

基于无人机的人脸识别追踪

王泽贤

中山大学数据科学与计算机学院

Email: wangzx6@sysu.edu.cn

龚志豪

中山大学数据科学与计算机学院

Email: zahirgong@gmail.com

Abstract—随着无人飞行器（UAV）的功能和潜力不断被发掘，多旋翼无人机已经进入高速发展的时代。本研究提出了多旋翼无人机应用的新思路，首次尝试将多旋翼无人机与人脸识别技术结合。通过多旋翼无人飞行平台上的单目摄像头采集到的图像数据通过图传回传到地面端，由地面端接收后对图像进行人脸识别处理，得到无人机目前所在的相对位置。由脸部大小判断出无人机与目标的相对距离，由脸部在图像中的位置来判断目标与无人机的相对位置，进而对无人机发出控制指令来实现对特定人的追踪，实现人群中识别并追踪特定人，不会因为人群的分散或者移动导致追踪失败，能达到对特定人的追踪和更高精度的跟踪的效果。本文着重于研究四轴飞行器中的姿态解算算法，互补融合算法和飞行控制算法，人脸识别技术中的 LBPH 算法。最终实验表明，四轴飞行器已经实现自稳悬停以及稳定飞行，并且能根据人脸识别处理后的数据跟踪目标。

关键词：多旋翼无人机，人脸识别，追踪

1. 引言

近年来，随着微机电系统（MEMS）研究的成熟，以及捷联惯性导航相关理论的完善，多旋翼飞行器的发展成了不可遏制的趋势，研究者纷纷开始研究多旋翼飞行器。多旋翼飞行器是一种小巧、稳定、可垂直起降、机械结构简单的飞行器。它利用空气动力来抵消自身的重量，可以实现自主飞行或者遥控飞行 [1]。它稳定性好，飞行平稳，相对于固定翼飞行器，多旋翼飞行器能做到悬停，相对于直升机，多旋翼飞行器结构更易于操控，因此非常实用于基于其做大量开发。早期，固定翼无人飞行器的发展已经相当成熟，而多旋翼无人机的发展相对来说较为缓慢，这是因为多旋翼无人机的控制较固定翼的复杂，早期的技术水平无法实现多旋翼飞行器的自主飞行控制 [2]。然而，多旋翼飞行器却拥有其自身独特的优点：（1）机械结构较为简单，只需要控制旋翼电机的转速就可以实现控制，飞行比较灵活；（2）不受

宽大的机翼的控制，能够应用于各种环境中；（3）具有自主起飞和降落的功能，系统高度智能化，可以实现较多的飞行姿态，例如：垂直悬停、俯仰升降、偏航转向等等，而且飞行器姿态的保持能力比较高 [3, 4]。目前，多旋翼飞行器技术已经比较成熟，相关姿态解算算法如四元数、余弦矩阵，相关控制算法如 PID 控制算法已经比较成熟，所以非常适合本研究基于其做特定人跟踪。

另一方面，人脸识别技术也趋于成熟，有多种方法可以应用到实践中。研究人脸识别在理论和技术上不仅可以推进对人类视觉系统本身的认识，还可以满足人工智能应用的需要。采用人脸识别技术，建立自动人脸识别系统，用计算机实现对人脸图像的自动识别有着广阔的应用领域和诱人的应用前景。人脸识别技术融合了数字图像处理、计算机图形学、模式识别、计算机视觉、人工神经网络、生物特征技术以及生理学、心理学等多个学科的理论和方法。人脸的识别研究，发展到能够动态实现人脸识别，目前正在向三维人脸识别的方向发展。在此过程中，人脸识别技术涉及的图像逐渐复杂，识别效果不断地得到提高。虽然人脸识别研究已积累了宝贵的经验但目前的识别技术仍然不能对诸如复杂背景中的人脸等进行有效的处理和自动跟踪。近年来，由于反恐、国土安全和社会安全的需要，世界上各个国家都对安防领域加大了投入。身份验证是安防的一个核心领域，因此人脸识别的技术与飞行器技术结合的方法显得尤为重要，我们也尝试将人脸识别技术应用到飞行器领域。

目前，将人脸识别技术与多旋翼飞行器结合起来的研究不多，主要的瓶颈在于机载处理器的处理能力无法达到人脸识别的处理能力标准。在多旋翼无人机行业处于领先地位的大疆创新（DJI）的最新产品 Phantom4 的自动跟踪功能也只是对于物体进行跟踪，存在的问题是无法对于人群中的特定人进行跟踪。本文中提出了将人工智能与多旋翼无人机技术结合在一起的应用。虽然

由于机载处理器的处理能力还是无法直接在飞行器上进行人脸识别处理（就算是树莓派的处理能力也是远远不够），但是提出了一种新的解决方案：通过图传将机载摄像头的模拟视频流传回地面端，由地面端的图传接收器接收后再由采集卡采集转换成数字信号后由计算机进行处理，然后由计算机通过数传模块发出控制指令来达到控制飞行器跟踪特定人的效果。在满足计算速度要求的数字信号处理器（DSP）支持的情况下，理论上是可以达到完全由机载的处理器进行控制和控制的。

2. 多旋翼无人机平台

2.1. 姿态解算

多轴飞行器的姿态角是控制的根据，首先必须有一个正确的系统的当前姿态才能准确地控制系统到目的姿态。得到水平的两个欧拉角最简单的方法就是用加速度计得到重力加速度在机体坐标系的投影，然后通过反三角函数即可得到欧拉角。但是这样的弊端就是无法得到第三个维度的姿态角，还有加速度计测量的数据会受到运动加速度、噪声的影响，而且电机振动的干扰对加速度计的影响是极大的，因此必须引入其他惯性单元：陀螺仪。陀螺仪的特点就是具有定轴性和进动性 [5]，它的定轴性决定了陀螺仪的数据在短时间内是准确的，而进动性决定了陀螺仪在长时间会存在漂移。而加速度计的特点是短时间内得到的数据是有噪声的，长时间考虑的话数据是比较准确的，这一特点和陀螺仪的特点正好互补，后续会介绍两者的数据融合算法。由陀螺仪与加速度组成惯性测量单元（IMU）是捷联式惯性导航的重要组成部分。姿态解算的核心在于旋转，而目前主流的用于姿态解算的组合旋转表示方法主要有余弦矩阵和四元数。

2.1.1. 余弦矩阵. 方向余弦矩阵也叫旋转矩阵，表示的是一个旋转组合，每一个旋转的组合都对应一个唯一的余弦矩阵。定义刚体绕 X,Y,Z 轴转的角度为 α, β, θ 。刚体的旋转可以等价于刚体的法向量的旋转，假设刚体法向量为：

$$\mathbf{v}_g = \begin{bmatrix} x & y & z \end{bmatrix}^T$$

刚体经过一个旋转组合后的法向量为：

$$\mathbf{v}'_g = \begin{bmatrix} x' & y' & z' \end{bmatrix}^T$$

假设 \mathbf{v}_g 绕 X 轴转 α 角度，定义绕 X 轴旋转的矩阵为 \mathbf{X} ，则：

$$\mathbf{v}'_g = \mathbf{X}\mathbf{v}_g \quad (1)$$

其中：

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha \\ 0 & -\sin \alpha & \cos \alpha \end{bmatrix} \quad (2)$$

假设 \mathbf{v}_g 绕 Y 轴转 β 角度，定义绕 Y 轴旋转的矩阵为 \mathbf{Y} ，则：

$$\mathbf{v}'_g = \mathbf{Y}\mathbf{v}_g \quad (3)$$

其中：

$$\mathbf{Y} = \begin{bmatrix} \cos \beta & 0 & -\sin \beta \\ 0 & 1 & 0 \\ \sin \beta & 0 & \cos \beta \end{bmatrix} \quad (4)$$

假设 \mathbf{v}_g 绕 Z 轴转 θ 角度，定义绕 Z 轴旋转的矩阵为 \mathbf{Z} ，则：

$$\mathbf{v}'_g = \mathbf{Z}\mathbf{v}_g \quad (5)$$

其中：

$$\mathbf{Z} = \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (6)$$

定义法向量的旋转顺序为 Z-Y-X，则刚体经过该旋转组合后的向量为：

$$\mathbf{v}'_g = \mathbf{XYZ}\mathbf{v}_g \quad (7)$$

则方向余弦矩阵定义为：

$$\mathbf{C}_{\mathbf{v}_g}^{\mathbf{v}'_g} = \mathbf{XYZ} \quad (8)$$

该矩阵的作用就是把向量 \mathbf{v}_g 按照 Z-Y-X 的旋转顺序分别旋转 α, β, θ 角度得到向量 \mathbf{v}'_g 。可以使用该矩阵来更新重力向量在机体坐标系的投影，从而得到姿态角。

2.1.2. 四元数. 四元数跟余弦矩阵一样都是表示一个旋转组合，也能唯一的表示一个旋转组合。对四元数的理解可以想象成把复杂的三维空间的运动放到四维空间里就成了简单运动。这里定义四元数为：

$$\mathbf{Q} = \begin{bmatrix} q_0 & q_1 & q_2 & q_3 \end{bmatrix}^T$$

四元数的表达方式有很多种，有矢量式、复数式、三角式、指数式和矩阵式，四元数的大小用范数来表示，四

元数也有加减乘除运算 [5]。四元数转化为方向余弦矩阵的公式如下：

$$\mathbf{C} = \begin{bmatrix} 1 - 2(q_2^2 + q_3^2) & 2(q_1q_2 + q_0q_3) & 2(q_1q_3 - q_0q_2) \\ 2(q_1q_2 - q_0q_3) & 1 - 2(q_1^2 + q_3^2) & 2(q_2q_3 + q_0q_1) \\ 2(q_1q_3 + q_0q_2) & 2(q_2q_3 - q_0q_1) & 1 - 2(q_1^2 + q_2^2) \end{bmatrix} \quad (9)$$

初始化四元数为：

$$\mathbf{Q} = [1 \ 0 \ 0 \ 0]^T$$

根据余弦矩阵与欧拉角的定义，则余弦矩阵 \mathbf{C} 的第三列就是重力向量在机体坐标系的投影：

$$\mathbf{v}_g = \begin{bmatrix} 2(q_1q_3 - q_0q_2) \\ 2(q_2q_3 + q_0q_1) \\ 1 - 2(q_1^2 + q_2^2) \end{bmatrix} \quad (10)$$

所以使用校准零偏后的陀螺仪的数据更新四元数就是等同于旋转重力向量到机体坐标系，使用一阶龙格库塔法¹更新四元数的公式如下：

$$\mathbf{Q} = \begin{bmatrix} q_0 + 0.5(-q_1G_x - q_2G_y - q_3G_z)\Delta t \\ q_1 + 0.5(q_0G_x - q_3G_y + q_2G_z)\Delta t \\ q_2 + 0.5(q_3G_x + q_0G_y - q_1G_z)\Delta t \\ q_3 + 0.5(-q_2G_x + q_1G_y + q_0G_z)\Delta t \end{bmatrix} \quad (11)$$

其中 G_x, G_y, G_z 为校准零偏后的陀螺仪的数据， Δt 为更新周期，更新四元数后再归一化即可得到姿态角，归一化后四元数到欧拉角的转换公式如下：

$$\begin{bmatrix} \alpha \\ \beta \\ \theta \end{bmatrix} = \begin{bmatrix} \arctan(2q_2q_3 + 2q_0q_1, 1 - 2(q_1^2 + q_2^2)) \\ -\arcsin(2q_1q_3 - 2q_0q_2) \\ \arctan(2q_1q_2 + 2q_0q_3, 2(q_0^2 + q_1^2) - 1) \end{bmatrix} \quad (12)$$

上述公式只对应 Z-Y-X 的欧拉角顺序，对于 12 种不同的欧拉角顺序，有不同四元数与方向余弦矩阵之间的转化公式和不同的四元数与欧拉角之间的转化公式 [7, 8]。

2.2. 数据融合

前面已经提到过，加速度计由于噪声比较大，而且在飞行过程中，受机体振动影响比陀螺仪明显，短时间内的可靠性不高。陀螺仪噪声小，但是由于积分是离散的，而且陀螺仪得到的角度是在没有平衡点为参照的基础上进行的，长时间的积分必然会出现漂移的情况。因此需要将用加速度计求得的姿态来矫正陀螺仪积分姿态的漂移。在机体坐标系上，加速度计测出来的是重力加速度在机体坐标系上的投影为：

$$\mathbf{acc} = [a_x \ a_y \ a_z]^T$$

1. 龙格库塔法是一种高精度单步算法，用于模拟常微分方程的解的重要的一类迭代法

而陀螺仪积分后的角度来推算出的重力向量是：

$$\mathbf{v}_g = [g_x \ g_y \ g_z]^T$$

这两个向量在一定的前提下都可以表示姿态角的向量，但是这两个向量存在误差，都不能准确地表示当前的姿态角。所以需要一种算法对这两个向量做融合以得到最逼近真实姿态向量的向量。

2.2.1. 权值融合法。 假设融合中向量 \mathbf{acc} 的权值为 acc_{cf} ，则融合后的姿态角向量 \mathbf{cf} 为：

$$\mathbf{cf} = \mathbf{acc} \cdot acc_{cf} + \mathbf{v}_g(1 - acc_{cf}) \quad (13)$$

根据加速度在短时间内的测量值存在噪音的特性， acc_{cf} 的取值一般小于 $1 - acc_{cf}$ 。得到的融合后的向量用来更新 \mathbf{v}_g ，这样做的效果在短时间的时段内，融合后的姿态角向量是以陀螺仪得到的姿态角向量为准，但是在长时间来看，加速度计得到的姿态角向量在纠正陀螺仪的偏移。

2.2.2. 最大梯度法。 假设 \mathbf{acc} 与 \mathbf{v}_g 两个向量的误差向量为：

$$\mathbf{e} = [e_x \ e_y \ e_z]^T$$

向量间的误差也就是两个向量的叉积，即：

$$\mathbf{e} = \mathbf{acc} \times \mathbf{v}_g \quad (14)$$

可以看到这个叉积的大小与陀螺仪的积分误差成正比，可以拿来修正陀螺仪的积分误差。使用 PI 控制器对 \vec{v}_g 进行修正，即：

$$\mathbf{v}_g = \mathbf{v}_g + k_p \mathbf{e} + k_i \sum \mathbf{e} \quad (15)$$

其中 k_p, k_i 为 PI 参数，通过调节这两个参数可以改变加速度计修正陀螺仪漂移的速度。该方法与权值融合法的本质是一样的，只是表达的效果不同。

2.3. PID 控制器

PID 控制器是最常见，应用最为广泛的自动反馈系统。PID 控制器由比例（P）控制器、积分（I）控制器和微分（D）控制器组成，通过 P,I,D 三个参数来对系统进行控制。PID 控制器又分为开环 PID 控制器和闭环 PID 控制器。

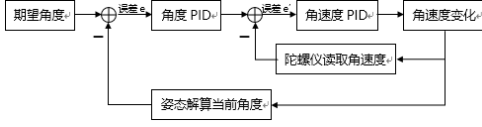


图 1: 闭环 PID 算法流程

2.3.1. 开环 PID. 假设系统在 t 时刻的误差为 $e(t)$, k_p, k_i, k_d 分别为 P,I,D 三个参数, $P_{out}, I_{out}, D_{out}$ 分别为 P,I,D 控制器的输出, $u(t)$ 为系统输出。

比例 (P) 控制的输出与输入成比例关系, 所以比例控制器的输出为:

$$P_{out} = k_p \cdot e(t) \quad (16)$$

积分 (I) 控制器的输出与输入的积分成正比关系, 所以积分控制器的输出为:

$$I_{out} = k_i \int_0^t e(\tau) d\tau \quad (17)$$

微分 (D) 控制器的输出与输入的微分成正比关系, 所以微分控制器的输出为:

$$D_{out} = k_d \frac{de(t)}{dt} \quad (18)$$

所以系统的输出为:

$$\begin{aligned} u(t) &= P_{out} + I_{out} + D_{out} \\ &= k_p \cdot e(t) + k_i \int_0^t e(\tau) d\tau + k_d \frac{de(t)}{dt} \end{aligned} \quad (19)$$

比例控制器能够提高系统的响应速度, 随着参数 k_p 的增大, 系统响应速度变快, 但是 k_p 过大会导致过调而引起系统震荡, 因此需要加入微分控制器。微分控制器调节的是误差的变化率, 能够实现系统的超前控制, 有提前抑制误差速率变化过快的效果, 有效的抑制震荡。 k_d 越大, 系统的阻尼越大, 则系统震荡快速衰减。对于有差系统 (系统存在静态误差) 需要引入积分控制器。 I_{out} 取决于误差对时间的积分, 在静态误差存在的时候, $e(t)$ 一般很小, 但是随着时间的推移, 积分项逐渐增大, I_{out} 则能够进一步减小误差直到静态误差消除。

2.3.2. 闭环 PID. 闭环 PID 也叫串级 PID, 在系统的运行中, 由于外界的干扰, 使得加速度计的采集数据失真导致姿态解算出现错误。而由于陀螺仪的原理所定, 一般不存在受到外界干扰的情况, 因此引进角速度环组成闭环 PID 控制算法。闭环 PID 算法的流程如图 1。

闭环 PID 是在外环 (角度环) 后加入内环 (角速度环), 把外环的输出作为内环的输入, 然后与陀螺仪的

数据做差得到角速度误差, 经过角速度 PID 控制器后得到系统输出。系统的输出又可以由姿态解算得到当前的角度输出给外环, 从而形成闭环。

闭环 PID 相比于单环 PID, 不仅抗干扰能力强, 并且回复迅速, 增强了系统的鲁棒性。

3. 人脸检测与识别

长期以来, 计算机以及其他嵌入式设备在与人类的交互方面都非常受限, 通常情况下是接受键盘输入, 而难以做到像人类一样游刃有余的接收并处理各种视觉信息。科学家们为了让机器能够更友好的与人类进行交互, 在机器视觉领域不断探索, 并且至今也得到了快速发展, 取得了令人欣慰的成果。

迄今为止, 机器视觉的发展已经历了一个漫长的过程。经过研究者们不懈努力, 新的信息技术和媒体手段的出现, 使得更加有效和友好的人机交互方式得到了发展, 新型的人机交互将不再依赖传统的输入设备。而且, 计算性价比的提高和视频获取成本的降低, 使得计算机视觉系统能够向桌面级和嵌入式系统发展, 这意味着计算机视觉系统能够安装在一切电子系统之中。

计算机视觉要处理的一个重要内容, 就是对人脸的视觉处理。人脸分析的相关研究希望用户的身份、状态和意图的信息能够从图象中提取出来, 然后由计算依此做出反应 (比如通过观察用户脸部表情来分析心情并进行相应反应)。由于人脸和脸部表情已经被心理学家、神经科学家和工程师们研究了多年, 所以人脸和脸部表情识别的研究得到了更多的关注。

3.1. 人脸检测

在 2001 年, Viola 和 Jones 发表了经典的《Rapid Object Detection using a Boosted Cascade of Simple Features》[9] 和《Robust Real-Time Face Detection》[10], 在 AdaBoost 算法的基础上, 使用 haar-like 小波特征和积分图方法进行人脸检测, 他俩不是最早使用提出小波特征的, 但是他们设计了针对人脸检测更有效的特征, 并对 AdaBoost 训练出的强分类器进行级联。

3.1.1. Haar 特征. 在一幅图像中检测出人脸所在位置, 其实是属于一个分类任务, 也就是对人脸与背景进行分类。完成这个任务需要知道人脸与背景特征有哪些不同, 这是进行分类的关键。为了避免在像素层面上做计算, Paul Viola 和 Michael Jones 使用了 haar 特征,

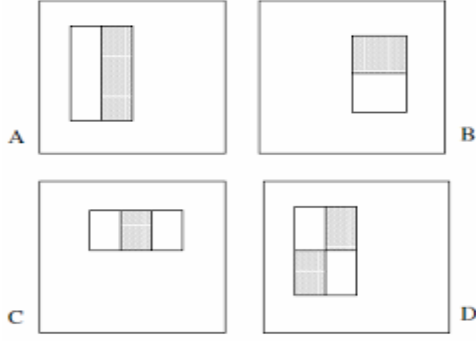


图 2: 矩形特征与检测窗口相关, 特征值为白色区域内像素之和减去灰色区域的像素之和。(图片引用自 viola 的论文 [9])

如图 2 所示。之后 Rainer Lienhart 和 Jochen Maydt 对这一特征进行了拓展, 得到性能更好的分类器, Rainer Lienhart 和 Jochen Maydt 提出的特征如图 3。这些算子与原始图像进行卷积运算便可以得到一种特征。有了计算区分人脸的特征的方法之后就可以着手解决问题了, 但是仔细观察可以发现计算某一区域的特征值就是做卷积, 并且这个计算的结果可以用更加有效的方法来实现, 这种方法就是积分图。积分图在 x, y 位置的值等于 x, y 位置左上角的所有像素的和:

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y')$$

其中 $ii(x, y)$ 是积分图, $i(x, y)$ 是原始图。使用下面两条递归方程可以快速计算出积分图:

$$s(x, y) = s(x, y-1) + i(x, y) \quad (20)$$

$$ii(x, y) = ii(x-1, y) + s(x, y) \quad (21)$$

在实现上我们使用了 Rainer Lienhart 提出的方法来做实际的检测算法。

3.1.2. AdaBoost. 有了区分人脸和背景的 haar 特征值, 下一步就要对不同的区域进行分类, 也就是判断是否是人脸。在所有计算出的特征中, 大部分是不相关的。如有些特征是判断眼睛的, 有些特征是判断鼻子的, 如何最好地综合这些特征进行分类, 就要用 Adaboost 算法来实现了。我们使用一种 AdaBoost 算法来训练分类器 [12]。Adaboost 算法的主要思想是综合多个简单的学习算法 (弱分类器) 来提高分类器的性能。Freund 和 Schapire 证明综合多个弱分类器能够使得训练分类错误率以指数速率趋近于 0 [13]。其实每个弱分类器做的事都比较简单, 就是比较输入图像的 haar-like 特征值与最初该分类器的特征值比较, 以此来判断输入图像是否

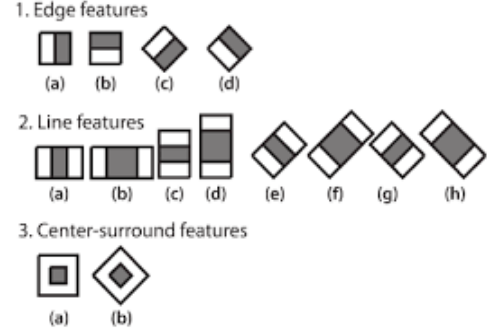


图 3: 扩展 haar-like 特征, 包括边缘特征、线特征和中心特征 (图片引用自 Rainer Lienhart 的论文 [11])

为人脸, 当然, 此分类器是非常简单的。具体数学模型如下:

$$h_j(x, f_j, p_j, \theta_j) = \begin{cases} 1 & \text{if } p_j f_j(x) < p_j \theta_j \\ 0 & \text{otherwise} \end{cases}$$

其中 x 表示图像的一个子窗口, f_j 表示特征, p_j 指示不等号方向, θ_j 表示分类器阈值。然而, 在一幅图像中, 大部分区域都不是人脸, 所以更好的思路是用一个简单的方法来判断一个区域是否是人脸, 如果不是的话接下来就不用再处理这个区域了。这样就不会花费过高的计算在非人脸区域上, 能大大提升性能。基于此想法 viola 提出了级联分类器的概念 [9], 他们把特征分成不同的级别, 然后将图像窗口的特征依次应用到每一个级别中。通常前面几级分类器包含的特征数较少, 这样只需要少数的计算就可以排除大部分非人脸区域了。

3.2. 人脸识别

3.2.1. Local Binary Pattern Histogram. 最初 LBP 操作算子是 Ojala 等人提出来的 [14], 这是纹理描述的有效方式。LBP 是提取局部特征作为判别依据, 这种方法的显著的优点是对光照不敏感。LBP 是定义在像素 3×3 邻域内的, 以邻域中心像素为阈值, 将相邻的 8 个像素的灰度值与其进行比较, 若周围像素值大于中心像素值, 则该像素点的位置被标记为 1, 否则为 0。这样, 3×3 邻域内的 8 个点经比较可产生 8 位二进制数 (通常转换为十进制数即 LBP 码, 共 256 种), 即得到该邻域中心像素点的 LBP 值, 并用这个值来反映该区域的纹理信息。如下图 4 所示。更加形式化的描述如下:

$$LBP(x_c, y_c) = \sum_{p=0}^{p-1} 2^p s(i_p, i_c) \quad (22)$$

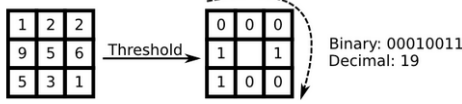


图 4: 基本 LBP 操作

其中 (x_c, y_c) 代表 3×3 邻域的中心元素, 像素值为 i_c , 邻域的其他像素值为 i_p , $s(x)$ 是符号函数, 定义如下:

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases}$$

基本的 LBP 算子的最大缺陷在于它只覆盖了一个固定半径范围内的小区域, 这显然不能满足不同尺寸和频率纹理的需要。为了适应不同尺度的纹理特征, 并达到灰度和旋转不变性的要求, Ahonen 等对 LBP 算子进行了改进 [15], 将 3×3 邻域扩展到任意邻域, 并用圆形邻域代替了正方形邻域, 改进后的 LBP 算子允许在半径为 R 的圆形邻域内有任何多个像素点。从而得到了诸如半径为 R 的圆形区域内含有 P 个采样点的 LBP 算子。对于一个给定邻域中心点 (x_c, y_c) , 其邻域采样点的位置通过以下公式计算得到:

$$x_p = x_c + R \cos\left(\frac{2\pi p}{P}\right) \quad (23)$$

$$y_p = y_c + R \sin\left(\frac{2\pi p}{P}\right) \quad (24)$$

其中 R 是圆半径, P 为采样点数, $p \in P$ 。如果上式计算得到的坐标点和图像坐标不对应 (结果不是整数), 则可以通过双线性插值来得到:

$$f(x, y) \approx \begin{bmatrix} 1-x & x \end{bmatrix} \begin{bmatrix} f(0,0) & f(0,1) \\ f(1,0) & f(1,1) \end{bmatrix} \begin{bmatrix} 1-y \\ y \end{bmatrix}$$

得到 LBP 图可以很好地捕捉图像的细节, 并对光线不太敏感。因而可以从此图中提取特征进行处理。我们选择统计直方图作为分类特征, 图像 $f(x, y)$ 的统计直方图定义如下:

$$H_i = \sum_{x,y} I\{f(x, y) = i\} \quad (25)$$

其中 $i \in \{0, 1, \dots, n-1\}$, n 为 LBP 操作得到的不同数值的数量, $I\{A\}$ 为指示函数, 定义如下:

$$I\{A\} = \begin{cases} 1 & A \text{ 为真} \\ 0 & A \text{ 为假} \end{cases} \quad (26)$$

直方图包含了图像整体的分布信息, 另一方面, 对于人脸来说, 通常需要局部特征, 如眼睛、鼻子、嘴巴等。

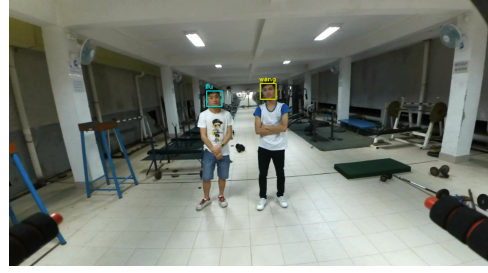


图 5: 图中结果为飞行器上的摄像机拍摄的一帧图片, 矩形框出了检测到的人脸区域, 矩形上方则为识别对象的文字标识。不同人用不同的颜色矩形框住

基于这个想法 Timo Ahonen 等人把图像分为不同的区域 R_0, R_1, \dots, R_{m-1} , 将空间增强直方图定义如下:

$$H_{i,j} = \sum_{x,y} I\{f(x, y) = i\} I\{(x, y) \in R_j\} \quad (27)$$

其中 $i \in \{0, 1, \dots, n-1\}$, $j \in \{0, 1, \dots, m-1\}$ 。我们把这些不同区域的直方图拼接起来就可以得到人脸的全局描述特征。

3.2.2. 分类准则。有了描述人脸的特征之后便可以基于此特征对人脸进行分类识别了, 为了实现的简单, 我们采取了最近邻算法作为分类器, 相似度测量准则有下面几个备选方案:

– 直方图交叉核方法:

$$D(\mathbf{S}, \mathbf{M}) = \sum_i \min(S_i, M_i) \quad (28)$$

– 对数似然统计:

$$L(\mathbf{S}, \mathbf{M}) = - \sum_i S_i \log M_i \quad (29)$$

– 卡方统计:

$$\chi^2(\mathbf{S}, \mathbf{M}) = \sum_i \frac{(S_i - M_i)^2}{S_i + M_i} \quad (30)$$

其中 S 表示样本直方图, M 表示模型直方图, i 表示直方图的第 i 个 bin, 直观来讲, 通常人脸的某些部位可能比其他部位更具有可分辨性, 如我们更倾向于通过眼睛、嘴巴和鼻子等特征来辨别一个人, 而不是额头、脖子等。基于这个启示, 可以对不同区域进行加权来表示该区域特征对识别结果的重要程度。因此 χ^2 统计方法的一个变形-加权 χ^2 统计定义如下:

$$\chi_w^2(\mathbf{S}, \mathbf{M}) = \sum_{j,i} w_j \frac{(S_{i,j} - M_{i,j})^2}{S_{i,j} + M_{i,j}} \quad (31)$$

其中 w_j 表示第 j 个局部区域的权重。

4. 实验

4.1. 人脸识别训练数据

在人脸识别过程中的一个重要步骤是获取用于训练模型的数据，本文采取直接从视频流中提取人脸的方法来获得。为了实验操作的方便，我们只对两人进行识别，此次提取人脸过程如下：

- 1) 分别拍摄两位目标对象的视频
- 2) 使用 haar cascade 方法进行人脸检测
- 3) 把检测到的人脸图像大小都改变为统一大小
- 4) 保存人脸图像

本次采用的训练数据总共有 500 个代表两位实验者正脸的灰度图像样本，250 张实验者 A 的头像，250 张实验者 B 的头像。图像包含不同光照、不同面部表情和姿势的图像，这样可以提高模型识别的鲁棒性。

4.2. Haar 人脸检测效果

人脸检测部分我们使用了开源的计算机视觉库 opencv 来实现，opencv 库中包含了已经训练好的人脸检测模型，包括复杂度低精度低的模型也有复杂度高精度高的模型，我们选择了折中的方案，使用 haarcascade_frontalface_alt2.xml 来作为人脸检测器。参数 minNeighbors 设置为 3，scaleFactor 设置为 1.3。在这部分我们更加在意的是有人脸的地方需要检测到，所以要尽量做到不漏检，错检关系不是很大，因为后面还有 LBPH 人脸识别的步骤判断目标人所在的位置，所以错检的区域不太可能判断为目标人。基于这个模型的人脸检测的效果如图 5。

4.3. LBPH 人脸检测效果

4.3.1. LBP 参数设定. 基于 LBP 的算法可以通过选择一些参数来使得性能最优化，其中包括 LBP 操作算子、图像划分为区域 R_0, \dots, R_{m-1} ，最近邻分类器距离准则还有 χ^2 统计权重的选取（方程 31）。我们选取最优窗口大小和 LBP 操作算子时发现 LBP 算法对于这些参数的选择有较强的鲁棒性，改变参数可能会对特征向量长度有较大的影响，但总体表现不会有太大的变化。这里我们选择 (8,2) 的 LBP 操作算子，即在半径为 2 的圆上采样 8 个点。另外，在实验比较中发现，使用 χ^2 准则效果会比直方图交叉核还有对数似然统计稍好一点，因此我们选用 χ^2 统计方法。

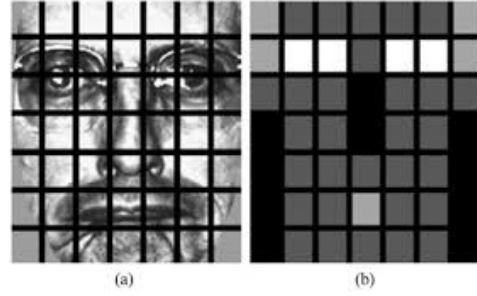


图 6: (a) 一幅面部图像分割成 7×7 的区域，(b) 加权 χ^2 统计权重。黑色方块表示权重 0.0，暗灰色表示权重 1.0，亮灰色表示权重 2.0，白色表示权重 4.0。（图片引用自 Timo Ahonen 的论文 [16]）

为了确定加权 χ^2 统计方法的权重 w_j ，只需要做些简单的计算就可以了。每次只要单独使用一个区域 R_j 来做识别，权重的确定则基于对此区域做识别的正确率，如图 6。虽然这个方法可能不是最优的，但要比不加权 χ^2 统计方法识别正确率稍微高一点。最终识别结果如图 5。

4.4. 位置 PID 整定

多旋翼飞行平台使用的 MCU 是 STM32F407VG，气压计使用的是 MS5611，IMU 使用的是 MPU6050。根据前面的人脸识别结果输出到位置 PID 控制器，以识别目标人脸中心位于图像的位置控制飞行器的横滚以及油门，以识别目标人脸在图像的大小控制飞行器的俯仰。设定目标位置为 (320,240)（图像大小为 640*480），方框目标边长度为 25 像素，然后可以得到人脸识别的结果与目标的误差，继而由位置 PID 得到控制量。控制原理为由机载摄像头得到的目标人脸位于中心位置偏右，说明飞行器位于目标的左边，则需控制横滚以控制飞行器往右飞行。假设机载摄像头得到的目标人脸位于中心位置偏上，说明飞行器高度过低，则需控制油门以控制飞行器往上飞行。假设机载摄像头得到的目标人脸大小大于设定目标值，说明飞行器离目标过于接近，则需控制俯仰以控制飞行器往后飞行。

对于位置 PID 使用单环 PID 控制器来控制飞行器相对于目标的横向以及距离的控制。所以在横向上，对于不同的 PID 参数，飞行器相对于目标的位置误差波形如图 7，在距离上，对于不同的 PID 参数，飞行器相对于目标的距离误差波形如图 8。从两图可以看到，对于参数 P 过大，飞行器在横向轴或距离轴上震荡，对于 D 过大，飞行器移动到目标位置的速度很慢（跟踪速度过慢），对于 I 过大，则飞行器在目标位置附近小幅度

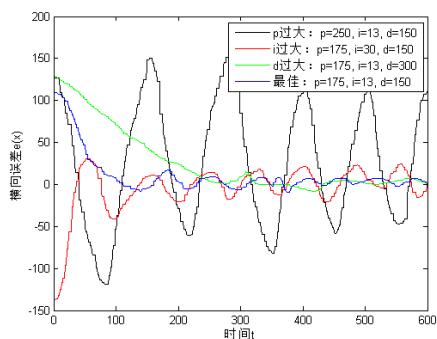


图 7: 对于不同 PID 参数飞行器相对目标的横向位置的误差变化波形图

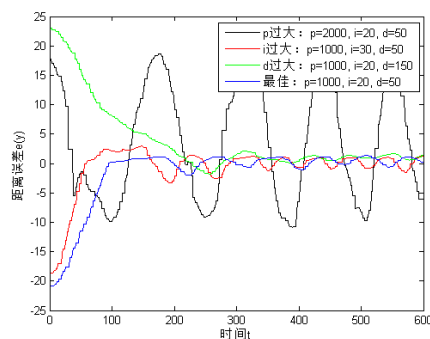


图 8: 对于不同 PID 参数飞行器相对目标的距离的误差变化波形图

的震荡。当横向以及距离两组位置 PID 参数整定完毕, 飞行器即可以实现对目标的跟踪。

5. 总结

本研究将多旋翼无人机平台与人脸识别技术结合, 对机载单目摄像头获取的图像进行人脸检测识别, 利用人脸识别的结果控制多旋翼无人机的飞行, 以达到对目标进行跟踪的目的, 并且进行了实验。从实验结果可以看出, 在单目摄像头的条件下, 飞行器能对目标进行精度较高(分米级)的全方位的近距离跟踪。由于单目摄像头硬件上的缺陷, 无法获取深度数据, 追踪精度不够理想, 但是根据脸部大小为参考进行相对距离的控制的精度也是比较可观的。由于硬件上的限制, 无法在机载平台上对实时图像处理, 只能回传到地面端进行处理。以本研究为基础, 可以考虑加入其他传感器为参考来源, 如使用双目视觉或者超声波传感器获取深度数据来进行追踪定位, 追踪的精度将会大大提高。使用 GPS 以及光流传感器辅助定位, 能够实现室外以及室内全环境的追踪定位。使用符合计算速度要求的 DSP 在机载平台对图像进行处理以及对飞行器进行控制, 真正实现自主追踪。

References

- [1] Unmanned Aircraft Systems Roadmap, 2005-2030[R].office Secretary of Defence(USA)Report August, 2005.
- [2] 胡春华. 纵列式无人直升机建模及非线性控制 [D]. 博士学位论文, 北京, 清华大学, 2004.
- [3] Lilly White Kirt, Lee Dah-Jye, TIPPETTS Beau, An embedded vision system for an unmanned four-rotor

helicopter, Proceedings of SPIE, the International Society for Optical Engineering.

- [4] S.Salazar-Cruz, J.Escareo, D.Lara, R.Lozano, Embedded control system for a four-rotor UAV, International Journal of Adaptive Control and Signal Processing, Volume 21 Issue 2-3, 31 Jan 2007.
- [5] 秦永元. 惯性导航 [M]. 北京: 科学出版社, 2006.
- [6] 邓正隆. 惯性技术 [M]. 哈尔滨: 哈尔滨工业大学出版社, 2006.
- [7] Diebel J. Representing Attitude: Euler Angles, Unit Quaternions, and Rotation Vectors[J]. Matrix, 2006, 58.
- [8] Henderson D. M. Euler Angles, Quaternions, and Transformation Matrices[R]. NASA Technical Reports, 1977, NASA-TM-74839.
- [9] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]//Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. IEEE, 2001, 1: I-511-I-518 vol. 1.
- [10] Viola P, Jones M J. Robust real-time face detection[J]. International journal of computer vision, 2004, 57(2): 137-154.
- [11] Lienhart R, Maydt J. An extended set of haar-like features for rapid object detection[C]//Image Processing. 2002. Proceedings. 2002 International Conference on. IEEE, 2002, 1: I-900-I-903 vol. 1.
- [12] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In Computational Learning Theory: Eurocolt ' 95, pages 23-37. Springer-Verlag, 1995.
- [13] Robert E. Schapire, Yoav Freund, Peter Bartlett,

and Wee Sun Lee. Boosting the margin: A new explanation for the effectiveness of voting methods. In Proceedings of the Fourteenth International Conference on Machine Learning, 1997.

- [14] Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions[J]. Pattern recognition, 1996, 29(1): 51-59.
- [15] Ahonen T, Hadid A, Pietikäinen M. Face recognition with local binary patterns[M]//Computer vision-eccv 2004. Springer Berlin Heidelberg, 2004: 469-481.
- [16] Ahonen T, Hadid A, Pietikainen M. Face description with local binary patterns: Application to face recognition[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2006, 28(12): 2037-2041.