

交互式分割视频运动对象的研究与实现

韩军 熊璋 侯亚荣 蒲菊华

(北京航空航天大学计算机学系, 北京 100083)

摘要 在 MPEG-4 视频编码标准中, 为了实现基于视频内容的交互功能, 视频序列的每一帧由视频对象面来表示, 而生成视频对象面, 需要对视频序列中运动对象进行有效分割, 并跟踪运动对象随时间的变化. 在视频分割方法中, 交互式分割视频对象能满足分割的效率与质量指标要求, 因此提出了一种交互式分割与自动跟踪相结合的方式分割视频语义对象. 即在初始分割时, 依据用户的交互与形态学的水分数线分割算法相结合提取视频对象轮廓, 并用改进的轮廓跟踪方法有效提高视频对象轮廓的精度; 对后续帧的跟踪, 采用六参数仿射变换跟踪运动对象轮廓的变化, 用平移估算的运动矢量作为初始值, 计算六参数仿射变换的参数. 实验结果表明, 该方法能有效地分割并跟踪视频运动对象.

关键词 计算机图像处理(520·6040) 视频分割 视频对象 自动对象跟踪 交互

中图分类号: TP391 **文献标识码**: A **文章编号**: 1006-8961(2003)02-0169-07

The Research and Implementation of Semantic Video Objects by User Interaction and Automatic Objects Tracking

HAN Jun, XIONG Zhang, HOU Ya-rong, PU Ju-hua

(Beijing University of Aeronautics and Astronautics, Dept. of Computer Science and Engineering, Beijing 100083)

Abstract In order to support content-based functions in the visual standard of MPEG-4, each frame of video sequences should be represented in terms of video object planes (VOP's). Segmentation and tracking of video moving objects are used for guiding the extraction of video object plane from video sequences. Interactive video objects segmentation is a good trade-off between efficiency and robustness. The proposed method comprises user interaction and automatic objects tracking of semantic video objects. First, the semantic video objects boundary can be found by using a combination of human assistance and a morphological segmentation tool. Accurate of video object boundary is effectively increased by updating contour tracking method. Second, the semantic video objects in the remaining frames are obtained using affine motion estimation and compensation of the previous semantic video objects. six-parameter motion vector is estimated on the basis of estimation of two-parameter translational motion vector. Experimental results demonstrate that this method is effective to segment and track video semantic objects.

Keywords Computer image processing, Video segmentation, Video objects, Automatic objects tracking, Interaction

0 引言

MPEG-4 和 MPEG-7 标准的制定, 能实现基于视频内容的存取、粘贴、查询、检索等功能, 使基于内容的交互式多媒体应用成为可能. 以前的视频压缩标

准 MPEG-1, MPEG-2 是以一帧中的像素点为处理基准的; 而在 MPEG-4 视频压缩标准中, 视频内容由视频对象来表示, 一帧画面由不同的视频对象所组成, 并定义了景物描述的二进制格式(BIFS)来组织视频对象, 即 MPEG-4 的景物描述用树形的层次化结构来表示, 树的每个结点是一个视频对象, 使用适

基金项目: 国防科工委“九五”项目(4.3.3.1); 国家教委跨世纪人才计划; 航空基金(00F51057)

收稿日期: 2001-12-05; 改回日期: 2002-05-27

当的对象描述结点可构造复杂的景物. 因此 MPEG-4 景物描述实现了将不同的视频对象按一定时空关系组合在一起, MPEG-4 编码器是对不同的视频对象单独编码, 进而实现对视频内容的随机存取.

视频分割的目的就是在视频序列中, 分割出在语义上有意义的视频对象, 视频分割要求精确提取运动对象的轮廓, 并跟踪其轮廓的变化. 将这些分割出的视频对象作为 MPEG-4 视频编码器的数据源, 能有效提高视频压缩编码的效率和实现基于视频内容的随机存取.

评价视频分割的性能主要有两个指标^[1]: 分割及跟踪算法的自动化程度和分割视频对象轮廓的精度. 由于视频序列的数据量很大, 完全采用人工方式分割出每帧中的视频对象, 其工作效率甚低, 因此希望能自动分割并跟踪视频序列中的运动对象, 以提高分割算法的效率. 自动分割视频运动对象的方法是基于视频序列中的时空一致特性, 可分为时序分割, 空间分割或组合时空分割. 时序分割确定在视频序列中运动对象的位置; 空间分割确定组成视频对象区域的精确轮廓; 时空分割融合时序分割和空间分割的结果, 进而产生 MPEG-4 所需要的视频对象面 (VOP). 在自动分割算法中, 需要对分割的区域进行有效的分裂与合并处理, 自动分割算法的关键是如何有效确定视频内容中的运动一致性.

自动分割算法只能适合于一些简单的景物及特定的视频序列, 自动分割算法虽然有可能实时分割视频序列, 但分割结果不能达到满意的效果, 主要有两个原因: 一是视频对象很难由低级特性 (如: 颜色, 直方图及运动特性) 来描述, 自动分割算法依据的是视频内容的时空一致性, 不能真正理解视频序列中语义意义上的视频内容; 二是, 现有大多数自动分割算法获取运动对象的轮廓是通过区域的分裂与合并获得的, 这样处理的结果使得运动对象轮廓会多出或丢失一些小的区域, 人眼的视觉对视频对象的轮廓极其敏感, 所以通过自动分割获得的运动对象轮廓精度难于满足 MPEG-4 的交互式应用的要求.

通过上述分析, 为了在分布式环境下, 实现 MPEG-4 交互应用, 保证分割视频对象轮廓精度是极其重要的, 在保证分割视频对象轮廓精度的情况下, 应尽量提高分割算法的自动化程度. 因此, 采用适当的人机交互能保证视频分割的质量, 提高分割算法的灵活性. 人机交互的作用有两点: 使计算机真正理解语义意义的视频内容; 使计算机能知道视频

对象轮廓的位置.

本文采用交互分割与自动跟踪相结合的方法来实现对视频语义对象的分割, 图 1 为交互分割与自动跟踪视频对象的原理框图. 初始的交互分割也称为 I 帧分割^[2], 对后续帧的自动跟踪也称为帧间 P 帧分割. 在初始分割时, 由用户的交互, 确定视频语义对象的大致轮廓, 进而采用形态学的分水线分割算法, 在特定的区域内分割视频对象的轮廓, 对于轮廓周围的一些毛刺, 采用改进的轮廓跟踪算法进行平滑, 提取出精确的视频对象轮廓, 由分割出的对象轮廓生成二进制掩模, 最终获得视频对象面 (VOP); 对后续帧的跟踪, 通过对仿射运动的六参数进行估算和补偿, 确定出运动一致的区域, 来实现对运动对象的自动跟踪.

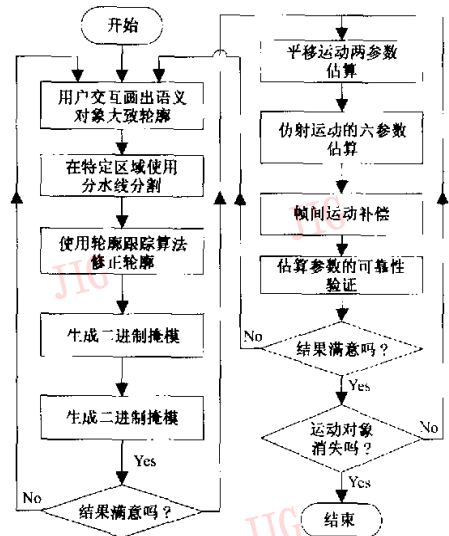


图 1 交互分割与自动跟踪视频对象原理图

1 交互分割视频对象

无论采用何种自动或半自动视频分割算法, 对视频图象的空间分割都起着关键的作用, 空间分割的精度强烈地影响着最终分割结果的质量. 一般认为, 视频图象是随亮度变化的彩色图象, 视频分割的实质是提取视频序列中运动对象的轮廓, 而不关心运动对象内部纹理的变化. 提取运动对象的轮廓产生对象的二值掩模, 即将对象区域标注为“1”, 而将图象的其他区域标注为“0”, 再将二值掩模图象与原始视频图象进行与运算, 就能提取视频语义对象.

在提取运动对象轮廓时,视频分割的精度受视频对象内部及图象背景的纹理、颜色、亮度变化的影响,视频图象亮度的变化,会使背景的局部区域与运动对象的局部区域融为一个区域,使分割的视频对象轮廓出现不连续;而视频对象内部纹理的不同,在分割时,会产生许多小的区域,同时图象背景也会分割出许多小的区域,在进行分裂合并时,会影响提取的运动对象轮廓的精度。

在视频图象空间分割算法中,采用数学形态学的分水线算法^[5]能有效、快速地提取运动对象的轮廓.分水线算法由标记提取和判决两部分组成,其中标志提取就是将视频图象进行化简,提取一致性的区域,将每个一致性区域的中心点标注为种子点,再在原视频图象上,分割出运动对象的轮廓;而分水线算法的判决过程是为了确定分割区域的边界点,这是一个区域增长过程.其判决准则有两种:一种是采用图象的梯度算子;而另一种是基于彩色图象的多值均值算子。

梯度算子的分水线算法会产生许多有噪声或量化错误的局部小块,而且还会使分割过细,而基于彩色图象的多值分水线算法不同于基于梯度的分水线算法,它不强调空间的一致性,而是依据空间颜色的相似性.实现这个算法需要分两步计算:首先,需要评估多值表示的标志区域,然后,计算相邻像素点与多值表示值之间的差.实际上,标志的表示被计算成平均彩色值,距离函数定义为绝对距离

$$d_i = |r - r_i| + |g - g_i| + |b - b_i| \quad (1)$$

其中, r_i, g_i, b_i 为一相邻像素点的红、绿、蓝分量, r, g, b 是一区域内所有像素红、绿、蓝分量的平均值。

在自动分水线算法的基础上,提出交互式提取运动对象的轮廓算法,使分割算法不受运动对象内部纹理和亮度变化的影响.其先采用图形工具在感兴趣的运动对象周围画出运动对象的大致轮廓,再比较运动对象的纹理与背景纹理的复杂度,有时图象背景的纹理及颜色一致性较好,能很容易分割图象背景和提取运动对象的轮廓.判决一个区域的一致性,可采用一个简单的方法^[4]:如果仅在区域 R 内最大颜色值与最小颜色值之差小于一个门限 T

$$(\max_{p \in R} (I_p) - \min_{p \in R} (I_p)) \leq T, \text{ 则一致性较好} \quad (2)$$

上述的交互视频分割算法归结如下:

(1) 使用图形工具,用鼠标在视频图象上画出语义视频对象的大致轮廓;

(2) 跟踪上述画出的大致轮廓,形成连续的、闭合的轮廓跟踪点;

(3) 以每个轮廓点为中心,在 10×10 区域内填充为“1”,扩展轮廓点,使所画的轮廓线变粗;然后,再判决像素点“1”的八邻域点的和是否等于8,形成所画轮廓线的内边缘和外边缘;

(4) 经上述处理后,在运动对象的轮廓周围形成了3条曲线,即鼠标画的线及内、外边缘线,这时,分析所画轮廓线与内边缘线之间及所画轮廓线与外边缘线之间这两个区域的颜色一致性情况,以确定在采用分水线图象分割算法时,是以内边缘作为种子点,还是以外边缘作为种子点;

(5) 采用多值彩色分水线算法,仅对内边缘与外边缘之间的区域判决边缘点,提取视频对象的轮廓;

(6) 上述提取的轮廓线不平滑,有一些毛刺,需采用轮廓跟踪算法,形成平滑的、精确的语义对象轮廓;

传统的二值图象闭合边界的轮廓跟踪算法是:首先按从上到下、从左到右的顺序搜索,找到的第一个为“1”的点,一定是最左上方的边界点,记为 A ,它的右、右下、下、左下4个邻点中至少有一个是边界点,记为 B .从 B 开始找起,按右、右上、上、左上、左、左下、下、右下的顺序找相邻点中的边界点 C .如果 C 就是 A 点,则表明已经转了一圈,程序结束.否则从 C 点继续找,直到找到 A 为止.这种算法是最简单的轮廓跟踪算法,如轮廓存在断点或局部回路,将无法跟踪,不能形成闭合的、完整的轮廓.将这种轮廓跟踪算法进行改进,使其能跟踪断点或存在局部回路的情况,改进的算法是判断轮廓的局部极值点,进行有向跟踪,以避免局部回路;如轮廓线存在断点,继续向前搜索,并插补存在的断点.具体按如下设计的结构进行搜索:

从左到右按如下结构搜索:

```
int ForwardDirect[12][2]=
  {{1,0},{1,1},{0,1},{-1,1},
   {2,0},{2,1},{2,2},{1,2},{0,2},{-1,2},
   {-2,2},{-2,1}}
```

从右到左按如下结构搜索:

```
int BackwardDirect[12][2]=
  {{-1,0},{-1,-1},{0,-1},{1,-1},
   {-2,0},{-2,-1},{-2,-2},{-1,-2},
   {0,-2},{1,-2},{2,-2},{2,-1}}
```

2 自动跟踪视频对象

以交互方式实现视频对象分割后,产生视频对

象的二值掩模、对后续帧的分割是自动跟踪对象区域的变化,当自动跟踪、提取运动对象的轮廓精度不可接受时,用户能暂停自动跟踪处理,以交互分割处理的方法改正或再定义这一帧中对象的轮廓.然后,再进行自动跟踪处理,用户中断分割处理次数越少,这种跟踪算法越好.自动跟踪的依据是视频对象运动的一致性.

由于镜头的移动或对象本身在运动,所以需考虑用运动参数模型的估计与补偿来跟踪视频运动对象的变化,并实现从当前帧视频运动对象区域到下一帧的映射.由于两参数运动矢量估算仅考虑了平移运动情况,因此对视频运动对象的旋转、伸缩、变形等情况,要使用六参数的仿射变换.由于仿射运动的六参数解具有不唯一性,其解的可靠性需要验证,为了可靠地估计仿射运动模型的运动参数,采用了层次化的迭代算法,即从低次的运动模型到高次的运动模型,使用的低次运动模型是两参数的平移运动模型,而六参数的仿射运动模型用于高次运动估算,所以,自动跟踪运动对象的算法由如下 4 步组成:

(1) 两参数平移运动估计

对视频运动对象进行跟踪,首先使用两参数平移运动模型

$$p(x, y, t) = a, q(x, y, t) = d \quad (3)$$

其中, $(p(x, y, t), Q(x, y, t))$ 是在点 (x, y) 处估算的 $t, t+1$ 帧之间对象运动的位移,假定连续帧之间图象亮度是不变的,那么通过使用泰勒展开式,能得到等式

$$pf_x + qf_y + f_t = 0 \quad (4)$$

其中, f_x, f_y, f_t 是亮度 $I_{x,y,t}$ 对应于 x, y, t 的局部偏导数.

在分割的视频运动对象掩模 R 区域上,估计平移运动参数 $v = (a, d)^T$ 就是对光流误差函数求偏导,令偏导等于零,使光流误差最小,得到两参数平移运动矩阵

$$\begin{pmatrix} \sum f_x^2 & \sum x f_x^2 & \sum y f_x^2 & \sum f_x f_y & \sum x f_x f_y & \sum y f_x f_y \\ \sum x f_x^2 & \sum x^2 f_x^2 & \sum x y f_x^2 & \sum x f_x f_y & \sum x^2 f_x f_y & \sum x y f_x f_y \\ \sum y f_x^2 & \sum x y f_x^2 & \sum y^2 f_x^2 & \sum y f_x f_y & \sum x y f_x f_y & \sum y^2 f_x f_y \\ \sum f_x f_y & \sum x f_x f_y & \sum y f_x f_y & \sum f_y^2 & \sum x f_y^2 & \sum y f_y^2 \\ \sum x f_x f_y & \sum x^2 f_x f_y & \sum x y f_x f_y & \sum x f_y^2 & \sum x^2 f_y^2 & \sum x y f_y^2 \\ \sum y f_x f_y & \sum x y f_x f_y & \sum y^2 f_x f_y & \sum y f_y^2 & \sum x y f_y^2 & \sum y^2 f_y^2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \end{pmatrix} = - \begin{pmatrix} \sum f_x f_t \\ \sum x f_x f_t \\ \sum y f_x f_t \\ \sum f_y f_t \\ \sum x f_y f_t \\ \sum y f_y f_t \end{pmatrix} \quad (9)$$

$$\begin{pmatrix} \sum f_x^2 & \sum f_x f_y \\ \sum f_x f_y & \sum f_y^2 \end{pmatrix} \begin{pmatrix} a \\ d \end{pmatrix} = - \begin{pmatrix} \sum f_x f_t \\ \sum f_y f_t \end{pmatrix} \quad (5)$$

设 2×2 的矩阵为 A_i , 等式右边为 b_i , 则 $A_i v_i = b_i$, 因此, 平移运动矢量为

$$v_i = A_i^{-1} b_i = [v_i^x, v_i^y]^T \quad (6)$$

使用平移运动的两参数作为仿射运动估算的初始点,即将获得的平移运动参数作为估计六参数仿射运动模型参数的初始值 a 和 b .

(2) 修改运动对象掩模

从当前运动对象的轮廓得到运动对象的区域,在运动对象区域内,实现连续两帧 F_{i-1}, F_i 之间的两参数平移运动补偿,删除运动补偿差大于给定门限的像素点来改进预测的运动对象区域,再用预测的运动对象区域来计算仿射运动参数.

(3) 六参数仿射运动估算

六参数仿射运动方程为

$$\begin{aligned} p(x, y, t) &= a + bx + cy \\ q(x, y, t) &= d + ex + fy \end{aligned} \quad (7)$$

在运动对象区域上,用均方预测误差最小的方法来估计仿射运动的六参数,均方预测误差函数 $E(R, t)$ 为

$$E(R, t) = \sum_{(x,y) \in R} DFD^2(x, y, t) \quad (8)$$

其中, DFD 是连续帧之间的位移差: $DFD(x, y, t) = I(x, y, t) - I(x-p, y-q, t-1)$, 使用高斯-牛顿算法^[5], 得到六参数矩阵(式(9)).

设 6×6 的矩阵为 A , 等式右边为 b , 则 $Av = b$, 因此, 仿射运动矢量为 $v = A^{-1} b$.

(4) 补偿六参数矢量

用估算的仿射六参数矢量(式(9))映射对象区域,以作为相应下一帧的视频对象.计算仿射运动的六参数矢量是一个迭代的过程,其估算过程如下:

- ① $v^{(i)} \leftarrow [v_i^x, 0, 0, v_i^y, 0, 0]^T, i \leftarrow 0;$
- ② 使用运动参数 $v^{(i)}$, 以当前帧 F_{i-1} 修正预测帧 F_i ;

③ 计算六参数矩阵 A, b , 在帧 F_i 与 F_{i+1} 之间, 估算六参数向量 $\Delta v = A^{-1}b$;

④ 迭代计算 $v^{(i+1)} = v^{(i)} + \Delta v, i \leftarrow i + 1$;

⑤ 对于 $i = 1, \dots, 6$, 如果 $\|\Delta v_i\| < T_1$, 则停止迭代, 否则转入第②步, 继续计算;

门限 T_1 选为 10^{-3} , 如对象运动较慢, 则需要迭代的次数较少, 小于 10 步迭代; 如对象运动较大, 则需要迭代的次数多些, 10~20 步之间. 当跟踪的运动对象突然消失, 或发生大的变形, 使用六参数仿射运动模型不能正确跟踪对象的运动时, 则需要使用交互空间分割, 重新确定新的运动对象.

3 实验结果

采用标准测试序列 Forman、Mother-daughter、Coastguard 来进行交互式提取运动对象的实验.

图 2 是采用交互式分割的只有头肩活动的视频

对象, 图 2(a)是在原 Forman 测试序列第 30 帧上, 用鼠标在感兴趣语义对象上画出的大致轮廓, 图 2(b)是在所画轮廓周围, 使用多值分水线算法所分割的对象轮廓, 再进行轮廓平滑处理, 所形成的结果, 图 2(c)是由图 2(b)轮廓生成区域的二值掩模, 所提取的语义对象. 图 3 和图 4 是采用仿射参数模型自动实现对分割的运动对象的跟踪, 图 3 是原测试序列 Forman 的第 30、35、40、45 帧图象, 图 4(a)是上述交互式分割的 Forman 第 30 帧运动对象, 图 4(b)、图 4(c)、图 4(d)是通过六参数仿射运动估算, 自动实现对运动对象的跟踪.

图 5, 图 6 采用 Mother daughter 测试序列交互式分割所需的视频对象, 在这个测试序列中 Mother 的运动比 Daughter 大. 在自动分割算法中, 采用将这两个对象作为一个对象进行分割, 如采用交互式分割所需的对象, 图 5 就是初始交互分割出的 Mother 对象, 在图 7 中显示了只跟踪 Mother 对象的变化.



(a) 用鼠标画出的轮廓



(b) 提取的对象轮廓



(c) 交互式提取的对象

图 2 采用交互式分割的视频对象(Forman 序列)



(a) 第 30 帧



(b) 第 35 帧



(c) 第 40 帧



(d) 第 45 帧

图 3 原测试序列 Forman 图象



(a) 交互式分割的对象(第 30 帧)



(b) 自动跟踪运动对象(第 35 帧)



(c) 自动跟踪运动对象(第 40 帧)

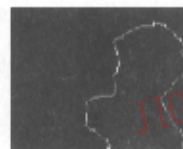


(d) 自动跟踪运动对象(第 45 帧)

图 4 采用仿射参数模型跟踪分割的运动对象(Forman 序列)



(a) 用鼠标画的轮廓



(b) 提取对象的轮廓



(c) 交互式提取的对象

图 5 交互分割出的 Mother 对象

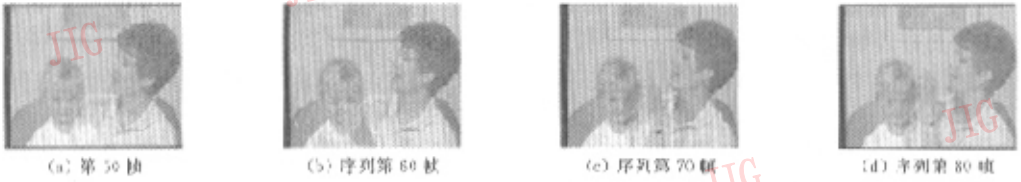


图 6 原测试序列 Mother-daughter 图象



图 7 自动跟踪的 Mother-daughter 对象(只显示了 Mother 对象)

图 8,图 9,图 10 是采用 Coastguard 测试序列交互式分割从左到右行驶的一条船,在这个测试序列中,背景比较复杂,要考虑室外背景的亮度与纹理的变化,如采用交互式分割可有效降低背景纹理对分割

的影响,图 8 就是在初始交互分割出“船”对象,图 10 为只跟踪“船”对象的运动.采用鼠标交互的方式分割视频语义对象,分割质量较好,但不具有实时性,而且对象轮廓需要采用适当的平滑算法.

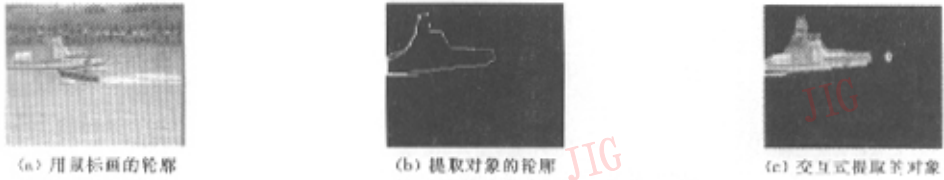


图 8 Coastguard 测试序列用交互式分割的一条船

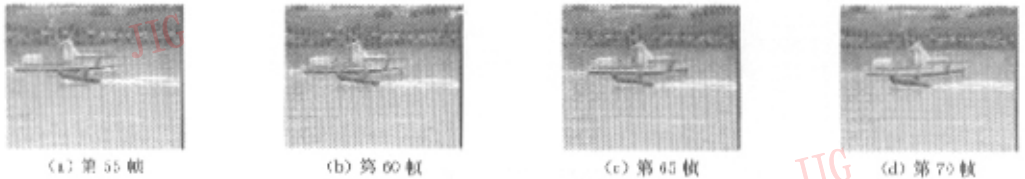


图 9 原测试序列 Coastguard 图象

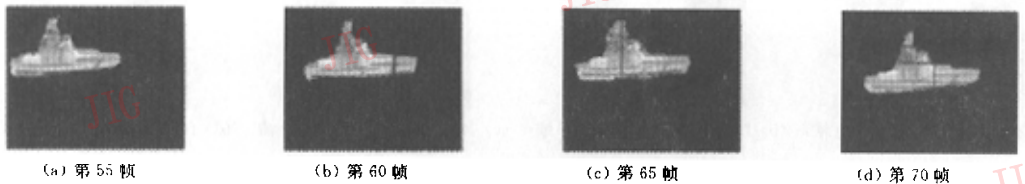


图 10 自动跟踪的船对象

4 结 论

为实现 MPEG-4 在分布环境下基于内容的应用,采用了交互式视频语义对象分割方法,取得较好

的分割质量,如对分割的结果不满意,还能进行灵活地调整.这种方法的分割结果能应用于 MPEG 4 的内容制作、MPEG-4 对象编码器的数据源,还能应用于将来 MPEG-7 基于内容的查询和检索数据库中.

由于本文采用的交互式分割方法是在语义对象

轮廓的周围形成内外轮廓,在内外轮廓构成的区域内,使用分水线算法进行空间分割,进而提取语义对象的精确轮廓,但由内外轮廓所形成的区域内,颜色、亮度的变化还会影响分割结果,使提取的轮廓有一些毛刺;运动对象的跟踪采用六参数仿射运动模型,只适合刚体运动的情况,对非刚体还需要进一步修整对象轮廓,下一步的工作是将视频分割的自动方法与交互式分割方法相结合,进一步完善视频分割算法的效率,以获得更加满意的分割质量。

参 考 文 献

- 1 Marcotegui B, Zanoguera F, Correia P *et al*. A video object generation tool allowing friendly user interaction[OL]. <http://intl.ieeexplore.ieee.org/xplore/conferences.jsp>.
- 2 Munchurl Kim, Joon J G, Kwak J S *et al*. Moving object segmentation in video sequences by user interaction and automatic object tracking[J]. *Image and Vision Computing*, 2001,19(2):245~260.
- 3 Luc Vincent, Pierre Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991,13(6):583~598.
- 4 Chuang Gu, Ming-Chieh Lee. Semiautomatic segmentation and tracking of semantic video objects[J]. *IEEE Transactions on Circuits and Systems For Video Technology* 1998, 8(5):572~584.
- 5 Jean-Luc Dugelay, Henri Sanson. Differential methods for the identification of 2D and 3D motion models in image sequences [J]. *Signal Processing: Image Communication* 1995,7(1):105~127.



韩 军 1965年生,高级工程师,1990年获西安电子科技大学通信与电子系统硕士学位,现为北京航空航天大学博士研究生,曾获航空部科技进步一等奖一项,目前主要研究方向为视频处理和视频传输。



熊 璋 1956年生,教授,博士生导师,曾获国家科技进步一等奖,航空部科技进步一等奖各一项,近年发表了学术论文40余篇,目前主要研究方向为多媒体技术和分布式信息处理。



侯亚荣 1974年生,现为北京航空航天大学计算机科学与工程系博士研究生,主要研究方向为分布式多媒体数据库系统、模式识别。



蒲菊华 1976年生,现为北京航空航天大学计算机科学与工程系博士研究生,主要研究领域为视频压缩、视频信息的自适应传输。