

# 基于压缩域的视频关注对象提取

潘琤雯<sup>1)</sup> 张兆扬<sup>1,2)</sup> 石旭利<sup>1,2)</sup> 沈礼权<sup>1,2)</sup>

<sup>1)</sup>(上海大学通信与信息工程学院, 上海 200072)

<sup>2)</sup>(新型显示技术与应用教育部重点实验室, 上海 200072)

**摘要** 针对视频压缩域的对象分割问题,提出了一种视频对象精确提取方法。先将视频对象的颜色空间进行均值偏移和区域生长,得到分块的视频图像。同时对视频编码过程中所产生的运动矢量统计熵值,提取出视频中能引起人眼关注的运动对象。最后利用提出的对象轮廓精确分割方法,提取视频关注对象。实验表明该算法能精确和完整地分割出视频关注对象,并对视频中关注对象的切换、物体的非刚性形变有很好的鲁棒性。

**关键词** 均值偏移 信息熵 运动矢量 对象提取

中图法分类号: TN941.1 文献标识码: A 文章编号: 1006-8961(2009)10-2145-04

## An Approach to Video Object Extraction in the Compressed Domain

PAN Cheng-wen<sup>1)</sup>, ZHANG Zhao-yang<sup>1,2)</sup>, SHI Xu-li<sup>1,2)</sup>, SHEN Li-quan<sup>1,2)</sup>

<sup>1)</sup>(College of Communication and Information Engineering, Shanghai University, Shanghai 200072)

<sup>2)</sup>(Key Laboratory of Advanced Display and System Application, Ministry of Education, Shanghai 200072)

**Abstract** In this paper, we propose a new method to segment video object. Firstly, we smooth each frame with the mean-shift and region-growing method. And then, the motion vectors which are released by video coding are used to extract a moving object. Finally, the object segmentation method proposed in this paper is used to extract the video attention object. Experimental results show that the proposed method can effectively separate the video object from the sequence and has the strong robustness in multi-targets tracking especially for switching targets and deformed objects.

**Keywords** mean-shift, information entropy, motion vector, object extraction

## 1 引言

视频对象的分割是机器人视觉的重要研究部分,同时也是 MPEG-4 基于对象编码的基础<sup>[1]</sup>。视频对象的分割主要依据视频图像的时间、空间信息来进行。

视频对象的分割集中在人眼感兴趣的对象上。Ahmed 等人通过定位视频中的人脸部分并结合针对颜色空间的处理来确定待分割的对象<sup>[2]</sup>; Hieu Tat Nguyen 等人根据给定的跟踪对象的定位模型参

数来提取物体的轮廓<sup>[3]</sup>; Chan 等人针对具有复杂纹理且变化剧烈的背景的视频,提出了实时更新背景信息来固定当前瞬时静态背景的内容,从而检测出在当前相对静止背景下的运动对象<sup>[4]</sup>; Mezaris 等人针对具有快速运动物体且运动对象更替频繁的视频提出了基于颜色、运动和位置信息定位初始运动对象的方法<sup>[5]</sup>; 以上的视频分割方法都各有针对性,也各有利弊。为了实现高效而精准的视频关注对象分割,提出了一种基于 H. 264 编码压缩域的视频关注对象提取方法。

基金项目:国家自然科学基金项目(60832003)

收稿日期:2009-06-18; 改回日期:2009-07-28

第一作者简介:潘琤雯(1985 ~ ),女。上海大学通信与信息学院电路与系统专业硕士研究生。研究方向为数字视频处理。

E-mail: pancat0622@126.com

## 2 算法描述

### 2.1 算法原理及框图

图 1 为所提出算法的基本框图,此算法利用两路输入数据,一路为视频编码过程中所产生的运动矢量  $Mv(x, y, t)$ , 用来提取关注运动对象。另一路输入为视频图像数据  $Img(x, y, t)$ , 在经过均值偏移平滑算法和区域生长之后可得到相应的分块图像  $Reg(x, y, t)$ 。在获得两路处理数据后,根据提出的关注运动对象分割方法完整地分割出人眼关注对象。

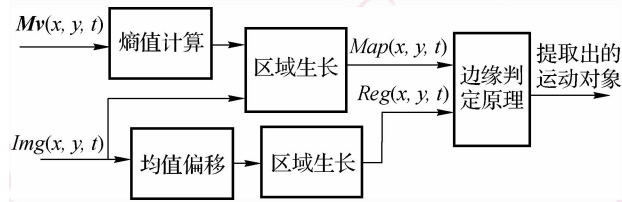


图 1 本算法流程图

Fig. 1 Flowchart of the proposed algorithm

### 2.2 人眼关注对象提取

人眼关注运动对象提取是视频研究中的一个重要部分。分割方法主要可按照所分析信息的不同分为基于空间和基于时间的两种方法。

由于本算法是针对视频压缩域所提出的一种视频关注对象提取方法,因此在确定运动对象时另辟蹊径,利用编码过程中产生的运动矢量作为分析数据<sup>[6]</sup>;通过计算运动矢量的信息熵来衡量视频中对象的运动情况。具体步骤如下:

(1) 利用编码过程中得到的运动矢量计算运动矢量的方向;设第  $n$  帧第  $(i, j)$  个宏块的运动矢量为  $PV(i, j) = (x_{n,i,j}, y_{n,i,j})$ , 该运动矢量的方向为  $\theta_{n,i,j} = \arctan(y_{n,i,j}/x_{n,i,j})$ 。

(2) 利用统计直方图对  $n \times n$  窗口内宏块的运动矢量值进行统计归类,并计算出每一类运动矢量值出现的概率。

$$P_s(k) = \frac{SH_{i,j}^w(k)}{\sum_{l=1}^M SH_{i,j}^w(l)} \quad (1)$$

式中,  $SH()$  表示  $n \times n$  窗口内宏块的运动矢量值的种类和每个种类所占宏块个数的函数关系;  $w$  表示  $n \times n$  的搜索窗口尺寸;  $M$  为运动矢量值的最终分类数;可知其中自变量  $k$  满足  $0 \leq k \leq M$ ;  $P_s(k)$  表示第

$k$  种类别宏块在总宏块的出现概率。

(3) 依据所得的概率分布情况按式(2)计算每一个宏块的运动矢量值的空间相关性熵值。

$$Cs(i, j) = - \sum_{k=1}^M P_s(k) \times \log(P_s(k)) \quad (2)$$

式中,  $Cs(i, j)$  表示当前  $(i, j)$  宏块运动矢量的空间信息熵;  $P_s(k)$  为式(1)中得到的概率分布函数。

(4) 在视频序列中,位于背景区域的宏块通常具有相同的运动矢量。对于背景静止的序列,背景块运动矢量的理论值为零;而背景非静止序列中背景块的运动矢量应和摄像头的变化保持一致,所以可以认定背景块具有较低的信息熵值。据此,可以通过找出信息熵值高的宏块组来提取人眼所关注的运动对象;同时利用区域生长法找出所占连续高熵值宏块数最多的连通区域,将这些联通区域中的宏块所代表的运动对象作为关注运动对象的初始定位信息。

### 2.3 图像平滑分块

采用 mean-shift 算法对图像进行平滑, mean-shift 算法是一个通过迭代计算局部密度极大值的过程。这种平滑算法的特点是,在将具有相同纹理特性的图像内容归为一类的同时,能够很好地保留图像中各个物体的轮廓信息<sup>[7]</sup>。

而在将图像进行平滑之后,利用区域生长法<sup>[8]</sup>,将纹理相似的点合并到同一个封闭类块中,完成图像基于颜色空间的分块工作。

### 2.4 关注运动对象分割

在得到初步的关注运动图像定位图和平滑分块后的视频图像后,提出一种能精确提取运动对象所占图像区域的选择算法,从而准确地分割出运动关注对象。具体步骤如下:

(1) 提取第 2.2 节得到的多个独立联通运动对象的位置信息,作为运动对象分割的初始定位信息。

(2) 由于在视频编码过程中得到的运动矢量存在着一定的误差,因此在经过第 2.2 节熵值计算后提取出的多个运动对象的独立连通区域中也存在由误差产生的非运动对象块,但是这些宏块却占据较少的像素点。基于以上两点原因,本算法在自动提取运动对象过程中利用式(3)~式(6)定义的一系列运动对象显著系数,来筛选最终分割出的一个最受关注的运动对象联通区域。运动对象显著系数定义如下:

$$\alpha(j) = \alpha_L(j) \alpha_{VRate}(j) \alpha_M(j) \quad (3)$$

$$\alpha_L(j) = \frac{1}{\sqrt{(x_c(j) - x_{pic_c})^2 + (y_c(j) - y_{pic_c})^2}} \quad (4)$$

$$\alpha_{NRate}(j) = \frac{\alpha_{num}(j)}{N_{all}} \quad (5)$$

$$\alpha_M(j) = \frac{\|MV(j)\|}{\|MV_{all}\|} \quad (6)$$

式(3)中,  $\alpha(j)$  为通过计算后得到的第  $j$  个独立运动对象的运动显著性系数;  $\alpha_L(j)$  为第  $j$  个运动对象联通区域中心到图像中心的距离的倒数; 式(4)中的  $(x_c(j), y_c(j))$  为第  $j$  个运动对象的联通区域的中心位置坐标;  $(x_{pic_c}, y_{pic_c})$  为图像的中心点的坐标; 式(5)中  $\alpha_{num}(j)$  为第  $j$  个运动对象联通区域所占的图像像素点的个数;  $N_{all}$  为一帧图像中像素点的总数;  $\alpha_{NRate}(j)$  为第  $j$  个运动对象像素点个数占总像素点个数的比率; 式(6)中  $\alpha_M(j)$  为第  $j$  个运动对象的运动显著系数,  $\|MV(j)\|$  为第  $j$  个运动对象所占宏块的运动矢量的模的均值,  $\|MV_{all}\|$  为当前这一帧图像中所有宏块的运动矢量模的均值。

(3) 得到各运动对象的显著性系数后, 对这些系数进行排序, 找出显著性系数最大的运动联通区域, 并确定此运动区域所对应的运动对象为最终需精确分割的关注对象。

(4) 根据步骤(3)得到的关注运动对象, 需通过步骤(4)和(5)进一步确定运动对象区域所占的平滑分块图区域。先统计出分块图像中第  $i$  个区域所占图像像素点的个数  $N_{all}(i)$  和显著联通区域所占第  $i$  个区域的像素点个数  $N_{object}(i)$ 。

(5) 利用第 2.3 节中得到的平滑分块图和最显著运动对象联通区域图及由步骤(4)得到的两个统计系数值  $N_{all}(i)$  和  $N_{object}(i)$  来精确提取出显著运动对象, 实现最终的运动对象分割。为此, 定义式(7)中的  $Obj_{Reg}(i)$  来进一步标示宏块是否为分割后运动对象所占的分块区域。式中的  $\varepsilon_1$  和  $\varepsilon_2$  为区分是否为运动对象所占区域的阈值。

$$Obj_{Reg}(i) = \begin{cases} 0, & \text{若 } N_{all}(i) > \varepsilon_1 N_{object}(i) \\ \text{且 } N_{object}(i) < \varepsilon_2 \sum_i N_{object}(i) \\ 1, & \text{其他} \end{cases} \quad (7)$$

整个运动对象分割过程可由图 2 所示, 图 2(a) 为经 mean-shift 算法后平滑的图像, 每个方块中的像素点被平滑成相同的值。图 2(b) 中的阴影框为由运动矢量得到的多个独立连通运动区域, 反映了

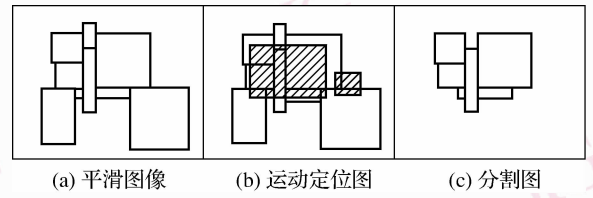


图 2 运动对象分割

Fig. 2 Video object segmentation

运动块所坐落的经平滑后图像位置。在将式(3) ~ 式(6)定义的一系列运动对象显著系数进行筛选后, 确定位于中心且占像素点多的运动联通块为最终关注对象。图 2(c) 为综合考虑了提出的两个分割阈值  $\varepsilon_1$  和  $\varepsilon_2$  后所得到的关注运动对象分割图。

### 3 实验结果讨论

在 JM10.2 编码平台上进行了多组视频序列的实验。下面对其中的一组视频序列 mother and daughter 进行讨论。

以下列出了 3 组实验结果图, 其中图 3 ~ 图 5 的每一组图像中的图(a)为由运动矢量所得到的运动对象区域图; 图(b)为通过均值平滑和区域生长法所得到的分块图; 图(c)为最终的精细运动对象分割图。

当运动矢量图不能完整反映运动对象整体(如图 3(a)), 而图像分块能较正确地反映图像运动对象的整体性(如图 3(b))时, 可以得到完整的运动对象(如图 3(c))。当色度空间分割也并不能完整地把所关注的对象分成一整块, 而运动矢量提取图 4(b)却能够很好体现运动对象(图 4(a))时, 提出的方法也能得到完整的对象(如图 4(c))。当运动矢量并不能很好地反映运动对象整体(如图 5(a)), 而色度分块空间(图 5(b))却很好地弥补了它的不足, 仍旧能较准确地找出完整的运动对象(图 5(c))。

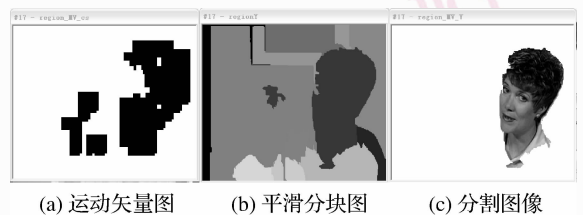


图 3 第 17 帧处理图像

Fig. 3 Processing picture of the 17th frame



(a) 运动矢量图 (b) 平滑分块图 (c) 分割图像

图 4 第 12 帧处理图像

Fig. 4 Processing picture of the 12th frame



(a) 运动矢量图 (b) 平滑分块图 (c) 分割图像

图 5 第 11 帧处理图像

Fig. 5 Processing picture of the 11th frame

## 4 结 论

提出了一种利用压缩域所产生的运动矢量以及图像色度域平滑后的分块图像来分割关注对象的方法。提出的基于压缩域的视频关注对象提取方法能很好地结合运动矢量对象提取及图像色度空间对象提取,使最终的对象分割取两家之长,达到更加精确和完整的分割效果,比之前的各类分割算法,具有更高的精度和更低的算法复杂度。它利用压缩视频流中的运动矢量信息,经过简单的运算就可得到关注

对象的初始位置信息,并结合图像处理中的分割方法,实现精确分割运动对象的目的。

在色度空间的图像平滑处理上,使用了均值偏移的方法,但处理效果仍不够理想。因此,今后的研究重点将放在如何更好地在色度域中实现视频图像的平滑上。

## 参考文献 (References)

- 1 ISO/IEC JTC1/SC29/WG11 N4668, MPEG Video Group Overview of the MPEG-4 standard [S].
- 2 Fan Jian-ping, Ahmed K. An automatic algorithm for semantic object generation and temporal tracking [J]. Signal Processing: Image Communication, 2002, 17(2):145-164.
- 3 Hieu T N, Worring M. Tracking nonparameterized object contours in video [J]. IEEE Transactions on Image Processing, 2002, 11(9):1081-1091.
- 4 Chan Wei-kai, Chien Shao-yi. Real-time memory-efficient video object segmentation in dynamic background with multi-background registration technique [A]. In: Proceedings of IEEE 9th Workshop on Multimedia Signal Processing [C], Crete, Greece, 2007:219-222.
- 5 Mezaris V, Kompatsiaris I, Strintzis M G. Video object segmentation using bayes-based temporal tracking and trajectory-based region merging [J]. IEEE Transactions on circuits and Systems for Video Technology, 2004, 14(6):782-795.
- 6 Toshiki I. Moving object detection using motion vectors in MPEG video data [J]. IPSJ SIG Technical Reports, 2006, 2006(93):167-174.
- 7 Fashing M, Tomasi C. Mean shift is a bound optimization [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(3):471-474.
- 8 Adams R, Bischof L. Seeded region growing [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(6):641-647.