

# 基于 SURF 算法的单目转双目视觉定位<sup>①</sup>

韩 峰

(上海应用技术学院 电气与电子工程学院, 上海 201400)

**摘 要:** 针对三维空间中目标物体定位的问题, 提出了一种结构简单、操作方便、性价比较高的单摄像机实现双目立体视觉定位的方法. 在对目标物体的识别和定位中, 利用各方面性能和指标都比较好的 SURF 算法对所获取的图像进行特征点的提取和匹配. 实验结果表明, 文中使用的基于 SURF 算法的单目转双目视觉定位的方法, 不论是在定位的精度, 还是在时间速度方面都表现出了很好的可行性与实用性, 具有一定的现实利用价值.

**关键词:** 视觉定位; 摄像机; SURF 算法; 特征提取; 匹配

## Monocular to Realize Binocular Visual Positioning Based on SURF Algorithm

HAN Feng

(School of Electrical and Electronic Engineering, Shanghai Institute of Technology, Shanghai 201400, China)

**Abstract:** On the issue of the target object location in three-dimensional space, this paper proposes a method to realize binocular visual positioning with monocular, this method is simple, convenient operation and cost-effective. Because SURF algorithm has better performance and indicators in all respects, so we make use of SURF algorithm to realize the image feature point extraction and matching in the process of object target recognition and locating. The experimental result indicates, the method of monocular to realise binocular visual positioning based on SURF algorithm, both in accuracy of positioning, and in speed of time shows good feasibility and practicability, it has a certain practical value.

**Key words:** vision localization; camera; SURF algorithm; feature extraction; matching

随着智能机器人技术的全面发展, 机器视觉作为其中的一个分支, 人们对它的研究也在不断深入, 双目立体视觉作为机器视觉的一种, 是近年来研究的热点方向之一. 该视觉系统模拟人眼, 利用成像设备获取被测物体的二维图像, 再对图像进行目标物体的识别与匹配, 最后利用视差原理<sup>[1]</sup>计算物体的三维坐标进行定位. 双目视觉相对于单目视觉虽然具有不可比拟的优势, 但双目视觉是由两个摄像机组成的, 结构相对复杂, 造价也较昂贵, 而且在每次不同的使用时, 都要对整个系统重新标定, 操做起来较复杂. 本文提出的单目转双目视觉定位的方法, 是利用一台摄像机, 通过给定方式的运动, 让其在不同的位置来观测空间中已知位置的静止目标物体, 从而获取不同的观测图像, 这样就可以在两次不同的位置图像之间建立对应的约束关系, 从而达到了双目立体视觉定位的功能,

其系统结构简单, 造价较低, 具有一定的现实意义. 通过摄像机获取的图像相对于真实物体往往有缩放、旋转以及遮挡的变化, 这就需要对图像进行特征匹配来识别目标物体. SURF 算法就是一种对提取的特征具有尺度不变、旋转不变、对光照变化和仿射、透视变换也具有部分不变性能的方法, 它在重复度、独特性、鲁棒性等方面, 均超越或接近以往提出的同类方法, 并且在计算速度上具有明显的优势, 本文正是基于 SURF 算法来实现单目转双目视觉定位的.

## 1 SURF 算法原理

SURF 算法是由 Herbert Bay 等人在 SIFT<sup>[2]</sup>算法的基础上提出的一种速度快、鲁棒性好的特征点提取与匹配算法<sup>[3]</sup>. SURF 算法不仅对图像的旋转、缩放、噪声以及平移具有较好的鲁棒性, 而且它在 SIFT 算法的

<sup>①</sup> 收稿时间:2015-03-16;收到修改稿时间:2015-05-04

基础上,使用了近似的 Hessian 矩阵检测兴趣点,并引入了积分图像<sup>[4]</sup>和盒子滤波器,使计算速度比 SIFT 要快 3 倍左右<sup>[5]</sup>,综合性能也优于 SIFT 算法。

### 1.1 特征点提取

SURF 特征点提取,引入了 Hessian 矩阵,通过计算 Hessian 矩阵行列式的局部极值点,来确定特征点的位置.对于图像 I 中的任意点, X 点处  $\sigma$  尺度上的 Hessian 矩阵定义为:

$$H(X, \sigma) = \begin{bmatrix} L_{xx}(X, \sigma) & L_{xy}(X, \sigma) \\ L_{xy}(X, \sigma) & L_{yy}(X, \sigma) \end{bmatrix} \quad (1)$$

其中,  $L_{xx}(X, \sigma)$  表示高斯二阶偏导在 X 处与图像 I 的卷积,  $L_{xy}(X, \sigma)$ 、 $L_{yy}(X, \sigma)$  具有相似的含义. SURF 算法利用盒子滤波器代替高斯二阶微分,并引入了积分图像,极大的提高了运算速度和精度.将盒子滤波器与图像进行卷积运算,其结果分别为  $D_{xx}$ 、 $D_{xy}$  和  $D_{yy}$ . 于是, Hessian 矩阵的行列式可以简化为:

$$\text{Det}(H) = D_{xx}D_{yy} - (wD_{xy})^2 \quad (2)$$

式中  $w$  是权重系数,实际应用中取 0.9 就可以了,按照式(2)中近似 Hessian 矩阵行列式的计算方法,对图像中每一点求取响应并记录下来,这样就得到了在尺度  $\sigma$  上的响应图。

### 1.2 SURF 特征描述子

与 SIFT 算法类似, SURF 特征点描述子描述的仍然是特征点某个小邻域范围内的灰度分布信息, SURF 使用一阶 Haar 小波,把在 x、y 方向的响应用做构建特征向量的分部信息.为了确保特征矢量具有旋转不变性,需要给每个特征点分配一个主方向.为此,在以特征点为圆心,以  $6\sigma$  ( $\sigma$  为特征点的尺度)为半径的圆形邻域内,用尺度为  $4\sigma$  的 Haar 小波模板对图像进行处理,求 x 和 y 两个方向的 Haar 小波响应运算.使用尺度为  $2\sigma$  的高斯加权函数对 Haar 小波的响应值进行高斯加权,离特征点越近,响应越大.用圆心角为  $\pi/3$  的扇形,以特征点为中心旋转一周,计算该扇形每转  $1^\circ$  时,所包含图像点的 Haar 小波响应之和,最大 Haar 响应累加值所对应的方向即为特征点的主方向.沿主方向,将以特征点为中心,边长为  $20\sigma \times 20\sigma$  的图像区域划分为  $4 \times 4$  的子区域.在每个子区域内,用尺寸为  $2\sigma \times 2\sigma$  的 Haar 小波模板进行处理,为了确保几何变换的鲁棒性,同样赋予以特征点为中心的高斯权重系数,然后对响应值进行统计  $\sum dx$ 、 $\sum |dx|$ 、 $\sum dy$ 、 $\sum |dy|$ . 这样,在每个子区域内,就得到了一个 4 维描

述向量:

$$V(\sum dx, \sum dy, \sum |dx|, \sum |dy|) \quad (3)$$

式中,  $dx$ 、 $dy$  分别表示水平方向和垂直方向的 Haar 小波响应.把  $4 \times 4$  个子区域的向量连起来,就会得到  $4 \times 4 \times 4 = 64$  维的特征向量,对特征向量进行归一化处理,就确保了描述子对光照的不变性。

### 1.3 特征点匹配

特征点匹配就是把 A、B 两幅图中具有相同物理位置的点对应起来,匹配的依据是 SURF 算法为每个特征点所构建的描述子向量.我们设  $P_A$  是图 A 中任意一点,  $P_B$  是图 B 中任意一点,它们的描述子向量分别为  $D_A$ 、 $D_B$ .  $D_A(i)$ 、 $D_B(i)$  分别是它们的第  $i$  个分量,则  $P_A$  和  $P_B$  之间的距离为:

$$D(P_A, P_B) = \sqrt{\sum_{i=1}^n (D_A(i) - D_B(i))^2} \quad (4)$$

按公式(4),把 A 图中某点  $P_A(i)$  与 B 图中的所有点之间的距离计算一遍,利用 Lowe 在文献[6]中使用的最近邻 ND(Nearest Distance)比次近邻 NND(Next Nearest Distance)的方法来求匹配点.我们把两者之间的比值表示为  $m = ND/NND$ . 设定阈值  $0 < threshold < 1$ ,若  $m \leq threshold$ ,则判定该点就是要找的匹配点,反之不是.根据文献[6]中的结论,阈值  $threshold$  的取值范围在 0.5~0.7 时,特征点匹配的效果是最理想的.对本文中的图像进行 SURF 特征点匹配时,当  $ND/NND < 0.6$  时,匹配的效果最好,把符合这个条件的最近距离的点,我们确定为最终的匹配特征点.用 SURF 算法所做的图像匹配情况如图 2、图 3 及表 1 所示。

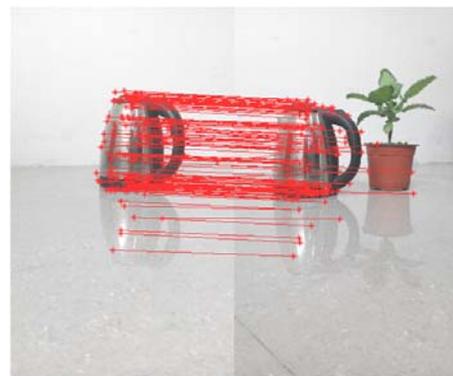


图 2 目标物体与左图像的匹配



图 3 目标物体与右图像的匹配

表 1 SURF 算法对的特征点匹配

SURF 算法	匹配点对	误匹配点对	所用时间
目标物体与左图像的匹配	250	13	1.433s
目标物体与右图像的匹配	250	15	1.386s

从匹配图和表 1 的统计结果可以看出, 用 SURF 算法进行图像匹配时, 在时间方面, 整个过程所花费的时间很短, 能满足实时性的要求. 在准确度方面, 大多数的匹配都是准确可靠的. 但由于客观原因的存在, 匹配过程中仍存在误匹配的现象. 当误配点对的数目影响到测量精度时, 就要采取一定的措施, 比如, 在保证能够解决问题的前提下, 不要设置过多的匹配点对, 也可通过加入剔除误配点对的算法, 来消除误匹配的影响.

## 2 双目立体视觉成像原理

双目立体视觉成像原理, 最早是由美国麻省理工学院的 Marr 博士提出的<sup>[7,8]</sup>, 它是模拟人的双眼, 从不同的位置来获取目标物体的两幅图像, 然后通过计算图像对应点间的位置偏差, 即利用视差原理<sup>[9]</sup>来获取物体三维几何信息的方法模型.

图 4 所示即为简单的平视双目立体成像原理图, 两摄像机投影中心的连线距离即为基线距  $b$ . 摄像机坐标系的原点在摄像机镜头的光心处. 事实上, 摄像机的成像平面是在镜头光心后的, 上图中我们将成像平面绘制在镜头的光心前  $f$  处, 虚拟的图像平面坐标系  $O_1uv$ 、 $O_2uv$  的  $u$  轴和  $v$  轴与摄像机坐标系的  $x$  轴和  $y$  轴方向一致, 这样可以简化计算过程. 左右图像坐标系的原点是摄像机光轴与平面的交点  $O_1$  和  $O_2$ . 空间中某点  $P$  在左图像和右图像中相应的坐标分别为

$p_1(u_1, v_1)$  和  $p_2(u_2, v_2)$ . 假定两摄像机的图像在同一平面上, 则点  $P$  图像坐标的  $Y$  坐标相同, 即  $v_1 = v_2$ , 由三角几何关系得到:

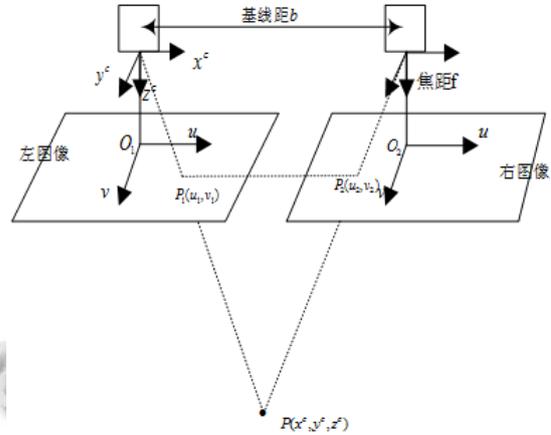


图 4 双目立体视觉成像原理图

$$u_1 = f \frac{x^c}{z^c} \quad u_2 = f \frac{(x^c - b)}{z^c} \quad v_1 = v_2 = f \frac{y^c}{z^c} \quad (5)$$

式中  $(x^c, y^c, z^c)$  为点  $P$  在左摄像机坐标系中的坐标,  $b$  为基线距,  $f$  为两摄像机的焦距. 视差为点  $P$  在两图像中相应点的位置差:

$$d = (u_1 - u_2) = \frac{f * b}{z^c} \quad (6)$$

由此根据相似三角形的关系, 可计算出空间中点  $P$  在左摄像机坐标系中的坐标为:

$$x^c = \frac{b * u_1}{d} \quad y^c = \frac{b * v}{d} \quad z^c = \frac{b * f}{d} \quad (7)$$

其中  $z^c = \frac{b * f}{d}$  就是目标点  $P$  的深度距离. 因此, 只要能够找到空间中某点在左右两个摄像机像面上的相应点, 并且通过摄像机标定获得摄像机的内外参数, 就可以确定这个点的三维

## 3 单目转双目立体视觉定位的测定

一般的双目立体视觉测距, 是用两台规格和型号都相同的摄像机, 对称的放置在固定的位置上, 来对被测物体进行拍摄, 以得到目标物体多幅不同的图像. 本次单目转双目试验, 利用的测距方法是运动立体视觉测距法<sup>[10]</sup>, 即利用一台摄像机沿着一条设定的基准线, 相对被测物体以给定方式运动, 这样就可以得到一系列不同的图像, 这种通过相对运动产生立体视觉的方法就称为运动立体测距法, 它是一种用一个成像传感器在不同时间和不同的空间位置来获取一系列物

体图像的方法,也即是一种用时间换取空间的方法.这种方法使得测量系统结构简单,操作方面,更重要的是具有很高的性价比,在控制成本方面有很大的优势.

这里,我们采用运动立体测距法最简单的情况来操作,让摄像机沿着固定的基线相对于目标物体作转动,在两次不同的位置上对目标物体进行拍摄,以得到两幅不同的图像,这两幅图像就可以看作是双目立体视觉的左右图像.试验中采用的摄像机是普通的针孔摄像机,焦距是 35mm,采集图像的尺寸为 900×1600 像素,水平和垂直分辨率是 96dpi.基线是以目标物体水壶为圆心、以设定的长度为半径的同心圆上不同两点之间的距离,为了实现双目测量,这两点之间的距离不可以太大.试验中选取的两点,基线  $b$  为 300mm,在这两点处对目标物体进行拍照,然后在 MATLAB8.2 中利用 SURF 算法对两幅图中的目标物体进行匹配,所得的仿真结果如图 5、图 6.



图 5 目标物体与左图像匹配结果



图 6 目标物体与右图像匹配结果

图中的黄花点是利用 SURF 算法匹配出的目标物体各个特征点坐标的平均值所在的点,黄花点就可以看做目标物体的坐标点.由仿真结果可以得出目标物体在左右图像中的坐标值分别为 (1325.7,765.3)、

(1345.1,764.9),根据公式(6),(7)可以得出目标物体水壶的深度距离为 541.8mm,在实际场景中测得的实际距离为 558mm.然后在保持同长度基线的情况下对不同的深度距离进行测量,所得到的结果如表 3 所示.

表 3 基线  $b=300\text{mm}$  时测定的数据

标定距离/mm	实测距离/mm	匹配时间/s
600	617.8	1.365
1000	1018.5	1.386
1500	1516.7	1.402
2000	2018.5	1.387

从表 3 中可以看出,基于 SURF 算法和运动立体测距法的单目转双目立体视觉的定位,在一定的精度范围内,完全可以实现双目立体视觉定位的功能.试验结果所产生的误差是有多方面原因的,比如摄像机自身的内部参数、拍照时的光线强度以及摄像机的摆放位置,还有进行特征点提取及匹配时误匹配点的影响等.

#### 4 结语

在立体视觉的研究中,多数是采用双目或者多目系统来进行定位研究,但当在物体多个特征点之间的位置关系基本上已知的情况下,我们可以采用单目立体视觉的方法来实现.单目立体视觉仅需要一台摄像机,系统结构设计简单,所占用的空间较小,操作起来也比较方便,更重要的是整个系统造价较低,具有很高的性价比,有一定的现实意义. SURF 作为一种特征提取算法,它具有尺度不变、旋转不变、对光照变化、仿射以及透视变换都具有部分不变的特性,在重复度、独特性、鲁棒性三个方面均超越以往提出的同类方法,且在计算速度上也具有明显的优势,本文将 SURF 算法运用在目标物体的识别与匹配上,得到了较好的效果,当然,中间也存在一定的不足,误匹配点的存在影响了定位的精度,对如何完全消除误匹配点对,需要我们在以后的研究中继续探讨.

#### 参考文献

- 1 魏振忠,张广军,徐园.线结构光传感器标定点获取方法研究.北京航空航天大学学报,2004,30(1):72-76.
- 2 Lowe DG. Distinctive image features from scale invariant key points. International Journal of Computer Vision, 2004, 60(2): 91-110.

- 3 Bay H, Ess A, Tuytelaars T, VanGool L. Speeded up robust features(SURF). Computer Vision and Image. Understanding, 2008, 110(3): 346–359.
- 4 Viola P, Jones M. Rapid Object Detection using a boosted cascade of simple. Proc. of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). Kauai Marriott, Hawaii. 2001. 511–518.
- 5 Bay H, Tuytelaars T, Van Gool L. SURF speeded-up robust features. 9th European Conference on Computer Vision, Graz, Austria. ECCV. 2006. 404–417.
- 6 Lowe DG. Distinctive image features from scale-invariant key points. International Journal of Computer Vision, 2004, 60(2): 91–110.
- 7 Tuvi E. Single-track gray codes and sequences. Lecture Notes in Computer Science, SSC, 2007, 4893: 129–133.
- 8 王伟,钟堰利,吴建坤.广角镜头图像畸变校正系统的研究.光学技术,2004,30(4): 489–490.
- 9 邱茂林,马颂德,李毅.计算机视觉中摄像机定标综述.自动化学报,2000,26(1):43–55.
- 10 张杨,基于双目立体视觉的 CCD 测距系统设计[硕士学位论文].长春:长春理工大学,2010.