



(12) 发明专利

(10) 授权公告号 CN 112287906 B

(45) 授权公告日 2021.04.09

(21) 申请号 202011509884.5

G06K 9/62 (2006.01)

(22) 申请日 2020.12.18

G06N 3/04 (2006.01)

(65) 同一申请的已公布的文献号

G06N 3/08 (2006.01)

申请公布号 CN 112287906 A

G06T 7/246 (2017.01)

(43) 申请公布日 2021.01.29

(56) 对比文件

(73) 专利权人 中汽创智科技有限公司

CN 110472608 A, 2019.11.19

地址 211106 江苏省南京市江宁区秣陵街

US 10552964 B2, 2020.02.04

道胜利路88号(江宁开发区)

CN 109671102 A, 2019.04.23

审查员 叶旭庆

(72) 发明人 陈志轩

(74) 专利代理机构 南京泰普专利代理事务所

(普通合伙) 32360

代理人 张磊

(51) Int. Cl.

G06K 9/00 (2006.01)

G06K 9/32 (2006.01)

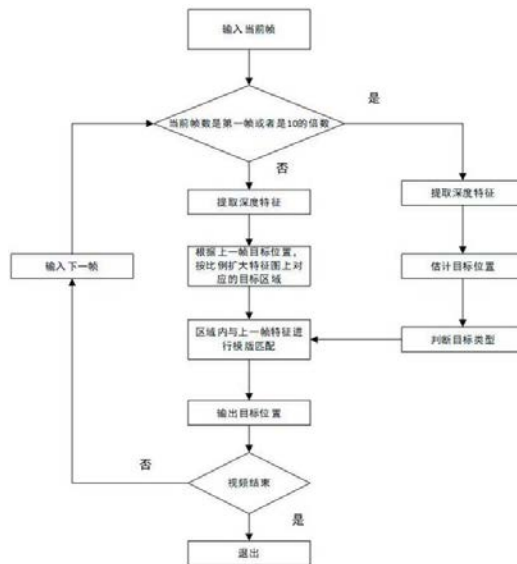
权利要求书3页 说明书6页 附图3页

(54) 发明名称

一种基于深度特征融合的模板匹配跟踪方法及系统

(57) 摘要

本发明提出了一种基于深度特征融合的模板匹配跟踪方法及系统,并通过以下步骤实现:步骤一、获取视频数据,并将视频第一帧图像输入深度卷积网络;步骤二、深度卷积网络接收图像帧信息,进行特征提取并输出;步骤三、获取根据特征信息估计出的目标位置;步骤四、判断目标类别,并根据类别信息修正目标框;步骤五、提取下一帧视频数据并输入深度卷积网络,进行特征提取;步骤六、根据步骤五提取的特征图,利用上一帧的目标特征进行模板匹配;步骤七、输出当前帧中目标位置;步骤八、判断视频读取是否结束,满足结束条件时完成目标跟踪,不满足时跳转至步骤五。本发明通过对目标的检测,针对视频中目标所在位置的进行判定,实现目标跟踪。



CN 112287906 B

1. 一种基于深度特征融合的模板匹配跟踪方法,其特征在于,包括:

步骤一:获取视频数据,并将视频第一帧图像输入深度卷积网络;

步骤二:深度卷积网络接收图像帧信息,进行特征提取并输出;

步骤三:获取根据特征信息估计出的目标位置;

步骤四:判断目标类别,并根据类别信息修正目标框;

步骤五:提取下一帧视频数据并输入深度卷积网络,进行特征提取;

步骤六:根据步骤五提取的特征图,利用上一帧的目标特征进行模板匹配;

步骤七:输出当前帧中目标位置;

步骤八:判断视频读取是否结束,满足结束条件时完成目标跟踪,不满足时跳转至步骤五;

所述步骤二进一步为:

用于提取图像帧信息特征的深度卷积网络,采用欧氏距离作为回归的损失函数,并选择focal loss作为分类损失函数对深度卷积网络进行学习能力的优化;所述欧氏距离作为回归的损失函数进一步为:

$$E_{\text{reg}} = \frac{1}{2N} \sum_{n=1}^N \|\hat{y}_n - y_n\|_2^2$$

式中,N表示目标个数,n表示当前目标索引,上标2表示范数的平方,下标2表示向量元素绝对值的平方和在开方, y_n 表示深度卷积网络中当前图像帧作为计算样本在类别划分中对应的位置参数, \hat{y}_n 表示深度卷积网络中目标图像帧作为划分类别对应的位置参数;

所述focal loss作为分类损失函数进一步引出二分类交叉损失函数,加入参数因子,并将网络关注度置于困难的、错分的样本;其中所述二分类交叉损失函数为:

$$L = -y \log y' - (1 - y) \log(1 - y) = \begin{cases} -\log y', (y = 1) \\ -\log(1 - y'), (y = 0) \end{cases}$$

式中, y' 表示经过激活函数后的输出;所述参数因子为alpha和gamma,即:

$$E_{\text{cls}} = \begin{cases} -\alpha(1 - y')^\gamma \log y', (y = 1) \\ -(1 - \alpha)y'^\gamma \log(1 - y'), (y = 0) \end{cases}$$

式中,alpha表示平衡因子, γ 表示样本重要性;

所述步骤四进一步为:

根据设定的参数类型,将步骤三中处于目标位置的跟踪目标,划分为便于区分的不同的形态、不同类型的类别;根据划分出的类别信息,输入目标真实尺寸,根据距离,估算目标长宽,模版匹配的结果,修正目标框的位置。

2. 根据权利要求1所述的一种基于深度特征融合的模板匹配跟踪方法,其特征在于,所述步骤六进一步为:

根据上一帧的目标位置,在当前帧特征图对应的目标位置附近,扩大已定义大小的区域进行目标跟踪;首先,通过上一帧选定目标所在位置,确定当前帧中初始目标框的候选区位置;其次,按照预设像素值,进行目标框的移动,产生领域目标框;再次,利用相似度匹配,

计算领域目标框中的图像特征与跟踪目标的图像特征相似度值；从次，将计算出的相似度值与预定阈值进行对比，当满足阈值时，将计算出的领域目标框作为当前帧目标所在位置的候选；最后，当相似度匹配不低于预定次数时，终止目标框的移动，并选取相似度值最高的领域目标框，作为当前帧中目标所在的位置框。

3. 一种基于深度特征融合的模板匹配跟踪系统，用于实现权利要求1~2任意一项方法，其特征在于，包括：

用于获取目标图像信息的第一模块；

用于提取特征信息的第二模块；

用于判断目标类型的第三模块；

用于获取目标位置第四模块；

用于进行数据处理的第五模块；

所述第一模块进一步包括提取模块、预处理模块；其中所述提取模块对视频中的目标信息进行读取；所述预处理模块用于对提取模块提取出的图片信息进行预处理；所述预处理进一步将待输入第二模块中的图片信息尺寸，设置为符合第二模块使用的大小；

所述第二模块进一步接收第一模块中进行预处理后的图片信息，并将其输入至深度卷积网络中进行特征提取；

用于提取图像帧信息特征的深度卷积网络采用欧氏距离作为回归的损失函数，并选择 focal loss 作为分类损失函数对深度卷积网络进行学习能力的优化；所述欧氏距离作为回归的损失函数进一步为：

$$E_{\text{reg}} = \frac{1}{2N} \sum_{n=1}^N \|\hat{y}_n - y_n\|_2^2$$

式中，N表示目标个数，n表示当前目标索引，上标2表示范数的平方，下标2表示向量元素绝对值的平方和在开方， y_n 表示深度卷积网络中当前图像帧作为计算样本在类别划分中对应的位置参数， \hat{y}_n 表示深度卷积网络中目标图像帧作为划分类别对应的位置参数；

所述 focal loss 作为分类损失函数进一步引出二分类交叉损失函数，加入参数因子，并将网络关注度置于困难的、错分的样本；其中所述二分类交叉损失函数为：

$$L = -y \log y' - (1 - y) \log(1 - y) = \begin{cases} -\log y', & (y = 1) \\ -\log(1 - y'), & (y = 0) \end{cases}$$

式中， y' 表示经过激活函数后的输出；所述参数因子为alpha和gamma，即：

$$E_{\text{cls}} = \begin{cases} -\alpha(1 - y')^\gamma \log y', & (y = 1) \\ -(1 - \alpha)y'^\gamma \log(1 - y'), & (y = 0) \end{cases}$$

式中， α 表示平衡因子， γ 表示样本重要性；

所述第三模块进一步根据设定的参数类型，将跟踪目标，划分为便于区分的不同的形态、不同类型的类别；同时，利用第五模块，根据划分出的类别信息，利用目标真实尺寸，根据距离，估算目标长宽，模版匹配的结果，修正目标框所在的位置。

4. 根据权利要求3所述的一种基于深度特征融合的模板匹配跟踪系统，其特征在于，所述第四模块进一步包括：初始位置获取模块、当前模块；所述初始位置获取模块用于根据第

一帧在第一模块获取的特征信息,利用第五模块进行目标位置的确定;所述当前模块用于上一帧中目标位置,通过目标框的移动得到候选目标框,并采用第五模块进行数据处理,从而获取当前的帧中,目标所在位置的目标框。

5. 根据权利要求3所述的一种基于深度特征融合的模板匹配跟踪系统,其特征在于,所述第五模块用于对数据进行处理,进一步为包括相似度计算模块、比例计算模块、NMS处理模块、误差计算模块;所述相似度计算模块用于计算图片信息之间的相似程度;所述比例计算模块用于根据目标真实大小合理规划目标框的选定的大小;所述NMS处理模块用于对目标值进行处理,保留最大置信度的候选框为目标框;所述误差计算模块用于计算误差值,并将其作为检测修正的依据。

一种基于深度特征融合的模板匹配跟踪方法及系统

技术领域

[0001] 本发明涉及一种基于深度特征融合的模板匹配跟踪方法及系统,特别是涉及目标跟踪技术领域。

背景技术

[0002] 随着计算机技术的发展,对于目标的检测、分类、分割等领域,基于深度学习的检测方法逐渐占据主导地位。由于实际场景中,遮挡、光照、目标非刚性等因素影响,目标跟踪的准确性和鲁棒性仍然存在问题。

[0003] 现有技术中,在采用深度特征流处理图像时,对过滤静止物体建立运动物体的特征采用关键帧移动物体特征传播到当前帧的措施,该技术处理过程容易出现漂移,导致鲁棒性不够。同时,在获取深度帧以及特征信息数据库需要额外的存储花销,计算所有参考轮廓的特征信息集合逐一计算匹配,导致实时性得不到实际应用过程中的满足。

发明内容

[0004] 发明目的:一个目的是提出一种基于深度特征融合的模板匹配跟踪方法,以解决现有技术存在的上述问题。进一步目的是提出一种实现上述方法的模板匹配跟踪系统。

[0005] 技术方案:一种基于深度特征融合的模板匹配跟踪方法,包括以下步骤:

[0006] 步骤一:获取视频数据,并将视频第一帧图像输入深度卷积网络;

[0007] 步骤二:深度卷积网络接收图像帧信息,进行特征提取并输出;

[0008] 步骤三:获取根据特征信息估计出的目标位置;

[0009] 步骤四:判断目标类别,并根据类别信息修正目标框;

[0010] 步骤五:提取下一帧视频数据并输入深度卷积网络,进行特征提取;

[0011] 步骤六:根据步骤五提取的特征图,利用上一帧的目标特征进行模板匹配;

[0012] 步骤七:输出当前帧中目标位置;

[0013] 步骤八:判断视频读取是否结束,满足结束条件时完成目标跟踪,不满足时跳转至步骤五。

[0014] 在进一步的实施例中,所述步骤二进一步为:

[0015] 用于提取图像帧信息特征的深度卷积网络,采用欧氏距离作为回归的损失函数,并选择focal loss作为分类损失函数对深度卷积网络进行学习能力的优化;所述欧氏距离作为回归的损失函数进一步为:

$$[0016] \quad E_{\text{reg}} = \frac{1}{2N} \sum_{n=1}^N \|\hat{y}_n - y_n\|_2^2$$

[0017] 式中, N 表示目标个数, n 表示当前目标索引,上标2表示范数的平方,下标2表示向量元素绝对值的平方和在开方, y_n 表示深度卷积网络中当前图像帧作为计算样本在类别划分中对应的位置参数, \hat{y}_n 表示深度卷积网络中目标图像帧作为划分类别对应的位置参数;

[0018] 所述focal loss作为分类损失函数进一步引出二分类交叉损失函数,加入参数因子,并将网络关注度置于困难的、错分的样本;其中所述二分类交叉损失函数为:

$$[0019] \quad L = -y \log y' - (1 - y) \log(1 - y) = \begin{cases} -\log y', & (y = 1) \\ -\log(1 - y'), & (y = 0) \end{cases}$$

[0020] 式中, y' 表示经过激活函数后的输出;所述参数因子为alpha和gamma,即:

$$[0021] \quad E_{cls} = \begin{cases} -\alpha(1 - y')^\gamma \log y', & (y = 1) \\ -(1 - \alpha)y'^\gamma \log(1 - y'), & (y = 0) \end{cases}$$

[0022] 式中, α 表示平衡因子, γ 表示样本重要性。

[0023] 在进一步的实施例中,所述步骤四进一步为:

[0024] 根据设定的参数类型,将步骤三中处于目标位置的跟踪目标,划分为便于区分的不同的形态、不同类型的类别;根据划分出的类别信息,输入目标真实尺寸,根据距离,估算目标长宽,模版匹配的结果,修正目标框的位置。

[0025] 在进一步的实施例中,所述步骤六进一步为:

[0026] 根据上一帧的目标位置,在当前帧特征图对应的目标位置附近,扩大已定义大小的区域进行目标跟踪;首先,通过上一帧选定目标所在位置,确定当前帧中初始目标框的候选区位置;其次,按照预设像素值,进行目标框的移动,产生领域目标框;再次,利用相似度匹配,计算领域目标框中的图像特征与跟踪目标的图像特征相似度值;从次,将计算出的相似度值与预定阈值进行对比,当满足阈值时,将计算出的领域目标框作为当前帧目标所在位置的候选;最后,当相似度匹配不低于预定次数时,终止目标框的移动,并选取相似度值最高的领域目标框,作为当前帧中目标所在的位置框。

[0027] 一种基于深度特征融合的模板匹配跟踪系统,用于实现一种基于深度特征融合的模板匹配跟踪方法,包括:

[0028] 用于获取目标图像信息的第一模块;

[0029] 用于提取特征信息的第二模块;

[0030] 用于判断目标类型的第三模块;

[0031] 用于获取目标位置第四模块;

[0032] 用于进行数据处理的第五模块。

[0033] 在进一步的实施例中,所述第一模块进一步包括提取模块、预处理模块;其中提取模块对视频中的目标信息进行读取。预处理模块用于对提取模块提取出的图片信息进行预处理;所述预处理进一步将待输入第二模块中的图片信息尺寸,设置为符合第二模块使用的大小。

[0034] 在进一步的实施例中,所述第二模块进一步接收第一模块中进行预处理后的图片信息,并将其输入至深度卷积网络中进行特征提取。

[0035] 在进一步的实施例中,所述第三模块进一步根据设定的参数类型,将跟踪目标,划分为便于区分的不同的形态、不同类型的类别;同时,利用第五模块,根据划分出的类别信息,利用目标真实尺寸,根据距离,估算目标长宽,模版匹配的结果,修正目标框所在的位置。

[0036] 在进一步的实施例中,所述第四模块进一步包括:初始位置获取模块、当前模块;所述初始位置获取模块用于根据第一帧在第一模块获取的特征信息,利用第五模块进行目

标位置的确定;所述当前模块用于上一帧中目标位置,通过目标框的移动得到候选目标框,并采用第五模块进行数据处理,从而获取当前的帧中,目标所在位置的目标框。

[0037] 在进一步的实施例中,所述第五模块用于对数据进行处理,进一步为包括相似度计算模块、比例计算模块、NMS处理模块、误差计算模块;所述相似度计算模块用于计算图片信息之间的相似程度;所述比例计算模块用于根据目标真实大小合理规划目标框的选定的大小;所述NMS处理模块用于对目标值进行处理,保留最大置信度的候选框为目标框;所述误差计算模块用于计算误差值,并将其作为检测修正的依据。

[0038] 有益效果:本发明提出了一种基于深度特征融合的模板匹配跟踪方法及实现该方法的系统,针对目标跟踪中不稳定且占用耗时间长的的问题,利用深度学习提取有效特征,并输出类别信息,在一定的有效区域内利用上一帧的特征以及类别信息对应的尺度,进行跟踪,在目标跟踪的鲁棒性方面产生质变的提高,且同时抑制了跟踪目标的抖动与漂移。

附图说明

[0039] 图1为本发明的方法实现流程图。

[0040] 图2为本发明的数据流向图。

[0041] 图3为本发明的网络结构图。

具体实施方式

[0042] 本发明通过一种基于深度特征融合的模板匹配跟踪方法及实现该方法的系统,实现目标跟踪的目的。下面通过实施例,并结合附图对本方案做进一步具体说明。

[0043] 在本申请中,我们提出了一种基于深度特征融合的模板匹配跟踪方法及实现该方法的系统,其中包含的一种基于深度特征融合的模板匹配跟踪方法,如附图1所示,为本发明方法实现流程图,具体划分为以下步骤:

[0044] 步骤一:获取视频数据,并将视频第一帧图像输入深度卷积网络;该步骤进一步对获取的视频数据进行预处理,具体为将待输入深度卷积网络中的图像尺寸,处理为深度卷积网络可接受的大小。

[0045] 步骤二:深度卷积网络接收图像帧信息,进行特征提取并输出深度特征图F0;该步骤中用于提取图像帧信息特征的深度卷积网络,采用欧氏距离作为回归的损失函数,并选择focal loss作为分类损失函数对深度卷积网络进行学习能力的优化。

[0046] 其中,欧氏距离作为回归的损失函数进一步为:

$$[0047] \quad E_{\text{reg}} = \frac{1}{2N} \sum_{n=1}^N \|\hat{y}_n - y_n\|_2^2$$

[0048] 式中, N 表示目标个数, n 表示当前目标索引,上标2表示范数的平方,下标2表示向量元素绝对值的平方和在开方, y_n 表示深度卷积网络中当前图像帧作为计算样本在类别划分中对应的位置参数, \hat{y}_n 表示深度卷积网络中目标图像帧作为划分类别对应的位置参数。

[0049] 所述focal loss作为分类损失函数进一步引出二分类交叉损失函数,加入参数因子,并将网络关注度置于困难的、错分的样本;其中所述二分类交叉损失函数为:

$$[0050] \quad L = -y \log y' - (1 - y) \log(1 - y) = \begin{cases} -\log y', & (y = 1) \\ -\log(1 - y'), & (y = 0) \end{cases}$$

[0051] 式中, y' 表示经过激活函数后的输出, 值域在 0~1 内, 由于普通的交叉熵对于正样本而言, 输出概率越大损失越小; 对于负样本而言, 输出概率越小则损失越小。所以此时的损失函数在大量简单样本的迭代过程中比较缓慢且可能无法优化至最优。为减少易分类样本的损失, 使得整个网络更关注困难的、错分的样本, 引入 α 和 γ 两个引子, 即:

$$[0052] \quad E_{cls} = \begin{cases} -\alpha(1 - y')^\gamma \log y', & (y = 1) \\ -(1 - \alpha)y'^\gamma \log(1 - y'), & (y = 0) \end{cases}$$

[0053] 式中, α 表示平衡因子, 用来平衡正负样本重要性, γ 表示样本重要性, 优选为 0.25。

[0054] 步骤三: 获取根据特征信息估计出的目标位置; 该步骤进一步为根据步骤二中获取的特征图对应的 anchors 生成候选区域, 并对生成的候选区域进行判断。当候选区域为目标时, 进行位置的偏差预测。随后, 对预测为目标的 proposal 进行遍历, 并进一步做 NMS 处理, 保留重叠阈值内最大置信度的框, 将其作为目标框, 从而确定当前目标的位置。

[0055] 步骤四: 判断目标类别; 该步骤根据设定的参数类型, 将步骤三中处于目标位置的跟踪目标, 如附图 3 所示, 输入至卷积神经网络中, 进行分类。具体划分为便于区分的不同的形态、不同类型的类别; 根据划分出的类别信息, 输入目标真实尺寸, 根据距离, 估算目标长宽, 模版匹配的结果, 修正目标框的位置。其中目标框的修正还包括每经历过 10 帧, 就进行一次检测修正。

[0056] 步骤五: 提取下一帧视频数据并输入深度卷积网络, 进行特征提取;

[0057] 步骤六: 根据步骤五提取的特征图, 利用上一帧的目标特征进行模板匹配; 匹配的过程具体为根据上一帧的目标位置, 在当前帧特征图 F0 对应的目标位置附近, 扩大已定义大小的区域进行目标跟踪。首先, 通过上一帧选定目标所在位置, 确定当前帧中初始目标框的候选区位置, 并根据目标类型的真实宽高, 距离信息估算出在图像中的尺度大小, 从而对候选区位置进行精修; 其次, 按照预设像素值, 进行目标框的移动, 产生领域目标框; 再次, 利用相似度匹配, 计算领域目标框中的图像特征与跟踪目标的图像特征相似度值; 从次, 将计算出的相似度值与预定阈值进行对比, 当满足阈值时, 将计算出的领域目标框作为当前帧目标所在位置的候选; 最后, 当相似度匹配不低于预定次数时, 终止目标框的移动, 并选取相似度值最高的领域目标框, 作为当前帧中目标所在的位置框。

[0058] 步骤七: 输出当前帧中目标位置;

[0059] 步骤八: 判断视频读取是否结束, 当满足结束条件时, 输出目标的最终位置, 并完成目标跟踪, 不满足时跳转至步骤五, 继续进行目标跟踪。

[0060] 基于上述方法, 提出一种基于深度特征融合的模板匹配跟踪系统, 用于上述方法, 包括:

[0061] 用于获取目标图像信息的第一模块; 该模块进一步包括提取模块、预处理模块; 其中提取模块对视频中的目标信息进行读取。预处理模块用于对提取模块提取出的图片信息进行预处理; 所述预处理进一步将待输入第二模块中的图片信息尺寸, 设置为符合第二模块使用的大小。

[0062] 用于提取特征信息的第二模块; 该模块进一步接收第一模块中进行预处理后的图片信息, 并将其输入至深度卷积网络中进行特征提取。

[0063] 用于判断目标类型的第三模块;该模块根据第二模块提取出的目标特征信息,将其输入至卷积神经网络,进一步根据设定的参数类型,将跟踪目标,划分为便于区分的不同的形态、不同类型的类别;同时,利用第五模块,根据划分出的类别信息,利用目标真实尺寸,根据距离,估算目标长宽,模版匹配的结果,修正目标框所在的位置。

[0064] 用于获取目标位置第四模块;该模块进一步包括:初始位置获取模块、当前模块;所述初始位置获取模块用于根据第一帧在第一模块获取的特征信息,利用第五模块进行目标位置的确定;所述当前模块用于上一帧中目标位置,通过目标框的移动得到候选目标框,并采用第五模块进行数据处理,从而获取当前的帧中,目标所在位置的目标框。

[0065] 用于进行数据处理的第五模块;该模块进一步为包括相似度计算模块、比例计算模块、NMS处理模块、误差计算模块;所述相似度计算模块用于计算图片信息之间的相似程度;所述比例计算模块用于根据目标真实大小合理规划目标框的选定的大小;所述NMS处理模块用于对目标值进行处理,保留最大置信度的候选框为目标框;所述误差计算模块用于计算误差值,并将其作为检测修正的依据。

[0066] 为便于对本发明的理解,以下提供一具体优选实施例,对本方案做进一步具体说明,但不作为对本发明自身的限制。例如当目标为行人时,对目标人物的跟踪。

[0067] 首先读取视频中目标人物在第一帧中的图片信息,并将图片缩放至深度卷积网络可接受的 256×256 图片尺寸;其次,输入图片经过 3×3 的卷积核以及预定量的卷积层提取特征;再次,根据提取出的图片特征信息,以及对应的anchors生成候选区域,利用NMS处理,保留置信度最大的候选区域,圈选目标所在区域,即目标框;再次,将目标框输入分类卷积神经网络,利用 3×3 、 1×1 的卷积核以及Global average pooling,对不具有固定输出尺寸的目标框所含信息,进行目标类型的判定,并输出;从次,接收视频下一帧的图片信息,通过深度卷积网络提取特征,并根据上一帧中目标框的位置进行匹配;最后,根据匹配结果输出目标位置,实现跟踪。

[0068] 其中,匹配的方式具体为以上一帧目标人物的位置为中心,1像素为步长,领域为方向,进行目标框的移动,得到8个候选目标框,将获取的候选目标框与上一帧目标人物的特征图,利用余弦距离的计算方式,进行相似度的匹配。当余弦距离数值越大,则两者之间的相似度就越大。将不低于阈值的候选目标框,且同时满足当前8个领域数值最大的候选目标框,作为当前扩张中匹配成功的目标框。在扩张过程中,以8领域中数值最大的候选目标框为基准,进一步进行扩张。当满足三次以上匹配成功的时候,停止扩张,并将三次匹配成功的结果中,选取对应相似度值最大的候选目标框为当前帧中的最终目标所在位置框。

[0069] 对于目标的类型判断在行人为目标的实施例中,在年龄上划分小孩,成年人,老年人;从姿态上,将行人分为站立、弯腰以及蹲着的行人,运动状态分为站立静止、弯腰静止、半蹲静止、行进;对应的行人真实身高应有区分,其中,成年人和老年人的身高在1.7m,小孩的身高在1.4m,站立的行人实际高度为1.7m,弯腰为1.2m,半蹲为1m。根据分类得到的真实身高,在根据影像中的相似三角形原理,可以推算得到: $F/d=h/H$,其中F为相机焦距,和相机内参相关,窄角相机的焦距在1280左右,d为利用天地消失点可以获得的行人离摄像头的距离,H为分类后实际行人高度,h为待求的图像中行人高度。依靠上述公式可以求得对应的行人高度,从而作为后续行人尺度的规范。

[0070] 对于目标的框的修订在行人为目标的实施例中具体为根据上述获取的行人高度

以及分类的行人状态,依据不同姿态对应的长宽比,站立行人5:1,弯腰行人3:2,半蹲行人1:1,走路行人1.8:1,推算行人宽度,得到图像中行人的长和宽。根据得到的目标位置中心,修正目标尺寸大小,将获取的最终目标所在位置框与当前分类信息推导得到的框进行融合,即判断最终目标所在位置框得到的长 h_2 和宽 w_2 与分类推导的到的长 h_3 和宽 w_3 的差值,相比检测得到的长 h_0 和宽 w_0 与 h_3 和 w_3 的差值,是否在阈值10%内,即是否满足 $90\% * (h_3 - h_0) < (h_3 - h_2) < 110\% * (h_3 - h_0)$ && $90\% * (w_3 - w_0) < (w_3 - w_2) < 110\% * (w_3 - w_0)$,若满足,即以当前 h_2 和 w_2 输出,若不满足,则利用检测进行修正。

[0071] 本发明提出的跟踪技术,与现有技术相比,呈现如下表1所示的有益效果。

[0072]

跟踪方法对比	OTB-100 AUC (%)	Speed (FPS)
BACF	62.98%	35ms (CPU)
CREST	62.3%	1ms (GPU)
PTAV	63.5%	25ms (GPU)
EA	62.9%	23ms (CPU)
本发明	64.1%	20ms (CPU)

[0073] 如附图2所示,本发明利用卷积神经网络提取特征跟踪目标的同时,额外使用分类的网络分支,对目标进行类别判断,在跟踪后续框的修正上,起到防止抖动漂移的作用,增强鲁棒性。其中,利用深度卷积神经网络识别的目标,在检测基础上进行目标识别,对目标跟踪位置的估算产生判决依据的作用;同时,在原有卷积神经网络做检测任务的基础上,加入分支对提取到的特征进行分类,有利于距离测算以及抑制误报和漏报的拓展。

[0074] 目前大部分主流的方法,在利用深度卷积神经网络检测出第一帧之后,用粒子滤波或者卡尔曼滤波对于每个目标框进行跟踪,这种检测结果之后再加入滤波器两个独立模块的方式,跟踪时耗费的时间与目标个数呈线性相关,多个滤波器线性叠加对资源占用较大,本发明提出的方法,将多个滤波时间全部缩减到深度卷积神经网络提取的模版匹配上,相比于传统跟踪方法,本发明不需要针对每个检测到的目标再进行滤波或者光流映射特征矢量,直接在原图深度特征上进行匹配,节省了时间。

[0075] 如上所述,尽管参照特定的优选实施例已经表示和表述了本发明,但其不得解释为对本发明自身的限制。在不脱离所附权利要求定义的本发明的精神和范围前提下,可对其在形式上和细节上做出各种变化。

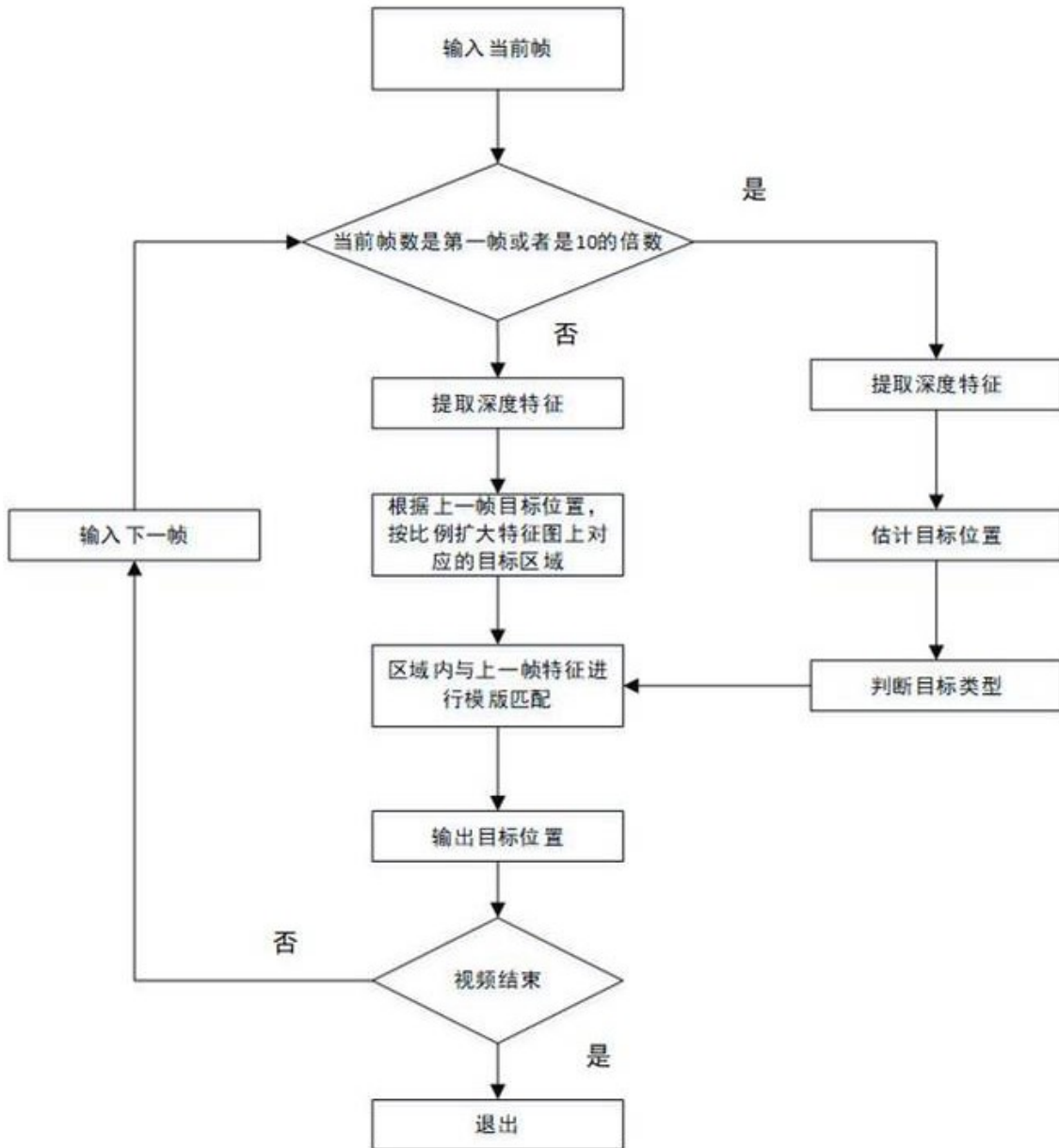


图 1



图 2

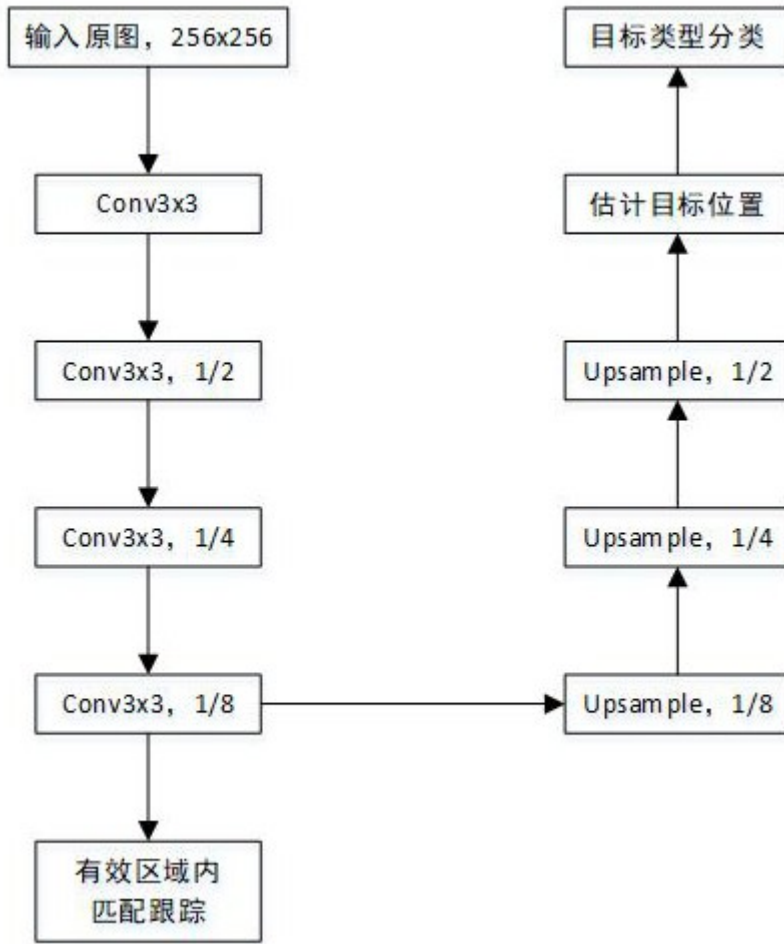


图 3