光电技术及应用

DOI：10.16818/j.issn1001－5868.2022120901

微型仿生扑翼飞行器机载单目视觉系统的设计

赵　燕[1,2]，张卫平[1*]，郭倾城[1,2]，牟家旺[1,2]，郭彦含[1,2]

（1. 微米纳米加工技术全国重点实验室，上海 200241；2. 上海交通大学 微纳电子学系，上海 200241）

摘　要： 近年来仿生扑翼飞行器利用视觉系统自主飞行成为一个具有广泛前景的研究方向，然而，其有限的带载能力对视觉传感器的类型、尺寸和重量提出了严格要求。目前商用图像处理模块的尺寸和重量较大，且需要回传图像信息至地面控制系统处理，文章旨在设计一款轻量化机载单目视觉系统，帮助微型仿生扑翼飞行器获取外界信息并实现智能自主的飞行。相比于其他图像处理模块，此系统以国产高算力芯片K210为核心进行设计，可脱离电脑端完成图像处理，尺寸仅为2.2 cm×2.3 cm，重量仅为3 g，内部兼容轻量化网络模型实现分类识别，通过串口进行信息交互，控制扑翼飞行器实现手势识别和目标追踪。

关键词： 仿生飞行器；单目视觉系统；卷积神经网络；轻量化；目标识别；机载图像处理；串口通信；自主飞行

中图分类号： TP391.41；V276　**文章编号：** 1001－5868(2023)02－0257－08

Design of An Onboard Monocular Vision System for Miniature Bionic Flapping Wing Aircraft

ZHAO Yan[1,2], ZHANG Weiping[1], GUO Qingcheng[1,2], MOU Jiawang[1,2], GUO Yanhan[1,2]

(1. National Key Laboratory of Science and Technology on Micro/Nano Fabrication, Shanghai Jiao Tong University, Shanghai 200241, CHN; 2. Department of Micro/Nano Electronics, Shanghai Jiao Tong University, Shanghai 200241, CHN)

Abstract: In recent years, the autonomous flight of bionic flapping wing aircraft using vision systems has become a widely promising research direction. However, its limited carrying capacity puts forward strict requirements for the type, size and weight of vision sensors. At present, the size and weight of commercial image processing modules are large, and the image information needs to be transmitted back to the ground control system for processing. Therefore, this paper aims to design a lightweight onboard monocular vision system to help the miniature bionic flapping wing aircraft obtain external information and realize intelligent and autonomous flight. Compared with other image processing modules, this system was designed with domestic high-computing power chip K210 as the core, which could complete image processing without the computer side. The size is only 2.2 cm×2.3 cm, and the weight is only 3 g. The internal compatible lightweight network model realizes classification recognition, and information interaction is achieved through the serial port, which can control the flapping wing aircraft to achieve gesture recognition and target tracking.

Key words: bionic aircraft; monocular vision system; convolution neural network; lightweight; target recognition; onboard image processing; serial communication; autonomous flight

收稿日期：2022－12－09.

*通信作者：张卫平　**E-mail：**zwp37@163.com

0 引言

微型仿生扑翼飞行器是指翼展尺寸小于30 cm、重量低于50 g、通过翅膀往复拍打产生升力的飞行器，其轻质量和小尺寸的特点对机载传感器的重量、尺寸和功耗提出了严格要求。为了使仿生扑翼飞行器具备视觉传感能力，2014年荷兰的代尔夫特大学在20 g的微型仿生飞行器Delfly[1]上搭载了重量为24 g的具有图像处理功能的立体视觉系统，该系统由两个数字摄像头和一个Stm32F405处理器组成，实现了扑翼飞行器室内自主飞行巡航功能，然而，Stm32系列芯片的处理能力有限，不能完成目标检测等工作。2016年，Ryu等在一款仿鸟飞行器上搭载了低功耗的单目摄像头和回传系统，实现了循迹运动[2]。此外，华盛顿大学经过数十年的发展，将一个微小型视觉系统集成到微型爬行机器人背部[3]，限制帧率为1～5 f/s，可将图像回传至手机端。

上海交通大学微系统集成研究所王昕鹏等以商用无线视觉模块为硬件主体，针对仿生扑翼飞行器飞行过程中的高频抖动和人像识别算法做了改进[4]；舒童等针对仿生扑翼飞行器进行轻量化目标识别系统的设计，在保证训练精度的前提下减小了网络计算量[5]。

借用地面计算机进行图像处理限制了扑翼飞行器远距离飞行，实用性和实时性会有所降低[6]，利用微小型处理器自主设计可脱离地面计算机的机载视觉感知系统，是一个巨大的挑战。本文针对仿生飞行器载重有限的问题设计了一款微小型、可机载、低延迟的视觉系统，内部可利用神经网络模型实现图像实时处理。该视觉系统与扑翼飞行器通过信息交互完成目标识别和目标追踪等功能，未来有希望作为仿生机器人感知未知环境的主要传感器。

1 仿生飞行器平台

微型仿生扑翼飞行器如图1所示，包括由碳纤维构成的骨架、作为执行器的电机、电池和飞控电路板。扑翼飞行器的机翼往复拍打产生升力，由翅膀拍打幅值的差异产生多轴力矩，可以实现悬停和任意方向飞行，具有广泛的飞行包线。该飞行器主控为Stm32F446，具备串口通信功能和遥控功能，剩余载荷为5 g，飞行器和摄像头的参数见表1。

针对图1中的仿生飞行器平台，视觉处理系统的重量和功耗都要尽量小。目前市面上的图像传感器大多需要将图像信息回传至电脑端处理，易受干扰且远距离通信受限，因此，为扑翼飞行器自主设计视觉处理系统非常具有研究意义。

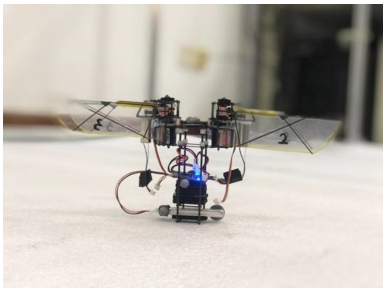


图1 微型仿生扑翼飞行器

表1 仿生飞行器参数

仿生飞行器主体		摄像头 OV2640	
构件	参数	性能	参数
重量/g	17	重量/g	0.5
长度/cm	17	像素	320×240
翼展/cm	25	数据位/bit	8
机翼角度/(°)	45	类型	CMOS
飞行速度/(flaps/s)	19	协议	SCCB
电池/mAh	250	视场角/(°)	180
主控芯片	Stm32F446	分辨率	UXGA

2 硬件设计

本文设计的单目视觉处理系统包括两个模块，即单目视觉主模块(包括图像处理模块、图像采集模块和视频存储模块)和程序及模型下载模块。

视觉系统在运行之前需要先采集图片训练模型，之后将模型下载到K210，这是系统识别图像的前提条件。

视觉处理系统的功能实现流程如图2所示。在飞行过程中，图像采集传感器OV2640将采集到的信息传给视觉处理器，将模型识别后的输出结果实时传输给扑翼飞行器主控。飞行器根据视觉信息控制微型机器人的运动，飞行器在飞行过程中可以通过SD卡保存视频信息，实现存储功能。

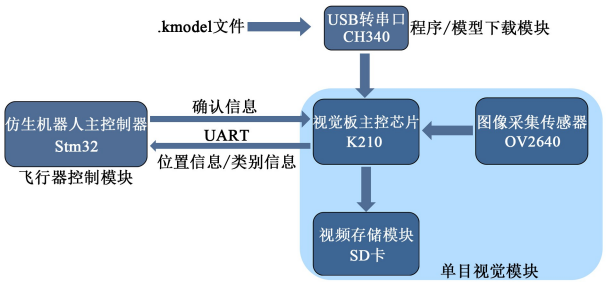


图2 硬件信息处理流程图

本节对单目视觉处理系统两个部分的电路设计

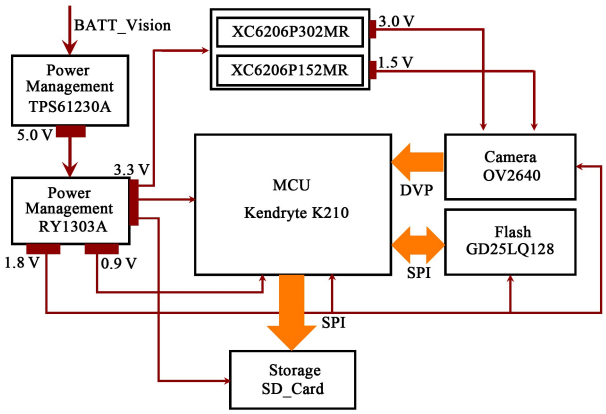
和优化进行介绍，为小型化的单目视觉系统实现手势识别和目标检测提供硬件载体。

2.1　单目视觉主模块

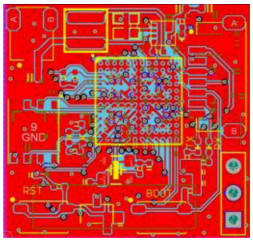
K210 作为一款国产 AI 图像处理芯片，算力高达 1 TFLOPS，功耗仅为 300 mW，芯片采用 BGA 封装，尺寸仅为 8 mm×8 mm，满足仿生扑翼飞行器视觉系统的小尺寸、低功耗和高算力要求，可以作为飞行器图像处理的主控芯片。本文基于 K210 芯片设计了一款单目视觉处理系统作为仿生扑翼飞行器的视觉传感器。

视觉系统主板的电路设计在矽速商用模块系列开发板 Maixbit 的基础上进行了二次开发，通过修改器件型号、封装、紧凑布局、物理分割等方法，使电路板的尺寸由原来商用模块的 5 cm×2.7 cm 缩小为 2.2 cm×2.3 cm，重量由原来的 8 g 缩小为 3 g。

单目视觉主板采用四层板设计，内部集成了图像处理模块，flash 扩容模块，电源管理模块以及 SD 卡存储接口模块，具体硬件电路见图 3 所示。



（a）框架结构



（b）PCB



（c）实物

图 3　视觉主板硬件

在图像采集模块中，本文的视觉处理系统搭载的摄像头为 OV2640。OV2640 是一款支持低电压的 CMOS 图像传感器，在小尺寸的封装内部提供单芯片摄像头和图片处理器。模块中出现的 XC6206 模块用于稳压，为摄像头接口提供 1.5 V 供电和 3.0 V 供电，分别作为模拟地和数字地。

在图像处理模块中，K210 外部接 128 M 的 flash 芯片来扩大存储容量，留出 RST 复位接口和 BOOT 模式选择接口，用于下载程序时的调试。另外，外接两个 LED 灯作为 RX 和 TX 收发信号的指示标识，可以直观感受硬件当前收发信息的状态。

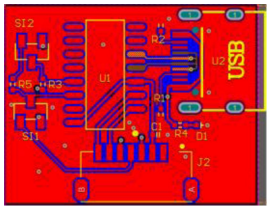
在电源管理模块中，主控芯片为隔离不同信号，需要多个稳压模块。电源管理模块中用到的 RY1303A 作为多路降压 DC-DC 芯片，内部包含三路稳压，可以在实现功能的前提下保证更小尺寸。此外，电路中的 SGM803 电源监管芯片，在上电时会自动提供一个复位信号。

在视频存储模块中，提供 SD 卡插入接口，数据传输采用 SPI 协议控制读写，进行高速同步通信。仿生机器人在运动过程中所采集到的所有视觉信息都可以通过 SD 卡进行保存。

2.2　程序及模型下载模块

为实现系统轻量化，本文将单目视觉板核心部分与下载部分实现物理分割，设计了可插拔电路。下载固件或模型时，视觉系统和下载系统依靠连接线传输数据，其他情况下单目视觉系统可脱离固件下载系统单独供电运行。

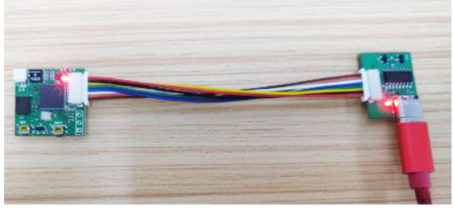
下载板采用 CH340 作为 USB 转串口芯片进行电平转换，此外，USB 的数据差分线采用了阻抗匹配技术[7]，差分阻抗设置为 90 Ω，提高了数据下载速度，具体电路硬件如图 4 所示。



（a）PCB



（b）实物照片



（c）与视觉板硬件连接实物照片

图 4　下载板硬件

3　模型训练

K210 芯片内存资源有限，要求网络模型尽可能轻量化和高效。K210 的模型训练采用了 Tensorflow 作为训练框架，在其本地训练环境中集

成了 YOLOv2[8]和 MobileNet[9]两种训练方式。单独使用 MobileNet 可用于目标分类，使用以 MobileNet 作为骨干网络的 YOLOv2 网络可用于目标检测。图 5 表示模型训练的流程框图。

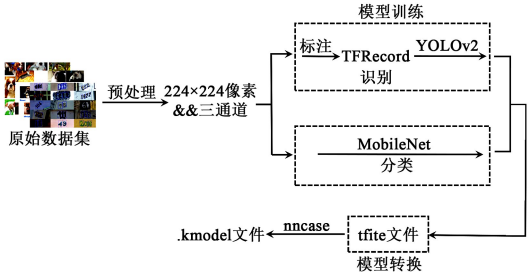


图 5 模型训练流程

3.1 目标分类

本文中目标分类采用深度可分离卷积[10]实现对卷积的轻量化操作。可分离卷积中的深度可分离卷积为将普通卷积拆分成为一个深度卷积和一个逐点卷积，深度可分离卷积与标准卷积的参数量之比和计算量之比 R 均为：

$$R=\frac{1}{N}+\frac{1}{D_K^2} \tag{1}$$

式中，k 为卷积核边长大小，N 为卷积核个数。当我们用到的卷积核大小为 3×3 的时，对应的计算量大约会下降到原来的 1/9。

MobileNet 输出模型的参数仅为 4 M 左右，模型经 nncase 量化之后可以从 Float 类型压缩到 Int 类型，使得模型资源再缩小几倍，满足 K210 的模型资源限制。

3.2 目标检测

近年来目标检测算法发展较为迅速，YOLO 系列算法作为主力军占据重要地位。相比于 YOLOv1 利用 GoogleNet 为骨干网络，YOLOv3/v4/v5 利用 darknet53 作为骨干网络，YOLOv2 中用到的 darknet-19 只有 19 层卷积神经网络，运算量更小，fps 是 darknet-53 的两倍之多，可以满足系统实时性；输出模型参数相对较少，满足 K210 的内存限制；相对于后期出现的 YOLOv6 和 YOLOv7，YOLOv2 较为成熟，便于算法移植和检错。

基于上述描述，本文中涉及到的目标检测采用 YOLOv2 作为训练网络，输出类别信息和位置信息，并结合 MobileNet 轻量化及高精度的特点，用 MobileNet 代替传统的 darknet-19 作为的 YOLOv2 的骨干网络，可以进一步缩短训练时间并减小模型参数。

3.3 模型训练

3.3.1 数据集采集

本文的数据集包括 paper，rock，scissors 三种手势，有两种方法采集数据，一种是从网络上下载背景较为简单的数据集（AI_GUY 数据集），另一种是利用单目视觉板中自带的 OV2640 完成图像采集。数据集采集完成后，训练前将图片预处理为像素点为 224×224 且通道数为 3 的固定尺寸。

考虑到 K210 的轻量化网络及训练时长，每种手势分别采集 100 张，120 张，150 张，260 张，320 张，420 张图片进行训练。数据采集情况及训练结果如表 2 所示。

表 2 数据采集情况

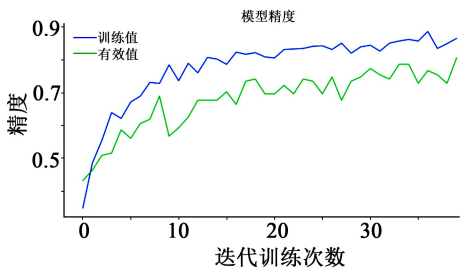
数据集来源	Paper 数量	Rock 数量	Scissor 数量	训练损失（loss）	训练精度（accuracy）
OV2640 拍摄	100	100	100	0.383 1	0.900 0
	120	120	120	0.406 1	0.879 9
	150	150	150	0.332 2	0.866 1
	260	260	260	0.374 6	0.898 4
	320	320	320	0.380 1	0.863 7
	420	420	420	0.351 7	0.888 3
AI_GUY 数据集	100	100	100	0.218 1	0.983 3
	120	120	120	0.233 2	0.968 3
	150	150	150	0.231 9	0.960 3
	260	260	260	0.201 1	0.953 0
	320	320	320	0.175 2	0.960 1
	420	420	420	0.169 8	0.957 0

针对不同采集方法下的不同数量的数据集，给出最优模型训练结果。

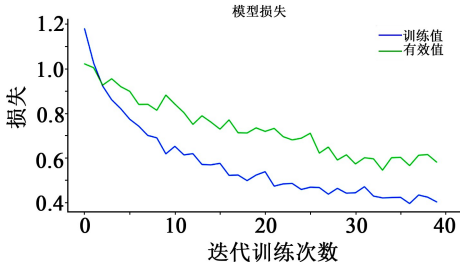
3.3.2 数据集训练

根据表 2 的训练结果，对于开发板摄像头采集到的数据集，每种手势图片为 260 张时模型效果较好。图 6 给出了此种情况下的模型训练结果，包括模型精度，模型损失及混合预测矩阵。

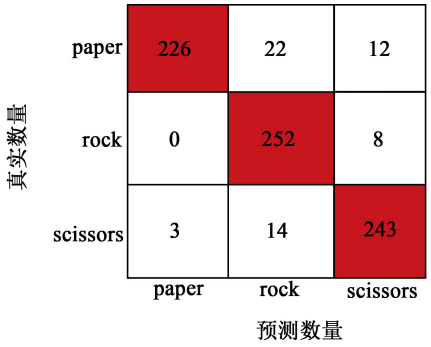
根据上面的实验结果可以看出，随着迭代次数的增加，模型精度和损失在不断优化，但是训练值和有效值有一定的差距。其原因主要是自采集数据集的背景杂乱，给轻量化网络的特征提取带来了难度。这表明在模型训练和实际识别时需要尽量保证特征明显。



（a）模型精度和迭代次数的关系



(b) 模型损失和迭代次数的关系



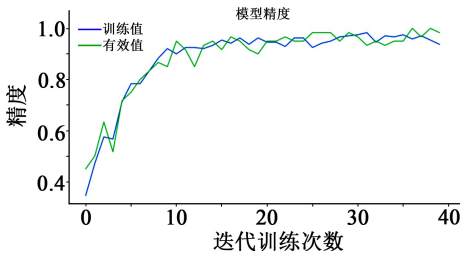
(c) 混合预测矩阵

图 6　自采集数据集的训练结果

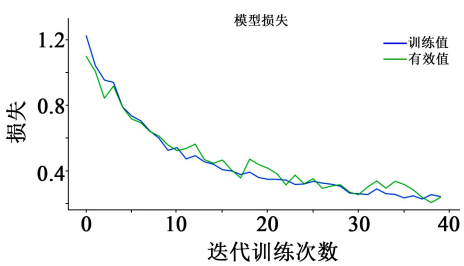
相比于摄像头自采集的数据，AI_GUY 数据集的背景简单很多。根据表 2 的训练结果，对于 AI_GUY 训练集，当每种手势图片为 100 张时，训练结果最优，此种情况下的结果如图 7 所示。

训练结果直观地显示了更换数据集之后，模型训练结果中的有效值和训练值的差距在不断缩小。模型精度最大可以达到 0.983 3，训练损失最低可以下降到 0.169 8，保证了误差在可接受范围内，在轻量化小模型的网络上属于较为可观的训练效果。

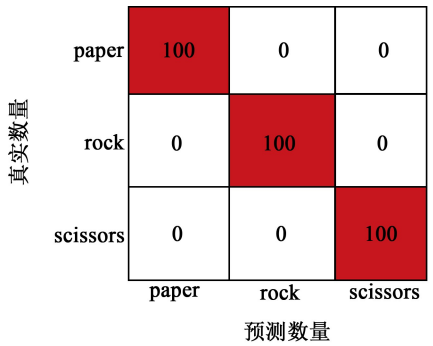
针对目标识别，本文以人脸为识别目标，旨在实现人脸追踪的功能。在模型训练的过程中使用



(a) 模型精度和迭代次数的关系



(b) 模型损失和迭代次数的关系



(c) 混合预测矩阵

图 7　网络数据集的训练结果

VOTT 作为数据集标注工具，之后送入 YOLOv2 网络进行模型训练。模型训练方法和目标分类类似，将训练结果下载入单目视觉板的主控芯片，可以观察到人脸识别结果。

4　视觉信息传递与运动控制

考虑到仅传输识别结果，数据量较小，所以无需采用复杂协议传输，K210 芯片采用串口协议[11]与 stm32 进行通信，串口通信结构如图 8 所示。在串口通信中设置波特率为 115 200，设置帧头 0xb2 和 0xb3 以及帧尾 0x5b 分别表示数据发送的起始和结束。每帧数据传输结束后，标志位 Rxflag 置位，进入中断根据相应的传输内容进行逻辑处理，在飞控的操作系统 FreeRTOS 中开启新的进程控制扑翼飞行器的飞行。算法部分的框架如图 9 所示。



图 8　串口通信

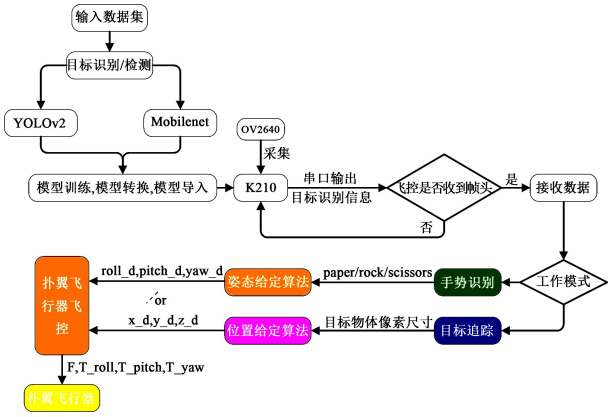


图 9　算法框图

图 9 中的 roll_d，pitch_d，yaw_d 是 pitch，roll 和 yaw 的期望角度。x_d，y_d，z_d 为空间中的期望

位置。扑翼飞行器视觉控制算法主要由两个部分协作完成，即姿态给定算法和位置给定算法，二者可实现视觉系统对扑翼飞行器的运动控制。

算法框图中提到的姿态给定算法的主要逻辑是根据识别到的手势类别对 pitch，yaw，roll 的值进行调整，使其经过控制系统后，输出对应角度期望的力矩和拉力，从而控制扑翼飞行器姿态。算法框图中提到的位置给定算法主要由两部分组成，第一部分是根据识别到的物体中心点坐标完成初始位置的调整，将其调整到中心位置；第二部分是根据识别物体的像素点乘积大小控制扑翼飞行器跟进、悬停或后退，实现目标追踪功能。

针对扑翼飞行器飞行过程中的抖动问题，上海交通大学牟家旺等提出的基于模型的抗扰控制算法[12]和多轴自适应控制器[13]在扑翼飞行器实际飞行中可以很好的抑制外部环境及自身抖动对控制产生的干扰，减少抖动对图像处理产生的负面影响，增加图像识别的准确率。

4.1 手势识别

将训练出的模型导入到飞行器中可以实现手势分类的功能，手势分类图片的左上角表示物品类别概率，左下角数据表示识别该物体所需要的时间，效果如图 10 所示。



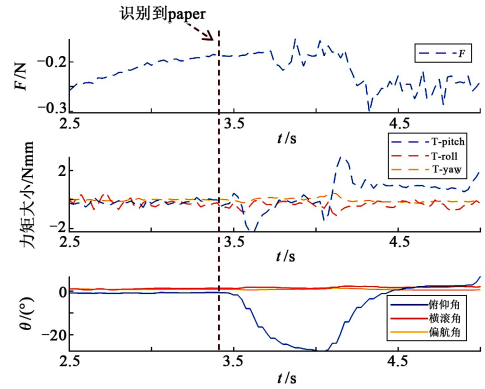
图 10 手势分类测试

由图 10 可以看出，对于 rock 和 paper，识别效果较好，对于 scissor 识别精度偏低，这种情况的原因可能是数据采集时 scissors 的特征不明显。

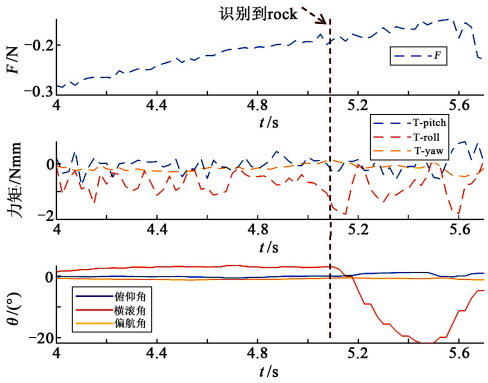
在手势识别命令下，根据图 9 使用姿态给定算法。paper，rock 和 scissors 经串口通信后对应到仿生飞行器的状态依次为：向前飞，向左飞，向左转并恢复。仿生飞行器识别到不同手势之后地面站输出的各个姿态角变化，升力和三轴力矩如图 11 所示。

以识别到 paper 为例，当摄像头识别到 paper 之后，视觉控制主板通过串口传输类别序号及概率值到扑翼飞行器控制主板，当概率值大于阈值 0.5 时即认为当前识别到的物体确定为 paper。

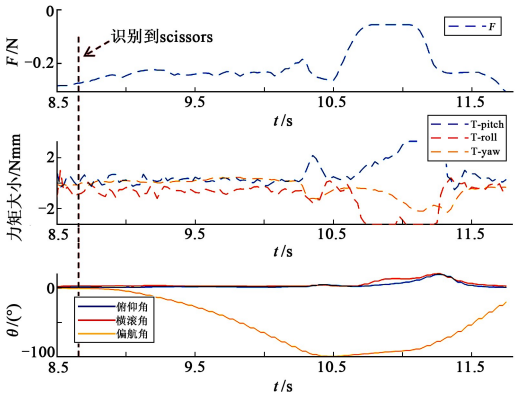
根据 11(a)中俯仰角 pitch 的变化可以看出，当确认识别到 paper 之后扑翼飞行器迅速做出反应，通过改变 pitch_d 改变自身的俯仰角度，从而使飞控完成了向前飞的指令。手势识别中对 rock 和 scissor 的识别原理与此类似，分别通过改变 roll_d 和 yaw_d 来改变 roll 和 yaw，实现向左飞和向左转并恢复的功能。



(a) 识别 paper



(b) 识别 rock



(c) 识别 scissors

图 11 手势分类对姿态的影响

4.2 人脸跟踪

对于人脸识别，方框框出的是人脸的位置，图片左下角显示结果为识别时间。单目视觉系统识别到目标后的结果如图 12 所示。

在人脸追踪的命令下，根据图 9 使用位置给定

算法。这种情况下，摄像头识别到物体之后向扑翼飞行器传输物体类别信息、识别框在整体像素点的中心位置以及识别框的高度和宽度信息。扑翼飞行器接收到信息之后修改自身 pitch，roll，yaw 的期望值，并按照各个方向的比例系数往此位置偏移，使人脸识别框尽量保证处在整体像素点的中心位置，之后再根据“近大远小”的原则控制飞行器调节的前进状态完成目标跟踪。

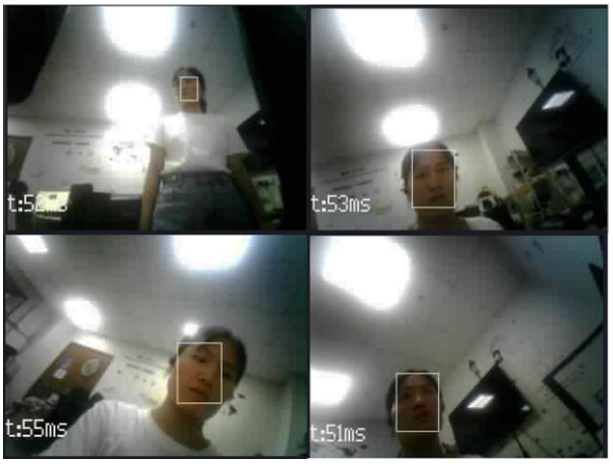


图 12　目标追踪测试

本文用 vicon 相机捕捉系统对扑翼飞行器的运动轨迹。图 13 截取了一段扑翼飞行器飞行过程中识别到人脸之后的自身运动状态的变化情况，包括升力、三轴力矩和空间中的位置变化。

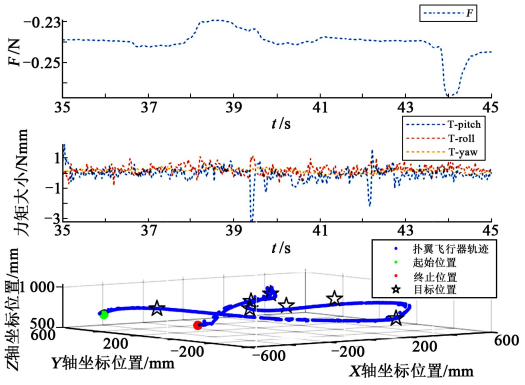


图 13　目标追踪中升力、三轴力矩和位置的变化

图 13 最下方的图反映了扑翼飞行器对人脸的追踪情况，星号表示不同时间点人脸的位置，蓝色线为扑翼飞行器在三维空间的运动轨迹。由图 13 的实验结果来看，追踪效果良好。

5　结束语

本文为扑翼飞行器设计了一个可脱离电脑端完成图像处理的单目视觉系统，扑翼飞行器可以借助此系统实现视觉感知和自主运动。本文的主要工作在于：

（1）设计了一款尺寸为 2.2 cm×2.3 cm，重量仅为 3 g，无需图像回传至电脑端的低功耗、小尺寸、高集成度的机载实时图像处理系统。通过将视觉处理芯片和模型下载板进行物理分割，紧凑化布局及选取微小型封装等方式，使其满足扑翼飞行器带载限度。

（2）自采集多种数量的数据集，在轻量化网络上训练了小参数的手势识别模型和人脸检测模型，将最优模型成功导入视觉系统，为扑翼飞行器完成后续指令提供基础。

（3）采用 UART 通信协议实现了仿生飞行器控制主板和视觉板之间的通信，完成了飞行器内部的算法控制代码编写，仿生飞行器根据摄像头采集到的图像处理对自身姿态进行调整和控制，完成特定功能。

参考文献：

[1] De Wagter C, Tijmons S, Remes B D W, et al. Autonomous flight of a 20-gram flapping wing mav with a 4-gram onboard stereo vision system [C]// 2014 IEEE Inter. Conf. on Robotics and Automation (ICRA), 2014: 4982-4987.

[2] Ryu S, Kwon U, Kim H J. Autonomous flight and vision-based target tracking for a flapping-wing MAV[C]// 2016 IEEE/RSJ Inter. Conf. on Intelligent Robots and Systems (IROS), 2016: 5645-5650.

[3] Iyer V, Najafi A, James J, et al. Wireless steerable vision for live insects and insect-scale robots[J]. Sci. Robot., 2020, 5 (44): eabb0839.

[4] 王昕鹏，张卫平，牟家旺，等. 应用于微型扑翼飞行器的实时视觉系统设计[J]. 半导体光电，2020，41(1)：114-117.
Wang Xinpeng, Zhang Weiping, Mou Jiawang, et al. Design of real-time vision system applied in flapping-wing micro air vehicle[J]. Semiconductor Optoelectronics, 2020, 41(1): 114-117.

[5] 舒　童，张卫平，牟家旺，等. 基于单目的轻量化目标识别系统设计[J]. 半导体光电，2021，42(2)：275-280.
Shu Tong, Zhang Weiping, Mou Jiawang, et al. Lightweight design of target recognition system based on monocular vision [J]. Semiconductor Optoelectronics, 2021, 42(2): 275-280.

[6] 付　强，陈向阳，郑子亮，等. 仿生扑翼飞行器的视觉感知系统研究进展[J]. 工程科学学报，2019. 41(12)：1512-1519.
Fu Qiang, Chen Xiangyang, Zheng Ziliang, et al. Research progress on visual perception system of bionic flapping-wing aerial vehicles[J]. Chinese J. of Engineering, 2019, 41(12): 1512-1519.

[7] Pues H F, Van De Capelle A R. An impedance-matching technique for increasing the bandwidth of microstrip antennas [J]. IEEE Trans. on Antennas and Propagation, 1989, 37 (11): 1345-1354.

[8] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C]// 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2017: 6517-6525.

[9] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications [EB/OL]. 2017-4-17[2023-5-23]. http:// arXiv. org/pdf/1704.04861.pdf.

[10] Sheng T, Feng C, Zhuo S, et al. A quantization-friendly separable convolution for mobilenets [C]// 2018 1st Workshop on Energy Efficient Machine Learning and Cognitive Computing for Embedded Applications (EMC2), 2018: 14-18.

[11] Nanda U, Pattnaik S K. Universal asynchronous receiver and transmitter (UART) [C]// Inter. Conf. on Advanced Computing and Communication Systems (ICACCS), 2016: 1-5.

[12] Mou J, Zhang W P, Zheng K X, et al. More detailed disturbance measurement and active disturbance rejection altitude control for a flapping wing robot under internal and external disturbances[J]. J. of Bionic Engineering, 2022, 19: 1722-1735.

[13] Mou J, Zhang W P, Wu C F, et al. Adaptive control of flapping-wing micro aerial vehicle with coupled dynamics and unknown model parameters[J]. Appl. Sciences, 2022, 12 (18): 9104.

作者简介：

赵　燕(1998—),女,硕士研究生,主要研究方向为仿生机器人的单目视觉、目标识别及仿生机器人的自主运动;

张卫平(1971—),男,教授,博士生导师,主要从事微机电系统及智能无人微系统、仿生微机器人信息处理技术、微惯性技术的研究;

郭倾城(1997—),女,博士研究生,主要研究方向为微型仿生飞行器、机器人运动控制、鲁棒控制,以及自适应控制。