



(12)发明专利申请

(10)申请公布号 CN 111158491 A

(43)申请公布日 2020.05.15

(21)申请号 201911423743.9

G06T 7/11(2017.01)

(22)申请日 2019.12.31

(71)申请人 苏州莱孚斯特电子科技有限公司
地址 215300 江苏省苏州市昆山市玉山镇
祖冲之南路1699号工研院北楼辅2-32
号,33号工位

(72)发明人 王震洲

(74)专利代理机构 石家庄新世纪专利商标事务
所有限公司 13100

代理人 甄伊宁 董金国

(51)Int.Cl.

G06F 3/01(2006.01)

G06K 9/00(2006.01)

G06K 9/62(2006.01)

G06T 5/40(2006.01)

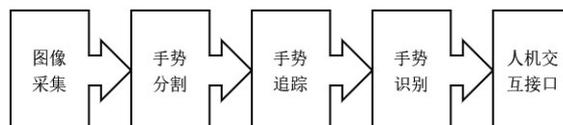
权利要求书1页 说明书19页 附图6页

(54)发明名称

一种应用于车载HUD的手势识别人机交互方法

(57)摘要

本发明公开了一种应用于车载HUD的手势识别人机交互方法,涉及人机交互领域,包括以下步骤:通过手势信息采集模块获取驾驶员的手势信息;通过数据处理识别模块将所述手势信息处理为手势指令;通过车载HUD模块执行所述手势指令,并将操作信息展示在所述驾驶员的驾驶视野内的一个物体上。本发明将手势识别技术与车载HUD结合起来,改变了传统车载HUD的交互方式,解决了现存的驾驶分心的问题,从而营造了一个安全驾驶的环境。



1. 一种应用于车载HUD的手势识别人机交互方法,其特征在于,包括以下步骤:通过手势信息采集模块获取驾驶员的手势信息;通过数据处理识别模块将所述手势信息处理为手势指令;通过车载HUD模块执行所述手势指令,并将操作信息展示在所述驾驶员的驾驶视野内的一个物体上。

2. 根据权利要求1所述的一种应用于车载HUD的手势识别人机交互方法,其特征在于:所述的摄像头为单目摄像头。

3. 根据权利要求1所述的一种应用于车载HUD的手势识别人机交互方法,其特征在于:所述通过数据处理识别模块为车载中控台。

4. 根据权利要求1所述的一种应用于车载HUD的手势识别人机交互方法,其特征在于:所述展示操作信息的物体为挡风玻璃。

5. 根据权利要求1至4任一项所述的一种应用于车载HUD的手势识别人机交互方法,其特征在于,通过以下方法将手势信息处理为手势指令:

使用背景差分法和肤色检测法将包含手势信息的图像分割为手势部分和背景部分;

使用Hu矩结合手势轮廓凸紧性提取所述手势部分的手势特征;

对所述手势特征通过支持向量机进行分类,将手势特征的至少一个类对应于一个手势定义,并以其对应手势定义作为手势指令。

6. 根据权利要求5所述的一种应用于车载HUD的手势识别人机交互方法,其特征在于:所述肤色检测法包括基于直方图的自适应肤色分割法。

7. 根据权利要求6所述的一种应用于车载HUD的手势识别人机交互方法,其特征在于:所述基于直方图的自适应肤色分割法中的直方图包括颜色直方图和色度直方图。

8. 根据权利要求1至4任一项所述的一种应用于车载HUD的手势识别人机交互方法,其特征在于,通过以下方法将手势信息处理为手势指令:

对包含手势信息的图像中角点进行流光检测得到手势区域,通过消除背景干涉光流的方法得到手势区域的手势部分;

通过提取多个所述手势部分的手势中心点得到手势轨迹,通过动态时间规整算法对所述手势轨迹进行分类,将手势轨迹的至少一个类对应于一个手势定义,并以其对应手势定义作为手势指令。

9. 根据权利要求8所述的一种应用于车载HUD的手势识别人机交互方法,其特征在于:所述包含手势信息的图像为动态手势序列中的关键帧。

10. 根据权利要求8所述的一种应用于车载HUD的手势识别人机交互方法,其特征在于:通过结合卡尔曼滤波器的光流法对包含手势信息的图像中包含手势部分的手势区域进行预测。

一种应用于车载HUD的手势识别人机交互方法

技术领域

[0001] 本发明涉及人机交互领域,具体涉及一种具有手势识别功能的车载HUD系统。

背景技术

[0002] 人机交互(Huamn-Comeputer Interaction)是指系统与用户的交互关系,人机交互系统可以看作一个带有输出输入功能的机器。人机交互方式也从最开始的命令行,转换为表情、语言、手势、体态等符合人类自然习惯的交互方式,大大提升了交互的便利性。为了获得更贴近人的行为习惯的交互方式,研究的重点也从以计算机为中心转变为以人为中心。随着计算机视觉、信号处理等领域的发展,对手势识别、人脸识别、声音识别的研究也越来越深入。

[0003] 在现代交通中,一些司机在驾驶的过程中接打电话的行为存在很大的安全隐患,许多导致重大交通事故的原因是由于驾驶员在驾驶中操作手机。一些法律中,明令禁止驾驶员在驾驶中使用手机,避免驾驶员注意力分散从而保障行车安全。伴随着计算机视觉技术的高速发展,汽车行业为满足驾驶需求,设计开发辅助行车系统使得驾驶员在驾车过程中保持视线不离开路面的同时,通过非接触的操作控制设备。这样可以有效地减少因注意力不集中而导致的交通事故。此外,无人驾驶汽车逐渐成为研究的热点,手势识别也是其中重要的组成部分。

[0004] 抬头显示仪(Heads up display),简称HUD。最早运用在军事驱动器上,飞行员无需低头就可以获取仪表盘信息,不妨碍正常飞行视野观察,减少了分心驾驶的问题。车辆在行驶的过程中,驾车者需要查看仪表盘或操作中控触摸屏幕时,所有的注意力都集中在车内,容易造成交通事故。为了解决这一问题,随着智能车载系统的出现,很多汽车在出厂或在后续改装中均可以安装车载HUD系统,通过外加设备将车内仪表盘信息投影到挡风玻璃上,有效解决了由于对道路注意力分散导致的交通事故。在不同系统手机中均可安装具有HUD功能的软件,即在移动端实现车载抬头显示仪的功能。目前软件市场上,IOS系统有HUDWAY的应用程序,Android系统有Navier HUD平视导航。此外,车载HUD可以采用外置或者OBD的连接方式,同样有不同的系统选择。

[0005] HUD最早应用在汽车上是在1988年,福特公司在“Oldsmobile”汽车上首次应用了HUD,此后越来越多的HUD应用在高级汽车上。宝马汽车是第一家运用车载HUD的公司,在之后宝马也在更多类型的汽车上配备了抬头显示设备。此外,丰田、本田、奔驰等汽车的部分车型了也提供车载HUD功能。目前车载HUD大多数采用触摸式操作的人机交互方式,这种方法的优势在于便捷、直观的操作,而缺点是接触式操作需要驾驶员视线集中在触摸界面才能实现功能,这样分散了驾驶员的注意力,加大了驾驶危险系数。

[0006] 计算机视觉是一门让机器拥有“眼睛”的科学技术,随着社会科学的发展,这一技术也逐渐成为研究的热点。更直观的说,就是采用摄像头或其他图像采集设备模仿人眼的功能,计算机通过获取的图像进行处理、分析,然后将结果反馈为具体功能的实现。主要应用图像处理技术,从而更方便获取信息的图像。将手势识别技术应用到车载HUD上形成完整

的人工交互系统,系统通过摄像头获取原始图像,然后经过算法处理原始图像,从而达到识别别人手的静态动作或者动态轨迹的效果,最后将其转化为实时、准确的指令。

[0007] 手势是各个手指之间进行不同姿势的组合或是手的一个或多个部位在空间中运动形成的轨迹,采用独立表达或者组合表达语义的方式予以表示。手势识别技术于1983年由Grimes提出,通过数据手套获得手势特征并取得专利,也因此被认定为手势识别技术的创始人。在之后直到90年代才有了较大发展, Christopher LEE和Xu于1999年研究出手势识别系统用来控制机器人。Tew与 Gray利用动态规整时间算法,建立了手势识别系统,可识别10种手势,准确率为98.3%。Mahmoud Elmezain等人基于肤色分割法和Camshift跟踪算法,建立了动态手势识别系统。在国内,研究手势识别起步较晚。台湾大学研究人员基于数据手套研究出手势识别系统,成功率为90%。哈尔滨工业大学高文等人将人工神经网络与马尔隐夫算法相结合,获得了较高的识别率。

[0008] 现有技术的手势识别系统,包括手势图像采集、手势分割、手势跟踪、特征提取、手势识别及人机交互接口。其处理数据的步骤包括:第一步,进行手势图像的获取并对其预处理;第二步,为提取完整的手势区域,进行手势区域分割;第三步,手势区域跟踪主要用于动态手势识别中,避免出现在有遮挡物的情况下手势丢失的情况;第四步,从手势区域中提取手势特征并进行分类从而完成手势识别;最后一步,将手势识别模块通过人机交互接口送入计算机,从而得到手势对应的计算机命令实现人机交互。

[0009] 目前已经存在大量对手势识别技术的研究,但仍存在这许多亟需解决的问题,重难点主要为以下几方面:

[0010] 1) 周围光照变化:行驶中的车辆会经过环境光照会产生变化,可能经过城市鳞次栉比的高楼大厦,也可能行驶在空旷的田野小路上,又或者是三回九转,时而迎光时而背光的盘山路。当驾驶员行驶在不同的道路时,车载HUD系统所处环境的光线明暗变化,使得摄像头提取的手势图像也存在明暗不同的情况,这就加大了手势分割的难度,容易产生识别错误的情况;

[0011] 2) 复杂的背景:驾驶员在行驶过程中,摄像头捕捉到的图像背景复杂,具有不确定性,影响手势识别结果及识别正确率高低,对手势识别系统具有重要意义。从有其他物体干扰的复杂背景中,将手势区域分割出来,是手势识别重要且关键的一步。复杂的背景会给手势分割带来影响,尤其是在有近似肤色的部分存在,会导致手势识别结果出现误差。仅采用单一阈值的肤色分割法,容易造成手势目标和背景的混淆,错将类肤色区域判断为肤色区域,影响最后手势识别的结果;

[0012] 3) 手的复杂性与高自由度:人的手势是一个有超过20个自由度(DOF)的物体,手势具有较大的状态空间,可以产生平移、尺度变化以及旋转等动作。这为实现精准的手势识别带来了极大的挑战性;

[0013] 4) 实时性与精确性的平衡:简单的算法具有较快的运算速度,但无法保证识别的精确性。于是,研究者提出了各种复杂的算法来提高识别的精确度,由于这些复杂算法的计算量巨大,使得系统无法保证实时性。如何找到精确度和实时性之间的平衡点,也是亟需解决的问题。

[0014] 在实际手势识别应用中,在满足高鲁棒性的要求下解决上述问题,是手势识别系统的技术难点。

发明内容

[0015] 本发明目的在于,提供一种基于普通单目摄像机获取的二维信息图像的人机交互系统,尤其适用于车载HUD。大多假设受外界环境影响较小。此外,摄像机捕捉到的手势与人脸重合的几率较大,且无法避免这种情况的发生,这时手势颜色区域、边缘信息都变得模糊,这大大增加了手势分割的难度。

[0016] 本发明提供的技术方案是:

[0017] 第一方面的技术方案提供了一种应用于车载HUD的手势识别人机交互方法,包括以下步骤:通过手势信息采集模块获取驾驶员的手势信息;通过数据处理识别模块将所述手势信息处理为手势指令;通过车载HUD模块执行所述手势指令,并将操作信息展示在所述驾驶员的驾驶视野内的一个物体上。

[0018] 上述技术方案的第一优选为,所述的摄像头为单目摄像头。

[0019] 上述技术方案的第二优选为,所述通过数据处理识别模块为车载中控台。

[0020] 上述技术方案第三优选为,所述展示操作信息的物体为挡风玻璃。

[0021] 上述各技术方案的一个适于静态手势识别的改进为,通过以下方法将手势信息处理为手势指令:

[0022] 使用背景差分法和肤色检测法将包含手势信息的图像分割为手势部分和背景部分;

[0023] 使用Hu矩结合手势轮廓凸紧性提取所述手势部分的手势特征;

[0024] 对所述手势特征通过支持向量机(SVM)进行分类,将手势特征的至少一个类对应于一个手势定义,并以其对应手势定义作为手势指令。

[0025] 一些适于静态手势识别的改进的具体实施例中,所述肤色检测法包括基于直方图的自适应肤色分割法。进一步的,所述基于直方图的自适应肤色分割法中的直方图包括颜色直方图和色度直方图。

[0026] 上述各技术方案的一个适于动态手势识别的改进为,通过以下方法将手势信息处理为手势指令:

[0027] 对包含手势信息的图像中角点进行流光检测得到手势区域,通过消除背景干涉光流的方法得到手势区域的手势部分;

[0028] 通过提取多个所述手势部分的手势中心点得到手势轨迹,通过动态时间规整算法对所述手势轨迹进行分类,将手势轨迹的至少一个类对应于一个手势定义,并以其对应手势定义作为手势指令。

[0029] 一些适于动态手势识别的改进的方法实施例中,所述包含手势信息的图像为动态手势序列中的关键帧。

[0030] 一些适于动态手势识别的改进的方法实施例中,通过结合卡尔曼滤波器的光流法对包含手势信息的图像中包含手势部分的手势区域进行预测。

[0031] 第二方面的技术方案提供了一种基于手势识别的车载HUD人机交互系统,包括用于获取携带驾驶员的手势信息的序列帧的手势信息采集模块、用于将所述手势信息处理为手势指令的数据处理识别模块以及用于执行所述手势指令的车载 HUD模块,其中车载HUD模块将操作信息展示在所述驾驶员的驾驶视野内的一个物体上。

[0032] 上述技术方案的第一优选为,所述的手势信息采集模块为单目摄像机。

- [0033] 上述技术方案的第二优选为,所述通过数据处理识别模块为车载中控台。
- [0034] 上述技术方案的第三优选为,所述展示操作信息的物体为挡风玻璃。
- [0035] 上述各技术方案的一个适于静态手势识别的改进为,包括以下将手势信息处理为手势指令模块:
- [0036] 手势分割模块,用于使用背景差分法和肤色检测法将包含手势信息的图像分割为手势部分和背景部分;
- [0037] 手势特征提取模块,用于使用Hu矩结合手势轮廓凸紧性提取所述手势部分的手势特征;
- [0038] 手势识别模块,用于对所述手势特征通过支持向量机进行分类,将手势特征的至少一个类对应于一个手势定义,并以其对应手势定义作为手势指令。
- [0039] 一些适于静态手势识别的改进的具体实施例中,所述肤色检测法包括基于直方图的自适应肤色分割法。进一步的,所述基于直方图的自适应肤色分割法中的直方图包括颜色直方图和色度直方图。
- [0040] 上述各技术方案的一个适于动态手势识别的改进为,包括以下将手势信息处理为手势指令模块:
- [0041] 手势分割模块,用于对包含手势信息的图像中角点进行流光检测得到手势区域,通过消除背景干涉光流的方法得到手势区域的手势部分;
- [0042] 手势识别模块,通过提取多个所述手势部分的手势中心点得到手势轨迹,通过动态时间规整算法对所述手势轨迹进行分类,将手势轨迹的至少一个类对应于一个手势定义,并以其对应手势定义作为手势指令。
- [0043] 一些适于动态手势识别的改进的系统实施例中,所述包含手势信息的图像为动态手势序列中的关键帧。
- [0044] 一些适于动态手势识别的改进的系统实施例中,通过结合卡尔曼滤波器的光流法对包含手势信息的图像中包含手势部分的手势区域进行预测。
- [0045] 本发明将手势识别技术与车载HUD结合起来,改变了传统车载HUD的交互方式,解决了现存的驾驶分心的问题,从而营造了一个安全驾驶的环境。

附图说明

- [0046] 图1为本发明一个实施例的手势识别系统工作原理示意图;
- [0047] 图2为本发明一个实施例中背景差分法和肤色检测法进行手势分割的过程示意图;
- [0048] 图3为本发明一个实施例中背景实时更新时的背景差分法和肤色检测法鲜果示意图;
- [0049] 图4为本发明一个实施例中手势轮廓及其凸包的示意图;
- [0050] 图5为本发明一个实施例中金字塔流光法示意图;
- [0051] 图6为本发明一个实施例中FAST算子示意图;
- [0052] 图7为本发明一个实施例中光流法中动态手势光流矢量效果对比示意图;
- [0053] 图8为本发明一个实施例中距离变化图像及手势中心点的示意图;
- [0054] 图9为本发明一个实施例中动态手势轨迹提取过程的流程示意图;

[0055] 图10为本发明一个实施例中动态手势轨迹识别的流程示意图；

[0056] 图11为本发明一个实施例中一种基于手势识别的车载HUD人机交互系统的系统框图。

具体实施方式

[0057] 首先应当说明的是,根据输入手势特征分为静态手势识别和动态手势识别两方面。普通单目摄像机仅能获取二维信息图像,不具备深度摄像机获取深度信息,这也为手势识别加大了难度。现存的基于普通单目摄像头的人机交互研究,大多假设受外界环境影响较小,不能考虑环境影响。此外,摄像机捕捉到的手势与人脸重合的几率较大,且无法避免这种情况的发生,这时手势颜色区域、边缘信息都变得模糊,这也增加了手势分割的难度。

[0058] 本发明通过对现有技术以下方面的改进形成的总的发明构思,实现了一种应用于车载HUD的手势识别人机交互方法及一种基于手势识别的车载HUD人机交互系统。

[0059] 第一方面的改进是,提供了新的复杂背景下的手势分割。本发明技术方案通过实时更新背景图减少光照变化的影响,建立新的肤色模型用来排除类肤色物的干扰,并将背景差分法和肤色检测法结合起来进行手势分割;通过金字塔 Lucas-Kanade光流法进行动态手势分割,并通过提取关键帧及其角点减少计算量,提高分割速率。

[0060] 第二方面的改进是,提供了新的动态手势跟踪方法解决手势运动速度及方向发生剧烈变化时跟踪手势丢失的问题。本发明在金字塔Lucas-Kanade光流法与卡尔曼滤波器相结合的方法,提升手势跟踪准确性。

[0061] 第三方面的改进是,提供了手势特征提取与手势分类方法。选择基于Hu矩和手势轮廓凸紧性的支持向量机法进行静态手势识别;针对动态手势识别,通过提取手势中心点获取手势轨迹特征,最后通过动态时间规整算法对动态手势进行分类。

[0062] 第四方面的改进是,提供了手势—车载HUD人工交互系统。实现了人与车载 HUD的手势交互,该系统让驾驶员通过非接触式操作即可完成对车载系统的各种指令。

[0063] 下面通过多个方法实施例描述本发明提供的方法技术方案以便本例关于技术人员予以实施和改进。总的来说,这些方法实施例基于适用于静态手势识别的方法步骤S110、S120、S130和S140,或者,基于适用于动态手势识别的方法步骤S210、S220和S230。本领域技术人员可以根据这些实施例实现如图1所示的手势识别系统,其包括手势图像采集、手势分割、手势跟踪、特征提取、手势识别及人机交互接口。第一步,进行手势图像的获取并对其预处理。第二步,为提取完整的手势区域,进行手势区域分割。第三步,手势区域跟踪主要用于动态手势识别中,避免出现在有遮挡物的情况下手势丢失的情况。第四步,从手势区域中提取手势特征并进行分类从而完成手势识别。最后一步,将手势识别模块通过人机交互接口送入计算机,从而得到手势对应的手势指令或者计算机命令实现人机交互。

[0064] 步骤S110,静态手势分割。具体的,本步骤中,结合肤色信息与运动信息进行手势分割,是一种多模式融合的手势分割方法,一个实施例中,可以包括以下步骤S111至S112:

[0065] 步骤S111,对摄像头获取的第一图像分别进行肤色检测以及背景差分处理后,进行与运算,获得第二图像。第二图像中包括手势部分和背景部分。

[0066] 步骤S112,对所述第二图像的手势部分进行连通域分析,将手势部分分成几个连

通域部分,计算各连通域像素点总数,过滤像素点较少部分,从而得到手势区域。

[0067] 在步骤S110的一个实施例中,所述肤色检验步骤具体为步骤S310至步骤 S330:

[0068] 步骤S310,选择颜色空间,所述颜色空间在本步骤的第一实施例中为RGB、XYZ、CMY、CMYK等基色混合型颜色空间,在本步骤的第二实施例中为HSV、HSL 等色调、饱和度、强度型颜色空间,在本步骤的第三实施例中为YUV、YCbCr、YIQ、Lab等亮度、色度型颜色空间。

[0069] 步骤S320,设置肤色空间模型,通过肤色空间可以将像素点分为肤色与非肤色,且通过代数形式描述,得到像素点和肤色的近似度。所述肤色空间模型包括:1) 阈值模型,阈值由颜色空间各分量的边界值表示,体现出肤色在颜色空间中的聚类性。当像素点的分量均在阈值所限定的范围内,判定该像素点为肤色点。2) 参数模型,包括椭圆模型和高斯模型。椭圆模型认为在rg、IQ、CbCr 等颜色空间上,肤色分布近似于椭圆;高斯模型通过概率密度函数估计像素点属于肤色点的概率,又可以分为单高斯模SGM和高斯混合模型GMMs。3) 非参数模型,在一些实施例中包括直方图模型以及基于模式识别的分类器等,其中,直方图模型把颜色空间转化为一组直方图bin,bin和颜色方位相对应,通常分外查找表法和贝叶斯法两种;基于模式识别的分类器可以将数据一般化,采用逼近复杂的非线性输入输出关系的方法。

[0070] 步骤S330,基于肤色空间对肤色进行自适应分割获取手势部分,一些实施例中也称目标部分,以及背景部分。

[0071] 在步骤S330一个优选的实施例中,通过在YCbCr颜色空间中建立基于直方图的动态模型,建立了一种自适应的肤色分割法。该实施例中,根据手势的颜色信息实时更新肤色模型。通过统计手势样本的肤色,得到初始CbCr颜色直方图,通过高斯分布计算直方图提高分割速度。根据某一像素点的色度值Cb' 与Cr' 找到相应的P(Cb',Cr') 单元,如图2所示。累计所述单元P邻域单元的高斯权重,由此可得到色度直方图。接下来对直方图阈值化可以获得查找表,若某一单元值大于预设阈值则设为1,否则为0。通过多次训练样本可得在直方图最大值的0.05倍时分割肤色区域效果较好。

[0072] 在步骤S110的一个实施例中,所述背景差分方法,指基于动态建模的背景差分法去检测分割手势区域,并实时更新背景图,其具体包括步骤S410至S420:

[0073] 步骤S410,通过如下式(2-14)与式(2-15)处理摄像头采集的一个图像帧,即前帧图像,并作为输入图获得背景部分的差分图像,即利用当前帧图像f与背景图像B做差分运算获得差分图像D,在这里我们同时考虑三个色道:

$$D(x, y) = |f_Y(x, y) - B_Y(x, y)| + |f_{Cb}(x, y) - B_{Cb}(x, y)| + |f_{Cr}(x, y) - B_{Cr}(x, y)| \quad (2-14)$$

$$M(x, y) = \begin{cases} 1 & \text{if } D(x, y) > T \\ 0 & \text{otherwise} \end{cases} \quad (2-15)$$

[0076] 式中, $f_Y f_{Cb} f_{Cr}$ 表示输入图的Y,Cb,Cr, B_Y, B_{Cb}, B_{Cr} 表示背景图值的Y,Cb,Cr,。

[0077] 比较固定阈值法和动态阈值法,p-tile法和双峰法具有很大的局限性,其中p-tile法的前提是得到手势的占比,而双峰法要求两个峰分别为背景和目标,且需要有明显的波谷,因此不考虑。

[0078] 本发明的一些实施例中,可以基于固定阈值法、Otsu法、迭代法或者均值聚类法的分割结果,如图3所示。在一个优选的实施例中,选择自适应阈值法实施手势分割,在该实施

例中,为降低运算压力,优选通过迭代法得到自适应阈值 T 。

[0079] 步骤S420,背景建模,获取背景图像,设置累计图Acc和计数图Num两个储存区域,Acc累加每帧的差分图像,Num储存像素点的累加次数,包括步骤S421 至S424:

[0080] 步骤S421通过第1帧图像初始化背景图像与累计图,Num中所有像素点的初始值为1。

[0081] 步骤S422,从第2帧开始,用当前帧 $f_k(x,y)$ 与上一帧 $f_{k-1}(x,y)$ 做帧间差分。假如得到 $|f_k(x,y)-f_{k-1}(x,y)|$ 较小,则把对应点Acc(x,y)值上加 $f_k(x,y)$, Num(x,y)的值加1。这些计算只针对非目标区域,从而避免将前景信息纳入背景模型中。

[0082] 步骤S423,重复步骤S422,直到第N帧。用下式(2-16)更新背景图:

$$[0083] \quad B(x,y) = \text{Acc}(x,y) / \text{Num}(x,y) \quad (2-16)$$

[0084] 步骤S424,设置累计图为背景图像,Num设为1。重复2)、3),实现背景图像的更新。

[0085] 本发明的一个实施例中,上述的动态的背景建模是通过帧差法实现的,所述帧差法在步骤S210的动态手势分割实施例中公开的方法实现。

[0086] 在步骤S110的一个具体实施例中,以单目摄像头手势信息采集模块获取包含手势信息的相继的多帧图像。如图2所示,分割过程中各阶段的图像结果。首先,对一个具体的图像帧进行肤色检测处理,处理后图像如图2,a)所示,同时对其进行本经差分处理,处理后图像如图2,b)所示,然后将两个图像进行与运算,运算结果如图2,c)所示,其中白色区域为手势部分,黑色区域为背景部分。最后白色部分进行连通域分析,将前景的白色部分分成几个连通域部分,计算各连通域像素点总数,过滤像素点较少部分,从而得到手势区域。如图2,d)中的方框所示处即为手势区域,用于候选手势部分。

[0087] 图3显示在一个动态变化的背景环境的实施例中,通过上述处理后得到的手势区域。其中每一横行为一个图像帧的处理过程,第一列为原始图像帧,第二列为肤色检测结果,第三列为背景差分结果,第四列为求与后结果,第五列为连通域分析后的结果。

[0088] 对比更新与不更新的肤色检测法与背景差分法,当背景中出现类肤色以及人的干扰时,上述算法具有较好的分割效果。

[0089] 在一个具体的实施例中,肤色检测法与背景差分法通过以下步骤实施,通过摄像头获取的图像建立背景图像B,差分图像D利用当前帧图像f与背景图B做差分运算获得,如下式(2-2)。将差分图像D进行二值化,其中T表示目标检测过程中分割背景和前景的合适阈值。

$$[0090] \quad D(x,y) = |f(x,y) - B(x,y)| \quad (2-2)$$

$$[0091] \quad M(x,y) = \begin{cases} 1 & \text{if } D(x,y) > T \\ 0 & \text{otherwise} \end{cases} \quad (2-3)$$

[0092] 环境噪声对背景差分法的影响较大,光照和背景场景变化都会导致前景目标的错误检测。背景模型选取的好坏直接影响到手势目标分割的结果是否精确,因此如何增强背景建模的鲁棒性以及及时更新背景图像是我们研究的重点。时间平均模型是根据相连帧图像求平均,其中图像序列中的低频分量被选作背景图像。设 $B_t(x,y)$ 和 $f_t(x,y)$ 为t时刻的背景图与图像帧, $B_t(x,y)$ 按下式(2-4)更新。

$$[0093] \quad B_t(x,y) = \alpha B_{t-1}(x,y) + (1-\alpha) f_t(x,y) \quad (2-4)$$

[0094] α 表示学习率,会直接影响背景更新,并且环境对时间平均模型的影响较大,造成

其有很大的局限性。

[0095] 自适应的背景模型指,在连续采集的图像流中,首先将第一帧、第二帧图像分别当作背景图像 $B(x, y)$ 和目标图像 $T(x, y)$,其次通过边缘分割法对目标图像 $T(x, y)$ 进行边缘提取,即可获得可能的手势区域。接下来用目标图像生成掩模图 $Mask(x, y)$,再去检测之前得到的可能的手势区域。若区域中有大于 $2/3$ 的像素点分布在肤色范围内,将范围内的像素点值设为1,其余为0。最后,根据下式(2-5)更新背景图,保留值为1的像素点,用目标图相应的点对值为0的像素点进行替换。

$$[0096] \quad B_t(x, y) = \begin{cases} B_{t-1}(x, y) & \text{if } Mask(x, y) = 1 \\ T_{t-1}(x, y) & \text{if } Mask(x, y) = 0 \end{cases} \quad (2-5)$$

[0097] 背景模型或者肤色模型的建模方法包括单高斯模型、混合高斯模型或者核密度估计等背景模型。在一个改进中,用高斯模型对背景像素点的值进行描述,如式(2-6)所示。该方法适用于比较稳定的环境。

$$[0098] \quad p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (2-6)$$

[0099] 其中, μ 代表均值, σ 代表标准差。每当获取到新的图像帧,首先会判断像素点:

$$[0100] \quad |x-\mu| < 2.5\sigma \quad (2-7)$$

[0101] 若像素点满足式(2-7),可以判定该像素点为背景点,否则为前景点。在实际应用中背景可能是在变化的,所以对背景模型也就是参数进行更新,如下式所示:

$$[0102] \quad \mu_{i+1} = (1-\alpha)\mu_t + \alpha x_{i+1} \quad (2-8)$$

$$[0103] \quad \Sigma_{i+1} = (1-\alpha)\Sigma_t + \alpha(x_{i+1}-\mu_{i+1})(x_{i+1}-\mu_{i+1})^T \quad (2-9)$$

[0104] 式中, μ_t 更新前的高斯分布均值为 μ_t , μ_{t+1} 更新后的高斯分布均值为 μ_{t+1} ,更新前的协方差矩阵用 Σ_t 表示,更新后的协方差矩阵用 Σ_{t+1} 表示, x_{t+1} 为 $t+1$ 时刻像素点值, α 为学习率,且 α 的值在0到1之间,直接影响背景更新速度。 α 太小导致背景更新速度太慢,会将背景中的静态物体误认为手势目标; α 太大则导致背景更新速度太快,运动的物体目标会被视为背景,并且噪声影响增大。灰度图像中 Σ_t 为 σ^2 ,在彩色图像中,各个像素点的颜色分量是独立的,所以 Σ_t 化简为 $diag[\sigma_R^2, \sigma_G^2, \sigma_B^2]$ 。

[0105] 本发明的一个改进在于二值化阈值的确定,上述背景差分法中,式(2-2)中的 T 表示目标检测过程中,分割背景和前景的合适阈值,若 T 为固定值则无法满足所有的情况,为获得更好的手势目标分割效果,本文采用自适应的阈值获取法。下面几种常见的自适应阈值获取方法适用于本发明:

[0106] 1) p-tile法

[0107] 这种方法假设已知手势目标占整张图像面积比例,设手势区域占整图面积的 $1/p$ 。通过累加直方图上的像素点,得到下一阈值 T ,因此灰度值小于 T 的像素点占总像素点的 $1/p$ 。

[0108] 2) 双峰法

[0109] 手势区域与背景区域间灰度差较大时,在图像直方图上呈双峰状分布。其中,图像的边缘处在直方图上表现为波谷,阈值设定根据波谷的灰度值所决定。在得到双峰的灰度值后,首先对直方图进行平滑处理,避免得到局部最大值。双峰法简单且易实现,但当手势

和背景有重叠时,不适用该方法。

[0110] 3) 类间差分法

[0111] 类间差分法通过灰度直方图与最小二乘理论得到最佳阈值。选定一个阈值 λ 将统计到的灰度直方图划分为两部分,再通过式(2-10)得到两部分的类间方差值:

$$[0112] \quad g = w_0(\mu_0 - \mu)^2 + w_1(\mu_1 - \mu)^2 \quad (2-10)$$

[0113] 式中, w_0 表示小于 λ 的部分所占比例,该部分的平均值为 μ_0 , w_1 为大于 λ 部分所占比例,该部分的平均值为 μ_1 ,直方图的平均值为 μ 。当 g 为最大值时 λ 为最佳阈值。该方法在图像直方图具有双峰特性时取得的效果最好,但在手势目标与背景灰度差较小时不适用。

[0114] 4) 基于均值聚类的最优阈值法

[0115] 基于均值聚类法,计算图像中主要灰度区域的最小误差,从而得到分割阈值。

[0116] 具体步骤如下:

[0117] a. 通过设定初始阈值 $T(0)$,图像可分为 C_1 和 C_2 两类,分别计算其均值与方差,设为 $\mu_1, \mu_2, \sigma_1, \sigma_2$ 。

[0118] b. 重新对像素点进行分类,若满足下式(2-11)条件:

$$[0119] \quad |f(i, j) - \mu_1| \leq |f(i, j) - \mu_2| \quad (2-11)$$

[0120] $f(i, j)$ 属于 C_1 , 否则属于 C_2 。

[0121] c. 重新计算分类后的 C_1 和 C_2 的均值与方差,若满足下式(2-12):

$$[0122] \quad [p_1\sigma_1^2 + p_2\sigma_2^2]_{T(k)} \geq [p_1\sigma_1^2 + p_2\sigma_2^2]_{T(k-1)} \quad (2-12)$$

[0123] 输出当前时刻的阈值 $T(k)$ 。如不满足,则重复b,c两步。

[0124] 5) 基于迭代的最优阈值法

[0125] 通过迭代得到最优阈值,步骤如下:

[0126] a. 设初始阈值 $T(0)$,通常为图像的灰度平均值;

[0127] b. 根据阈值 $T(k)$ 分割图像,其中 k 为迭代次数。分割后图像分为2个区域,分别记作 $C_1(k)$ 与 $C_2(k)$;

[0128] c. 通过计算 $C_1(k)$ 和 $C_2(k)$ 的平均灰度值 u_1 与 u_2 ,可得新的阈值:

$$[0129] \quad T(k+1) = \frac{\mu_1(k) + \mu_2(k)}{2} \quad (2-13)$$

[0130] d. 重复b,c步,在 $T(k+1)$ 与 $T(k)$ 的差值小于一定值是停止迭代。

[0131] 步骤S120,静态手势提取,即手势特征的提取,是指采集可以表示手势的特征,从而可以通过特征集合,判别是否为手势。在静态手势识别中,选择合适的特征直接影响分类器的结果。部分手势特征会受环境变化的影响,导致手势特征提取的难度增大。目前常见的静态手势识别所用特征如下:

[0132] 1) 原始图像

[0133] 可将检测分割后手势图像的灰度、颜色等信息作为特征,这种特征提取方法简单,但是鲁棒性较差,环境及光照变化等外界因素对其影响较大。在基于深度学习算法的手势识别中,有不错的效果。

[0134] 2) 统计特征

[0135] 统计分析图像中灰度及梯度分布信息,得到图像的统计特征,其包括 Haar-like 特征、梯度方向直方图(HOG)以及主成分分析(PCA)。其对环境变化的鲁棒性较强。利用机器

学习法获得的特征代表性较好,能够有效提高识别率。

[0136] 3) 手掌、手指结构

[0137] 手掌和手指最能够表达手势的特征,手掌与手指相互间的结构特征经常被用做手势姿势的简单识别。该方法的缺点是只能识别张开的手指,对于拳头等手势不适用。

[0138] 4) 基于数学变换的特征

[0139] 经过数学变换的原始手势图像得到的系数,也可作为特征。比如,常用的傅立叶描述子特征。傅里叶描述子算法具有平移与旋转不变的特性,且计算简单。但基于数学变换的特征具有较差的鲁棒性,且描述手势的信息不够完备。

[0140] 5) 统计矩

[0141] 统计矩可以表示手势目标的形状特征,且具有平移与旋转不变得特性。

[0142] 6) 手势轮廓凸紧性

[0143] 轮廓的凸紧性常用于图像识别中,尤其物体轮廓凹凸性较强。如图4,图中外圈的线为手势的凸包,可以很好的表示手势的形状特征。

[0144] 本发明的一个实施例中,手势特征的提取通过用Hu矩结合手势轮廓凸紧性提取所述手势部分的手势特征,包括几何矩的求解和主成分分析。

[0145] $p+q$ 阶标准几何矩在离散数字图像 $f(x,y)$ 中定义为:

$$[0146] \quad m_{pq} = \sum_{x=1}^M \sum_{y=1}^N x^p y^q f(x,y) \quad (2-17)$$

[0147] 其中, N,M 表示图像的高度与宽度, p,q 取自然数。图像的整体特征用低阶矩表示,而图像的细节用高阶矩来表示,比如扭曲度、峰态等。若 $f(x,y)$ 是灰度图像,图像的灰度值总和用零阶矩 m_{00} 来表示;若 $f(x,y)$ 是二值图时,图像的面积用 m_{00} 来表示;若 $f(x,y)$ 是手势轮廓图时,轮廓周长用 m_{00} 来表示。图像的质心 (x_c, y_c) 是由一阶矩 m_{10} 与 m_{01} 来表示,如下式2-18:

$$[0148] \quad x_c = \frac{m_{10}}{m_{00}}, y_c = \frac{m_{01}}{m_{00}} \quad (2-18)$$

[0149] 二阶矩可以用作计算目标的长短轴、方向角。

[0150] 中心距通过将坐标原点移动到质心来表示:

$$[0151] \quad \mu_{pq} = \sum_{x=1}^M \sum_{y=1}^N (x - x_c)^p (y - y_c)^q f(x,y) \quad (2-19)$$

[0152] 中心矩具有平移不变的特性,但是会根据图像大小而变化。通过对中心距进行归一化使得中心距具有尺度不变的特性:

$$[0153] \quad \eta_{pq} = \frac{\mu_{pq}}{(\mu_{00})^\gamma} \quad (2-20)$$

[0154] 在区域矩中, $\gamma = (p+q+2)/2$;在边界矩中, $\gamma = p+q+1$ 。

[0155] Hu矩由7个不变量矩组成,如下式(2-21)所示。其显著特点是平移、旋转和尺度不变性。

$$[0156] \quad \left\{ \begin{array}{l} M_1 = \eta_{20} + \eta_{02} \\ M_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\ M_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\ M_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\ M_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\ \quad (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ M_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + \\ \quad 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\ M_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - \\ \quad (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{array} \right. \quad (2-21)$$

[0157] 通过之前的研究表明,低阶矩中包含了手势图像中大部分的有效信息,高阶矩中的细节信息受噪声影响较大,并且计算复杂程度较大。一个实施例中,用 Hu矩的前四个矩 $M_1 \sim M_4$ 来表示手势特征。

[0158] 主成分分析(PCA)为一个线性变换矩阵,来源于特征手算法的特征提取部分,将原始的数据从高维空间投影到低维子空间。所以为了降低计算量,可以通过选择特征来实现。通常用一个1维向量 x , 表示一幅样本图像,训练样本集为 $\{x_i | i=1, 2, \dots, N\}$, 由 N 个样本组成。可通过下式(2-22)计算该样本集的平均图像:

$$[0159] \quad \mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (2-22)$$

[0160] 式中, x_i 表示样本图, $x_i - \mu$ 表示 x_i 相对于 μ 的均差,从而构成了矩阵 $X = [x_1 - \mu, x_2 - \mu, \dots, x_N - \mu]$ 。协方差矩阵 $Q = XX^T$ 可由构造的训练样本集得到,协方差矩阵的正交特征向量会组成新的特征空间,就得到了特征手。

[0161] 通过引入奇异值分解原理(SVD),可以直接求解得到 Q 的特征向量。若人手的样本数 N 跟像素点数 i 相比小很多,可以通过求解 $Q' = X^T X$ 的特征向量 v 得到 Q 的特征向量 λ 。 Q' 的大小为 $N \times N$,大大降低了计算的复杂程度。利用 Q' 的特征向量 $V = [v_1, v_2, \dots, v_N]$ 与 X 计算得到 Q 的特征向量 $U = [u_1, u_2, \dots, u_N]$:

$$[0162] \quad U = X \cdot V = [x_1, x_2, \dots, x_N] [v_1, v_2, \dots, v_N] \quad (2-23)$$

[0163] 这就是所谓的特征手,或者手势特征的一种,本发明的手势识别中不需要考虑到所有的 u_k ,绝大多数有用信息保存在一部分特征向量中。特征按照 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ 从大到小排列,线性变化矩阵由前 n ($n < N$) 个对应的特征向量构成。 n 通过下式(2-24)确定:

$$[0164] \quad n = \underset{r}{\operatorname{argmin}} \left(\sum_{i=1}^r \lambda_i / \sum_{j=1}^N \lambda_j > \theta_\lambda \right) \quad (2-24)$$

[0165] 式中, θ_λ 表示阈值。 θ_λ 的取值为样本集在前 n 个轴上的能量分配比重。

[0166] 步骤S130,静态手势分类。手势特征提取后进行手势分类,最后将手势与手势定义对应起来,从而实现具体功能。根据获取的手势特征,可以采取不同的分类器进行分类,分类器的效果受手势特征的质量、不同手势差异、手势的几何变化等多种因素的影响。本发明的实施例中,分类器包括基于最小距离以及基于判别函数两类,基于最小距离的分类器,将手势模式识别为距离其最近的样本所在的类别;基于判别函数的分类器是根据输出值进行

手势分类,最大输出值的类别为手势类别。

[0167] 步骤S140,静态手势识别。静态手势识别方法,即对单帧图像上的手势进行识别,不需要考虑一系列帧图像中的运动信息。各类识别方法的区别,在于手势建模的方式以及特征距离的计算方法。常用的静态手势识别方法,主要有基于模板匹配的方法、基于决策树的方法、基于K最邻近的方法、基于支持向量机的方法、基于主成分分析的目标识别方法等。

[0168] 1) 基于模板匹配的方法

[0169] 基于模板匹配的方法,首先将连续的手势图像分解成单帧图像,然后将需要识别的手势目标与已有的手势模板进行比较,计算出两者之间的相似程度,从而识别出手势。这是一种简单方便的识别方法,但由于手势变化较多,需要大量的手势模板作支撑,故识别计算量较大。Y Li与J P Wachs提出一种具有权重机制的模板匹配算法,可精确识别手势类别。

[0170] 2) 基于决策树的方法

[0171] 决策树是一种树型结构,一个对象对应树中的一个节点,可能的属性值对应分叉路径。例如赵雅静等人建立了基于决策树的两级识别系统,进行对自定义的八种手势的分类,实现了很好的识别效果。决策树容易实现,对背景知识要求较小。但其错误率会随着识别种类的增多而加大,且有连续图像的分类所需时间较多。

[0172] 3) K最邻近算法

[0173] K最邻近算法(K-nearest neighbor,KNN)是机器学习领域一种经典的分类算法,无需参数且手势分类结果效果显著,通过比较数据间的距离或相似度得到分类结果。其实现手势分类原理:若一个待分类手势在特征空间中的距离最近,即特征空间最邻近的K个手势样本都属于一种类别。D Dahmani与S Larabi融合离散正交矩、Hu矩以及手势凸包组成手势特征,并通过K邻近算法(KNN)和支持向量机(SVM)进行手语识别。该方法计算简单,但需要大量的测试数据才可以得到K个最邻近点,会产生大量的计算,对于车载HUD系统来说不适用。

[0174] 4) 卷积神经网络

[0175] 卷积神经网络(CNN),是仿照人的大脑神经建立的非线性数学模型,它可以通过原始图像的非线性核函数映射,生成高层次的特征表达。CNN具有较强的学习能力,精确的分类结果。因此,卷积神经网络广泛适用于各种图像和视频分类识别中,如图像分类、目标分割、行为定位与识别等。李杰在卷积神经网络的基础上,使用三个模态间联合训练的方式达到手势识别的目的。CNN的训练过程比较复杂,对CPU的要求较高,并不适用于本文的系统环境。

[0176] 5) 支持向量机

[0177] 支持向量机(SVM)是一种基于统计学习理论的机器学习方法,是静态手势识别常用的方法。通过非线性变化,将不可分的特征向量转换为可分的特征向量,寻找最优分类面从而完成手势分类。比起其他机器学习法,稀疏性以及特征相关性对SVM无影响,并且很好的解决了非线性、高维数、局部极小等问题。但在数据量较大时,SVM的训练时间较长。

[0178] 6) 目标识别

[0179] 主成分分析是通过分析特征 u_1, u_2, \dots, u_N ,从而构造一个低维空间,该空间的每个点对应一副手势图像。通过手势图像向该空间的投影或得坐标系数,从而实现手势识别。基

于特征手的主成分分析方法分为两步,即训练阶段和测试阶段。

[0180] 训练过程中,通过2.2.2节介绍的主成分分析手势类别建立特征手空间,通过映射得到一个n维向量 Ω ,见下式(2-25):

$$[0181] \quad \Omega = U^T(x-\mu) \quad (2-25)$$

[0182] 总类别数为 N_c ,每个类别均有一个均值向量 $\Omega_k, k=1,2,\dots,N_c$ 。利用下式(2-26)可以得到估计距离特征子空间的值 θ_c :

$$[0183] \quad \theta_c = \frac{1}{2} \max_{j,k} (\|\Omega_j - \Omega_k\|) \quad j, k = 1, 2, \dots, N_c \quad (2-26)$$

[0184] 在测试阶段,将手势图像 x 映射到手势特征空间,通过式(2-27)得到一个特征向量 Ω 。再根据距离函数进行手势分类。可利用欧式距离或Mahalanobis距离,在欧氏距离中:

$$[0185] \quad \varepsilon_k^2 = \|\Omega - \Omega_k\|^2, \quad k = 1, 2, \dots, N_c \quad (2-27)$$

[0186] 测试图像 x 通过特征空间重建的图像 x_f 为:

$$[0187] \quad x_f = U\Omega + \mu \quad (2-28)$$

[0188] 估计 x 与 x_f 的距离:

$$[0189] \quad \varepsilon^2 = \|x - x_f\|^2 \quad (2-29)$$

[0190] 如果 $\varepsilon < \theta_c$ 且 $\varepsilon_{k^*} = \min_k \{\varepsilon_k\} < \theta_c$,那么手势图像属于第 k^* 类;否则,判定输入的手势图像不属于现有种类。

[0191] 上述静态手势识别的多个方法实施例提供了静态手势分割、静态手势特征提取以及静态手势分类方法。

[0192] 针对静态手势分割技术,提供了基于肤色的手势分割方法,以及基于背景差分法的手势分割方法。分析目前在静态手势分割方面需要解决的问题,结合本发明手势识别应用的车内环境的复杂度,提出一种结合基于肤色与背景差分法的静态手势分割法。排除了其他物体的干扰,解决了光照变化等问题。

[0193] 针对静态手势特征提取技术,提供了几何矩以及主成分分析特征提取方法,基于在车载HUD的应用环境中各种特征的适用性。在手势识别部分对特征以及分类器进行评估,最后选定Hu矩与手势轮廓凸紧性相结合的特征表示方法。高效的表示了手势的特点,提高识别正确率。

[0194] 针对静态手势分类技术,即分类器的选择。本发明结合Hu矩、手势轮廓凸紧性特征且用支持向量机实现静态手势识别的方法,有较高的识别正确率和较强的鲁棒性。上述实施例是优选实施例,其处理方案并不局限于静态手势的识别,在本发明的另一些实施例中,也可以利用上述方法对动态手势进行识别。

[0195] 动态手势是由一系列的帧图像组成,具有手势运动速度快、单帧图像上手势特征不明显的特点,在动态手势识别中重点需要关注手势运动的轨迹路线,而不再关注手势本身的特点。动态手势对比静态手势更具有灵活性,接近人们生活习惯,也可以在人机交互中表达更多的含义。本发明提供的动态手势识别方法主要包括动态手势分割、动态手势跟踪、动态手势轨迹特征提取以及动态手势分类。

[0196] 本发明的一个实施例中,通过关键帧提取方法对手势信息采集模块获取的原始帧序列进行处理,得到手势运动的关键帧。利用FAST角点检测算法提取关键帧图像中的角点信息,通过光流法计算角点的光流矢量,并通过设定阈值排除背景中运动物体的影响。在此

基础上,通过卡尔曼滤波器与光流法的结合进行手势跟踪,在有遮挡物手势消失的情况下仍可以得到手势完整的运动轨迹。最后将关键帧手势部分的中心点作为手势运动轨迹,并用动态时间规整算法进行最后的动态手势识别。通过光流法分析运动目标,计算相邻图像帧的光流场差值判断帧图像间是否存在运动。当判定不存在运动时可以删除前一帧图像,图下式(3-1)所示

$$[0197] \quad \text{Keyframe}(i) = \begin{cases} 1 & \text{if } I_{Flow} > T \\ 0 & \text{otherwise} \end{cases} \quad (3-1)$$

[0198] 式中,Keyframe(i)的值为1时,则该帧为关键帧。若为0时,则删除该帧。 I_{Flow} 为第i帧和i+1帧图像的光流场差值,阈值用T表示。

[0199] 步骤S210,动态手势分割。在动态手势中,本发明实施例主要利用手势的运动信息分割手势区域,具体的,通过光流法得到手势的运动信息,为了提高光流估计的准确性,我们采用基于多尺度的金字塔Lucas-Kanade光流法,或称金字塔光流法。

[0200] 一个实用光流法提取原始帧序列的关键帧的具体实施例中,金字塔光流法建立过程分三步:首先,建立金字塔,顶层分辨率最低,底层分辨率最高;然后,基于金字塔进行光流跟踪;最后,进行迭代过程得出图像角点光流矢量结果。如步骤S211、S212、S213所述。

[0201] 步骤S211,利用原始帧序列建立金字塔。

[0202] $I^0 = I$ 是第0层的图像,图像的宽度和高度分别为 $n_{x0} = n_x$ 和 $n_{y0} = n_y$,L代表金字塔的层数。本实施例中算法中L取3。图像 I^L 如下式(3-7)得:

$$[0203] \quad I^L = \frac{1}{4} I^{L-1}(2x, 2y) + \frac{1}{8} (I^{L-1}(2x-1, 2y) + I^{L-1}(2x+1, 2y) + I^{L-1}(2x, 2y-1) + I^{L-1}(2x, 2y+1)) \\ + \frac{1}{16} (I^{L-1}(2x-1, 2y-1) + I^{L-1}(2x+1, 2y-1) + I^{L-1}(2x-1, 2y+1) + I^{L-1}(2x+1, 2y+1)) \quad (3-7)$$

[0204] 步骤S212,金字塔跟踪。

[0205] 金字塔跟踪算法是将最顶层图像的仿射矩阵和光流作为初始值传递到下一层图像,直到传递到最后一层即原始图像层,将最后一层的光流及仿射变化矩阵作为结果,如图5所示。定义图像中角点u在第L层对应点的坐标为 $u^L = [u_x^L, u_y^L]$,计算出 $u^L = \frac{u}{2^L}$ 。光流计算初值为 $g^L = [g_x^L, g_y^L]$,随后计算仿射变换矩阵, $G^L = [1 + g_{xx}^L, g_{xy}^L, g_{yx}^L, 1 + g_{yy}^L]$ 。在L层上计算光流和仿射变换矩阵需重新定义L层上的匹配误差 ε^L :

$$[0206] \quad \varepsilon^L(d^L, A^L) = \varepsilon^L(d_x^L, d_y^L, d_{xx}^L, d_{xy}^L, d_{yx}^L, d_{yy}^L) \\ = \sum_{x=-W_y}^{W_y} \sum_{y=-W_y}^{W_y} (I_{comp}^L(x) - J_{comp}^L(A^L x + d^L))^2 \quad (3-8)$$

[0207] 其中,图像 $I_{comp}^L(x)$ 和 $J_{comp}^L(x)$ 是L层上的采样图像,基于L层光流 g^L 和仿射矩阵初值 G^L 计算出两个对应图像 $I_{comp}^L(x)$ 和 $J_{comp}^L(x)$:

$$[0208] \quad \begin{cases} I_{comp}^L(x) = I^L(x + u^L) \\ J_{comp}^L(x) = J^L(G^L x + g^L + u^L) \end{cases} \quad (3-9)$$

[0209] 将结果传递给下一层,计算下一层的初值:

$$[0210] \quad \begin{cases} g^{(L-1)} = 2(g^L + G^L d^L) \\ G^{(L-1)} = G^L A^L \end{cases} \quad (3-10)$$

[0211] 再次将 g^{L-1} 和 G^{L-1} 作为初值,继续循环以上步骤,直到传递到最后一层即原始图像。

[0212] 步骤S213,迭代过程。

$$[0213] \quad I_x = [I(x+1, y) - I(x-1, y)] / 2 \quad (3-11)$$

$$[0214] \quad I_y = [I(x, y+1) - I(x, y-1)] / 2 \quad (3-12)$$

[0215] 根据上式得出空间梯度矩阵:

$$[0216] \quad G \leftarrow \sum_{x=-W_x}^{W_x} \sum_{y=-W_y}^{W_y} \begin{pmatrix} I_x^2 & \cdots & y I_x I_y \\ \vdots & \ddots & \vdots \\ y I_x I_y & \cdots & y^2 I_y^2 \end{pmatrix} \quad (3-13)$$

[0217] 更新光流 $v = 2 * v$,计算后一帧图像中对应角点的灰度 $J(x) \leftarrow J^L(Ax+v)$,将两帧图像对应角点的灰度值做差 $\delta I(x) \leftarrow I(x) - J(x)$,并计算两帧图像间的误差向量:

$$[0218] \quad \bar{b} \leftarrow \sum_{x=-W_x}^{W_x} \sum_{y=-W_y}^{W_y} \begin{bmatrix} I_x \delta I \\ \vdots \\ y I_y \delta I \end{bmatrix} \quad (3-14)$$

[0219] 最后计算针对仿射光流 $[\eta_x \quad \eta_y \quad \eta_{xx} \quad \eta_{xy} \quad \eta_{yx} \quad \eta_{yy}]^T$ 更新跟踪结果。

$$[0220] \quad \begin{bmatrix} A & v \\ 0 & 1 \end{bmatrix} \leftarrow \begin{bmatrix} A & v \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 + \eta_{xx} & \eta_{xy} & \eta_x \\ \eta_{yx} & 1 + \eta_{yy} & \eta_y \\ 0 & 0 & 1 \end{bmatrix} \quad (3-15)$$

[0221] 直到 $\sqrt{\eta_x^2 + \eta_y^2}$ 小于本文所设阈值,结束在这一层的迭代过程。

[0222] 在一个改进的实施例中,步骤S210通过以下方法步骤S215、S216和S217实现。这些步骤实现了对包含手势信息的图像中角点进行流光检测得到手势区域,通过消除背景干涉光流的方法得到手势区域的手势部分。该方法实施例中,运用FAST角点检测算法结合金字塔光流法进行动态手势检测,提高手势分割效率。再通过Mean Shift算法,消除检测过程中背景部分干扰角点光流,提高分割准确率。以下具体实现步骤。

[0223] 步骤S215,手势角点特征提取。角点是手势重要的局部特征,集合了大量的重要图像信息,具有旋转不变性以及对照的鲁棒性。本发明算法通过角点检测,优化光流法检测速度与精度。优选的,实用FAST角点检测算法。

[0224] FAST角点检测算法使用圆周长为16个像素点(半径为3的Bresenham圆)来判定其圆心像素是否为角点,如图6所示。将与圆心点差异较大的点连成连续圆弧,且圆弧长度大于整圆周长的3/4,那么判断圆心点为角点,该圆上的其他点不需再做判断,因此算法的计算速度较快。

[0225] FAST角点检测采用非极大值抑制方法,需要通过指定角点响应强度的方法,角点响应强度为连续圆弧上像素点与圆心点差值的绝对值之和,如式(3-16)所示。

$$[0226] \quad E = \sum_{x \in \text{circle}(p)} |I(x) - I(p)| > \varepsilon_d \quad (3-16)$$

[0227] 式中,E为响应强度, $I(x)$ 为圆周上任意一点的像素值, $I(p)$ 为候选点的像素值, ε_d 为角点阈值。

[0228] 步骤S216,光流法手势分割,运用金字塔光流法进行手势检测。金字塔光流法建立过程分三步:首先,建立金字塔,顶层分辨率最低,底层分辨率最高;然后,基于金字塔进行光流跟踪;最后,进行迭代过程得出图像角点光流矢量结果。优选的,在背景中存在其他运动物体的情况下,检测结果包括背景中非手势运动部分时,采用Mean Shift算法对背景光流进行搜索,角点光流矢量与区分阈值进行比较,从而判定该角点属于手势区域还是背景干扰区域。

[0229] 步骤S217,消除背景光流干扰,选择光流矢量图中一点 x 作为圆心,圆的半径为 h ,得到一个向量 $\overline{M_{h,g}(x)}$,设 $\overline{M_{h,g}(x)} = 0$ 可得到新的球心,更新球心值后得到新的向量 $\overline{M_{h,g}(x)}$,循环到向量的模小于给定误差时结束。新的球心为光流矢量中密度最大的点,其坐标就是背景的光流矢量。设待检测的光流矢量为 $G(x,y)$, $B(x,y)$ 为通过Mean Shift算法搜索到背景光流矢量,将背景光流矢量作为判断阈值,有 $T(x,y) = B(x,y)$,运动目标的二值图像 $O(x,y)$ 为:

$$[0230] \quad O(x,y) = \begin{cases} 1 & G(x,y) \neq T(x,y) \\ 0 & G(x,y) = T(x,y) \end{cases} \quad (3-17)$$

[0231] 通过式(3-17)将背景光流和动态手势光流分割开来,同时提高了光流矢量计算的实时性及准确性,将背景的光流矢量去除,同时提高了分割中运算速度。图7中a)为普通光流法获取手势区域和角点的结果图像,b)为实用本实施例获取手势区域和角点的结果图像。

[0232] 步骤S220,动态手势跟踪。在动态手势跟踪算法中,需要考虑到动态手势常有模糊运动、目标遮掩、干扰环境等影响因素,跟踪结果直接影响手势识别的实时性和鲁棒性。跟踪算法包括Meanshift算法、Camshift算法、光流法。Meanshift算法是基于无参密度估计方法,通过迭代实现对目标的跟踪,但其检测结果容易出现漏检情况。Camshift算法的运行效率比较高,其优势在于可调整目标区域适应手势大小变化,但在背景颜色与手势目标相接近时,容易跟踪失败。使用光流法进行跟踪的前提是假设跟踪目标一直存在,即无遮挡物情况,这在实际情况中很难实现。

[0233] 动态手势跟踪中,动态手势视频中手势的运动方向为二个或者三个。用时间 $K-1$ 的运动状态函数表示时间 K 的状态描述,如式(3-18)所示。

$$[0234] \quad x_k = Fx_{k-1} + Bu_k + w_k \quad (3-18)$$

[0235] 其中, x_k 表示当前运动状态的 n 维向量表示, F 为一个 $n \times n$ 的传递矩阵, u_k 为输入控制的 c 维向量, B 表示关联输入控制和状态改变的 $n \times c$ 矩阵, w_k 为过程噪声,服从高斯分布。

[0236] Z_k 表示 m 维观测值,观测值 Z_k 与状态变量的直接测量 x_k 不成等价关系。观测值 Z_k 的 m 维向量可用式(3-19)表示。

$$[0237] \quad Z_k = H_k x_k + v_k \quad (3-19)$$

[0238] 式中, H_k 为一个 $m \times n$ 的观测模型矩阵,从现实状态映射为观测状态,观测噪声用 v_k 表示,服从协方差矩阵 R_k ($m \times m$)的高斯分布,即 $v_k \sim N(0, R_k)$ 。

[0239] 根据状态 x 在每一个时间步长更新转移矩阵 F 和输入控制矩阵 B ,卡尔曼滤波器通过预测下一帧图像中手势的位置和速度,减少检测区域,检测时间仅仅是对整个区域检测所用时间的1/6。

[0240] 本实施例的追踪算法采用卡尔曼滤波器估计下一帧中跟踪手势所在区域,采用光流法进行手势跟踪,跟踪效果如图8所示,其中,外框所示区域为卡尔曼滤波器进行手势预测区域,而后利用光流法进行检测跟踪的准确手势区域为内框区域。通过卡尔曼滤波器预测出手势可能存在的区域,同时减小了角点检测的范围,从而优化检测精度以及速度的要求。

[0241] 步骤S220的一些实施例中,为了提取动态手势运动轨迹,在光流法分割出手势区域中进行中心点的检测。对手势区域进行肤色分割得到手势区域的二值图像,参考之前的手势中心点提取,人们通常认为手掌的重心即为手掌的中心点,下式(3-20)给出了手掌重心的计算公式:

$$[0242] \quad \bar{x} = \frac{\sum_i \sum_j i \times f(i, j)}{\sum_i \sum_j f(i, j)}, \bar{y} = \frac{\sum_i \sum_j j \times f(i, j)}{\sum_i \sum_j f(i, j)} \quad (3-20)$$

[0243] 式中,像素点 (i, j) 处的像素值为 $f(i, j)$, \bar{x} 和 \bar{y} 代表重心的纵横坐标。

[0244] 这种提取手势中心的办法在手指张开时,检测误差较大,这会对后期手势轨迹的采集产生不利影响。设手势部分的像素值为1,其余区域像素值设为0,得到手势二值图像。

[0245] 本发明的一个实施例中,采用对手势二值图像进行距离变化得到手势中心点,图像的距离变换是通过计算手势目标和背景部分的距离,通过得到的距离矩阵以及灰度图像,从而得到距离图像。目标与背景之间其中离背景越远的点其数值越大,在图像中越亮。因此,认为在手势的距离图像中最亮的点即为手势的中心点。

[0246] 图像I的大小为 $m \times n$,其中行坐标为 $x (1 \leq x \leq m)$ 、列坐标为 $y (1 \leq y \leq n)$,目标部分为 $o = \{(x, y) | I(x, y) = 1\}$,背景部分为 $B = \{(x, y) | I(x, y) = 0\}$,Dist 为距离图像,如下式(3-21)所示。

$$[0247] \quad Dist(x, y) = \begin{cases} 0 & (x, y) \in B \\ \min(d((x, y), (x_0, y_0))) & (x, y) \in o, (x_0, y_0) \in B \end{cases} \quad (3-21)$$

[0248] 距离变换包括两类,欧式距离变换和非欧式距离变换。本发明的一些实施例中,距离变换采用欧式距离、棋盘距离、城市街区距离及近欧式距离。取图像上的两点 x 和 y ,距离表达式如式(3-22)至式(3-25)所示:

[0249] 欧式距离:

$$[0250] \quad d((x, y), (x_0, y_0)) = \sqrt{(x - x_0)^2 + (y - y_0)^2} \quad (3-22)$$

[0251] 棋盘距离:

$$[0252] \quad d((x, y), (x_0, y_0)) = |x - x_0| + |y - y_0| \quad (3-23)$$

[0253] 城市街区距离:

$$[0254] \quad d((x, y), (x_0, y_0)) = \max(|x - x_0|, |y - y_0|) \quad (3-24)$$

[0255] 近欧式距离:

$$[0256] \quad d((x, y), (x_0, y_0)) = \begin{cases} |x - x_0| + (\sqrt{2} - 1)|y - y_0| & \text{if } |x - x_0| > |y - y_0| \\ (\sqrt{2} - 1)|x - x_0| + |y - y_0| & \text{otherwise} \end{cases} \quad (3-25)$$

[0257] 步骤S220的一个具体实施例中,选择近欧式变换对手势部分作处理,获得手势的中心点。图8为手势区域二值化处理后进行距离变换的图像,图中空心远点表示手势中心点。

[0258] 提取到关键帧中的手势部分后,对手势部分进行距离变换得到手势中心点,将手势图像序列中的中心点组合起来就得到了手势的轨迹特征。步骤S220的一个具体实施例中,通过一些步骤提取手势运动轨迹特征。在步骤S220的一个而具体实施例中,第一手势图像序列包括向下运动的手势运动轨迹特征,第二手势图像序列包含画圈的手势运动轨迹特征。将第一手势图像序列部分帧中偏离了整体的轨迹路线的手势中心点判定为错误点,然后将其删除。将第二手势图像序列部分帧出现了偏离轨迹线的手势中心点判定为误差点,通过线性滤波将其修正。

[0259] 经过上述处理的手势图像序列的手势中心点组成轨迹线,设轨迹线上的第k 个点为 $t_k(x_k, y_k)$,那么以 $t_k(x_k, y_k)$ 为中心的 $2n+1$ 个点位置为 $t_0(x_0, y_0)$,当轨迹点与 $t_0(x_0, y_0)$ 的间距大于设定的阈值T时判定为错误点,若其余点均有误差,可根据前一点位置对其进行修正。因此,该式同时具有一定的容错能力。如下式 (3-26) 所示。

$$[0260] \quad t_k(x_k, y_k) = \begin{cases} \text{delete} & \text{if } \|t_k(x_k, y_k) - t_0(x_0, y_0)\| > T \\ \alpha t_{k-1}(x_{k-1}, y_{k-1}) + \beta t_k(x_k, y_k) & \text{otherwise} \end{cases} \quad (3-26)$$

$$[0261] \quad t_0(x_0, y_0) = \left(\frac{1}{2n+1} \sum_{i=k-n}^{k+n} x_i, \frac{1}{2n+1} \sum_{i=k-n}^{k+n} y_i \right) \quad (3-27)$$

[0262] 取n值为2,通过上式对获取的手势原始轨迹线进行修正。

[0263] 此外,手势运动轨迹相同但起点位置不同对手势识别产生影响,为了排除起点位置不同对轨迹识别的干扰,更容易判断手势运动方向,本发明的一个实施例对手势轨迹做起点归一化处理,将手势起点位置平移到坐标的原点(0,0)位置,因此称其为手势归零化处理,其他点的位置根据与起始点的差值进行重新排列。

[0264] 步骤S230,动态手势识别。如图9所示,动态手势轨迹特征提取过程框图,通过该处理过程对获取的动态手势轨迹特征进行分类,从而实现动态手势的识别。

[0265] 本发明的一个实施例中,使用动态时间规整算法(Dynamic Time Warping, DTW)来对比两个时间序列的相似度。该方法起初运用于语音识别中,解决了语音长短不同的问题,被用于独立词语的识别。随着发展DTW算法广泛运用于各类线性序列的分析处理。在动态手势识别中,常用的方法有隐马尔可夫算法和动态时间规整算法。隐马尔可夫算法需要庞大的训练数据以及复杂的迭代算法,综合考虑本文的应用环境及设备成本,动态时间规整算法在运算时间和计算量上都更适合本发明的动态手势识别。

[0266] 本发明的一个实施例中,采用了预采集的动态手势数据库。该数据库共有 20名测试者,采集8组不同的动态手势,总共得到2000组动态手势样本。动态手势样本序列中,单帧图像大小为 920×1080 像素。动态手势采集背景环境较为复杂,包含其他运动物体。在动态手势识别中,首先进行关键帧的提取。对连续帧中进行光流场估计,当光流场的值小于阈值时,将删除前一帧,否则保留为手势运动的关键帧。使用FAST角点提取手势目标角点特征,其次使用金字塔光流法计算运动角点的光流矢量,减少检测过程中计算量,增加实时性与

准确性。利用Mean Shift算法去除背景干扰角点光流矢量,从而检测出动态手势,最后利用光流法与卡尔曼滤波器相结合,对手势位置进行预测跟踪,解决了动态手势目标跟踪丢失问题,同时提高了光流法的跟踪实时性。通过对手势部分进行距离变换处理得到手势中心点,将所有关键帧中的手势中心点表示手势运动特征,修正误差点删除错误点,并对轨迹做起点归零化处理。最后我们采用DTW算法,针对本实施例采集的动态手势库中的8种动态手势定义,进行动态手势识别分类。动态手势的模板库通过选取8种手势的标准轨迹构建,其余手势样本均作为测试集,用来检验动态识别算法的正确率。当得到一段新的动态手势视频时,对其进行处理后得到其轨迹特征,通过动态时间规整算法对待分类手势轨迹与轨迹模板进行匹配,得到规整距离最短的手势模板,就完成对手势的识别分类。动态手势轨迹识别过程如图10所示。

[0267] 上述动态手势识别的多个方法实施例提供了动态手势分割、手势跟踪以及动态手势识别方法。

[0268] 针对动态手势分割方法,通过提取动态手势序列中的关键帧,减小计算量。对图像中角点进行光流检测得到图像中可能的手势区域,并通过消除背景干扰光流的方法得到手势区域部分。

[0269] 针对手势跟踪方法,将卡尔曼滤波器与光流法结合起来预测手势位置,提高手势位置检测的准确度,并且预防手势丢失等情况的发生。

[0270] 针对动态手势的方法,通过提取手势中心点得到手势运动的轨迹,剔除手势轨迹中错误点,修正手势轨迹中的误差点,对手势轨迹做起点归零化处理。最后通过动态时间规整算法对手势轨迹进行分类,最终识别出动态手势。

[0271] 本发明还提供了一种适用本发明各方法实施例的基于手势识别的车载HUD 人机交互系统,下面通过多个系统实施例以及具体的模块实施例对此予以做出便于实施和改进的描述或者说明。

[0272] 如图11所示,在本发明的一个系统实施例中,包括在同一个局域网中客户端和服务端以及与服务端通讯连接的普通单目摄像机,其中,普通单目摄像机作为手势信息采集模块,通过普通单目摄像机获取驾驶员的手势信息;服务器端作为数据处理识别模块,对普通单目摄像机获取的携带驾驶员的手势信息的手势数据进行处理;车载HUD模块做为客户端,执行服务器端提供的具体手势指令,并将操作信息展示在挡风玻璃上。在一个具体实施例中,服务器端为具有数据处理能力的车载中控台。

[0273] 在一个具体的系统实施例中,为了增加手势系统应用的灵活性,方便不同的车载HUD系统自定义手势功能,即手势定义,在服务器端实现本发明方法的预设程序中分别定义6种静态手势及6种动态手势共对应客户端的16种具体功能,即识别后的一个手势指令,可以结合客户端的当前状态产生两种以上的功能响应,如对于一个手势指令,当客户端接收该手势指令时为未运行应用A的状态,则启动应用A作为响应,当客户端接收该手势指令时应用A的已经运行,则关闭应用 A作为响应。

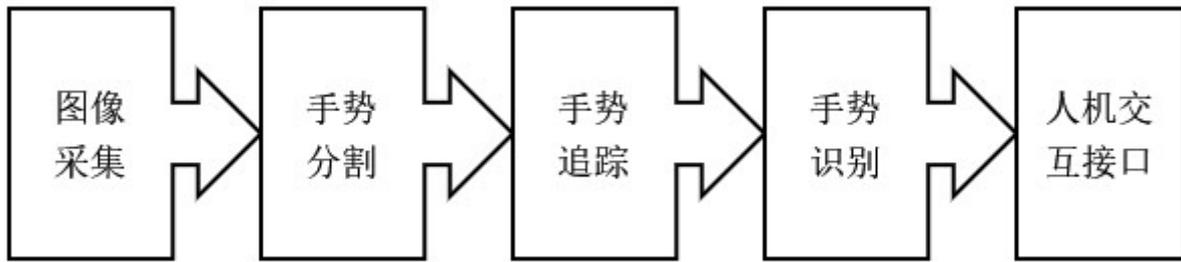


图 1

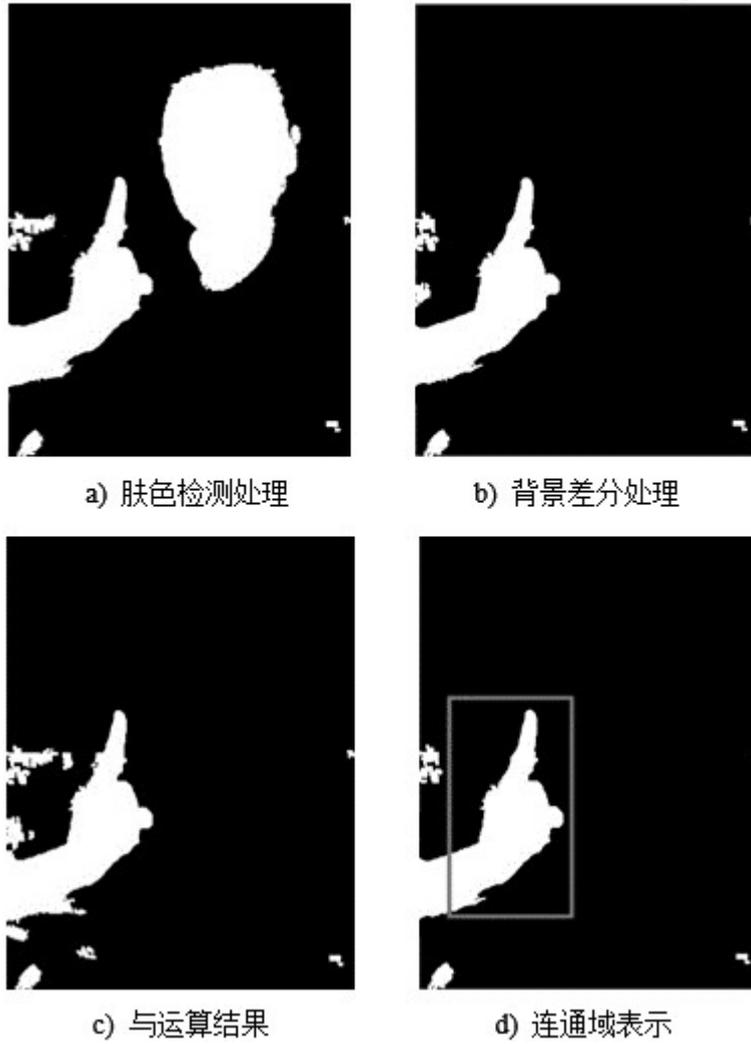


图 2

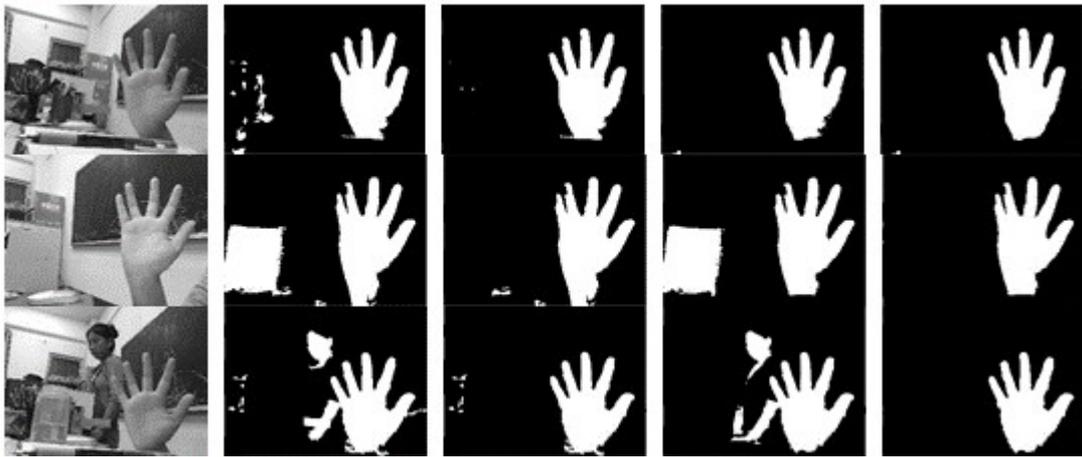


图 3



图 4

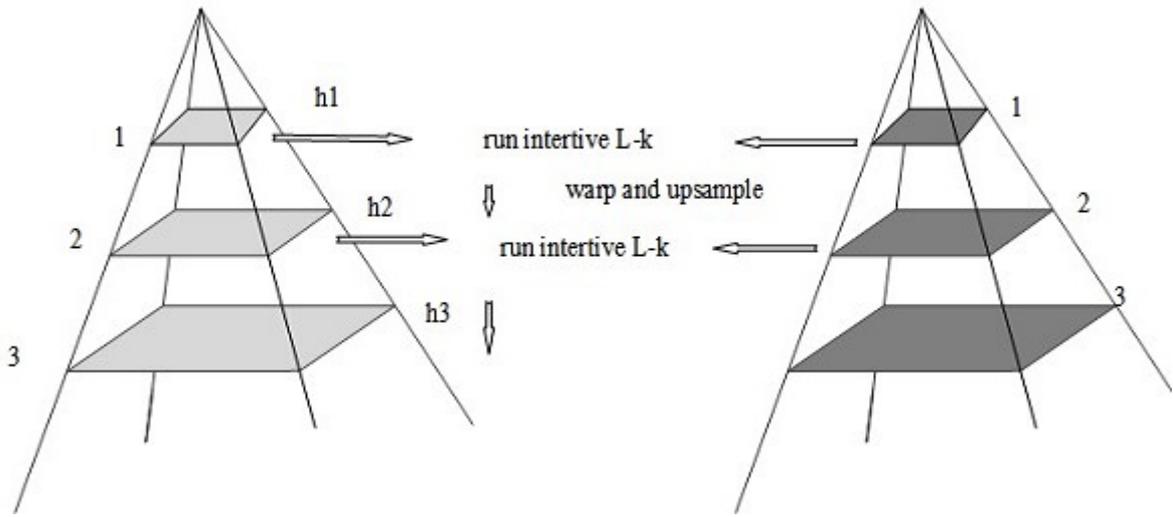


图 5

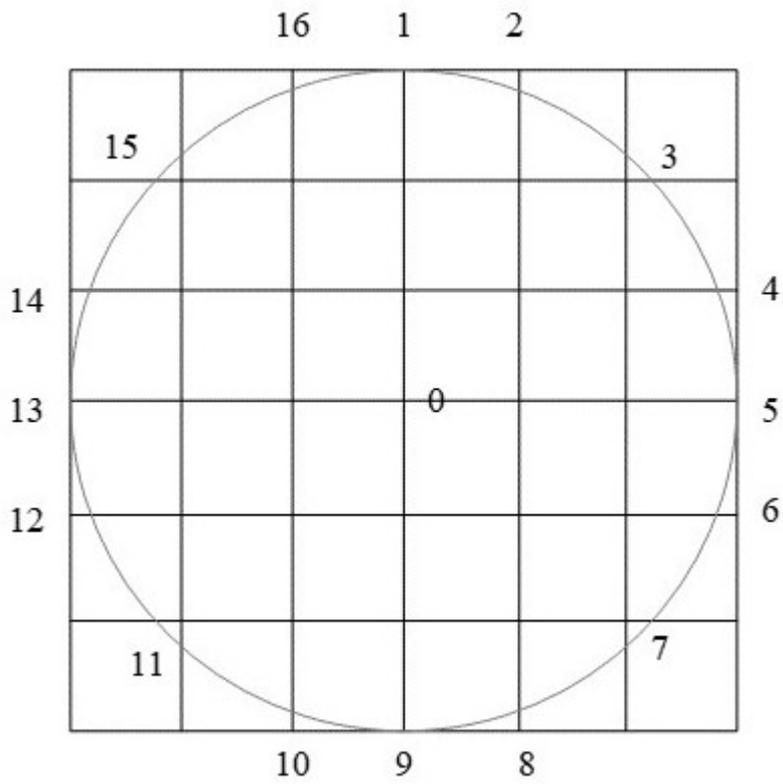


图 6

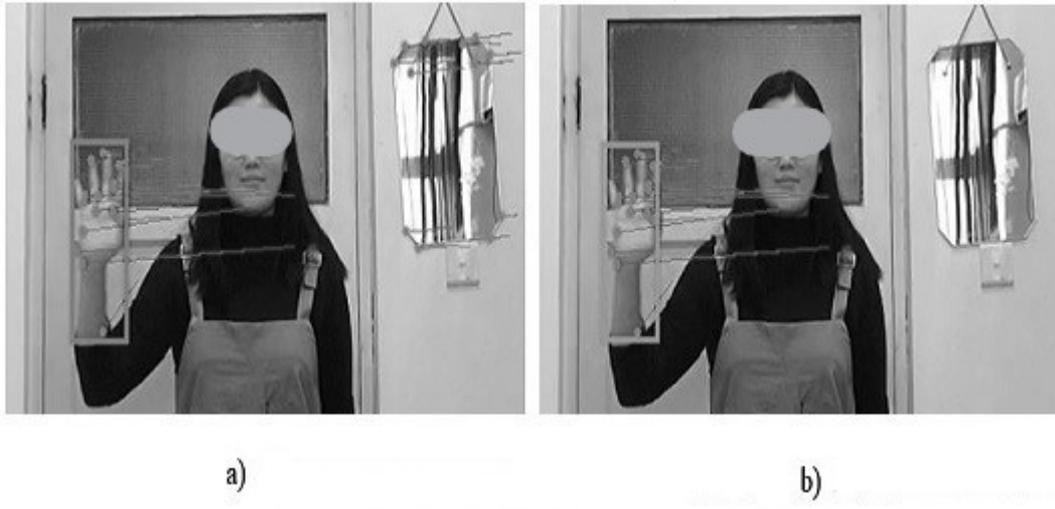


图 7

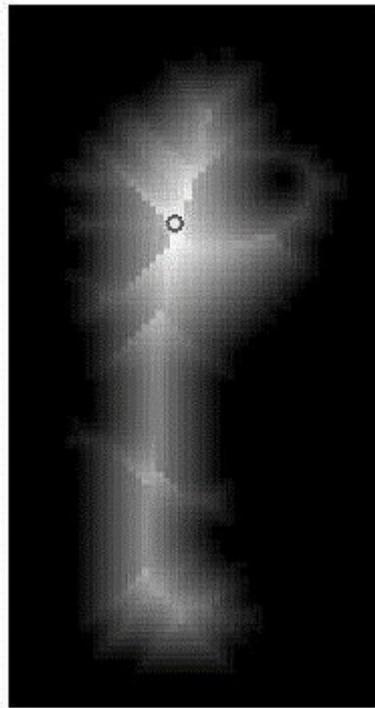


图 8

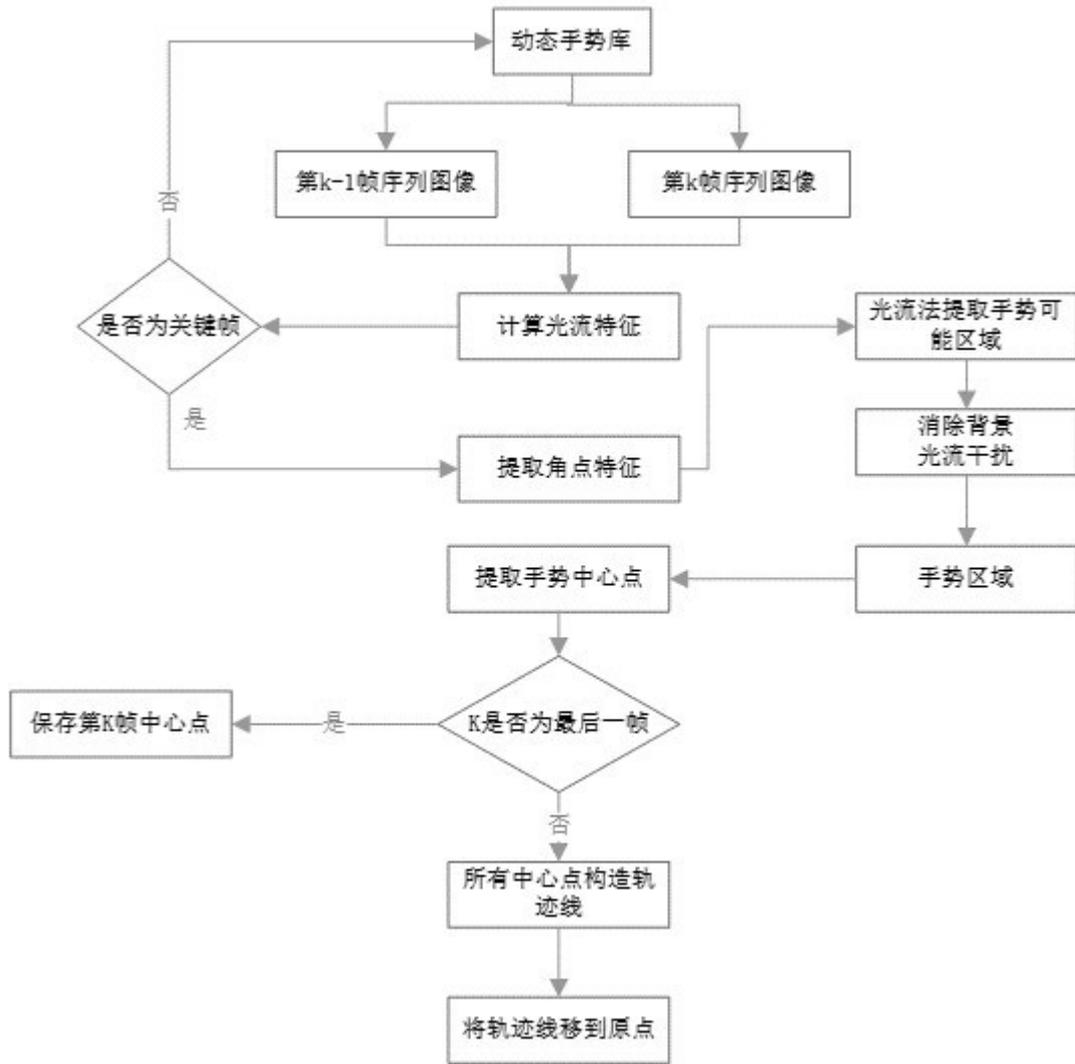


图 9

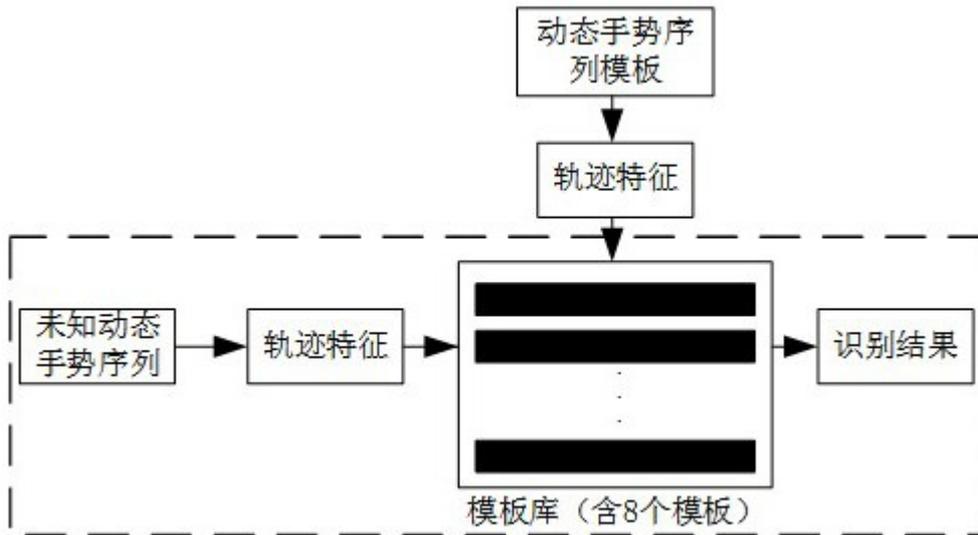


图 10

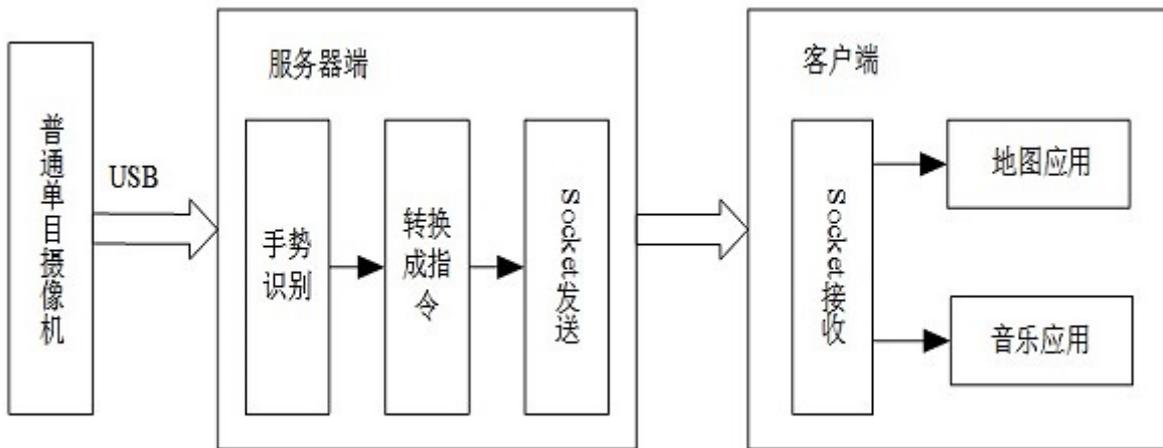


图 11