

视频编码参数对目标识别性能影响的研究

吴泽民 刘涛* 姜青竹 胡磊

(中国人民解放军理工大学通信工程学院 南京 210007)

摘要: 国内外研究人员对图像目标分类识别和视频编码传输问题都分别进行了大量研究,但是对于视频编码参数对目标识别性能影响的定量关系,还没有公开的文献报导。针对这一问题,该文选择典型的目标识别算法可变部件模型(DPM)和最常用的视频编码方法 H.264/AVC 作用测试对象,通过设计的编码和检测实验,研究了码率和分辨率参数对视频目标识别性能的影响,并拟合了识别性能随码率和分辨率变化的函数关系。通过选取编码器合适的码率和分辨率工作参数,可以获得信道带宽与视频目标识别性能的折中,为设计不同视频应用的编码优化目标函数提供了依据。

关键词: 计算机视觉; 目标识别; 视频编码; 码率; 分辨率

中图分类号: TP391.4

文献标识码: A

文章编号: 1009-5896(2015)08-1906-07

DOI: 10.11999/JEIT141613

Video Coding Parameters Effect on Object Recognition

Wu Ze-min Liu Tao Jiang Qing-zhu Hu Lei

(College of Communications Engineering, PLA University of Science and Technology, Nanjing 210007, China)

Abstract: Researchers have done a great number of studies on the object recognition and the video coding transmission respectively. However, there are still no public reports about the influence on the object recognition raised by the video encoding parameters. For this issue, the Deformable Part Model (DPM), a typical object recognition algorithm and the most commonly-used video encoding methods-H.264/AVC are chosen as the test objects. In order to study how the code rates and the resolution affect the performance of video object recognition, the coding and detection experiments are designed and the function of recognition performance changes caused by the code rates and the resolution is fitted. The result shows that the compromise can be achieved between the channel bandwidth and the video object recognition performance through selecting the appropriate the code rates and the resolution parameters for the encoder which provides basis for encoding optimization object function of different video applications.

Key words: Computer vision; Object recognition; Video code; Code rates; Resolution

1 引言

目标识别是计算机视觉的热门领域,在军事和民用领域都受到越来越多的重视,得到极大发展和应用。视频目标识别是其重要应用之一。但是在网络化应用环境中,由于信道实际带宽的限制,视频必须经过压缩编码再传输。不同的视频编码参数形成的接收端重构数据,它对目标识别性能的影响还不太清楚。这也是本文的中心研究问题。

在战场、抢险等低传输带宽的应用环境中,视频码率是关键的性能参数,它与分辨率、感知失真、量化步长等参数交互作用,构成系统最优化问题。

当考虑视频目标识别这样的应用时,应用性能指标与视频编码参数的关系变得更加复杂。需要寻找应用性能指标与编码参数的函数关系,从而在原有的优化框架中实现视频应用的编码目标。本文考虑实验参数的可选择性和编解码工具的可控性,重点分析了目标识别性能与编码速率与分辨率的关系。第2节回顾相关工作,并介绍目标识别算法和视频编码器;在第3节和第4节,分别试验和总结了码率、分辨率对目标识别性能的影响;在第5节,完成了码率和分辨率对目标识别性能的联合影响曲线的实验和拟合。

2 目标识别和视频编码器

目前,图像目标识别技术主要采用词包^[1,2](Bag Of Feature, BOF)、方向梯度直方图^[3,4](Histogram

2014-12-18 收到, 2015-01-22 改回, 2015-05-11 网络优先出版

航空科学基金(18265)资助课题

*通信作者: 刘涛 ltaoliu_tao@foxmail.com

of Oriented Gradient, HOG)和可变部件模型^[5](Deformable Part Model, DPM)3种方式,它们在视频目标识别^[6,7]和视频人体目标识别^[8]得到了广泛的应用,并实现了无监督方法对视频目标识别^[9,10];视频编码参数对视频质量的影响^[11,12]也得到研究人员的关注。但是,对于视频编码参数对目标识别性能影响的定量关系,还没有公开的文献报导。

2.1 目标识别算法

词包模型基于中层特征对场景语义建模,它无需分析场景图像中的具体目标组成,而是应用图像场景的整体统计信息,将量化后的图像低层特征视为单词,通过图像的单词分布来表达图像场景内容^[1]。方向梯度直方图模型使用 HOG 特征来表达人体,提取人体的外形信息和运动信息,形成丰富的特征集^[3]。HOG 的主要思想为在一幅图像中,局部目标的表象和形状能够被梯度或边缘的方向密度分布很好地描述。

可变部件模型为这几年最为流行的图像目标检测算法。文献[13]为了在特征描述阶段定义物体形变而提出了 DPM 模型。DPM 模型使用星型结构的部件模型,此模型由一个根滤波器、一系列部件滤波器以及相应的可变形模型构成。根滤波器包含目标的整体信息,而部件滤波器采用高分辨率的细节建模。星型模型在图像特定位置和尺度的得分,等于根滤波器的得分加上各个部件的得分的总和。每个部件的得分等于此部件在所有空间位置的得分的最大值,而部件在某位置的得分等于部件滤波器在此位置的得分减去此位置的变形代价。变形代价衡量了部件偏离其理想位置的程度,即部件偏离与根滤波器的最优相对位置程度。

通过定位每个部件和定量部件之间的相对位置关系,DPM 模型容许目标出现较大程度的外观形变。从而使得 DPM 模型可以识别各种多变的目標,大大提高了检测正确率。文献[5]利用 DPM 进行图像目标识别,并利用级联技术加速目标识别^[14]。为了提高目标识别精度,文献[15]将语法模型添加到 DPM 目标识别算法中。模型使用判别分类器完成目标识别,有监督的训练过程需要用到图片集中标注的目标矩形框。训练完成的分类器既高效又精确,能够在 PASCAL VOC 2006, 2007, 2008 等测试集上达到目前最佳结果^[5]。故本文选 DPM 作为识别算法进行实验。

2.2 视频编码器

H.264/AVC 是由联合视频组(JVT)所开发的最新的视频编码技术国际标准,被国际电信联盟标准化组织(ITU-T)的视频编码专家组(VCEG)和国际

标准化组织(ISO/IEC)的运动图像专家组(MPEG)共同认可。它以计算复杂度增加和编码结构复杂为代价,取得了比其他标准更高的编码效率^[16]。因 H.264/AVC 具有压缩视频图像清晰、压缩比率大、低码率易于传输等优点,使其成为被广泛使用的视频编码标准。故本文选 H.264/AVC 作为编码工具进行实验。

本文中使用的视频编码器是 X.264,是由法国巴黎中心学校的中心研究所于2004年6月发起的项目,由许多视频爱好者共同完成的项目。X.264 注重实效性。在不明显降低编码性能的前提下,降低编码的复杂度^[17],同时它在码率控制算法中具有灵活性、实时性和高效性的优点。而且 X.264 编码器控制简单,可以根据预设的编码速率,自动生成指定速率的视频码流。故本文选 X.264 作为视频编码器进行实验。

3 码率对识别性能影响

码率是视频编码最常见最重要的参数之一,表示视频流在单位时间内产生的数据流量,其单位通常为 kbps(千比特每秒)。码率控制就是有效地控制视频编码器,使其输出码流的速率稳定在一定范围,满足传输信道实际带宽的限制。通过试验的方式,统计出感兴趣物理量之间的关系,在视频分析和应用是普遍的方法^[18,19]。本节通过码率控制,测试码率参数与目标识别性能的关系。

3.1 码率对识别性能影响实验方法

本文实验使用 VIRAT 视频数据库,它由美国国防高级研究计划局(DARPA)下属信息处理技术办公室(IPTO)资助建立。VIRAT 视频数据库主要提供目标为街道行人、道路车辆及停车场等的监控视频,还包括目标为坦克、汽车等的无人机航拍视频,视频分辨率大小为 1280×720。数据库中同时还包含各个视频的手工标注信息,包括目标的种类和位置。如果目标为人,则还包括目标行为类别。本实验使用该数据库的街道行人和车的监控视频及其位置标注信息,通过测试行人和车的检测成功率,建立视频编码参数与检测器性能的关系。

本文实验对 DPM 模型的训练分别使用 Pascal VOC 2012 和 INRIA Datasets 两个图像数据集。Pascal VOC 数据集有 10000 幅图像,包含有人、汽车等 20 个种类的目标,并给出了目标的种类和位置人工标注,本文使用该图像集分别训练出人和汽车的 DPM 模型。INRIA Datasets 含有人、汽车和马的图像集,本文使用其人和汽车图像集分别训练出人和汽车的 DPM 模型。本文分别使用 Pascal VOC

和 INRIA Datasets 图像集训练出的 DPM 模型检测 VIRIA 视频中的行人和车。

实验分为两部分,如图 1 所示。第 1 部分直接使用 DPM 算法对原 YUV 视频进行目标识别,将识别结果作为对比的基准。第 2 部分,在其它编码参数相同的条件下,对原视频采用 H.264/AVC 编码器进行不同的码率下的压缩编码,使得生成的码率能适应不同的传输信道带宽。然后使用 H.264 解码器完成对视频帧的重建,在重建后的图像帧中使用 DPM 算法进行目标识别的测试,用平均正确率 (Average Precision, AP) 指标值衡量目标识别性能。本实验中,视频编码速率 R 取值从 50 kbps 开始,直到 1000 kbps 结束,每次调整编码速率时步进 50 kbps。

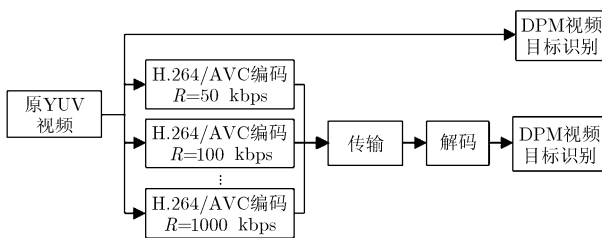


图 1 码率对识别性能影响的实验流程

3.2 码率对识别性能影响实验结果

通过实验,我们得到原始 YUV 视频直接识别的性能指标 $AP_0 = 0.73$ 。在不同编码码率下,DPM 模型的识别性能如图 2 所示。从图中可以看到,随着编码码率的增大,识别性能总体为上升趋势,在接近 AP_0 后识别性能趋于稳定。经过观察,对实验数据进行负指数和多项式函数拟合,如表 1 所示。对比得到负指数函数拟合式(1)拟合效果更好,其拟合曲线如图 2 所示。

$$AP = AP_0 - 0.49e^{-0.005125R} \quad (1)$$

3.3 编码速率对识别性能影响的实验结果分析

由式(1)和图 2 得到,当码率在区间 50~300 kbps 时,随着码率的上升,识别性能指标 AP 值上

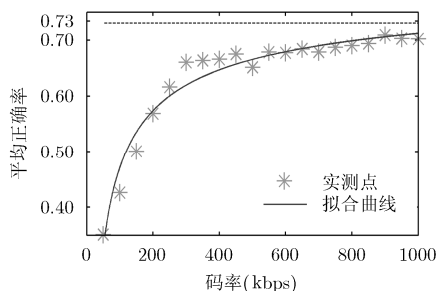


图 2 不同视频编码码率的解码后识别性能

表 1 不同编码码率对识别性能影响的函数拟合结果

	拟合结果	相关系数	标准差
负指数函数拟合	$AP = AP_0 - 0.49 \cdot e^{-0.005125R}$	0.9735	0.0256
多项式函数拟合	$AP = AP_0 - 8.77R^{-0.776}$	0.9686	0.0270

升较快;当码率在区间 300~700 kbps 时,随着码率的上升,AP 值上升缓慢;当码率超过 700 kbps 后,随着码率的上升,AP 值趋于稳定在 0.71。

对于本文使用的监控视频,在实验中使用的最大码率 1000 kbps 时,由 YUV 得到的 H.264 文件压缩比达到 330 倍,实验中使用的最小码率 50 kbps 时,压缩比更是达到了 3640 倍。压缩比越大,解码后的视频失真越严重。当编码参数变化时,本文也观察并分析了 DPM 识别算法性能下降的原因。

如图 3(a)所示,码率 50 kbps 编码后的视频解码后每帧图像失真非常严重,块效应严重,而且图中人体出现了大的块效应。DPM 算法进行全局模板匹配时,在大的块效应作用下,得到的人体边缘梯度直方图误差非常大,影响人体目标整体识别,而人体目标整体识别得分在人体识别最终得分比例最高,所以大的块效应对识别性能影响非常大。故监控视频以码率 50 kbps 编码,其传输解码后的人体目标识别性能远低于原始 YUV 视频,识别性能指标 $AP_1 = 0.35$ 低于 AP_0 一半,大的块效应为最大影响因素。

图 3(b)为码率 300 kbps 编码后的与图 3(a)相同帧的图像。与图 3(a)相比,它的图像质量显著上升,虽然没有明显的大的块效应,但小的块效应广泛存在。DPM 算法进行局部部件模板匹配时,小的块效应干扰小尺度精细特征的提取,使算法不能得到正确的人体部件和部件位置。所以小的块效应影响人体目标部件的识别,因为部件识别得分占人体识别最终得分相对低于整体识别得分,所以小的块效应对识别性能影响小于大的块效应。编码码率在区间 50~300 kbps 时,影响 AP 值的主要因素为大的块效应,它对识别性能影响非常大,因此随着码率的上升识别性能指标 AP 值上升较快。

图 3(c)为码率 700 kbps 编码后的与图 3(a)相同帧的图像。将它与图 3(b)进行仔细对比,发现图 3(c)的图像质量上升,小的块效应也基本消失。显然编码码率在区间 300~700 kbps 时,小的块效应为识别性能的最大影响因素,而小的块效应对识别性能影响小于大的块效应,因此在区间内随着码率的上

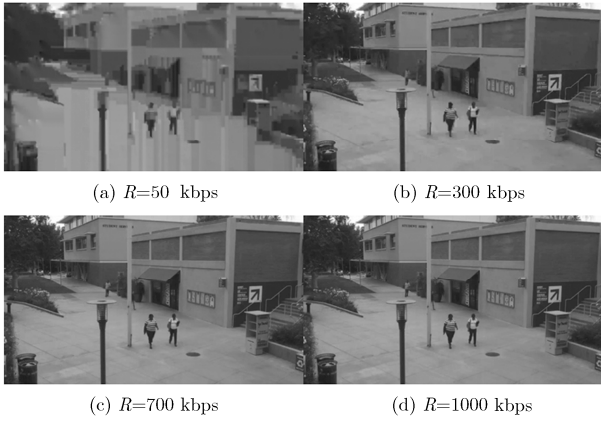


图 3 不同编码码率的解码视频同一帧的图像

升 AP 值上升相对较慢。

图 3(d)为码率 1000 kbps 编码后的与图 3(a)相同帧的图像，它的图像质量与图 3(c)基本相同，小的块效应彻底消失。因此编码码率在区间 700~1000 kbps 时影响识别性能的主要因素为 DPM 目标识别算法和原始视频，随着码率的上升，AP 值趋于稳定，略小于 AP_0 。

根据以上分析，在原视频及其它编码参数一定的条件下，通过在视频编码时选取合适的码率，则既可以节约信道，又能保证视频具有较高的识别性能。

4 分辨率对识别性能影响

视频分辨率代表每帧图像的精细程度，通常用每帧图像的像素个数表示。本文实验采用的 VIRAT 数据库视频分辨率大小为 $D_0 = 1280 \times 720$ 。在视频编码器的工作参数中，可以选择不同的视频分辨率，它对最终的编码速率和率失真影响也很大^[20]。这里，在编码前对原始视频进行下采样，减小视频的分辨率，为视频质量的控制增加一个可调参数。解码重建后的视频需要进行上采样恢复相同的视频分辨率，然后再进行目标识别检测。下采样时使用高斯平滑滤波器进行像素滤波，上采样时使用双线性滤波进行插值。实验的其它步骤与第 3 节相同，这里不再描述。

4.1 分辨率对识别性能影响的实验原理

如图 4 所示，本文首先将原 YUV 视频 A 进行下采样，下采样尺度比为 S_D ，得到分辨率大小为 $D_0 \times S_D$ YUV 视频 A'；然后使用 H.264/AVC 算法将 A' 压缩编码，限制码率大小为 R，得到 H.264 文件 H；将传输后的文件 H' 解码得到 YUV 文件 B'，显然 B' 分辨率大小和 A' 相同；最后将 B' 上采样，上采样尺度比为 $S_U = 1/S_D$ ，得到与 A 分辨率大小相同的 YUV 视频 B。

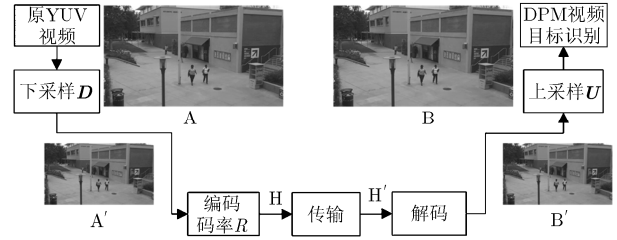


图 4 采样编码流程图

设 δ_c 为编码产生的 AP 值误差，则视频 B 识别性能表示为

$$AP = U(D^T(AP_0) + \delta_c) = UD^T(AP_0) + U\delta_c \quad (2)$$

其中 D 为下采样数据， U 为上采样数据， δ_c 为编码产生的误差。

视频 B 识别性能 AP 被分解为采样 AP 项和编码 AP 误差项。为了研究分辨率大小对识别性能的影响，则需要将编码误差降到最小。在理想状态下，编码误差为 0。在本节实验中，通过去除图 4 实验流程中的编码和解码模块，可以模拟编码误差为 0 的条件。此时的目标识别性能只是采样尺度比的函数：

$$AP = UD^T(AP_0) \quad (3)$$

实验中，本文取离散的下采样尺度比 $S_D = (s_1, s_2, s_3, \dots, s_8) = (1, 7/8, 6/8, \dots, 1/8)$ ，解码后对应的上采样尺度比 $S_U = (1/s_1, 1/s_2, 1/s_3, \dots, 1/s_8) = (1, 8/7, 8/6, \dots, 8/1)$ 。本文采用 JSVM 编码器中的数据抽取和差值函数完成对视频进行上、下采样。

下采样后的编码视频经解码后，如果形成的 YUV 图像不经过上采样而直接进行目标识别，得到的识别性能很差，其中在采样尺度比为 3/8 时 AP 值已经等于 0。经过分析，这与 VIRAT 数据库监控视频和 DPM 模型相关。由于监控画面的角度大，范围广，人在视频中所占的比例较小，当对视频下采样分辨率大小减小后，人体目标所占的像素随之减小；使用的 DPM 模型需要对人体各个部分进行识别，当人体所占像素较小时则人体部件所占像素过小而识别性能变差甚至不能识别。所以在实验中，将解码后的视频进行上采样，使其分辨率大小恢复原始的尺度。

4.2 分辨率对识别性能影响实验结果及分析

经过实验，得到的拟合结果如图 5。根据样本点的分布情况，分别对其进行负指数函数和多项式函数拟合，拟合结果如表 2 所示。通过对比得到负指数函数拟合式(4)拟合效果更好。

结合图 5 与式(4)，可以直观得到分辨率大小与 AP 值之间的关系。随着 S_D 的增大，AP 值先增加，

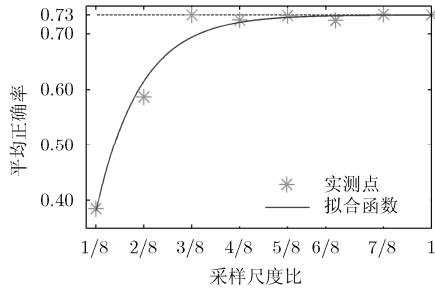


图 5 视频在不同采用尺度比的识别性能

$$AP = AP_0 - 1.053e^{-8.729S_D} \quad (4)$$

后平稳趋近于 AP_0 。分析原视频 A 和不同 S_D 解码出的视频 B, 得到与第 3 节类似的结论。即在 $S_D \in (s_1, s_3)$ 时, 视频 B 解码出的帧图像存在块效应, 并随着 S_D 的增大块效应逐渐减少, 在 S_3 时块效应基本消失。由第 3 节分析知道, 块效应会影响 DPM 算法全局模板和部件模板的匹配, 降低人体目标识别性能。因此 $S_D \in (s_1, s_3)$ 时, 随着 S_D 的增大 AP 值上升; $S_D \in (s_3, s_8)$ 时, AP 值在 AP_0 附近保持平稳。

视频下采样后, 生成的码流速率将减小, 对传输信道的带宽要求降低。因此, 在原视频及其它参数不变的条件下, 通过对视频进行合适采样尺度比的下采样后再传输, 既可以节约信道, 又能保证视频具有较高的识别性能。这样可以为视频编码参数的优化提供一个新的思路。

表 2 采样尺度比对识别性能影响的函数拟合结果

	拟合结果	相关系数	标准差
负指数函数拟合	$AP = AP_0 - 1.053e^{-8.73S_D}$	0.9888	0.0205
多项式函数拟合	$AP = AP_0 - 0.00682S_D^{-1.9}$	0.9775	0.0298

5 码率和分辨率对识别性能的联合影响

经过测试和分析, 发现通过选取合适码率编码或者合适分辨率大小采样, 可以在保持较高识别性能情况下减小视频传输所需的带宽。本节结合前面的实验, 同时改变码率和分辨率大小, 希望在识别性能和视频传输带宽之间寻找更合适的平衡点。

5.1 码率和分辨率对识别性能联合影响的实验方法

本节的实验流程设计与第 4 节相同。因为需要同时研究码率和分辨率对识别性能影响, 因此实验操作步骤中, 与第 4 节的主要区别是编码和解码模块不再省略。为了方便对比, 离散选取的编码速率与第 3 节相同, 而采样尺度比 S_D 选取与第 4 节相同。

5.2 码率和分辨率对识别性能联合影响的结果及分析

实验结果如图 6 中点所示。图 6(a)以码率为横坐标, 为了便于观察, 本文只画出了采样尺度比 S_D 为 1/8, 3/8, 5/8 和 7/8 的实验结果。图 6(b)以采样尺度比为横坐标, 画出了码率 R 为 100 kbps, 300 kbps, 500 kbps 和 700 kbps 的实验结果。第 4 节由式(3)知, 视频 B 的识别性能由视频采样 AP 项 $UD^T(AP_0)$ 和编码 AP 误差项 $U\delta_c$ 组成, 结合式(4)得到

$$\begin{aligned} AP &= UD^T(AP_0) + U\delta_c \\ &= (AP_0 - 1.053e^{-8.73S_D}) + U\delta_c \end{aligned} \quad (5)$$

在只进行编码而不进行下采样的模式下, 由式(1)可得编码误差项:

$$U\delta_c = -0.49e^{-0.005125R} \quad (6)$$

当编码与采样联合时, 显然编码误差项 $U\delta_c$ 将受码率和采样尺度的共同影响。仿照式(6), 假设编码误差项受采样尺度影响的形式为

$$U\delta_c = -\{0.49 + f_1(S_D)\}e^{-R\{0.005125 + f_2(S_D)\}} \quad (7)$$

其中 $y_1 = f_1(S_D)$, $y_2 = f_2(S_D)$ 是 S_D 的函数。将实验数据代入式(7), 得到 y_1 , y_2 的统计值, 如图 7 中样点所示。

根据 y_1 和 y_2 值的分布情况分别对其进行拟合, 得到式(8), 式(9):

$$y_1 = f_1(S_D) = -0.00022e^{-7.413S_D} \quad (8)$$

$$y_2 = f_2(S_D) = 0.0013S_D^{-1} \quad (9)$$

将 $f_1(S_D)$, $f_2(S_D)$ 拟合结果式(8), 式(9)代入式(7), 得到

$$\begin{aligned} U\delta_c &= -\left\{0.49 - 0.00022e^{-7.413S_D}\right\} \\ &\quad \cdot e^{-R\{0.005125 + 0.0013S_D^{-1}\}} \end{aligned} \quad (10)$$

将式(10)代入式(5), 得到采样尺度比、编码速率及 AP 之间的关系拟合式(11)。验证得到式(11)与原数据的相关系数为 0.9688。图 6(a)和图 6(b)分别是以码率和采样尺度比为横坐标的拟合曲线。

$$\begin{aligned} AP &= AP_0 - 1.053e^{-8.73S_D} \\ &\quad - \left\{0.49 - 0.00022e^{-7.413S_D}\right\} \\ &\quad \cdot e^{-R\{0.005125 + 0.0013S_D^{-1}\}} \end{aligned} \quad (11)$$

当采样尺度比 S_D 一定时, 式(11)为变量 R 的负指数函数。因此在图 6(a)中, S_D 确定的曲线随着 R 增大先上升后趋于稳定。即 S_D 确定条件下, $R \in (50, 300)$ 时, 随着 R 的增大, 编码产生的块效应逐渐减小, 目标识别性能提高。 $R > 300$ kbps 时, S_D 为影响识别性能的主要因素, AP 趋近于 AP_0 -

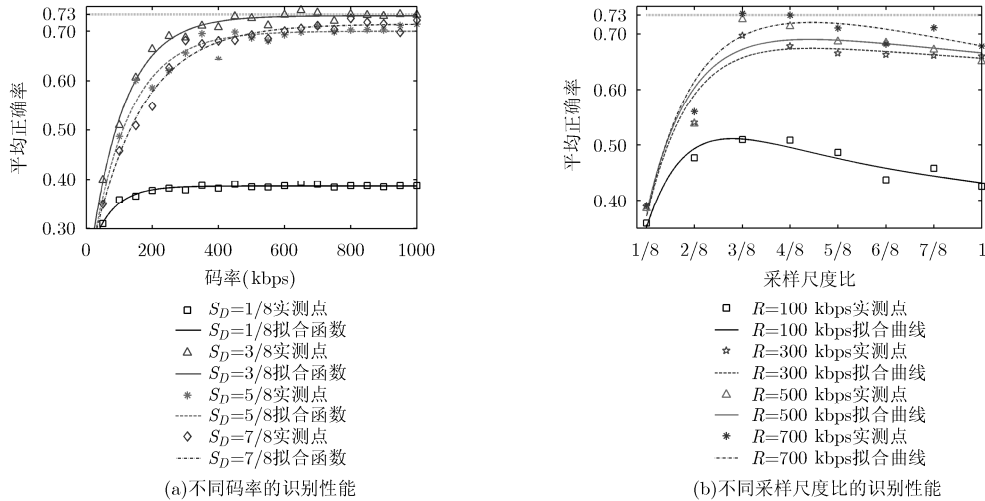


图 6 视频在不同码率和分辨率大小条件下编码传输解码后的识别性能

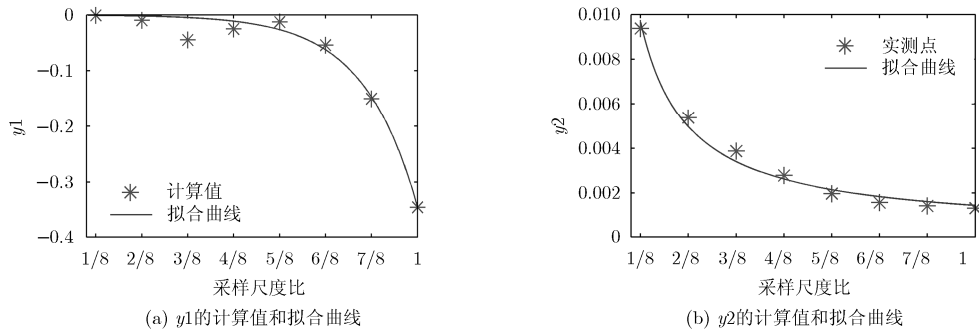


图 7 y_1, y_2 的计算值和拟合曲线

$1.05e^{-8.73S_D}$ 。

当编码码率 R 一定时，式(9)为只含变量 S_D 的函数。分析这个函数，可以得到 S_D 较小时式(11)值变化趋势主要由视频采样 AP 项 $UD^T(AP_0)$ 决定； S_D 较大时，视频采样 AP 项 $UD^T(AP_0)$ 随 S_D 增大保持稳定，式(9)变化趋势主要由编码 AP 误差项式(7)决定。因此在图 6(b)中， R 确定的曲线随着 S_D 增大先快速上升后缓慢下降，在 S_D 为 3/8 处出现峰值。即 R 确定的条件下，当 $S_D \in (1/8, 3/8)$ 时，随着 S_D 的增大，采样产生的块效应显著减少，AP 值快速上升。当 $S_D \in (3/8, 1)$ 时，随着 S_D 的增大，码率与采样尺度比共同作用的编码块效应缓慢增加，AP 值平缓下降，在 S_D 为 3/8 处出现峰值。

综合分析码率和分辨率大小，本文发现：在编码前先将视频用合适的尺度比进行下采样，然后在接收端进行上采样恢复，能够在保证视频具有较高识别性能前提下进一步降低码率。因此，选取合适的码率和分辨率大小，可以在传输带宽和目标识别性能之间获得平衡。

6 结束语

在网络化应用环境中，视频编码的编码速率和分辨率参数，对接收端目标识别性能有重要影响。本文通过实验，获得了检测性能与编码速率和分辨率的函数关系。在实际应用中，可以通过选取合适的码率和分辨率大小，来降低传输所需的带宽，同时保证视频具有较高的识别性能。这也为针对不同的视频应用，设计更合理的率失真优化函数提供了依据。但是由于该函数关系具有较强的非线性，为优化算法提出了新的问题。下一步将继续细化统计特征分析和验证，另一方面也将进行联合优化算法的研究。

参考文献

[1] Li L J and Li F F. What, where and who? classifying events by scene and object recognition[C]. Proceedings of the IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 2007: 1-8.

[2] Lei B, Wang T, Chen S, et al.. Object recognition based on adaptive bag of feature and discriminative learning[C].

- Proceedings of the 20th IEEE International Conference on Image Processing, Melbourne, Australia, 2013: 3390–3393.
- [3] Dalal N and Triggs B. Histograms of oriented gradients for human detection[C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, USA, 2005, 1: 886–893.
- [4] Wei D, Zhao Y, Cheng R, *et al.* An enhanced histogram of oriented gradient for pedestrian detection[C]. Proceedings of the 4th IEEE International Conference on Intelligent Control and Information Processing, Beijing, China, 2013: 459–463.
- [5] Felzenszwalb P F, Girshick R B, McAllester D, *et al.* Object detection with discriminatively trained part-based models[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(9): 1627–1645.
- [6] Ding Y, Zhang J, Li J, *et al.* A bag-of-feature model for video semantic annotation[C]. Proceedings of the 6th IEEE International Conference on Image and Graphics, Hefei, China, 2011: 696–701.
- [7] Huang D K, Chen K Y, and Cheng S C. Video object detection by model-based tracking[C]. Proceedings of the 20th IEEE International Symposium on Circuits and Systems, Beijing, China, 2013: 2384–2387.
- [8] Blair C, Robertson N M, and Hume D. Characterizing a heterogeneous system for person detection in video using histograms of oriented gradients: power versus speed versus accuracy[J]. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 2013, 3(2): 236–247.
- [9] Liu Y, Jang Y, Woo W, *et al.* Video-based object recognition using novel set-of-sets representations[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, USA, 2014: 533–540.
- [10] Sharma P, Huang C, and Nevatia R. Unsupervised incremental learning for improved object detection in a video[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, USA, 2012: 3298–3305.
- [11] Wu Q and Li H. Mode dependent down-sampling and interpolation scheme for high efficiency video coding[J]. *Signal Processing: Image Communication*, 2013, 28(6): 581–596.
- [12] Wang T, Chen Y, He Y, *et al.* A real-time rate control scheme and hardware implementation for H. 264/AVC encoders[C]. Proceedings of the 5th IEEE International Congress on Image and Signal Processing, Chongqing, China, 2012: 5–9.
- [13] Felzenszwalb P F and Huttenlocher D P. Pictorial structures for object recognition[J]. *International Journal of Computer Vision*, 2005, 61(1): 55–79.
- [14] Felzenszwalb P F, Girshick R B, and McAllester D. Cascade object detection with deformable part models[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 2010: 2241–2248.
- [15] Girshick R B, Felzenszwalb P F, and McAllester D A. Object detection with grammar models[C]. Proceedings of the 25th IEEE Conference on Advances in Neural Information Processing Systems, Granada, Spain, 2011: 442–450.
- [16] 袁武, 林守勋, 牛振东, 等. H. 264/AVC 码率控制优化算法[J]. *计算机学报*, 2008, 31(2): 329–339.
- Yuan W, Lin S X, Niu Z D, *et al.* Efficient rate control schemes for H.264/AVC[J]. *Chinese Journal of Computers*, 2008, 31(2): 329–339.
- [17] 魏江, 刘迪. 基于 DM642 的 X.264 编码器优化[J]. *现代电子技术*, 2011, 34(14): 68–70.
- Wei J and Liu D. Optimization of X.264 encoder based on DM642 platform[J]. *Modern Electronics Technique*, 2011, 34(14): 68–70.
- [18] Huang Y H, Ou T S, and Su P Y. Perceptual rate distortion optimization using structural similarity index as quality metric[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2010, 20(11): 1614–1624.
- [19] Ou T S, Huang Y H, and Chen H H. SSIM-based perceptual rate control for video coding[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2011, 21(5): 682–691.
- [20] Wang R, Huang C, and Chang P. Adaptive downsampling video coding with spatially scalable rate-distortion modeling [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2014, 24(11): 1957–1968.
- 吴泽民: 男, 1973 年生, 博士, 副教授, 研究方向为数据融合、图像分析.
- 刘涛: 男, 1991 年生, 硕士生, 研究方向为图像分析、目标识别.
- 姜青竹: 男, 1987 年生, 硕士生, 研究方向为视频编码与传输.