



# (12) 发明专利

(10) 授权公告号 CN 113608663 B

(45) 授权公告日 2023. 07. 25

(21) 申请号 202110783266.8

G06N 3/0464 (2023.01)

(22) 申请日 2021.07.12

G06N 3/08 (2023.01)

(65) 同一申请的已公布的文献号  
申请公布号 CN 113608663 A

(56) 对比文件

EP 3686772 A1, 2020.07.29

US 2016180195 A1, 2016.06.23

(43) 申请公布日 2021.11.05

US 2019347817 A1, 2019.11.14

(73) 专利权人 哈尔滨工程大学  
地址 150001 黑龙江省哈尔滨市南岗区南通大街145号哈尔滨工程大学科技处知识产权办公室

李智娴;陈正鸣.1种基于Kinect深度图像的指尖检测与跟踪算法.江苏农业科学.2015,(第05期),全文.

审查员 姜涛

(72) 发明人 孟浩 王玥 田洋 邓艳琴

(51) Int. Cl.

G06F 3/04883 (2022.01)

G06V 10/44 (2022.01)

G06V 10/56 (2022.01)

G06V 10/82 (2022.01)

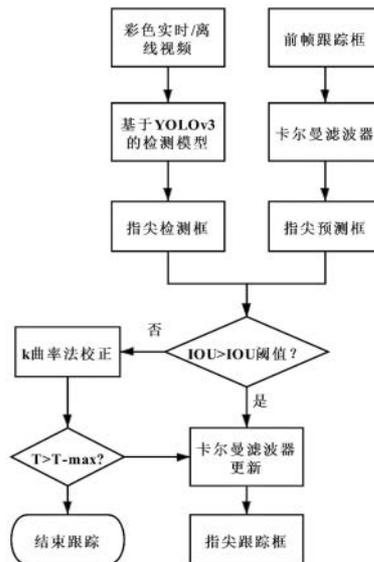
权利要求书3页 说明书8页 附图3页

## (54) 发明名称

一种基于深度学习和K-曲率法的指尖跟踪方法

## (57) 摘要

本发明公开了一种基于深度学习和K-曲率法的指尖跟踪方法,首先利用YOLOv3网络模型训练预处理后的数据集,获取指尖检测模型;再利用摄像头获取视频流,输入检测模型并检测出检测框信息,初始化卡尔曼滤波器;然后利用卡尔曼滤波器得到预测框,计算出本帧检测框和预测框的IOU,设定IOU阈值,判断该IOU是否大于IOU阈值,若该IOU大于IOU阈值则更新卡尔曼滤波器得到指尖跟踪框;否则,利用K-曲率法对指尖位置进行校正,并更新卡尔曼滤波器;最后设定一个时间阈值T-max,在该时间阈值帧内未检测跟踪信息,则终止跟踪。本发明减弱了复杂环境对检测准确性的影响,提升了检测速度,增加了准确性和鲁棒性。



1. 一种基于深度学习和K-曲率法的指尖跟踪方法,其特征在于,包括以下步骤:

S1: 获取手部数据集,进行预处理;

S2: 利用深度学习神经网络模型YOLOv3对数据集进行训练,获取指尖检测模型;

S3: 获取视频流,将当前帧输入指尖检测模型中,利用Darknet53网络对当前帧的图像进行多尺度特征提取,检测出目标类别和检测框位置信息,并根据此信息初始化卡尔曼滤波器;

S4: 读取下一帧图像,利用卡尔曼滤波器得到预测框,计算出本帧检测框和预测框的IOU,设定IOU阈值,判断该IOU是否大于IOU阈值,若该IOU大于IOU阈值则进行S5步骤,否则进行S6步骤;

S5: 用本帧匹配到的预测框去更新卡尔曼滤波器,并将状态更新值作为本帧的跟踪框输出,该帧跟踪完成,回到步骤S4;

S6: 利用K-曲率法得到指尖点的位置信息;

S7: 计算该指尖点与检测框的中心点的欧式距离,并设定一个阈值,当该距离小于阈值时,则初始化卡尔曼滤波器得到新预测框,并重新开始匹配;否则删除掉此帧的跟踪信息,读取下一帧;

S8: 设定一个时间阈值T-max,在该时间阈值帧内未检测跟踪信息,则终止跟踪。

2. 根据权利要求1所述的一种基于深度学习和K-曲率法的指尖跟踪方法,其特征在于:所述步骤S1包括以下步骤:

S11: 在不同的场景、光照、人、指尖角度、手部数量条件下采集大量包含露出指尖的手部彩色图;

S12: 利用数据增强的方法对手部彩色图像进行数据扩充;

S13: 标记出目标指尖的区域框,添加标签信息,生成数据标签文件。

3. 根据权利要求1所述的一种基于深度学习和K-曲率法的指尖跟踪方法,其特征在于:所述步骤2包括以下步骤:

S21: 将手部数据集输入图片转化成416\*416大小的网格图片,增加灰度条防止失真;

S22: 将处理后的图片通过Darknet53特征提取网络降采样5次,生成多尺度的特征图;

S23: 利用多尺度融合的方式在13\*13、26\*26、52\*52三个尺度的特征图上分别进行卷积检测,得到3次检测结果,并进行非极大值抑制,获取最终结果;

S24: 生成训练后的指尖检测模型。

4. 根据权利要求1所述的一种基于深度学习和K-曲率法的指尖跟踪方法,其特征在于:所述步骤S4包括以下步骤:

S41: 读取下一帧图像,利用卡尔曼滤波器得到所有预测到的指尖预测框;

S42: 计算出本帧检测框和所有预测框的IOU;

S43: 利用匈牙利算法得到IOU最大的唯一匹配;

S44: 设定IOU阈值,判断该IOU是否大于IOU阈值;

S45: 该IOU大于IOU阈值则进行S5步骤,否则进行S6步骤。

5. 根据权利要求4所述的一种基于深度学习和K-曲率法的指尖跟踪方法,其特征在于:所述步骤S41: 读取下一帧图像,利用卡尔曼滤波器得到所有预测到的指尖预测框具体包括:

使用的卡尔曼滤波器包含七个状态变量,四个观测输入:

$$\hat{x}_{\bar{k}} = A\hat{x}_{k-1} + Bu_{k-1}$$

$$\hat{P}_{\bar{k}} = A\hat{P}_{k-1}A^T + Q$$

$$\hat{x}_k = \hat{x}_{\bar{k}} + K_k(Z_k - H\hat{x}_{\bar{k}})$$

$$K_k = (\hat{P}_{\bar{k}} - H^T) / (H\hat{P}_{\bar{k}}H + R)$$

$$\hat{P}_k = (I - K_kH)\hat{P}_{\bar{k}}$$

输出目标状态:

$$X = [u, v, x, y, \dot{u}, \dot{v}, \dot{x}, \dot{y}]^T$$

式中,  $\hat{x}_{\bar{k}}$  表示k时刻先验状态估计值,  $\hat{x}_k$ 、 $\hat{x}_{k-1}$  分别表示k时刻、k-1时刻后验状态估计值, A表示状态转移矩阵, B表示可选的控制输入  $u \in R^1$  的增益,  $u_{k-1}$  表示k-1时刻的控制增益,  $\hat{P}_{\bar{k}}$  表示k时刻的先验估计协方差,  $\hat{P}_{\bar{k}}$ 、 $\hat{P}_{k-1}$  分别表示k时刻、k-1时刻的后验估计协方差, Q表示过程激励噪声的协方差,  $K_k$  表示卡尔曼增益,  $Z_k$  表示测量值, H表示测量矩阵, u表示当前帧目标中心的水平像素位置, v表示当前帧目标中心的垂直像素位置, x表示目标区域的面积, y表示目标区域的高宽比,  $\dot{u}$  表示预测后的下一帧图像中目标中心的水平像素位置,  $\dot{v}$  表示预测后的下一帧图像中目标中心的垂直像素位置,  $\dot{x}$  表示预测后的下一帧图像中目标区域的面积,  $\dot{y}$  表示预测后的下一帧图像中目标区域的高宽比。

6. 根据权利要求1所述的一种基于深度学习和K-曲率法的指尖跟踪方法, 其特征在于: 所述步骤S6包括以下步骤:

S61: 选择YCbCr颜色空间对本帧图像进行手部肤色提取:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 65.738 & 129.057 & 25.06 \\ -37.945 & -74.494 & 112.43 \\ 112.439 & -94.154 & -18.28 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix}$$

式中, R表示红色通道, G表示绿色通道, B表示蓝色通道; Y为明亮度; Cb为RGB输入信号蓝色部分与RGB信号亮度值之间的差异; Cr为RGB输入信号红色部分与RGB信号亮度值之间的差异;

S62: 求高斯模型的马氏距离:

$$dis_s = (X - m_s)^T C_s^{-1} (X - m_s)$$

式中,  $m_s$  表示肤色单高斯模型的平均矢量,  $C_s$  为肤色单高斯模型的协方差矩阵, X为像素点;

S63: 利用马氏距离分别给肤色和非肤色区域建立单高斯模型, 然后求某一像素在肤色模型和非肤色模型下的马氏距离来判断是否属于肤色点, 并分割出手部图像:

$$dis_s - dis_{ns} = (X - m_s)^T C_s^{-1} (X - m_s) - (X - m_{ns})^T C_{ns}^{-1} (X - m_{ns}) \leq \tau_{diff}$$

式中,  $\tau_{diff}$  为设定的阈值,  $m_{ns}$  表示非肤色单高斯模型的均值,  $C_{ns}$  为高斯模型的协方差;

S64: 对图像进行二值化处理和双边滤波处理;

S65: 取轮廓点 $k_i$ 与其前面的第 $m$ 个点 $k_{i+m}$ 和后面的第 $m$ 个点 $k_{i-m}$ ,使用向量 $\overline{k_i k_{i+m}}$ 和向量 $\overline{k_i k_{i-m}}$ 的夹角余弦值作为点 $k_i$ 的曲率:

$$\cos \theta_i = \frac{\overline{k_i k_{i+m}} \times \overline{k_i k_{i-m}}}{|\overline{k_i k_{i+m}}| \times |\overline{k_i k_{i-m}}|}$$

S66: 检测具有局部最大K余弦的凸轮廓点作为指尖点:

$$d_i > (d_{i-s} + d_{i+s}) / 2$$

式中, $d_i$ 表示从手掌心最大圆的质心到点的距离。

## 一种基于深度学习和K-曲率法的指尖跟踪方法

### 技术领域

[0001] 本发明属于目标检测及跟踪技术,具体涉及一种基于深度学习和K-曲率法的指尖跟踪方法。

### 背景技术

[0002] 人手姿态检测与跟踪是目前人机交互与计算机视觉领域的一个热门方向,其子方向指尖检测与跟踪技术是人手姿态检测与跟踪技术的重要组成部分,通过检测并跟踪指尖可以为空中手写、空中点击虚拟屏幕、手势识别、智慧教学等人机交互行为提供良好的基础。

[0003] 在指尖检测与跟踪技术中,指尖检测的正确性及跟踪的快速性与准确性都至关重要;目前基于目标检测的算法主要分为传统检测算法和基于深度学习的检测算法,传统目标检测算法主要包括DPM(Deformable Parts Model)、选择性搜索等,这些方法在实际应用中都存在时间复杂度高、环境复杂导致鲁棒性不好,泛化性差等缺点,对于快速变化及发生短暂遮挡的指尖很难进行有效的检测与跟踪。

[0004] 基于深度学习的指尖检测与跟踪方法可以很好的增加准确率及鲁棒性,主流的深度学习目标检测算法主要分为双阶段检测算法和单阶段检测算法,其中单阶段检测算法中的YOLO系列很好的兼顾了快速性和准确性;跟踪算法目前广泛应用的有基于滤波理论的卡尔曼滤波、粒子滤波等,但单纯的基于滤波的跟踪算法在实际应用中,对于被短暂遮挡的指尖的漏检率较高,在实时性上也有很大提升空间。

### 发明内容

[0005] 针对上述现有技术存在问题,本发明要解决的技术问题是提供一种基于深度学习和K-曲率法的指尖跟踪方法,减弱复杂环境对检测准确性的影响,提升检测速度和准确性。

[0006] 为解决上述技术问题,本发明的一种基于深度学习和K-曲率法的指尖跟踪方法,包括以下步骤:

[0007] S1:获取手部数据集,进行预处理;

[0008] S2:利用深度学习神经网络模型YOLOv3对数据集进行训练,获取指尖检测模型;

[0009] S3:获取视频流,将当前帧输入指尖检测模型中,利用Darknet53网络对当前帧的图像进行多尺度特征提取,检测出目标类别和检测框位置信息,并根据此信息初始化卡尔曼滤波器;

[0010] S4:读取下一帧图像,利用卡尔曼滤波器得到预测框,计算出本帧检测框和预测框的IOU,设定IOU阈值,判断该IOU是否大于IOU阈值,若该IOU大于IOU阈值则进行S5步骤,否则进行S6步骤;

[0011] S5:用本帧匹配到的预测框去更新卡尔曼滤波器,并将状态更新值作为本帧的跟踪框输出,该帧跟踪完成,回到步骤S4;

[0012] S6:利用K-曲率法得到指尖点的位置信息;

[0013] S7:计算该指尖点与检测框的中心点的欧式距离,并设定一个阈值,当该距离小于阈值时,则初始化卡尔曼滤波器得到新预测框,并重新开始匹配;否则删除掉此帧的跟踪信息,读取下一帧;

[0014] S8:设定一个时间阈值T-max,在该时间阈值帧内未检测跟踪信息,则终止跟踪。

[0015] 本发明还包括:

[0016] 1.步骤S1包括以下步骤:

[0017] S11:在不同的场景、光照、人、指尖角度、手部数量条件下采集大量包含露出指尖的手部彩色图;

[0018] S12:利用数据增强的方法对手部彩色图像进行数据扩充;

[0019] S13:标记出目标指尖的区域框,添加标签信息,生成数据标签文件。

[0020] 2.步骤2包括以下步骤:

[0021] S21:将手部数据集输入图片转化成416\*416大小的网格图片,增加灰度条防止失真;

[0022] S22:将处理后的图片通过Darknet53特征提取网络降采样5次,生成多尺度的特征图;

[0023] S23:利用多尺度融合的方式在13\*13、26\*26、52\*52三个尺度的特征图上分别进行卷积检测,得到3次检测结果,并进行非极大值抑制,获取最终结果;

[0024] S24:生成训练后的指尖检测模型。

[0025] 3.步骤S4包括以下步骤:

[0026] S41:读取下一帧图像,利用卡尔曼滤波器得到所有预测到的指尖预测框;

[0027] S42:计算出本帧检测框和所有预测框的IOU;

[0028] S43:利用匈牙利算法得到IOU最大的唯一匹配;

[0029] S44:设定IOU阈值,判断该IOU是否大于IOU阈值;

[0030] S45:该IOU大于IOU阈值则进行S5步骤,否则进行S6步骤。

[0031] 4.步骤S41:读取下一帧图像,利用卡尔曼滤波器得到所有预测到的指尖预测框具体包括:

[0032] 使用的卡尔曼滤波器包含七个状态变量,四个观测输入:

$$[0033] \hat{x}_{\bar{k}} = A\hat{x}_{k-1} + Bu_{k-1}$$

$$[0034] \hat{P}_{\bar{k}} = A\hat{P}_{k-1}A^T + Q$$

$$[0035] \hat{x}_{\bar{k}} = \hat{x}_k + K_k(Z_k - H\hat{x}_{\bar{k}})$$

$$[0036] K_k = (\hat{P}_{\bar{k}} - H^T) / (H\hat{P}_{\bar{k}}H + R)$$

$$[0037] \hat{P}_{\bar{k}} = (I - K_kH)\hat{P}_{\bar{k}}$$

[0038] 输出目标状态:

$$[0039] X = [u, v, x, y, \dot{u}, \dot{v}, \dot{x}, \dot{y}]^T$$

[0040] 式中,  $\hat{x}_{\bar{k}}$ 表示k时刻先验状态估计值,  $\hat{x}_k$ 、 $\hat{x}_{k-1}$ 分别表示k时刻、k-1时刻后验状态估计值,A表示状态转移矩阵,B表示可选的控制输入  $u \in \mathbb{R}^1$  的增益,  $u_{k-1}$ 表示k-1时刻的控制增

益,  $\hat{P}_k$  表示k时刻的先验估计协方差,  $\hat{P}_k$ 、 $\hat{P}_{k-1}$  分别表示k时刻、k-1时刻的后验估计协方差, Q表示过程激励噪声的协方差,  $K_k$  表示卡尔曼增益,  $Z_k$  表示测量值, H表示测量矩阵, u表示当前帧目标中心的水平像素位置, v表示当前帧目标中心的垂直像素位置, x表示目标区域的面积, y表示目标区域的高宽比,  $\hat{u}$  表示预测后的下一帧图像中目标中心的水平像素位置,  $\hat{v}$  表示预测后的下一帧图像中目标中心的垂直像素位置,  $\hat{x}$  表示预测后的下一帧图像中目标区域的面积,  $\hat{y}$  表示预测后的下一帧图像中目标区域的高宽比。

[0041] 5. 步骤S6包括以下步骤:

[0042] S61: 选择YCbCr颜色空间对本帧图像进行手部肤色提取:

$$[0043] \begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 65.738 & 129.057 & 25.06 \\ -37.945 & -74.494 & 112.43 \\ 112.439 & -94.154 & -18.28 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix}$$

[0044] 式中, R表示红色通道, G表示绿色通道, B表示蓝色通道; Y为明亮度; Cb为RGB输入信号蓝色部分与RGB信号亮度值之间的差异; Cr为RGB输入信号红色部分与RGB信号亮度值之间的差异。

[0045] S62: 求高斯模型的马氏距离:

$$[0046] dis_s = (X - m_s)^T C_s^{-1} (X - m_s)$$

[0047] 式中,  $m_s$  表示肤色单高斯模型的平均矢量,  $C_s$  为肤色单高斯模型的协方差矩阵, X为像素点。

[0048] S63: 利用马氏距离分别给肤色和非肤色区域建立单高斯模型, 然后求某一像素在肤色模型和非肤色模型下的马氏距离来判断是否属于肤色点, 并分割出手部图像:

$$[0049] dis_s - dis_{ns} = (X - m_s)^T C_s^{-1} (X - m_s) - (X - m_{ns})^T C_{ns}^{-1} (X - m_{ns}) \leq \tau_{diff}$$

[0050] 式中,  $\tau_{diff}$  为设定的阈值,  $m_{ns}$  表示非肤色单高斯模型的均值,  $C_{ns}$  为高斯模型的协方差。

[0051] S64: 对图像进行二值化处理和双边滤波处理;

[0052] S65: 取轮廓点  $k_i$  与其前面的第m个点  $k_{i+m}$  和后面的第m个点  $k_{i-m}$ , 使用向量  $\overline{k_i k_{i+m}}$  和向量  $\overline{k_i k_{i-m}}$  的夹角余弦值作为点  $k_i$  的曲率:

$$[0053] \cos \theta_i = \frac{\overline{k_i k_{i+m}} \times \overline{k_i k_{i-m}}}{|\overline{k_i k_{i+m}}| \times |\overline{k_i k_{i-m}}|}$$

[0054] S66: 检测具有局部最大K余弦的凸轮廓点作为指尖点:

$$[0055] d_i > (d_{i-s} + d_{i+s}) / 2$$

[0056] 式中,  $d_i$  表示从手掌心最大圆的质心到点的距离。

[0057] 本发明的有益效果: 本发明可应用在以下领域: 1. 空中手写; 2. 智慧教学; 3. 手势检测与跟踪; 4. 基于手部姿态的人机交互。本发明检测部分利用深度神经网络YOLOv3算法对指尖数据集进行训练并得到检测模型, 减弱了复杂环境对检测准确性的影响, 提升了检

测速度,增加了准确性和鲁棒性;跟踪部分利用卡尔曼滤波、匈牙利算法对指尖进行跟踪,并加入K-曲率法对未跟踪到的指尖进行校正,提升了跟踪的实时性和准确性,减少了因指尖移动过快或出现遮挡等情况对跟踪有效性的影响。本发明提供的上述基于深度学习和K-曲率法的指尖跟踪方法,可以有效降低对摄像设备的要求,提升指尖跟踪的准确性及有效性,在实际应用中有较好的应用价值。

### 附图说明

- [0058] 图1为本发明的整体网络结构图;
- [0059] 图2为YOLOv3的网络结构图;
- [0060] 图3为IOU说明及不同IOU情况示例图;
- [0061] 图4为K-曲率法示意图。

### 具体实施方式

[0062] 下面结合说明书附图和具体实施方式对本发明做进一步说明。

[0063] 参见图1,为本发明的整体网络结构图。首先利用YOLOv3网络模型训练预处理后的数据集,获取指尖检测模型;再利用摄像头获取视频流,输入检测模型并检测出检测框信息,初始化卡尔曼滤波器;然后利用卡尔曼滤波器得到预测框,计算出本帧检测框和预测框的IOU,设定IOU阈值,判断该IOU是否大于IOU阈值,若该IOU大于IOU阈值则更新卡尔曼滤波器得到指尖跟踪框;否则,利用K-曲率法对指尖位置进行校正,并更新卡尔曼滤波器;最后设定一个时间阈值T-max,在该时间阈值帧内未检测跟踪信息,则终止跟踪。

[0064] 本发明的基于深度学习和K-曲率法的指尖跟踪方法,它包括以下步骤:

[0065] S1:获取手部数据集,进行预处理;

[0066] 手部数据集需要含有大量不同状态及条件下的手部彩色图,并包含有每张图唯一对应的标签信息;

[0067] 步骤S1包括以下子步骤:

[0068] S11:在不同的场景、光照、人、指尖角度、手部数量及遮挡条件下采集大量包含露出指尖的手部彩色图;

[0069] S12:利用旋转、变形、平移和添加噪声的方法分别以25%的概率对手部彩色图像进行数据扩充;

[0070] S13:标记出目标指尖的区域框,添加标签信息,生成数据标签文件。

[0071] S2:利用深度学习神经网络模型YOLOv3的数据集进行训练,获取指尖检测模型;

[0072] 在目标检测与跟踪领域,目标检测的精确性和快速性至关重要,指尖检测模型的鲁棒性和准确性很大程度决定了后续跟踪部分的准确性;YOLOv3在前两个版本YOLOv1、YOLOv2的基础上进行了改良,在精度达到与R-CNN等two-stage系列算法匹敌的同时,极大的提高了检测速度;Yolov3的网络结构如图2所示;

[0073] 步骤S2包括以下子步骤:

[0074] S21:将手部数据集输入图片转化成416\*416大小的网格图片,增加灰度条防止失真;

[0075] S22:将处理后的图片通过Darknet53特征提取网络降采样5次,生成多尺度的特征

图;Darknet53网络框架参数如表1所示:

[0076] 表1 Darknet53网络框架参数

残差块	层类型	输出通道数	卷积核	输出特征图大小
	卷积	32	3x3	256X256
	卷积	64	3x3/2	128X128
1x 残差块	卷积	32	1x1	
	卷积	64	3x3	
	残差			128x128
[0077] 2x 残差块	卷积	128	3x3/2	64x64
	卷积	64	1x1	
	卷积	128	3x3	
	残差			64x64
8x 残差块	卷积	256	3x3/2	32x32
	卷积	128	1x1	
	卷积	256	3x3	
	残差			32x32
	卷积	512	3x3/2	16x16
8x 残差块	卷积	256	1x1	
	卷积	512	3x3	
	残差			16x16
[0078] 4x 残差块	卷积	1024	3x3/2	8x8
	卷积	512	1x1	
	卷积	1024	3x3	
	卷积			8x8
	平均池化	1024	全局	
	全连接		1000	
	Softmax			

[0079] S23:利用多尺度融合的方式在13\*13、26\*26、52\*52三个尺度的特征图上分别进行卷积检测,得到3次检测结果,并进行非极大值抑制,获取最终结果;

[0080] S24:生成训练后的指尖检测模型。

[0081] S3:利用摄像头获取实时视频流,将当前帧输入指尖监测模型中,利用Darknet53网络对当前帧的图像进行多尺度特征提取,检测出目标类别和检测框位置信息,并根据此信息初始化卡尔曼滤波器;

[0082] S4:利用卡尔曼滤波器得到预测框,计算出本帧的检测框和预测框的IOU,设定IOU阈值,判断该IOU是否大于IOU阈值,当该IOU大于IOU阈值则进行S5步骤,否则进行S6步骤;

[0083] 当某个检测框和所有已有目标在当前帧的预测框之间的IOU都小于了设定的阈值,则认为出现了两种情况:一是检测模型没有检测到指尖;二是出现了新的指尖或者出现误检。简单的利用IOU判断可能出现误检、漏检等情况出现,增加K-曲率法对指尖检测框进行监督和校正,有效的提高了跟踪的正确性。

[0084] 如图3所示,为IOU的示意图;IOU每个检测框与本帧指尖的所有预测框指尖的交并比(IOU)阈值的大小决定了跟踪的准确性,本案例设置IOU阈值为0.7;

[0085] 步骤S4包括以下子步骤:

[0086] S41:读取下一帧图像,利用卡尔曼滤波器得到所有预测到的指尖预测框;

[0087] 使用的卡尔曼滤波器包含七个状态变量,四个观测输入:

$$[0088] \quad \hat{x}_k = A\hat{x}_{k-1} + Bu_{k-1}$$

$$[0089] \quad \hat{P}_k = A\hat{P}_{k-1}A^T + Q$$

$$[0090] \quad \hat{x}_k = \hat{x}_k + K_k(Z_k - H\hat{x}_k)$$

$$[0091] \quad K_k = (\hat{P}_k - H^T)/(H\hat{P}_kH + R)$$

$$[0092] \quad \hat{P}_k = (I - K_kH)\hat{P}_k$$

[0093] 输出目标状态:

$$[0094] \quad X = [u, v, x, y, \dot{u}, \dot{v}, \dot{x}, \dot{y}]^T$$

[0095] 式中,  $\hat{x}_k$  表示k时刻先验状态估计值,  $\hat{x}_k$ 、 $\hat{x}_{k-1}$  分别表示k时刻、k-1时刻后验状态估计值,A表示状态转移矩阵,B表示可选的控制输入  $u \in R^1$  的增益,  $u_{k-1}$  表示k-1时刻的控制增益,  $\hat{P}_k$  表示k时刻的先验估计协方差,  $\hat{P}_k$ 、 $\hat{P}_{k-1}$  分别表示k时刻、k-1时刻的后验估计协方差,Q表示过程激励噪声的协方差,  $K_k$  表示卡尔曼增益,  $Z_k$  表示测量值,H表示测量矩阵,u表示当前帧目标中心的水平像素位置,v表示当前帧目标中心的垂直像素位置,x表示目标区域的面积,y表示目标区域的高宽比,  $\dot{u}$  表示预测后的下一帧图像中目标中心的水平像素位置,  $\dot{v}$  表示预测后的下一帧图像中目标中心的垂直像素位置,  $\dot{x}$  表示预测后的下一帧图像中目标区域的面积,  $\dot{y}$  表示预测后的下一帧图像中目标区域的高宽比。

[0096] S42:计算出本帧检测框和所有预测框的IOU;

[0097] S43:利用匈牙利算法得到IOU最大的唯一匹配;

[0098] S44:设定IOU阈值,判断该IOU是否大于IOU阈值;

[0099] S45:该IOU大于IOU阈值则进行S5步骤,否则进行S6步骤。

[0100] S5:用本帧匹配到的预测框去更新卡尔曼滤波器,并将状态更新值作为本帧的检测框输出,该帧跟踪完成,回到步骤S4;

[0101] S6:利用K-曲率法得到指尖点的位置信息;

[0102] 在进行肤色区域分割时,由于YCbCr颜色空间对肤色具有很好的聚类效果,比较容易实现RGB到YCbCr颜色空间的转换,选择YCbCr颜色空间对手势区域提取;而得到的二值化手势图像带有大量噪声,因此选用双边滤波对其进行平滑处理;由于手的轮廓有明显的高低差,可以通过高值来检测指尖,而K-曲率法能有效的测量曲线点的夹角,因此选用K-曲率法检测出指尖点。如图4所示,为K-曲率法示意图;

[0103] 步骤S6包括以下子步骤:

[0104] S61:选择YCbCr颜色空间对本帧图像进行手部肤色提取;

$$[0105] \quad \begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 65.738 & 129.057 & 25.06 \\ -37.945 & -74.494 & 112.43 \\ 112.439 & -94.154 & -18.28 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix}$$

[0106] 式中,R表示红色通道,G表示绿色通道,B表示蓝色通道;Y为明亮度;Cb为RGB输入信号蓝色部分与RGB信号亮度值之间的差异,取 $77 < Cb < 127$ ;Cr为RGB输入信号红色部分与RGB信号亮度值之间的差异,取 $133 < Cr < 173$ 。

[0107] S62:求高斯模型的马氏距离:

$$[0108] \quad dis_s = (X - m_s)^T C_s^{-1} (X - m_s)$$

[0109] 式中, $m_s$ 表示肤色单高斯模型的平均矢量, $C_s$ 为肤色单高斯模型的协方差矩阵,X为像素点。

[0110] S63:利用马氏距离分别给肤色和非肤色建立单高斯模型,然后求某一像素在肤色模型和非肤色模型下的马氏距离来判断是否属于肤色点,并分割出手部图像:

$$[0111] \quad dis_s - dis_{ns} = (X - m_s)^T C_s^{-1} (X - m_s) - (X - m_{ns})^T C_{ns}^{-1} (X - m_{ns}) \leq \tau_{diff}$$

[0112] 式中, $\tau_{diff}$ 为设定的阈值, $m_{ns}$ 表示非肤色单高斯模型的均值, $C_{ns}$ 为高斯模型的协方差。

[0113] S64:对图像进行二值化处理和双边滤波处理;

[0114] S65:取轮廓点 $k_i$ 与其前面的第m个点 $k_{i+m}$ 和后面的第m个点 $k_{i-m}$ ,使用向量 $\overline{k_i k_{i+m}}$ 和向量 $\overline{k_i k_{i-m}}$ 的夹角余弦值作为点 $k_i$ 的曲率:

$$[0115] \quad \cos \theta_i = \frac{\overline{k_i k_{i+m}} \times \overline{k_i k_{i-m}}}{|\overline{k_i k_{i+m}}| \times |\overline{k_i k_{i-m}}|}$$

[0116] 式中,m的大小决定了指尖点计算的准确性,本案例中取m为5。

[0117] S66:检测具有局部最大K余弦的凸轮廓点作为指尖点:

$$[0118] \quad d_i > (d_{i-s} + d_{i+s}) / 2$$

[0119] 式中, $d_i$ 表示从手掌心最大圆的质心到点的距离。

[0120] S7:分别计算该指尖点和检测框的中心点的欧式距离,并设定一个阈值,当该距离小于阈值时,则初始化卡尔曼滤波器得到新预测框重新开始匹配;否则删除掉此帧的跟踪信息,读取下一帧;

[0121] 对于使用检测框信息初始化新目标的位置信息,速度设置为0,由于此时无法观测到速度,因此速度分量的协方差设置较大初始值,反应不确定性。新的追踪目标需要经历一段待定时间去和检测结果进行关联以累计新目标的置信度,这可以有效防止目标检测的虚警造成的新追踪目标误创建。

[0122] S8:设定一个时间阈值T-max,在该时间阈值帧内未检测跟踪信息,则终止跟踪;

[0123] 如果连续T-max帧没有已有指尖的预测位置和检测框的IOU匹配,则认为指尖消失,终止轨迹。这可以防止跟踪器数量的无限增长,以及由于长时间的预测而导致的定位误差。本案例设置T-max为1。

[0124] 上述实施例为本发明的最佳实施方式,但本发明的实施方式不受限于上述实施

例,其他的任何不脱离本发明原理所作的诸如简化、改变、替换、组合的置换方式,都包含在本发明的保护范围之类。

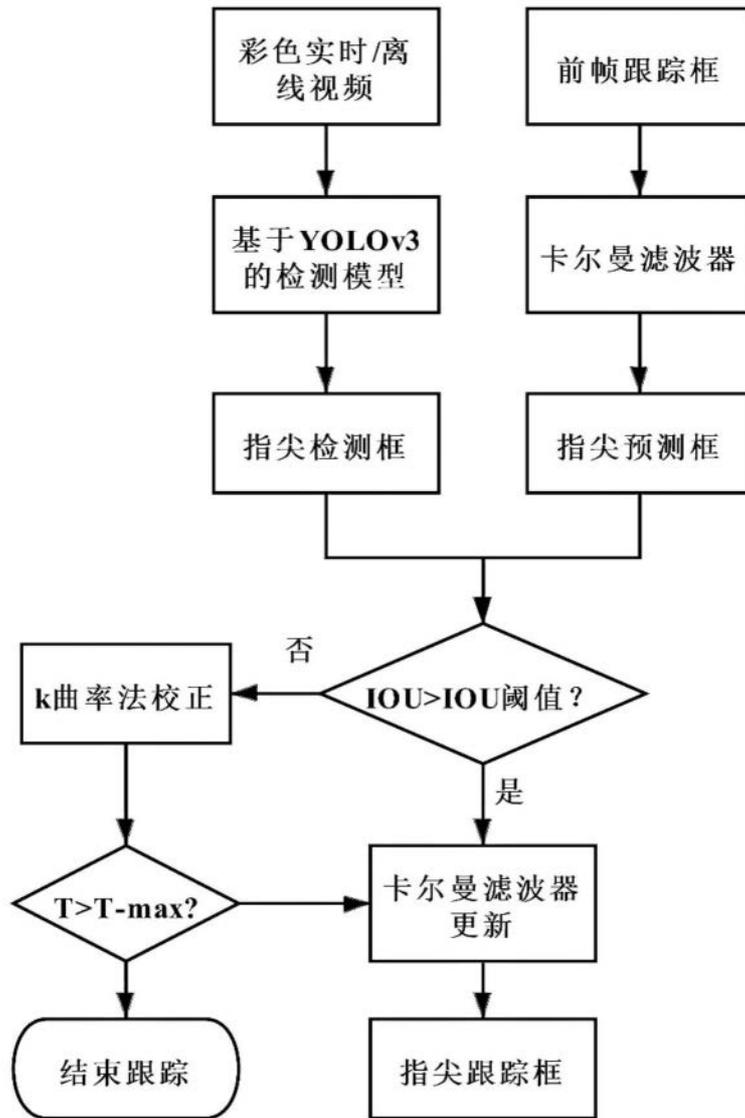


图1

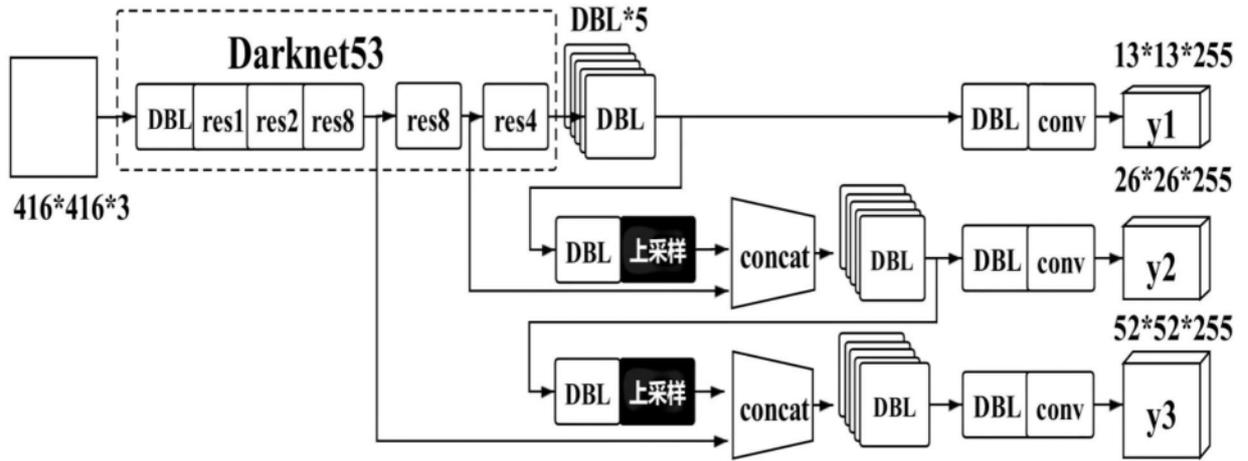


图2

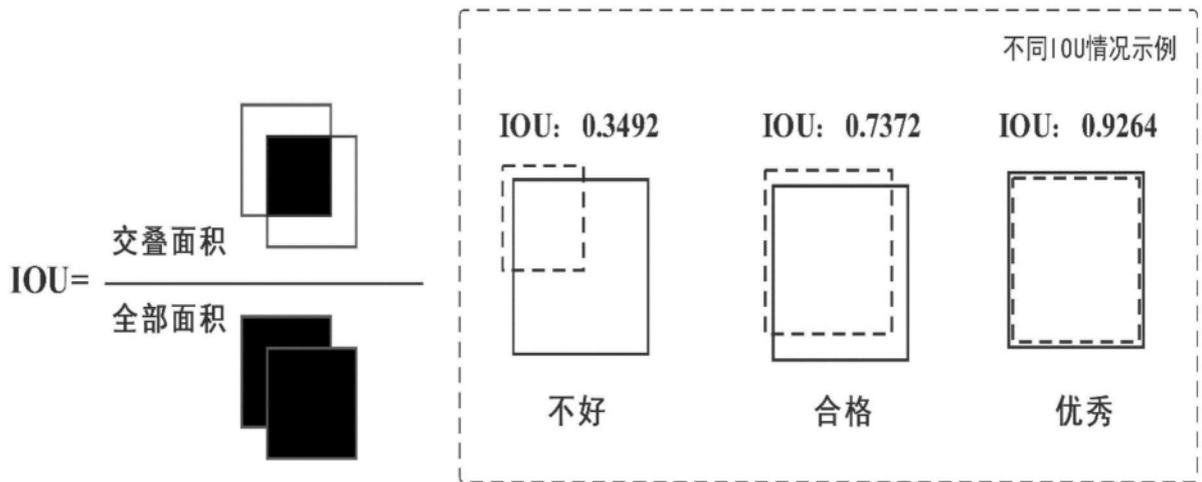


图3

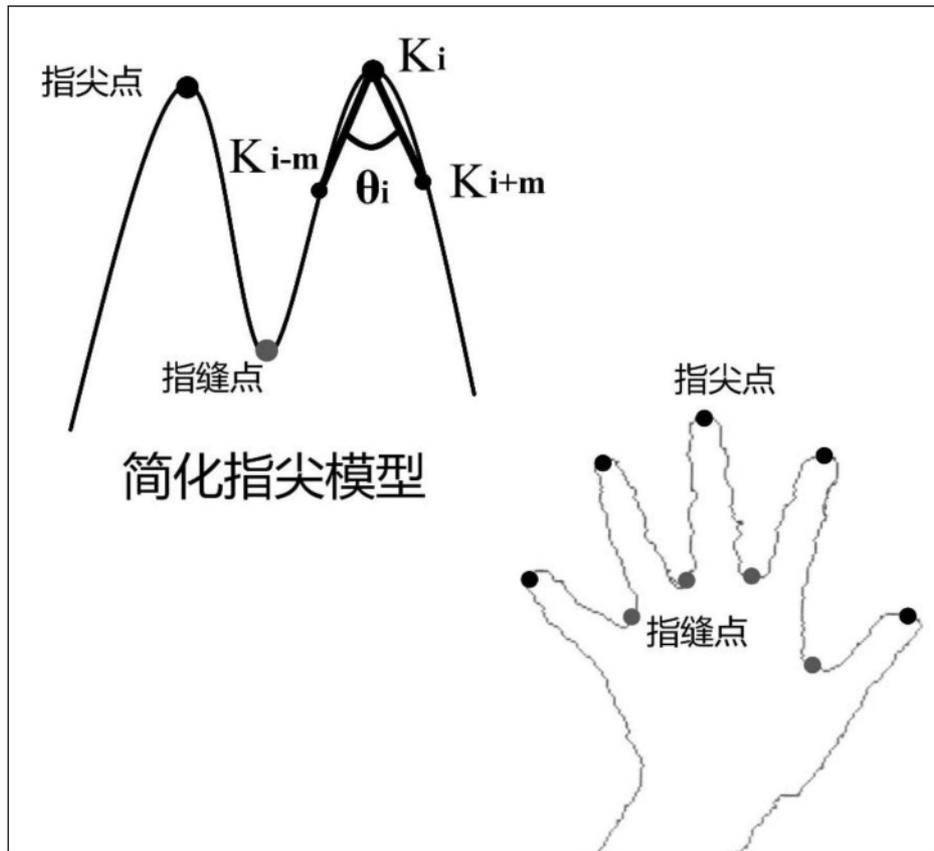


图4