无线发射装置传输至体外的无线接收器,从而获得消化道系统各个部位的视频<sup>[1]</sup>。一般而言,每个病人的 WCE 视频由 5 000 帧图像组成,帧速率为 2 帧/s。

自从胶囊内窥镜技术在临床使用后,它在消化道出血位置确定<sup>[2]</sup>,节段性回肠炎<sup>[3]</sup>,以及乳糜泻<sup>[4]</sup>等疾病的诊断中起到重要作用。但是该技术却需要消耗大量的人力资源,因为即使是一名接受过专业培训的工作人员也需要 1~2 小时从 WCE 视频中找到病灶,例如出血点或者溃疡<sup>[5]</sup>。为了缓解该问题,需要对应于消化道系统中不同部位将 WCE 视频序列分类为 4 段:入口、胃、小肠,以及大肠。图 1 中所示分别为这 4 个不同部位的图像帧。因为某些病灶只会出现在特定的部位,为了使医生只关注某些特定的图像帧,快速进行诊断,因此对 WCE 视频进行分类是 WCE 数据快速分析的一个关键问题。

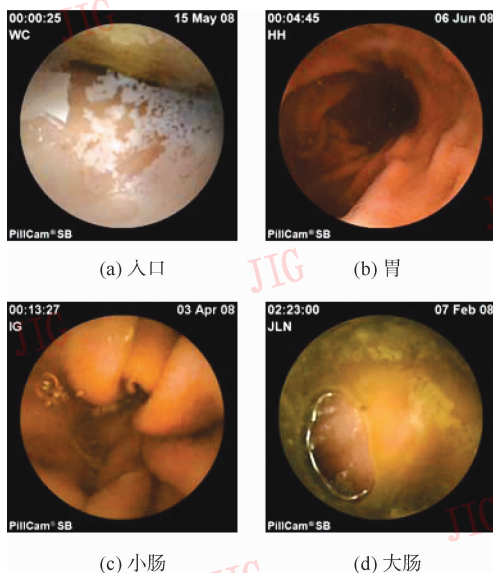


图 1 消化道系统中不同器官

Fig. 1 Four parts of digestive

WCE 视频分类方法一般可以分为 2 类。第 1 类方法侧重于鲁棒性特征的提取(具有更强描述特性的颜色特征或纹理特征)<sup>[6-10]</sup>,第 2 类方法侧重于分类器的研究,例如贝叶斯分类器和支持向量机<sup>[11-12]</sup>。

文献[6]中,使用 MPEG-7 中的颜色和纹理描述子对视频进行分类。文献[7]中,通过在 6 个颜色通道( $R, G, B, H, S, V$ )中分别提取纹理特征向量进行视频分类。在文献[8]中,通过合并颜色、纹理

和动作特征,在图像块中获取可视组织。

文献[11]中, Spyridonos 从图像中提取梯度张量特征,使用支持向量机检测器官褶皱从而对器官进行分类。文献[12]中, Cunha 分别使用贝叶斯分类器和支持向量机进行 WCE 视频分类,并进行对比。Mackiewicz<sup>[8]</sup>在假设消化道系统不同器官(入口、胃、小肠和大肠)遵循确定的概率分布的前提下,在隐马尔科夫模型框架中使用支持向量机进行分类。

在上述所有方法中,都是利用有监督学习方法对 WCE 视频进行分类,即这些方法需要大量正确标记的数据库的基础上才能进行分类器的训练。但是由于 WCE 是一种新兴的技术,并不能获得大量 WCE 视频。更进一步,由于各个消化器官的颜色和纹理的区别不大,人工标注数据既耗时也容易出错。因此,提出一种有效的无监督学习方法进行 WCE 视频分类。该方法在分析视觉词汇的基础上获得语义,从而可以在不知道类别种类信息的前提下将其分成不同的类别。使用概率隐语义分析模型( $pLSA$ )<sup>[13]</sup>,采用尺度不变特征变换(SIFT)<sup>[14]</sup>代替  $pLSA$  模型中的词汇,为了获得更好的结果,将 SIFT 特征与颜色特征进行融合。

## 1 视觉词汇提取及视觉字典建立

对 WCE 视频进行处理的第 1 个步骤为从原始视频帧中去除黑色背景和文字标注信息。因为 WCE 摄像机的特殊构造,只有图像中央的圆形区域为兴趣区域(ROI)。而兴趣区域以外的信息都与消化道器官无关,因此可以安全地去除。

下一步则为视觉词汇提取,即从 WCE 的每一帧图像中的兴趣区域中提取 SIFT 特征。因为在 bag-of-words 模型中,特别是在场景分类与目标识别应用中, SIFT 特征具有很大的优势。SIFT 特征是一种具有高区别性的局部特征,它具有尺度和旋转不变性以及光照和视觉的部分不变性。SIFT 特征中的兴趣点是从多层高斯差分(Difference of Gaussians)过滤器中提取的局部极值,因此具有尺度不变性。确定兴趣点后,利用具有 36 簇的直方图来描述其梯度方向,其中每一个簇由对应梯度权值加权。在每一个簇中,将相当于主峰值 80% 能量的峰值方向视为该关键点方向,其中梯度方向可有多个。接下来以兴趣点为中心取  $16 \times 16$  的窗口,其中包括 16 个  $4 \times 4$  子窗口,由

图 2(b)所示。由于有 16 个子窗口,且每个窗口上计算 8 个方向的梯度方向直方图,因此每一个兴趣点的描述子获得一个 128 维向量。

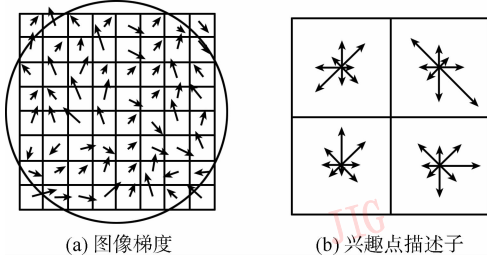


图 2 图像梯度和兴趣点描述子

Fig. 2 Image gradient and keypoint descriptor

与灰度图像相比颜色信息可以提供更易分辨的信息,因此在内窥镜检测图像中将在 SIFT 特征的基础上融合颜色特征。在所有颜色空间中,虽然 RGB 颜色空间是最简单且最常用的方式,但是 HSV (hue, saturation, value) 颜色特征却具有最好的分辨特性<sup>[15]</sup>。同时,Coimbra<sup>[16]</sup>研究得出 MPEG-7 颜色特征(HSV 颜色特征)在 WCE 图像中能够获得最佳的分类结果。在 HSV 颜色空间的所有 3 个通道上面提取 SIFT 特征,因此 SIFT 特征的维数将会变为原先的 3 倍,且颜色空间中的兴趣点描述子将有 384 (3 × 128) 种成分。

由于分类结果既取决于颜色空间的选择,也取决于颜色特征的提取方式,为了比较不同的颜色空间,因此在实验中同时也采用 RGB 颜色空间进行对比。与 HSV 颜色空间一样,在 RGB 颜色模型中的 3 个通道中进行 384 维的 SIFT 特征提取。

因为无监督学习方法是基于 bag-of-words 模型(BOW)<sup>[17-21]</sup>,而 BOW 模型最初是用于文本中进行语义内容分析,因此需要在视觉词汇的基础上建立视觉字典。特征向量描述子不能直接用来作为视觉词汇,这是因为特征向量中的每一个成分的跨度为无限集合,而文本中的词汇都是属于有限集合<sup>[22-23]</sup>。最常用的解决方法是通过向量量化来限制特征向量的数目。利用  $K$  均值聚类算法,将从 WCE 视频帧中提取的特征向量聚类,则聚类中心表示为视觉词汇,每个类别的特征数目较多但并不是无限集合。这里  $K$  为视觉词典的容量,每一个特征向量都唯一映射至词典中的一个视觉词汇。依据实验 3 的结果,最终选取视觉字典的容量为 600 个视觉词汇(量化向量)。

## 2 pLSA 模型

pLSA 模型最初用于自然语言或文本文档的语义内容分析,它是统计生成模型。假设需要分析  $M$  个文档集合  $d_j \in \mathbf{D} = \{d_1, \dots, d_M\}$  中的每一个文档,其包含的词语为  $w_i \in \mathbf{W} = \{w_1, \dots, w_N\}$ ,这里  $N$  为词汇词典  $\mathbf{W}$  中包括的词汇的数量。对文档的分析过程就是从文档中获得隐语义信息,即类别信息(话题)  $z_k \in \mathbf{Z} = \{z_1, \dots, z_K\}$ 。由于在文档中会有多义词的存在,因此很难从词汇中获得文档的语义信息,且由于文档  $d_j$  可能是若干类别信息的混合,更进一步增加了分析难度。pLSA 模型的应用是对文档进行分析,特别是将其分解为依据话题信息混合分布的词汇。

$d_j$  对应于 WCE 图像帧, $\mathbf{D}$  对应与整段视频, $z_k$  对应消化系统中的器官,而  $w_i$  为视觉词汇即量化特征向量。由于在消化道系统中包含 4 种器官,因此  $K$  的数值这里为 4。

在将 pLSA 模型映射至本文的问题后,可以采用 Hofmann<sup>[23]</sup>的 pLSA 公式。假设每一个视频帧  $d_j$  服从隐话题的混合分布,定义为  $p(z_k | d_j)$ ,  $p(z_k)$  为话题分布的先验概率。令  $p(w_i | z_k)$  为表示话题  $z_k$  的多项式分布,则视觉词汇和图像帧的联合概率分布定义为

$$p(w_i, d_j) = \sum_{k=1}^K p(z_k) p(w_i | z_k) p(z_k | d_j) \quad (1)$$

因为话题分布具有隐性特征,所以这里通过在迭代过程中最大化对数似然采用期望最大化算法(EM)对参数  $p(z_k)$ ,  $p(w_i | z_k)$ , 和  $p(z_k | d_j)$  进行估计:

$$L = \log P(\mathbf{D}, \mathbf{W}) = \prod_{i=1}^M \prod_{j=1}^N n(w_i, d_j) \log P(w_i, d_j) \quad (2)$$

这里  $n(w_i, d_j)$  表示视觉词汇  $w_i$  在图像  $d_j$  中出现的次数。

EM 算法具有两个步骤:

在 E 步骤中,计算隐话题的条件概率分布  $z_k$ :

$$P(z_k | d_i, w_j) = \frac{p(z_k) P(w_j | z_k) P(d_i | z_k)}{\sum_{l=1}^K P(w_j | z_l) P(z_l | d_i)} \quad (3)$$

在 M 步骤中,依据在 E 步骤中获得的期望值  $P(z_k | d_i, w_j)$  对话题概率  $p(z_k)$ , 视觉词汇概率  $P(w_i | z_k)$  以及图像概率  $P(d_i | z_k)$  进行更新。

$$p(z_k) = \frac{\sum_{i=1}^M \sum_{j=1}^N n(w_i, d_j) P(z_k | d_j, w_i)}{\sum_{i=1}^M \sum_{j=1}^N n(w_i, d_j)} \quad (4)$$

$$P(w_j | z_k) = \frac{\sum_{i=1}^N n(d_i, w_j) P(z_k | d_i, w_j)}{\sum_{m=1}^M \sum_{i=1}^N n(d_i, w_m) P(z_k | d_i, w_m)} \quad (5)$$

$$P(z_k | d_i) = \frac{\sum_{j=1}^M n(d_i, w_j) P(z_k | d_i, w_j)}{n(d_i)} \quad (6)$$

在初始时,EM 算法分别对于  $p(z_k)$ ,  $p(w_i | z_k)$ , 和  $p(z_k | d_j)$  设置随机初始值(对应变量的总和为 1)。在 EM 的每一次迭代中,计算式(4)(6),并可以通过式(3)计算获得似然  $L$  的数值。如果在当次迭代中获得的  $L$  的数值大于前次所获得的最大值,则更新似然的数值,同时  $p(z_k | d_j)$  (本文算法的核心向量)也被存储。在迭代过一定次数后,如果似然值不再有变化,则算法收敛,且对于每一个文档的核心向量  $p(z_k | d_j)$  的数值用来确定类别信息。图 3 所示为算法流程图。首先对于随机抽取的测试集提取视觉词汇,在 pLSA 框架中获得最大似然,在系统运行一定次数,算法收敛后,从所获得的对应概率值可以获得分类结果。

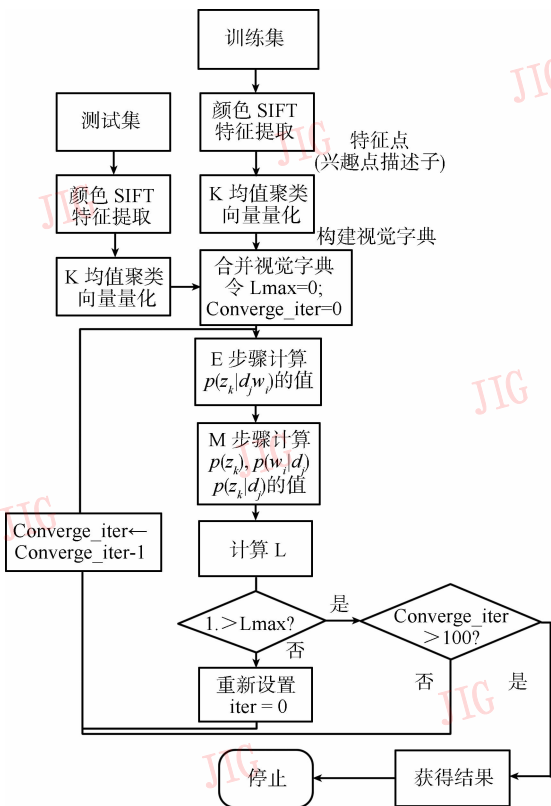


图 3 无监督学习 WCE 视频帧分类算法流程图  
Fig. 3 Flowchart of WCE video classification by unsupervised learning algorithm

### 3 实验分析

对十段 WCE 视频进行分类,共计 5 000 帧图像,并根据医学常识将视频帧分为 4 段:入口(P1),胃(P2),小肠(P3)和大肠(P4)。在实验中,将随机选取半数的视频帧作为训练集,剩余半数作为测试集。如图 4 所示,本文的实验结果是对输入的王CE 视频序列,经过训练后自动分类为 4 种对应消化道器官,并可根据需要输出对应器官的视频。下面对于本文进行的 4 个实验进行分析。

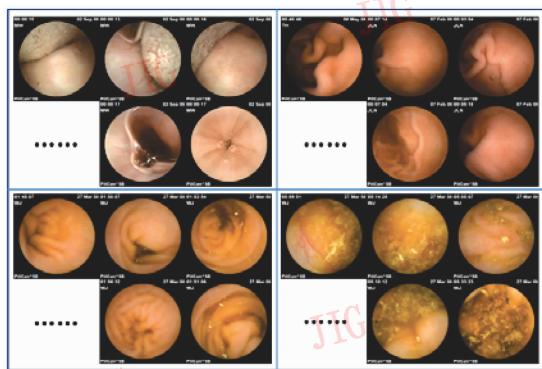


图 4 WCE 视频分类结果

Fig. 4 WCE video classification result

#### 3.1 分类器性能对比实验

因为分类的结果同时与特征集选取和分类器的选取相关,因此第 1 个实验是在相同的特征集(灰度 SIFT 特征)基础上研究分类器的分类性能。该实验的目的是将无监督方法与支持向量机方法进行比较,而支持向量机方法是现有有监督分类方法中最有效的。因此,将同时使用线性核和非线性核(径向基函数)支持向量机。对提取过 SIFT 特征的人体的 4 个部位进行 3 次一对一支持向量机进行分类。但是支持向量机(其他有监督学习分类器也一样)上采用 SIFT 特征时,它将每一帧图像视为一个整体,但是特征向量却是从不同帧中提取出来。可以通过将所有帧中兴趣点的特征向量连接起来的方法进行解决,即相当于对所有帧构建特征向量。但是由于每一帧图像中检测到的兴趣点数目不同,则帧特征向量(连接所有兴趣点的特征向量后获得的特征)的尺寸将会不同。为了解决此问题,将最大尺寸的帧特征视为标准特征向量。这里,将随机抽取一半 WCE 视频图像帧作为训练集来训练分类器,剩余一半用来测试分类器。各分类器结果如表 1 所示可以看出无监督学习

方法同样可以得到较好的结果。

表 1 实验 1 结果比较

Tab. 1 Experiment 1 results comparison

	P1/P2	P2/P3	P3/P4
SIFT-SVM-lin	75.6	73.2	69.8
SIFT-SVM-rad	78.3	77.3	71.9
SIFT-pLSA	85.0	84.2	80.4

从表 1 结果中不难看出,入口/胃和胃/小肠分类结果比小肠/大肠分类结果要好。这是因为前 3 部分的组织相似性没有后两部分的相似性强。从上述分类结果可以看出,非线性(径向基核)SVM 分类结果比线性 SVM 分类结果要好。而本文方法的准确率又比上述两种分类器要高,但是并不能就此说该方法比 SVM 方法要好。这是因为 SIFT 特征不是专门为 SVM 构建的,而是特别为 bag-of-words 模型构建,所以 pLSA 可以获得更好的结果。

### 3.2 特征向量性能对比实验

第 2 个实验希望在对 WCE 视频分类时,利用 pLSA 分类器分别对采用灰度 SIFT 特征,RGB-SIFT 颜色特征以及 HSV-SIFT 颜色特征进行比较。实验结果如表 2 所示,RGB-SIFT 颜色特征比其他两个特征可以获得更好的结果。

表 2 实验 2 结果比较

Tab. 2 Experiment 2 results comparison

	P1/P2	P2/P3	P3/P4
Gray-SIFT	85.0	84.2	80.4
HSV-SIFT	94.4	92.5	91.0
RGB-SIFT	98.3	94.1	93.9

由于在 SVM 分类器中使用 SIFT 特征具有一定的不适应性,在表 3 中显示了使用 RGB-SIFT 颜色特征的 pLSA 分类器与 Mackiewicz<sup>[8]</sup> 提出的最佳分类结果进行比较。在文献[8]中,作者使用了两种 SVM 分类器,而采用了以下不同的特征:1)局部二元模式(LBP)直方图,在 3 种颜色通道上有 343 簇,并利用主成分分析约减;2)从 41 帧连续图像中提取的 6 元组的运动特征,经过离散傅里叶变换后,再利用主成分分析约减;3)32×32 簇的 HS(Hue 和 Saturation)直方图,在通过离散傅里叶变换和主成分分析约减。从表 3 可知,基于 RGB-SIFT 颜色特征的 pLSA 分类器的分类结果略低于当前最好的有监督学习方法。但

是必须注意的一点是,RGB-SIFT 颜色特征的 pLSA 分类器不需要在图像预处理过程中进行图像标注,同时分类结果的细微差异也有可能是由于使用特定图像库或者在特征提取过程造成的。

表 3 实验 2 结果比较

Tab. 3 Experiment 2 results comparison

	P1/P2	P2/P3	P3/P4
pLSA-RGB-SIFT	98.3	94.1	93.9
SVM-Radial	99.9	98.3	94.7
SVM-Linear	99.7	96.4	89.1

### 3.3 视觉词汇数目对比实验

最后一个实验是对 pLSA 模型进行测试。由于使用 RGB-SIFT 颜色特征的 pLSA 模型的性能依赖于视觉字典的尺寸,因此希望研究本文的词汇字典(相对应的在 K 均值聚类方法进行向量量化的聚类数目)对于视频分类精确性的影响。视觉字典的尺寸从 200~600 依次增加 100。平均分类精度如图 5 所示。从上述结果可知,当词汇字典的尺寸分别为 300,400 和 500 时,都可以获得较高的分类精度。

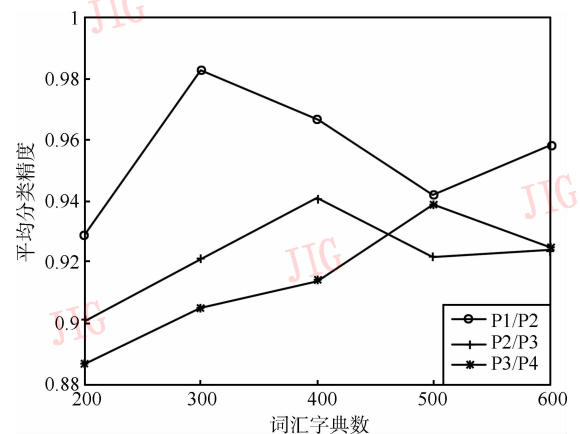


图 5 实验 3 结果比较

Fig. 5 Experiment 3 results comparison

## 4 结论

提出一种新颖的 pLSA 方法应用于无线胶囊内窥镜视频帧分类。该方法利用 SIFT 算法提取局部特征构建视觉字典,利用 pLSA 模型训练实现视频帧分类。实验结果表明,方法优于传统的基于 SIFT 特征的 SVM 分类器方法。然而对于 SVM 分类器构建 SIFT 特征向量具有一定的难度,且实验

表明 SIFT 特征并不是最适合 SVM 或者其他有监督学习方法的特征。基于颜色特征的 SIFT 特征比纯灰度 SIFT 特征的性能大大提升,但是与有监督学习分类器结果相反<sup>[27-28]</sup>,基于 RGB 颜色特征的 SIFT 特征的分类性能高于基于 HSV 颜色特征的 SIFT 特征。将本文方法与当前最优的 SVM 分类器进行无线胶囊内窥镜视频分类的性能相比较,无监督学习方法可以达到有监督学习方法的性能,同时适当的选择视觉词汇字典的尺寸可以帮助于获得较好的特征集合。

## 参考文献 (References)

- [ 1 ] Krishnan S, Wang P, Kugean C, et al. Classification of endoscopic images based on texture and neural network [ C ] // Proc. of 23rd Annual IEEE Int. Conf. in Engineering in Medicine and Biology. Washington DC; IEEE Press, 2001 : 3691-3695.
- [ 2 ] Gay G, Delvaux M, Key J. The role of video capsule endoscopy in the diagnosis of digestive diseases; A review of current possibilities [ J ]. Endoscopy, 2004, 36 ( 3 ) : 913-920.
- [ 3 ] Swain P. Wireless capsule endoscopy and Crohn's disease [ J ]. Gut, 2005, 54 ( 4 ) : 323-326.
- [ 4 ] Culliford A, Daly J, Diamond B, et al. The value of wireless capsule endoscopy in patients with complicated celiac disease [ J ]. Gastrointestinal Endoscopy, 2005, 62 ( 2 ) : 55-61.
- [ 5 ] Maieron A. Multicenter retrospective evaluation of capsule endoscopy in clinical routine [ J ]. Endoscopy, 2004, 36 ( 3 ) : 864-868.
- [ 6 ] Coimbra M T, Cunha J P S. MPEG-7 visual descriptors—contributions for automated feature extraction in capsule endoscopy [ J ]. IEEE Transactions on Circuits and Systems for Video Technology, 2006, 16 ( 4 ) : 628-637.
- [ 7 ] Boulougoura M, Wadge E, Kodogiannis V S, et al. Intelligent systems for computer-assisted clinical endoscopic image analyses [ C ] // Proc. of 2nd Int. Conf. Biomed. Engineering. Innsbruck, Austria : IEEE Press, 2005 : 405-412.
- [ 8 ] Mackiewicz M, Berens J, Fisher M. Wireless capsule endoscopy color video segmentation [ J ]. IEEE Transactions on Medical Imaging, 2008, 27 ( 2 ) : 1123-1130.
- [ 9 ] Howarth P, Yavlinsky A, Heesch D, et al. Medical image retrieval using texture, locality and color [ C ] // Proc. Cross Language Evaluation Forum. Washington DC, USA; IEEE Press, 2005 ( 2 ) : 740-749.
- [ 10 ] Wang P, Krishnan S, Kugean C, et al. Classification of endoscopic images based on texture and neural network [ C ] // Proc. 23rd Annu. Int. Conf. IEEE Eng. Med. Biol. Sci. Washington DC, USA; IEEE Press, 2001 : 3691-3695.
- [ 11 ] Spyridonos P, Vilarino F, Vitria J, et al. Anisotropic feature extraction from endoluminal images for detection of intestinal contractions [ C ] // Proc. MICCAI. Washington DC, USA; IEEE Press, 2006 : 161-168.
- [ 12 ] Cunha J, Coimbra M, Campos P, et al. Automated topographic segmentation and transit time estimation in endoscopic capsule exams [ J ]. IEEE Trans. Medical. Image, 2008, 27 ( 1 ) : 19-27.
- [ 13 ] Hofmann T. Unsupervised learning by probabilistic latent semantic analysis [ J ]. Machine Learning, 2001, 42 ( 2 ) : 177-196.
- [ 14 ] Lowe D G. Distinctive image features from scale-invariant keypoints [ J ]. International Journal of Computer Vision, 2004, 60 ( 2 ) : 91-110.
- [ 15 ] Berens J. Image Indexing Using Compressed Color Histograms [ M ]. Springer-Verlag, Norwich; Springer, 2002.
- [ 16 ] Coimbra M, Campos P, Cunha J P S. MPEG-7 visual descriptors—Contributions for automated feature extraction in capsule endoscopy [ J ]. IEEE Trans. Circuits Syst. Video Technology, 2006, 16 ( 4 ) : 628-637.
- [ 17 ] Jurie F, Triggs B. Creating efficient codebooks for visual recognition [ C ] // Proceedings of International Conference on Computer Vision. Washington DC, USA; IEEE Press, 2005 : 458-465.
- [ 18 ] Bosch A, Zisserman A, Muoz X. Scene classification using a hybrid generative discriminative approach [ J ]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30 ( 3 ) : 712-727.
- [ 19 ] Linde Y, Buzo A, Gray R M. An algorithm for vector quantizer design [ J ]. IEEE Transactions on Communications, 1980, 28 ( 3 ) : 84-95.
- [ 20 ] Oliva A, Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope [ J ]. International Journal of Computer Vision, 2001, 42 ( 1 ) : 145-175.
- [ 21 ] Niebles J, Wang H, Li F. Unsupervised learning of human action categories using spatial-temporal words [ J ]. International Journal of Computer Vision, 2008, 32 ( 4 ) : 1223-1230.
- [ 22 ] Lloyd S P. Least square quantization in PCM [ J ]. IEEE Transactions on Information Theory, 1982, 28 ( 2 ) : 129-137.
- [ 23 ] Li Feifei, Perona P. A Bayesian hierarchical model for learning natural scene categories [ C ] // IEEE Conference on Computer Vision and Pattern Recognition. Washington D C, USA; IEEE Press, 2005 : 524-531.