

# 一种鲁棒的视频分割算法

李凌娟 贾振堂 贺贵明

(武汉大学软件工程国家重点实验室, 武汉 430072)

**摘要** 无论是在图象识别,还是在基于 MPEG-4 的图象压缩编码等应用领域,视频对象分割提取是其中一个很重要的技术环节。为了在静止背景的情况下,能很好地解决多目标分割的问题,同时能进行单目标的分割,提出了一种鲁棒性较好的视频分割算法。该算法通过对图象序列中每连续 3 帧图象进行对称差分,首先检测出目标的运动范围;然后通过对差分结果进行聚类分析来确定该帧图象中视频对象的个数;接着再利用在二值差分图象上收缩的活动轮廓,把视频对象的轮廓精确地包围起来,即得到该帧分割结果;最后利用光流法来对视频对象进行投注跟踪、修正。另外还利用多个图象序列对该方法进行了试验。实验结果表明,在静止背景下,该算法无论是对运动的单目标,还是对运动的多目标,均能较好地从静止背景中分离出来,即能得到理想的分割结果,故具有一定的鲁棒性和实用性。

**关键词** 视频分割 对称差分 聚类 活动轮廓 光流场

**中图法分类号:** TP391.41 TN919.81 **文献标识码:** A **文章编号:** 1006-8961(2002)11-1128-06

## A Robust Video Segmentation Algorithm

LI Ling-juan, JIA Zhen-tang, HE Gui-ming

(State Key Laboratory of Software Engineering, Wuhan University, Wuhan 430072)

**Abstract** Video segmentation is an important step in image recognition and in video coding based on MPEG-4 standard. This paper proposes a robust video segmentation algorithm which can greatly segment not only multiple VOs but also single VO from the static background. From the research, it is found that the raw contours of the active objects in the sequential video frames can be gained fast and exactly by using symmetrical DFD. So first use every three continuous video frames to get symmetrical DFD, then acquire the motion area of the moving object. Second analyze the result of the symmetrical DFD by clustering to inspect the number of VOs in the video frame. During clustering analyzing, not employing the traditional fuzzy-C clustering mode, but employing a new clustering mode with using the body's characteristic which is proposed firstly in this paper. Third use the active contour algorithm in the binary image to get the contours that enclose the VOs accurately. By this step acquire the result of segmentation. Last, track the VOs in the video frames by the vectors of optical-flow and also modify the result in some special conditions. All tests receive the ideal segmentation effects by using this algorithm with several image sequences. The experimental results show that this algorithm can complete video segmentation both for multiple VOs and for single VO in the static background correctly, so this algorithm has some robusticity and practicality.

**Keywords** Video segmentation, Symmetrical DFD, Clustering, Active-contour, Optical-flow

## 0 引言

1997年,ISO制定MPEG-4、MPEG-7标准框

架以后,由于面向对象、基于内容的概念被引入到多媒体领域,因此,基于对象、面向内容的视频压缩和视频检索成为目前该领域研究的热点,而且视频对象的分割提取算法作为这些研究中最基础、最关键

基金项目:武汉大学软件工程国家重点实验室建设基金项目

收稿日期:2001-11-09;改回日期:2002-01-07

的部分,也成为目前的讨论焦点。

目前经过研究所提出的自动分割算法有许多种,而通常是利用视频图象在时间和空间轴上的信息来进行分割。其中一类为利用相邻两帧间的时域差分来得到目标的运动信息,再用统计学方法来消除掉噪音的干扰<sup>[1,2]</sup>;另一大类是基于视频图象的运动估计和分割,即首先把图象中存在的运动估计出来,再根据运动一致性准则,把序列图象分割成不同的区域<sup>[3]</sup>。另外,还有其他的一些视频分割提取技术,如遗传算法、神经网络等方法。

本文在视频处理的研究过程中,提出了一种较为有效的视频分割算法,该算法能较好地由静止背景中分割出运动的单目标或多目标。该算法的基本思路是,首先通过连续相邻 3 帧图象的对称差分,快速地检测出一个或多个运动的视频对象轮廓;然后通过对称差分结果二值图象中的所有像素块进行聚类分析来确定该帧图象中视频对象的个数;再通过对二值差分图象上活动轮廓进行收缩,把视频对象的轮廓精确地包围起来,最后利用光流法来对视频对象进行投注跟踪、修正。

## 1 目标检测

众所周知,在对相邻两帧源图象进行差分,利用图象帧之间的显著差异,能快速地检测出目标的运动范围,而连续 3 帧序列图象通过对称差分相与运算,则能较好地检测出中间帧运动目标的形状轮廓<sup>[4]</sup>。由于这种方法,不但速度快,而且在视频对象连续运动时,其得到的目标形状轮廓相当精确,所以本文使用这种方法将运动的视频对象从静止的背景中分离出来,即表现为对称差分二值图象上的“白色”像素块群。在此基础上,再通过对这些像素块进行“聚类”分析来确定图象帧中的视频对象个数。

### 1.1 运动范围检测

检测时,首先分别计算视频序列连续 3 帧源图象  $f_{(t-1)}(x, y)$ ,  $f_t(x, y)$  和  $f_{(t+1)}(x, y)$  的相邻两帧源图象的绝对差灰度图象

$$d_{t-1,t}(x, y) = |f_t(x, y) - f_{(t-1)}(x, y)|$$

$$d_{t,t+1}(x, y) = |f_{(t+1)}(x, y) - f_t(x, y)|$$

然后可分别使用基于图象距离度量的阈值选取方法来选取适当阈值<sup>[5]</sup>,同时对以上  $d_{t-1,t}(x, y)$ 、 $d_{t,t+1}(x, y)$  两个绝对差灰度图象进行二值化。这样,就得到两个绝对差二值图象  $b_{t-1,t}(x, y)$  和  $b_{t,t+1}(x, y)$ ,再在其每一个像素位置进行相“与”运算,即得到对称差分结果二值图象,而该图中所获得结果即为检测到的视频对象的形状轮廓。在这一阶段中,二值化阈值的确定很重要,因为阈值过大,使目标像素块太零碎,将加大后面的分割难度,而阈值过小,又容易引入噪声块的影响。

### 1.2 目标个数确定

对于经过对称差分得到的二值图象,由于计算机还不能自动数出所检测出来的视频对象个数,因此,还要应用聚类的思想来对得到的视频对象像素块进行分析,以确定图象帧中视频对象的个数。

在聚类前,首先在对称差分结果图象中,对每个像素按照八邻域进行一次扫描,确定每个连通像素块的最小包围矩形框,以用于后面的聚类分析。在聚类过程中,本文抛弃了传统的、经常被广泛采用的模糊 C 聚类方法,而是依据人体宽高特性,采用了一种较为直观的聚类分析方法。该方法就是将视频像素块聚类在最小包围矩形框之间,若在水平方向上互相交叠或距离相差在 3 个像素之内,而且在垂直方向上互相交叠或距离相差在图中最大矩形框长度的 1/5 的距离内,则认为框内的像素块同属一个目标,就可合并为一个新的大矩形框。如此在图中对各矩形框进行循环比较合并,直到图中再没有矩形框



图 1 两种方法聚类结果比较图

合并为止,最后得到的矩形框个数及范围,即认为是所确定的视频对象个数及所在范围.试验结果证明,这种矩形框的聚类分析方法,对于确定视频对象的个数相当有效,而使用传统的模糊C聚类方法来对二值差分图象进行分析时,则首先需要使用密度函数方法来确定初始的聚类中心<sup>[6]</sup>,而密度函数分析的范围为以样本集合最中间的样本到距离最远的样本之间的距离的二分之一为半径的圆<sup>[7]</sup>,这里的样本为二值图象中的视频对象像素块,而距离的远近则是通过像素块重心之间的距离来求得,因为没有考虑到像素块之间相邻的物理意义,而只是单纯考虑到像素块重心之间距离远近,所以对于人体这种垂直长度远大于水平宽度的目标,其错误划分视频对象个数的概率比较大,效果不是很理想.

## 2 目标轮廓界定

对于经过对称差分得到的二值图象,其中视频对象轮廓并不封闭、连续,为了得到对象的精确封闭边缘,本文利用活动轮廓的基本原理,设计了一种二值图象的收缩型活动轮廓,用来从残缺不全的二值差分图象中获得视频对象形状的矢量描述,在进行目标界定时,首先以每个视频对象从聚类过程中所获得的外接矩形框作为活动轮廓的初始轮廓,再通过顶点增删、曲线平滑滤波、主动收缩(粘定图象)等局部化计算步骤来完成每次迭代,以得到目标的正确边缘.

### 2.1 活动轮廓的一般模型

通常,活动轮廓是一条闭合的曲线.在确定活动轮廓时,首先给定初始轮廓,由于初始轮廓受内力和外力的共同作用而变化,其中外力推动活动轮廓向物体的边缘靠近,而内力则保持轮廓的光滑性和拓扑性,因此当活动轮廓达到平衡位置时,对应能量最小,活动轮廓就收敛到目标物体的边缘<sup>[8]</sup>.

由于力的作用过程也可用能量的概念来表述,因此,活动轮廓的收敛过程就转化为“能量极小化”的问题,这是优化理论中的常见问题,它可以用迭代的方式来实现.其能量函数通常定义为:

$$E = \int_0^1 [E_{\text{contour}}(v(s)) + E_{\text{image}}(v(s)) + E_{\text{control}}(v(s))] ds$$

其中, $E_{\text{contour}}$ 表示曲线的内部能量,是由模型拉伸、弯曲而产生的内能而提供平滑性约束; $E_{\text{image}}$ 是由图象

本身特征所决定的能量,用于将活动轮廓吸引到图象某特征点处(如边缘),一般定义为图象 $I(x, y)$ 的负梯度; $E_{\text{control}}$ 是由用户根据应用需要而自行定义的外部能量,它的作用是帮助活动轮廓朝着物体边缘移动.

### 2.2 二值图象下的收缩活动轮廓模型及其局部化计算

对于二值图象,梯度的模只有两个值:要么为0,要么为1,而且图象中大部分区域是“梯度盲区”(梯度模值为0),在这种情况下,因为轮廓的推进过程无法依靠梯度的渐变来进行,所以必须给轮廓的演变指定方向,要么收缩,要么膨胀.根据视频分割的实际情况,本文选用收缩方式.处理时,与每个视频对象对应的活动轮廓,首先以所在的最小包围矩形框作为初始轮廓,然后作适当的收缩,即可获得视频对象的精确轮廓.在目标轮廓确定时,活动轮廓的迭代变化过程,由以下几个步骤组成,它们将被循环执行(在具体编程实现时,有些步骤是合在一起、通过一次扫描完成的).

#### 2.2.1 控制点重新分配和平滑

当直接对对称差分二值图象作处理时,某些活动轮廓点在收缩的过程中,如果在它的 $(3 \times 3)$ 邻域小窗内遇到图象中属于视频对象的白色像素点,则该轮廓点就被固定,且以后便不再移动,而另外的轮廓点则继续移动,由于活动轮廓点移动的情况不尽相同,所以可能会造成相邻轮廓点之间的距离增大或减小,如果相邻轮廓点距离过大,则在活动轮廓包围时,容易遗漏图象点;如果距离过小,则这些相邻轮廓点可简化为一个点,这样就便于节省运算时间.因为活动轮廓点之间的距离在收缩过程中会发生变化,所以在轮廓演变的过程中,必须对活动轮廓点间距离 $d$ 进行检查,如果距离过大( $d > 1.5H$ ),则插入一个(或多个)点,如果过小( $d < 0.5H$ ),则删除一个(或多个)点.这里, $H$ 是标准采样距离,可以人为指定,或根据目标的大小和目标边缘的曲率来确定.

对于重新分配过后的点,还要进行平滑.这里平滑的目的就是要使轮廓有一定的硬度,起到抗弯曲的作用.如果不经过平滑处理,而直接进行收缩变换,则会出现扭曲现象.本文采用高斯滤波的方法来进行处理,以滤除掉轮廓中的高频成分.该方法就是以某一轮廓点为中心,首先从该点出发,分别沿顺时针和逆时针方向各取两个顺序挨着的点,再按照下面公式进行加权平均计算,以便达到平滑效果:

$$v_i^{(k+1)} = \sum_{j=-2}^2 \omega_j v_{i+j}^{(k)}, \sum_{j=-2}^2 \omega_j = 1$$

其中,  $v_i$  表示活动轮廓上的第  $i$  个控制点,  $k$  表示迭代次数,  $\omega = [\omega_{-2}, \omega_{-1}, \omega_0, \omega_1, \omega_2]$  为窗口函数,  $\omega_j$  的不同分布情况相当于不同的轮廓硬度, 类似高斯分布, 比如,  $\omega = [0.05, 0.2, 0.5, 0.2, 0.05]$ , 这样, 不仅可使轮廓有一定硬度, 而且还可以维持活动轮廓的重心在迭代后不变。

### 2.2.2 收缩

收缩定义为: 使每个顶点按“内法向” $n$  所确定的方向移动一定步长  $T_{move}$ . 这样可保持形状的相似收缩。

$$v_i = v_i + T_{move}n$$

其中, 法向量  $n$  可通过相邻的前后点来近似确定. 假设以逆时针方向作为轮廓运动的方向, 设函数  $u(\cdot)$  表示取单位向量, 则“内法向” $n$  表示为:  $n = u(n_0 + n_1)$ , 其中,  $n_0 = Au(v_i - v_{i-1})$ ,  $n_1 = Au(v_{i+1} - v_i)$ , 其中

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \text{ 是一个旋转矩阵.}$$

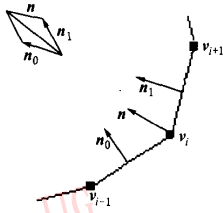


图 2 收缩方向——内法向

### 2.2.3 收敛标准

对每个轮廓, 可以通过考察如下两个指标来作为活动轮廓最后收敛的判据: ①长度, 看活动轮廓顶点数有无变化, 如果有变化, 则继续迭代, 否则考察第 2 项; ②曲线距离, 即本次活动轮廓到上一次活动轮廓之间距离的总和, 如果该项小于某一阈值, 则结束迭代。

每个视频对象的活动轮廓收敛后, 再把最后得到的各个视频对象的活动轮廓顶点相连, 就得到了各个视频对象的精确封闭边缘. 当把这些封闭边缘包围的图象内容从源图象帧中提取出来, 即得到了该帧的视频对象分割结果。

## 3 目标投注

在多目标分割算法中, 为了保持视频对象平面

的连续性, 还应该对每个视频对象进行运动跟踪分析, 以确定这一帧视频对象与下一帧视频对象之间的对应关系. 在此, 本文使用了文献[9]、[10]的方法, 即通过光流场来计算图中每个视频对象活动轮廓点的点矢量, 再通过光流矢量, 把这一图象帧中的视频对象轮廓投注到下一帧, 以达到对视频对象进行跟踪的目的。

当每分割完一帧后, 都把该帧每个视频对象的活动轮廓点保留, 而且保留这一帧到下一帧的光流矢量, 以便于目标投注跟踪. 这样当下一帧目标检测结束, 即获得了下一帧的视频对象个数及其所在最小矩形框, 然后就可通过对光流矢量进行偏移, 把已分割帧的每个视频对象的活动轮廓点投注到下一图象帧的位置上, 即得到各个视频对象的预测活动轮廓. 如果这个视频对象的预测活动轮廓的重心在其中某个最小矩形框内, 则认为这个视频对象与该最小矩形框内的视频对象相对应. 在目标投注过程中, 一般情况下, 每一帧的视频对象之间都能顺利对应, 但是也会出现以下特殊情况:

(1) 该帧的某个最小矩形框, 若无上一帧视频对象的预测活动轮廓投注过来, 则认为是出现了新的视频对象, 此时需增加一个新的视频对象标志号(图 3(a)所示);

(2) 该帧的某个最小矩形框, 若有多于一个上一帧视频对象的预测活动轮廓投注过来, 则认为是原来的某几个视频对象相叠, 此时应将原来的几个视频对象标志号删除, 并产生一个新的视频对象标志号(图 3(b)所示);

(3) 若上一帧视频对象的预测活动轮廓虽投注过来, 但没有投注在任何一个最小矩形框内, 且又投注在图象帧边界外, 则认为是视频对象正常消失, 此时应删除该视频对象标志号(图 3(c)所示)。

(4) 上一帧视频对象的预测活动轮廓虽投注过来, 但没有投注在任何一个最小矩形框内, 且仍投注在图象帧边界内, 则认为是出现了对称差分误操作, 此时应将该视频对象预测活动轮廓作为该视频对象在这一帧的活动轮廓, 并提取出视频对象内容, 以便对该帧的分割结果起到修正作用. 这种情况下, 该视频对象标志号保持不变. 由于对称差分误操作是指其中某一帧上该视频对象运动变化很小, 致使差分运算时, 无法检测出该目标, 从而造成视频对象的“假消失”(图 3(d)所示);

(5) 若上一帧视频对象的预测活动轮廓投注过

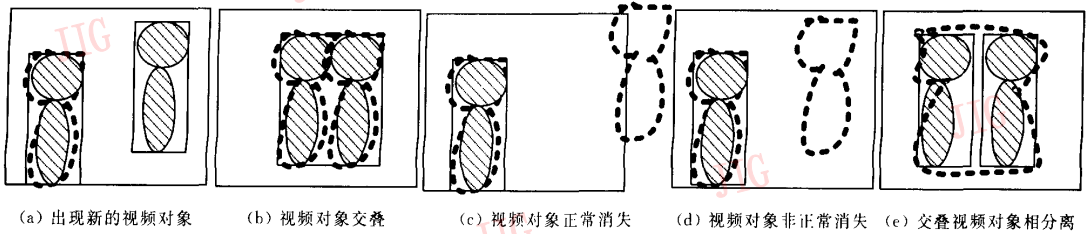


图3 目标投注特殊情况示意图

(黑色虚线表示从上一帧投注过来的目标轮廓, 黑线矩形框表示为本帧包围目标的最小矩形框, 阴影线部分为本帧经过活动轮廓收缩所得到的目标轮廓所包围的范围)

来, 且其中某一个预测活动轮廓又投注在多于一个小矩形框内, 即认为是交叠在一起的视频对象分离, 此时应删除该预测活动轮廓的视频对象标志号, 而增加分离后的多个视频对象的标志号(图3(e)所示).

### 4 试验结果

取 192×144 格式的视频序列图象, 在 CPU 主频 866 的微机进行了试验. 图 4、图 5 为用该算法对多目标、单目标视频序列进行视频对象提取的演示图, 实验的效果如图 4、图 5 所示.

从图中可以看出, 图 4(c) 有 8 个, 图 5(c) 有 2 个独立的连通像素块.

图 4(d) 为用各像素块最小包围矩形框聚类合并得到的结果, 其最后获得的运动视频对象数目为

2, 从源图象中可以看出, 在大的矩形框内, 左边的人与右边两个紧挨的人还有一定距离, 但由于背景中人的影子的影响, 这 3 个人还是聚合为一个视频对象. 图中的灰白色曲线为从上一帧投注过来的最小矩形框内的两个预测活动轮廓曲线, 这表明上一帧的两个视频对象在这一帧相交叠, 表现为一个视频对象.

图 5(d) 中显示了用各像素块最小包围矩形框聚类得到的结果, 其合并得到运动视频对象数目为 1.

图 4(e)、图 5(e) 为在对称差分二值图象上, 由活动轮廓收缩所得到的视频对象的轮廓, 由该图可以看到, 收缩后结果并没有歪曲视频对象本身的轮廓形状.

图 4(f)、图 5(f) 为经过以上处理以后, 从背景中分离出来的多个或单个视频对象.

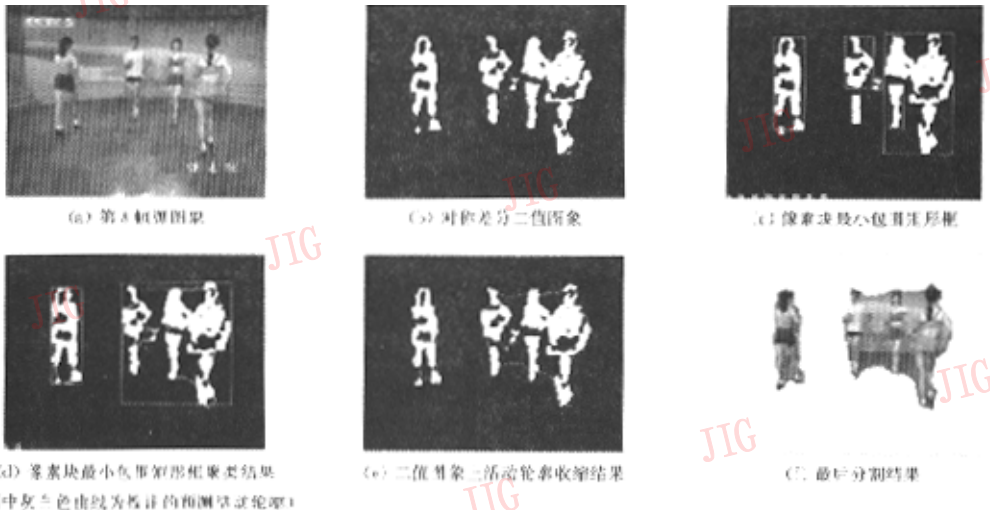


图4 多目标分割示意图

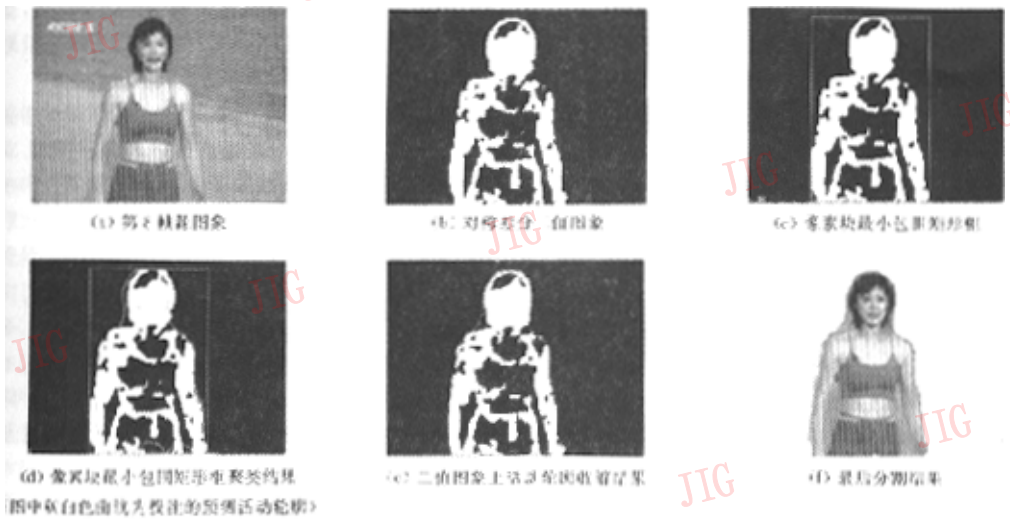


图 5 单目标分割示意图

### 5 结 论

从上面的实验结果可以看出,本文提出的这种分割算法,不但能较好地分割多目标,而且也能用于分割单目标,另外,即使在背景复杂,但在随时间变化不大的情况下,效果仍然比较好。从总体性能上来说,该算法鲁棒性好,能用于模式识别、面向内容的图象检索、基于对象的压缩编码等领域,具有一定实用性。但是,由于对称差分本身所具有的缺陷性,即当目标部分运动不明显时,其所出现的局部缺损现象,还将会在某些时候为分割结果带来一些瑕疵,为此,下一步的工作将是研究如何更好地解决对称差分本身所具有的缺陷,以及研究如何加快光流场、活动轮廓算法的计算时间,并进一步提高这些算法的计算精度,以便使该算法能满足各应用领域的更高要求。

### 参 考 文 献

- 1 Colonese S, Russo G. Automatic moving objects and background segmentation by means of higher order statistics [A]. In: IS&T Electronic Imaging'97, Conference: Visual Communication and Image Processing[C], 1997, 3024:8~14.
- 2 Jae Cark Choi, Munchurl Kim, Myoung Ho lee et al. Automatic segmentation based on spatio-temporal information[S]. MPEG 97/1743, 1997.
- 3 Ayer S, Bigun L. Segmentation of moving objects by robust motion parameter estimation over multiple frames [A]. In: Proceedings of the 4<sup>th</sup> Int. Workshop on Time-Varying Image Processing and Moving Object Recognition[C], 1993:295~302.
- 4 王积分,张新荣. 计算机图象识别[M]. 西安:中国铁道出版社,

- 1988:208~233.
- 5 付忠良. 基于图象差度量度的阈值选取方法[J]. 计算机研究与发展,2001,38(5):563~567.
- 6 李金宗. 模式识别导论[M]. 北京:高等教育出版社,1994:294~352.
- 7 裴继红,谢维信. 聚类的密度函数方法[J]. 西安电子科技大学学报,1997,24(4):463~467.
- 8 夏利民,谷士文. 基于形变模型的图象分割和非刚性目标的跟踪[J]. 模式识别与人工智能,2000,13(1):64~68.
- 9 罗新根,吴立明. 一种改进的光流场计算方法[J]. 电子学报,2000,28(1):130~131.
- 10 李介谷. 计算机视觉的理论和实践[M]. 上海:上海交通大学出版社,1991:40~47.



**李凌娟** 1976年生,武汉大学计算机学院在读硕士。研究方向为多媒体通信,现主要研究 H.263、MPEG-4 视频压缩及传输技术。



**贾振堂** 1969年生,河南城建高等专科学校计算机工程系讲师,现为武汉大学软件工程重点实验室博士研究生。研究方向为视频处理及多媒体网络通信。



**贺贵明** 1946年生,1981年获武汉大学水利电力大学计算机应用专业工学硕士学位,现为武汉大学软件工程重点实验室教授、博士生导师。研究方向为多媒体网络通信。