

基于改进活动轮廓的视频对象 自动分割及跟踪算法

张晓燕¹⁾ 赵荣椿¹⁾ 马志强²⁾

¹⁾(西北工业大学计算机学院计算机信息与工程系, 西安 710072) ²⁾(空军工程大学电讯工程学院网络工程系, 西安 710077)

摘 要 为了能够从视频序列图像中准确地提取出运动视频对象, 提出了一种基于改进活动轮廓的视频对象自动分割及跟踪方法。该方法首先采用连续帧间差的 4 次统计量假设检验来确定视频对象的运动区域, 并使用形态滤波消除残余噪声和空洞; 然后根据 3 帧序列图像得到的前后运动区域的相与运算来有效地解决运动视频对象前后帧的遮挡问题, 以获得视频对象模板, 当提取出视频对象模板的边缘轮廓后, 再用梯度向量流场作为外力的改进活动轮廓算法来获得视频对象的精确轮廓; 最后以此视频对象的轮廓为基础进行运动补偿, 以得到下一帧图像的初始曲线, 再使用改进的活动轮廓算法对下一帧图像进行分割, 即可实现视频对象的跟踪。该方法不仅能够消除差分图像中的显露背景, 得到运动视频对象精确的轮廓, 并且可进行多目标的分割与跟踪。

关键词 视频对象 活动轮廓 假设检验 形态滤波 运动补偿

中图分类号: TP391.41 **文献标识码**: A **文章编号**: 1006-8961(2007)03-0438-06

Automatic Video Object Segmentation and Tracking Using An Improved Active Contour

ZHANG Xiao-yan¹⁾, ZHAO Rong-chun¹⁾, MA Zhi-qiang²⁾

¹⁾(Department of Computer Information and Engineering, Computer School, Northwestern Polytechnical University, Xi'an 710072)

²⁾(Department of Network Engineering, Telecommunication Engineering Institute, Air Force Engineering University, Xi'an 710077)

Abstract In order to extract the moving video object accurately, a novel algorithm for automatic segmentation and tracking moving object is proposed in this paper. Firstly, a fourth-order statistic hypothesis testing in inter-frame difference is used to automatically separate the moving areas from the background in a general video sequence. Then the morphological filter is used to remove the noises and holes. Based on the intersection operation of the two moving areas which obtained from three consecutive frames, the covered/uncovered background may be detected and moving object model can be obtained. Secondly, after determining the initial contour of the moving object, an improved active contour which uses the gradient vector as the external force guides the initial contour moving to the video object contour. Lastly, the motion compensation is made to the video object contour to obtain the initial contour of the next frame and the improved active contour is used again to correct the video object contour of the next frame. Thus, video object tracking can be realized. Our method can not only obtain the accurate contour of the video object but also make the segmentation and tracking for multiple objects.

Keywords video object, active contour, hypothesis testing, morphological filter, motion compensation

1 引 言

众所周知, 视频对象分割是视频国际标准

MPEG-4 及 MPEG-7 的首要步骤和前提。在 MPEG-4 中, 人们通过对分割出的视频对象进行编码来有效地提高视频压缩编码的效率; 在 MPEG-7 中, 则要求视频数据必须以视频对象的形式存在。此外, 由于

基金项目: 国家自然科学基金项目(60141002)

收稿日期: 2005-12-14; 改回日期: 2006-03-22

第一作者简介: 张晓燕(1970 ~), 女。2000 年获西安理工大学硕士学位, 现为西北工业大学计算机学院计算机信息与工程系博士研究生。主要研究方向为序列图像处理、模式识别和计算机视觉。E-mail: zxyxjwxxj@163.com

视频对象分割也是由视频处理到视频分析的关键环节,其在模式识别、计算机视觉等领域也得到了广泛应用,因此,对视频对象进行自动分割及跟踪是非常重要的。但是,由于视频对象具有语义意义,且在提取时发生的任何错误都可能导致视频对象的无效,所以精确分割视频对象,并跟踪其变化已成为一个亟待解决的问题。

视频对象的分割一般从帧内和帧间两个方面来考虑。基于帧内信息分割方法^[1,2]的主要思想是首先根据灰度或色彩的相似性,将视频的每一帧划分成若干区域,然后估计各区域的运动参数,并按运动相似性对区域进行合并,以实现视频对象分割,但这种方法运算量巨大,分割速度慢,而且当视频对象的各部分存在不一致运动时,无法得到完整的视频对象。虽然采用帧间差分图像^[3,4]来检测运动目标是基于帧间信息分割的最直接方法,但由于帧间差分获得的变化检测模板(change detection mask, CDM),通常只能检测到含有遮挡区域的运动区域,而并不是整个运动对象,而且帧间差分会减少空间边缘信息,因此很难得到准确的视频对象轮廓。

本文提出了一种基于改进活动轮廓的视频对象自动分割及跟踪方法。该方法能够消除帧间差分图像中的显露背景,以获得运动视频对象模板。采用梯度向量流场作为外力的改进活动轮廓来对视频对象进行分割,不仅有效地克服了传统梯度势能场的弱点,而且能得到视频对象精准的外轮廓。本文在进行视频对象的跟踪中是采用简单易行、效率较高的新三步搜索算法^[5](new three-step search, NTSS)进行运动估计,其可以准确地捕捉较细小的运动。实验表明,本文方法不仅能够准确地进行多目标的分割与跟踪,而且获得的视频对象的闭合轮廓曲线还可以进行编码传输和基于形状的检索。

2 获取初始运动视频对象及轮廓

2.1 自动分离运动区域与背景

在假设摄像机静止的前提下,若连续帧灰度差为非零值,则认为帧间灰度变化值是由噪声和运动对象的变化引起的。但由于简单地通过对差分图像进行阈值化来获取运动区域的方法过于依赖指定的阈值,因此不具有通用性。文献[6]指出,视频序列的噪声主要包括随机噪声、室内亮度变化、室外背景纹理(如草地、海浪、云等)的慢变化等,它们的统计量一般符合高斯

特性,而运动对象则具有很强的结构性。所以,为了提高算法的自适应性,使得阈值的设定能够自动化,可采用将帧间差分图像的局部4阶矩的平均值与根据其背景区域估计的高斯噪声的方差进行比较的方法来确定运动对象的位置,以自动分离运动与背景区域。

设帧间差分图像像素值为 $d(x, y)$,定义一个 3×3 的矩形窗口,使其在差分图像上逐点移动,若 $m_{i,j}^{(4)}$ 和 $\mu_{i,j}$ 分别代表以第 i 行第 j 列像素点为中心窗口像素的四阶矩和均值,则

$$m_{i,j}^{(4)} = \frac{1}{N} \sum_{(x,y) \in W_{i,j}} (d(x,y) - \mu_{i,j})^4 \quad (1)$$

$$\mu_{i,j} = \frac{1}{N} \sum_{(x,y) \in W_{i,j}} d(x,y) \quad (2)$$

其中, $W_{i,j}$ 表示以第 i 行第 j 列像素点为中心的大小为 $N = 3 \times 3$ 的窗口, (x, y) 代表窗口内像素的坐标。先通过窗口移动逐像素计算 $m_{i,j}^{(4)}$,再进行如下判断就可得到运动区域。

若 $m_{i,j}^{(4)} > c(\sigma_s^2)^2$,则为 H_1 ,

若 $m_{i,j}^{(4)} \leq c(\sigma_s^2)^2$,则为 H_0 。

其中, H_0 表示像素是静止的, H_1 表示像素是运动的。 c 是一个常量,可根据视频序列设置。而在静止背景区域 S 内的噪声均方差 σ_s^2 则按下式计算:

$$\sigma_s^2 = \frac{1}{N_s} \sum_{(x,y) \in S} (d(x,y) - \mu_s)^2 \quad (3)$$

静止区域 S 为属于背景的一块区域,一般在边缘地带选取, N_s 是区域 S 的大小, μ_s 是区域 S 内差分图像的平均值。具体实现时,取差分图像的4个角区域,并对获得的4个区域的噪声方差取中值,再将其作为背景噪声方差,即可取得较好的效果。

2.2 获取视频对象模板及轮廓

使用上述的互帧差4次统计量判决方法虽然可分离出运动的区域与背景,但提取的运动区域模糊、不完整,而且还存在一些孤立噪声点。为了消除这些干扰,可选择一个合适的结构元素进行形态学处理,即先消除运动区域外的一些孤立点,之后进行开启运算,去除一些细小的突出,最后进行闭合运算,通过填补一些小洞和缝隙和融合边界上一些较窄的缺口来得到完整的运动区域。由于运动区域包含了运动对象区域以及遮挡区域,所以本文利用3帧图像来提取中间帧运动视频对象,即先将中间帧与前后帧分别进行互帧差4次统计量假设检验,再对经过形态学滤波后的运动变化区域进行相与运算,即可去除掉遮挡区域和得到运动视频对象的初始分割

模板。这种利用 3 帧图像来提取中间帧运动视频对象的方法不一定必须是连续的 3 帧,可根据视频序列情况适当加大帧之间对象的位移。然后通过对比

动视频对象的模板进行形态学运算去除掉内部像素,即可得到运动视频对象的初始轮廓。图 1 显示了获取的初始运动视频对象及轮廓。



图 1 Akiyo 第 3 帧的原图、初始运动视频对象及轮廓

Fig. 1 Source image (Akiyo frame3), initial moving video object and its contour

3 基于改进活动轮廓的分割

3.1 活动轮廓的基本概念

活动轮廓是定义在图像范围内的一条曲线或一个表面,它能够在曲线或表面本身的内力影响和由图像产生的外力影响共同作用下发生变形。变形的最终目的是使曲线或平面本身与目标物体的边缘或者是图像中期望检测到的特征的形状相一致。

第 1 个活动轮廓模型^[7]是由 Kass 在 1988 年提出的,其能量表达式为

$$E = \int_0^1 \{ E_{\text{int}}(X(s)) + E_{\text{ext}}(X(s)) \} ds \quad (4)$$

其中, s 表示弧长的参数,活动轮廓在图像中的位置可通过参数向量 $X(s) = [x(s), y(s)]$ 表示, E_{int} 是 Snake 的内部能量函数,它用于保持 Snake 的连续性和平滑性,其定义为

$$E_{\text{int}} = \frac{\alpha |X'(s)|^2 + \beta |X''(s)|^2}{2}, s \in [0, 1] \quad (5)$$

式中, $X'(s)$ 表示参数向量 $X(s)$ 关于 s 的一阶微分, $X''(s)$ 表示参数向量 $X(s)$ 关于 s 的二阶微分。 α 为弹力系数,若系数 α 的值越大,则使得曲线在力的作用方向上收缩得越快。 β 为强度系数,如果 β 值很大,则曲线就会变得很僵硬而不容易发生弯曲,而如果 β 值较小,则就会容许曲线弯曲成一个角。

E_{ext} 表示由于图像的性质而产生的外部能量,它能够使曲线向着目标移动。

根据式(4)和式(5),可以写出以下完整的能量表达式

$$E = \int_0^1 \left\{ \frac{1}{2} (\alpha |X'(s)|^2 + \beta |X''(s)|^2) + E_{\text{ext}}(X(s)) \right\} ds \quad (6)$$

当曲线到达目标物体(通常是物体的边缘线)的时候,能量函数 E 也应达到最小值。若要使得可变轮廓线能量 E 最小,则式(6)应当满足欧拉方程,于是有

$$\alpha X''(s) - \beta X''''(s) - \nabla E_{\text{ext}} = 0 \quad (7)$$

式中的 $X''''(s)$ 表示参数向量 $X(s)$ 关于 s 的四阶微分。 ∇E_{ext} 表示外部能量关于 s 的微分。

这个方程可以看作是一个作用力平衡方程

$$F_{\text{int}} + F_{\text{ext}} = 0 \quad (8)$$

其中, $F_{\text{int}} = \alpha X''(s) - \beta X''''(s)$, $F_{\text{ext}} = -\nabla E_{\text{ext}}$, 内在力 F_{int} 用于阻止曲线的伸展和弯曲,外在力 F_{ext} 用于驱使可变轮廓向着期望的图像边缘移动。

为了对式(7)进行动态求解,可以把 X 看作是时间 t 以及弧长 s 的函数。于是, X 关于 t 的偏微分等于式(7)的左端,即

$$X_t(s, t) = \alpha X''(s, t) - \beta X''''(s, t) - \nabla E_{\text{ext}} \quad (9)$$

当方程的解 $X(s, t)$ 趋于平稳时, $X_t(s, t)$ 趋于零,此时就可以得到式(7)的解。

3.2 改进的活动轮廓

在传统的活动轮廓中,外力通常由边缘点的梯度势能场给出。这不仅导致力量场的有效捕捉范围很小,而且使得缓变区域不能够获得外力。若采用梯度向量流场 V 来替代式(9)中的势能场函数 $-\nabla E_{\text{ext}}$, 则可得

$$X_t(s, t) = \alpha X''(s, t) - \beta X''''(s, t) + V \quad (10)$$

一般将通过解上述动态方程而得到的参数曲线

称为梯度向量流 (gradient vector flow, GVF) 可变形轮廓^[8]。

梯度向量流场的向量 $V(x, y) = (u(x, y), v(x, y))$ 被定义为能够使得下式的能量最小的向量:

$$\varepsilon = \iint k(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |V - \nabla f|^2 dx dy \quad (11)$$

其中, k 为参数, 它取决于图像中的噪声大小, 噪声越大, 其值应越小。 u_x, u_y, v_x, v_y 分别为梯度向量流场的向量分量 u, v 在 x, y 方向上的偏导数。 ∇f 为图像 f 的梯度。式(11)右端第 1 项为平滑项, 它用于使得向量场缓慢变化; 第 2 项为数据项, 它用于使得向量场 V 尽量地接近图像梯度 ∇f 。要使得整体能量最小, 应取 $V = \nabla f$ 。这样就满足梯度向量流场的设计目标: 当向量场的值较大时, 向量 V 可尽量保持与边界图的梯度相一致, 而在图像的缓变区, 向量 V 的变化则较平缓。

向量 V 的两个分量 u, v 可以利用变分法, 通过解以下的 Euler 方程得到:

$$\begin{aligned} k \nabla^2 u - (u - f_x)(f_x^2 + f_y^2) &= 0 \\ k \nabla^2 v - (v - f_y)(f_x^2 + f_y^2) &= 0 \end{aligned} \quad (12)$$

其中, ∇^2 为 Laplace 算子。 f_x, f_y 分别是图像 f 在 x, y 方向的梯度。式(12)对于梯度向量流场提供了更加直观的理解, 在均匀区域中, 因为图像 f 的梯度 ∇f 为零, 所以每个方程的第 2 项都是零, 在这样的区域内, 由于 u, v 由 Laplace 方程所决定, 而梯度向量流场是从该区域的边界插值获得, 其在一定程度上反映了边界向量的竞争, 因此梯度向量流场可以生成指向凹边界的向量。

3.3 基于改进活动轮廓的分割

分割时, 首先对需分割的图像使用文献[9]的方法进行边缘检测, 该算法适合于检测强边缘点, 并可去除一些弱边缘短线, 这正好符合视频对象边缘轮廓线的特点; 然后使用式(2)计算出边缘图像的 GVF 场; 最后以图 1(c) 所示的轮廓曲线为初始轮廓, 通过迭代收缩来获得运动视频对象精确的轮廓曲线。图 2 显示了 Akiyo 第 3 帧的边缘图像、轮廓收缩过程以及分割结果。



图 2 Akiyo 第 3 帧分割结果

Fig. 2 Akiyo (frame 3) segmentation result

4 运动视频对象的跟踪

4.1 运动估计

运动估计是为了得到相邻帧间的运动向量, 而且通过运动估计, 不仅可以有效地去除冗余, 保留帧间的有效信息, 还可根据运动估计结果对运动对象轮廓进行补偿。本文选用新三步搜索法进行运动估计。对于通常的序列图像而言, 由于所含的运动大多幅度很小, 其块运动场经常是平稳、光滑和变化缓慢的, 所以全局最小值一般是基于中心分布的, 而不是均匀分布的。这恰好符合了新三步搜索法关于运动物体在帧间的运动较小的假设。此外, 该块匹配

算法还具有简单易行、效率较高的特点。

图 3 是 NTSS 算法的原理图, 图中数字表示搜索顺序, 用黑圈圈出的数字表示搜索到的最小块误差 (block distortion measure, BDM) 值的点。

NTSS 算法对于三步搜索算法 (three-step search, TSS) 的改进之处在于对外围大模板进行搜索时, 同时对内侧的小模板进行搜索。外围大模板由原点周围步长为 4 的 8 个点组成, 内侧小模板由原点以及原点周围步长为 1 的 9 个点组成。NTSS 算法在第 1 步就同时搜索大小模板, 通过进行块匹配计算并比较, 以找到最小 BDM 值的点。如果最小 BDM 值的点是原点, 则停止搜索。如果最小的 BDM 点是原点外的小模板上的点, 则以最小 BDM

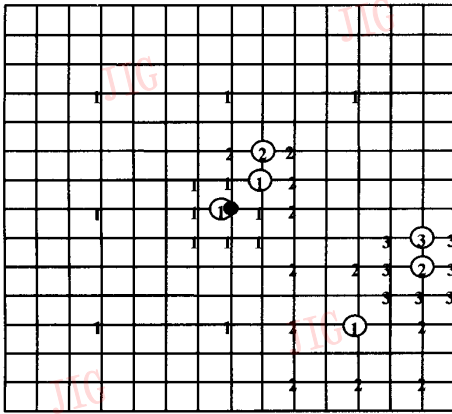


图 3 新的三步搜索算法示意图
Fig. 3 NTSS algorithm

值的点为中心计算其 8 邻域点,在找出最小 BDM 值的点后就停止搜索。如果最小 BDM 值的点为大模板中的点,则后面的步骤同 TSS 算法一样,即步长减半搜索其 8 邻域点,直到步长为 1 时,即找到最小

BDM 值的点。

NTSS 算法最多要测试 33 个点,但是由于它采用了半途终止的策略,因此常常在第 1 步或是第 2 步就终止,即对于静止的块只要测试 17 个点,对于亚静止的块(运动范围在 ± 2 pixels 内)只要测试 20 或 22 个点。所以说 NTSS 算法对于估计静止及亚静止的块的运动向量是非常有效的。NTSS 算法相对于 TSS 算法的加速因子为 18%。

4.2 运动视频对象轮廓的跟踪

本文先利用运动估计获得的相继帧间的运动向量来对分割后得到的准确的视频对象轮廓进行运动补偿,以得到下一帧的跟踪轮廓曲线;然后用数学形态学方法进行膨胀、细化等处理后得到下一帧闭合的初始轮廓曲线,再使用 3.3 节给出的基于改进活动轮廓的分割方法对下一帧进行分割,即可得到精确的边缘轮廓,从而实现了运动视频对象的跟踪。图 4 显示了运动补偿得到的 Akyio 第 4 帧的初始轮廓、收缩过程以及跟踪结果。



图 4 Akyio 第 4 帧的跟踪结果
Fig. 4 Tracking result of Akiyo(frame4)

5 实验结果

为验证本文算法的效果,用本文方法对 Trevor

和 Table tennis 标准序列图像进行了检测,试验结果如图 5 和图 6 所示。Trevor 序列图像是背景较复杂的头肩序列,从图 5 可以看出,初始分割结果较准确,且只需要很少的迭代收缩就可得到准确轮廓,但

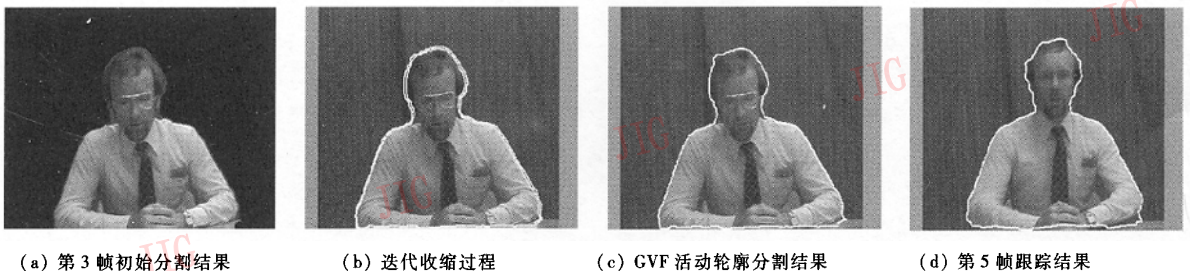


图 5 Trevor 序列图像分割与跟踪结果
Fig. 5 Trevor segmentation and tracking result

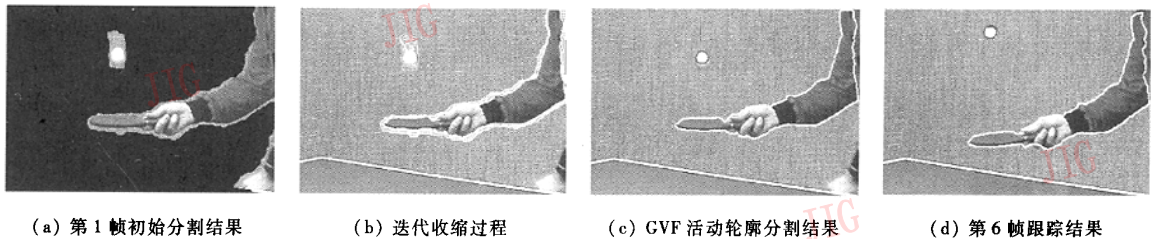


图6 Table tennis 序列图像分割与跟踪结果

Fig.6 Table tennis segmentation and tracking result

由于左手肘部的边缘受到桌子边缘的影响,所以稍有偏差。Table tennis 图像序列虽然获得的初始视频对象轮廓较大,不准确(如图6所示),但是经过GVF活动轮廓迭代收缩后仍然得到了准确的分割和跟踪结果。从图5和图6可看出,本文算法对单个视频对象和多个视频对象都能进行很好的分割,而且经过运动补偿后也获得了很好的跟踪结果。

6 结论

本文提出了一种基于改进活动轮廓的视频对象自动分割及跟踪算法。该算法首先依据视频序列图像帧间差4次统计量的假设检验,自动地将运动区域与背景分离,并经过形态学处理去除掉噪声和空洞,同时利用3帧序列图像得到的前后运动区域进行相与运算来提取出视频对象的模板及轮廓;然后用基于改进活动轮廓的分割算法来得到视频对象的精确轮廓;最后以分割出的视频对象的轮廓为基础,通过使用NTSS算法估计出的运动向量进行运动补偿来得到视频对象在下一帧的初始曲线,再使用改进的活动轮廓算法进行分割,以实现视频对象的跟踪。由于本文方法使用了梯度向量流场为外力的活动轮廓,这不但加大了力量场的捕捉范围,并且能够收缩于边缘的凹边界,因此得到了较准确的视频对象轮廓曲线。另外,由于对视频对象初始轮廓曲线的确定比较准确,所以改进的活动轮廓分割算法只需很少次数的迭代收缩就可得到准确的轮廓曲线,从而提高了分割效率。对于含有多个视频对象的序列图像,通过对运动区域相与后获得的对象模板进行聚类分析来确定出视频对象的个数以及各个视频对象的初始轮廓曲线,就能够实现多个对象的分割与跟踪。本文实验结果令人基本满意,但是对于运动较大的对象如何获得更好的跟踪轮廓以及如何减

少背景边缘在迭代收缩过程中的影响则是需要进一步研究的问题。

参考文献 (References)

- 1 Demin W. Unsupervised video segmentation based on watersheds and temporal tracking[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1998, 8(5): 539~542.
- 2 Liu Zhi, Yang Jie, Peng Ning-song. Video object segmentation based on hypothesis testing and region merging [J]. Journal of Data Acquisition and Processing, 2004, 19(2): 124~129. [刘志,杨杰,彭宁嵩.基于假设检验和区域合并的视频对象分割[J].数据采集与处理,2004,19(2):124~129.]
- 3 Kim C, Hwang J N. Fast and automatic video object segmentation and tracking for content-based application[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(2): 122~129.
- 4 Yang Li, Zhang Hong, Li Yu-shan. Automatic segmentation of moving objects in video sequences[J]. Journal of Computer Aided Design and Computer Graphics, 2004, 16(3): 301~306. [杨莉,张弘,李玉山.视频运动对象的自动分割[J].计算机辅助设计与图形学学报,2004,16(3):301~306.]
- 5 Li R X, Zeng B, Liou M L. A new three-step search algorithm for block motion estimation [J]. IEEE Transactions on Circuits and Systems for Video Technology, 1994, 4(4): 438~442.
- 6 Neri A, Colonese S, Talone P. Automatic moving object and background separation [J]. Signal Processing, 1998, 66(2): 219~232.
- 7 Kass M, Witkin A, Terzopoulos D. Snakes: active contour models [J]. International Journal Computer Vision, 1988, 1(4): 321~331.
- 8 Xu C Y, Prince J L. Snake, shapes and gradient vector flow [J]. IEEE Transactions on Image Processing, 1998, 7(3): 359~369.
- 9 Jiang Xiao-yue, Zhao Rong-chun. Application of the B-SPLINE wavelet in edge detection [J]. Chinese Journal of Stereology and Image Analysis, 2002, 7(4): 198~201. [蒋晓悦,赵荣椿. B-样条子波在图象边缘检测中的应用[J].中国电视学与图象分析,2002,7(4):198~201.]