

联合特征累积分析的压缩域运动对象检测*

李 乐^{1,2}, 张茂军¹, 张文琪³, 李永乐¹

(1. 国防科技大学 信息系统与管理学院, 湖南 长沙 410073;
2. 武警工程大学 信息工程系, 陕西 西安 710086;
3. 酒泉卫星发射中心, 甘肃 酒泉 732750)

摘 要:依据 H. 264 压缩域中能够反映景物运动变化的 MV 和 DCT 系数特征, 本文提出了一种多特征联合累积分析的压缩域运动对象检测方法。该方法对压缩码流中各宏块的运动信息进行时空域滤波, 并使用雅克比矩阵描述全局运动参数和宏块 MV 之间的关系, 简化参数求解过程, 通过比较局部运动和全局运动之间的差异初步检测运动对象; 选取宏块周围可靠的运动特征用于宏块 DCT 系数能量的投影累积, 并通过熵能原理在压缩域中选取各个宏块的自适应阈值, 检测运动对象的边缘及纹理显著区域; 通过一定的逻辑准则将 MV 和 DCT 系数的检测结果结合起来, 最终检测出视频中运动对象。实验结果表明, 本文算法可准确地检测压缩视频中的运动对象, 且检测结果具有较高的查全率和查准率。

关键词: H. 264 压缩域; 运动矢量; DCT 系数; 特征累积; 熵能; 运动对象
中图分类号: TP391.4; TN941.1 **文献标志码:** A **文章编号:** 1001-2486(2012)06-0054-07

Associated features accumulation and analysis for motion object detection in compressed domain

LI Le^{1,2}, ZHANG Maojun¹, ZHANG Wenqi³, LI Yongle¹

(1. College of Information Systems and Management, National University of Defense Technology, Changsha 410073, China;
2. Department of Information Engineering, Engineering University of Armed Police Force, Xi'an 710086, China;
3. Satellite Launching Center in Jiuquan, Jiuquan 732750, China)

Abstract: A novel motion object detection method in H. 264 compressed domain is proposed, based on multi-features accumulation and analysis, which can reflect the motion of object and change in the area of borders. Firstly, the MVs of macroblocks were accumulated and filtered. The Jacobian matrix was used to describe the relationship between parameters of global motion and MV in each macroblock, which makes the computation of global motion parameters easier than before. Then the motion areas were detected by the similarity of local motion and global motion. Secondly, the remarkable blocks of DCT energy were accumulated in temporal by the reliable motion around them. Then the border and texture area was found by the accumulations of DCT energy with the local self-adaptive threshold which selected by entropy. Finally, the results which had been detected by MV and DCT respectively were combined. The experiments shows that the method proposed can detect the motion object in compressed video accurately, and the result of can achieve good recall and precision simultaneously.

Key words: H. 264 compressed domain; motion vectors; DCT coefficients; feature accumulation; entropy energy; motion object

视频运动对象检测是智能视频处理的基础问题之一, 为后续的其他应用如对象跟踪、分类、行为理解等提供支持。传统的检测算法主要是依据视频像素域中单个像素的运动信息进行检测的, 虽然结果精确, 但是计算量大、速度慢。目前, 随着 H. 264、MPEG 等视频编码标准的推广, 绝大多数视频都是以压缩格式存储和传输的。如果直接利用视频压缩域中蕴含的运动补偿信息进行检测, 不但可以避免视频全解码和再编码过程中引

起的视频质量退化, 又可以减少计算量, 更适合于实时性要求较高的场合。

压缩域中运动对象的检测方法主要是利用压缩码流中的运动矢量 (Motion Vector, MV) 和离散余弦变换 (Discrete Cosine Transfer, DCT) 系数来实现的。其中, MV 可以看作是压缩域中的光流场, 反映了物体的运动信息; 而 DCT 系数则反映了图像的能量和纹理信息。因此, 压缩域的运动对象检测方法可以分以下几类:

* 收稿日期: 2012-06-06
基金项目: 国家 863 高技术研究发展计划资助项目 (2009AA01Z328); 国家自然科学基金资助项目 (60705013, 60902091)
作者简介: 李乐 (1983—), 男, 河南焦作人, 博士研究生, E-mail: happylee613@gmail.com;
张茂军 (通信作者), 男, 教授, 博士, 博士生导师, E-mail: maojun.z@gmail.com

基于 MV 的检测方法是通过分析视频压缩域中各宏块的 MV 来检测运动对象的。Venkatesh 等^[1]和刘方青等^[2]都采用时域累积等方法估计视频帧对应的运动矢量场,并通过期望最大化(EM)算法对 MV 进行聚类以检测视频中各个运动对象。Niu 等^[3]利用局部各宏块的运动矢量与全局运动矢量之间的差异检测运动对象,并利用帧内预测模式等信息对检测结果进行修正。Cipres 等^[4]则通过模糊逻辑来分析视频的运动矢量场,降低了编码噪声的影响,并通过实时的更新以保证算法的鲁棒性。徐剑锋等^[5]首先对运动矢量进行中值滤波,然后利用最大熵原理计算自适应阈值检测运动对象。由于视频压缩时的运动补偿过程是为了达到最优的压缩效果,导致个别宏块的 MV 不能反映其真实的运动信息,因此单独利用 MV 信息来检测运动对象有时会出现较大的误差。

基于 DCT 系数的检测方法则是通过分析压缩域内 DCT 系数来检测景物边缘和纹理显著区域,从而获取视频中的运动对象。Zeng 等^[6]先建立原始图像对应的 DCT 图像,再利用前景和背景信号的不同分布特征,结合高阶矩的方法对 DCT 图像进行运动检测。Sukmarg 等^[7]在 DCT 域上采用 K 均值算法将不同的块组合成匀质区域,再按照区域的平均变化量区分运动区域或背景区域。在实际应用中,对于视频背景纹理简单和景物运动规则的情况,编码时宏块的运动补偿较准确,各宏块 DCT 系数值较小,此时仅通过分析 DCT 系数难以准确地检测运动对象。

鉴于单独使用 MV 或 DCT 系数检测运动对象存在不足,Eng 等^[8]将二者结合起来检测运动对象,先利用 MV 信息进行投影定位相邻帧内运动对象的疑似区域,再对该区域宏块 DCT 系数采用最大熵失真聚类准则来分割运动对象,该方法需要复杂的建模和迭代运算,计算量大。冯镔等^[9]使用邻帧 DCT 系数之差提取运动对象的残差掩码,并依据 MV 将检测结果后向投影提取对象的运动掩码,最后将二者直接进行简单的合并得到运动对象。Poppe^[10]发现与运动对象相关的宏块由于包含了 MV 和 DCT 信息,在压缩编码后数据量更大,以此为依据可以快速检测运动对象。该方法由于不需要对宏块进行完全解码,速度较快,但是缺乏对压缩域信息的分析,检测结果查准率较低。

1 本文算法

本文提出一种联合 MV 和 DCT 系数特征累

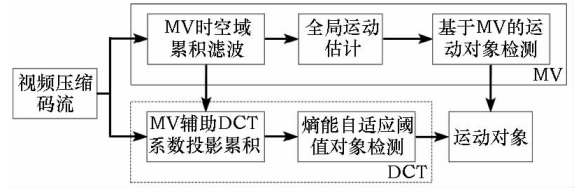


图1 本文算法流程

Fig. 1 The flowchart of our algorithm

积分析的 H. 264 压缩域运动对象检测方法。如图 1 所示,该方法首先对压缩域中各宏块 MV 进行时空域累积滤波,以消除噪声;再使用雅克比矩阵描述全局运动参数和宏块 MV 之间的关系,简化参数求解过程,并通过比较宏块 MV 和全局运动之间的差异初步检测视频运动区域;随后,联合前面检测结果内可靠的运动特征,对 DCT 系数特征进行时域投影累积,消除局部噪声;并将熵能选取自适应阈值原理应用于压缩域中,通过选取合适的局部阈值,检测运动对象的边缘和纹理显著区域,提高算法的鲁棒性;最后通过一定的逻辑准则将基于 MV 和 DCT 系数的检测结果结合起来,得到视频中的运动对象。实验表明,本文算法可以将视频压缩域中 MV 和 DCT 系数特征结合起来,准确提取压缩视频的运动对象,鲁棒性强,检测结果具有较高的查全率和查准率。

2 基于运动特征的对象检测

2.1 MV 累积滤波

视频压缩域中, MV 反映了当前帧内各宏块相对于前一帧的运动信息,以此为依据可以识别视频中的运动对象。但是,由于 H. 264 标准采用的是块匹配的运动估计技术,偏重于码流的压缩效率,而非运动估计的准确性。压缩域中原始的 MV 包含了许多噪声,需要通过对运动矢量场的时空域累积滤波,消除噪声 MV 的干扰。

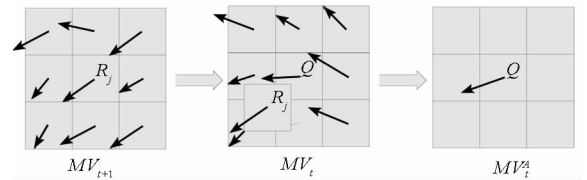


图2 运动矢量时域累积

Fig. 2 The accumulation of MV in temporal domain

在时域上,利用各宏块的 MV 在相邻帧之间进行时域投影累积,强化正确的运动信息。如图 2 所示,以 $t+1$ 帧中宏块 R_j 为例,依据其运动矢量 $MV_{t+1}(R_j)$ 将其投影到 t 帧中,它将与 t 帧中宏块 Q 发生重叠,那么以二者重叠面积的比例为权

重可以将宏块 R_j 的运动矢量叠加到宏块 Q 上,得到 t 帧中宏块 Q 的运动矢量累积结果 $MV_t^A(Q)$ 。

在空域上,对各宏块 MV 进行中值滤波,先选取滑窗内与矢量中间值最接近的若干 MV ,并计算其平均值作为滤波结果对滑窗中央宏块 MV 进行修正,从而去除局部宏块 MV 的毛刺。

2.2 全局运动估计

由于视频的多样性,对于存在背景运动的视频,难以直接从运动矢量场中提取运动对象。这时,通过全局运动补偿,可以消除视频背景运动对检测结果的影响。本文结合压缩视频的特点,提出了一种适用于压缩域的全局运动参数快速求解方法。首先,这里使用 6 参数仿射模型来描述视频全局运动。

$$\begin{bmatrix} x_{t-1} \\ y_{t-1} \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & r_y & u \\ r_x & s_y & v \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} x_t \\ y_t \\ 1 \end{bmatrix} \quad (1)$$

上式描述了从第 t 帧到第 $t-1$ 帧的全局运动模型,其中 s_x, s_y 为缩放因子, r_x, r_y 为旋转因子, u, v 为位移因子。利用上述全局运动模型可以计算第 t 帧到第 $t-1$ 帧视频背景区域在 x 和 y 方向上的运动矢量为

$$\begin{aligned} MV_x &= x_{t-1} - x_t = (s_x - 1)x_t + r_y y_t + u \\ MV_y &= y_{t-1} - y_t = r_x x_t + (s_y - 1)y_t + v \end{aligned} \quad (2)$$

此时,令相邻宏块的 MV 作差,可以发现全局运动参数和 MV 之间的关系。

$$MV_x(x_t, y_t) - MV_x(x_t - 1, y_t) = s_x - 1 \quad (3)$$

同理,通过背景区域内 MV 的雅克比矩阵可以描述其与全局运动参数之间的关系:

$$J(MV) = \begin{vmatrix} \frac{\partial(MV_x)}{\partial(x_t)} & \frac{\partial(MV_x)}{\partial(y_t)} \\ \frac{\partial(MV_y)}{\partial(x_t)} & \frac{\partial(MV_y)}{\partial(y_t)} \end{vmatrix} = \begin{vmatrix} s_x - 1 & r_y \\ r_x & s_y - 1 \end{vmatrix} \quad (4)$$

因此,通过考察背景区域内各个宏块 MV 分别在 x 和 y 方向上的变化,可以求解全局运动模型中的旋转、缩放因子,然后再结合式(2)求解位移因子。本文通过 RANSAC 算法迭代求解,可以有效地去除当前帧中不属于背景区域的宏块对全局运动参数估计的影响,经过 3~5 次迭代运算,模型将收敛,完成全局运动参数估计过程。

2.3 基于运动相似性的对象检测

在获得稳定可靠的运动矢量场之后,本文使用一种兼顾方向性和大小的矢量相似性度量方法^[11]来比较局部运动矢量和全局运动矢量之间

的差异,以提取运动对象。在比较之前,将 MV 的角度映射到 $[0^\circ, 90^\circ]$,相似性度量公式如下:

$$dist(MV, GMV) = \frac{MV \cdot GMV}{\max(\|MV\|^2, \|GMV\|^2)} \quad (5)$$

通过上式考察局部宏块运动矢量 MV 和全局运动矢量 GMV 之间的差异,可以检测视频中的运动区域,如图 3 所示。此外,由于目标局部运动的不连续性,可能会导致检测结果出现误差,还需要通过形态学的方法(如目标内部空洞的扫描填补)进行优化。

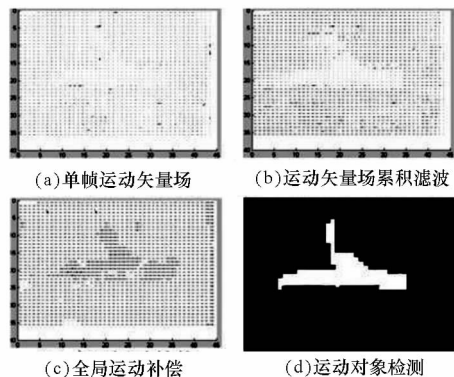


图 3 基于运动特征分析的对象检测

Fig. 3 Motion detection by MV analysis

3 联合运动特征的 DCT 系数累积分析对象检测

视频中景物的边缘和纹理显著区域在发生不规则运动时,如果使用统一的 MV 来描述宏块内所有像素的运动,则会出现较大的误差,产生较大的 DCT 系数。此时需要通过分析 DCT 系数特征来检测运动对象。

3.1 联合运动信息累积 DCT 系数

DCT 系数是由运动补偿不准确产生的残差经过变换和量化得到的。其分布具有一定的随机性,不但每帧视频内各宏块 DCT 系数的分布比较稀疏,而且时域上相关宏块的 DCT 系数值也不连续,难以直接用于检测运动对象。因此,本文参考 MV 时域投影累积原理,提出了联合 DCT 系数显著区域周围可靠的运动信息对其进行投影累积的方法,以获取稳定可靠的 DCT 系数特征。

$$E = \sum_{i=1}^{4 \times 4} w(i) \times (DCT(i))^2 \quad (6)$$

H. 264 编码过程是以 4×4 像素块为单位进行 DCT 变换的,本文依据上面的公式来量化各像素块的 DCT 系数能量 E 。其中, $w(i)$ 为像素块内各位置 DCT 系数的权重,权重大小与其在像素块内的 Zig-Zag 扫描顺序有关。由于图像从时域到

频域转换后,能量集中于低频部分,因此低频位置的权重较大,高频位置的权重较小。

由于 DCT 系数能量较大的块一般分布于运动对象的边缘和纹理显著的区域,而与其相邻的运动对象内部区域则主要由运动补偿较准确、DCT 系数能量较低的宏块组成。因此,依据前面基于运动特征的对象检测结果,利用 DCT 系数显著区域周围属于运动对象内部的可靠的 MV 信息对其进行投影累积,具体过程如下:

(1) 根据式(6)计算当前帧内各宏块的 DCT 系数能量,选取 DCT 系数能量大于阈值 T 的宏块作为 DCT 显著宏块 Q ;

(2) 依据前面基于 MV 的运动对象检测结果,以选中的 DCT 显著宏块 Q 为中心,查找半径为 r 的邻域内所有属于已检测出的运动对象的宏块 R_j ;

(3) 由于宏块 MV 的可信度(即运动补偿的准确性)和 DCT 系数能量之间成反比例关系,那么以所有属于运动对象宏块 R_j 的 DCT 系数能量 E_j 的倒数为权值,加权平均计算该 DCT 显著宏块 Q 对应的投影累积运动矢量 $MV_{DCT}(Q)$ 。

$$MV_{DCT}(Q) = \sum_j MV(R_j) \times \frac{1/E_j}{\sum_j 1/E_j} \quad (7)$$

(4) 根据 MV 投影累积原理,联合 $MV_{DCT}(Q)$ 对宏块 Q 进行时域前向投影,搜索其在上一帧中匹配位置,并对宏块 Q 的 DCT 系数能量进行累积,见图 4。

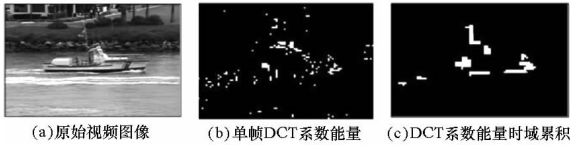


图 4 联合运动特征的 DCT 系数能量累积
Fig. 4 The accumulation of DCT with motion feature

3.2 熵能自适应阈值的对象检测

在检测运动对象的过程中,由于视频内各宏块的噪声分布是不均匀的,采取单一的阈值进行检测可能会出现较大的误差,因此本文将熵能选取自适应阈值原理应用到压缩域中,降低噪声干扰,提高检测精度。

根据香农定理,一个离散信号 O 的熵 H 定义为

$$H = - \sum p_i \log p_i \quad (8)$$

文献[12]认为,可以将图像 O 看作是有用信号 X 和高斯白噪声 G 的叠加(即 $O = X + G$),假定噪声的能量很低但分布于整个图像,而有用信号

的幅度虽大但分布范围很小,这样就符合熵能的特点。根据香农对信息熵的定义,依据观察信号 O 的熵 H 估算信息熵的标准差 σ 。

$$\sigma = \frac{\exp(H)}{\sqrt{2\pi e}} \quad (9)$$

此时,认为阈值 θ 与标准差呈 σ 正比关系,即

$$\theta = k \cdot \sigma \quad (10)$$

在进行压缩编码时,视频中背景宏块都应得到较好的运动补偿,即 DCT 系数基本为 0;而对于运动对象的边缘和纹理区域,则由于运动补偿的不准确而产生较大的 DCT 系数。因此,可以将每帧压缩视频的 DCT 系数看作是属于运动对象的有效信号与背景噪声信号之和,符合文献[12]中熵能原理应用的假设。当视频场景中没有运动目标时,各 4×4 像素块的 DCT 系数能量 $E(Q)$ 应小于阈值 θ ,根据式(10)可推出 $k > E(Q)/\sigma$ 。依据该原理,可利用背景图像序列对每个 4×4 像素块的 k 进行训练,选取每个块的自适应阈值。其方法如下:

(1) 统计第 i 帧内所有 4×4 像素块的 DCT 系数能量直方图,并依据式(8)计算该帧的熵 H_i 。

(2) 按照式(9)计算第 i 帧视频的信息熵标准差 σ_i 。

(3) 根据第 i 帧内各 4×4 像素块 Q 的能量 $E(Q)$,依据式(10)计算其对应的训练参数 $k_i(Q) = E(Q)/\sigma_i$ 。

按照以上步骤,经过 f 帧的背景图像训练后,对每个 4×4 像素块都可得到一个 k 值序列 $(k_1, k_2, \dots, k_f)$ 。根据噪声的分布特点,可将 k 值序列看作一个高斯分布,初始分布参数 (μ_0, δ_0) 可通过似然估计法得到,即变量 $\eta = (k - \mu_0)/\delta_0$ 满足标准正态分布。那么,可设定像素块的初始阈值参数为 $k_0 = \mu_0 + \eta\delta_0$,根据正态分布理论, η 取 1.3 时样本落于分布内的概率大于 90%,实际应用中, η 的值在 $[2, 3]$ 间选取。那么,对后续各帧视频可以采用如下的步骤进行对象检测:

(1) 计算压缩码流中第 j 帧 DCT 系数能量的信息熵标准差 σ_j 。

(2) 当 4×4 像素块 Q 的 DCT 系数能量 $E(Q) > k(Q) \times \sigma_j$ 时,标记该块为运动对象。

(3) 计算当前第 j 帧中每个 4×4 像素块 Q 对应的阈值参数 $k_j(Q) = E(Q)/\sigma_j$ 。

(4) 对满足 $|k_j - \mu_j| \leq \eta\delta_j$ 的块,依据下面的公式更新对应的高斯分布参数 (μ, δ) 和阈值参数 k 。

$$\begin{aligned}\mu_{j+1} &= (1 - \alpha)\mu_j + \alpha E(Q) \\ \delta_{j+1}^2 &= (1 - \alpha)\delta_j^2 + \alpha(E(Q) - \mu_j)^2 \\ k_{j+1} &= \mu_{j+1} + \eta\delta_{j+1}\end{aligned}\quad (11)$$

3.3 联合特征的对象检测

基于 DCT 系数可以检测出对象的边缘和纹理显著区域,而基于 MV 可以得到对象内部可靠的运动区域。因此,将基于 MV 和 DCT 系数的检测结果相结合,可以较完整地检测视频中的运动对象。设第 n 帧基于 MV 的检测结果为 R_n^{MV} ,基于 DCT 系数的检测结果为 R_n^{DCT} , $N(B_n(i,j))$ 为 4×4 像素块 $B_n(i,j)$ 的 4 邻域, $B_{ref}(i,j)$ 则为 4×4 块 $B_n(i,j)$ 在前一帧中对应的像素块, O 为运动对象。二者结合的逻辑准则如下:

(1) 对于 MV 检测结果 R_n^{MV} 中的块 $B(i,j)$,若周围相邻的块 $N(B_n(i,j))$ 及其在前一帧中对应的块 $B_{ref}(i,j)$ 都属于运动对象时,则认为该块位于可靠的运动对象区域之中。即

$$\begin{aligned}\{B_n(i,j) \in R_n^{MV} \wedge B_{ref}(i,j) \in R_{n-1}^{MV} \wedge \\ \{ \forall B_n(m,n) \in N(B_n(i,j)) \rightarrow B_n(m,n) \in R_n^{MV} \} \} \\ \Rightarrow B_n(i,j) \in O\end{aligned}$$

(2) 对于 MV 检测结果 R_n^{MV} 中的块 $B(i,j)$,若其在前一帧的对应块 $B_{ref}(i,j)$ 属于运动对象且其周围相邻块 $N(B_n(i,j))$ 中还存在属于 DCT 系数检测结果 R_n^{DCT} 的像素块,则可认为该块属于运动对象,且位于运动对象内部靠近边缘的区域。即

$$\begin{aligned}\{B_n(i,j) \in R_n^{MV} \wedge B_{ref}(i,j) \in R_{n-1}^{MV} \wedge \\ \{ \exists B_n(m,n) \in N(B_n(i,j)) \rightarrow B_n(m,n) \in R_n^{DCT} \} \} \\ \Rightarrow B_n(i,j) \in O\end{aligned}$$

(3) 对于 DCT 系数检测结果 R_n^{DCT} 中的块 $B(i,j)$,如果其相邻块 $N(B_n(i,j))$ 中存在运动块,且这些运动块在前一帧中所对应的也为运动块时,说明该块与可靠的运动对象区域相邻,应属于运动对象的边缘区域。即

$$\begin{aligned}\{B_n(i,j) \in R_n^{DCT} \wedge \{ \exists B_n(m,n) \in N(B_n(i,j)) \rightarrow \\ \{B_n(m,n) \in R_n^{MV} \wedge B_{ref}(m,n) \in R_{n-1}^{MV} \} \} \} \\ \Rightarrow B_n(i,j) \in O\end{aligned}$$

4 实验

本文选取 PetsD2TeC2、Coastguard 等几段 YUV 4:2:2、CIF 格式的标准测试视频序列进行实验,实验中使用 JM11.0 版本的 H.264 编码器对实验视频进行编码,帧结构是 IPPPP,帧率为 30fps,参考帧数为 1 帧,量化步长 $QP = 28$ 。三段实验视频各有特点,对算法提出了不同的要求,其中 PetsD2TeC2 是静态背景的视频,由于该视频背景中局部区域存在树枝摇摆的现象,要求检测算法能够屏蔽局部噪声的干扰;Coastguard 视频中使用摄像机对景物进行跟踪拍摄,具有明显的全局运动,需要通过准确的全局运动补偿来解决视频背景运动对景物检测的影响;Bus 视频场景的景物构成比较复杂,而且存在全局运动,导致编码时运动补偿的准确性较差,码流中包含了大量 DCT 系数显著宏块,需要将 MV 和 DCT 系数相结合用于检测运动对象。此外,本文使用查全率和查准率两个指标来衡量算法的性能。查全率是指检测结果中属于真实运动对象的区域占整帧视频中实际运动对象区域的百分比;查准率是指检测结果中属于真实运动对象的区域占检测结果的百分比。查全率越高,表示运动对象的检测结果越完整,即漏检较少;而查准率越高,则表示检测的伪运动区域越少,即误检较少。为了统计方便,本文以 4×4 块为单位来统计算法的查全率和查准率。

仅使用运动特征、使用固定阈值的 MV 和 DCT 系数相结合的方法以及使用 MV、DCT 系数和自适应阈值的方法分别对实验视频进行运动对象检测的结果如图 5 所示。图中第一至第三行

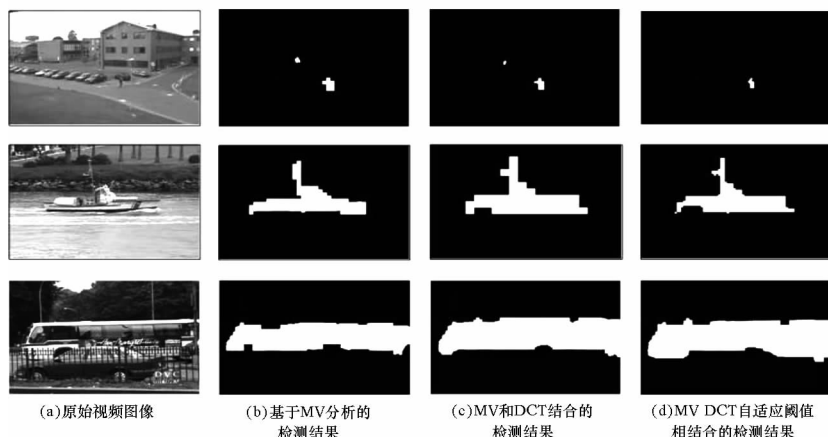


图 5 基于 MV、DCT 系数和自适应阈值算法效果比较

Fig. 5 The comparison of motion detection by MV, DCT and self-threshold

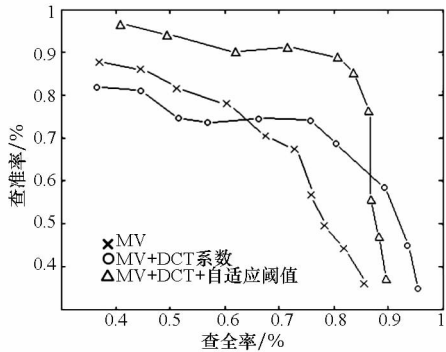


图 6 基于 MV、DCT 系数和自适应阈值算法
查全率、查准率比较

Fig. 6 The comparison of recall and precision which
is detected by MV, DCT and self-threshold

分别为 PetsD2TeC2、Coastguard、Bus 三段视频的处理结果。其中,第一列为原始视频图像,第二列为单独使用 MV 的检测结果,第三列为固定阈值条件下 MV 和 DCT 系数相结合的检测结果,第四列为联合 MV 和 DCT 系数特征并使用熵能选取自适应阈值的检测结果。从图 5 可以看出,单独

依靠 MV 检测运动对象还存在较大的不足,很多边缘和纹理显著区域出现漏检和误检的现象;将 MV 与 DCT 系数相结合之后,这些纹理显著区域基本都被检测出来了,但是使用固定阈值进行运动对象判决,算法容易受到噪声的干扰,出现误检的情况;而在此基础上使用熵能原理以块为单位选取自适应阈值判别运动对象,可以有效地屏蔽局部噪声的干扰。图 6 是图 5 所示三种检测结果之间“查全率—查准率”曲线的比较,可以看出单独基于 MV 检测运动对象的查全率和查准率都相对较低;而将 MV 和 DCT 系数结合之后,运动对象检测的查全率有所提高,但是由于噪声宏块的干扰,有时查准率有所下降;在使用自适应阈值判别运动对象之后,可有效地屏蔽局部宏块的噪声,使得检测结果同时具有较高的查全率和查准率。因此,通过前面的比较可以发现,使用 MV、DCT 系数特征以及熵能自适应阈值相结合的方法检测运动对象,比单独使用某一类特征进行检测的效果更好,具有明显的优势。

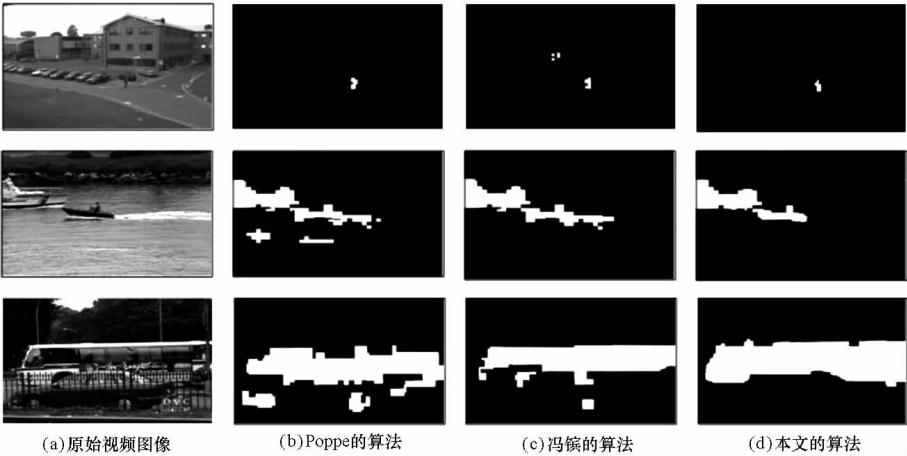


图 7 压缩域运动对象检测结果比较
Fig. 7 The comparison of motion detection

使用本文算法与冯镔^[9]和 Poppe^[10]的算法分别对 3 段实验视频进行运动对象检测,结果如图 7 所示。从图中可以看出,使用本文的算法得到的检测结果在直观效果上要优于 Poppe 和冯镔的算法。这是因为,Poppe 的方法仅通过宏块的数据量来判断运动对象,容易受到噪声的干扰,尤其是视频的全局运动会对检测结果产生较大影响。冯镔的算法使用 MV 和 DCT 系数差进行检测,基本可以准确地检测视频中的运动对象,但是当视频中包含大量噪声时,无法有效地剔除噪声宏块的干扰,鲁棒性较差。而本文的算法得到的检测结果不但能较完整地检测出视频中的运动对象,而且使用自适应阈值进行前景判决可以有效

屏蔽局部噪声干扰,提高算法的鲁棒性。

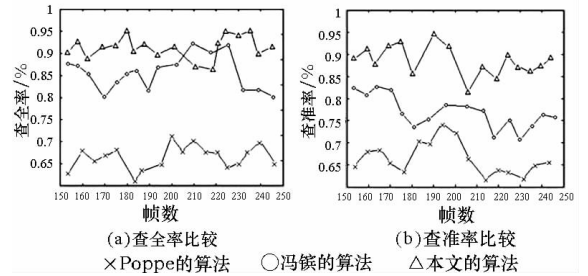


图 8 压缩域运动对象检测算法查全率与查准率比较
Fig. 8 The comparison of recall and precision

使用本文算法与冯镔^[9]和 Poppe^[10]的算法分别对 Bus 视频的 150 帧到 250 帧之间的图像序列进行运动对象检测,并统计比较各个算法对应

的查全率和查准率,如图 8 所示。由于 Poppe 仅使用编码后宏块的数据量进行检测,对于具有全局运动或者背景区域包含明显噪声的情况,无法准确地区分运动对象和背景区域,导致检测结果的查全率和查准率都比较低。而冯镔在算法中使用 DCT 系数差和 MV 进行检测,二者互补使得算法的查全率有了较大提升;其中,邻帧 DCT 系数作差可以去除恒定噪声的干扰,但也会弱化边缘和纹理显著区域的 DCT 系数特征,从而影响检测结果的查全率;同时,使用单一阈值进行景物判决,使得算法容易受到局部突发噪声的干扰,影响检测结果的准确性,降低算法的查准率。相比之下,本文的算法将 DCT 系数和 MV 多特征相结合用于检测运动对象,提高了检测结果的查全率;并通过对 MV 和 DCT 进行时域投影累积,强化了有用的 DCT 系数和 MV 特征,以及使用熵能原理选取自适应阈值,增强了算法的抗干扰能力,提高运动对象检测结果的查准率。

通过上面的实验对比与分析,可以发现用本文算法检测运动对象具有更高的查全率和查准率,检测结果更加可靠。与其他算法相比,本文的算法具有以下几点优势:(1)充分利用视频压缩域中蕴含的 MV 和 DCT 系数特征,二者互补共同用于对象检测;(2)通过分析压缩码流中 DCT 系数产生的原因,选取其周围可靠的运动特征辅助宏块 DCT 系数能量进行投影累积,强化了有用的 DCT 系数特征,有助于运动对象边缘区域的检测;(3)将熵能自适应阈值选取原理应用于压缩视频中,有效地屏蔽了视频压缩域中局部宏块噪声的干扰。

5 结论

本文提出了基于多种特征联合累积分析的压缩域运动对象检测方法。该方法主要利用视频时空域信息分别对各宏块的 MV 和 DCT 系数进行投影累积,消除局部宏块噪声。并将基于熵能自适应阈值选取原理应用到压缩域中,为每个宏块选取合适阈值,增强了算法的抗干扰能力。最后依据一定的逻辑准则将 MV 和 DCT 系数的检测结果进行合并。实验表明,使用本文的算法得到的运动对象检测结果同时具有较高的查全率和查准率。但是,目前算法中只是依据简单的逻辑将 MV 和 DCT 系数检测结果结合起来,对于某些特殊的情况可能无法得到满意的结果。

参考文献 (References)

[1] Venkatesh B R, Ramakrishnan K R, Srinivasan S H. Video

- object segmentation: a compressed domain approach[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2004, 14(4): 462-474.
- [2] 刘方青,石旭利,张兆扬. 基于 EM 聚类的 H. 264 压缩域视频对象实时分割算法[J]. 中国图像图形学报, 2007, 12(10): 1819-1822.
- LIU Fangqing, SHI Xuli, ZHANG Zhaoyang. Real-time video object segmentation in the compressed domain based on EM clustering[J]. Journal of Image and Graphics, 2007, 12(10): 1819-1822. (in Chinese)
- [3] NIU C F, LIU Y S. Moving object segmentation based on video coding information in H. 264 compressed domain[J]. Image and Signal Processing, 2009, 1-5.
- [4] Solana-Cipres C, Fernandez-Escribano G, et al. Real-time moving object segmentation in H. 264 compressed domain based on approximate reasoning [J]. International Journal of Approximate Reasoning, 2009, 51: 99-114.
- [5] 徐剑锋,刘志,张兆杨. 基于熵模型的压缩域运动对象检测[J]. 中国图像图形学报, 2007, 12(10): 1815-1818.
- XU Jianfeng, LIU Zhi, ZHANG Zhaoyang. Compressed domain moving object detection based entropy model [J]. Journal of Image and Graphics, 2007, 12(10): 1815-1818. (in Chinese)
- [6] Zeng W, Gao W, Zhao D. Automatic moving object extraction in MPEG video [C]//Proceedings of the IEEE International Symposium on Circuits and Systems 2003, Bangkok, Thailand, 2003, 11: 524-527.
- [7] Sukmarg O, Rao K R. Fast object detection and segmentation in MPEG compressed domain [C]//Proceeding of IEEE TENCON. 2000, 3: 364-368.
- [8] Eng H L, Ma K K. Spatiotemporal segmentation of moving video objects over MPEG compressed domain [C]//Proceedings of the IEEE International Conference on Multimedia and Expro 2000. New York, 2000, 3: 1531-1534.
- [9] 冯镔,肖非,朱光喜,等. 一种基于 H. 264/AVC 的压缩域运动对象分割方法[J]. 中国图像图形学报, 2009, 14(7): 1327-1333.
- FENG Bin, XIAO Fei, ZHU Guangxi, et al. Moving object segmentation in compressed domain for H. 264/AVC [J]. Journal of Image and Graphics, 2009, 14(7): 1327-1333. (in Chinese)
- [10] Poppe C, De Bruyne S, et al. Moving object detection in the H. 264/AVC compressed domain for video surveillance applications[J]. Journal of Visual Communication and Image Representation, 2009, 20(6): 428-427.
- [11] Yeo C, Ahammad P, Ramchandran K, et al. High-speed action recognition and localization in compressed domain video [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2008, 18(8): 1006-1015.
- [12] Luthona F, Liévin M, Faux F. On the use of entropy power for threshold selection [J]. Signal Processing. 2004, 84(10): 1789-1804.
- [13] Mezaris V, Kompatsiaris I, Boulgouris N V. Real-time compressed-domain spatiotemporal segmentation and ontologies for video indexing and retrieval [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2004, 14(5): 606-621.