Doggi 道吉智能云台-相机架一体机设计与开发报告

1. **项目管理**

* 项目经理：于港
* 造型设计师：徐一支、项雨桐
* 机械工程师：徐一支、于港
* 开发工程师：项雨桐

1. **设计背景**

摄影，记录也。

生活中，我们通过摄影记录点点滴滴，把时间的光与影定格为美好的回忆；工作上，我们通过摄影记录分分秒秒，把成功的苦与乐封缄为永恒的记念。

数字技术的发展已然为人们随时随地进行记录创造了极大的便利，但在当下，无论是一位热心的亲人或朋友，还是专业摄影师，许多摄影相关场景中“摄影师”的角色仍然必不可少。然而，我们并不希望在亲朋好友合影留念时因为操作相机而缺少一位成员、留下遗憾，有时也不希望外界的专业摄影师进入私密的生活或工作场景；又如在企业年会、婚庆典礼等情景中，摄影工作往往单调枯燥却成本高昂；而且尤其是在那些中小型会议和活动里，插满数据和电源线、架有沉重摄像机的拍摄点经常成为全场的“路障”——先进的摄影技术并没有使我们的摄影情景得到完全解放，我们仍然受到“摄影师”这一必要角色的限制。

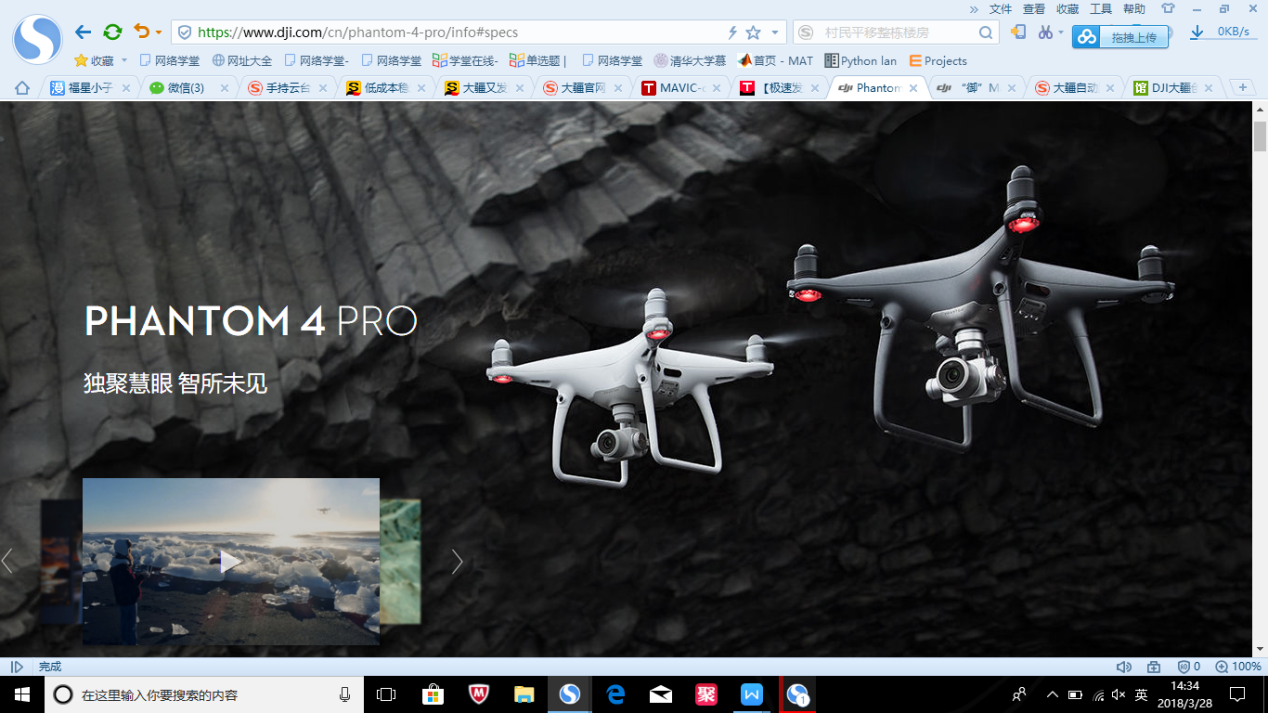
不仅如此，那些以求助于诸如自拍杆和往来游人等外界条件解决“人手”问题的种种手段并非是完美的解决方案，因为这只是对特定场景进行“摆拍”，而这种照片、录像的内容通常僵硬死板，失去了生活影像记录的真实和自然，丧失了影像对日常点滴瞬间的记录意义。

从摄影从业人员的角度来说，如果大多数人的日常工作只是肩扛摄像机在各类会议、活动中穿梭，完成一些简单枯燥却不得不做的信息记录，那么“摄影师”作为一种职业也没有被完全解放：摄影之于人类的最高意义是艺术创作的手段，摄影师应该是是艺术家而非办公室文员，我们不应当让包括摄影从业者在内的任何人将时光虚掷于无意义的简单重复劳作中。

因此，本着解放人类摄影的理念，我们希望能设计出一种能够根据语音和图像识别算法来自动识别拍摄对象、并且能够根据拍摄内容的变化而进行自主移动，自动转变拍摄取景、拍摄角度的智能摄像云台-相机架一体机，以在日常生活与会议活动的记录中给予人们最大的便捷和自由。

1. **竞争产品调研及分析**

**（1）无人机类——大疆 PHANTOM 4 PRO**



价格：

* ¥10998（官网）
* ¥8000~¥13000（电商）

重量（含电池及桨）：1388 g

轴距： 350 mm

最大水平飞行速度：

* 运动模式：72 km/h
* 姿态模式：58 km/h
* 定位模式：50 km/h

最大飞行时间：约30分钟

产品特点：指定位置飞行，指定目标跟拍，智能识别障碍物并避让

优点：

1、可以实现跟拍功能，尤其适合在复杂，危险地形进行跟拍。

2、从高空中进行拍摄，避开了地面的遮挡物，视角更广阔。

3、操作性强，适合大型活动记录或者新闻拍摄。

缺点：

1、续航时间短，不能进行长时间拍摄。

2、价格较高，不适合在家庭中推广，更多是针对无人机爱好者和企业。

3、仪器精密，维护难度和成本较高。

**（2）手持稳定器类——智云 SMOOTH 4**



价格：

¥799（官网）

¥600~¥900（电商）

俯仰轴机械动作范围：

标准： 240 °

横滚轴机械动作范围：

标准： 240 °

负载重量：

最大： 210 g

最小： 75 g

产品裸机尺寸(mm)：123\*105\*328 (W\*D\*H)

产品净重(g)： 547/台

续航时间： 12h

产品特点：慢动作拍摄，防抖，跟随拍摄，云台共享

优点：

1、体积重量较小，方便携带。

2、操作简单，容易上手，入门难度低。

3、续航时间常，适合长时间拍摄。

4、配合手机即可完成拍摄，拍摄成本低。

缺点：

1、需要拍摄者手持拍摄，不能完全解放拍摄者。

2、功能有限，跟拍还是需要拍摄者主动移动。

**（3）智能摄像机类——360智能摄像机**



价格：

¥299（官网）

¥230~350 （电商）

机身尺寸：高119mm，头部宽53mm，底座宽89mm，底座高31.5mm

机身重量：裸机310g

旋转角度：水平360° ，垂直260°

供电方式：USB口输入

产品特点：

* 1080p拍摄
* 手机连网控制
* 夜间拍摄
* 人工智能追踪拍摄对象
* 安全报警

优点：

1、价格较低，安装方便简单。

2、联网操作，手机，pc端可以实时监控。

3、拥有夜间拍摄模式，拍摄时域更广。

缺点：

1、智能化程度较低。

2、不具备移动功能，更多是实现“监控”功能。

1. 供电方式固定，不适合外出携带。

**（4）总结：**

根据能否自主移动跟随拍摄对象，目前市面上的智能跟拍设备可分成两大类：可自主移动跟随（以大疆无人机为代表）和不可自主移动跟随（以智能三脚架、智能摄像机为代表）。前一种普遍存在成本高、续航时间短、技术要求高的缺点，而后一种则普遍存在移动能力差、需要人工操作、智能化程度低的缺点。

1. **用户需求调研与分析**

为了充分了解用户在摄影方面的需求，精准定位产品的核心功能和目标用户，我们设计了一组问卷展开调研（问卷原稿及数据分析见附录）。

问卷有效填写人数为220人，女性约占，以18~25和41~50岁人群为主。

问卷显示，人们对摄影普遍有较强烈需求：有60%以上的被调查者表示摄影在生活中必不可少（题4），有超过75%的人认为自己日常摄影频率较高（题3）。这一点在女性身上表现得更为强烈，且随年龄增大需求呈较弱正相关。

在对摄影的心理需求上，人们普遍追求较为随性、自然的摄影，这一点充分体现在了人们对摄影作品和摄影形式的偏好上（题5-10），并与年龄和性别相关程度均较弱。

在拍摄习惯上，当遇到需要合影的情景时，超过75%的人选择请路人来拍，超过40%的人会选择抽出一个人来拍（题11）。但在请人来拍摄时人们会普遍担心对方的摄影水平并且感到不好意思（题12）；而如果抽出一个人进行拍照，近90%的人会为同伴感到遗憾（题13）；同时，缺乏合适的支撑（包括自带架子）、调节高度、角度、距离的繁琐等等均成为使用外界支撑的掣肘（题14）以上统计结果与性别和年龄关系不大。

与前面的统计呼应，题15的统计结果显示超过60%的人对家庭影像记录有较强烈需求，这一点在女性身上体现得更为明显；但这其中有超过的人并不能实现，其中有超过半数人受限于没有合适的设备（题17）；有能力拍摄的人也多半选择手持DV进行拍摄（题16）。

基于以上调研结果，我们认为，各性别和各年龄段用户对一个能够替代人力进行全自动摄像，并能在拍出高质量影像的同时保证拍摄真实性、自然性的摄影自动化设备会有较强烈兴趣，这些将会是我们开展设计与研发的基本指导方针。

1. **用户体验设计——应用场景构思**

基于我们对产品功能的构想和对用户需求调研的分析，我们设想了如下的产品应用场景：

**（一）家庭/生活**

用户需求调研显示，人们对摄影记录生活的需求普遍比较强烈，而且普遍希望这样的记录能够随性、自然，还原人们真实的生活状态。然而现有的摄影形式如摆拍都有僵硬刻意之嫌，而且由于有时必须有人充当摄影师，影像中总是少了一个人的存在，这是人们最大的遗憾。

道吉的存在可以顺利解决人手问题；此外，在不失真实性、随意性的前提下，利用AI算法道吉仍能保证为人们提供高质量的影像。

例如，道吉可以凭借其定位功能与图像识别的良好配合，无障碍跟拍家中活泼的小孩子，让被工作缠身的新爸爸新妈妈不再有失去见证孩子成长精彩瞬间的遗憾；或者在携手漫步于夕阳海岸的情侣旁默默跟随，依托强大的图像识别算法，以最完美的角度、用最动人的光影诠释最完美的爱情。

**（二）工作会议（中小型）**



在通讯手段极为发达的今天为什么我们仍然需要面对面的工作会议？我们认为，会议中人们不仅仅通过语言传达信息，更重要的是通过肢体语言和表情进行交流。

然而在各类中小型会议中，由于场地所限，并不是所有的人都能充分地从他人的肢体与表情中获取这些信息，而我们通常也没有条件专门配置摄影人员和设备来传递实时画面；此外，虽然有固定的会议记录仪解决此类问题，但缺乏移动性使得我们不能根据实时情况为每一位与会者提供最好的拍摄角度。

道吉体积小而轻便，适合在中小型会议桌上移动，它通过声音识别分辨发言者的方向，然后移动至合适的位置，调整拍摄高度、角度、距离，并定焦在发言者身上，进行详细记录，并能够随发言者的走动、停留而自动调整合适的拍摄画面；当另一位参会人员发言时，道吉会通过识别新的声音来源而定位发言者的方向，重新定焦、调整。

**（三）大型会议与活动：**

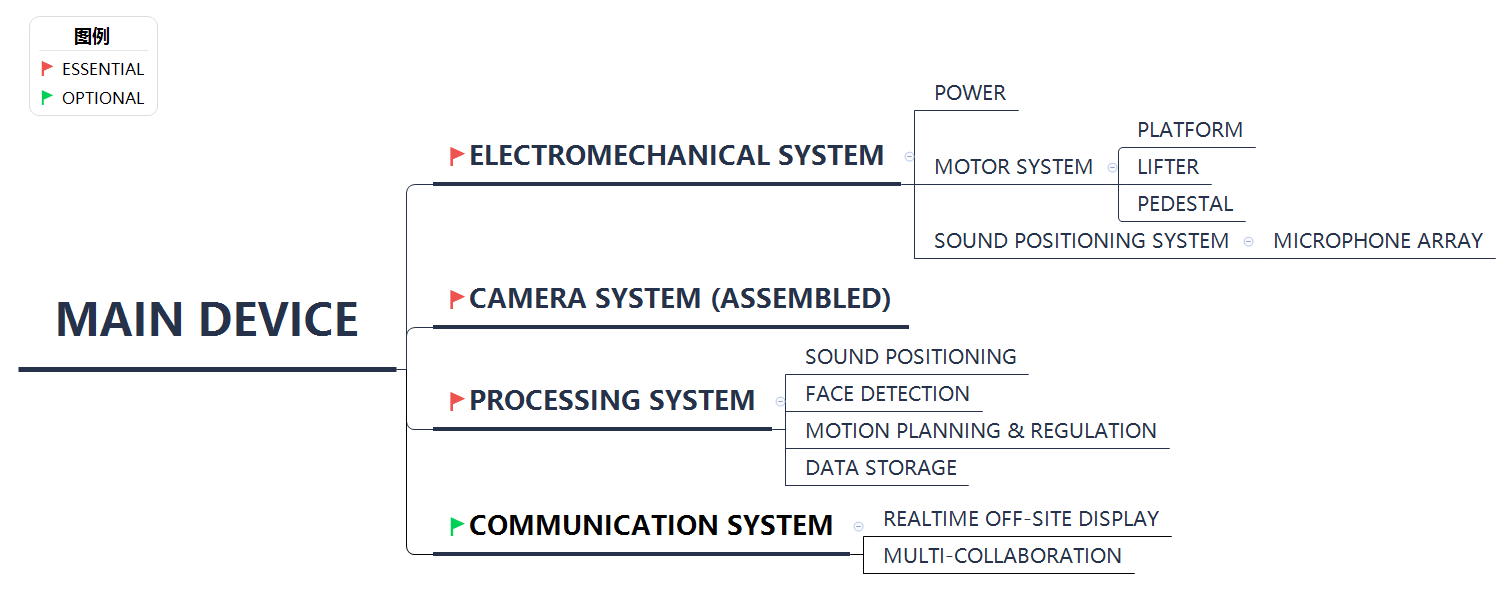


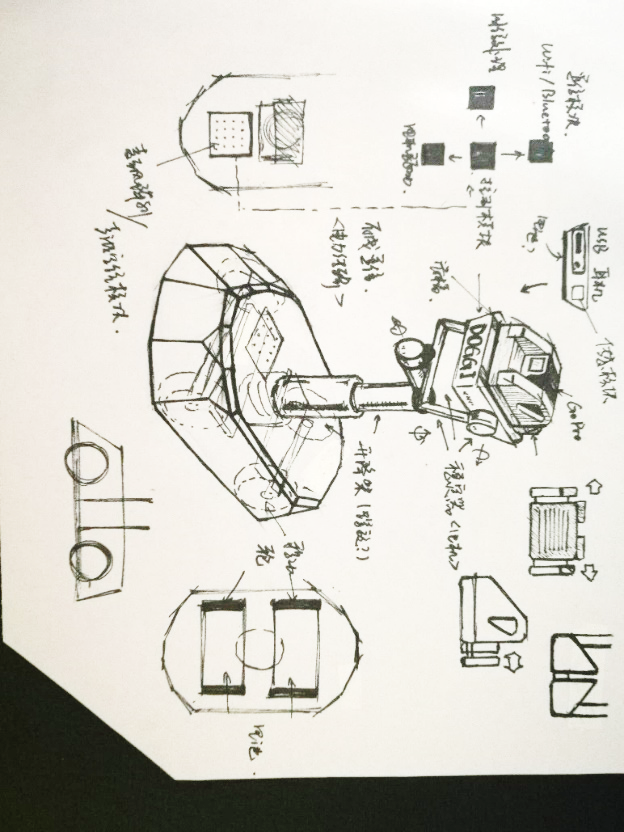
目前的大型会议或活动通常会雇佣专门的摄影公司进行记录，而摄影形式多是一人一机、在整个会场布置多台机器，通常会浪费许多活动空间，造成现场取景点附近活动不便；并且在诸如春晚直播这种需要多机实时协调的场景中，如何实现场景的自动切换十分考验幕后人员的功力。

道吉一体机的结构相较于传统的一人一机组合更为灵活、机动性更强，可极大地节约场地空间；更重要的是，以AI算法为依托实现了摄影全自动的道吉可以极大地节省摄影人力成本；而且，如果我们能很好地实现多设备联网，那么由道吉构成的机组便有希望通过实施通讯+AI算法胜任多机协作的场景。或许，有了道吉的帮助，未来摄影公司的组织形式将会由传统的“设备+人”转变为“算法+设备”。

1. **整体设计构想**

经过一系列的市场调研、用户需求调研和用户体验设计，我们最终将解放摄影师、实现摄影自动化作为我们产品的核心定位。以声源定位、图像识别、导航定位算法和物联网技术为依托，辅以精密的机电结构设计和优美造型，打造一系列可以胜任家庭、会议、户外等多种环境是相机架-云台一体机，让摄影——人类活动记录真正达到真实、随性、自然的理想境界。

整体结构图思维图和设计草图如下：

****

在Demo的设计实现中，我们对功能和结构做了相应的简化，经过反复论证、实验和调试，最终确定Demo的基本结构如下：

(MINDMAP)

1. **机电结构：**
2. 云台：

* 单自由度稳定器（）：乐高积木
* 摄影设备：Huawei P9

1. 升降架：乐高积木
2. 底座：

* 四驱轮式车：乐高积木+4EV3 Large Motor
* 升降架驱动：2EV3 Large Motor
* 总控制台：2EV3 Brick

1. **传感/通讯：**
2. 声源定位：4NXT Sound Sensor
3. 通信传输:

* Wi-Fi 视频获取与传输：Android ip webcam v1.13.7
* EV3遥控：RPyC

1. **算法支持：**
2. 声源定位算法：自定义
3. 实时视频帧接收与提取：OpenCv 3.4.1
4. 人脸检测算法：OpenCv 3.4.1
5. 人脸跟随算法：自定义
6. **BOM表**
7. **乐高部分：**

* 2EV3 Brick
* 6EV3 Large Motor
* 4NXT Sound Sensor
* 乐高积木与接线若干

1. **外部支持及其他：**

* 声音收集：4塑料漏斗，海绵若干
* 摄影器材： LEICA SUMMARIT H 1:2.2/27 ASPH for Huawei P9

1. **整体结构LDD**

**//**

1. **机械设计与优化**

**//**

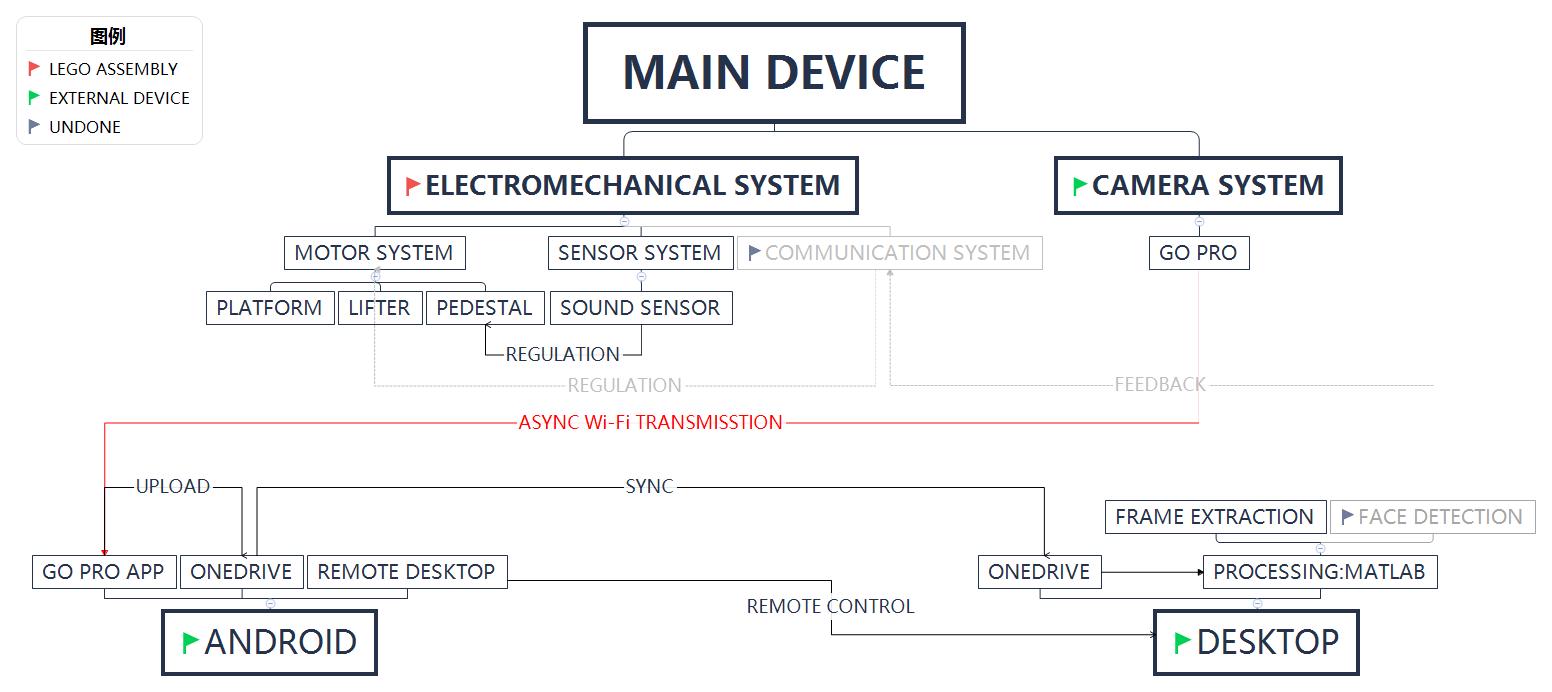
1. **程序设计与优化**
2. **第一代设计**

在demo制作环节，鉴于EV3处理器运算性能有限、Android快速开发难度较大,我们设想利用PC处理器，将视频数据和控制信号以无线方式在摄像机、PC和EV3主机间传输以实现相应功能。

因此，程序可以分为如下3个模块：

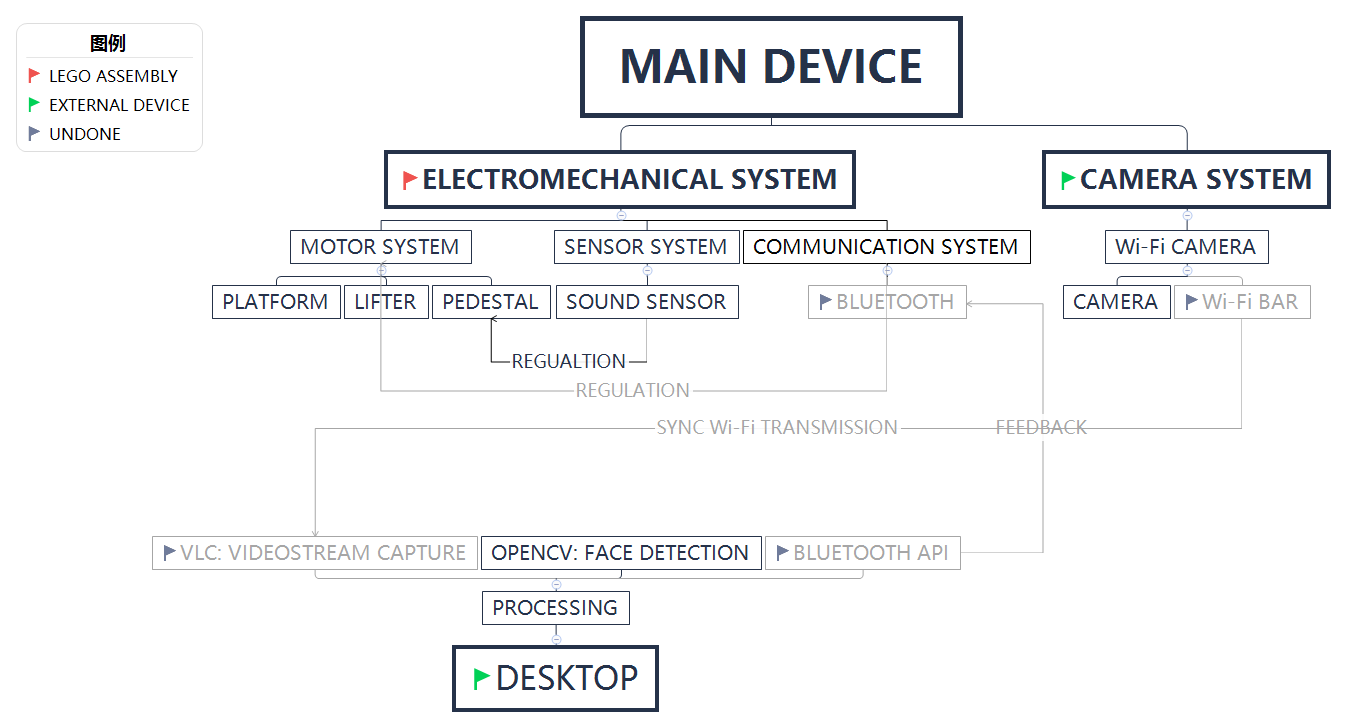
* 实时视频帧接收与提取
* 人脸检测
* 人脸坐标发送

由于GoPro为非开源软件，无法直接从其APP中获取实时视频帧，所以在第一代设计中我们只实现了视频的非实时传输（即生成完整的视频文件后再进行无线传输）和利用MATLAB的视频帧提取；为了减少控制终端，我们采用了远程桌面技术，利用手机进行APP的视频上传、控制PC进行视频帧等操作。

****第一代设计的思维图如下：

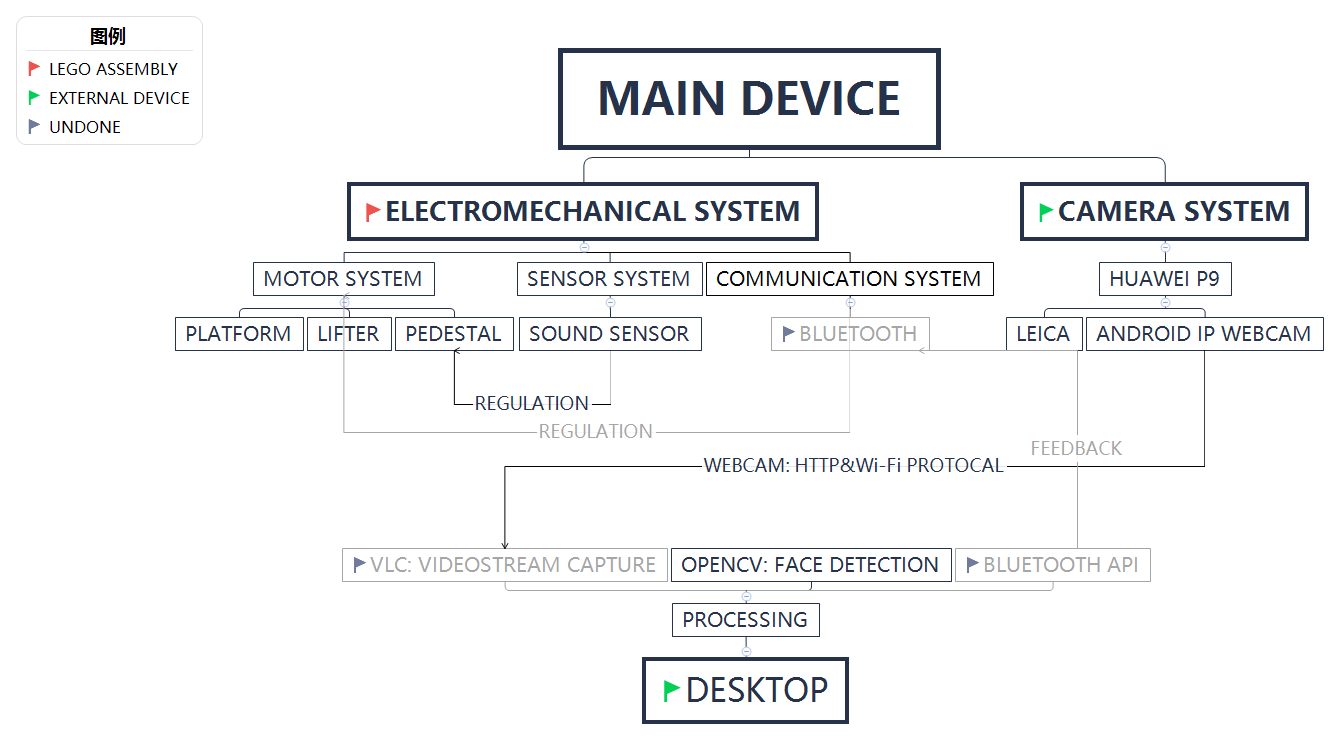
1. **第二代设计**

为了得到实时视频帧，在老师的提醒与帮助下，我们放弃了GoPro，采用市面所售Wi-Fi摄像头作为我们的视频采集设备，并采用开源播放器VideoLAN Client（VLC）的SDK完成实时视频帧接收和获取模块的开发；利用OpenCV实现人脸检测；初步拟定利用蓝牙实现人脸坐标的发送与接收。

第二代设计的思维图如下：

1. **第三代设计**

由于始终无法正确调试Wi-Fi摄像头，我们将其替换为Android手机，利用IP webcam APP实现实时视频帧的发送，并成功地应用VLC软件实现了实时视频帧的接收。

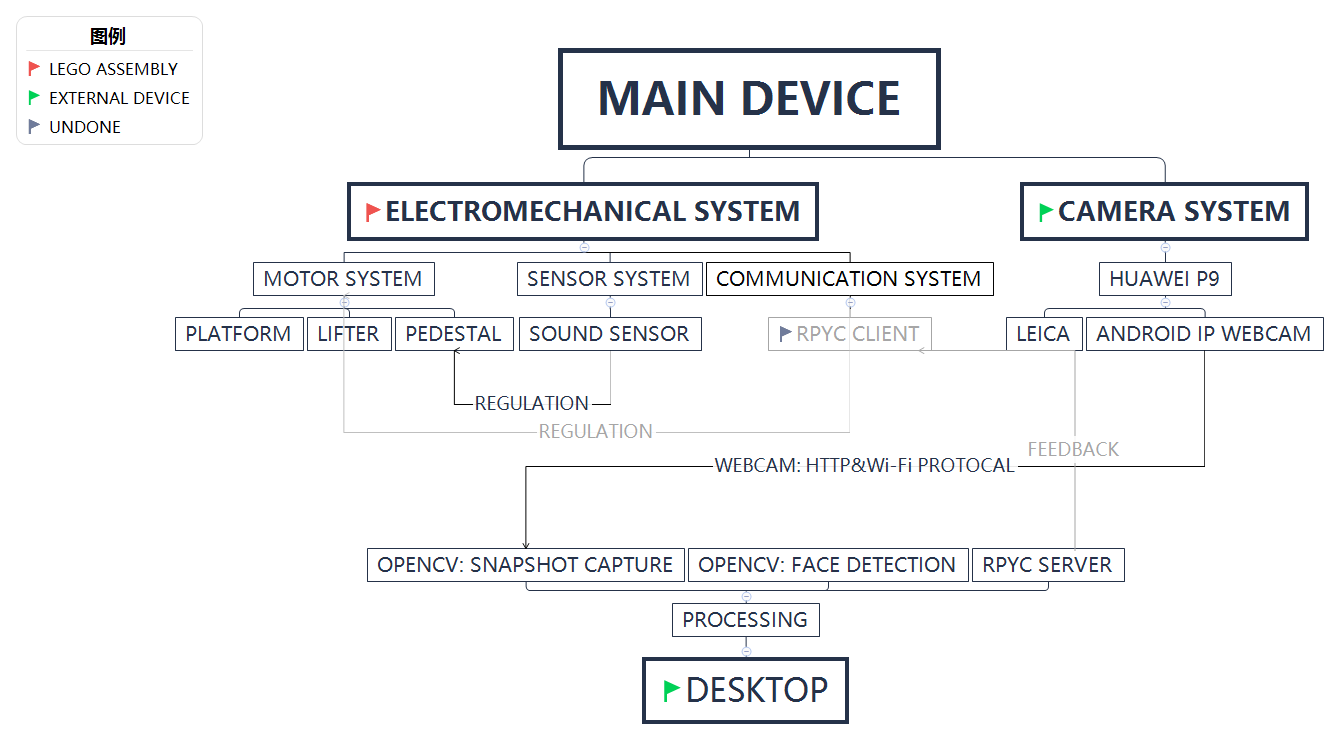
****第三代设计的思维图如下：

1. **第四代设计**

由于未知原因，VLC SDK程序模块始终正确无法接收实时视频帧。我们发现IP webcam同时提供当前视频截图的URL，而OpenCV 3提供截取网络图片的函数，故将VLC的SDK替换为OpenCV来捕捉视频截图，结合其原有的人脸检测算法，成功实现了实时视频中的人脸捕捉。

由于蓝牙编程难度较高且缺乏合适的开源项目，我们最终决定采用Remote Python Call（RPyC）来根据图像中人脸坐标位置对乐高机器人在PC端进行直接遥控。

由于RPyC在EV3 linux系统由于不明原因无法正确安装，故最终没有实现，但大量资料表明该方案在所有对EV3进行无线控制的方法中最为简单易行、成功率较高，故我们仍采用RPyC作为demo制作中无线控制环节的最终解决方案。

第四代设计的思维图如下：

1. **关键技术**
2. **IP Webcam：**

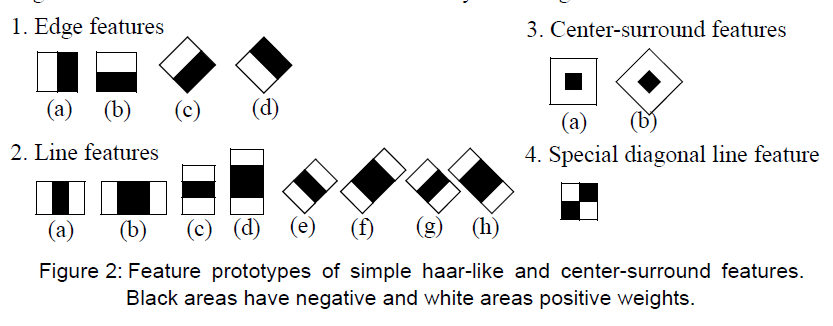
这款软件可以把手机变成一个功能极为强大的无线摄像头。

* 支持多种分辨率，本次采用640480
* 支持多种视频编码格式，本次采用MPEG4。
* 支持多协议传输，本次采用HTTP。
* 其他多样的功能。

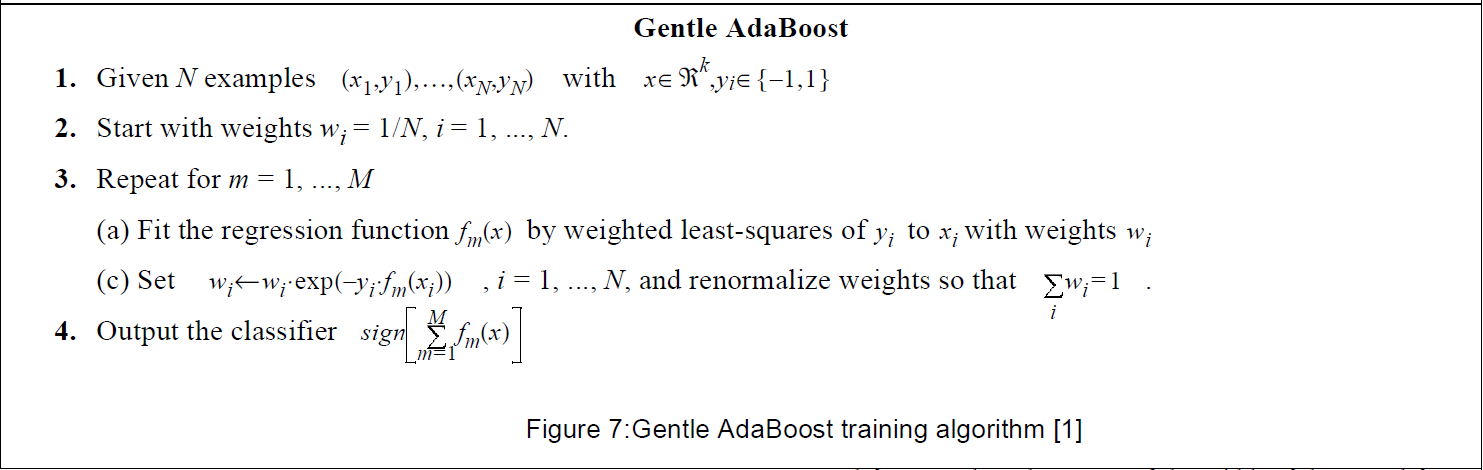
1. **OpenCV：人脸检测-Cascade Classifier 级联分类器**

Cascade Classifier最早由Paul Viola发表在2001年的CVPR上，并在2002年由Rainer等人进行改进。OpenCV利用由Rainer改进的Gentle AdaBoost算法，将Haar弱分类器集合成强分类器并进行级联，从而极大地提高了识别准确率和速度。

1. **经Rainer改进的类Haar特征（Haar-like feature）的弱分类器**（外围黑色框表示平滑窗，窗内固定的位置有个黑白相间的矩形区域，黑色表示该位置上的像素每个权重都是-1，白色位置上的像素权重都是+1，权重与像素卷积得到的值就是特征值）：

****

1. **利用Rainer等人改进的Gentle Adaboost算法将弱分类器集合成强分类器：**

Adaboost是一种迭代算法，其核心思想是针对同一个训练集训练不同的弱分类器，然后把这些弱分类器集合起来，构成一个强分类器。由Rainer等人改进的Gentle Adaboost算法被证明为是同类算法中速度和准确率最高的算法。

1. **级联Adaboost：**

将若干个小的Adaboost分类器级联起来，并且每一级的分类器，都具有非常高的检测率（接近100%）和误检率（约50%）。这样在经过多级级联之后，可以得到准确率非常高的人脸检测器。

同时，一旦在某一级分类器中，检测出不是人脸，人脸检测立刻终止，这样可以极大地提高检测速率。

1. **RPyC：远程控制**

RPyC 是一个可以实现对称远程程序调用（symmetrical remote procedure calls）、聚类和分布式计算的透明Python库。它使用对象代理，利用了Python的动态特性来克服处理器和计算机之间的物理界限，使远程对象被像本地对象一样被操控。

1. **项目总结**
2. **调研分析：**
3. **竞争产品调研：**数据来源有限，仅限于电商平台提供的相关信息；该类产品的行研报告平均售价7000元以上，囿于财力所限无法有效获得。
4. **用户需求调研：**形式较为单一，缺乏实地调研环节；问卷数据来源在性别、年龄以及地域分布中有明显偏斜，加之没有用科学的统计方法进行分析，得出的结论较为模糊，统计结果的准确性偏低。
5. **应用场景：**

场景设想较模糊，需要进一步细化

对需求的深层次挖掘

1. **机械设计：**

**//**

1. **程序设计：**

由于受硬件设备等多方面限制，我们在无线传输部分、也即实时视频帧接收和远程控制的调试上耗时过长，并且最终也没能实现远程控制，可以说留下了不小的遗憾。

1. 在一番尝试之后我们仍然选定智能手机作为摄影设备，在这种情况下，如果一开始我们能够尝试Android开发，充分利用其摄影设备和处理器集成的固有特性而避开无线传输的问题，或许demo开发会进行得更加顺利、效率会更高。
2. 经测试，OpenCV自带的Haar分类器的准确度仍然有不足，应当进一步训练以得到更高的检测正确率。
3. 当前程序设计语言为C/C++，若RPyC在EV3 Linux上可以顺利安装，则可以用一个Python工程实现全部功能，无需进行语言的衔接。
4. **未来设想**
5. **机械结构**

//

【提升云台自由度】

【升降架技术】

1. **硬件设备**
2. **声源检测：**

在后续设计中我们拟采用麦克风阵列来代替现有的乐高声音传感器以提高准确率；同时利用麦克风阵列的音频采集功能实现音频的存储和异地播放。

1. **影像采集及处理：**

在后续设计中应为各种摄影设备提供USB等接口，并保证能获取不同摄影设备的实时视频帧访问权限；不借助外部设备支持，在一体机的嵌入式系统中完成人脸检测和运动控制等操作。

1. **软件/算法**
2. **人脸检测：**

进一步训练OpenCV的级联分类器，或者利用开源平台的人脸检测算法，进一步提升人脸检测的准确度。

1. **运动控制：**

在运动控制与调节模块，可以考虑机器人自主导航算法，使之能适应更复杂的应用场景。

1. **多机协作**

可以考虑进一步开发无线网络通信模块以实现必要的数据传输和多机协作功能

1. **参考文献**

[1] 目标检测（Object Detection）原理与实现(五):

<https://blog.csdn.net/marvin521/article/details/9173645>

[2] Lienhart, R. Kuranov, A. Pisarevsky, V. (2002). Empirical analysis of detection cascades of boosted classifiers for rapid object Detection [PDF file]. *MRL Technical Report.* Available from <http://www.lienhart.de/Prof._Dr._Rainer_Lienhart/Publications_files/MRL-TR-May02-revised-Dec02_1.pdf>

[3] Adaboost\_百度百科：<https://baike.baidu.com/item/adaboost>

[4] Viola Jones Face Detector: <http://www.cnblogs.com/hrlnw/p/3374707.html>

[5] RPyC - Transparent, Symmetric Distributed Computing: [http://rpyc.readthedocs.io/en/latest/#](http://rpyc.readthedocs.io/en/latest/)

1. **附录**

[1]问卷原稿：<https://www.wjx.cn/jq/22064800.aspx>

[2]问卷统计结果：<https://www.wjx.cn/report/22064800.aspx>

[3]程序源代码：<https://github.com/Iceblaze9527/Doggi_Fundamentals-of-Shaping>