一种面向图象语义的主要区域提取方法

王惠锋 孙正兴
(南京大学计算机软件新技术国家重点实验室,计算机科学与技术系,多媒体技术研究所,南京 210093)

摘 要 图象主要区域的提取是图象语义抽取及其应用的基础.为了更好地进行图象语义的抽取,提出了一种面向图象语义的图象主要区域自动提取方法.该方法首先将图象划分成固定大小的子块,并通过对子块特征进行聚类来获得图象的初始区域分割;而后,经过一系列的后处理来优化分割结果,并实现前景和背景区分;最后通过分析每个背景区域的重要程度,去除掉不相关的背景区域.通过对包含有显著对象的户外图象进行的实验表明:该方法不仅可以去除图象中,大量与图象语义不相关的内容,而且能保留图象的主要信息,这就为进一步的图象语义应用打好了基础.

关键词 图象语义 图象分割 聚类分析 图象划分 背景分割
中图法分类号: TP391.41 **文献标识码**: A **文章编号**: 1006-8961(2003)01-0027-06

A Method of Main-Region Extraction for Semantic Image Retrieve

WANG Hui-feng, SUN Zheng-xing
(*State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing* 210093)
(*Department of Computer Science and Technology, Nanjing University, Nanjing* 210093)

Abstract Semantic image retrieval is one of the key technologies to find useful multimedia information more efficiently on Internet or in multimedia database. Extraction of main regions in an image is a precondition for semantic image retrieval. In this article, an automatic approach to extract those main regions is proposed. It first partitions an image into fixed sized blocks, and an elementary segmentation is achieved by clustering the visual features of all the blocks of the image. Then the result of the original segmentation is improved by some extra processing. After that, a special method is employed to distinguish the foreground regions and background regions. Finally, the regions, which are considered not important to the image content, are eliminated, and it is done by analyzing the importance of every region. Our experiments for outdoor images containing relatively salient objects show that, the approach proposed in this paper can get rid of lots of information, which are not related to the image content, and at the same time can also reserve the main useful information for image semantics. It gives a better foundation for the further applications such as image retrieval and image understanding.

Keywords Image semantics, Image segmentation, Cluster analysis, Image partition, Background removal

0 引 言

大家知道,图象低层的物理视觉特征与人的高层认识之间不存在明显的直接联系,这就是视觉信息处理中的"语义鸿沟(Semantic Gap)"[1],且由于这种语义鸿沟,使得基于图象全局特征的检索结果与人的主观感觉大相径庭.要缓解"语义鸿沟"问题,一个直接的方法是在低层的视觉特征和高层的主观语义之间建立多个中间处理过程[2],但是,枚举所有的主观语义是不可能的,而只能采取渐进过渡的方式,并保证每一步处理结果都要更加有利于主观语义的辨认,显然,这是十分困难的.实际上,人在观察图象时,首先关注的是前景对象和主要背景区域,即以主要区域的过滤为前提.也就是说,图象语义处理的首要工作就是提取图象中的主要区域,并据此区分前景和背景.

传统的图象分割方法,均是力求把图象中所有的

基金项目:国家自然科学基金(69903006),教育部高等学校骨干教师资助计划(教技司(2000)65 号;中国博士后科学基金(中博基(1997)11 号)
收稿日期:2001-11-27;改回日期:2002-07-29

区域都分割出来，这就导致太多的冗余信息. 例如，对于复杂图象，若使用基于边缘提取的方法，则将产生过多的边缘和区域，这样虽然把握住了细节，却忽略了图象整体的组成信息；而使用区域增长等方法，虽然可以控制区域的数目，但是同一区域中往往含有视觉上明显不相似的部分. 由此可见，这些方法对于基于区域的图象检索来说，复杂度大大增高，并会返回许多不相关结果；对于图象理解来说，则会干扰对象的识别，并会导致整幅图象语义的判定错误.

笔者认为：图象的语义更多地体现在主要区域的组成上，而不在于每个区域的细节上，而且图象前景和背景是两个层次的语义，其在不同的应用中，起着不同的作用. 据此，本文采取了一种先进行基于子块聚类的图象分割和分割结果优化，再通过分析不同区域在整个图象中的作用来鉴别图象整体区域分布合理性(即进行区域重要程度分析)的方法，以便据其去除与图象内容关系不大的信息，这样就形成了一种无监督、全自动，并具有一定自适应能力的图象主要区域提取方法，其工作流程如图 1 所示. 这样，既保证了处理速度，也便于进一步的处理.

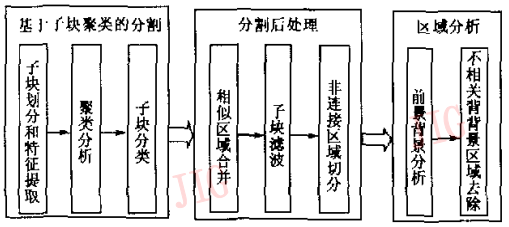


图 1 主要工作流程图

1 基于子块聚类的图象分割

本文利用了一个包含 3 400 个图象的图象库，所有试验用的图象均从 corel 光盘图象库中[3]选出，且都是户外的彩色照片，每一幅图象中都包含一个或几个前景对象，图象的大小是 384×256 或者 256×384.

由于一个图象可看成是由一系列区域组成的，每个区域又可看成是具有一定粒度的子块集合，因此同一个区域的子块在特征空间和几何空间具有一定的聚合性. 其中，特征空间的聚合性表现为，区域内的子块在颜色、纹理等特征上的相似性；而几何空间的聚合性就是同一区域的子块在图象中的位置比较靠近. 据此，图象的分割就可看成是先将图象划分成适当大小的子块，再对子块进行适当聚类的处理过程.

1.1 子块大小的确定

图象的聚类分割最直接的方法是基于像素的方法[4]. 一般像素级的聚类是将每个像素都当作样本，由于图象较大，如果直接计算的话，那么聚类速度明显比子块聚类要慢得多，因此，一般情况下，需要先将图象缩小，再进行处理. 这种处理与基于子块的聚类相比，实质上没有太大的区别，反而丢失了局部的纹理细节.

为了了解子块大小选择对聚类的影响，分别采用 32×32、24×24、16×16、8×8 子块，对本系统使用的 384×256 的图象进行划分比较. 4 种划分相应的子块数目是 96、160、384、1 536. 由于聚类方法需要保证一定数量的样本数目，由实验结果比较可见，32×32、24×24 子块分割的聚类效果明显不如 16×16 子块分割的效果好；16×16 子块的分割效果略好于 24×24 子块；8×8 子块产生的分割，则由于结合了太多的局部细节，使得区域划分过细而出现太多的分类，因此，过小的子块对于户外图象这种同类背景区域特征变化较大的图象不合适. 同时，由于要从子块提取纹理特征，若子块太大，则纹理的一致性不容易满足；而子块太小，则子块的纹理提取也没有意义，所以权衡两方面的要求，针对本系统的图象，本方法采用的子块大小是 16×16.

1.2 子块特征向量的形成

颜色和纹理是户外图象的主要语义视觉特征，也就是说，每个子块的特征是由颜色和纹理两类特征构成的. 其中颜色特征可通过提取子块的 HSV 颜色空间的一阶和二阶矩来获得[5]；而纹理特征则可借助于提取子块的对比度和粗糙度[6]来获取.

子块的特征用特征向量来表示. 由于随后的聚类分析需要有足够的样本数量，但这个数量与特征向量的维数有关，即维数越大，要求的样本数越多，因此，在已经根据固定的子块大小确定了样本数的前提下，必须限制特征向量的维数.

对于颜色矩来说，实验表明：一、二阶颜色矩已可较好地描述子块的颜色特征，并且每阶只有 3 个值，比较适合于目前的场合，没有必要再取高阶矩. 同样，粗糙度和对比度都只具有一个值，经过实验检验，其对于纹理也具有较好的区分度.

此外，本文的方法主要针对的是户外图象，其细节比较复杂，若在特征向量中，不利用子块的位置信

息，则会使大量几何上不连续的子块归成一类，从而导致分割结果出现众多的零星的小区域．为此，在特征向量中，需要加进子块的位置信息，这样，具有相近的几何位置的子块就更容易聚成一类，也更符合区域的几何一致性条件．

于是，一、二阶的 HSV 矩（6 个值）和粗糙度（1 个值）、对比度（1 个值）和子块在图象中的位置（x，y 两个值）形成了 10 维的子块特征向量，如图 2 所示．

H 均值	H 方差	S 均值	S 方差	V 均值	V 方差	粗糙度	对比度	位置 x	位置 y

图 2　子块特征向量

1.3　图象子块的聚类和子块分类

从聚类的角度来看，经过子块划分的每个区域的特征可以看成是一个多维高斯分布，而整个图象就是混合多维高斯分布，且每一个子块的特征向量就是一个样本．这样图象的区域分割就演变为混合多维高斯分布参数的确定问题．

混合多维高斯模型需要估计的参数包括每个多维高斯子类的参数和混合高斯分布的子类数．本文首先采用 Expectation Maximization（EM）来估计这些参数[7]，然后用 Minimum Description Length（MDL）来估计合适的聚类数目[8]，并利用 Cluster 软件包[9]来具体完成参数估计过程．

经过聚类分析后，每一个类别就对应于图象的一个区域．但是目前只获得了每一个区域的特征分布描述，具体哪个子块对应于哪个区域，还需要通过估计出的先验概率和每一类的类条件概率密度以及计算出每一个子块属于各个分类的后验概率来判断．这样就获取了初始的图象分割结果．

2　子块分割后处理

由于聚类分割方法内在的一些特性，致使子块分割的初始结果还存在许多不足：（1）由于利用了位置信息，因而对于一个比较大的区域，则会因子块位置差别比较大而往往被分成几个较小的区域；（2）对于区域之间交界处的子块，常常会用单独聚类出的一个类来代表，或者产生许多零星小块，而对于特征上特别相似的区域，虽然空间上不连续，但由于有时也会被分到同一类别中，因此需要进行一系列处理来优化初始分割．

2.1　合并同类区域

对于实际是一个区域，但因在特征向量中，加入了位置而被分成多个区域的情况，应予以合并．直观的想法是，若聚类时不利用子块的位置，则这些区域应该属于一类．本文的合并方法是首先将特征向量中的子块位置去掉，仅使用剩下的 8 维向量，并利用 1.3 节的方法再进行一次聚类分割；然后，对于第 1 种聚类获得的分割，再通过判断其任意两个区域，即判断在第 2 种聚类获得的分割中，是否大多数子块属于一个区域，若是的话，就将两个区域合并．通过这种简单的方法，就可把大多数应合并的区域合并起来．

2.2　子块滤波

对比较复杂的图象进行聚类分割时，由于不可避免会产生一些零星小区域，且其中有些是对所属区域隶属度不高的子块（类别边界子块），有些是几个大区域交界处的临界子块（空间边缘子块），因此应利用基于子块的滤波来消除以上两种情况中的大部分的零星小区域．具体方法是，对于每一个子块，统计它本身及其 8 邻接子块所对应的类别，若 9 个子块中，属于第 j 个分类的最多，则将其也归入第 j 类．

2.3　非连接区域区分

由于前景对象一般比较复杂，有时候背景的一个区域会与前景的一个部分在特征上非常接近，虽然它们在空间上不连续，可仍然被分到一个类别，但是对接下来的前景背景分析来说，它必须将几何上分离的每一个区域都看成是不同的，否则会导致错误的区分，因此，将这些几何上分离，但被分到同一类别的区域转换成不同的区域是必要的．本文采用了连通成分标记算法来进行区分和转换，具体请见文献[10]．

3　图象区域重要度分析

根据人们的一般认识习惯，图象中的不同区域，其重要性是不同的．为更好地进行语义分析，将图象的前景和背景区分开来很有必要，而对于图象的背景来说，由于只需要图象的主要背景对象，因此需要通过一些分析来去掉与图象内容不太相关的背景区域．

3.1　图象前景背景分析

前景对象一般主要位于图象中央，而背景主要处于图象的其他部分．这种假设对于由包含显著对象的户外场景照片组成的图象集更加适用．在进行前景背景分析时，一般是将图象划分成如图 3 所示的 9 个分块[11]，其中，分块 5 称为中央分块，分块 1，

3,7,9 称为角部分块,分块 2,4,6,8 称为边缘分块,分别标记为 $\boldsymbol{B}_i, i=1,2,\cdots,9$. 设图象宽度为 W,高度为 H,则中央分块的大小为 $(\omega_1\times W)\times(\omega_2\times H)$. ω_1、ω_2 分别表示中央分块的宽度和高度与整个图象的宽度和高度的比例.

1	2	3
4	5	6
7	8	9

图 3 将图象分成 9 个部分做背景分析

经过前述的图象分割和后处理,对于生成的任何一个区域 $\boldsymbol{R}_i$,定义如下变量:$r_{i,j}$ 是 $\boldsymbol{R}_i$ 在 $\boldsymbol{B}_j$ 中的子块数与 $\boldsymbol{B}_j$ 的大小的比值,$\boldsymbol{B}_j$ 的大小指的是它的所有子块数目;若 $r_{i,j}>\theta_a$(θ_a 是一个阈值),则称 $\boldsymbol{R}_i$ 占据 $\boldsymbol{B}_j$;N_{i_c} 是 $\boldsymbol{R}_i$ 占据的角部分块数;N_{i_e} 是 $\boldsymbol{R}_i$ 占据的边缘分块数;p_{i_5} 是 $\boldsymbol{R}_i$ 在 $\boldsymbol{B}_5$ 中的子块数与 $\boldsymbol{R}_i$ 大小的比值.

在进行图象前景背景分析时,首先设定 ω_1 和 ω_2 的初始值,再使用下面的步骤来判断其是否是背景:

(1) 若 $N_{i_c}>2$ 并且 $p_{i_5}<\theta_1$,或者 $r_{i_5}<\theta_2$,其中 θ_1 和 θ_2 为某个阈值,则标注该区域为背景.

(2) 若 $N_{i_c}\geq 1$ 或者 $N_{i_e}\geq 1$,并且 $p_{i_5}<\theta_3$,其中 θ_3 为阈值,则标注该区域为背景.

(3) 若 $N_{i_c}\geq 1$ 或者 $N_{i_e}\geq 1$,并且 $r_{i_5}<\theta_4$,其中 θ_4 为阈值,则标注该区域为背景.

(4) 若 $N_{i_c}=0$ 或者 $N_{i_e}=0$,并且 $\boldsymbol{R}_i$ 不占据 $\boldsymbol{B}_5$,$p_{i_5}<\theta_5$,其中 θ_5 为阈值,则标注该区域为背景.

(5) 若所有的区域都被标注成背景,则增大 ω_1 和 ω_2 的值,返回步骤 2,否则算法结束.

这样,前景和背景就可以大致分割出来了,但对于复杂的前景对象,由于其分割出来的前景可能是多个区域的组合,因此还需在其上进行进一步的处理.

以上的前景背景区分方法,需要根据所处理的图象集来确定参数的设置. 本文针对使用的户外场景照片图象集,同时根据试验的效果,将 ω_1 和 ω_2 的初值都设为 0.6. 在进行前景背景区分的迭代运算中,可以认为当 ω_1 和 ω_2 的值超过一定范围(>0.9),这样图象就不包含前景对象,但是总的来说,本方法主要是针对前景和背景区分度较大的图象而进行的前景、背景区分,并不是一种通用的图象解决方案.

3.2 背景不相关区域去除

对于图象检索和理解,人们总是希望分割出来的背景区域只包含主要的背景对象. 由于与这些主要背景对象对应的区域一般都比较大,同时区域内部的子块之间的相似性也比较高,因此,要找出满足以上特点的背景区域,并加以保留,同时要过滤掉其他的背景.

对于每一个背景区域,首先计算其面积,并去除面积小于一定阈值的区域,由于从聚类角度来说,区域内部子块之间的相似性与子块样本相对于子块均值的离散度成反比[12],且对于样本的离散度而言,它与样本协方差矩阵的行列式的平方根成正比,因此可将剩余的每一个区域当成多维正态分布的模型来估计其协方差矩阵,再将协方差矩阵行列式大于一定阈值的区域去除. 最后获得的就是背景的主要区域.

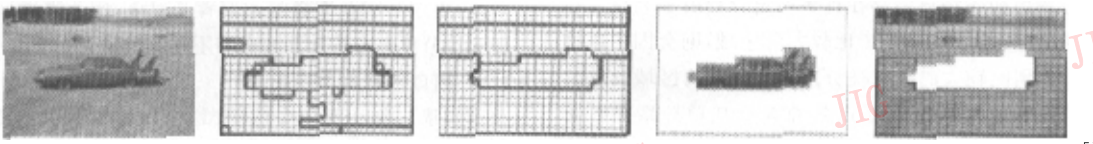
4 实验结果和比较

4.1 实验结果和实验系统

上述工作属于图象语义预处理,它构成了图象复杂检索和图象语义处理的基础. 下面是两组代表性图象的实验结果.

图 4 为本文方法对简单图象的提取结果,其初始分割结果(图 4(b))已较好地对应了对象的区域,经过后处理后的分割结果(图 4(c))和区分出的前景(图 4(d))背景(图 4(e)),与人的主观感觉比较接近.

图 5 是一幅复杂图象,图 6~图 9 是本文方法对该复杂图象的提取结果,其初始分割结果(图 6(a))是杂乱无章的,而直接基于此分割所获得的前景(图 6(b))和主要背景(图 6(c))也是没有意义的;而后面的每一步处理,都使分割结果更加合理,也间接改善

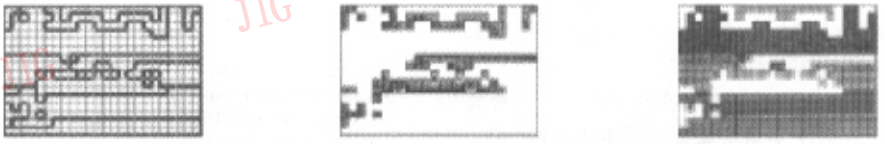


(a) 原始图象 (b) 经过子块聚类后得到的初始分割结果 (c) 图 4(b)经优化后的结果 (d) 分离出来的前景区域 (e) 主要的背景区域

图 4 一种简单图象主要区域提取结果

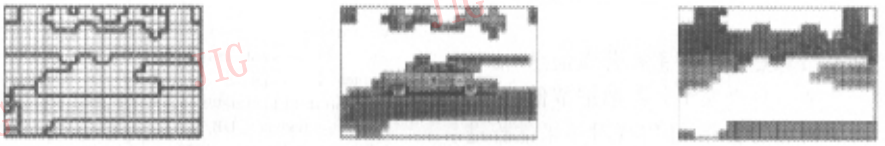
图 5　一幅复杂的原始图象

了图象的前景背景分析：经过第 1 次子块滤波，前景的汽车已经合成为一个整体(图 7(b))，其提取出来的主要背景(图 7(c))也相对完整，且每个区域的组成也比较单一．经过相似区域合并，再进行一次滤波后，分离前景时(图 8(b))还会产生不相关区域，而此时主要背景(图 8(c))都基本出现；最后经过非连接区域切分，分离出的前景(图 9(b))仅仅包含汽车，而分离出的主要背景(图 9(c))则包含了所有背景区域，并且每个区域都只包含一类背景对象．经过这一系列处理，总的来说，获得的结果比较合理．



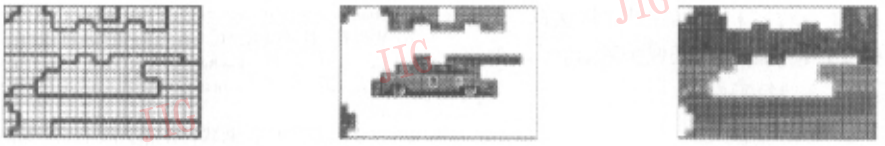
(a) 初步分割结果　(b) 经前景、背景分析后得到的前景区域　(c) 去除了不相关背景后的主要背景区域

图 6　图 5 经过子块聚类后的初始分割结果



(a) 初步分割结果　(b) 经前景、背景分析后得到的前景区域　(c) 去除了不相关背景后的主要背景区域

图 7　图 5 进行了一次子块滤波后的分割结果



(a) 初步分割结果　(b) 经前景、背景分析后得到的前景区域　(c) 去除了不相关背景后的主要背景区域

图 8　图 5 进行了相似区域合并后，再进行一次滤波后的结果

(a) 初步分割结果　(b) 经前景、背景分析后得到的前景区域　(c) 去除了不相关背景后的主要背景区域

图 9　图 5 经过非连接区域切分后的结果

4.2　相关工作比较

本文的工作包括图象分割和区域分析，下面就从这两个方面与其他的方法进行比较．

文献[4]提出的 Blobworld 方法中使用了基于像素的图象聚类分割方法，输入系统的图象必须较小，对于 192×128 大小的图象，在 Pentium 4，256M 的机器上要计算好几分钟，而本文的算法则是实时的．在没有经过分割后处理前，本文方法的分割效果虽不如文献[4]方法，但是经过后处理，效果就基本差不多．另外，文献[4]方法没有将聚类后的同类区域合并，这一点也不如本文的方法．

文献[13]中提出的 JSEG 方法将分割分成颜色

量化和空间分割两步. JSEG 方法的分割结果是像素轮廓,其处理速度也比较快,单从分割角度来论,优于本文的方法,但是 JSEG 方法的分割效果与它的参数设置关系很大,即需要根据不同的图象类型,通过实验来手工确定;而本文的方法则是全自动和自适应的. 另外,单纯 JSEG 的自动分割,仍然无法获得对象的精确轮廓,而用子块表示的粗分辨率区域,一个内在的优点是便于人工干预,即可以精确指定任一个子块归属的区域. 也就是说,用人工方法修正了错误分割的子块后,再用活动轮廓模型或可变模板,可以获得对象的精确轮廓.

文献[11]的分离前景和背景的方法,由于参数是限定的,因此对于前景对象太大或者太小的情况,不能很好地处理. 本文的前景背景区分,不仅能检测小的前景对象,并且通过迭代,还可不断改变中央子块的大小,也就是说,对于大的前景对象,也有很好的检测效果.

本文的分割结果是图象的主要分布信息. 由于子块分割的一个缺点是丢失了分割的细节信息,因此对于需要精确轮廓的场合,还需要对结果进行进一步的处理. 但是在把握全局的区域组成方面,本文的方法更为出色. 另外,对于图象检索而言,区域的子块表示一般已经满足要求了. 以往的方法中,一般都没有将与图象内容不相关的背景区域去除. 实验证明,将这些没有意义的背景区域去除后,可以在一定程度上提高检索的精确性.

5 结束语

图象的主要区域及其组成关系是图象语义提取的基础. 本文针对户外图象,提出了一种面向图象语义的主要区域提取方法. 归结起来,该方法具有以下特点:(1)由于利用固定划分的子块进行聚类,能使得图象分割的重点放在图象的整体区域分布上,而通过一系列后处理,又使得每个区域的组成比较完整,并且还尽量包含了一致的视觉特征,因而可确保对图象语义的有效抽取;(2)由于该方法将图象的前景和背景区域进行了区分,因而使得进一步的语义处理和应用比较方便;(3)由于它去除了与图象内容相关性不大的背景区域,因而使得结果只包含图象的主要组成部分. 总之,由于该方法更多地考虑了图象语义应用的可能需求,因而不仅避免了传统图象分割产生的不必要信息处理和冗余,而且包含了更多有用的信息. 本文所提出的方法已经嵌入到笔者等开发的"视觉语义分析与建模系统 VisEngine"[14]中,实验结果表明,该方法还可以比较明显地提高对象的识别和检索效果.

参 考 文 献

1 Gudivada V N, Raghavan V V. Content-based image retrieval system[J]. IEEE Computer, 1995,28(9):18~22.

2 王惠锋,孙正兴. 基于内容的图象检索中的语义处理方法[J]. 中国图象图形学报,2001,6A(10):945~952.

3 Wang J Z, Li J, Wiederhold G. Corel1m database used in SIMPLIcity [DB/OL]. http://wang.ist.psu.edu/docs/related/, 2000.

4 Carson C, Belonge S, Greenspan H *et al*. Blobworld: Image segmentation using expectation-maximization and its application to image querying [DB/OL]. http://citeseer.nj.nec.com/45960.html, 1999.

5 Stricker M, Orengo M. Similarity of color images [A]. In: Proc. of SPIE storage and retrieval for image and video Databases Ⅲ [C], San Jose, CA, IS&T and SPIE, 1995,2420: 381~392.

6 Hideyuki T, Shunji M, Takashi Y. Textural features corresponding to visual perception[J]. IEEE Trans on Systems, Man, and Cybernetics, 1978, SMC-8(6):460~1473.

7 Bilmes J. A gentle tutorial of the EM algorithm and its application to parameter estimation gaussian mixture and hidden Markov models [DB/OL]. http://citeseer.nj.nec.com/bilmes98gentle.html, 1998.

8 Rissanen J. Modeling by shortest data description [J]. Automatica, 1978,14:465~471.

9 Bouman C A. An unsupervised algorithm for modeling gaussian mixtures [DB/OL]. http://dynamo.ecn.purdue.edu/~bouman/software/cluster/, 2000.

10 贾云得. 机器视觉[M]. 北京:科学出版社,2000:36.

11 Lu Y, Guo H. Background removal in image indexing and retrieval[DB/OL]. http://citeseer.nj.nec.com/332981.html, 1999.

12 边肇祺,张学工等. 模式识别(第二版)[M]. 北京:清华大学出版社,2000:26.

13 Deng Y, Manjunath B S, Shin H. Color image segmentation [A]. In Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)[C], Fort Collins, Colorado USA, IEEE CS press, 1999:446~451.

14 王惠锋. 基于语义的图像检索系统及其关键技术研究[D]. 南京:南京大学计算机科学与技术系, 2002.

王惠锋 1976 年生,南京大学计算机科学与技术系硕士研究生. 主要研究方向为图象理解、图象检索、视觉信息挖掘.

孙正兴 1964 年生,1996 年获南京航空航天大学博士学位,现为博士后研究人员,副教授、硕士生导师. 研究方向为视觉信息挖掘、多媒体辅助工程. 发表论文 40 余篇.