基于运动窗生成的时空视频分割

朱 辉[1,2], 李在铭[1], 蔡 毅[2]
(1. 电子科技大学通信与信息工程学院, 四川成都 610054; 2. 四川移动通信有限责任公司, 四川成都 610081)

摘 要: 本文提出了一种基于运动窗生成的时空视频分割方法. 首先通过检测运动变化区域而生成运动窗, 然后只对视频图像在运动窗内的部分进行时空视频分割, 从而大大减少了运算量, 提高了运算速度. 此外, 在空间分割中, 本文提出了一种符合人眼视觉特性的逐级划分的空间分割方法; 最后根据运动相似性将区域合并, 完成视频分割. 实验结果表明, 本文的方法运算速度快并能有效地进行空间分割, 最后能取得较好的视频分割结果.

关键词: 运动窗; 运动变化区域; 视频分割; 时空特性
中图分类号: TN919. 8 文献标识码: A 文章编号: 0372-2112 (2004) 03-0480-05

Spatio-Temporal Video Segmentation Based on the Generation of Motion Window

ZHU Hui[1,2], LI Zai-ming[1], CAI Yi[2]
(1. School of Communication and Information Engineering, UEST of China, Chengdu, Sichuan 610054, China;
2. Sichuan Mobile Communication Company Limited, Chengdu, Sichuan 610081, China)

Abstract: This paper presents a spatio-temporal video segmentation algorithm based on the generation of motion window. Motion window is generated through detecting motion-changed region, then video segmentation is proceeded in the interior of motion window and computation can be greatly reduced. In addition, in spatial segmentation, a hierarchical partition algorithm is presented which is accordant with visual properties. Finally, regions are merged based on the motion similarity. Experimental results demonstrate the performance of our algorithm.

Key words: motion window; motion-changed region; video segmentation; spatio-temporal information

1 引言

新一代视频编码标准 MPEG-4 采用基于对象的视频编码, 基于对象的视频编码要求首先把视频图像划分成不同的对象或把运动对象从背景中分离出来, 然后对不同的视频对象采用不同的编码方法. 基于对象的视频编码不仅能大大提高压缩比, 而且允许用户对多媒体数据按内容进行交互式操作. 视频对象的分割是基于对象的编码的基础, 因此视频对象的分割技术具有重要的研究价值和应用意义.

近年来, 国内外学者对视频对象的分割进行了许多研究, 其中主要采用的是基于视频的时空特性来分割视频对象, 例如文献[1~6]. 这种方法的基本思想是, 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, 从而实现视频分割. 现有的基于时空特性的视频分割方法都是在整幅视频图像上进行的, 因而其运算量巨大, 运算速度慢. 考虑到视频分割主要指的是视频中运动对象的分割, 而运动对象存在于运动变化区域内, 因此如果我们能首先检测出运动变化区域, 然后只对运动变化区域内的部分进行视频分割, 则能大大减少运算量, 提高运算速度.

本文提出了一种基于运动窗生成的时空视频分割算法, 算法的流程图如图 1 所示: 首先通过检测运动变化区域而生成运动窗, 然后只对视频图像在运动窗内的部分按时空特性进行分割, 这样使运算量大大减少, 提高了运算速度. 另外, 在空间分割中, 本文提出了一种符合人眼视觉特性的逐级划分的空间分割方法, 实验结果表明这是一种有效的空间分割方法. 最后根据运动相似性将空间分割后形成的区域合并, 完成视频分割. 全文安排如下: 第二节是运动窗生成, 第三节是空间分割, 第四节是运动估计与区域合并, 第五节是实验结果与分析, 第六节是全文总结.

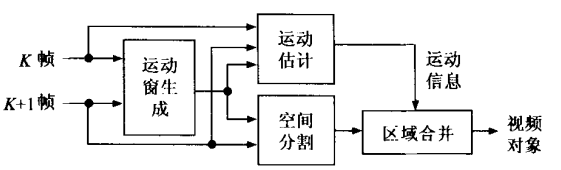

图 1 算法流程图

2 运动窗生成

运动窗的生成是建立在运动变化区域基础上的, 为此需要先检测出运动变化区域, 然后再生成运动窗.

收稿日期: 2002-05-16; 修回日期: 2003-11-06

211　运动变化区域检测

设 $f(x, y, t)$ 表示一个视频序列, $f_k(x, y)$、$f_{k+1}(x, y)$ 分别表示其中的第 k、k+1 帧, 假定视频场景不存在由摄像机运动而引起的全局运动, 则 $f_k(x, y)$、$f_{k+1}(x, y)$ 可表示成:

$$f(x, y) = b(x, y) + m(x, y) + p(x, y) + n_k(x, y) \quad (1)$$

$$f_{k+1}(x, y) = b(x, y) + m(x + \$x, y + \$y) + q(x, y) + n_{k+1}(x, y) \quad (2)$$

其中 $b(x, y)$ 表示第 k、k+1 帧的共同背景, $m(x, y)$、$m(x+\$x, y+\$y)$ 分别对应第 k、k+1 帧中的运动目标, $\$x$、$\$y$ 表示运动目标从第 k 帧到 k+1 帧的位移矢量分量, $p(x, y)$、$q(x, y)$ 分别表示由运动目标引起的遮挡和显露的背景, $n_k(x, y)$、$n_{k+1}(x, y)$ 分别表示第 k、k+1 帧中的噪声. 令式(1)、(2)相减, 可得差分图像:

$$\begin{aligned} d(x, y) &= f_{k+1}(x, y) - f_k(x, y) \\ &= [m(x + \$x, y + \$y) - m(x, y)] + q(x, y) \\ &\quad - p(x, y) + [n_{k+1}(x, y) - n_k(x, y)] \end{aligned} \quad (3)$$

上式中, $m(x+\$x, y+\$y) - m(x, y)$、$p(x, y)$ 和 $q(x, y)$ 三部分都是由目标运动引起的运动变化区域, 令 $MR(x, y) = [m(x+\$x, y+\$y) - m(x, y)] + q(x, y) - p(x, y)$ 表示运动变化区域, 令 $n(x, y) = n_{k+1}(x, y) - n_k(x, y)$ 表示相邻两帧之间的相对噪声, 可得到:

$$d(x, y) = MR(x, y) + n(x, y) \quad (4)$$

式(4)表明, 差分图像中包括相对噪声和由运动目标所引起的运动变化区域两部分, 其中运动变化区域又包括真正的运动目标、遮挡背景和显露背景三部分. 大多数文献采用将差分图像和一个固定阈值相比较的方法来区分运动变化区域和相对噪声, 其中阈值是人工预先设定的, 具有很强的主观性, 因此有必要寻找一种方法能自动地对运动变化区域和相对噪声进行划分. 由于视频序列的多样性, 使得很难用某种具体的特征对运动变化区域进行描述, 因此直接提取运动变化区域是比较困难的. 考虑到在成像过程中噪声 $n_k(x, y)$ 和 $n_{k+1}(x, y)$ 的分布具有一定规律性, 通常可假设服从高斯分布[3], 由于服从高斯分布的两个随机变量之差仍然服从高斯分布, 则相邻两帧之间的相对噪声 $n(x, y)$ 也服从高斯分布, 因此我们可以通过估计相邻两帧之间的相对噪声的特征参数对差分图像中的相对噪声进行滤波, 从而提取出运动变化区域.

设第 k、k+1 帧之间的相对噪声 $n(x, y)$ 服从均值为 L、方差为 R^2 的高斯分布: $n(x, y) \sim N(L, R^2)$. 为检测出相对噪声, 需要先估计出相对噪声的特征参数, 包括其均值和方差. 由于差分图像中的相对噪声具有相同的分布, 而运动变化区域不具有此性质, 因此我们可以首先估计出一个较为可信的相对噪声的均值和方差的初始估计值, 并计算差分图像中各象素的灰度值与相对噪声均值之间的误差, 然后采用迭代加权法, 即根据误差的大小在下一次的相对噪声均值估计过程中对各个象素分配相应的权值, 误差小的象素分配的权值大, 误差大的象素分配的权值小, 即:

$$L_{k+1} = \frac{\mathop{E}_{x, y} d(x, y) \# w_{xy}^k}{\mathop{E}_{x, y} w_{xy}^k} \quad (5)$$

$$R_{k+1}^2 = \frac{\mathop{E}_{x, y} |d(x, y) - L_{k+1}|^2 \# w_{xy}^k}{\mathop{E}_{x, y} w_{xy}^k} \quad (6)$$

其中 L_{k+1}、R_{k+1}^2 分别表示第 k 次迭代后得到的相对噪声的均值和方差, w_{xy}^k 表示第 k 次迭代中象素 (x, y) 的权值. 概率论中高斯分布的/3R0 特性表明, 服从高斯分布的随机变量主要分布在以均值为中心的 3R 范围内, 因此当 $|d(x, y) - L_k| > 3R_k$ 时, 表明差分图像中的象素 (x, y) 属于相对噪声的可能性极小, 该象素在相对噪声均值估计过程中所起的作用也极小, 因此在相对噪声均值估计的下一次迭代过程中赋予该象素最小的权值/00; 当差分图像中象素 (x, y) 的灰度值与相对噪声均值之间的误差不超过相对噪声的方差时, 即当 $|d(x, y) - L_k| \leqslant R_k$ 时, 表明该象素属于相对噪声的可能性较大, 因此在相对噪声均值估计的下一次迭代过程中赋予该象素最大的权值/10; 当 $|d(x, y) - L_k| > R_k$ 时, 随着差分图像中象素 (x, y) 的灰度值与相对噪声均值之间的误差的逐渐增大, 象素 (x, y) 属于相对噪声的可能性将逐渐减小, 因此在下一次迭代过程中赋予它的权值也逐渐减小, 为简便计, 设计当 $R_k < |d(x, y) - L_k| \leqslant 3R_k$ 时, 随着 $|d(x, y) - L_k|$ 的增大, 赋予象素 (x, y) 的权值为一个线性递减函数, 即 w_{xy}^k 由下式确定:

$$w_{xy}^k = \begin{cases} 1, & |d(x, y) - L_k| \leqslant R_k \\ \dfrac{3}{2} - \dfrac{1}{2R_k}|d(x, y) - L_k|, & R_k < |d(x, y) - L_k| \leqslant 3R_k \\ 0, & |d(x, y) - L_k| > 3R_k \end{cases} \quad (7)$$

式中 L_k、R_k 是第 k-1 次迭代后的相对噪声的均值和均方差. 当 $L_{k+1} = L_k$, $R_{k+1}^2 = R_k^2$ 或相差很小时迭代结束, 且相对噪声的最优均值和方差为:

$$L_{opt} = L_{k+1} \quad (8)$$

$$R_{opt}^2 = R_{k+1}^2 \quad (9)$$

该方法性能好坏的关键在于相对噪声均值和方差的初始估计值的可靠程度. 在相邻两帧之间的运动变化区域与整个视频图像的面积之比不太大的假设下(大多数实际的视频序列满足此假设条件), 由于差分图像中相对噪声区域较大且具有相同分布而运动变化区域不具有此性质, 因此可选择差分图像的均值和方差作为相对噪声的均值和方差的初始估计值, 并以此为基础采用迭代加权法估计出相对噪声的最优均值 L_{opt} 和方差 R_{opt}^2.

利用估计的相对噪声特征参数对差分图像中的相对噪声进行滤波可提取出运动变化区域. 由于相对噪声服从高斯分布, 根据高斯分布的/3R0特性可知, 相对噪声中的绝大部分集中在以均值为中心的/3R0 范围内, 即相对噪声主要分布在 $[L_{opt} - 3R_{opt}, L_{opt} + 3R_{opt}]$ 范围内, 因此可以设计相对噪声判决准则如下:

$$L_{opt} - T \leqslant d(x, y) \leqslant L_{opt} + T \quad (10)$$

虽然相对噪声主要集中在以均值为中心的/3R0 范围内, 但并不完全处于/3R0 范围内, 因此 T 应比 $3R_{opt}$ 略大, 在本文中 T 取为 $4R_{opt}$.

若 d(x, y) 满足式(10)所示的准则, 则(x, y) 判决为相对噪声并进行滤除. 通过上述方法可将差分图像中绝大多数的相对噪声滤除, 剩下的孤立噪声可用形态滤波的方法滤除掉, 最后得到运动变化区域, 即:

$$MR = O_m\{dc(x, y)\} \tag{11}$$

其中 dc(x, y) 表示按式(10)所示的准则对差分图像进行滤波后的结果, MR 表示运动变化区域, $O_m\{\#\}$ 表示形态操作算子.

2.2 运动窗生成

由于视频中可能存在多个运动目标, 因此可能存在多个运动变化区域. 设 MR_i 表示第 i 个运动变化区域, 定义对应于运动变化区域 MR_i 的运动窗为 MR_i 的外接矩形. 设 $RECT_i$ 表示 MR_i 的外接矩形, $RECT_i$ 的中心为 (x_i, y_i), 在 x, y 方向上的长度分别为 l_x, l_y. 由于运动变化区域检测过程中可能存在误差而使检测出的区域可能比实际的小, 为保证生成的运动窗的准确性, 考虑将 $RECT_i$ 适当扩大, 故对应于运动变化区域 MR_i 的运动窗 W_i 为:

$$W_i = rect\left(\frac{x - x_i}{K\#l_x}, \frac{y - y_i}{K\#l_y}\right) \tag{12}$$

其中 K 值应略大于 1, 在本文中 K 值取为 1.1.

当对应于每个运动变化区域的运动窗生成后, 空间分割、运动估计和区域合并等过程就只在视频图像对应于运动窗内的部分进行, 从而大大减少了计算量, 提高了运算速度.

3 空间分割

空间分割是按照人的视觉特性根据灰度或色彩的相似性将视频图像划分成若干区域. 分水岭算法[7]是一种著名的空间分割方法, 该方法首先确定种子域, 然后将种子域以外的象素按照一定准则划分到与它最相似的相邻区域中, 它是一种区域增长的方法.

由分水岭算法可知, 符合人眼视觉特性的空间分割后形成的区域 R_i 满足如下三个特性:

(1) 区域 R_i ($i = 1, 2, \cdots, k$, k 表示区域数目) 内的象素具有灰度或色彩相似的特点;

(2) 对于任一个区域 R_i ($i = 1, 2, \cdots, k$), 一定存在着一个对应的种子域 S_i, 使得 $R_i \supseteq S_i$, 特别地, $R_i = S_i$ 表示该区域完全由其种子域构成;

(3) 对于任意两个不同的种子域 S_i, S_j ($i \neq j$), 不存在一个区域 R_m, 使得 $R_m \supseteq S_i$ 和 $R_m \supseteq S_j$ 同时成立.

特性(2) 表明, 任何一个区域都是由种子域经一定方式增长而成; 特性(3) 表明, 任何一个区域内不可能包含两个或以上的种子域, 即一个区域只能由一个种子域增长而成.

根据上述特性, 我们提出一种逐级划分的方法对视频图像进行空间分割, 使得最后形成的区域 R_i 满足上述三个特性. 与分水岭算法采用自下而上的区域增长的思想不同, 本文的方法是一种自上而下的逐级划分的空间分割方法, 每一次的划分过程包含两个阶段: 第一, 根据灰度或色彩相似性进行二分类划分; 第二, 根据空间连接性进一步划分.

设 I(x, y) 表示视频图像, 定义 I(x, y) 中的每个种子域 S_i 为空间相邻且图像亮度梯度满足局部极小值条件的象素构成的区域:

$$S_i = \{(x, y) \mid Grad(x, y) \text{ 满足局部极小值且 } (x, y) \text{ 在空间上相邻}\} \tag{13}$$

其中 Grad(x, y) 表示图像在 (x, y) 处的亮度梯度, 然后开始对 I(x, y) 进行划分. 首先在灰度特征空间中根据一定的分类准则将 I(x, y) 中的象素按灰度相似性划分为两类, 分别记为 C_1, C_2, 即:

$$I(x, y) = \bigcup_{i=1}^{2} C_i \tag{14}$$

其中 $\bigcup$ 表示联合, 接下来在每一类 C_i ($i = 1, 2$) 中, 根据象素的空间连接性将 C_i 划分为若干区域 R_{ik}, 且有:

$$C_i = \bigcup_{k=1}^{m} R_{ik} \quad (m \text{ 表示对 } C_i \text{ 划分的区域数目}) \tag{15}$$

$$R_{ik} = \{(x, y) \mid (x, y) \in C_i \text{ 且满足空间连接性}\} \tag{16}$$

对每一个区域 R_{ik}, 若其中包含的种子域数目为 1, 则定义该区域为基区域, 并停止对该区域进一步划分; 若其中包含的种子域数目大于 1, 则定义该区域为一个父区域, 并进入下一级的划分过程. 在对一个父区域进行划分时, 仍然采用和上述相同的划分过程: 首先根据一定的分类准则按灰度相似性对其进行二分类划分, 形成两个不同的类别; 接下来在每一类中按空间连接性进一步划分成若干新的区域, 称为该父区域的子区域. 若子区域满足基区域的条件, 则停止对该子区域进行划分, 否则将该子区域作为下一级划分的父区域, 并进入下一级的划分过程中. 重复此过程, 直到所有区域都满足基区域条件.

在本文中, 按灰度相似性将父区域 R 中的象素划分为两类时采用的分类准则是类间方差最大准则, 详细描述见文献[8].

空间分割能够比较准确地得到运动对象的边界, 但一般容易造成过分割, 即图像被分割成数目众多的小区域, 因此空间分割往往得不到具有语义的视频对象. 要得到具有语义的视频对象, 必须联合时间分割对空间分割后形成的区域进行合并. 为此还需要进行运动估计, 然后再根据运动相似性对区域进行合并.

4 运动估计与区域合并

运动估计是基于运动相似性的区域合并的基础, 常用的运动估计方法有光流场法和块匹配法, 本文采用块匹配法估计视频图像的运动窗内的各个象素的运动矢量.

在区域合并之前, 需要先计算各区域的运动参数, 常用的运动模型有八参数模型、六参数模型等[9], 本文采用如下所示的六参数仿射运动模型:

$$\begin{cases} d_x = a_1 x + a_2 y + a_3 \\ d_y = a_4 x + a_5 y + a_6 \end{cases} \tag{17}$$

其中 d_x, d_y 是象素 (x, y) 的运动向量.

对于每个区域 R_i, 定义其误差函数为:

$$E(a) = \sum_{(x, y) \in R_i} [(d_x - a_1 x - a_2 y - a_3)^2 + (d_y - a_4 x - a_5 y - a_6)^2] \tag{18}$$

其中 $a = (a_1, a_2, \cdots, a_6)$, 则区域 R_i 的最优运动参数就是使式(18)达到最小的矢量 a:

$$a_{opt} = \arg\min_{a} E(a) \quad (19)$$

每个区域的最优运动参数可通过最小二乘法对误差函数求最小值而得到.

求得每个区域的运动参数后, 根据运动相似性对区域进行合并, 详细方法见文献[2, 6].

5　实验结果与分析

根据本文的方法, 我们对不同的视频序列进行了测试, 下面列出两组测试结果: 图 2 和图 3 分别表示/ hall monitor0 序列和/ toy vehicle0 序列, 其中(a) 表示前一帧, (b) 表示当前帧, (c) 是在差分图像中根据估计的相对噪声的均值和方差按式(10)所示的准则对相对噪声滤波后的结果, (d) 是对(c) 进行形态滤波去除掉孤立噪声和小区域后生成的运动窗, (e) 是对当前帧的运动窗内的部分进行的空间分割, (f) 是将空间分割后生成的各区域按运动相似性聚类合并生成的视频对象. 实验结果表明, 本文的方法能有效地生成运动窗, 大大减少了运算量, 并且本文提出的空间分割方法能有效地进行空间分割, 最后能取得较好的视频分割结果.

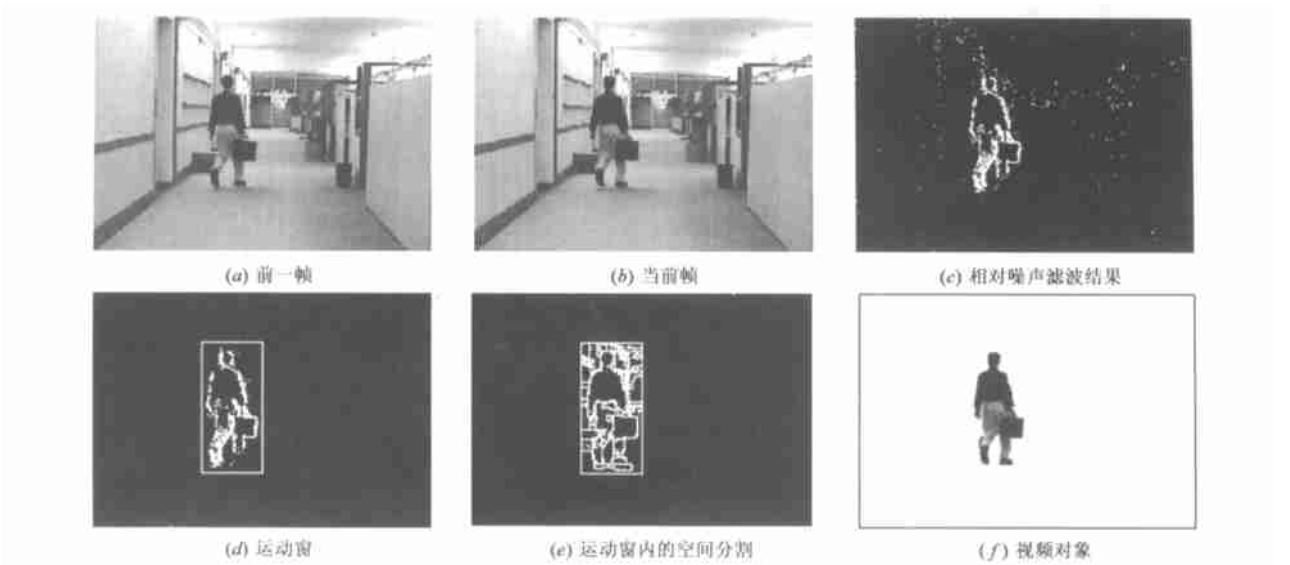

图 2　hall monitor 图像序列实验结果

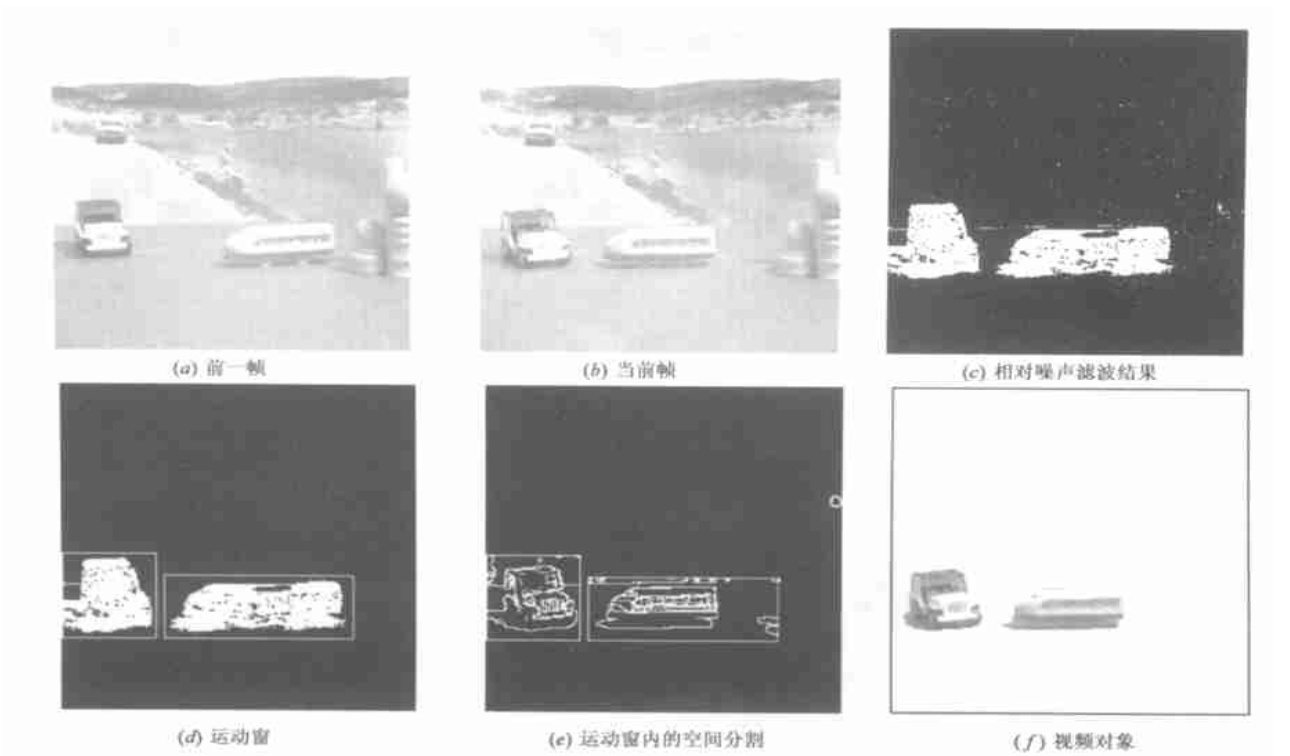

图 3　toy vehicle 图像序列实验结果

6 结论

本文提出了一种基于运动窗生成的时空视频分割方法. 首先在差分图像中通过估计相对噪声的特征参数而检测出运动变化区域并生成运动窗, 然后在运动窗内按时空特性进行分割, 从而大大减少了运算量, 提高了运算速度. 在空间分割中, 本文提出了一种符合人眼视觉特性的逐级划分的空间分割方法, 最后根据运动相似性将区域合并, 实现视频分割. 实验结果表明, 本文的方法能大大减少运算量并能有效地进行空间分割, 最后能取得较好的视频分割结果. 关于本文的方法和其他方法的性能比较, 以及如何更有效地进行区域合并, 将是我们今后进一步的研究工作.

参考文献:

[1] J Fan, J Yu, et al. Spatiotemporal segmentation for compact video representation[J]. Signal Processing, 2001, 16: 553-566.

[2] 黄波, 杨勇, 等. 一种基于时空联合的视频分割算法[J]. 电子学报, 2001, 29(11): 1491-1494.

[3] M Kim, J G Choi, D Kim, et al. A VOP generation tool: Automatic segmentation of moving objects in image sequences based on spatio-temporal information[J]. IEEE Trans, 1999, CSVT-9(8): 1216-1226.

[4] J B Pineau, F Morier, et al. Hierarchical segmentation of video sequence for content manipulation and adaptive coding[J]. Signal Processing, 1998, 66: 181-201.

[5] A A Alatan, L Onural, et al. Image sequence analysis for emerging interactive multimedia services-the European COST 211 framework[J]. IEEE Trans, 1998, CSVT-8(7): 802-813.

[6] J G Choi, S W Lee and S D Kim. Spatio-temporal video segmentation using a joint similarity measure[J]. IEEE Trans, 1997, CSVT-7(2): 279-286.

[7] L Vincent, P Soille. Watersheds in digital spaces: an effective algorithm based on immersion simulations[J]. IEEE Trans, 1991, PAMI-13(6): 583-598.

[8] N Otsu. A threshold selection method from gray-level histograms[J]. IEEE Trans, 1979, SMC-9(1): 62-66.

[9] A M Teklap. Digital Video Processing[M]. USA: Prentice Hall, 1995.

作者简介:

朱 辉 男, 1974 年 6 月生于四川泸州, 2003 年在电子科技大学获通信与信息系统博士学位, 主要研究方向为多媒体通信与图像处理.

李在铭 男, 1939 年 4 月生于重庆, 教授, 博士生导师, 主要研究方向为多媒体通信与信号处理.

蔡 毅 男, 1971 年 11 月生于四川, 北京邮电大学电子与信息工程硕士毕业, 高级工程师, 从事 IT 系统建设方面的工作.