



(12) 发明专利申请

(10) 申请公布号 CN 114363617 A

(43) 申请公布日 2022.04.15

(21) 申请号 202210266889.2

(22) 申请日 2022.03.18

(71) 申请人 武汉大学

地址 430072 湖北省武汉市武昌区珞珈山  
街道八一路299号

(72) 发明人 王中元 易鹏 江奎 肖晶  
涂卫平 杨玉红 李登实 肖进胜

(74) 专利代理机构 武汉科皓知识产权代理事务  
所(特殊普通合伙) 42222

代理人 肖明洲

(51) Int.Cl.

H04N 19/137 (2014.01)

H04N 19/139 (2014.01)

H04N 19/42 (2014.01)

H04N 19/70 (2014.01)

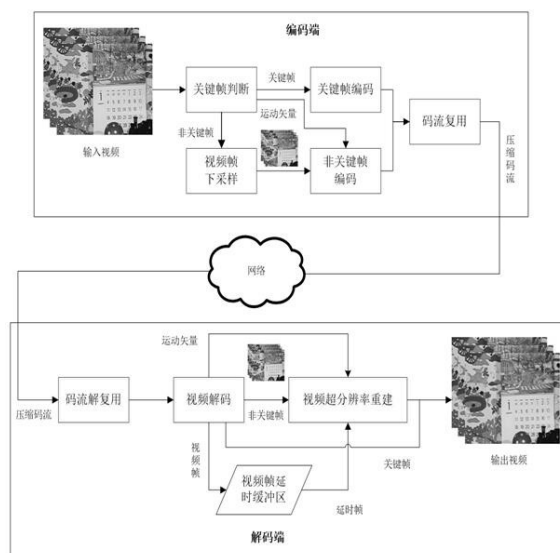
权利要求书3页 说明书6页 附图2页

(54) 发明名称

一种网络轻量级视频流传输方法、系统及设  
备

(57) 摘要

本发明公开了一种网络轻量级视频流传输方法、系统及设备,编码端将原始视频进行空间分辨率下采样,然后对降分辨率视频进行编码,具体包含关键帧判断、视频下采样、关键帧编码、非关键帧编码、码流复用五个步骤;解码端对解码视频通过超分辨率重建恢复视频应有的空间分辨率,具体包含码流解复用、视频解码、视频超分辨率重建三个步骤及一个视频帧延时缓冲区。本发明通过帧间运动复杂度判断关键帧,通过边缘保持的下采样对视频帧进行空间抽样,通过视频超分辨率重建网络重构高分辨率视频帧。本发明基本不牺牲视频清晰度,但显著降低了视频流的数据量,有利于压缩视频的网络传输。



1. 一种网络轻量级视频流传输方法,其特征在于,包括编码过程和解码过程;  
所述编码过程,具体实现包括以下步骤:

步骤1:针对输入视频,进行关键帧判断;

当帧间运动复杂度 $C$ 超过预设门限 $T$ 时,则判断为关键帧,否则判断为非关键帧;

步骤2:针对关键帧,直接进行编码;针对非关键帧,首先进行下采样,然后结合运动矢量进行编码;

步骤3:对关键帧和非关键帧的压缩码流进行传输前的封装,使得接收端能够对二者进行区分;

所述解码过程,具体实现包括以下步骤:

步骤4:将关键帧与非关键帧的码流拆分;

步骤5:视频解码,获得关键帧与非关键帧;

解码的视频帧送入视频帧延时缓冲区缓存起来,用于供给视频超分辨率重建时需要的多个连续视频帧;同时,将解析出的运动矢量参数送入视频超分辨率重建网络中;

步骤6:利用视频超分辨率重建网络,进行视频超分辨率重建,获得超分辨率的非关键帧;

步骤7:根据步骤5解码得到的关键帧和步骤6中超分辨率的非关键帧,恢复出视频并输出。

2. 根据权利要求1所述的网络轻量级视频流传输方法,其特征在于:步骤1中所述帧间

运动复杂度  $C = \sum_{i=1}^N (|xMV_i| + |yMV_i| + \lambda SAD_i)$ , 其中, $N$ 为 $16 \times 16$ 像素尺寸的宏块数目, $xMV_i$ 、 $yMV_i$ 分别表示宏块的水平、竖直方向运动矢量, $SAD_i$ 代表宏块的帧间运动估计误差, $\lambda$ 为预设权重。

3. 根据权利要求1所述的网络轻量级视频流传输方法,其特征在于:步骤2中,所述针对非关键帧进行下采样,采用的是图像边缘保持的下采样方法,首先令 $f_0, f_1, f_2, f_3$ 为空间上连续的一组像素,考察像素点 $f_0, f_1, f_2, f_3$ 之间的相关性,将其分成两组I( $f_0, f_1, f_2$ )和II( $f_1, f_2, f_3$ ),分别求出它们的二阶差分如下:

$$\Delta^2 f_0 = f_2 - 2f_1 + f_0;$$

$$\Delta^2 f_1 = f_3 - 2f_2 + f_1;$$

将相邻3个点的二阶差分的绝对值作为衡量其相关性大小的标准,绝对值越小,相关性越大,反之亦然;选用相关性大的近邻像素进行插值,通过如下二阶插值公式计算插值像素 $I$ :

$$I = \begin{cases} f_0 + \Delta f_0 t + \frac{\Delta^2 f_0}{2!} t(t-1), & |\Delta^2 f_0| \leq |\Delta^2 f_1| \\ f_1 + \Delta f_1(t-1) + \frac{\Delta^2 f_1}{2!} (t-1)(t-2), & |\Delta^2 f_0| > |\Delta^2 f_1| \end{cases};$$

这里,  $t$  为插值像素点  $I$  与源像素点  $f_0$  间的距离,  $1 \leq t \leq 2$ ; 一阶差分  $\Delta f_i = f_{i+1} - f_i \quad i=0, 1$ 。

4. 根据权利要求1所述的网络轻量级视频流传输方法, 其特征在于: 步骤2中, 针对非关键帧首先进行下采样, 然后结合运动矢量进行编码; 其中所述运动矢量, 来自关键帧判断步骤。

5. 根据权利要求1所述的网络轻量级视频流传输方法, 其特征在于: 步骤5中所述视频超分辨率重建网络, 包括Bicubic上采样、运动估计层、运动补偿层、特征提取层、多记忆细节融合层、特征重构层、亚像素放大层、残差叠加操作; 首先通过运动估计和运动补偿, 将输入的低分辨率帧转换为补偿帧, 接着对这些补偿帧依次执行特征提取, 多记忆细节融合, 特征重构和亚像素放大, 最后将Bicubic上采样的帧与像素放大结果相加, 得到重建的高分辨率帧。

6. 根据权利要求5所述的网络轻量级视频流传输方法, 其特征在于: 所述将输入的低分辨率帧转换为补偿帧, 是直接利用解码获得的运动矢量进行运动补偿。

7. 根据权利要求5所述的网络轻量级视频流传输方法, 其特征在于: 所述特征提取, 是采用残差块结构实现, 其中残差块由一系列卷积层组成; 具体过程如下:

$$I_{n+1} = \{I_n, O_n\} = \{I_n, Conv_n(I_n)\};$$

其中,  $Conv_n$  表示残差块中的第  $n$  个卷积层,  $I_n$  和  $O_n$  代表第  $n$  个卷积层的输入和输出。

8. 根据权利要求5所述的网络轻量级视频流传输方法, 其特征在于: 所述多记忆细节融合, 采用多记忆残差块结构实现, 其中多记忆残差块由一系列卷积长短期记忆层组成; 当一个低分辨率帧通过残差块时, 卷积长短期记忆层的细胞状态将保留该帧的特征图信息; 当下一帧进入残差块时, 它将接收从上一帧继承的特征图; 卷积长短期记忆层的处理过程表示为:

$$i_t = I(X_t, H_{t-1}, C_{t-1}),$$

$$f_t = F(X_t, H_{t-1}, C_{t-1}),$$

$$C_t = C(f_t \circ C_{t-1} + i_t \circ (X_t, H_{t-1})),$$

$$o_t = O(X_t, H_{t-1}, C_t),$$

$$H_t = o_t \circ \tanh(C_t),$$

其中,  $i_t, f_t, C_t, o_t$  以及  $H_t$  分别代表输入门、遗忘门、细胞状态、输出门以及隐状态;  $X_t$  表示特征图,  $\circ$  表示哈达玛乘积;  $I(\cdot), F(\cdot), C(\cdot)$  以及  $O(\cdot)$  分别表示输入门、遗忘门、细胞状态以及输出门的函数, 由标准的长短期记忆网络LSTM定义,  $\tanh(\cdot)$  表示双曲正切激活函数;

一个多记忆残差块包含3个卷积长短期记忆层, 每层均采用  $3 \times 3$  大小的卷积核。

9. 一种网络轻量级视频流传输系统, 其特征在于, 包括编码端和解码端;

所述编码端, 包括以下模块:

模块1, 用于针对输入视频, 进行关键帧判断;

当帧间运动复杂度  $C$  超过预设门限  $T$  时, 则判断为关键帧, 否则判断为非关键帧;

模块2, 用于针对关键帧, 直接进行编码; 针对非关键帧, 首先进行下采样, 然后结合运动矢量进行编码;

模块3, 用于对关键帧和非关键帧的压缩码流进行传输前的封装, 使得接收端能够对二者进行区分;

所述解码端, 包括以下模块:

模块4, 用于将关键帧与非关键帧的码流拆分;

模块5, 用于视频解码, 获得关键帧与非关键帧;

解码的视频帧送入视频帧延时缓冲区缓存起来, 用于供给视频超分辨率重建时需要的多个连续视频帧; 同时, 将解析出的运动矢量参数送入视频超分辨率重建网络中;

模块6, 用于利用视频超分辨率重建网络, 进行视频超分辨率重建, 获得超分辨率的非关键帧;

模块7, 用于根据模块5解码得到的关键帧和模块6中超分辨率的非关键帧, 恢复出视频并输出。

10. 一种网络轻量级视频流传输设备, 其特征在于, 包括:

一个或多个处理器;

存储装置, 用于存储一个或多个程序, 当所述一个或多个程序被所述一个或多个处理器执行时, 使得所述一个或多个处理器实现如权利要求1至8中任一项所述的网络轻量级视频流传输方法。

## 一种网络轻量级视频流传输方法、系统及设备

### 技术领域

[0001] 本发明属于多媒体技术领域,涉及一种网络视频流传输方法、系统及设备,具体涉及一种网络轻量级视频流传输方法、系统及设备。

### 背景技术

[0002] 突如其来的新冠肺炎使得全球多数国家采取社会隔离措施,人们的生活、工作、学习、社交突然转移到网络空间,导致网络流量暴增,给远程可视化办公、网课等依赖严重视频的应用带来挑战。由于社会已经习惯和享受到了网络空间工作和学习的优点和便捷性,未来网络化远程办公、学习、会议、社交将成为一种常态,随之而来的因网络长期处于重载引起的流量拥塞也会成为一种常态。

[0003] 提高视频流传输效率的有效途径就是视频编码。自80年代以来,国际标准组织制定了系列视频编码标准,形成了以块为单元的预测加变换的混合编码框架。然而,视频编码技术的压缩效率差不多十年才能提高一倍,国际视频编码标准的演进周期也需要十年。例如,从2003年发布H.264标准到2013年发布H.265标准,经历了整整十年。可见,编码效率的进步明显跟不上视频数据量的增长趋势,为此,必须为高负荷网络视频业务时代寻求新的解决之道。

[0004] 图像超分辨率(Super-Resolution,SR)是指由一幅低分辨率图像或图像序列恢复出高分辨率图像的技术。当前基于深度学习的视频图像的超分辨率技术获得巨大成功。2017年,韩国首尔大学研发了一种增强型深度超分辨率网络,其性能表现超越了之前SR方法。同年,谷歌推出的RAISR(Rapid and Accurate Image Super-Resolution)技术,利用机器学习将低分辨率图像转化为高分辨率图像,能够在节省75%带宽的情况下效果达到甚至超过原图。2019年,华为推出了HiSR(Hisilicon Super-Resolution)超分辨率技术,借助深度学习算法将低分辨率图片转化生成高清图片,并在移动端实现了快速预览高清图片的效果。

### 发明内容

[0005] 本发明的目的在于基于超分辨率技术强大的图像细节重建能力,提供一种网络轻量级视频流传输方法、系统及设备。首先通过空域下采样降低原始视频帧的空间分辨率,对下采样后的视频进行压缩使得压缩码流体量更小,然后通过超分辨率重建恢复视频帧原有的空间分辨率,从而在基本不牺牲视频清晰度的前提下显著减少带宽占用。

[0006] 本发明的方法所采用的技术方案是:一种网络轻量级视频流传输方法,包括编码过程和解码过程;

所述编码过程,具体实现包括以下步骤:

步骤1:针对输入视频,进行关键帧判断;

当帧间运动复杂度 $C$ 超过预设门限 $T$ 时,则判断为关键帧,否则判断为非关键帧;

步骤2:针对关键帧,直接进行编码;针对非关键帧,首先进行下采样,然后结合运

动矢量进行编码；

步骤3:对关键帧和非关键帧的压缩码流进行传输前的封装,使得接收端能够对二者进行区分;

所述解码过程,具体实现包括以下步骤:

步骤4:将关键帧与非关键帧的码流拆分;

步骤5:视频解码,获得关键帧与非关键帧;

解码的视频帧送入视频帧延时缓冲区缓存起来,用于供给视频超分辨率重建时需要的多个连续视频帧;同时,将解析出的运动矢量参数送入视频超分辨率重建网络中;

步骤6:利用视频超分辨率重建网络,进行视频超分辨率重建,获得超分辨率的非关键帧;

步骤7:根据步骤5解码得到的关键帧和步骤6中超分辨率的非关键帧,恢复出视频并输出。

[0007] 本发明的系统所采用的技术方案是:一种网络轻量级视频流传输系统,包括编码端和解码端;

所述编码端,包括以下模块:

模块1,用于针对输入视频,进行关键帧判断;

当帧间运动复杂度 $C$ 超过预设门限 $T$ 时,则判断为关键帧,否则判断为非关键帧;

模块2,用于针对关键帧,直接进行编码;针对非关键帧,首先进行下采样,然后结合运动矢量进行编码;

模块3,用于对关键帧和非关键帧的压缩码流进行传输前的封装,使得接收端能够对二者进行区分;

所述解码端,包括以下模块:

模块4,用于将关键帧与非关键帧的码流拆分;

模块5,用于视频解码,获得关键帧与非关键帧;

解码的视频帧送入视频帧延时缓冲区缓存起来,用于供给视频超分辨率重建时需要的多个连续视频帧;同时,将解析出的运动矢量参数送入视频超分辨率重建网络中;

模块6,用于利用视频超分辨率重建网络,进行视频超分辨率重建,获得超分辨率的非关键帧;

模块7,用于根据模块5解码得到的关键帧和模块6中超分辨率的非关键帧,恢复出视频并输出。

[0008] 本发明的设备所采用的技术方案是:一种网络轻量级视频流传输设备,包括:

一个或多个处理器;

存储装置,用于存储一个或多个程序,当所述一个或多个程序被所述一个或多个处理器执行时,使得所述一个或多个处理器实现所述的网络轻量级视频流传输方法。

[0009] 本发明的优点与积极效果:

(1)本发明通过将原始视频帧空间分辨率抽样后再编码,显著降低了压缩码流的体量,使得网络传输更流畅;同时,通过在解码端采取超分辨率重建恢复出原始分辨率,与直接编解码的传统方式相比,基本上不牺牲视频质量。

[0010] (2)本发明对下采样方法和视频超分辨率重建网络进行了独创设计,设计的边缘



保持的下采样策略以及基于深度学习的超分辨率网络,具有优良的性能,从抽样和重建两个方面保障了目标边缘与纹理细节的还原质量。

### 附图说明

[0011] 图1为本发明实施例的方法流程图;

图2为本发明实施例的视频超分辨率重建网络结构示意图。

### 具体实施方式

[0012] 为了便于本领域普通技术人员理解和实施本发明,下面结合附图及实施例对本发明作进一步的详细描述,应当理解,此处所描述的实施例仅用于说明和解释本发明,并不用于限定本发明。

[0013] 请见图1,本发明提供一种网络轻量级视频流传输方法,包括编码过程和解码过程;

本实施例的编码过程,具体实现包括以下步骤:

步骤1:针对输入视频,进行关键帧判断;

当帧间运动复杂度 $C$ 超过预设门限 $T$ 时,则判断为关键帧,否则判断为非关键帧;

压缩视频的关键帧不涉及帧间预测,因而能阻止帧间预测误差的累积,提高视频解码还原质量。通常有两种开启关键帧编码的策略,定时关键帧和场景切换的强制关键帧。定时关键帧的定时间隔一般在10s;场景切换指视频画面是否出现大幅度剧烈运动。当发生场景切换时,帧间编码的效率并不高,不如采用帧内编码提高码流的容错性能,因此,场景切换时往往采用关键帧编码。场景切换的判断依据是帧间运动复杂度,含运动变化幅度和内容变化强度,运动变化幅度基于累计运动矢量衡量,内容变化强度则通过累计帧差衡量。为此,帧间运动复杂度 $C$ 计算如下:

$$C = \sum_{i=1}^N (|xMV_i| + |yMV_i| + \lambda SAD_i)$$

其中, $N$ 为 $16 \times 16$ 像素尺寸的宏块数目, $xMV_i$ 、 $yMV_i$ 分别表示宏块的水平、竖直方向运动矢量, $SAD_i$ 代表宏块的帧间运动估计误差, $\lambda$ 为预设权重。

[0014] 宏块的运动矢量和帧间误差通过现有的运动估计算法求取。运动矢量在非关键帧编码中也会用到,为了避免不必要的重复运动估计的计算代价,同时将运动矢量提供给非关键帧编码操作。

[0015] 步骤2:针对关键帧,直接进行编码;针对非关键帧,首先进行下采样,然后结合运动矢量进行编码;

考虑到传统的下采样算法容易模糊物体边缘,对目标的清晰度损伤过大,本实施例采取一种图像边缘保持的下采样算法。

[0016] 令 $f_0, f_1, f_2, f_3$ 为空间上连续的一组像素,考察像素点 $f_0, f_1, f_2, f_3$ 之间的相关性,将其分成两组I( $f_0, f_1, f_2$ )和II( $f_1, f_2, f_3$ ),分别求出它们的二阶差分如下:

$$\Delta^2 f_0 = f_2 - 2f_1 + f_0;$$

$$\Delta^2 f_1 = f_3 - 2f_2 + f_1;$$

将相邻3个点的二阶差分的绝对值作为衡量其相关性大小的标准，绝对值越小，相关性越大，反之亦然。相关性越大，表明像素处在同质图像区域的可能性越大。因此，选用相关性大的近邻像素进行插值更为合理。基于这一原理，通过如下二阶插值公式计算插值像素 $I$ ：

$$I = \begin{cases} f_0 + \Delta f_0 t + \frac{\Delta^2 f_0}{2!} t(t-1), & |\Delta^2 f_0| \leq |\Delta^2 f_1| \\ f_1 + \Delta f_1(t-1) + \frac{\Delta^2 f_1}{2!} (t-1)(t-2), & |\Delta^2 f_0| > |\Delta^2 f_1| \end{cases}$$

这里， $t$ 为插值像素点 $I$ 与源像素点 $f_0$ 间的距离， $1 \leq t \leq 2$ ；一阶差分

$$\Delta f_i = f_{i+1} - f_i \quad i=0, 1。$$

[0017] 本实施例中，关键帧和非关键帧编码，采用成熟的H.264或H.265编码技术，其中关键帧对应的是编码标准中的帧内帧，非关键帧对应的是预测帧。关键帧用原始空间分辨率编码，非关键帧用降低的空间分辨率编码。非关键帧编码模块不再重复计算运动矢量，需要的运动矢量来自关键帧判断步骤。

[0018] 步骤3：对关键帧和非关键帧的压缩码流进行传输前的封装，使得接收端能够对二者进行区分；本实施例在网络传输带宽资源紧张时，优先封装关键帧。

[0019] 本实施例的解码过程，具体实现包括以下步骤：

步骤4：将关键帧与非关键帧的码流拆分；

本实施例将关键帧与非关键帧的码流拆分，便于后面对非关键帧进行解码及超分辨率重建，而对非关键帧则直接解码输出。

[0020] 步骤5：视频解码，获得关键帧与非关键帧；

本实施例根据压缩码流的编码标准，执行对应标准的解码。解码的视频帧送入视频帧延时缓冲区缓存起来，用于供给视频超分辨率重建时需要的多个连续视频帧。同时，将视频解码器解析出的运动矢量参数送入视频超分辨率重建网络，以节省计算量。

[0021] 本实施例的视频超分辨率重建网络负责恢复解码出的非关键帧的空间分辨率，以弥补编码端对非关键帧抽样引起的细节损失。采用基于深度学习的视频超分辨率方案，从一系列相邻的低分辨率帧重建出对应的高分辨率帧。

[0022] 请见图2，本实施例的视频超分辨率重建网络，包括Bicubic上采样、运动估计层、运动补偿层、特征提取层、多记忆细节融合层、特征重构层、亚像素放大层、残差叠加操作。首先通过运动估计和运动补偿，将输入的低分辨率帧转换为补偿帧，接着对这些补偿帧依次执行特征提取，多记忆细节融合，特征重构和亚像素放大，最后将Bicubic上采样的帧与像素放大结果相加，得到重建的高分辨率帧。



[0023] 本实施例的运动估计和补偿用于处理连续低分辨率帧之间的时间相关性。这里不执行新的运动估计,而是直接利用解码获得的运动矢量进行运动补偿,从而节省复杂运动估计的计算开销。

[0024] 本实施例的特征提取功能采用残差块结构实现,这种残差块由一系列卷积层组成。过程描述如下:

$$I_{n+1} = \{I_n, O_n\} = \{I_n, Conv_n(I_n)\};$$

其中, $Conv_n$ 表示残差块中的第 $n$ 个卷积层, $I_n$ 和 $O_n$ 代表第 $n$ 个卷积层的输入和输出。残差块会保留来自先前卷积层的信息,并将其传递给所有后续的卷积层。特征重构中使用的残差块同样是特征提取残差块这种结构,但它们在网络中的位置不同,因而所起的作用也不同。

[0025] 本实施例的多记忆细节融合功能采用多记忆残差块结构实现,这种残差块由一系列卷积长短期记忆层组成。当一个低分辨率帧通过残差块时,卷积长短期记忆层的细胞状态将保留该帧的特征图信息;当下一帧进入残差块时,它将接收从上一帧继承的特征图。通过这种方式,卷积长短期记忆层学会该记住哪部分有效信息,以及该遗忘哪些无效信息。卷积长短期记忆层的处理过程表示为:

$$i_t = I(X_t, H_{t-1}, C_{t-1}),$$

$$f_t = F(X_t, H_{t-1}, C_{t-1}),$$

$$C_t = C(f_t \circ C_{t-1} + i_t \circ (X_t, H_{t-1})),$$

$$o_t = O(X_t, H_{t-1}, C_t),$$

$$H_t = o_t \circ \tanh(C_t),$$

其中, $i_t, f_t, C_t, o_t$ 以及 $H_t$ 分别代表输入门、遗忘门、细胞状态、输出门以及隐状态; $X_t$ 表示特征图, $\circ$ 表示哈达玛乘积; $I(\cdot), F(\cdot), C(\cdot)$ 以及 $O(\cdot)$ 分别表示输入门、遗忘门、细胞状态以及输出门的函数,由标准的长短期记忆网络LSTM定义, $\tanh(\cdot)$ 表示双曲正切激活函数;

一个多记忆残差块包含3个卷积长短期记忆层,每层均采用 $3 \times 3$ 大小的卷积核,但数量不同。由于卷积长短期记忆层会消耗较大的GPU显存(大约是普通卷积层的4倍),所以首先将输入的特征图从64层映射到16层,以降低GPU显存成本和计算复杂度。

[0026] 卷积神经网络中,最常见的放大特征图的方法是卷积转置,Caballero等人提出了一种亚像素放大的方法来放大特征图。本实施例选择亚像素放大,因为这种方法需要较少的计算成本,并且在类似的网络中性能更好。

[0027] 步骤6:利用视频超分辨率重建网络,进行视频超分辨率重建,获得超分辨率的非关键帧;

步骤7:根据步骤5解码得到的关键帧和步骤6中超分辨率的非关键帧,恢复出视频

并输出。

[0028] 应当理解的是,上述针对较佳实施例的描述较为详细,并不能因此而认为是对本发明专利保护范围的限制,本领域的普通技术人员在本发明的启示下,在不脱离本发明权利要求所保护的范围情况下,还可以做出替换或变形,均落入本发明的保护范围之内,本发明的请求保护范围应以所附权利要求为准。

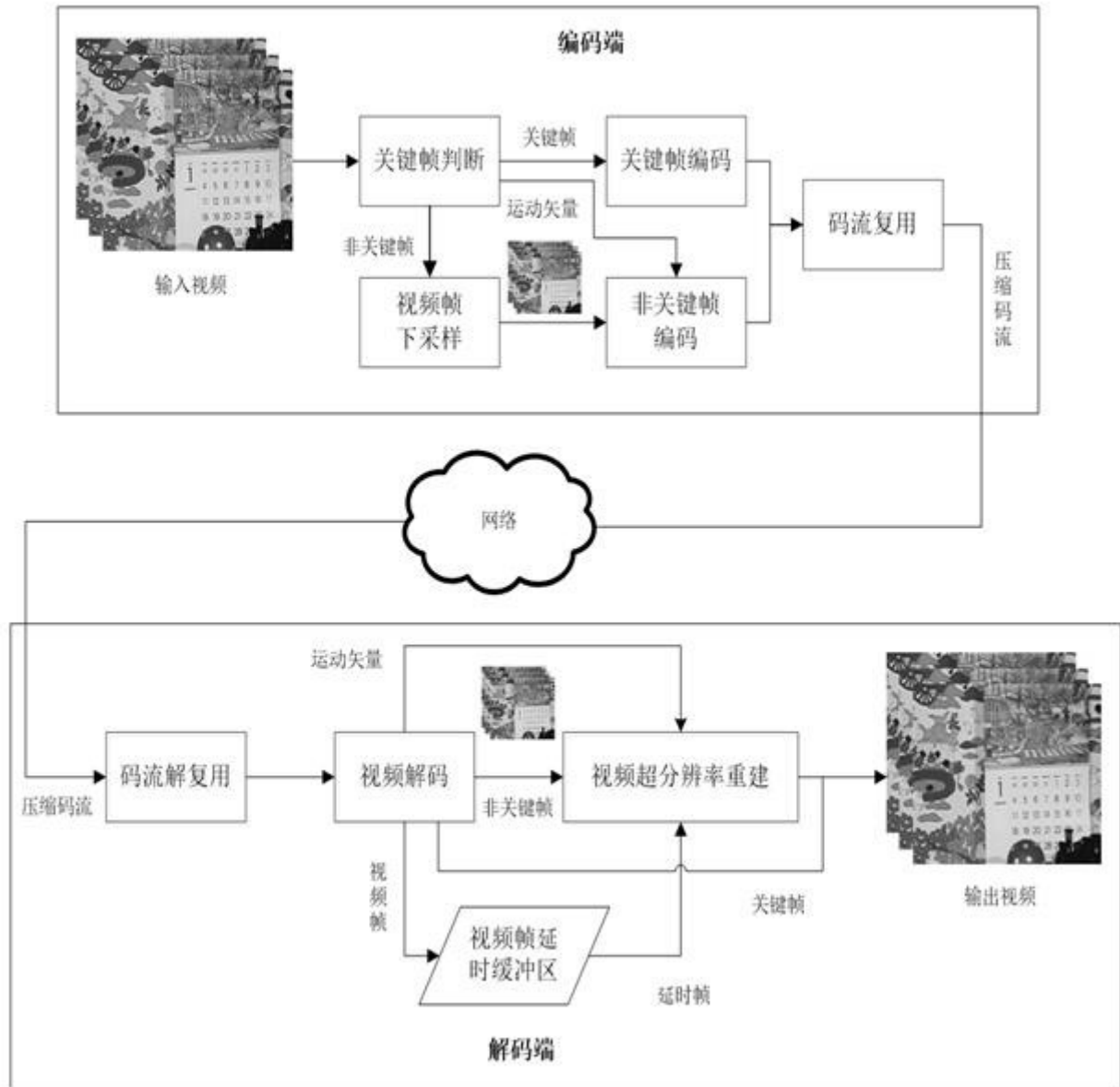


图 1

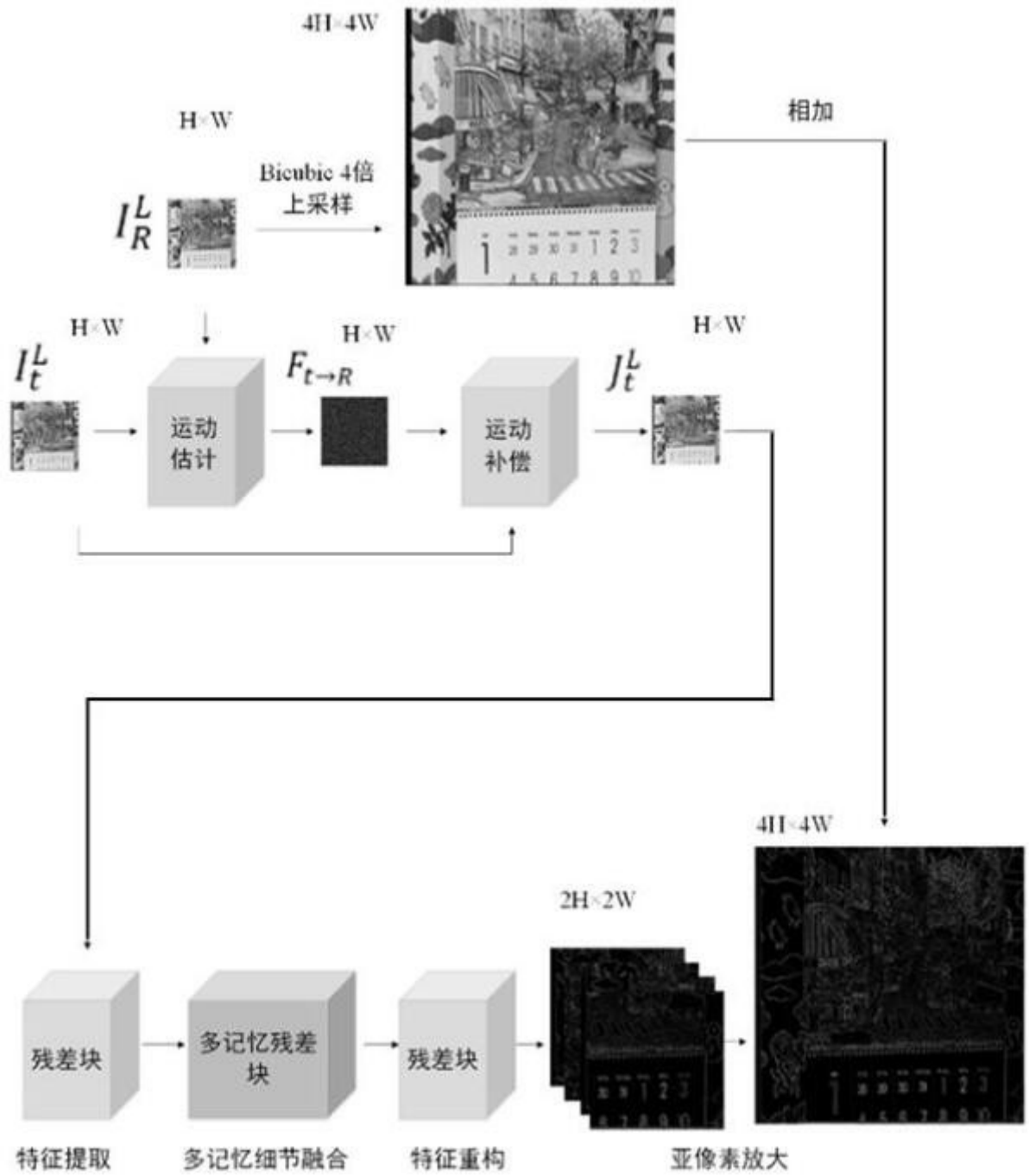


图 2