



(12) 发明专利

(10) 授权公告号 CN 103379266 B

(45) 授权公告日 2016. 01. 20

(21) 申请号 201310280431. 3

(22) 申请日 2013. 07. 05

(73) 专利权人 武汉烽火众智数字技术有限公司

地址 430074 湖北省武汉市洪山区邮科院路 88 号

(72) 发明人 贺波涛 余少华 王峰 杨波 李华民

(74) 专利代理机构 武汉宇晨专利事务所 42001 代理人 黄瑞棠

(51) Int. Cl. H04N 5/225(2006. 01) H04N 7/18(2006. 01)

(56) 对比文件 CN 102724485 A, 2012. 10. 10, CN 103020624 A, 2013. 04. 03,

CN 102982311 A, 2013. 03. 20, CN 101778260 A, 2010. 07. 14, CN 102724485 A, 2012. 10. 10, CN 101902617 A, 2010. 12. 01, US 2003/0081564 A1, 2003. 05. 01, US 2005/0140787 A1, 2005. 06. 30,

审查员 黄益超

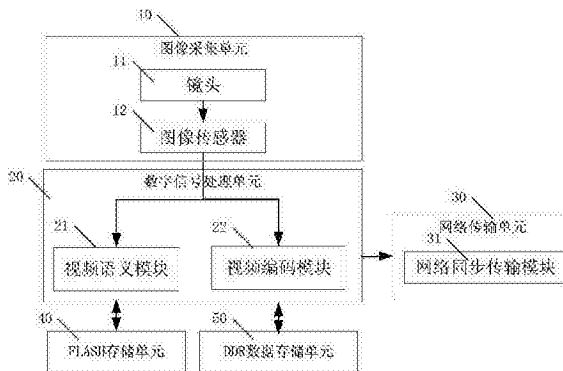
权利要求书3页 说明书6页 附图3页

(54) 发明名称

一种具有视频语义分析功能的高清网络摄像机

(57) 摘要

本发明公开了一种具有视频语义分析功能的高清网络摄像机,涉及图像识别技术领域。本摄像机的结构是:镜头与图像传感器通过结构件实现光连接;数字信号处理单元分别与图像传感器、FLASH 存储单元、DDR 数据存储单元和网络传输单元连接。本方法包括视频语义元数据生成及同步传输。本发明将视频语义分析模块前置,实现单台摄像机的视频语义分析功能,形成实时的全时间段的语义分析,同时将视频语义元数据与视频数据同步传输到后端平台,后端平台在此基础上能实现海量视频的快速检索和定位,节约海量视频目标对象查找过程中的大量人力和物力;适用于大规模的高清视频监控应用。



1. 基于摄像机的视频语义元数据生成及同步传输的方法，
所述的摄像机包括图像采集单元 (10)、数字信号处理单元 (20)、网络传输单元 (30)、FLASH 存储单元 (40) 和 DDR 数据存储单元 (50)；

图像采集单元 (10) 包括镜头 (11) 和图像传感器 (12)；

数字信号处理单元 (20) 包括语义分析模块 (21) 和视频编码模块 (22)；

镜头 (11) 与图像传感器 (12) 通过结构件实现光连接；

数字信号处理单元 (20) 分别与图像传感器 (12)、FLASH 存储单元 (40)、DDR 数据存储单元 (50) 和网络传输单元 (30) 连接；

所述的方法是：

1) 视频语义元数据生成：

①图像采集单元采集原始视频图像通过数模转换后将采集到的数字信号传输给语义分析模块；

②语义分析模块接收采集的连续原始视频图像序列；

③原始视频图像序列缩放到指定处理大小；

④利用缩放后的图像数据初始化背景模型；

⑤根据背景差分法检测运动目标，并且更新背景模型；

⑥对检测出来的运动目标进行形态学处理；

⑦对处理过的运动目标进行跟踪；

⑧计算运动目标的特征，特征包括类别、颜色、大小、运动方向、运动速度；

⑨提取目标特征后根据目标特征数据字典，实现目标语义数字化；

2) 同步传输：

①图像采集单元采集原始视频图像，且记录当前图像帧的时间戳，将图像数据和时间戳传输给视频语义模块和视频编码模块；

②视频语义模块接收到连续的图像序列，将视频目标进行语义转换，输出其对应特征的特征值；

③视频编码模块接收到连续的图像序列，进行编码，生成压缩后的 H. 264 视频流；

④通过时戳实现视频语义元数据与编码数据的对应关系，并将对应时戳的视频数据和语义数据传入同步传输模块；

⑤同步传输模块采用标准的 RTP 协议进行流媒体传输，同时通过每帧最后一个 RTP 分包的包头扩展位封装传入的视频语义元数据，实现视频数据与视频语义元数据的同步传输；

其特征在于：

所述语义分析模块 (21) 的工作流程是：

①视频图像缩放 (211)

将原始视频图像缩放到算法分析所需要的分辨率，即 352×288 像素；

②初始化背景模型 (212)

根据缩放后的前 n 帧视频图像初始化背景模型， n 是自然数， $5 \leq n \leq 10$ ；

③检测运动目标 (213)

根据当前的背景模型和当前的图像，进行高斯差分，获得运动目标；

④背景模型更新 (214)

利用高斯背景建模方法更新当前的背景模型；

⑤运动目标形态学处理 (215)

对检测出来的运动目标进行形态学处理,包括腐蚀膨胀处理,以及连通域标记,从而得到完整的运动目标；

⑥运动目标跟踪 (216)

利用最近邻方法对运动目标进行目标的跟踪；

⑦计算运动目标特征 (217)

所述的运动目标特征包括目标的类别、颜色、大小、运动方向和运动速度；

A、所述目标的类别其计算流程如下：

a、运用 canny 算子提取目标图像轮廓；

b、计算图像轮廓处的梯度方向,将梯度方向分成上、下、左、右四类,分别统计四类方向的总数,记为 $g(i)$, $0 < i < 5$, i 为整数；

c、分别对 $g(i)$ 归一化处理,归一化处理后的数据记为 $x(i)$, 公式为
$$x(i) = \frac{g(i)}{\sum_{i=1}^4 g(i)}$$
,

$0 < i < 5$, i 为整数；

d、计算目标的长宽比和占空比,分别记为 $x(5)$ 和 $x(6)$,其中占空比指目标的实际面积与外接矩形之比；

e、计算目标轮廓的长度与目标轮廓所包围的面积之比,记为 $x(7)$ ；

f、将 $x(j)$ 组成一个特征向量,并且对特征向量进行归一化处理,归一化后的特征向量记为 $Y(j)$, $0 < j < 8$, j 为整数；

g、利用支持向量机对已经计算好的特征向量 $Y(j)$, $0 < j < 8$, j 为整数,进行分类,达到目标分类的目的；

B、所述目标的颜色其计算流程如下：

a、将原始的目标图像 RGB 数据转换成 HSV 数据；

b、将目标的颜色划分成九类,分别是红、橙、黄、绿、青、蓝、紫、黑、白；

c、判断目标每个像素的颜色,判断的准则如下：

$$\text{颜色} = \begin{cases} \text{红} & h \in (330, 360] \cup [0, 22] \\ \text{橙} & h \in (22, 45] \\ \text{黄} & h \in (45, 70] \\ \text{绿} & h \in (70, 155] \\ \text{青} & h \in (155, 186] \\ \text{蓝} & h \in (186, 278] \\ \text{紫} & h \in (278, 330] \\ \text{白} & s \in [0, 0.25] \text{ 并且 } v \in (0.2, 1] \\ \text{黑} & h \in (0.25, 1] \text{ 并且 } v \in (0, 0.2] \end{cases}$$

d、统计目标中每类颜色对应像素个数之和,将包含像素最多的颜色类别作为目标的最终颜色,判断完成；

C、所述目标的大小通过目标面积计算得到；

D、运动方向和运动速度均通过目标跟踪轨迹进行计算得到。

一种具有视频语义分析功能的高清网络摄像机

技术领域

[0001] 本发明涉及图像识别技术领域,尤其涉及一种具有视频语义功能的高清网络摄像机。

背景技术

[0002] 随着安防视频监控行业的发展,网络化、高清化和智能化已成为这个行业的发展趋势,随着智慧城市的推进,监控摄像头已经遍布大街小巷,产生海量的非结构化的图像数据,仅仅将视频图像记录下来只是第一步,有了相关视频图像不等于就找到了目标信息;在海量非结构化的视频图像数据中查找视频、分析视频的工作常常会耗用大量的时间和人力,能否在海量视频中更方便和更省力地查找到相关信息,如何从杂乱的非结构化数据中提取准确无误的关键信息;同时,如何高效存储和访问利用这些信息都是目前亟待解决的问题;在现有技术条件下,要实现海量视频的检索,通常采用后台手动加注和自动生成相关目标低级语义,该方式需要耗费大量的计算量和人力且系统结构复杂,而该方式只能针对部分重点区域或重点时间段视频,不能对系统内所有视频进行语义转换。

发明内容

[0003] 本发明的目的就在于克服现有技术存在缺点和不足,提供了一种具有视频语义分析功能的高清网络摄像机。

[0004] 本发明的目的是这样实现的:

[0005] 将相关图像目标提取算法、目标特征(类别、颜色、大小、运动方向和速度等)分析算法植入到高清网络摄像机中,在视频压缩编码的同时生成视频语义元数据,使摄像机具有目标对象语义提取功能,同时采用基于时戳(ts)的视频数据与视频语义元数据的同步生成和传输方法将视频数据与视频语义元数据同步传输到后端平台,从而提供一种全新概念的高清网络摄像机,它输出的将不仅仅是一幅幅图像,而且同时还输出图像中被结构化语义描述的目标对象。

[0006] 具体技术方案为:

[0007] 一、具有视频语义分析功能的高清网络摄像机(简称摄像机)

[0008] 本摄像机包括图像采集单元、数字信号处理单元、网络传输单元、FLASH 存储单元和 DDR 数据存储单元;

[0009] 图像采集单元包括镜头和图像传感器;

[0010] 数字信号处理单元包括语义分析模块和视频编码模块;

[0011] 镜头与图像传感器通过结构件实现光连接;

[0012] 数字信号处理单元分别与图像传感器、FLASH 存储单元、DDR 数据存储单元和网络传输单元连接。

[0013] 二、视频语义元数据生成及同步传输的方法(简称方法)

[0014] 本方法基于上述的具有视频语义功能的高清网络摄像机,包括视频语义元数据生

成和同步传输；

[0015] 1) 视频语义元数据生成：

[0016] ①图像采集单元采集原始视频图像通过数模转换后将采集到的原始视频图像序列传输给语义分析模块；

[0017] ②语义分析模块接收采集的连续原始视频图像序列；

[0018] ③原始视频图像序列缩放到指定处理分辨率 352×288 ；

[0019] ④利用缩放后的图像数据初始化背景模型；

[0020] ⑤根据背景差分法(背景模型和当前帧的差)检测运动目标,并且更新背景模型；

[0021] ⑥对检测出来的运动目标进行形态学处理；

[0022] ⑦对处理过的运动目标进行跟踪；

[0023] ⑧计算运动目标的特征,特征包括类别(人、车和物)、颜色、大小、运动方向和运动速度；

[0024] ⑨提取目标特征后根据目标特征数据字典,实现目标语义数字化。

[0025] 2) 同步传输：

[0026] ①图像采集单元采集原始视频图像,且记录当前图像帧的时间戳,将图像数据和时间戳传输给视频语义模块和视频编码模块；

[0027] ②视频语义模块接收到连续的图像序列,将视频目标进行语义转换,输出其对应特征的特征值；

[0028] ③视频编码模块接收到连续的图像序列,进行编码,生成压缩后的 H. 264 视频流；

[0029] ④通过时戳实现视频语义元数据与编码数据的对应关系,并将时戳相同的视频数据和语义数据传入同步传输模块；

[0030] ⑤同步传输模块采用标准的 RTP 协议进行流媒体传输,同时通过每帧最后一个 RTP 分包的包头扩展位封装传入的视频语义元数据,实现视频数据与视频语义元数据的同步传输。

[0031] 由此可见,本发明是通过上述流程提供了一种全新概念的高清网络摄像机,它输出的将不仅仅是一幅幅图像,而且同时还输出图像中被结构化语义描述的目标对象(比如人、车、物等目标及其特征),同时很好地解决了视频语义元数据和视频数据的同步生成和传输问题。

[0032] 本发明具有以下优点和积极效果：

[0033] ①将视频语义分析模块前置,实现单台摄像机的视频语义分析功能,形成实时的全时间段的语义分析,同时将视频语义元数据与视频数据同步传输到后端平台,后端平台在此基础上能实现海量视频的快速检索和定位,节约海量视频目标对象查找过程中的大量人力和物力；

[0034] ②很好地解决了视频语义元数据和视频数据的同步生成和传输问题,便于后端平台的存储、目标的查找和准确定位。

[0035] ③与采用通过后端平台进行语义转换的方案相比,将计算单元前置,具有成本低的优点,且节约大量服务器资源,具有节能环保的积极作用；

[0036] 适用于大规模的高清视频监控应用。

附图说明

- [0037] 图 1 为摄像机的结构方框图；
- [0038] 图 2 为视频数据与视频语义同步生成流程图；
- [0039] 图 3 为视频语义分析模块工作流程图；
- [0040] 图 4 为网络同步传输模块工作流程图。
- [0041] 图中：
- [0042] 10—图像采集单元；
- [0043] 11—镜头,12—图像传感器；
- [0044] 20—数字信号处理单元，
- [0045] 21—语义分析模块,22—视频编码模块；
- [0046] 30—网络传输单元,31—网络同步传输模块；
- [0047] 40—FLASH 存储单元；
- [0048] 50—DDR 数据存储单元。
- [0049] 英译汉：
- [0050] RTP :实时传送协议；
- [0051] FLASH 存储 :闪存；
- [0052] DDR :双倍速率同步动态随机存储；
- [0053] Ts :时戳。

具体实施方式

- [0054] 下面结合附图和实施例详细说明：
- [0055] 一、摄像机
- [0056] 1、总体
- [0057] 如图 1,本摄像机包括图像采集单元 10、数字信号处理单元 20、网络传输单元 30、FLASH 存储单元 40 和 DDR 数据存储单元 50；
- [0058] 图像采集单元 10 包括镜头 11 和图像传感器 12；
- [0059] 数字信号处理单元 20 包括语义分析模块 21 和视频编码模块 22；
- [0060] 镜头 11 与图像传感器 12 通过结构件实现光连接；
- [0061] 数字信号处理单元 20 分别与图像传感器 12、FLASH 存储单元 40、DDR 数据存储单元 50 和网络传输单元 30 连接。
- [0062] 工作机理：
- [0063] 镜头 11 实现在图像传感器 12 表面成像，图像传感器 12 将转换后的原始的高清视频数字信号传送给数字信号处理单元 20。数字信号处理单元 20 实现视频图像的语义分析和压缩编码并通过网络传输单元 30 实现网络数据接口交换。FLASH 存储单元 40 和 DDR 数据存储单元 50 负责数字信号处理单元 20 的程序和数据保存。
- [0064] 2、功能块
- [0065] 1) 图像采集单元 10
- [0066] (1) 镜头 11
- [0067] 镜头 11 采用安防行业通用的镜头；负责光学系统成像。

[0068] (2) 图像传感器 12

[0069] 图像传感器 12 采用索尼公司的 IMX122 等高清图像传感器 ;负责将光学成像转换为原始的高清视频信号。

[0070] 2) 数字信号处理单元 20

[0071] 数字信号处理单元 20 采用美国 TI 公司的 TMS320DM8168 芯片,植入自定义的语义分析模块 21 和视频编码模块 22 ;负责图像的语义分析和压缩编码。

[0072] 如图 2,图像采集单元 10 将采集到的图像序列同时发送给语义分析模块 21 和视频编码模块 22,且同时携带该图像序列的时戳 (ts) 信息 ;

[0073] 语义分析模块 21 和视频编码模块 22 对图像序列进行同步处理,同时输出对应时戳的视频压缩数据和目标对象语义元数据,通过时戳实现视频编码数据和视频语义元数据的同步生成。

[0074] 3) 网络传输单元 30

[0075] 网络传输单元 30 采用 Atheros 公司的 AR8033 芯片,负责网络电平转换和网络数据接口交换,植入自定义的网络同步传输模块 31。

[0076] 4) FLASH 存储单元 40

[0077] FLASH 存储单元 40 采用美国 MICRON 公司的 MT29F2G16 芯片 ;负责数字信号处理单元 20 的程序保存和基本配置数据的保存。

[0078] 5) DDR 数据存储单元 50

[0079] DDR 数据存储单元 50 采用三星公司 K4B1G1646 芯片 ;负责数字信号处理单元 20 的运行数据的保存。

[0080] 二、方法

[0081] 1、语义分析模块 21

[0082] 如图 3,语义分析模块 21 其软件是由依次交互的视频图像缩放 211、初始化背景模型 212、检测运动目标 213、背景模型更新 214、运动目标形态学处理 215、运动目标跟踪 216、计算运动目标特征 217 和语义化输出算法模块 218 组成。

[0083] 具体地说,语义分析模块 21 的工作流程是 :

[0084] ①视频图像缩放 211

[0085] 将原始视频图像缩放到算法分析所需要的分辨率,即 352×288 像素 ;

[0086] ②初始化背景模型 212

[0087] 根据缩放后的前 n 帧视频图像初始化背景模型,n 为整数, $5 \leq n \leq 20$;

[0088] ③检测运动目标 213

[0089] 根据当前的背景模型和当前的图像,进行高斯差分,获得运动目标 ;

[0090] ④背景模型更新 214

[0091] 利用高斯背景建模方法更新当前的背景模型 ;

[0092] ⑤运动目标形态学处理 215

[0093] 对检测出来的运动目标进行形态学处理,包括腐蚀膨胀处理,以及连通域标记,从而得到完整的运动目标 ;

[0094] ⑥运动目标跟踪 216

[0095] 利用最近邻方法对运动目标进行目标的跟踪 ;

[0096] ⑦计算运动目标特征 217

[0097] 所述的运动目标特征包括目标的类别(人、车和物)、颜色、大小、运动方向和运动速度；

[0098] A、所述目标的类别其计算流程如下：

[0099] a、运用 canny 算子提取目标图像轮廓；

[0100] b、计算图像轮廓处的梯度方向，将梯度方向分成上、下、左、右四类，分别统计四类方向的总数，记为 $g(i)$, $0 < i < 5$, i 为整数；

[0101] c、分别对 $g(i)$ 归一化处理，归一化处理后的数据记为 $x(i)$ ，公式为

$$x(i) = \frac{g(i)}{\sum_{i=1}^4 g(i)}, \quad 0 < i < 5, i \text{ 为整数};$$

[0102] d、计算目标的长宽比和占空比，分别记为 $x(5)$ 和 $x(6)$ ，其中占空比指目标的实际面积与外接矩形之比；

[0103] e、计算目标轮廓的长度与目标轮廓所包围的面积之比，记为 $x(7)$ ；

[0104] f、将 $x(i)$ 组成一个特征向量，并且对特征向量进行归一化处理，归一化后的特征向量记为 $Y(i)$, $0 < i < 8$, i 为整数；

[0105] g、利用支持向量机（简称 SVM）对已经计算好的特征向量 $Y(i)$, $0 < i < 8$, i 为整数，进行分类，达到目标分类的目的；

[0106] B、所述目标的颜色其计算流程如下：

[0107] a、将原始的目标图像 RGB（红、绿、蓝）数据转换成 HSV（色调、饱和度、亮度）数据；

[0108] b、将目标的颜色划分成九类，分别是红、橙、黄、绿、青、蓝、紫、黑、白；

[0109] c、判断目标每个像素的颜色，判断的准则如下：

[0110]

$$\text{颜色} = \begin{cases} \text{红} & h \in (330, 360] \cup [0, 22] \\ \text{橙} & h \in (22, 45] \\ \text{黄} & h \in (45, 70] \\ \text{绿} & h \in (70, 155] \\ \text{青} & h \in (155, 186] \\ \text{蓝} & h \in (186, 278] \\ \text{紫} & h \in (278, 330] \\ \text{白} & s \in [0, 0.25] \text{ 并且 } v \in (0.2, 1] \\ \text{黑} & h \in (0.25, 1] \text{ 并且 } v \in (0, 0.2] \end{cases}$$

[0111] d、统计目标中每类颜色对应像素个数之和，将包含像素最多的颜色类别作为目标的最终颜色，判断完成；

[0112] C、所述目标的大小通过目标面积计算得到；

[0113] D、运动方向和运动速度均通过目标跟踪轨迹进行计算得到；

[0114] ⑧提取目标特征后根据目标特征数据字典，实现目标语义数字化输出 218。

[0115] 2、视频编码模块 22

[0116] 视频编码模块 22 是一种常用的功能模块，其功能是将图像采集单元 10 传入的数据通过视频编码算法进行 H264 编码生成标准 H264 视频数据。

[0117] 3、网络同步传输模块 31

[0118] 网络同步传输模块 31 的功能是：摄像机收到流媒体传输请求时，将视频数据和视频语义元数据同步传给网络同步传输模块 31，网络同步传输模块 31 将相同时戳的视频语义元数据和视频帧数据采用标准 RTP 方式打包同步发送。

[0119] 如图 4，网络同步传输模块 31 的工作流程如下：

[0120] ①输入相同时戳的视频语义元数据和视频帧数据 -41；

[0121] ②读取视频帧数据，生成标准 RTP 包 -42；

[0122] ③判断 RTP 包是否为视频帧最后一个分包，是则进入下一步骤④，否则跳转到步骤⑤；

[0123] 判断依据为未发送的视频帧数据长度是否 $\leq N$ ， N 为自然数， $N < 1480$ ，如果是则生成的 RTP 包为该帧最后一个 RTP 分包；

[0124] ④将视频语义元数据封装在该 RTP 包的包头扩展位 -44；

[0125] ⑤发送 RTP 包数据，完成视频帧数据与视频语义元数据的同步传输 -45。

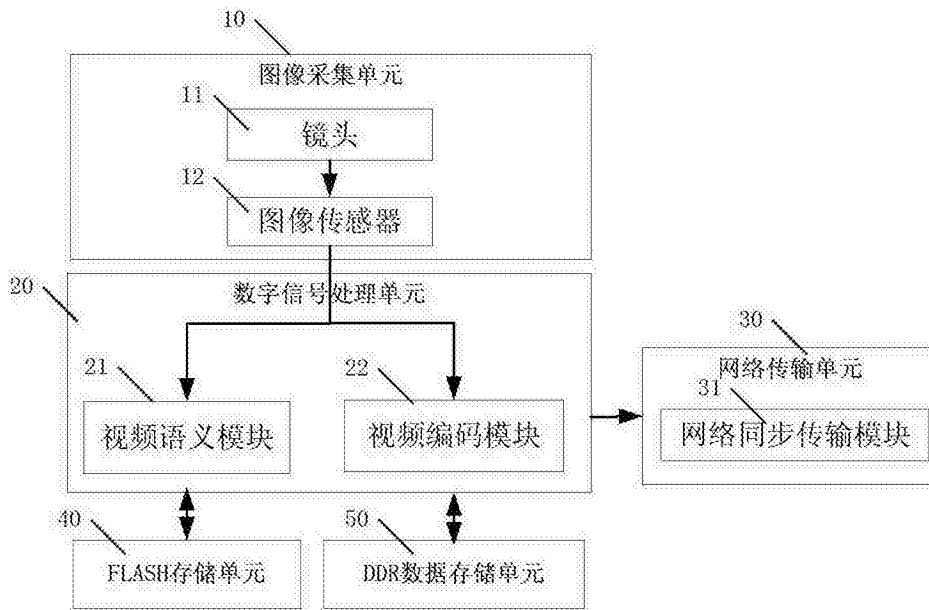


图 1

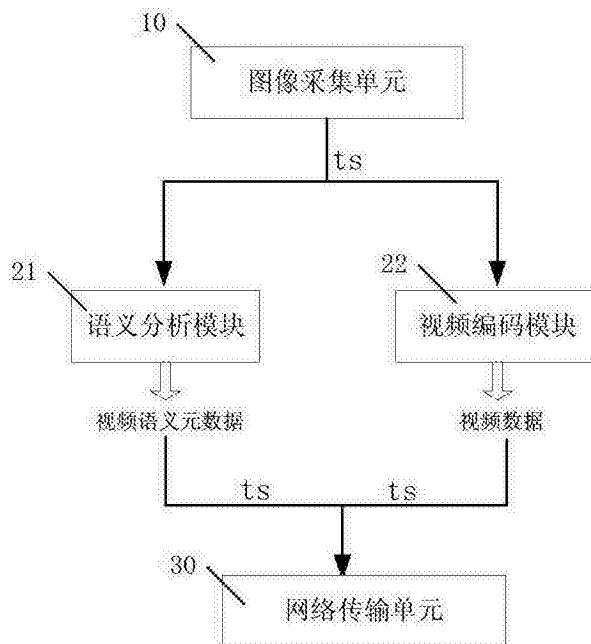


图 2

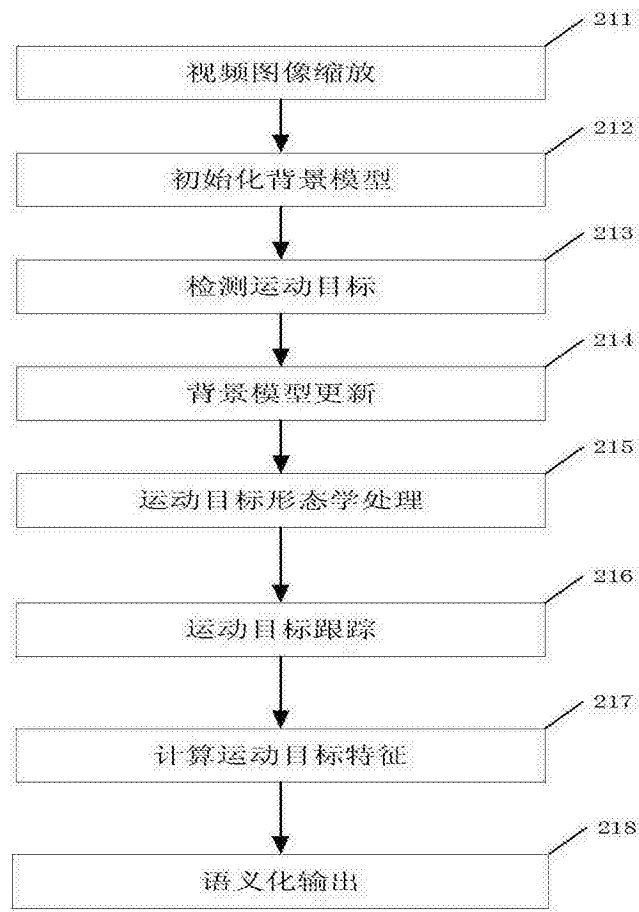


图 3

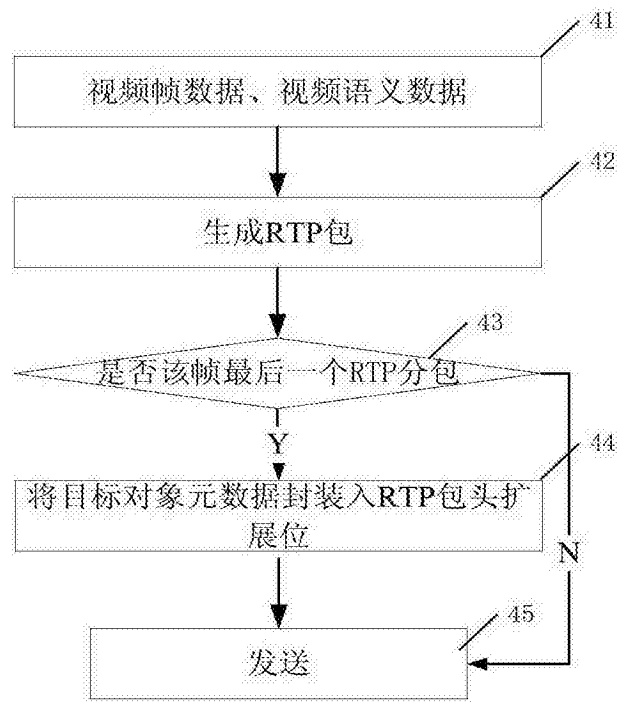


图 4