

# Guidance——机器人视觉传感平台

翻译: 刘佰鑫 苏琦

Guyue Zhou, Lu Fang, Ketan Tang, Honghui Zhang, Kai Wang, Kang Yang

{ guyue.zhou, ketan.tang, honghui.zhang, kevin.wang, kang.yang } @dji.com,

Dajiang Innovations Technology Co., Ltd., Shenzhen, P.R.China

fanglu@ustc.edu.cn, University of Science and Technology of China, P.R. China

**摘要:**视觉传感器作为智能机器人的关键技术,可以实现机器人定位、导航和目标跟踪。但是,当前市场上缺乏高性能的机器人视觉传感器平台。本文将介绍一个功能强、效率高的视觉传感平台Guidance,其包含一个中央处理器和多个(五个以上)伺服传感单元,内置的基础功能有:视觉定位、避障、景深图像生成。此外,Guidance配套有完整的SDK文档,可供使用者灵活使用并开发相关的项目,譬如自主导航、目标跟踪、定位和地图构建等。

**关键词:**Guidance 视觉传感

**DOI:** 10.13211/j.cnki.pstech.2016.03.003

## 1 简介

随着智能时代的到来,智能机器人开始走进千家万户,改变着人们的生活方式。例如,作为机器人的类型之一的无人机(DJI Phantom),将人们的视角由地面拓展到高空。由于智能机器人有着体积小、价格低、可操作性强和传感能力好等特点,在航拍、航测和精细农业等民用领域得到了广泛的应用。

视觉传感器是使机器人更加智能化的关键技术,具有输出高效且价格低廉的特点。目前,视觉传感器在智能机器人中如何应用已经被大量广泛研究。以定位为例,视觉测量(Visual odometry, VO)研究了如何基于视觉特征估计机器人局部运动状态<sup>[11, 18, 12]</sup>。即时定位与地图构建(Simultaneously localization and mapping, SLAM)扩展了VO的研究范围,即研究了如何在定位的过程中构建起全局地图<sup>[7, 14]</sup>。为了解决单目视觉SLAM不可测量物体尺寸的固有缺陷,将SLAM与惯性导航技术融合,已经逐渐成为一种趋势<sup>[15, 23]</sup>。与此同时,多目视觉SLAM技术也逐渐成熟,其可以通过标定摄像头的外部参数解决了尺寸测量的问题。

基于视觉导航是机器人在复杂环境下进行无碰撞移动的关键技术。为了将机器人导航通过可

行走的区域,需要构建出三维空间地图。使用扩展卡尔曼滤波更新三维空间地图<sup>[7]</sup>,同时采用逆深度参数化算法进行优化<sup>[6]</sup>。另外,也可采用光束平差算法<sup>[14]</sup>解决地图更新的问题。为了减少空间地图的存储负担,通常使用多尺度Octomap算法<sup>[13]</sup>。

基于视觉的目标跟踪是智能机器人的另一关键技术。在视频中跟踪特定物体的研究已开展了数十年<sup>[4, 30]</sup>,近年来,研究人员开始关注机器人应用中的实时跟踪技术<sup>[21, 24, 28]</sup>。但是,现有的方法都依赖于深度传感器<sup>[21]</sup>或离线计算<sup>[24, 28]</sup>。这意味着将额外增加成本、降低可靠性和适用性。

尽管智能控制系统的需求量越来越大,关于视觉传感器的研究也越来越多,但市场依然缺乏性能足够优异的产品。PX4FLOW<sup>[12]</sup>是目前市面上唯一一款单目视觉传感器,该设备采用声纳辅助测量景深。但它的光流计算仅能达到 $64 \times 64$ 的分辨率,这限制了它的量程及测量精度。VI-Senor<sup>[27]</sup>是一个多目视觉系统,它能以较高的分辨率实时检测和提取目标特征。但是,它只有一个工作面,检测的角度较窄。此外,该设备在某些环境下会失效,比如视觉盲区、玻璃窗和水下环境。同时,市场上还有一些娱乐级的视觉传感

器，比如微软的Kinect和ASUS的Xtion，由于它们使用红外传感器而且体积和重量较大，因而只能在室内环境使用。

本文介绍一款全新的板载级智能机器人视觉平台Guidance。它包含一个计算处理器和五个以上的空间传感单元（Stereo Sensing Units，SSU），如图1（a）所示。为了弥补基于视觉算法可能出现的错误，每一个SSU都配置一个超声波传感。Guidance平台提供了精准的、大量程的、宽视角的视觉测量，丰富的物理量输出和完善的SDK文档，使得基于Guidance的机器人开发变得更简单。

本文的第二节介绍Guidance内置的算法和功能。第三节介绍SDK的细节信息。第四节举例展示Guidance的潜在应用领域。

## 2 内在特征

一个完整的Guidance系统包含一个计算处理器和多个传感单元，见图1（a）。具体来说，一个传感单元含有两个全局光圈灰度摄像机（640×480分辨率）和一个超声波传感器，这些传感器被巧妙地安装在支架上。计算单元模块包括惯性传感器（MPU6050）、一个性价比高的SOC FPGA（Altera Cyclone V）和五个传感单元接口。该设备主要面向两类目标客户：（1）传统的无人机玩家；（2）机器人应用开发者。本节将从系统概述、算法细节来介绍针对无人机玩家而设计的Guidance内置特征。

### 2.1 系统概述

图1（b）展示了配置有Guidance的四旋翼无人机，该无人机可以在没有GPS的情况下进行悬停定位和避障。严格来讲，Guidance实际上是对Zhou等人研究<sup>[31]</sup>的补充完善。这项工作中，使用了四个摄像头和一个单核计算单元（Altera's SoC FPGA）完成地图构建。Guidance的内置功能框如图2所示，简要介绍如下：

使用五个相互独立的地图重绘核处理20Hz的图像数据，每次生成大约10张无畸形的校正图像（320×240 QVGA）。尽管摄像头可以提取VGA分辨率的图像，但为了方便实时处理，将采

样降低为QVGA图像。随后以下的两个线程同时运行，进行图像重绘：（1）使用视觉测量算法进行系统定位（2）为感知和避障系统重绘地图。



(a) Guidance的传感器和计算单元



(b) 安装Guidance的四旋翼飞机

图1

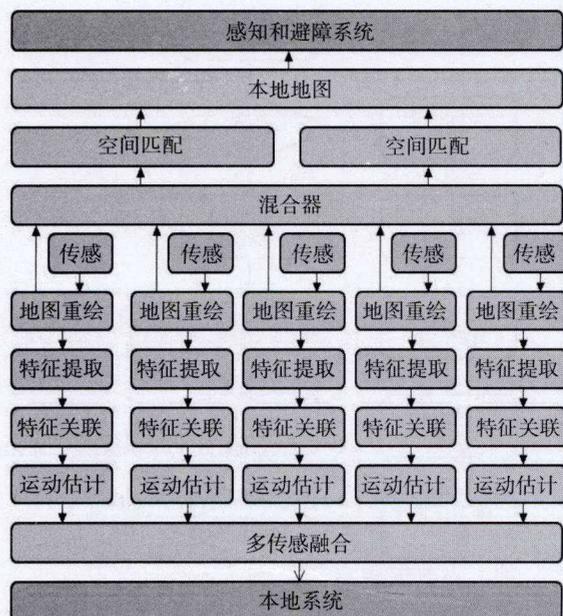


图2 Guidance的内置功能框图：灰色、蓝色、紫色和绿色分别表示系统输入，FPGA硬件核心、ARM软件核心和系统输出

在第一个线程中，运行FAST特征检测<sup>[25, 26]</sup>，BRIEF特征描述器<sup>[5]</sup>和双目匹配算法，提取并关联图像特征<sup>[31]</sup>。运动估计和多传感器融合算法将在后面章节详细介绍。在第二个线程中，深度图像将发挥重要作用。由于多张图像匹配会消耗过多FPGA计算资源，因此采用混合器预先筛选出两张最适合匹配的图像。筛选的标准与无人机移动的方向有关。

## 2.2 算法

内置的算法可以分为三类：视觉测量，视觉映射和多传感器融合。像素级的处理算法、视觉映射和多传感器融合已在文献<sup>[31]</sup>里介绍。这里我们要强调的是，Guidance的视觉测量采用了文献<sup>[32]</sup>介绍的惯性传感辅助空间测量的混合版。当空间测量失效时Guidance可以使用采用惯性传感辅助单目视觉测量算法<sup>[32]</sup>。

图3为视觉测量的算法流程图，引用自文献<sup>[32]</sup>。我们改进了“优化匹配质量”、“2D-2D运动估计”和“度量尺度滤波”等模块。

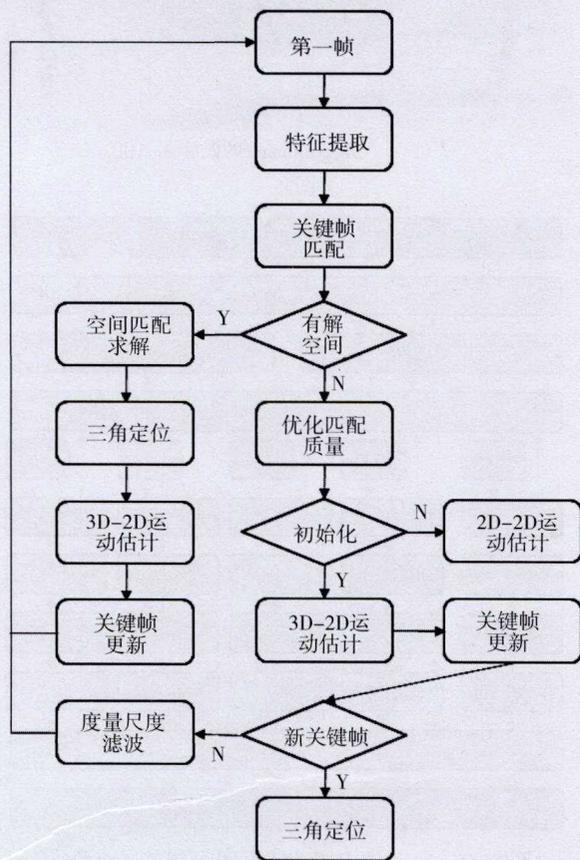


图3 视觉测量的流程图

### 2.2.1 优化匹配

由于单目相机不能像多目相机一样直接标示外部参数，其深度图像的质量也完全依赖于相机运动拍摄的质量。因而采用FAST特征检测器和BRIEF特征描述器提取出相匹配的特征，最后采用the non-pyramids Lucas-Kanade Tracker<sup>[19]</sup>优化匹配结果。

### 2.2.2 2D-2D运动估计

在单目视觉测量的初始化阶段，用2D-2D算法进行运动估计。每一对2D-2D算式可以用  $C_i = \{u_i, u'_i\}$  来表示，这里  $u = [u, v, 1]^T$  为无畸变图像的像素坐标。考虑到对极几何学的约束  $(K^{-1}u'_i)^T [t] \times R(K^{-1}u_i) = 0$ ，其中  $R$  直接从惯性传感器中获取， $K$  是相机的内部参数，令  $(K^{-1}u'_i)^T = [x_i, y_i, 1]$  和  $R(K^{-1}u_i) = [a_i, b_i, c_i]^T$ 。由于不能从单目视觉中获得绝对的尺度因子，所以将第二项平移因子  $t_y = 1$  归一化。平移因子  $t$  的其余两项的最小解可由下面的式子求得：

$$\begin{bmatrix} t_x \\ t_z \end{bmatrix} = \begin{bmatrix} b_1x_1 - a_1y_1 & c_1y_1 - b_1 \\ b_2x_2 - a_2y_2 & c_2y_2 - b_2 \end{bmatrix}^{-1} \begin{bmatrix} x_1c_1 - a_1 \\ x_2c_2 - a_2 \end{bmatrix} \quad (1)$$

### 2.2.3 度量尺度滤波器 (Metric Scale Filter)

使用扩展的卡尔曼滤波器 (由Weiss的文献<sup>[15, 29]</sup>改进) 获得视觉传感与惯性传感间的度量尺度。应重点关注的是：实际使用中视觉测量的结果有时候是无效的，因此对视觉测量结果赋予的权重需认真处理。定义状态空间向量  $x$  如下：

$$x = \begin{bmatrix} v_w \\ b \\ \lambda \end{bmatrix} \quad (2)$$

这里  $v_w$  表示空间坐标系里的速度， $b$  是加速度的有偏估计， $\lambda$  是单目测量的尺度因子。

惯性传感器的输出状态向量如下：

$$x_k = \begin{bmatrix} I & -\Delta t R & 0 \\ 0 & I & 0 \\ 0 & 0 & 1 \end{bmatrix} x_{k-1} + \begin{bmatrix} -\Delta t R a \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} R & -\Delta t R \\ 0 & I \\ 0 & 0 \end{bmatrix} \begin{bmatrix} n_a \\ n_b \end{bmatrix} \quad (3)$$

式中， $\Delta t$  是惯性传感的采样时间间隔， $R$  是间隔时间内的相对角速度， $a$  直接从惯性传感的加速

度中读取,  $n_a$  和  $n_b$  是读取加速度数据时应当考虑的高斯噪声的有偏估计。

视觉传感器的输出状态向量如下:

$$\begin{aligned} z &= \begin{bmatrix} v_{mono} \\ v_{stereo} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} \lambda R^T & 0 & \frac{1}{2} R^T v_w \\ R^T & 0 & 0 \end{bmatrix} x + n_z \end{aligned} \quad (4)$$

这里,  $v_{mono}$ ,  $v_{stereo}$  分别为单目和多目视觉测量,  $n_z$  表示为高斯噪声。

### 3 软件开发套件 (SDK)

调用Guidance的SDK可以由USB2.0接口获取各个传感器的原始数据和预处理数据。该SDK支持ROS<sup>[3]</sup>, OpenCV<sup>[2]</sup>和MAVlink<sup>[1]</sup>。同时, 例程和WIKI说明可以从<http://dev.dji.com>获取。

表格1罗列了从Guidance的SDK获取的数据详细信息。需要注意的是, 数据描述会随着版本更新而改变。为获取最新的信息, 软件开发者应随时关注官方网站。在官网除了可以获得如下大量详实的数据, 开发者还可以享受SDK带来的如下便利:

在带宽内自定义USB数据流, 即开发者可以调节帧率和通道的数量;

表1 Guidance SDK提供的的数据

数据类型	描述
图像 (10通道)	QVGA 分辨率 8位灰度 不失真校正 采样频率20Hz以上
深度图像 (2通道)	QVGA 分辨率 16位深度 OpenCV BM算法 采样频率20Hz以上
障碍物距离 (5通道)	0.1-20m 采样频率20Hz
超声波 (5通道)	0.1-8m 采样频率20Hz
惯性传感器	3轴陀螺仪 3轴加速度计 与图像同步 采样频率20Hz
视觉测量	当量速度 采用频率 10Hz

低频数据 (图像数据和深度图像除外) 通过UART传输, 因此Guidance可以非常方便地应用在现有的机器人上;

使用板载自标定模块或PC上的图像界面标定软件, 可以很方便地标定摄像机参数。

改变摄像机的曝光时间 (自动模式或者手动模式)。

### 4 应用

本节, 我们通过几个实例展示无人机玩家和机器人开发者是如何使用Guidance的。

#### 4.1 自主导航

通过内置的空间/单目视觉测量和多方向观测模块, Guidance可以实现可靠的位置控制和避障功能。这两个功能是自主导航的基础。

图4阐述了Guidance在开阔室外环境中的工作表现。飞行的时间为109s, 飞行距离大约300m, 高度大于20m。单目模式只在极低和极高的高度时可用, 这是因为在这两种情况下多目模式无法找到合适的基线。

假设GPS提供的是真值, 则可用计算出Guidance给出的速度的期望值和标准偏差分别为:

$$m_e = [0.0785, 0.0767, 0.0822]^T \text{m/s}$$

和:

$$\sigma_e = [0.0722, 0.0767, 0.0822]^T \text{m/s}$$

表2为各个算法模块的运行时间。

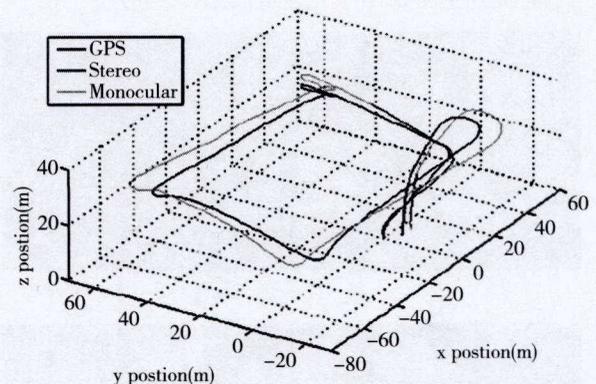


图4 Guidance视觉测量在开阔环境下的应用: 蓝色、红色和粉红色的轨迹线分别表示GPS的位置读数、空间视觉测量的读数和单目视觉测量的读数。注意到, 单目的方法仅在极低或极高的高度时才用, 因此粉红色的轨迹线仅出现一部分。

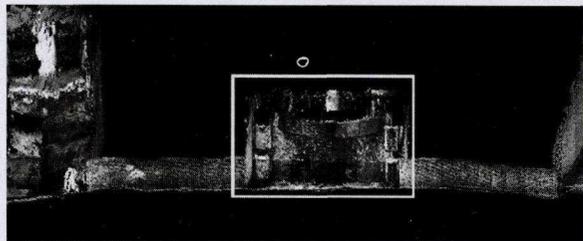
表2 算法模块运行时长

算法模块	运行时长
Re-mapping	N/A
FAST	N/A
BRIEF	N/A
Pixel-wise Pipeline	2.56ms
Block-based Stereo Match	11.14ms
Feature Matching	2.73ms
Matching Refinement	4.44ms
3D-2D Motion Estimation	1.61ms
2D-2D Motion Estimation	1.09ms
Triangulation	2.11ms
Multi-sensor Fusion	1.27ms
Visual Mapping	8.38ms
总运行时间	47.34ms

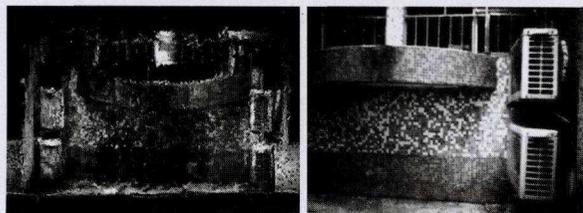
#### 4.2 定位与地图构建 (SLAM)

实现SLAM是智能机器人的一个最重要的任务。Guidance可以为此提供精确的VO结果以及经过校正的图像。所以，VO结果可以方便地使用扩展卡尔曼滤波<sup>[7]</sup>或者光束平差法<sup>[14]</sup>来构建全局地图。如果使用单目SLAM方法，则VO可以直接由校正后的图像计算出来。随后，使用全局最优姿态估计算法，比如 $g^2o$ <sup>[16]</sup>，构建出全局地图。

由于通过内置的匹配模块可以很方便获取到深度信息，Guidance提供更为可靠和精确的RGBD SLAM<sup>[8,9]</sup>。如图5所示，RGBD SLAM在典型室外场景下的性能表现。图5 (b)为SLAM的结果，其很真实地反映出了实际场景图5 (c)。



(a) 全局景深图像



(b) 局部景深图像

(c) 左相机获取的图像

图5 由Guidance生成的室外环境的RGBD SLAM结果

#### 4.3 基于深度的跟踪

Guidance可以提供深度图像，它可以为许多实际应用提供额外的有效信息。为了不失普遍性，接下来将讨论基于Guidance深度图像的目标跟踪应用例子。

我们使用了CAMshift跟踪算法<sup>[4]</sup>作为例子。该算法使用颜色直方图进行目标建模，然后不断地搜索该模型在全局颜色直方图中的分布。由于Guidance使用的是灰度摄像机，所以不能直接从照片中获取颜色信息。因此我们采用CAMshift算法，根据两通道灰度照片获得伪彩色图像。用极具挑战性的室内环境的图像时间序列，评估Guidance的基于深度的物体跟踪功能。室内环境饱和的照明使得物体成像像素在图像时间序列内几乎一致，这有可能导致传统的完全基于灰度图像的跟踪算法失效。如图6所示，使用深度图像可以获得更为精确和鲁棒的物体跟踪算法。

深度示意图灰度图像

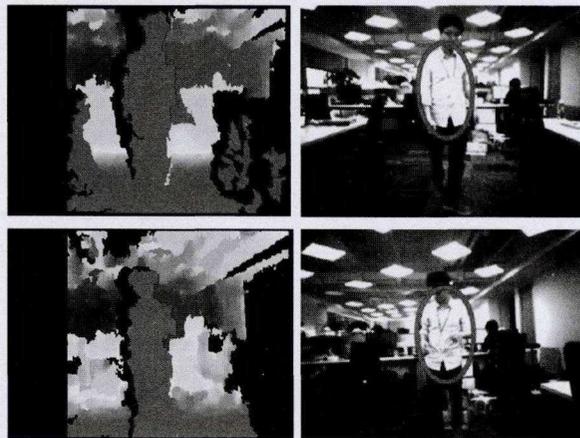


图6 基于深度的物体跟踪算法示意

#### 5 结论

本文介绍了Guidance，一个强大的机器人应用视觉传感平台。它有5个以上的空间传感单元和一个中央处理器。ARM+FPGA架构为实时运算内置算法提供了足够的运算资源，这些内置的算法包括：视觉测量、避障和深度图像生成。Guidance为自主导航，SLAM，物体跟踪等应用提供了一个柔性的开发平台。测试证明，使用Guidance将使一些原本很复杂的机器视觉基础问题变得简单化和易于操作。

## 参考文献

- [1] MAVLink: Micro Air Vehicle Communication Protocol. <http://qgroundcontrol.org/mavlink/start>.
- [2] OpenCV: Open Source Computer Vision Library. <http://opencv.org/>
- [3] ROS: Robot Operating System. <http://www.ros.org/>.
- [4] G. Bradsky. Computer vision face tracking for use in a perceptual user interface. Intel Technology Journal, (Q2), 1998.
- [5] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, and P. Fua. BRIEF: Computing a Local Binary Descriptor Very Fast. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(7):1281–1298, 2012.
- [6] J. Civera, A. Davison, and J. Montiel. Inverse depth parametrization for monocular slam. Robotics, IEEE Transactions on, 24(5):932–945, Oct 2008.
- [7] A. Davison. Real-time simultaneous localisation and mapping with a single camera. In Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on, pages 1403–1410 vol. 2, Oct 2003.
- [8] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, and W. Burgard. An evaluation of the RGB-D SLAM system. In Robotics and Automation (ICRA), 2012 IEEE International Conference on, pages 1691–1696, May 2012.
- [9] F. Endres, J. Hess, J. Sturm, D. Cremers, and W. Burgard. 3-D mapping with an RGB-D camera. Robotics, IEEE Transactions on, 30(1):177–187, Feb 2014.
- [10] J. Engel, T. Schops, and D. Cremers. LSD-SLAM: Large scale direct monocular SLAM. In ECCV, 2014.
- [11] C. Forster, M. Pizzoli, and D. Scaramuzza. SVO: Fast semidirect monocular visual odometry. In ICRA, 2014.
- [12] D. Honegger, L. Meier, P. Tanskanen, and M. Pollefeys. An open source and open hardware embedded metric optical flow cmos camera for indoor and outdoor applications. In Robotics and Automation (ICRA), IEEE International Conference on, pages 1736–1741, May 2013.
- [13] A. Hornung, K. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard. Octomap: an efficient probabilistic 3D mapping framework based on octrees. Autonomous Robots, 34(3):189–206, 2013.
- [14] G. Klein and D. Murray. Parallel tracking and mapping for small AR workspaces. In Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR), 2007.
- [15] L. Kneip, S. Weiss, and R. Siegwart. Deterministic initialization of metric state estimation filters for loosely-coupled monocular vision-inertial systems. In Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on, pages 2235–2241, Sept 2011.
- [16] R. Kummerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard. g<sup>2</sup>o: A general framework for graph optimization optimization. In Robotics and Automation (ICRA), IEEE International Conference on, pages 3607–3613, 2011.
- [17] T. Lemaire. Vision-based SLAM: Stereo and monocular approaches. International Journal of Computer Vision, 74(3):343–364, 2007.
- [18] A. Levin and R. Szeliski. Visual odometry and map correlation. In Computer Vision and Pattern Recognition, CVPR, volume 1, pages 1–611–1–618 Vol. 1, 2004.
- [19] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In Proceedings of the 7th International Joint Conference on Artificial Intelligence – Volume 2, pages 674–679, 1981.
- [20] D. Magree, J. Mooney, and E. Johnson. Monocular visual mapping for obstacle avoidance on uavs. Journal of Intelligent & Robotic Systems, 74(1–2):17–26, 2014.
- [21] T. Naseer, J. Sturm, and D. Cremers. Followme: Person following and gesture recognition with a quadcopter. In Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on, pages 624–630, Nov 2013.
- [22] D. Nister, O. Naroditsky, and J. Bergen. Visual odometry. In Computer Vision and Pattern Recognition, CVPR, volume 1, pages 1–652–1–659 Vol. 1, 2004.
- [23] G. N. utzi, S. Weiss, D. Scaramuzza, and R. Siegwart. Fusion of imu and vision for absolute scale estimation in monocular slam. Journal of Intelligent & Robotic Systems, 61(1–4):287–299, 2011.
- [24] J. Pestana, J. L. Sanchez-Lopez, P. Campoy, and S. Saripalli. Vision based gps-denied object tracking and following for unmanned aerial vehicles. In Safety, Security, and Rescue Robotics (SSRR), IEEE International Symposium on, pages 1–6, 2013.
- [25] E. Rosten and T. Drummond. Fusing points and lines for high performance tracking. In IEEE International Conference on Computer Vision, volume 2, pages 1508–1511, 2005.
- [26] E. Rosten and T. Drummond. Machine learning for high speed corner detection. In European Conference on Computer Vision, volume 1, pages 430–443, May 2006.
- [27] Skybotix. VI-Sensor. <http://www.skybotix.com/>.
- [28] C. Teuliere, L. Eck, and E. Marchand. Chasing a moving target from a flying uav. In Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on, pages 4929–4934, Sept 2011.
- [29] S. Weiss, M. W. Achtelik, S. Lynen, M. Chli, and R. Siegwart. Real-time onboard visual-inertial state estimation and self-calibration of mavs in unknown environments. In Robotics and Automation (ICRA), IEEE International Conference on, pages 957–964, 2012.
- [30] Y. Wu, J. Lim, and M.-H. Yang. Online object tracking: A benchmark. In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, pages 2411–2418, June 2013. 1
- [31] G. Zhou, A. Liu, K. Yang, T. Wang, and Z. Li. An embedded solution to visual mapping for consumer drones. In Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE Conference on, pages 670–675, 2014. 2
- [32] G. Zhou, J. Ye, W. Ren, T. Wang, and Z. Li. On-board inertial-assisted visual odometer on an embedded system. In ICRA, 2014. 2