

一种动态场景下基于时空信息的 视频对象提取算法

田宏阳 陈辉 马文静

(山东大学信息科学与工程学院, 济南 250100)

摘要 在实际应用中,许多视频序列具有运动背景,使得从其中提取视频对象变得复杂,为此提出了一种基于运动估计和图形金字塔的动态场景下的视频对象提取算法。该算法首先引入了相位相关法求取运动向量,因避免了视频序列中光照变化的影响,故可提高效率和稳健性;接着再根据参数模型进行全局运动估计来得到最终运动模板;然后利用图形金字塔算法对当前模板内图像区域进行空间分割,最终提取出语义视频对象。与现有算法相比,对于从具有动态场景的视频流中提取运动对象的情况,由于使用该算法能有效地避开精准背景补偿,因而不仅节省了计算量,而且提取出来的语义对象精度较高。实验表明,无论是对动态场景中刚性还是非刚性运动物体的分割,该算法都具有较好的效果。

关键词 动态场景 视频对象提取 相位相关 运动估计 图形金字塔

中图分类号: TP391.41 TN191.81 **文献标识码**: A **文章编号**: 1006-8961(2007)09-1653-06

Video Object Extraction Algorithm Based on Spatio-temporal Information in Dynamic Scene

TIAN Hong-yang, CHEN Hui, MA Wen-jing

(School of Information Science and Technology, Shandong University, Jinan 250100)

Abstract In practical applications, many video sequences have moving background, and then the extraction of video object becomes complicated. An algorithm is proposed in the paper to extract video object from dynamic scene based on motion estimation and the graph pyramid. Phase correlation is first used to obtain the motion vector with high efficiency and robustness, and to weaken the impacts of illumination in the video sequence. Then global motion estimation with parameter-model is used to find the final motion template. Finally, to extract the semantic video object, spatial segmentation using the graph pyramid is applied to the image region in the current motion template. Compared with some prevailing methods, in the case of extraction of moving object from video sequences of dynamic scene, our algorithm avoids precise background compensation and is very computationally efficient, while the extracted semantic object is of high precision. The experimental results show that both rigid and non-rigid moving objects in dynamic scene are well extracted by this algorithm.

Keywords dynamic scene, video object extraction, phase correlation, motion estimation, graph pyramid

1 引言

为了实现基于内容的视频编解码、基于内容的视频跟踪以及基于内容的视频索引和检索,视频对象(video object, VO)的概念在 MPEG-4 中被提了出

来^[1]。这样从视频序列中提取出语义对象就成了关键问题,而在视频流的记录过程中,为了扩大视频流中的信息量,摄像机不可避免产生缩放、平移和旋转等运动。由摄像机运动产生的全局运动使得视频序列具有了动态场景。较之静止背景,从动态场景中提取出运动对象,在视频跟踪、视频摘要和数字

基金项目:山东省优秀中青年科学家科研奖励基金(03BS001);山东省自然科学基金重点项目(Z2005G02)

收稿日期:2006-02-24;改回日期:2006-07-07

第一作者简介:田宏阳(1982~),男,硕士研究生。主要研究方向为视频分析、图像处理、计算机视觉。E-mail: hondri82@yahoo.com.cn

监控等领域有更加重大的实用价值。

根据分割过程中所利用的信息不同,视频分割算法可分为时间分割、空间分割和时空联合分割。为得到更好的分割效果,当前主流的视频对象提取算法都是将时间域连续性和空间域的相似性结合起来,以获得语义意义的视频对象。在动态场景下,由于背景的运动,使得直接分割变得非常困难。为此,Tsaig 提出了一种首先使用透射模型和基于梯度的技术进行帧间运动补偿,再进行分水岭分割,最后利用马尔可夫随机域模型进行聚类处理的方法^[2]。Hsu 提出了一种首先利用全局和局部结合的运动补偿方法将相邻帧配准,然后将强度变化检测信息和 Canny 边缘信息融合得到运动物体的方法^[3]。文献[4]则先利用基于自适应 Fisher 外点滤除的全局运动估计实现背景对准,然后根据显著性测试得到二值化帧差图,最终获得运动物体的粗略轮廓。在上述各方法中,处理时都要先进行精准背景补偿^[2~4],这往往是整个分割过程中最耗时的阶段。

本文提出一种基于时空联合的动态场景视频对象提取算法,该算法先利用相邻帧间的运动信息得到视频分割的大致轮廓,然后利用空域分割方法自动提取出视频对象。实验表明,本文算法能有效地避开精准背景补偿,不仅可节省计算量,而且在动态场景下,本文算法能较为准确地提取出前景物体。

2 算法概述

本文提出了一种基于运动估计和图形金字塔的动态场景视频分割算法。其系统流程图如图 1 所示。该算法在时间域采用了一种粗略的整体运动分析方法,即首先利用频域相位关系求得运动向量;然后根据前景与背景运动向量的不同,采用相应的全局运动参数模型,通过反复处理来得到当前帧的前景区域;最后进行形态学后处理,以去除干扰块和填充内部孔洞,即得到完整封闭的前景物体,即视频分割的大致轮廓。在得到前景物体大致区域以后,再根据图形金字塔算法,对运动模板内的图像区域进行空间域分割,即可提取出最终语义视频对象。

3 相邻帧间运动学分析

全局运动,又称为主运动,通常是由摄像机运动产生的背景运动,而视频序列中活动物体的运动则

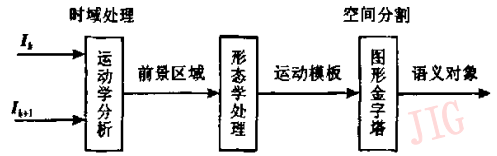


图 1 视频对象提取算法原理框图

Fig. 1 Algorithm structure map for extracting video object

是局部运动,又称为从运动^[5]。由于主从运动不一致,通常前景与背景间的运动向量是有显著偏离的,利用这些差异就可以实现前景区域的分割。本文采用一种基于块的方法,可以有效地从背景中提取运动物体。图 2 描述了此算法的步骤。

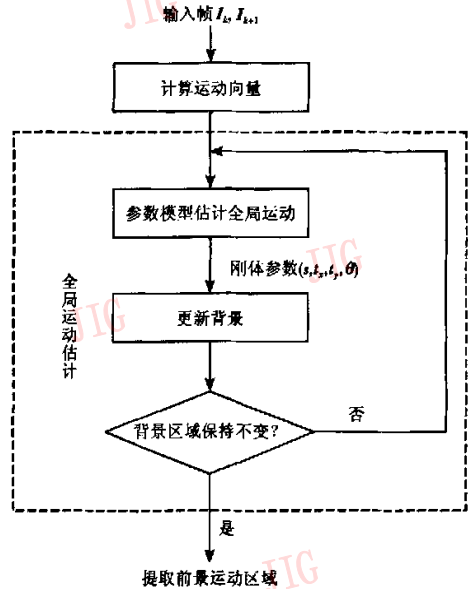


图 2 运动分析流程图

Fig. 2 Flow chart for motion analysis

3.1 计算运动向量

运动向量的计算是整个运动学分析的核心部分,也是利用参数模型进行全局运动估计的基础,其复杂度和准确性决定了前景运动物体提取的效率和稳健性。本文提出了一种基于相位相关^[6]的运动向量估计算法。不同于基于亮度匹配搜索的传统块匹配算法,相位相关法是通过两区域的相位来直接估计它们之间的运动。

假定两帧具备以下平移关系:

$$s(n_1, n_2, k) = s(n_1 + d_1, n_2 + d_2, k + 1) \quad (1)$$

其傅里叶变换为

$$S_k(f_1, f_2) = S_{k+1}(f_1, f_2) \cdot e^{j2\pi(d_1 f_1 + d_2 f_2)} \quad (2)$$

定义归一化互功率谱为

$$\tilde{C}_{k,k+1}(f_1, f_2) = \frac{S_{k+1}(f_1, f_2) \cdot S_k^*(f_1, f_2)}{|S_{k+1}(f_1, f_2) \cdot S_k^*(f_1, f_2)|} \quad (3)$$

由式(2)、式(3)得

$$\tilde{C}_{k,k+1}(f_1, f_2) = e^{-j2\pi(d_1 f_1 + d_2 f_2)} \quad (4)$$

对式(4)进行傅里叶逆变换得到

$$\tilde{c}_{k,k+1}(n_1, n_2) = \delta(n_1 - d_1, n_2 - d_2) \quad (5)$$

通过找寻式(5)中的冲激函数峰值所在的位置,就可以得到平移参数,该平移参数即为运动向量。初始处理时,当前帧被分成 8×8 大小的块。为了在不同帧间正确地估计对应块的互相关性,可先将块扩展为 16×16 大小,然后执行相位相关计算(如图3所示)。图4是两个 16×16 块的相位相关估计。一般而言,相位相关法能够更加准确地计算运动向量,特别是在运动平移较大的情况下。

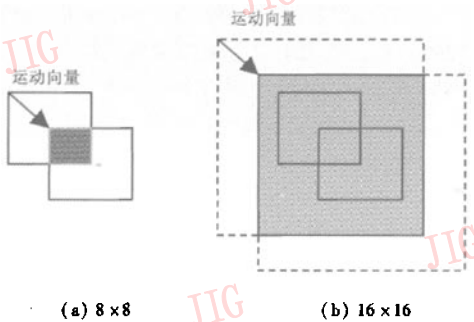


图3 块大小为 8×8 和 16×16 的相关区域
Fig. 3 Correlation area with size 8×8 and 16×16

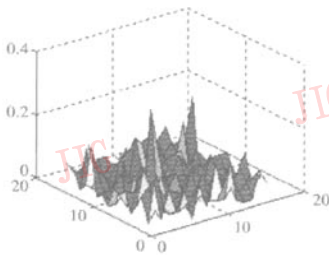
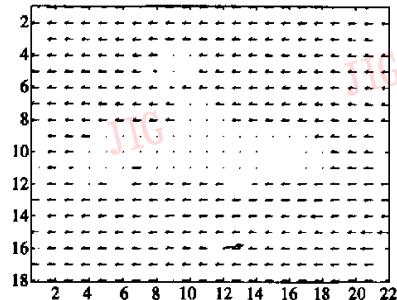


图4 两块间的相位估计
Fig. 4 Phase correlation between two blocks

图5(b)为通过对MPEG-4标准coastguard序列应用相位相关法得到的运动向量。由于在频谱计算过程中使用了快速傅里叶变换^[7],因而提高了运算速度,而且由于式(3)中对互功率谱进行了归一化,



(a) 236帧



(b) 236帧对应的运动向量

图5 标准coastguard序列236帧及其对应的运动向量

Fig. 5 Standard coastguard sequence 236 frame and its motion vector

因而在相位分析过程中还能够避免光照的影响和提高算法的稳健性。

3.2 全局运动估计提取运动物体

在得到当前帧运动向量后,即可采用刚体参数模型在整幅图像上粗略估计全局运动^[8]。

$$\begin{pmatrix} \hat{x}_j \\ \hat{y}_j \end{pmatrix} = s \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_j \\ y_j \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad (6)$$

(x_j, y_j) 和 (\hat{x}_j, \hat{y}_j) 分别为第 k 帧中第 j 个背景块及其在第 $k-1$ 帧中匹配块的中心点。 $s, \theta, (t_x, t_y)$ 分别表示尺度因子,旋转参数和平移向量。根据3.1节中得到的运动向量,就可以求出第 k 帧中第 j 个背景块在 $k-1$ 帧中相应的匹配块。假定起始时当前帧(即第 k 帧)内所有的块都属于背景区域,则误差函数 Φ 定义如下:

$$\Phi = \sum_{j=1}^n [(s \cdot \cos\theta x_j - s \cdot \sin\theta y_j + t_x - \hat{x}_j)^2 + (s \cdot \sin\theta x_j - s \cdot \cos\theta y_j + t_y - \hat{y}_j)^2] \quad (7)$$

全局运动估计时,先通过最小二乘方法最小化误差函数 Φ ;然后经过奇异值分解来获得全局运动参数^[9];最后再由求得的全局运动参数 s, θ 和

(t_x, t_y) 来预测 (x_j, y_j) 的匹配点 $(\tilde{x}_j, \tilde{y}_j)$ 。若运动向量预测值 (\hat{x}_j, \hat{y}_j) 与全局运动参数预测值 $(\tilde{x}_j, \tilde{y}_j)$ 之间的欧氏距离 d_j 大于给定的阈值, 则认为当前块运动不同于全局运动, 此时当前块被从背景区域中除去, 并被标定为前景物体。反复进行上述步骤, 直至背景不再变化, 即可得到前景区域 (如图 6(a))

所示)。

对图 6(a) 先进行连通域分析, 再通过最小尺度门限去除干扰块后来得到前景模板 (图 6(b)); 然后经过膨胀以及内部空洞填充等形态学处理后即得到最终的运动模板 (图 6(c)), 以保证模板 (图 6(c)) 内包括全部前景物体。



图 6 标准 coastguard 序列 236 帧前景提取及后处理

Fig. 6 Standard coastguard sequence 236 frame foreground extraction and subsequent processing

4 图形金字塔分割

在通过运动学分析得到前景物体的大致区域后, 图形金字塔分割算法被用来从当前活动区域内提取更精确的运动物体。

图形金字塔 (也称为不规则金字塔) 分割算法^[10] 是一种基于多分辨率处理的有效方法, 它是一种静止图像的区域生长分割算法。它的数据结构有其特殊性, 即每一层金字塔都由一个邻接图和一个分区集组成 (如图 7(a) 和图 7(b) 所示)。

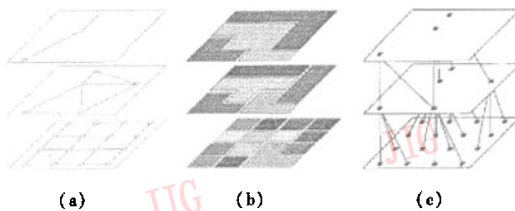


图 7 不规则金字塔的 3 种结构 (4 × 4 图像采样)

Fig. 7 Three structure of irregular pyramid (4 × 4 image sampling)

图像金字塔的不规则采样分级结构是自下而上的, 这样使其能够自动适应输入图像的内容。一般来说, 开始重建时, 在金字塔底层, 首先原始图像每个像素被认为是一个区域, 同时通过 4 或 8 连通关系产生邻接图; 然后在邻接图上通过设置顶点间的颜色差值门限来建立每一层顶点的类成员从属关

系, 即在邻接图上两个邻接区域 R_i 和 R_j , 如果它们之间的彩色空间距离小于给定的门限 T , 则定义为相似顶点。这样就可通过邻接图和其中顶点的属性, 建立一个相似度图来连接相似的邻接顶点。

一个基于相似度图的局部随机消减处理允许局部地选择下一层的顶点, 也就是本层的保留顶点, 而它们将成为下一层的顶点。在消减处理过程中, 为了能够并行局部处理, 有以下两个条件需要满足:

$$\forall i = 1, 2, \dots, u \quad (c_i[l], c_0[l]) \in E[l] \quad (8)$$

$$\forall c_0[l] \in V[l], d_0[l] \in V[l] \quad (c_0[l], d_0[l]) \notin E[l] \quad (9)$$

上式中, c_0, d_0 是 l 层的保留顶点, c_i 为 c_0 的邻接顶点, $V[l], E[l]$ 表示 l 层上的顶点集和邻接边集。

在由第 l 层顶点中选取保留点作为第 $l+1$ 层顶点的过程中, 式 (8) 表明, 由于 l 层的每一个非保留顶点在它的邻域至少有一个保留顶点, 从而使它能够通过局部选择分派给一个 $l+1$ 层顶点, 式 (9) 表明, 由于 l 层上的两个相邻接的顶点不能同时保留到第 $l+1$ 层, 从而保证了顶点数目能逐层递减。

在分组阶段, 非保留顶点归为其邻域中最相似的保留顶点, 即同类区域中的保留顶点, 这样通过子-父关系在下一层就可形成更大的区域 (如图 7(c) 所示)。在新的一层, 根据新的顶点邻域间关系, 邻接图被重新定义。新的一层可以通过同样处理来获得它的下一层, 直到不能被融合为止。

由于在运动学分析中已经得到活动对象的粗略区域(运动模板),本文采用了一种局部图形金字塔算法来进行分割。在局部图形金字塔中,只有部分图像像素与金字塔底层邻接图顶点一一对应,其余的像素只对应一个或少量几个根顶点。这样由于只有部分图像被处理,因此这种局部地分割不仅提高了处理速度,而且能显著地减小空域错分的概率和改善最终的分割效果。

运动分析中得到的活动对象区域是基于块的,与物体形状相关的有界多边形。由于多边形外的图像区域都属于背景物体,因此初始化时,其对应图形金字塔底层的一个根顶点,而多边形内部是待分割区域,其内部图像区域的每一个像素对应图像金字塔底层的一个顶点。这样在逐级分割过程中,就可以实现多边形内部区域的分割,即与背景相似度高的部分被融合到背景区域中去,而其余的则合并为一个或者几个区域,代表前景物体。

较之 Haris 改进的分水岭分割算法^[11]而言,两者同作为经典的图像分割工具,都是利用区域(像素)之间的邻接和相似性进行区域生长和区域合并。Haris 是先采用分水岭算法完成图像的初始分割,然后使用快速的区域合并算法将颜色距离最

相近的区域进行逐步合并,当图像中的区域达到设定数目时,则停止合并。虽然通过区域合并在一定程度上解决了经典分水岭算法^[12]的过度分割现象,但却不能直接从运动模板中精确地提取运动物体。本文采用图形金字塔算法则能够将当前帧中运动模板以外的背景区域归为金字塔底层一个根顶点,仅对运动模板内的图像区域进行空域分割,这不仅减少了处理复杂度,还提高了正确提取出语义对象的几率。这种处理不但能提取出来当前的运动物体,而且能将此运动物体分割为均一的区域。

5 实验结果

为检验算法的有效性,本文采用标准 MPEG-4 测试序列 coastguard, foreman 进行仿真实验。由于波浪和绿地,致使 coastguard 序列背景纹理比较复杂,并且船艇与背景有不同的运动;而 foreman 序列则不仅伴有摄像机旋转,并且活动前景是非刚性物体。两者均采用 QCIF 格式的视频序列,每帧图像大小为 176 × 144。具体实验结果如图 8 所示。

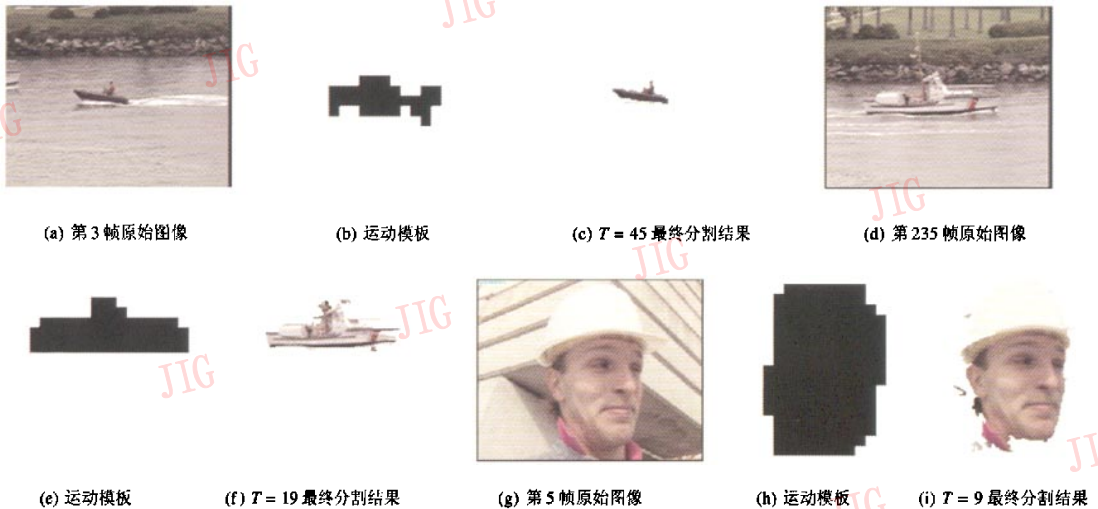


图 8 Coastguard 和 foreman 序列的实验结果

Fig. 8 Experimental results of sequence coastguard and foreman

T 为金字塔分割时判定邻接图中相似顶点的门限,但其要针对不同的序列进行调整。仿真实验在 Pentium IV 1.5GHz, 内存 256M 的机器上进行,对于

上述 QCIF 格式的视频序列,每帧平均处理时间为 1s。实验结果表明,本文算法不仅能有效地对动态场景中刚性与非刚性运动物体进行分割,并能提取

出具有语义意义的视频对象。

6 结 论

本文提出了一种基于动态视频场景的运动物体分割算法,它具有以下特点:(1)较之传统的动态场景分割算法,本文由于避开了精准背景补偿阶段,且在运动学分析中首先引入了相位相关法得到运动向量,然后通过全局运动估计获取运动模板,因而节省了计算量,也提高了运算速度;(2)采用局部图形金字塔算法进行空间分割,只有当前运动模板内的图形区域被处理,因而不仅减少了空域分割计算量,而且降低了错误分割的几率。实验表明,该算法不仅计算量小、稳健性高,而且能够从动态场景中快速准确地提取出刚性和非刚性前景物体。如何从动态场景中准确地提取出多个运动物体并进行识别和跟踪将是下一步要研究的问题。

参考文献 (References)

- 1 Sikora T. The MPEG-4 video standard verification model[J]. *IEEE Transactions on Circuits System for Video Technology*, 1997, 7(1): 19 ~ 31.
- 2 Tsaig Y, Averbuch A. Automatic segmentation of moving objects in video sequences: A region labeling approach[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2002, 12(7): 597 ~ 612.
- 3 Hsu Chiou-Ting, Tsai Yu-Sheng. Video object segmentation based on global motion estimation/compensation [A]. In: *Proceedings of International Conference on Consumer Electronics[C]*, Los Angeles, CA, USA, 2001: 168 ~ 169.
- 4 Wu Si, Zhang Yong-dong. Automatic extraction of foreground areas in video sequences of a dynamic scene[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2005, 17(2): 355 ~ 363. [吴思,张勇东. 动态场景视频序列中的前景区域自动提取[J]. *计算机辅助设计与图形学学报*, 2005, 17(2): 355 ~ 363.]
- 5 Sawhney Harpreet S, Ayer Serge. Compact representations of videos through dominant multiple motion estimation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1996, 18(8): 814 ~ 830.
- 6 Tekalp A M. *Digital Video Processing[M]*. Englewood Cliffs, NJ, USA: Prentice-Hall Inc, 1995.
- 7 Reddy B S, Chatterji B N. A FFT-based technique for translation, rotation, and scale invariant image registration [J]. *IEEE Transactions on Image Processing*, 1996, 5(8):1266 ~ 1271.
- 8 Liu S, Yan Z, Kim J W, et al. Global/Local motion-compensated frame interpolation for low bitrate video [A]. In: *Proceedings of Image and Video Communication and Processing[C]*, San Jose, CA, USA, 2000: 223 ~ 234.
- 9 Arum K S, Huang T S, Blostein S D. Least-Squares fitting of two 3-D point sets[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1987, 9(5): 698 ~ 670.
- 10 Montanvert A, Meer P, Rosenfeld A. Hierarchical image analysis using irregular tessellations [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991, 13(4): 307 ~ 316.
- 11 Haris K, Efstratiadis S, Maglaveras N, et al. Hybrid image segmentation using watersheds and fast region merging[J]. *IEEE Transactions on Image Processing*, 1998, 7(12): 1684 ~ 1699.
- 12 Vincent L, Soille P. 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.