物联网环境下基于场景要素的多码流变分辨率压缩传输技术研究

肖尚武，胡瑞敏，陈宇，肖晶
（武汉大学计算机学院国家多媒体软件工程技术研究中心，湖北 武汉 430072）

摘　要：针对城市监控覆盖面广、海量接入的需求，实现低带宽和低功耗性能是解决这一问题的重要研究方向。在智慧城市、安防监控等应用领域，基于场景要素，如人脸关键区域的视频监控尤为重要。实现场景要素的提取，以极低带宽传输关键信息，通过多码流区别编码策略，在物联网环境下实现视频技术的应用，是目前值得研究的可行方向。通过设计面向人脸的变分辨率混合编码算法，可大幅度节省带宽、降低功耗，满足窄带物联网的接入要求。通过基于深度学习 Caffe 框架的人脸检测算法，在关键帧获取人脸感兴趣区域，并以高分辨率编码人脸图像；通过设计码率自适应分配算法，合理利用带宽，区别编码人脸信息和全图背景内容；通过窄带传输编码后的混合码流信息，在接收端采用基于关键帧的人脸增强解码算法，得到人脸局部的高清监控画面。实验表明，采用所提方法在 120~160 kbit/s 窄带传输时，人脸画面可以保持与原始高清监控采集端同等清晰度，具有很强的实用性。

关键词：NB-IoT；监控视频；变分辨率；视频编码；人脸检测
中图分类号：TP37
文献标识码：A
doi: 10.11959/j.issn.2096-3750.2018.00075

Research on multi-stream variable resolution compression and transmission technology based on scene elements in Internet of things environment

XIAO Shangwu, HU Ruimin, CHEN Yu, XIAO Jing
National Engineering Research Center for Multimedia Software, Wuhan University, Wuhan 430072, China

Abstract: In response to the demand of wide coverage and massive access, low bandwidth and low power consumption is an important research direction to solve this problem. In smart cities, security monitoring and other application areas, video surveillance based on the region of interest of the face are particularly important. It is a feasible direction to realize the extraction of scene elements, transmission of key information with very low bandwidth and the application of video technology in the Internet of things environment through the strategy of multi-stream differential coding. By designing a face-oriented variable resolution hybrid coding algorithm, the bandwidth could be saved and the power consumption could be reduced greatly, the access requirements of narrowband Internet of things could be met. Through the face detection algorithm based on the deep learning Caffe framework, the face region of interest was acquired in key frames, and the face image was encoded with high resolution. By designing the code rate adaptive allocation algorithm, the bandwidth was utilized rationally, and the encoded face information and the full background content were distinguished. The encoded mixed code stream information was transmitted through the narrowband; the key frame-based face enhancement decoding algorithm was adopted at the receiving end to obtain a partial HD high-definition monitoring picture. Experiments show that when the video encoded by the proposed method is transmitted in a narrow band whose transmission rate is 120~160 kbit/s, the face image can maintain the same definition as the original HD monitoring acquisition end, which has strong practicability.

Key words: NB-IoT, surveillance video, variable resolution, video coding, face detection

收稿日期：2018-08-09；修回日期：2018-10-29

1 引言

随着城市视频监控系统的普及，监控在公共安全领域发挥着越来越重要的作用。安防监控系统作为预防和打击犯罪以及预防灾害事故发生的利器，是社会治安综合治理的重要一环。随着布控范围的增大、画面清晰度要求的提高，每日产生的监控视频数据量也在不断增加，使得以 4G 和 Wi-Fi 为主的现有监控视频接入方式所需的能耗及成本与日俱增。如何在保持监控视频可分析性的同时降低监控摄像机接入和网络传输成本，已成为必须解决的问题。

近年来，公安部在全国着力推广智慧小区工程，监控视频在物联网环境下的部署是亟待解决的实际问题。以 NB-IoT 为代表的物联网技术逐渐兴起，被广泛应用于智慧城市、智能抄表、智慧物流等领域。由于其广覆盖、海量连接、低功耗和低成本的特性，满足城市视频监控接入对低功耗、低成本的要求，可极大提升监控系统的布设范围和可扩展性。同时，由于基于人脸感兴趣区域的视频监控是安防领域的核心，因此，研究面向人脸的物联网视频监控技术具有很大的实用价值和科研意义。

视频清晰度与带宽要求是一对矛盾问题，在 NB-IoT 极低带宽下，视频传输困难。由于监控视频的拍摄呈现高清化趋势，加之物联网的上行带宽极窄，对监控视频编码效率提出了严峻挑战。因此，急需研究面向极低码率应用的监控视频编码方法。在实际视频监控中，行人的人脸画面质量较差、尺度较小，特别是在极低码率的监控视频传输业务中，为有效传输监控视频流，需要采用高压缩率的方式，导致视频质量严重受损，进一步限制了监控视频的人脸信息提取与辨识性能。考虑监控视频的特殊性，侦查人员通常关注某些特定的感兴趣区域，如人脸、车牌等，而背景等非感兴趣区域包含的刑侦辨识信息相对较少，所以可以根据不同区域的信息重要性采用相应的视频处理方法。对于感兴趣区域，需要在编码过程中尽可能降低失真以保障可分析性；对于非感兴趣区域，可采取有效方式降低编码所需的码率从而尽可能地节约传输资源。

针对监控视频降低码率的研究，当前主要有两大方向，一是前背景的分离，二是围绕感兴趣与非感兴趣区域区别处理[1]。在前背景分离方面，监控视频编码又可分为针对背景特性的编码方法和针对前景特性的编码方法。针对背景特性的编码方法主要利用背景长时微变特性，去除背景冗余；针对前景特性的编码方法主要利用前景对象的二维轮廓特征，去除前景对象的纹理冗余[2]。2014 年，北京大学的张贤国等[3]提出了基于背景建模的编码方法，将视频图像宏块分为前景宏块、背景宏块、前背景混合宏块 3 类。利用一段时间的视频图像，通过滑动平均法构造背景参考帧 G 帧，用于背景宏块预测，同时通过背景差分法构造差分参考帧，用于前背景混合宏块的预测。高质量的参考帧使得背景宏块与前背景混合宏块的预测精度显著提升，极大地降低了背景部分编码所需的比特数。2012 年 Tsai 等[4]提出了一种基于内容的编码框架，通过视频图像与估计得到背景图像的差异，将视频图像宏块划分为对象宏块、动态背景宏块、静态背景宏块以及不透明背景宏块，利用前景对象的二维结构特性，提升对象纹理的预测精度，从而提高对象的编码效率。在感兴趣与非感兴趣区域进行区别处理方面，包括提升质量和分配更多码率两种思路。2010 年 Ng 等[5]通过二值形状掩膜、灰度形状图以及深度图对轮廓特征进行修正，提升了基于对象的重建质量。2016 年 Liao 等[6]提出了基于特征相似度加权的 R-Lambda 模型，使感兴趣区域与非感兴趣区域之间的码率分配更精确，对非感兴趣区域采用较大量化步长编码，在保证感兴趣区域质量的同时，降低整体码率。

上述方法主要通过提升预测精度或者降低背景及非感兴趣区域的图像质量从而减少总数据量，未对视频分辨率进行调整。传统的基于感兴趣区域的视频编码技术，存在场景要素提取幅度大、不精准的情况，占用过多码率；并且仅通过量化步长的调整实现感兴趣区域与非感兴趣区域的区别编码，当量化步长过大时，背景出现“花屏”现象（大面积图像量化失真），严重影响整体视觉观感；量化步长调整视频压缩率的能力有限，达不到 NB-IoT 的要求，而通过降低背景分辨率，可极大地提高压缩效率。实验结果表明，在同等码率情况下，低分辨率的背景图像质量明显优于大量化步长的高分辨率图像质量。对于高清监控视频，编码前数据量较大，传统方法的编码效率难以满足 NB-IoT 的传输要求，急需研究更高效、适用于物联网环境的监控视频编码方法。本文将视频编码从传统的面向时空冗余的压缩，拓展到面向内容对象理解的多码流交叉编码模式。

2　系统方案

本文以理论算法为基础，通过设计实验系统进行验证的方式展开面向物联网的变分辨率监控视频编码技术的研究。系统结构分为编码端、窄带传输、解码端 3 个环节，系统结构模型如图 1 所示。在视频信息处理流程中，可以分为 4 个关键模块：基于深度学习的人脸检测模块、基于 NB-IoT 的码率自适应分配模块、变分辨率混合编码模块和基于 I 帧的人脸增强解码模块。

1) 在人脸检测模块中，针对采集的监控视频数据，指定图像组（GOP, group of pictures）大小，在关键帧进行人脸检测，区分人脸感兴趣区域与非感兴趣区域，并标定人脸轮廓，该阶段主要利用人脸检测、显著区域提取等技术。本文采用基于深度学习 Caffe 框架，有效提高了检测率和运算效率。

2) 在码率自适应分配模块中，优先保证人脸感兴趣区域的高质量编码，根据实际可用的窄带带宽上限，计算剩余带宽资源，以剩余码率编码全图并做好码率控制、平稳传输。

3) 在变分辨率混合编码模块中，根据关键帧中人脸检测环节得到的感兴趣区域信息，作为图片保持原始采集分辨率，采用 JPEG 高质量或无损编码，并二次编码为二进制数据流，以利于数据传输。在原始监控画面下采样，分辨率变换至适合于窄带的 CIF 级清晰度，以传统的 x264 编码器编码全图信息。

4) 在人脸增强解码模块中，与普通解码器有所不同，该系统基于双码流解码。先进行预解码，分别得到关键帧高清人脸图片和 CIF 级低清视频序列，将序列的分辨率变换至采集端原始分辨率；在关键帧根据参数信息恢复高清人脸区域，存入编码器缓冲区，根据前后两个关键帧增强中间帧人脸清晰度，得到最终解码序列。

在系统结构模型中，硬件系统分为编码端和解码端，通过监控采集设备得到视频画面，送入编码端，实验中采用 NVIDIA Jetson TK1 开发板，进行实时人脸检测和编码处理，经过 NB-IoT 的 Wi-Fi 模组接入移动数据网络，再通过承载网络传输到内部的监控后台，进行解码和存储。软件系统与硬件系统基本对应，采集的画面以 YUV 格式在混合编码系统中，实现人脸检测和多码流变分辨率编码；通过窄带传输后，在解码分析系统，实现混合码流的解析、解码，得到整个视频序列具备人脸高清、背景低清的监控画面，并进行结果显示输出。

3　系统实现

在系统结构模型的搭建中，核心算法主要包含 4 个方面：在原始监控采集画面中，采用基于深度学习 Caffe 框架的人脸检测算法；根据窄带带宽要求，自适应调整码率分配，以人脸图像为优先的码率分配算法；以原始高分辨率编码人脸图像，以较低分辨率编码采样后视频画面的混合编码算法；在解码端，基于 I 帧人脸高清，通过 I 帧增强 P 帧人

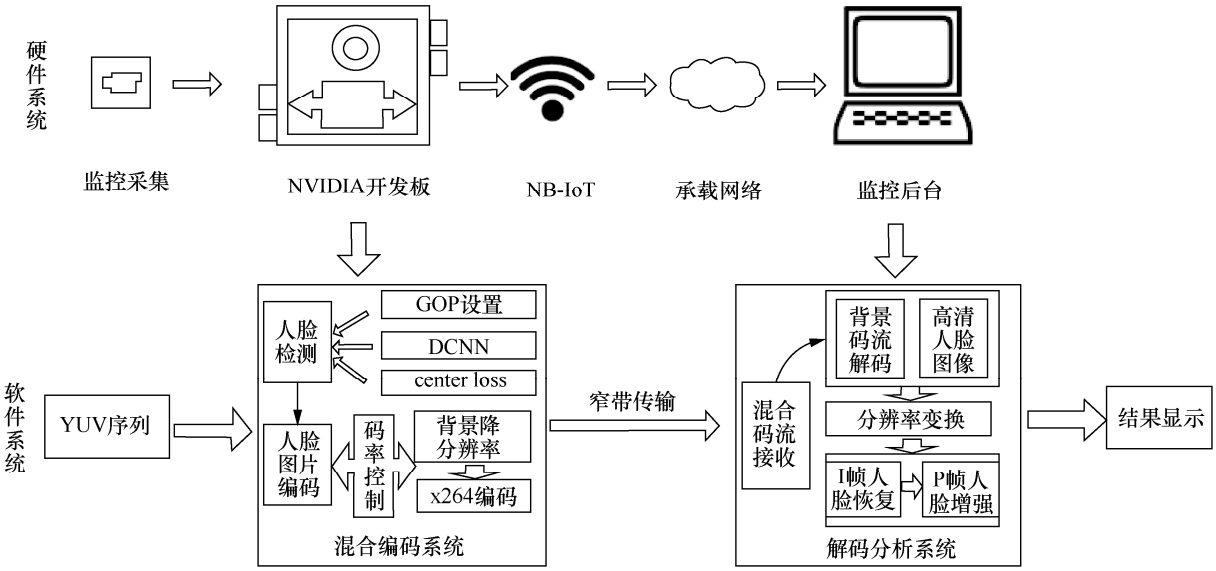

图 1　系统结构模型

脸区域的解码算法。

3.1 场景要素的分析与提取

分析场景中的关键要素，并实现有效提取是区分信息重要性的第一步。本文采用基于深度学习的人脸检测算法。

考虑人脸信息的时域冗余，相邻两帧图像变化极小；另一方面，如果保留所有帧的高清人脸信息，则会占用过多码率，而窄带带宽资源有限。所以此处仅在关键帧 I 帧进行人脸检测，编码传输 I 帧人脸高清信息。

本文采用基于深度学习 Caffe 框架的人脸检测方法，网络结构的主体采用 DCNN（深度卷积神经网络，deep convolutional neural network），模型来源于WIDER FACE数据库基础离线训练得到的人脸检测模型[7]，同时采集部分真实场景下的数据，即现场视频截图照片对原有模型进行微调。另外，本模块在原有网络结构的softmax loss函数的基础上加入center loss 函数，在保持较高检测准确率的同时，提高运算效率。

该人脸检测算法步骤如下。

1) 使用全卷积网络P-Net生成人脸的候选窗和边框回归向量，使用边界框回归方法校正该候选窗，使用非极大值抑制合并重叠候选窗。该网络中的人脸分类交叉熵损失函数根据式(1)计算

$$L_i^{\text{det}} = -\left(y_i^{\text{det}} \log(p_i) + \left(1 - y_i^{\text{det}}\right)\left(1 - \log(p_i)\right)\right)$$
$$y_i^{\text{det}} \in \{0,1\} \tag{1}$$

其中，p_i 是输入图像为人脸的概率，y_i^{det} 为输入图像的真实标签。

2) 使用添加了全连接层的 R-Net 改善人脸候选窗，使用边界框回归和非极大值抑制过滤错误的人脸候选窗。边界框回归通过欧氏距离计算回归损失

$$L_i^{\text{box}} = \| \hat{y}_i^{\text{box}} - y_i^{\text{box}} \|_2^2$$
$$y_i^{\text{box}} \in R^4 \tag{2}$$

其中，$\hat{y}_i^{\text{box}}$ 是通过网络预测得到的输入图像坐标，y_i^{box} 为实际的输入图像坐标。

3) 使用 O-Net 输出最终人脸框和相应特征点位置，计算网络预测的坐标位置和真实坐标的欧氏距离，并最小化该距离。

$$L_i^{\text{landmark}} = \left\| \hat{y}_i^{\text{landmark}} - y_i^{\text{landmark}} \right\|_2^2$$
$$y_i^{\text{landmark}} \in R^{10} \tag{3}$$

其中，$\hat{y}_i^{\text{landmark}}$ 是通过网络预测得到的坐标位置，y_i^{landmark} 为实际的真实坐标。

4) 计算 center loss 值，并根据该特征值与其对应的 center 距离调整惩罚函数。center loss 按照式(4)计算

$$L_c = \frac{1}{2}\sum_{i=1}^{m} \| x_i - c_{y_i} \|_2^2 \tag{4}$$

其中，x_i 是第 i 张图片的特征值，c_{y_i} 是该图片所属分类的中心，即特征值的中心。

至此，完成了感兴趣区域人脸信息的提取。

3.2 要素评估与码率估计

提取场景要素之后，根据其重要程度评估应分配的码率资源，在保证要素应达到的质量指标的前提下，尽可能提高背景质量。本文设计基于 NB-IoT 的码率自适应分配算法，具体过程如下。

在窄带传输过程中，期望平均码率不超过 160 kbit/s，本方案涉及人脸信息和全图信息的双码流传输，所以需要合理分配带宽资源。既要保证人脸信息的高质量编码，又要在充分利用带宽的前提下尽可能提高全图背景的清晰度。在编码过程中，首先需要确定人脸画面压缩等级，从最大可接受的有损压缩到无损压缩之间调节，选择合理的参数设置，然后在此基础上计算剩余带宽资源，用于编码视频背景画面。

$$B_i' = \frac{B_{\text{NB-IoT}} \times \text{GOP} - \sum_{k=1}^{n} d_{\text{JPEG}}}{\text{GOP}} \tag{5}$$

其中，$B_{\text{NB-IoT}}$ 为窄带最大带宽限制，GOP 为图像组长度，d_{JPEG} 表示一张人脸区域图像数据，n 为相应 I 帧的人脸数，由式(5)计算得到第 i 个GOP 的剩余码率为 B_i'。本文窄带带宽 $B_{\text{NB-IoT}}$ 取 160 kbit/s，以 B_i' 为最高码率限制，对背景全图的编码进行码率控制。

此处采用 CVBR（constrained variablebit rate）码率控制方式，兼顾恒定比特率（CBR, constant bit rate）和动态比特率（VBR, variable bit rate）两种方法的优点，即在图像内容静止时，节省带宽；当有移动发生时，利用前期节省的带宽来尽可能提高图像质量，达到同时兼顾带宽和图像质量的目的。限定编码过程的最大码率和最小码率，当画面静止时，码率稳定在最小码率附近；当画面运动时，码率大于最小码率，但是不超过最大码率，整体可以实现很好的码率平稳效果。

3.3　变分辨率混合编码算法

为适应 NB-IoT 的传输环境，系统采用变分辨率混合编码的技术路线。核心思想是将感兴趣区域和非感兴趣区域区分开，以不同分辨率分别进行编码处理，并设计具备同步时间戳（timestamp）的双码流结构。基于分辨率变换的监控视频编码框架如图 2 所示。

在监控采集端，首先确定 GOP 大小，并固定 I 帧间隔，即每隔 GOP 为一个关键帧 I 帧，此处需要说明，由于是实时监控视频，所以仅采用 I 帧和 P 帧编码，每个 GOP 由 1 个 I 帧和 GOP−1 个 P 帧组成。采集的视频数据输入编码器中，判断是否为 I 帧，若为 I 帧，则进行人脸独立处理；若不为 I 帧，则作为普通 P 帧编码，编码分辨率继承前面最邻近 I 帧的分辨率大小。

在 I 帧人脸处理环节中，首先对 I 帧进行人脸检测和人脸轮廓定位，抠出人脸轮廓，得到感兴趣区域的人脸图片和该图在视频帧中的位置信息。对于人脸图片，保持原始采集的高清分辨率大小不变，采用图片编码（或帧内编码）得到压缩后的高清人脸数据，本文采用常用的 JPEG 编码，在保证质量基本无损的同时，确保了较高的压缩率。但是，由于 I 帧可能存在多幅人脸图片，且进行图像编码后的数据整体不利于传输，所以还需要进行二次编码，基于统计学的熵编码，将图片编码成二进制码流信息，并通过截断处理，分成较小的码流片，避免出现码流阻塞或波动较大的情况。该帧中的人脸数、相对位置、帧戳以及分辨率档次等参数信息，则类似于图片流，标注时间戳后作为码流分组同步发送。

经过图片编码后，可以由带宽减去 I 帧中的总人脸图片数据，预估剩余带宽资源大小，具体过程如式(5)所示。在该 GOP 中，由摄像头采集的原始视频数据，需要根据由上述步骤得到的剩余带宽大小，自适应调整分辨率。在保证尽可能利用带宽资源而不超过带宽大小的前提下，适配最大分辨率档次，以最佳分辨率编码全图信息，此处的最佳分辨率即满足要求的最大分辨率档次。剩余码率适配分辨率档次如表 1 所示。

由表 1 得到编码全图所确定的分辨率后，对原始图像进行下采样，将分辨率降低到对应的档次。然后采用常规视频编码器进行编码，并且通过 CVBR 算法进行码率控制，保证码率不会出现波动较大的情况。最后与图片流（包括参数流）一起组

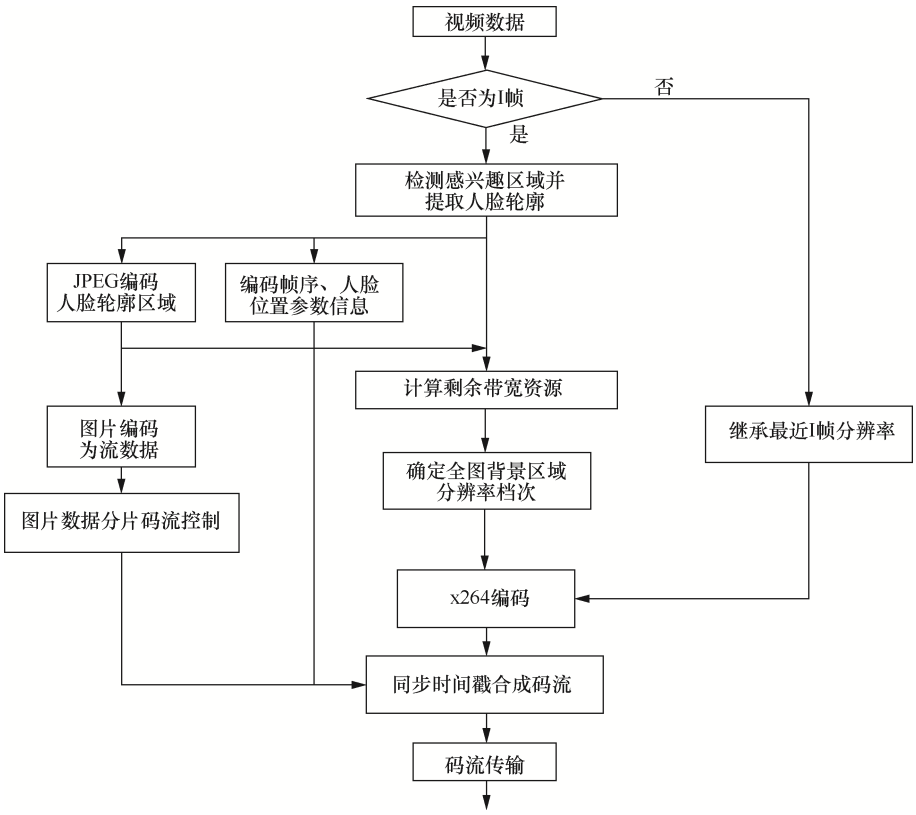

图 2　基于分辨率变换的监控视频编码框架

成双码流同步传输至解码端。

表 1　　剩余码率适配分辨率档次

剩余码率/(kbit·s^{-1})	图像格式	分辨率	Level
<70	sub-QCIF	128×96	5
70～90	QCIF	176×144	4
90～120	CIF	352×288	3
120～150	SIF	640×480	2
150～160	4CIF	704×576	1

3.4　基于 I 帧人脸增强的解码算法

码流传输至接收端后，首先解析码流分组，在一个 GOP 内，将得到低清视频流、人脸图片流以及这个 GOP 内第一个 I 帧的相关参数流。因为 GOP 相对独立，所以解码流程仅对一个 GOP 展开说明，详细过程如图 3 所示。

通常编码器会缓存一个 GOP 的数据，所以解码端接收的数据在信号时延的基础上滞后一个 GOP。由于编码端强制固定 GOP 大小，保证了时间上的人脸检测频率，也保证了解码端对 GOP 相对独立的解码。解析码流分组后，对图片数据流进行码流片拼图并解码，得到此 GOP 中 I 帧所有完整的人脸图像。对视频流数据，先判断是否为 I 帧，若为 I 帧，则将 I 帧标志 FLAG 置为 TURE，否则置为 FALSE。然后解码得到视频 YUV 序列，再通过分辨率变换到原始视频采集端的分辨率大小。

对于 I 帧解码的序列，需要进行高清人脸恢复处理。根据码流分组中解析的人脸相关参数信息，将对应的人脸填充至 I 帧对应的位置，得到人脸局部高清、背景低清的图像。此时视频序列排布为：在一个 GOP 内，一个具备高清人脸的 I 帧图像后面跟着 GOP−1 帧人脸低清模糊的 P 帧图像。为了提高时域上人脸持续清晰的视觉效果，需要对 P 帧进行人脸增强处理，本文采用 CrossNet 端到端基于参考图像的超分辨率网络（以下简称“超分网络”），实现 P 帧人脸增强[8-9]。时域迭代基于参考的图像超分辨率示意图如图 4 所示。

CrossNet 是完全卷积的深度神经网络，包含图像编码器、交叉尺度变形图层和融合解码器。与普通参考超分辨率（Ref-SR, reference super- resolution）相比，CrossNet 实现了超过 100 倍的加速，适用于视频图像的实时超分辨率（以下简称“超分”）处理。在 GOP 起始位置，输入低分辨率 P1 帧低分辨率（LR, low-resolution）图像和高分辨率 I 帧参考高分辨率（Ref-HR, reference high-resolution）图像，Ref-HR 图像和 LR 图像具有相似视点，但存在显著的分辨率差异。输出 LR 图像的超分辨率（4x

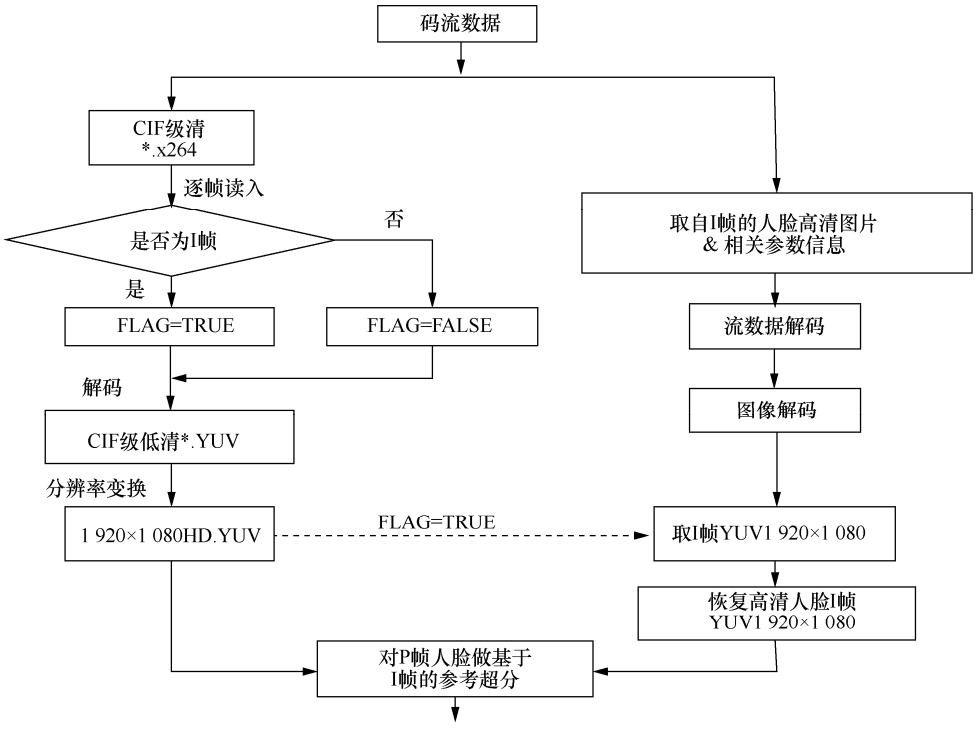

图 3　基于双码流的混合解码框架

或 8x）结果，得到 P1 帧超分辨率结果后，作为下一帧 P2 的高清参考帧 Ref-HR，然后与 P2 帧 LR 图像结合，输入网络、超分辨率后得到 P2 帧结果，以此类推，迭代处理，分别得到后续 P 帧高清人脸图像。由于距离 I 帧较远的 P 帧差异性较大，所以作为改进方案，采取双向迭代的方式。当 P 帧位置小于 GOP/2 时，从最近的前一个 I 帧为起点开始迭代超分处理；当 P 帧位置大于 GOP/2 时，从后面最近的一个 I 帧为起点，向前迭代超分处理。直至处理完所有 P 帧，得到全序列都具备人脸高清、背景低清的视频图像，最后存储或显示输出。

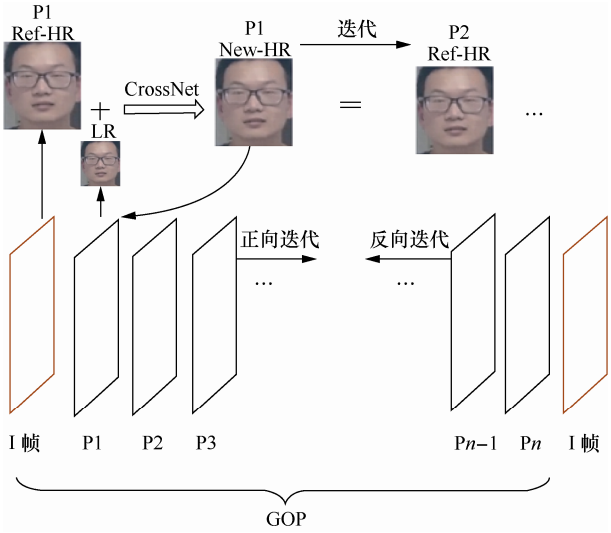

图 4　时域迭代基于参考的图像超分辨率示意图

4　实验结果及分析

为了验证本文设计的系统方案的效果，搭建了实景实验平台。下面从实验结果和数据对比分析两个方面展开说明。

4.1　结果演示

系统分为编码端和解码端两部分，系统演示界面如图 5 所示。视频采集环境为教学楼内过道，普通高清摄像头 1 920×1 080@30fps。在编码端处理平台为 NVIDIA Jetson TK1 开发板，模拟窄带环境为 120～160 kbit/s，解码端在 PC 上处理并显示。实验通过峰值信噪比（PSNR, peak signal to noise ratio）评价编解码后图像的客观质量。

实验过程：通过接入的摄像头实时采集视频数据，并通过 TK1 开发板实时处理，以无线方式发送码流数据。经过窄带传输后，PC 端接收实时数据流，通过特殊设计的解码器解码得到最终视频，保存到本地磁盘或直接调取显示。图 5(a)编码端可以根据实际带宽设置期望的平均码率值，同时在此基础上调整人脸区域的编码质量。解码端实时对比常规低清和经过本方案恢复的局部高清的人脸质量，并相较于原始采集画面，计算人脸局部的 PSNR。主观上，可以看到人脸清晰度得到了显著提升；客观上，人脸画面的 PSNR 值也得到大幅度提高。

4.2　对比分析

为体现变分辨率编码方案在相同带宽下的显著优势，设计如下对比实验。通过 5 路摄像头分别获取 10 段（1 920×1 080）YUV 视频序列：test_01.yuv、test_02.yuv、…、test_10.yuv，选取 10 段计算每个视频中所有人脸的平均 PSNR。在不同带宽限制下，分别经过常规编码和变分辨率方案编码，得到解码画面后，截取人脸局部进行画面质量

(a) 编码端

(b) 解码端

图 5　系统演示界面

对比。人脸画面对比如表 2 所示，人脸质量 PSNR 对比如表 3 所示。

表 2　　人脸画面对比

带宽/(kbit·s^{-1})	采集端	常规编码	变分辨率编码
100			
120			
160			

不同带宽下人脸质量 PSNR 对比如图 6 所示，从图 6 可以看出，当带宽变化时，不影响原始画面的人脸质量，将保持同样的质量传输至解码端；采用变分辨率编码方案时，人脸图像 PSNR 值相较于常规编码平均提升 20 dB 以上，人脸区域基本还原了原始采集端的清晰度。

5　结束语

本文针对当前安防监控低功耗、多接入、广覆盖的需求，与物联网技术相结合，提出了基于 NB-IoT 面向人脸的变分辨率视频监控技术。此方案与传统的基于感兴趣视频编码相比，更大程度上降低了视频编码码率，并且在人脸信息编码方面基本保证了与原始采集信息一致，甚至可以实现无损的效果。根据信息的关注程度、重要程度不同，本文采用不同的分辨率编码处理，设计了根据码率自适应匹配最佳分辨率的算法，与常规监控视频前背景分离的处理方式相比，算法复杂度更低，更适应物联网终端，在普通门禁系统、建筑物内、公共场所具有广泛的适用性。为了保证实时性和可靠性，本文只在 I 帧进行人脸检测处理，在解码端进行人脸增强恢复，未来的工作可以围绕改进码流结构、在 P 帧采用与时域相结合的基于 I 帧的人脸增强等方面展开。

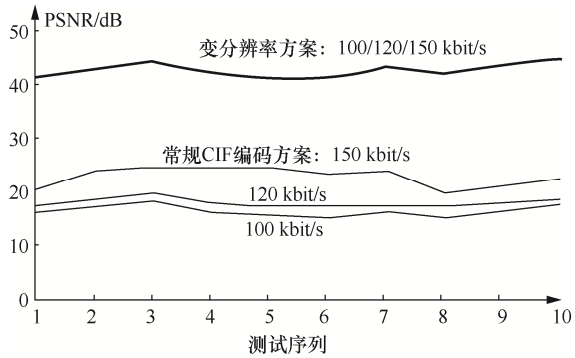

图 6　不同带宽下人脸质量 PSNR 对比

参考文献：

[1] LIU Y, LI Z G, SOH Y C. Region-of-interest based resource allocation for conversational video communication of H.264/AVC[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2008, 18(1):134-139.

[2] XU M, DENG X, LI S, et al. Region-of-interest based conversational HEVC coding with hierarchical perception model of face[J]. IEEE Journal of Selected Topics in Signal Processing, 2014, 8(3):475-489.

[3] ZHANG X, HUANG T, TIAN Y, et al. Background-modeling-based adaptive prediction for surveillance video coding[J]. IEEE Transactions on Image Processing, 2014, 23(2):769-784.

[4] TSAI T H, LIN C Y. Exploring contextual redundancy in improving

表 3　　人脸质量 PSNR 对比

序列	带宽/(kbit·s^{-1})					
	<100		100～130		130～160	
	CIF PSNR/dB	HD PSNR/dB	CIF PSNR/dB	HD PSNR/dB	CIF PSNR/dB	HD PSNR/dB
1	16.32	41.56	17.56	41.56	20.32	41.56
2	17.54	42.96	18.95	42.96	23.78	42.96
3	18.32	44.38	19.78	44.38	24.56	44.38
4	16.41	42.42	18.02	42.42	24.32	42.42
5	15.74	41.21	17.34	41.21	24.78	41.21
6	15.34	41.52	17.21	41.52	23.56	41.52
7	16.42	43.53	17.85	43.53	24.12	43.53
8	15.42	42.26	17.21	42.26	20.06	42.26
9	16.67	43.84	17.94	43.84	21.32	43.84
10	17.73	44.78	18.67	44.78	22.68	44.78

object-based video coding for video sensor networks surveillance[J]. IEEE Transactions on Multimedia, 2012, 14(3):669-682.

[5] NG K, WU Q, CHAN S, et al. Object-based coding for plenoptic videos[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2010, 20(4):548-562.

[6] LIAO L, HU R, XIAO J, et al. An analysis-oriented ROI based coding approach on surveillance video data[C]//Pacific Rim Conference on Multimedia, September 15-16, 2016, Xi'an, China. Berlin: Springer, 2016:428-438.

[7] RANJAN R, CASTILLO C D, CHELLAPPA R. L2-constrained softmax loss for discriminative face verification[J]. ArXiv Preprint, 2017.

[8] ZHENG H, JI M, WANG H, et al. CrossNet: an end-to-end reference-based super resolution network using cross-scale warping[C]//European Conference on Computer Vision. Berlin: Springer, 2018:87-104.

[9] ZHANG Z, WANG Z, LIN Z, et al. Reference-conditioned super-resolution by neural texture transfer[J]. ArXiv Preprint, 2018.

[作者简介]

肖尚武（1994-），男，武汉大学计算机学院国家多媒体软件工程技术研究中心硕士生，主要研究方向为视频编码、图像处理等。

胡瑞敏（1964-），男，教授，武汉大学计算机学院国家多媒体软件工程技术研究中心主任，主要研究方向为音视频编解码、多媒体网络传输、安防应急信息处理、大数据行为分析等。

陈宇（1990-），男，武汉大学计算机学院国家多媒体软件工程技术研究中心博士，主要研究方向为视频编码、图像处理、计算机视觉等。

肖晶（1986-），女，博士，武汉大学计算机学院副教授，任职于国家多媒体软件工程技术研究中心，主要研究方向为视频编码与传输、计算机视觉、多媒体大数据处理与分析等。