

中国科学技术大学

学士学位论文



面向移动服务机器人的行人追踪跟随的 研究与实现

作者姓名： 韦清
学科专业： 计算机科学与技术
导师姓名： 陈小平 教授
完成时间： 二〇一九年五月十二日

University of Science and Technology of China
A dissertation for bachelor's degree



Research and Implementation of People Tracking and Following with Mobile Service Robots

Author: Qing Wei

Speciality: Computer Science and Technology

Supervisor: Prof. Xiaoping Chen

Finished time: May 12, 2019

目 录

中文内容摘要	2
英文内容摘要	3
第 1 章 绪论	4
1.1 研究背景及意义	4
1.2 相关工作	5
1.2.1 基于视觉信息的行人检测与追踪	5
1.2.2 基于激光的行人检测与追踪	6
第 2 章 行人追踪算法	7
2.1 视觉行人追踪	7
2.1.1 基于检测的追踪	7
2.1.2 基于动态的追踪	16
2.2 激光行人追踪	18
参考文献	19
附录 A 补充材料	20
在读期间发表的学术论文与取得的研究成果	22

中文内容摘要

服务机器人是一种半自助或全自主工作的机器人。它能完成有益于人类健康的服务工作，但不包括从事生产的设备。服务机器人分为个人/家庭服务机器人和专业机器人。专业机器人一般在特定场景中使用，如商业服务、物流、医疗、救援等；而个人/家用服务机器人主要在日常生活场景中进行与人进行交互，提供家政服务、陪伴、娱乐、辅助学习等多种功能，包括家政机器人、娱乐休闲机器人、助老助残机器人等。其中，个人/家庭服务机器人为本文研究内容所适用的对象。

为了精准理解当前环境和有效执行指令，能够精确可靠地自动识别目标人物并对其进行追踪陪同，是移动服务机器人的人机交互中的一项重要且必要的功能。

本文将针对室内移动机器人的行人跟随问题做如下研究：

- (1) 常用目标跟随算法的原理与实现；
- (2) ROS 导航和可佳导航介绍；
- (3) 可佳机器人上行人跟随系统的实现。

关键词：计算机视觉；机器人；目标追踪；路径规划

Abstract

A service robot is a robot which operates semi- or fully autonomously to perform services useful to the well-being of humans and equipment, they exclude manufacturing operations, and they are capable of making decisions and acting autonomously in real and unpredictable environments to accomplish determined tasks. There are two types of service robots, personal/domestic service robots and professional robots. Professional robots are typically used in specific occasions, including business, delivery, medical, rescue, etc. Personal/domestic service robots, which include cleaning robots, elder care and medical companions, entertainment and leisure robots, home education and training robots, are the specific research objects in this paper.

In order to accurately understand the current environment and effectively execute the instructions, one of the important and necessary ability for a personal service robot, is to automatically recognize and track a person precisely and robustly.

This paper is consist of three parts:

- (1) The theories and implementation of various existing object tracking algorithms;
- (2) Introduction of ROS navigation and keaja navigation;
- (3) The implementation of the people following system on keaja robot.

Key Words: Computer Vision; Robotics; Object Tracking; Path Planning

第 1 章 绪论

1.1 研究背景及意义

服务机器人是一种半自助或全自主工作的机器人。它能完成有益于人类健康的服务工作，但不包括从事生产的设备。它们可以在真实且不可预测的环境中自动进行决策和行动来完成确定的任务。

服务机器人分为个人/家庭服务机器人和专业机器人。专业机器人一般在特定场景中使用，如商业服务、物流、医疗、救援等；而个人/家用服务机器人主要在日常生活场景中进行与人进行交互，提供家政服务、陪伴、娱乐、辅助学习等多种功能，包括家政机器人、娱乐休闲机器人、助老助残机器人等。其中，个人/家庭服务机器人为本文研究内容所适用的对象。

为了精准理解当前环境和有效执行指令，能够精确可靠地自动识别目标人物并对其进行追踪陪同，是移动服务机器人的人机交互中的一项重要且必要的功能。

移动机器人的核心技术包括导航定位、地图创建、路径规划、任务分配和目标跟踪等。移动机器人的智能指标包括三个方面：自主型、适应性和交互性。

国际机器人联合会 (International Federation of Robotics) 对服务机器人做了如下定义：

服务机器人是一种半自助或全自主工作的机器人。它能完成有益于人类和设备的服务工作，但不包括从事生产的设备。

服务机器人通常也是移动机器人。

从 20 世纪 80 年代中期开始，机器人已从工厂的结构化环境进入人的日常生活环境——医院、办公室、家庭和其他杂乱及不可控环境，成为不仅能自主完成工作，而且能与人共同协作完成任务或在人的指导下完成任务的智能服务机器人。

20 世纪 90 年代末，世界服务机器人协会 (International Service Robot Association) 才第一次定义了服务机器人的概念：能够进行感知、思考和行动，并以此来有益于和扩展人类的能力和人类的生产效率的机器。

服务机器人被看重的就是交互能力

随着人工智能与物联网技术不断发展，服务机器人作为智能硬件之一，不断地丰富其自身功能及其实现更强大的性能。在技术层面，我国服务机器人与国外

相比,仍存在较大差距比如在机器人基础算法、核心软件、人工智能硬件落地等方面存在短板。国内服务机器人总体尚处于初级发展阶段,半数以上的产品处于研发试验阶段,但其增长速度较快。根据相关数据显示,2017 年全年我国的整体机器人规模市场达到 1200 亿元,其中服务机器人占据 28% 的市场,服务机器人增长率明显高于工业机器人的发展。

现今,随着物联网的发展和人们对智能化的要求,在日常生活中协助或娱乐人类的个人服务机器人市场正在迅速发展。2017 年,个人/家庭服务机器人市场价值增长了 27%,达到 21 亿美元;总数增加了 25%,达到约 850 万台。据估计,近 610 万台机器人被用于家庭工作。

个人和国内应用中的机器人技术经历了强劲的全球增长,地板清洁机器人,机器人和用于家庭教育的机器人(后者越来越多地被称为社交机器人)越来越成为人们生活的一部分。未来的产品愿景指向具有更高复杂性,能力和价值的家用机器人,例如用于支持老年人的辅助机器人,帮助做家务和娱乐。

1. 随着人口老龄化趋势的加重,服务机器人市场迎来了爆发增长期。家庭用机器人,智能公共服务机器人应用场景和服务模式不断拓展。随着人类寿命的演唱,老龄化趋势的加重,给医疗健康机构带来越来越大的压力,养老问题兔罾家明显,社会对老年人护理的需求大大增加。智能养老设备,如智能服务机器人,的出现极大的弥补了由老年人口激增,护工、养老机构等养老资源匮乏所带来的养老服务供需缺口。此处可引用一段华为项目比赛的简介?)

2. 行人跟随是智能服务机器人的人机交互中的一项重要技术。它要求机器人能够准确识别指定目标,通过对目标的跟随来保证更好地完成人机交互,同时,在移动过程中强调安全。

1.2 相关工作

Literature 中已经有很多 following 相关的研究工作。常用的行人追踪方法分为基于视觉信息的行人检测追踪、基于激光信息的行人追踪,以及多传感器融合的方法。

1.2.1 基于视觉信息的行人检测与追踪

基于检测的追踪和基于追踪的算法:<https://zhuanlan.zhihu.com/p/32826719>

大部分方法使用了粒子滤波 (particle filters),多假设追踪 (multiple hypothesis

tracking), 卡尔曼滤波 (Kalman filters) 的方法

[1]

1. 基于人脸识别的跟踪

(2) 人脸识别的速度和正确率均已达到一个很高的层次，但在实际的激动机器人跟随场景中，人不是一直面对移动机器人。

(3) 基于模板匹配的跟踪

(4) 基于轮廓信息的跟踪

1.2.2 基于激光的行人检测与追踪

(1) 使用几何特征识别目标

(2) 基于运动检测识别目标

第 2 章 行人追踪算法

行人追踪算法 blabla

2.1 视觉行人追踪

计算机视觉中的行人追踪，主要包括密集跟踪方法，即基于行人检测和识别的追踪，以及稀疏跟踪方法，即基于目标动态的追踪。

在密集跟踪方法中，我们实际上并没有“跟踪”物体，而是在视频不同的时间点的一系列帧上扫描和检测物体的位置。由于每次的目标检测都是独立地在当前帧上进行的，所以每次检测时，都需要处理图像中的所有像素，所以以这种方法进行目标跟踪，计算量会比较大。

稀疏跟踪方法是根据物体的动态信息，对其可能的运动轨迹进行预测，并结合其上一帧所在位置和对当前帧的观察，得出其当前位置的算法。由于已知物体在上一帧时的位置，所以对当前帧识别时，只需要检测上一帧物体所在位置附近的像素，这样一来，相对于密集跟踪方法，就减少了大量的计算。此外，由于我们结合了对物体运动的预测和观察来进行估计，在一些情况下准确度也会较高，但在物体速度较快时，可能会失去对物体的追踪，当目标物暂时从视野中消失时，可能难以重新找回物体。

2.1.1 基于检测的追踪

基于机器学习的方法是现阶段行人检测算法的主流，下面将介绍常用行人检测方法原理。

1. 人工特征 + 分类器

(1) 常用特征描述子

特征描述子是一种对图片的表示方法，它通过提取图片中的关键信息并丢弃多余信息来对图片信息进行简化。通常地，特征描述子将一个 RGB 三通道的图片转化成一个特征向量。

为了做到精确地进行图像识别、目标检测，我们必须首先明确什么是关键的、有用的信息，什么是冗余信息。

a. 颜色直方图

颜色特征具有旋转不变性，且不受目标的大小和形状的变化影响，在颜色空

间中分布大致相同，从而具有较高的鲁棒性。

颜色直方图是描述颜色特征最常用的描述子，它是对目标表面颜色分布的统计，描述了不同色彩在图像中所占的比例，但无法描述图像中颜色的局部分布及每种色彩所处的空间位置，即无法描述图像中的某一具体的对象或物体。颜色直方图具有稳定性好、抗部分遮挡、计算方法简单和计算量小的特点。

颜色直方图可以基于不同的颜色空间，其中，最常用的是 RGB 空间和 HSV 空间。以 RGB 空间为例，分别统计每个像素的 R、G、B 数值落在 $[0, 255]$ 上每个点的频度，绘制出直方图，该图片的颜色特征即为三个长为 256 的向量，分别表示其的红色、绿色、蓝色的统计分布。以一张人物照片 2.1 为例，图 2.2 为其 RGB 颜色直方图。

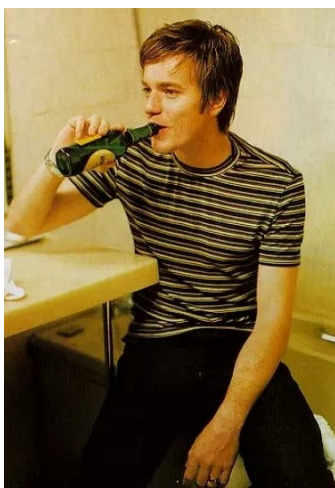


图 2.1 原图

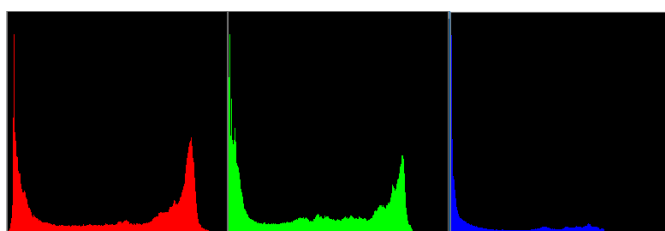


图 2.2 左：B 通道颜色直方图；中：G 通道颜色直方图；右：R 通道颜色直方图

但 RGB 颜色空间的均匀性非常差，且两种颜色之间的视觉差异色差不能表示为改颜色空间中两点间的距离，RGB 这三种颜色的分量的取值与所生成的颜色之间的联系并不直观。

在计算机视觉中，我们常采用 HSV 颜色空间来表示颜色。HSV 是一种将 RGB 色彩空间中的点在圆柱坐标系中的表示方法，相对于 RGB，它能够更加直观地表示色彩的明暗、色调以及鲜艳程度，方便进行颜色之间的对比。此外，由

于 HSV 单独提取了颜色的明暗，也可以一定程度上抵抗光照明暗带来的影响。Sural et al.^[2] 的实验显示，使用 HSV 直方图进行行人识别的结果相比 RGB 直方图有了明显提高。

HSV 即色相 (Hue)、饱和度 (Saturation)、亮度 (Value)。色相即表示物体的颜色，如红色、黄色等，在 0° 到 360° 的标准色轮上，按位置度量色相；饱和度是指颜色的强度或纯度，表示色相中灰色分量所占的比例，它使用从 0% (灰色) 至 100% (完全饱和) 的百分比来度量；亮度是颜色的相对明暗程度，通常使用从 0% (黑色) 至 100% (白色) 的百分比来衡量。

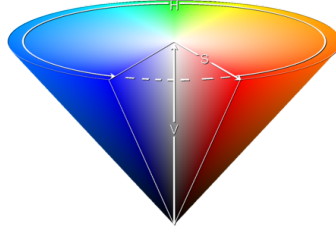


图 2.3 HSV 模型可以使用圆柱坐标系中的一个圆锥形子集表示

由于大部分数字图像都是基于 RGB 空间进行表示的，我们需要首先把 RGB 空间坐标映射到 HSV 空间。给定 (r, g, b) 分别是一个颜色的红、绿、蓝坐标，它们的值是在 0 到 1 之间的实数， max 为 r 、 g 和 b 之中的最大值， min 为其中的最小值，则从 (r, g, b) 到 (h, s, v) 的转换公式如下：^[3]

$$h = \begin{cases} 0^\circ & \text{if } max = min \\ 60^\circ \times \frac{g-b}{max-min} + 0^\circ, & \text{if } max = r \text{ and } g \geq b \\ 60^\circ \times \frac{g-b}{max-min} + 360^\circ, & \text{if } max = r \text{ and } g < b \\ 60^\circ \times \frac{b-r}{max-min} + 120^\circ, & \text{if } max = g \\ 60^\circ \times \frac{r-g}{max-min} + 240^\circ, & \text{if } max = b \end{cases}$$

$$s = \begin{cases} 0, & \text{if } max = 0 \\ \frac{max-min}{max} = 1 - \frac{min}{max}, & \text{otherwise} \end{cases}$$

$$v = max$$

HSV 直方图的计算与 RGB 类似，只是将颜色空间有所差异，我们同样使用图片 2.1 计算其 HSV 直方图，见图 2.4。

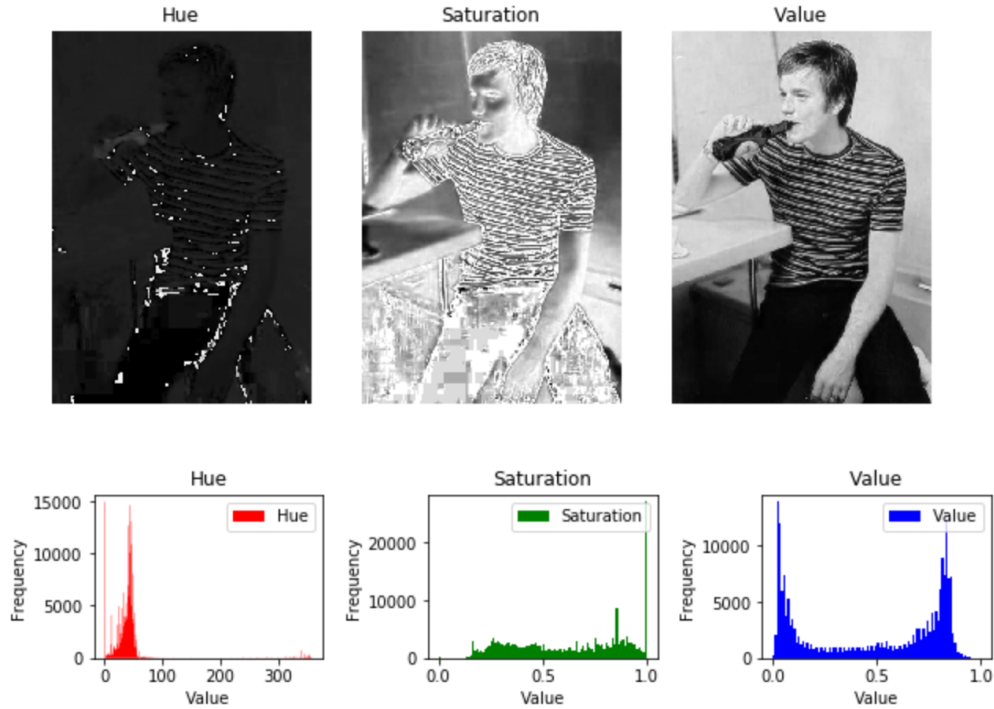


图 2.4 左：H 通道灰度图和颜色直方图；中：S 通道灰度图和颜色直方图；右：V 通道灰度图和颜色直方图

b. 局部二值模式

局部二值模式（Local Binary Pattern, LBP）是一种用来描述图像局部纹理特征的特征描述子，具有旋转不变性和对光照变化不敏感等优点，由 Ojala et al.^[4] 在 1994 年首次提出。

LBP 的计算方法非常简单。每个像素都根据它相邻的八个像素按规定的顺序（如顺时针、逆时针）作比较，来确定其特征值。对于中心像素大于某个相邻像素的，该像素对应的二进制位设置为 1，否则设置为 0，比较了中心点相邻的八个像素后，就得到了一个 8 位的二进制数，这个数字即为该中心像素的特征值，如图 2.5 所示，将每个点的 LBP 值使用灰度图表示，得到 LBP 图谱如 2.6 为例。

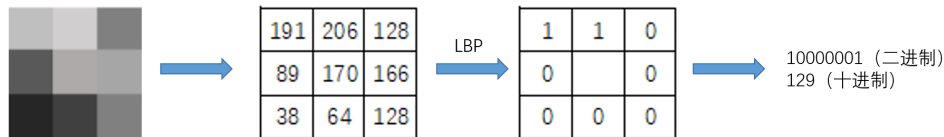


图 2.5 计算 3x3 像素块中中心点的 LBP 值

为了使得 LBP 描述子有旋转不变性，Ojala et al.^[5] 提出了一个 LBP 的具有旋转不变性扩展方法，即不断旋转其邻域，得到一系列的 LBP 值，取其最小值作为该点的局部二值模式值。

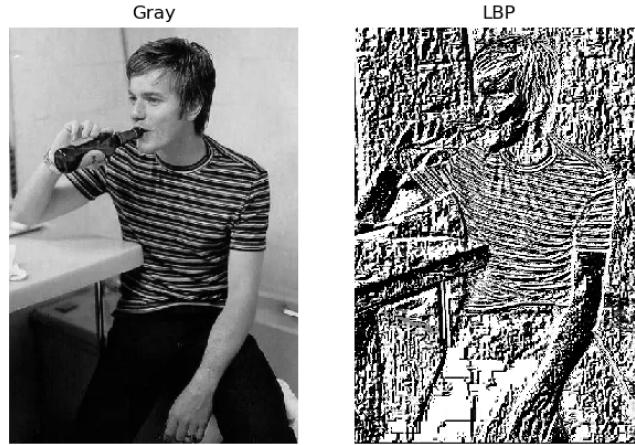


图 2.6 左：灰度图；右：由灰度图计算得到的 LBP 图谱

在计算一个图片的 LBP 描述子时，首先将图片分成固定大小的单元格（如 16×16 像素），在计算出每个像素的 LBP 值后，统计每个单元格内的 LBP 值直方图，再串联所有单元格的直方图，即可得到该图片的 LBP 特征向量。

c. 方向梯度直方图

方向梯度直方图（Histogram of Oriented Gradient, HOG）是目前行人识别中最广泛使用的特征描述子之一。Dalal et al.^[6] 在 2005 年提出 HOG 结合 SVM（支持向量机，support vector machine）进行行人检测的方法，在此之后，该方法被广泛应用到了图像识别中，并尤其在行人检测中获得了巨大的成功，也出现了很多改进和变体。

在 HOG 特征描述符中，它通过计算和统计图像局部区域的梯度方向直方图来构成特征。由于在物体的边缘和角落处图片的颜色会进行突变，故在这些区域，梯度的大小会很大，显然，边缘和角落比起平坦区域包含更多关于物体形状的信息。而通过对边缘和角度的描述，HOG 正可以很好地描述局部目标的表面质地和形状信息。但同时，由于梯度的性质，HOG 特征描述字对噪点比较敏感，且由于 HOG 主要描述了物体的轮廓，所以很难处理遮挡问题。

为了计算方向梯度，我们可以简单地使用内核（Kernel） $[-1, 0, 1]$ 和 $[-1, 0, 1]^T$ 对原图进行过滤，分别得到横向和纵向上的有向梯度。除了这种方法之外，还可以使用 $[-1, 1]$, $[1, -8, 0, 8, -1]$ 和 Sobel 算子等作为内核，不过根据 Dalal et al.^[6] 的实验，使用最简单的 $[-1, 0, 1]$ 进行计算的梯度，在以 HOG 为特征进行的图像识别中效果最佳。

在每个像素处，方向梯度都具有大小和方向。对于彩色图像，我们分别计算 RGB 三个通道的梯度。对原图片上的每个像素点 (x, y) ， $f(x, y)$ 为其 R、G、B 值

中的一个，该通道上的横向和纵向方向梯度为：

$$g_x(x, y) = [-1, 0, 1] * f(x, y) = -f(x + 1, y) + f(x - 1, y),$$

$$g_y(x, y) = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} * f(x, y) = -f(x, y + 1) + f(x, y - 1)$$

梯度大小和方向分别为：

$$|g(x, y)| = \sqrt{g_x(x, y)^2 + g_y(x, y)^2}$$

$$\theta(x, y) = \tan^{-1} \left(\frac{g_y(x, y)}{g_x(x, y)} \right)$$

使用以上公式在 RGB 颜色空间上计算图2.1的梯度值，如图2.7所示。这张梯度图像已经省略了图中很多不必要的信息，如颜色几乎一致的背景，且在同时突出了人物的轮廓。

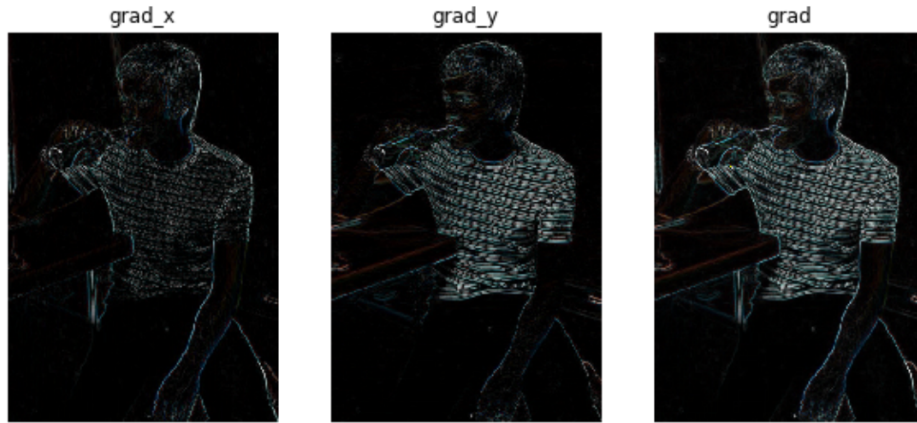


图 2.7 左：横向梯度绝对值；中：纵向梯度绝对值；右：梯度大小

图2.7中包括了 RGB 三个通道在每一个像素点上的梯度值，在计算 HOG 特征向量时，我们选取三个通道的梯度的最大值作为该点处的梯度大小，最大值对应的通道的梯度角度为该点处的梯度方向。

方向梯度直方图统计的实际上是梯度的方向。梯度的方向在 $[0^\circ, 360^\circ)$ 上，但是我们实际在统计方向时，采用的却是 $[0^\circ, 180^\circ)$ 的统计范围，计算 $\theta(x, y) \bmod 180^\circ$ 来代替原有的角度值，即将相差 180° 的两个角度视为同一个梯度方向。实验表明，这种统计方式得到的结果往往比采用 $[0^\circ, 360^\circ)$ 范围的原方向更好^[6]。在统计梯度方向时时，我们还需要使用梯度的大小作为对应方向的权重。

在计算直方图时，我们取 9 个组 (bins)，分别对应 $0^\circ, 20^\circ, 40^\circ, \dots, 160^\circ$ ，若一个像素处的梯度正好为 20 的整数倍，将其梯度大小加到对应的 bin 中；否则，

按照比例，将其加入相邻的两个 bins 中。以图2.8为例。这样一来，HOG 特征描述子即为一个长为 9 的向量，每一个分量的大小对应直方图中相应 bin 的高度。

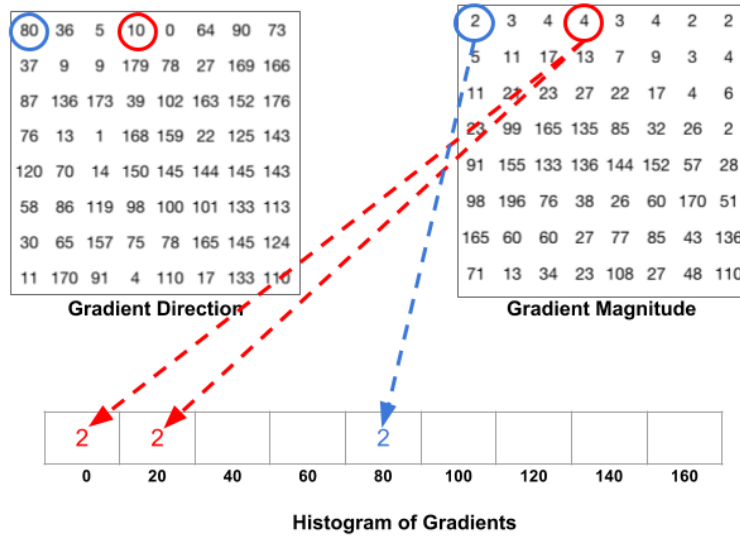


图 2.8 统计梯度方向直方图的方法示意

值得注意的是，由于图像的梯度是由各像素点周围的颜色值大小计算得到的，所以也会受光照的影响，例如，将所有像素值除以 2 来使图像变暗，这时所有梯度大小就也会减半，直方图每个 bin 的高度也会减半。而在一张图片中，每个局部区域的光照可能会有所不同，为了降低这些影响，在进行方向梯度统计时，并不会直接统计一整个图片的方向梯度直方图，而是以 8×8 像素的区域为一个单元格（cell）来分别进行统计，再在此基础上进行规范化（normalization）。这样会降低光照等噪音对特征描述子质量的影响，使 HOG 描述子更加稳定鲁棒。此过程如图2.9所示。



图 2.9 统计单元格梯度直方图的过程示意图

在进行规范化时，最常用的方法是在单元格的基础上取一个更大的块（block），每块的大小为 16×16 像素，即包括 4 个单元格，将每一个块的 4

个 HOG 向量作为一个整体进行 L2 规范化。即 $\mathbf{v} \leftarrow \mathbf{v} / \sqrt{\|\mathbf{v}\|_2^2 + \epsilon^2}$, 其中 ϵ 为一个足够小的正常数。在规范化一个块之后, 我们得到了一个长为 36 的向量, 即这个块最终的 HOG 向量。再以 8 像素的步长 (stride) 移动这个块, 对下一个块进行规范化, 即两个相邻块之间有 2 个单元格的重叠。如此循环, 直到整张图的每一个单元格都被计算, 再将所有经过的块的向量合并在一起, 作为整张图的 HOG 描述子。

d. 尺度不变特征变换

尺度不变特征变换 (Scale-invariant feature transform, SIFT) 是一种不受图片尺度和旋转影响, 并在一定程度上不受光照和相机视角影响的特征描述子, 它将图像数据转换为有关局部特征的尺度不变坐标。同时, 通过在空间和频率域的准确定位, SIFT 还可以减少遮挡和噪音带来的干扰。SIFT 可以使用有效的算法从图片中提取出大量的独特特征, 可以在所有的尺度和位置上密集地覆盖图像, 例如, 一个 500×500 像素的图像可以最多生成 2000 个稳定的特征。SIFT 方法中的关键点描述非常独特, 可以使单个特征从大型特征数据库中能得到高概率的匹配。但在较为杂乱的图像中, 背景中的很多特征可能不会从数据库中得到正确的匹配, 故而在正确的匹配之外, 生成错误的匹配, 不过通过识别关于目标物及其在新图像中的位置, 尺度和方向的关键点的子集, 可以从完整匹配集中过滤出正确的匹配^[7]。

为了最大限度地降低提取特征的成本, 使用级联过滤 (cascade filtering) 方法来检测关键点, 先使用高效的算法来检测出一些候选位置, 然后再进一步详细检查, 将更加耗时的计算只运用到通过了初始测试的候选点位置。

Lowe^[7] 提出的 SIFT 生成图像特征的主要步骤如下:

I. 尺度空间极值检测: 通过高斯差分方程 (difference-of-Gaussian function), 在所有可能的尺度下搜索稳定的特征, 来找出尺度和方向不变的潜在相关点。对于输入图像 $I(x, y)$ 和可变尺度高斯核函数 $G(x, y, \sigma)$, 可以计算出图像的尺度空间 $L(x, y, \sigma)$:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

“*” 为 x, y 上的卷积计算操作, $G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$, σ 为尺度空间因子, 反映了图像被模糊的程度, σ 越大, 对应的尺度也越大。

为了有效地检测出尺度空间中稳定的关键点位置, 使用高斯差分方程来与

图像进行卷积，由一个常数因子 k ，得出 $D(x, y, \sigma)$ ：

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma)$$

为了检测中 $D(x, y, \sigma)$ 函数的局部最大值和最小值，每个点都与它同尺度上的 8 个邻点，以及相邻的两个尺度上的各 9 个邻点相比较（相邻尺度上 $\sigma_{s+1} = k\sigma_s$ ）。只有它的值比相比较的 26 个相邻点都要小或者都要大时，才选取该点作为候选点。

II. 关键点精确定位：对每个候选点上，拟合一个精细的模型来确定其位置和尺度，并根据其稳定性来选择关键点。

对尺度空间函数 $D(x, y, \sigma)$ 进行最多二阶的泰勒展开：

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2} \mathbf{x}$$

其中 D 和其微分在给定候选点处被计算， $\mathbf{x} = (x, y, \sigma)^T$ 是相对于该点的偏移向量。为了求 D 的局部极值的位置，对上式求导并设 D' 为 0，则极值的位置的偏移量 $\hat{\mathbf{x}}$ 和 D 的局部极值点分别为：

$$\hat{\mathbf{x}} = -\frac{\partial^2 D^{-1}}{\partial \mathbf{x}^2} \frac{\partial D}{\partial \mathbf{x}}$$

$$D(\hat{\mathbf{x}}) = D + \frac{1}{2} \frac{\partial D^T}{\partial \mathbf{x}} \hat{\mathbf{x}}$$

为了保证极值点的稳定性，需要剔除低对比度的极值点。若 $|D(\hat{\mathbf{x}})|$ 的值小于 0.03（假定每个像素的值的大小在 $[0, 1]$ 之间），则将该极值点舍弃。

III. 方向赋值：根据局部的图像梯度方向，将一个或多个方向赋值给每个关键点位置。后续在该图像上进行的所有操作都将根据为每个特征所指定的方向、尺度和位置进行变换（transform），从而为这些特征提供方向、尺度不变性。

得到了每个关键点的尺度 σ 后，由特征点为中心，计算出同一尺度下周围区域每个点 (x, y) 的梯度大小 $m(x, y)$ 和方向 $\theta(x, y)$ ：

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$\theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y)))$$

使用直方图统计关键点邻域内像素对应的梯度方向和大小。

IV. 生成特征描述子：校正旋转主方向以确保旋转不变性，生成关键点描述子并进行归一化处理，以去除光照的影响。

SIFT 准确率较高，对于尺度和旋转不敏感，但由于求 SIFT 特征向量计算量较大，耗时较长，所以在需要实时图像识别时并不常被使用。

(2) 算法示例

a. HOG+SVM

b. HOG-LBP

<https://ieeexplore.ieee.org/abstract/document/5459207/>

2. 深度学习

随着深度学习的发展，通过神经网络提取特征得到了广泛的应用。

<https://www.jianshu.com/p/d94e558ebe26>

2.1.2 基于动态的追踪

1. 卡尔曼滤波

卡尔曼滤波是一种假定目标物体的运动服从线性高斯分布，以此对目标的运动状态进行预测，将预测结果与观察模型进行比较，根据误差更新预测模型，估计物体的当前位置的方法。

卡尔曼滤波器是一组提供最小二乘法的有效递归解的数学方程，它可以对于系统的过去、当前、甚至未来状态的进行估计^[8]。

卡尔曼滤波器对离散时间的控制过程的状态 $x \in \mathbf{R}^n$ 进行估计，该过程由以下线性随机差分方程控制：

$$x_{k+1} = A_k x_k + B u_k + w_k$$

同时，提供了对系统当前状态的测量 $z \in \mathbf{R}^m$ ：

$$z_k = H_k x_k + v_k$$

其中，随机变量 w_k 和 v_k 分别表示系统和测量误差，假定它们是互相独立的，并服从正态分布：

$$p(w) \sim N(0, Q),$$

$$p(v) \sim N(0, R).$$

$n \times n$ 的矩阵 A 将系统在时间 k 和 $k+1$ 时的状态相关联起来，不存在驱动函数或系统噪音。 $n \times l$ 的矩阵 B 将控制输入 $u \in \mathbf{R}^l$ 与系统状态 x 相关联。 $m \times n$ 的矩阵 H 将系统状态和对系统的测量 z_k 相关联。

我们根据时间 k 前的过程，计算 $\hat{x}_k^- \in \mathbf{R}^n$ 作为为时间 k 时的先验 (a priori) 状态估计，并根据对系统状态的测量 z_k 计算后验 (a posteriori) 状态估计 $\hat{x}_k \in \mathbf{R}^n$ 。

我们将先验和后验估计误差定义为：

$$e_k^- \equiv x_k - \hat{x}_k^-, e_k \equiv x_k - \hat{x}_k.$$

则先验和后验估计误差协方差分别为：

$$\begin{aligned} P_k^- &= E[e_k^- e_k^{-T}], \\ P_k &= E[e_k e_k^T] \end{aligned}$$

使用先验估计 \hat{x}_k^- 和实际测量 z_k 来计算后验状态估计 \hat{x}_k ：

$$\hat{x}_k = \hat{x}_k^- + \mathbf{K}(z_k - \mathbf{H}_k \hat{x}_k^-)$$

在上式中， $\mathbf{H}_k \hat{x}_k^-$ 是根据先验估计对测量值的预测， $(z_k - \mathbf{H}_k \hat{x}_k^-)$ 被称为测量残差（residual）。残差反映了先验估计及预测方法相对于实际测量的插值。 $n \times m$ 的矩阵 \mathbf{K} 是最小化后验误差协方差的增益（gain）。将上式代入求 P_k 的公式中，取结果相对于 \mathbf{K} 的导数，并设为 0，可以求得：

$$\mathbf{K} = \frac{P_k^- \mathbf{H}_k^T}{\mathbf{H}_k P_k^- \mathbf{H}_k^T + \mathbf{R}_k}$$

由上式可以得出， \mathbf{R}_k 为了测量误差协方差，当它趋于 0 时， \mathbf{K} 越大，特别地：

$$\lim_{\mathbf{R}_k \rightarrow 0} \mathbf{K}_k = \mathbf{H}_k^{-1}$$

另外，当先验估计误差协方差 P_k^- 趋于 0 时， \mathbf{K} 越小，特别地：

$$\lim_{P_k^- \rightarrow 0} \mathbf{K}_k = 0$$

所以，当测量误差协方差 \mathbf{R}_k 约接近 0 时，实际测量 z_k 更加被取信，而预测的测量值 $\mathbf{H}_k \hat{x}_k^-$ 更不被相信，而当先验估计误差协方差 P_k^- 趋近于 0 时则相反。

现在我们就有了根据先验估计和测量计算后验状态估计的方法。后验状态估计反映了状态分布的数学期望，根据 w, v 的分布，状态概率分布也应当满足正态分布，而后验估计误差协方差则反映了状态分布的误差。故：

$$p(x_k | z_k) \sim N(E[x_k], E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T]) = N(\hat{x}_k, P_k)$$

卡尔曼滤波分为两组方程：时间更新方程和测量更新方程。时间更新方程根据当前状态和误差协方差估计，预测下一时间的先验估计；测量更新方程用于根

据所获得的新的测量，再结合先验估计来获取一个已优化的后验估计，这个后验估计又被传回时间更新方程。如此循环，完成一个预测-矫正的过程，以自动化地对模型进行更新，对状态进行估计。

时间更新方程包括：

$$\begin{aligned}\hat{x}_{k+1}^- &= A_k \hat{x}_k + B u_k \\ P_{k+1}^- &= A_k P_k A_k^T + Q_k\end{aligned}$$

测量更新方程包括：

$$\begin{aligned}K_k &= P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \\ \hat{x}_k &= \hat{x}_k^- + K(z_k - H_k \hat{x}_k^-) \\ P_k &= (I - K H_k) P_k^-\end{aligned}$$

Q_k 和 R_k 均为常数，分别与 w 和 v 相关，估计误差协方差 P_k 和增益矩阵 K_k 将会在计算中迅速收敛，并保持不变。

2. 粒子滤波

首先对跟踪目标进行建模，并定义一种相似度度量确定粒子与目标的匹配程度。在目标搜索的过程中，它会先按照一定的分布（比如均匀分布或高斯分布）在全局撒一些粒子，统计这些粒子与目标的相似度，确定目标可能的位置。在可能性较高的位置上，下一帧加入更多新的粒子，确保在更大概率上跟踪上目标。

跟踪速度很快，而且能解决目标的部分遮挡问题。

可以对非线性、非高斯系统的动态进行建模。

粒子滤波器（particle filters）是一种基于概率密度的粒子表示的顺序蒙特卡罗方法 (sequential Monte Carlo methods)，它可以应用在任意状态-空间模型，是传统的卡尔曼滤波的一般化方法^[9]。

3. 点云法

4. MeanShift 和 CamShift

5. 相关滤波

6. 压缩追踪

7. 深度学习

2.2 激光行人追踪

参 考 文 献

- [1] Mucientes M, Burgard W. Multiple hypothesis tracking of clusters of people. *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006:692-697.
- [2] Sural S, Qian G, Pramanik S. Segmentation and histogram generation using the hsv color space for image retrieval. *Proceedings. International Conference on Image Processing*, 2002, 2:II-II.
- [3] Foley J D, Van Dam A, et al. Fundamentals of interactive computer graphics: volume 2. Addison-Wesley Reading, MA, 1982.
- [4] Ojala T, Pietikainen M, Harwood D. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. *Proceedings of 12th International Conference on Pattern Recognition*, 1994, 1:582-585.
- [5] Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2002(7):971-987.
- [6] Dalal N, Triggs B. Histograms of oriented gradients for human detection. *international Conference on computer vision & Pattern Recognition (CVPR'05)*, 2005, 1: 886-893.
- [7] Lowe D G. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 2004, 60(2):91-110.
- [8] Welch G, Bishop G, et al. An introduction to the kalman filter. 1995.
- [9] Arulampalam M S, Maskell S, Gordon N, et al. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on signal processing*, 2002, 50(2):174-188.

附录 A 补充材料

补充内容。

致 谢

致谢应置于目录前。

在研究学习期间，我有幸得到了三位老师的教导，他们是：我的导师，中国科大 XXX 研究员，中科院 X 昆明动物所马老师以及美国犹他大学的 XXX 老师。三位深厚的学术功底，严谨的工作态度和敏锐的科学洞察力使我受益良多。衷心感谢他们多年来给予我的悉心教导和热情帮助。

感谢 XXX 老师在实验方面的指导以及教授的帮助。科大的 XXX 同学和 XXX 同学参与了部分试验工作，在此深表谢意。

在读期间发表的学术论文与取得的研究成果

已发表论文

1. A A A A A A A A A
2. A A A A A A A A A
3. A A A A A A A A A

待发表论文

1. A A A A A A A A A
2. A A A A A A A A A
3. A A A A A A A A A

研究报告

1. A A A A A A A A A
2. A A A A A A A A A
3. A A A A A A A A A