# 基于 SIFT特征的图像检索 *

吴锐航, 李绍滋, 邹丰美
(厦门大学 计算机系人工智能研究所, 福建 厦门 361005)

**摘 要**: 提出一种多尺度图像检索算法,该算法基于 SIFT特征提取,它将一幅图像转换成特征向量的集合,图像间的相似距离是通过计算两幅图像特征向量间的欧氏距离来实现的。实验结果很好地说明了该算法具有尺度、平移、旋转不变性,一定的仿射、光照不变性以及算法能很好地应用在特定形状特征目标的检索中。

**关键词**: 基于内容的图像检索; 尺度不变特征变换特征; K-L变换; 近似最近邻搜索; 欧氏距离
**中图分类号**: TP391. 3 **文献标志码**: A **文章编号**: 1001-3695(2008)02-0478-04

# Image retrieval based on SIFT features

WU Rui-hang, LI Shao-zi, ZOU Feng-mei
(*Institute of Artificial Intelligence, Dept of Computer Science, Xiamen University, Xiamen Fujian* 361005, *China*)

**Abstract:** This paper presented a new image retrieval algorithm which used a new class of image features as the image descriptors, the SIFT features. It transformed the image into a set of SIFT features. The similarity was computed using Euclidean distance. Experiments show that the features are invariant to image scaling, translation, rotation, and change in viewpoint, and partially invariant to illumination changes and affine. It is quite applicable to typical image retrieval.

**Key words:** content-based image retrieval; SIFT(scale invariant feature transform) features; K-L transform; ANN searching; Euclidean distance

基于内容的视觉信息检索是由于视觉信息的飞速膨胀而得到关注并被提出来的。如何快速准确地提取视觉信息内容是视觉信息检索最关键的一步。当今流行的一些图像检索系统多利用图像的底层特征,如颜色、纹理、形状以及空间关系等。这些特征对于图像检索有着不同的作用,但是同时也存在着不足。颜色特征是一种全局特征,它对图像或图像区域的方向、大小等变化不敏感,所以颜色特征不能很好地捕捉图像中对象的局部特征,也不能表达颜色空间分布的信息。它也是一种全局特征,纹理特征只是物体表面的一种特性,并不能完全反映物体的本质属性。最严重的一个缺点是,当物体受到光照、反射等影响时,从二维图像中反映出来的纹理不一定是物体真实的纹理。这些虚假的纹理均会对检索造成误导。全局的特征描述并不能很好地适应在对目标感兴趣的图像检索。基于形状的特征常常可以利用图像中感兴趣的目标进行检索。但是形状特征的提取,常常受到图像分割效果的影响。空间关系特征可以加强对图像内容的描述和区分能力,但空间关系特征对图像或者目标的旋转、平移、尺度变换等比较敏感,并且不能准确地表达场景的信息。图像检索领域急需一种能够对目标进行检索,并且对图像目标亮度、旋转、平移、尺度甚至仿射不变的检索算法。本文就是在这样的基础上提出的。

SIFT是 David G. Lowe 2004年总结不变量技术的特征检测方法,提出的一种对尺度空间、图像缩放、旋转甚至仿射不变的图像局部特征描述算子。

SIFT特征向量有如下特性: a)局部性。SIFT特征向量是一种局部特征,其对旋转、尺度缩放、亮度变化保持不变,并且对图像目标受到部分遮挡的情况下也有很好的检索性能。b)特殊性。SIFT特征向量包含丰富的图像内容信息,适合在海量的目标数据库中进行检索、匹配。c)多量性。仅有少量的一个目标也可以产生大量的 SIFT特征向量。d)高效性。SIFT能够快速地从图像中提取出来,并且实现高效的特征匹配。

本文提出了一种基于 SIFT特征提取的图像检索算法,并结合 K-L变换对该特征向量进行主成分分析,降低向量空间维数,减少计算量。图像间的相似特征向量的搜索由 ANN搜索算法实现。文中描述了 SIFT特征向量的提取过程,并给出三个不同的实验来检测该算法的检索性能。

## 特征提取算法

### . 图像的多尺度表示

多尺度技术也称为多分辨率技术。多尺度图像技术指对图像采用多尺度的表达,并在不同尺度下分别进行处理。在很多情况下,图像中某种尺度下不容易看出或获取的特性在另外的尺度下很容易看出来或检测到。所以利用多尺度常可以更有效地提取图像特征,获取图像内容。对同一幅图像用不同的尺度表达后,相当于给图像数据的表达增加了一个新的坐标。即除了一般使用的空间的分辨率外,现在又多了一个刻画当前分辨率层次的新参数。如果用 σ来标记这个新的尺度参数,则包含一系列有不同分辨率的图像的数据结构可被称为尺度

**收稿日期**: 2006-12-15; **修回日期**: 2007-03-21 **基金项目**: 国家自然科学基金资助项目(60373080); 厦门大学"985"二期信息技术创新平台资助项目(2004-2007); 厦门大学院士启动基金资助项目

**作者简介**:吴锐航(1981-),女,福建福州人,硕士研究生,主要研究方向为图像检索(wrh. fang@gmail. com);李绍滋(1963-),男,湖南常德人,教授,博导,主要研究方向为人工智能和智能信息检索、网络多媒体与 CSCW;邹丰美(1969-),男,福建上杭人,副教授,博士,主要研究方向为代数不变量的计算和应用、计算机视觉.

空间(scale space)。可用 $G(x, y, \sigma)$来表示图像 $G(x, y)$的尺度空间。Lindeberg在文献[2]中证实在一些一般性的假设条件下,高斯核和高斯微分是尺度空间分析的惟一平滑核。

## . 特征描述

David G Lowe在文献[3]中提出一种称之为 SIFT的图像局部特征描述算子,并通过实验证实了该算子对图像的尺度缩放、平移、旋转变换保持不变,对图像的亮度变化以及仿射变换也具有相当的稳健性(robust)。

SIFT算法的本质就是从图像中提取 SIFT关键点(key-point)的过程。SIFT特征提取的算法包含四步骤。

### . . 尺度空间的极值检测

为了达到尺度不变性,理论上必须在图像所有可能的尺度空间中搜索不变的特征,但是实际上这是不可能的。解决方法就是对尺度空间图像进行亚采样。将平滑和亚采样重复进行,就可得到能构成金字塔的一系列图像。如下定义二维高斯滤波函数。其中 σ表示高斯函数的方差。

$$G(x, y, \sigma) = 1/2\pi\sigma^2 \, e^{-(x^2+y^2)/2\sigma^2} \tag{1}$$

一幅 $N \times N$的图像 $I(x, y)$,在不同尺度空间下的表示可以由图像与高斯核卷积得到 Gaussian图像:

$$L(x, y, \sigma) = G(x, y, \sigma) \times I(x, y) \tag{2}$$

其中:σ称之为尺度空间因子,其值越小表示图像被平滑得越少。大尺度对应图像的概貌,小尺度对应图像的细节。DoG算子定义为

$$D(x, y, \sigma) = [G(x, y, k\sigma) - G(x, y, \sigma)] \times I(x, y) \tag{3}$$

采用 DoG算子建立的尺度空间如图 1所示。

图 1左边显示的 Gaussian金字塔中,最底层就是没有作过高斯滤波的原始图像。高斯金字塔分为 $n$阶(octave),每一阶分为 $s^1$层。为了在这 $s$层上检测特征点,需要产生 $s+2$幅 DoG图像,对应 $s+3$幅 Gaussian图像。这 $s+3$幅 Gaussian图像从下到上尺度比例以 $k$递增,即是说,若当前的 Gaussian图像的尺度参数为 σ,则下一层的 Gaussian图像的尺度参数为 $k\sigma$。为了要让一阶之间有 $s$层,本文规定 $k = 2^{1/s}$。将同一阶中的 $s+3$层的 Gaussian图像,相邻的图像相减就产生了相应的 $s+2$层 DoG图像。

为了产生下一阶的 Gaussian图像,对上一阶产生的最后一幅 $N \times N$的 Gaussian图像进行 1∶2的亚采样,生成 $N/2 \times N/2$的图像。在亚采样图像基础上重复上述建立 $s+3$层 Gaussian图像的过程。

为了检测 $D(x, y, \sigma)$的局部极值点,需要比较 DoG图像中每个像素与其 26个近邻像素的值。若像素 $(x, y)$是一个可能的 SIFT关键点,则它必须在它周围的 26个近邻像素(上一个尺度的 9个点 +同尺度的 8个点 +下一个尺度的 9个点)中是极值点,如图 2所示。所有这样的局部极值点,就构成了一个 SIFT候选关键点的集合。

### . . 关键点的确定

极值检测得到的所有候选关键点,还必须通过两步检验才能确定是关键点:一是它必须与周围的像素有明显的差异,也就是说低对比度的点不要;二是它不能是边缘点。

1)低对比度的点

通过拟合三维二次方程,可以找出低对比度的点。对 $D(x, y, \sigma)$进行泰勒展开到二次项:

$$D(x) = D + (\partial D^{T}/\partial x)\, x + 1/2\,[x^{T}(\partial^2 D/\partial x^2)\, x] \tag{4}$$

其中:$D$是 DoG计算的结果;$x$是候选关键点之一。由 $D$和 $x$,可以找到一个偏移量:

$$\hat{x} = -(\partial^2 D^{-1}/\partial x^2)(\partial D/\partial x) \tag{5}$$

将式(5)代入(4)中,如果计算得到的 $|D\hat{x}| < 0.03$,则该点是低对比度的点[4]。

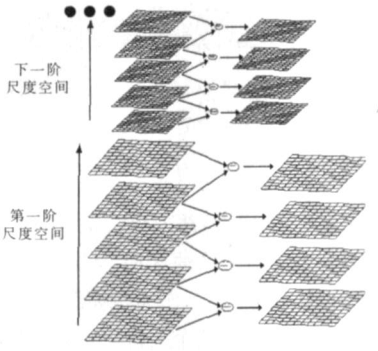


图 1 Gaussian 金字塔和 DoG 金字塔的建立过程

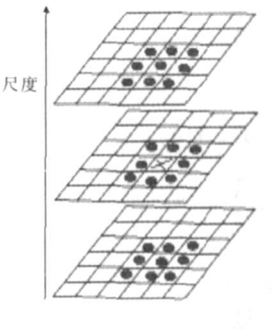


图 2 尺度空间极值点的确定

2)边缘处的点

类似 Harris边缘检测的方法,计算 Hessian矩阵:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \tag{6}$$

令

$$\begin{aligned} \mathrm{Tr}(H) &= D_{xx} + D_{yy} = \alpha + \beta \\ \mathrm{Det}(H) &= D_{xx}D_{yy} - (D_{yy})^2 = \alpha\beta \end{aligned} \tag{7}$$

$$\mathrm{Tr}(H)^2/\mathrm{Det}(H) = (\alpha+\beta)^2/\alpha\beta = (r\beta+\beta)^2/r\beta^2 = (r+1)^2/r$$

若点不满足式(8),则该点可能是边缘的点,可以将它剔除。

$$\mathrm{Tr}(H)^2/\mathrm{Det}(H) < (r+1)^2/r \tag{8}$$

### . . 关键点的大小和方向

为了使算子具备旋转不变性,采用梯度直方图来确定关键点的主方向。点 $(x, y)$处梯度的模值和方向的计算公式为

$$\begin{aligned} m(x, y) &= \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \\ \theta(x, y) &= \tan^{-1}((L(x, y+1) - L(x, y-1))/ \\ &\quad (L(x+1, y) - L(x-1, y))) \end{aligned} \tag{9}$$

对于每一个关键点,本文考虑它的邻近的一个邻域窗口内点的梯度方向,最多人投票的方向就当做该点的主方向。而每个邻近的点对中央这个点的权重由高斯分布再乘上该点的梯度的大小来决定。

如果梯度直方图中存在另一个相当于主峰值 80%的峰值时,则将整个方向认为是该关键点的辅方向。一个关键点可能被指定多个方向(一个主方向,一个以上的辅方向)。这种情况下,就可复制同一个关键点,使它们分别朝向不同的方向。

### 1.2.4 SIFT特征向量描述子

为了确保旋转不变性,首先将坐标轴旋转为关键点的方向。以一个关键点为中心,取 8×8的窗口,将该窗口切成 2×2的子窗口,统计每个子窗口中的方向直方图,如图 3所示。

每一个子窗口的方向由其上 4×4的小块的方向用之前的方法来决定。图中的每个关键点方向由 2×2共 4个种子点的方向决定,一个种子点有 8个方向的信息,则每个关键点就有 4×8=32维。

在实际计算过程中,为了增强匹配的稳健性,通常采用 4×4共 16个种子点来描述,这样一个关键点就有 16×8=128维的数据,形成 128维的 SIFT特征向量。图 4是一幅标记了 SIFT特征向量的图像,箭头的起点、长度、方向表示了关键点

的位置、该点的梯度大小以及梯度的方向。

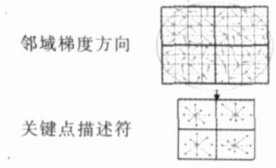


图 3 由关键点邻域的梯度信息生成的特征向量

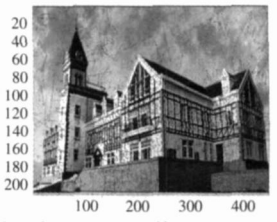


图 4 标定了 SIFT 特征向量的图例

## 图像的检索算法

128维的向量有力地刻画了图像的局部内容特征,并且具有很强的局部特征匹配能力。当两幅图像的 SIFT特征向量生成后,笔者采用关键点特征向量的欧氏距离作为两幅图像中关键点的相似性判断度量。取查询图中的某个关键点,找出其与数据库图像中欧氏距离最近的前两个关键点。在这两个关键点中,如果最近的距离除以次近的距离少于某个比例阈值(distance ratio),则接收这一对匹配点。降低比例阈值,匹配的点的数量就会减少,匹配的精确度就会提高。适当的阈值可以减少采样过程和图像背景因素引起的错误匹配数量。

### . 特征向量的匹配

一幅图像中可能有成千上万的 SIFT特征向量,对于这 128维的向量来说,要找出数据库图像中与之距离最小的向量,其计算量可想而知是十分巨大的。为了解决高维空间搜索问题,本文采用的解决方法是:a) K-L 变换降低向量的空间维数;b) 利用 ANN搜索(approximate nearest neighbor searching)算法实现图像中最近邻向量的快速搜索。

#### 2.1.1 K-L变换

K-L 变换又称为主成分分析(PCA),通常应用在图像的数据压缩中。离散的 K-L 变换是针对向量的变换。设 $N$ 维的矩阵 X,其均值向量为 $M_X$,设 K-L的变换矩阵 A是由 $X$的协方差矩阵的特征向量按特征值递减顺序排列组成的变换核矩阵。则典型的 K-L 变换写为

$$Y = A(X - M_X) \tag{10}$$

由于能量集中在特征值 λ比较大的系数中,因此可以舍弃对应特征值 λ较小的 $Y$系数而不会影响图像的质量。取 λ最大的前 $k$个特征向量组成的变换核矩阵,记做 $A_k$,则有

$$Y' = A_k(X - M_X) \tag{11}$$

这相当于原 $y$的 $k$维投影。在本文的实验中,取 $k = 20$,这样就使 128维的向量降到了 20维。

#### 2.1.2 ANN搜索算法

为了准确匹配特征向量,对于查询图中的每一个 SIFT特征向量,需要比较它与查询图中所有特征向量的欧氏距离。明显地,穷举搜索可以实现精确定位,但实际应用中却十分低效。在准确性与效率之间作个权衡,本文选择 ANN搜索算法。

ANN近似最近邻搜索算法,由 Arya等人[6]提出,为了解决高维空间的快速搜索问题。ANN搜索算法基本思想是:给定一个 $d$维空间的点集 $P$,算法将数据组织成 $kd$树结构,对于每一个查询的点集 $Q$,ANN算法能够快速返回 $P$与 $Q$中各点距离最近的 $k$个点对。

### . 图像的相似距离

若关键点匹配的数量越多,图像的相似程度就越高。在实验中,使用匹配程度百分数来表示相似程度。笔者规定,匹配程度百分数 =图像匹配的关键点总数 /查询图像中总的关键点的数量。

## 实验及其结果分析

首先,测试算法是否可以实现相似检索。选用四幅内容相关的图像。一幅全景图像,给出三幅包含全景图像中部分景色的图像。计算全景图像与其三幅片段图像之间的相似距离。算法给出的结果如图 5所示,结果按相似程度从大到小排序,图像结果下方显示了匹配程度百分数。很明显,算法给出的结果也符合人类的视觉感受。

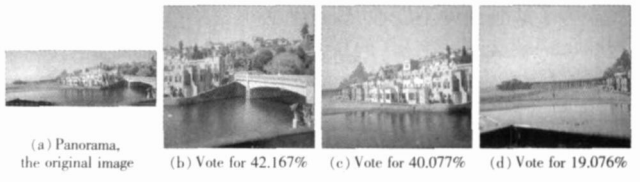


图 5 一幅全景图像与它的片段间的相似距离

从实验 1返回的结果可以看出,查询图与它的片段图像之间有很好的相似距离。如果仅是利用颜色、纹理、形状或者空间关系等图像特征,将有可能得到完全不同的结果。从实验返回的结果也可以看出,SIFT特征完全不同于颜色、纹理、空间关系等全局特征,是一种局部区域特征。

在实验 2中,本文选用一个特定的图像数据库进行检索,该数据库包含了近 300幅的军用飞机图像。数据库中图像的分辨率均为 1 024 ×1 536。检索结果按照相似度从左到右、从上到下排列,如图 6所示。

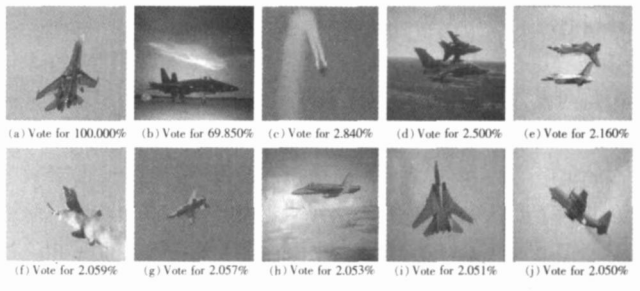


图 6 在军用飞机图像库中的检索结果

从实验 2的结果可以看出,该算法可以很好地描述查询图中目标的内容,并且确实具有尺度缩放、平移、旋转、亮度不变的优点,对于一定程度的仿射变换也保持较好的稳定性。

实验 3,笔者在一个包含 4 000幅各类图像的综合图像数据库进行检索,图像大小均为 1 024 ×1 536。库中包含了各类常见的图像,人物、花草、上水、建筑、军事等。实验返回前 10幅相似程度最高的图像,如图 7所示。最上方的图像为查询图,也是结果返回的拥有最近相似距离的图像。

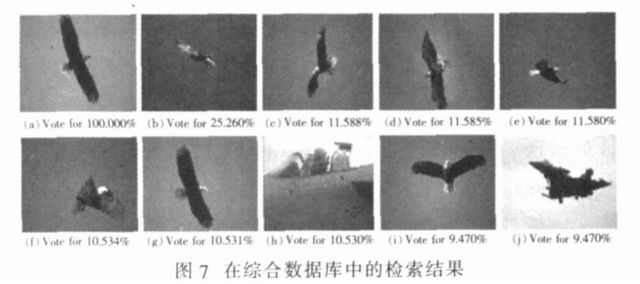


图 7 在综合数据库中的检索结果

实验 3的结果是令人兴奋的,除了第 8和第 9幅是不相关的图像外,其余 8幅都是正确的关联图像。图像库中有上百张

的飞禽图像,但仅有 15张鹰的图像。这里检出了 8幅,事实上,剩余的 7幅图像也在最相似的前 20幅图像中检出。结果中出现了两幅不相关的图像,其原因主要有以下几点: a)该算法与图像中目标形状关系密切,如果目标形状发生了较大的改变,则该算法就不容易将其检出。后 7幅的鹰的图像实际上就是与查询图中目标姿态上存在较大区别。 b)在计算两个关键点的相似距离时,比例阈值的选择对检索结果的精确性有很大影响。比例阈值越小,匹配的数量就越少,但匹配的精确度高,实验中,比例阈值选择为 0.6。 c)在特征向量的降维过程中,对向量包含的信息量不可避免地造成损失,这也对图像匹配过程中的稳健性造成影响。

## 结束语

本文针对图像检索领域在图像目标发生一定变化情况下不能很好实现检索问题,提出一种新颖的多尺度图像检索算法,基于 SIFT特征提取的图像检索算法。该算法将图像内容转换成 128维特征向量的集合,通过计算向量之间的欧氏距离实现匹配。实验结果表明该算法能较好地描述图像内容,具有较好的尺度缩放、平移、旋转、亮度不变性,以及一定程度的仿射不变性。

事实上, SIFT的局部特性对于图像中目标被部分阻隔的情况也有很好的检索效果。如果能够将该图像特征与图像的全局特征如颜色、纹理、空间关系等相结合,将能得到更好的检索效果。从实验 2和 3可以看出,该算法尤其适用于图像中有特定目标的检索,如果数据库图像中有部分片段与查询图中相同,或者说是相似,该算法就有很高的概率将其检索出。算法也有不适用的地方,比如对草图的查询,其主要原因是草图绘制过程中,对目标形状的描述可能会有很大的误差,并且草图提供的图像信息一般过于简单,缺少必要的区域梯度信息。

SIFT特征向量结合 ANN搜索算法使得整个检索过程稳定而且高效,甚至在标准配置的个人电脑上均可达到实时的效果。

该检索算法有着很大的应用前景,如对徽标的检测,给出一个徽标图像,就可以精确地查询到数据库中是否有相同的徽标;或者应用在指纹识别中,由于指纹图案有着明显的梯度信息,控制尺度空间参数以及采样频率,可以达到精确识别效果;也可以应用在武器库图像检索中,如果给定一张包含了某一武器设备的查询图像,该算法可以以很高的概率在武器图像库中检索该武器。

进一步的研究计划是将图像 SIFT特征与全局特征相结合,以寻求更好的检索效果和系统的稳定性。

**参考文献**:


[1] 章毓晋. 基于内容的视觉信息检索 [M]. 北京:科学出版社, 2003.

[2] TONG L. Scale-space theory: a basic tool for analyzing structures at different scales[J]. Journal of Applied Statistics, 1994, 21(2): 224-270.

[3] LOW D G. Object recognition from local scale-invariant features [C]//Proc of the 7th IEEE International Conference on Computer Vision. Kerkyra, Greece: [s.n.], 1999: 1150-1157.

[4] LOWE D G. Distinctive image features from sacle-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[5] JORDAN M I. Properties of kernels and the Gaussian kernel[M]. Topics in Learning and Decision Making, 2004.

[6] ARYA S, MOUNT D M, NETANYAHU N S, *et al*. An optimal algorithm for approximate nearest neighbor searching[J]. Journal of the ACM, 1998, 45(6): 891-923.

[7] BEIS J, LOWE D G. Shape indexing using approximate nearest neighbour search in high-dimensional spaces[C]//Proc of Conference on Computer Vision and Pattern Recognition. 1997: 1000-1006.

[8] HARRIS C, STEPHENS M. A combined corner and edge detector [C]//Proc of the 4th Alvey Vision Conference, 1988: 147-151.

[9] CROWLEY J L, PARKER A C. A representation for shape based on peaks and ridges in the difference of low-pass transform [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1984, 6(2): 156-170.

[10] FRIEDMAN J H, BENTLEY J L, FINKEL R A. An algorithm for finding best matches in logarithmic expected time[J]. ACM Trans on Mathematical Software, 1977, 3(3): 209-226.

[11] 章毓晋,图像工程(上册)——图像处理 [M]. 2版. 北京:清华大学出版社, 2006.

[12] 欧珊瑚,王倩丽,朱哲瑜. Visual C++.NET数字图像处理技术与应用 [M]. 北京:清华大学出版社, 2004.

[13] KANUNGO T, MOUNT D M, NETANYAHU N, *et al*. An efficient K-means clustering algorithm: analysis and implementation [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2002, 24(7): 881-892.

[14] 向友君,谢胜利. 图像检索技术综述 [J]. 重庆邮电学院学报:自然科学版, 2006, 18(3): 348-354.


(上接第 477页) 学习理论实现填充阈值的自适应调整,这将是本文后续工作中需要进一步深入研究的课题。

**参考文献**:


[1] VINCENT L, SOILLE P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1991, 13(6): 583-598.

[2] WANG De-min. Unsupervised video segmentation based on watersheds and temporal tracking[J]. IEEE Trans on Circuits and Systems for Video Technology, 1998, 8(5): 539-546.

[3] 卢官明. 一种计算图像形态梯度的多尺度算法 [J]. 中国图象图形学报, 2001, 6(3): 214-218.

[4] 王小鹏,郝重阳,樊养余. 基于形态学尺度空间和梯度修正的分水岭分割 [J]. 电子与信息学报, 2006, 28(3): 485-489.

[5] 罗希平,田捷,诸葛婴. 图像分割方法综述 [J]. 模式识别与人工智能, 1999, 12(3): 300-312.

[6] CANNY J. A computational approach to edge detection[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1986, 8(6): 678-698.

[7] NAJWA V S. On detecting edges[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1986, 8(6): 699-714.

[8] 林开颜,吴军辉,徐立鸿. 彩色图像分割方法综述 [J]. 中国图象图形学报, 2005, 10(1): 1-10.

[9] TANCHAROEN D, JITAPUNKUL S. Spatial segmentation based on modified morphological tools[J]. IEEE Conference Information Technology, 2001, 9(1): 478-482.