

一种改进的基于 SURF 的视频帧间匹配方法

武 频¹, 王 庆¹, 朱永华¹, 高洪皓²

(1. 上海大学 计算机工程与科学学院, 上海 200444;

2. 上海大学 计算中心, 上海 200444)

摘 要:针对视频连续帧间匹配不准确、错误率高、匹配速度慢的问题,提出了一种改进的基于 SURF (Speeded Up Robust Feature) 特征点的匹配方法。按照 SURF 算法进行特征点检测和描述;对视频连续帧利用改进的最近邻与次近邻的比的方法进行双向匹配,在匹配时仅在以相应位置为中心的邻域内寻找最近邻点和次近邻点,根据最近距离与次近邻距离的比值与预先设定阈值的比较结果确定是否接受这一匹配点对;用 RANSAC (Random Sample Consensus) 方法建立变换矩阵模型剔除错误匹配点,得到精确匹配的特征点对,完成匹配过程。在经典的视频数据集上进行实验,实验结果表明该方法不仅提高了视频连续帧间匹配的正确率,同时使匹配时间相对缩短了一半左右,显著提高了匹配效率,证明了算法的有效性。

关键词: SURF; 视频连续帧; 最近邻; 匹配; 随机抽样一致性

中图分类号: TP301.6

文献标识码: A

文章编号: 1673-629X(2017)02-0020-05

doi: 10.3969/j.issn.1673-629X.2017.02.005

An Improved Matching Method in Video Frames Based on SURF

WU Pin¹, WANG Qing¹, ZHU Yong-hua¹, GAO Hong-hao²

(1. School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China;

2. Computer Center, Shanghai University, Shanghai 200444, China)

Abstract: With the problem of inaccurate matching, high error rate and low speed in video frames, an improved matching method based on SURF (Speeded Up Robust Feature) is presented. The SURF features are detected and described. The improved method of ratio between the nearest and the next nearest neighbor is used for bidirectional matching. When matching, the nearest and next nearest neighbor points are searched only in the neighborhood of the corresponding points and the two matching points are accepted according to the comparison results between the distance ratio and the present threshold. The RANSAC method is applied to build the transformation matrix model to removing the error matches and get the exact match, completing the match process. The experiment is carried out on the classic video dataset, and the result shows that the method can improve the matching accuracy, and the matching time is relatively shortened by about half, significantly improving the matching efficiency and verifying the effectiveness of the algorithm.

Key words: SURF; continuous video frames; nearest neighbor; matching; RANSAC

0 引言

视频帧间特征点匹配是视频连续相邻两帧间的匹配,以图像特征点匹配为基础。目前图像特征点匹配方法广泛用于运动跟踪、动作识别和基于内容的视频检索等领域,是计算机视觉的重要研究热点和研究基础^[1]。有很多学者在基于特征点的匹配方法方面进行研究,并取得了显著效果。常用的匹配特征点有 Lowe 等提出并完善的 SIFT (Scale Invariant Feature Transform) 特征点^[2-3], Bay 提出的对 SIFT 的改进 SURF (Speeded Up Robust Features) 特征点^[4]等。文献[5]

对 SIFT 和 SURF 进行比较, SIFT 特征点在很多方面表现较好,但由于特征向量维度过高,实时性不好, SURF 与 SIFT 相比,性能相近但时间上是 SIFT 的三倍。为了加快匹配速度,有人对 SURF 特征点提取与描述方法进行改进^[6-7],也有对匹配方法进行改进^[8]。文中是对匹配方法进行改进。

随着运动跟踪和动作识别等领域的发展,越来越多的学者将图像的匹配方法用于视频连续帧匹配。如佟爱华^[9]利用 SIFT 算法提取视频帧的特征点,使用基于欧氏距离的最近邻与次近邻匹配的方法进行视频连

收稿日期:2015-07-16

修回日期:2015-12-15

网络出版时间:2017-01-10

基金项目:上海市自然科学基金(15ZR1415200)

作者简介:武 频(1975-),女,副教授,博士后,研究方向为数值计算、图像处理。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20170110.1010.026.html>

续帧匹配,但 SIFT 特征点提取时间长,且匹配时对视频帧中的每个特征点在下一帧的所有特征点中搜索,时间效率低且匹配正确率低。卓武汉等^[10]提取视频相邻帧的 SURF 特征点,同样采用最近邻与次近邻比的方法进行匹配,虽然使用 SURF 算法提高了特征点提取的速度,但匹配方法在时间效率和匹配正确率上同样较低。上述两种方法均是将图像匹配直接用于视频帧间匹配,未考虑视频帧自身的特点即视频帧连续间相差很小。为了解决这一问题, J. Yu 等^[11]使用 SURF 算法提取特征点,在匹配时利用欧氏距离度量法在下一帧的一个窗口中搜索距离最近的点作为匹配点。该方法虽在一定程度上缩短了搜索时间,但所使用匹配方法的匹配正确率较低。

针对上述问题,文中抓住 SURF 特征提取速度快及视频帧自身的特点,使用 SURF 算法提取特征点,提出一种改进的最近邻与次近邻比的方法,同时将双向匹配^[12]方法用于视频帧间匹配,利用随机抽样一致性 (RANDOM SAMPLE CONSENSUS, RANSAC) 算法^[13]剔除错误匹配点,不仅减少了特征提取与匹配的时间,同时一定程度上提高了匹配正确率。

1 SURF 特征提取与描述

SURF 算法采用积分图像和盒式滤波器的思想,提高了计算速度。这一节将重点介绍 SURF 特征点提取和描述的步骤。

1.1 SURF 特征点提取

特征点提取过程包括积分图像、极值点检测和特征点确定三个部分。

首先是积分图像,主要是利用其计算速度快的特征来做图像卷积。积分图像表示图像上所有像素的和,即对图像 I 和其上的像素点 $I(x, y)$, 积分图像可用式(1)表示:

$$I_{\Sigma}(x, y) = \sum_{i=0}^x \sum_{j=0}^y I(i, j) \quad (1)$$

由式(1)可见,其计算方法非常简单,速度快。例如,若图像区域由顶点 A, B, C 和 D 四点构成,则图像的像素值之和为 $\Sigma = A - B - C + D$ 。

其次,使用近似 Hessian 矩阵检测极值点。Hessian 矩阵具有良好的计算时间和精度表现, SURF 算法和 SIFT 算法一样都是基于尺度空间的,那么在尺度空间为 σ 的点 $I(x, y)$ 的 Hessian 矩阵定义如下:

$$H = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \quad (2)$$

其中, $L_{xx}(x, \sigma)$, $L_{xy}(x, \sigma)$, $L_{yy}(x, \sigma)$ 是点 $I(x, y)$ 处高斯函数二阶偏导数与图像 I 卷积的结果。

构建尺度空间的过程为:首先在图像上下采样,然

后将其与各个尺度的二维高斯函数进行卷积操作。在进行卷积时, SURF 算法使用方框滤波器代替二阶高斯滤波,这样大大提高了卷积计算速度,进而提高了整个算法的计算速度。仅仅通过改变方框滤波器的大小,就能建立尺度空间, Bay 等对 Hessian 矩阵行列式的值做了较精确的估计,表示为式(3):

$$\Delta H = D_{xx}(x) D_{yy}(x) - (0.9 D_{xy}(x))^2 \quad (3)$$

其中, ΔH 是点 $I(x, y)$ 周围邻域的方框滤波器响应值; D_{xx} , D_{xy} , D_{yy} 是图像和各个方框滤波器模板进行卷积操作后的结果。

判断该点是否为极值点主要是依据两类重要的值,即该矩阵行列式的值及其特征值,如果行列式的结果符号为负,并且特征值有不同的符号,就判定该点不是局部极值点;如果行列式的结果符号为正,且两个特征值同时为正或为负,就可以把该点归类为极值点。

最后确定选出哪些极值点作为最终的特征点。该过程是通过比较极值点的邻域信息进行判断的,具体实现方法是在一个以某极值点为中心的 $3 \times 3 \times 3$ 的立体邻域内,比较相邻的上下尺度和该尺度周围的 26 个邻域值,若该极值点的值比邻域值都大或都小,则判定其是最终的特征点。

1.2 SURF 特征点描述

特征描述过程主要包括两个方面:一是确定基准方向,二是描述子的生成过程。

(1) 基准方向确定。

为了保证特征矢量具有旋转不变性,需要为每一个特征点分配一个主方向即基准方向。具体方法是,首先在以特征点为圆心, $6s$ (s 为特征点所在尺度值) 为半径的区域内计算水平方向和垂直方向的加权 Harr 小波响应,令越靠近特征点的 Harr 小波响应的权值越大;然后用一个 60° 的扇形模板遍历该圆形区域,得到多个由扇形模板内 Harr 响应的累加之和构成的矢量,把最大的累加和所对应的方向作为该特征点基准方向。

(2) 生成描述子。

沿着基准方向构造以特征点为中心,长为 $20s$ 的正方形区域。如图 1 所示,将该区域划分成 4×4 的子区域,在每个子区域内采样 5×5 个点,计算每个点相对于基准方向的水平和垂直方向的加权 Harr 小波响应值,分别记做 d_x 和 d_y 。同样对响应值进行高斯加权,对每个子区域的响应值统计 $\sum d_x$, $\sum d_y$, $\sum |d_x|$, $\sum |d_y|$, 得到的四维矢量 $V = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|)$ 即为该子区域的描述符。特征描述符的生成过程如图 1 所示。那么整个区域的 $4 \times 4 \times 4 = 64$ 维的特征向量就组成了该特征点的 SURF

描述子。

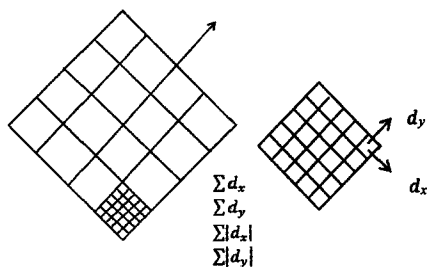


图 1 特征描述符的生成过程

2 视频连续帧间特征匹配

在提取特征之后,为了提高匹配正确率,减少匹配时间,文中对已经提取过特征点的连续视频帧,采用改进的最近邻与次近邻比的方法进行双向匹配,然后使用 RANSAC 算法剔除错误匹配点。其中,距离衡量方式采用欧氏距离,定义如下:

$D =$

$$\sqrt{(x_1 - x'_1)^2 + (x_2 - x'_2)^2 + \cdots + (x_{64} - x'_{64})^2} \quad (4)$$

其中, $(x_1, x_2, \cdots, x_{64})$ 和 $(x'_1, x'_2, \cdots, x'_{64})$ 为待匹配的两个特征点的 SURF 特征向量。

2.1 改进算法

基本思想为:对于前一帧中的每个特征点,在下一帧以这个位置为中心的 $N \times N$ (窗口范围很小) 的窗口范围内搜索最近点和次近点,得到两个距离:最近距离和次近距离。把最近距离与次近距离的比值和预先设定的阈值进行比较,若前者小于后者,则接受最近点作为该点的匹配点,完成单向匹配过程,得到单向匹配的候选匹配对集,然后对后一帧中已匹配的点用同样的方法与前一帧进行反向匹配,得到反向匹配的特征点对。对两个方向得到的匹配点对进行比较,若完全一样,则接受这一匹配点对,将其加入最终的匹配点对集。

具体步骤如下:

(1) 在视频连续帧间,对于前一帧 I_1 的特征点 p_1 , 在下一帧 I_2 中,在以对应位置为中心的 32×32 (该邻域大小对存在较大幅度相机抖动的视频可适当放大) 的邻域内搜索计算出欧氏距离最小的两个点作为最近邻点和次近邻点。最近点为 p_{11} , 次近点为 p_{12} , 得到两个距离,最近距离和次近距离,分别标记为 d_{11} 和 d_{12} 。若 d_{11} 与 d_{12} 的比值 k 小于给定阈值 r (r 取 0.8), 则接受这对匹配对 (p_1, p_{11}) 作为候选匹配对,否则剔除该特征点。接着继续匹配下一对特征点,完成单向匹配。

(2) 对于图像 I_2 中的已经接受的匹配点 p_{11} , 用同样的方法与 I_1 中的点进行匹配,得到其匹配点 p'_1 。对 p'_1 进行判断,若 $p'_1 = p_1$, 则 p_{11} 与 p'_1 相互匹配,否则剔除,进而完成双向匹配过程,得到最终匹配的点对集。

2.2 RANSAC 剔除错误匹配点

按照上述方法完成匹配后,为了进一步剔除错误匹配点对,可以使用 RANSAC 算法。其基本思想是,样本中包含两类数据,正确数据即可以被模型描述的数据,异常数据即偏离正常范围的数据。通常异常数据可能是由错误的测量、估算等产生。文中的异常数据是由匹配误差过大产生的。RANSAC 也假定,可以为一组正确的数据建立一个描述它的模型,其基本步骤如下^[14]:

(1) 初始化模型 M 。已知初始化 M 所需的最小参数为 n 和一个样本集 P , 其中 P 的样本个数大于 n 。从 P 中随机抽取 n 个样本,组成集合 S , 用 S 来初始化模型 M 。

(2) 构建内点集 S' 。将 P 中剩余样本中与模型 M 的误差小于给定阈值的样本组成的集合与集合 S 构成内点集 S' , 称 S' 是 S 的一致集 (consensus set)。

(3) 重复多次寻找一致集。若 S' 的样本数大于等于 N , 就认为得到了正确的模型参数,然后在集合 S' 上采用最小二乘等方法重新计算模型 M' 。重复步骤 (1) 和 (2)。

(4) 用一致集判断内外点。在重复一定次数的抽样后,若找到一致集,则选取抽样后得到的最大一致集判断内外点。否则算法失败。

文中运用 RANSAC 算法的思想是:从前面算法得到的匹配点对集中随机取出一些匹配点对来计算模型,然后根据计算出的模型遍历所有匹配点对,计算在某一特定阈值下满足该模型的百分比。重复上面的步骤,比较满足每个模型的百分比,取对应百分比最大的那个模型作为最终模型。实际实验中,首先根据上述思想建立一个 3×3 的最好的特征变换矩阵作为最终模型,然后在最好模型下剔除误差超过某个阈值的匹配对,得到精确匹配的特征点对。

3 实验结果与分析

实验环境为: Microsoft Visual Studio 2012 和 OpenCV, 在 Intel(R) Core(TM) i5-3230 M, 主频 2.6 GHz, 内存 4G 的主机上进行。

共进行两组测试,分别对 SURF 和文中匹配方法在匹配对数、匹配时间、匹配正确率上进行比较。

第一组测试使用中科院的行为数据集 CASIA 进行实验,每帧大小为 320×240 。该视频为静态相机拍摄,共包括 299 帧,帧率为 25 fps。测试步骤如下:

(1) 对于随机的连续两帧,如视频第 86, 87 帧 (左边 front 为 86 帧,右边 after 为 87 帧) 见图 2(a)。

(2) 使用 SURF 算法提取特征点并进行 64 维的特征点描述。

(3)利用一般的 SURF 特征匹配方法得到的匹配结果如图 2(b)所示。

(4)利用文中的匹配方法得到的匹配结果如图 2(c)所示。

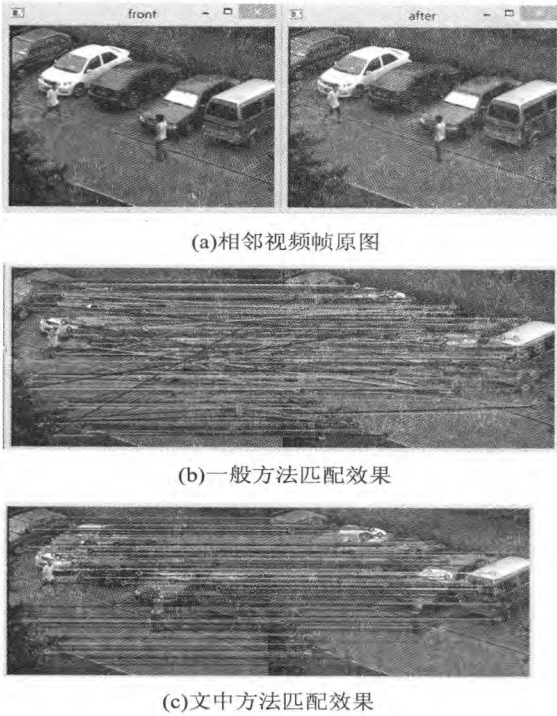


图 2 匹配结果(1)

由图 2(b)可看出,一般 SURF 特征匹配方法得到的匹配对比较杂乱,并且存在很多错误匹配点对。而由图 2(c)可看出,文中方法得到的特征点对相对图 2(b)有所减少,但匹配正确率较高。

下面在匹配点对数、匹配正确率和匹配时间这几个方面进行详细的分析比较。图 2(a)中前一帧有 255 个特征点,后一帧有 263 个特征点,对该视频的 299 帧综合分析所得平均匹配点数、平均错误匹配数、平均正确匹配率和平均匹配时间如表 1 所示。

表 1 算法比较(1)

方法	匹配点数量	错误匹配数量	正确匹配率/%	匹配时间/ms
SURF	230	37	83.9	113.6
文中	226	11	95.1	51.8

由表 1 可知,文中匹配算法对于静止相机拍摄的短视频的特征点匹配情况,不仅在匹配正确率上提高了大概 10%,同时显著缩短了匹配时间,使得匹配时间变为原来的一半左右,证明了文中算法的有效性。

第二组测试使用 YouTube 数据集进行实验,其中的视频是用运动相机拍摄,存在旋转和噪声,用于验证文中算法对于旋转和噪声的鲁棒性。每帧大小为 320 × 240,帧率为 29 fps,测试过程如下:

按照与第一组测试同样的步骤,对于随机连续的

两帧如 42,43 帧,原图如图 3(a)所示(左边 front 为前一帧,右边 after 为后一帧)。一般 SURF 特征匹配方法得到的结果如图 3(b)所示,文中匹配方法得到的结果如图 3(c)所示。

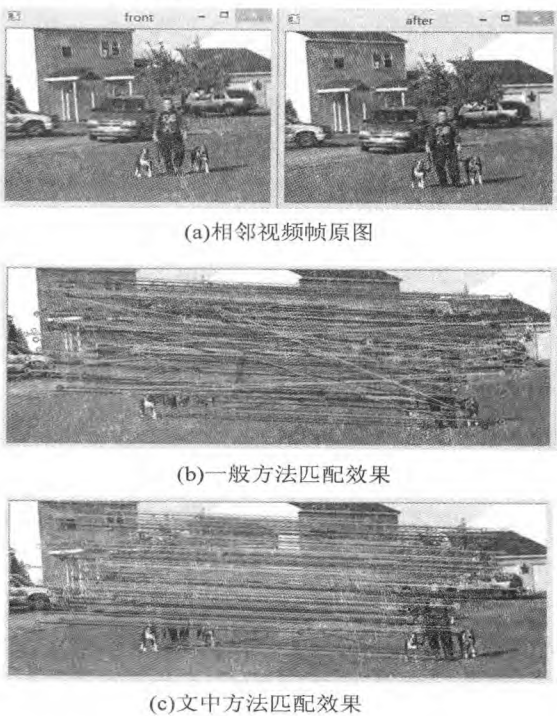


图 3 匹配结果(2)

由图 3 中的(b)和(c)可明显看出,匹配时虽然一般的 SURF 特征匹配方法得到的匹配对较多,但由于在全图片范围搜索对应特征点,存在较多的错误匹配对。而文中的匹配方法虽然匹配对数相对较少,但匹配正确率显著提高。

对这两帧进行综合分析,前一帧检测出 337 个特征点,后一帧 352 个特征点,对一般 SURF 特征匹配方法和文中匹配方法在特征点匹配数量、匹配正确率、匹配时间方面进行比较,结果如表 2 所示。

表 2 算法比较(2)

方法	匹配点数量	错误匹配数量	正确匹配率/%	匹配时间/ms
SURF	207	44	78.74	148.4
文中	171	13	92.39	79.5

由表 2 可知,文中算法在相机运动和噪声存在情况下不仅在匹配正确率方面优势明显,同时匹配时间大概缩短了一半,保证了实时性。对于该情况下的多帧实验,由于存在相机运动和噪声,从各帧提取的特征点数量相差较大,进而匹配时间也相差较多,对上述各比较项不能平均而论,但文中匹配方法的匹配正确率都在 90% 以上,且匹配时间也较一般的 SURF 特征匹配时间缩短了一半左右,从而证明了文中算法的有效性。

对两组实验进行综合分析,第一组测试针对静态相机拍摄视频,通过图 2 和表 1 可看出,文中算法不仅提高了匹配正确率,而且缩短了匹配时间,证明了在该情况下的有效性。第二组实验视频是在运动相机下拍摄且存在严重噪声,比较图 2(c)和图 3(c)可明显看出,匹配对间的连线在图 2(c)中是水平的,在图 3(c)中是斜的,这是由于存在相机运动,导致两帧间有一定程度的旋转,进而得到图 3(c)匹配对间的连线是斜的。该组实验验证了文中算法在该情况下的有效性。通过两组实验证明了文中算法对静态相机、相机抖动及噪声存在情况下拍摄的视频进行连续帧匹配的有效性。

4 结束语

研究了视频帧间的基于 SURF 的匹配方法。采用 Fast-Hessian 快速检测特征点,并对特征点进行描述,然后通过改进的最近邻与次近邻比的方法进行双向匹配,最后用 RANSAC 算法剔除错误匹配点。通过在静态相机、动态相机、噪声这几种不同情况下拍摄的视频进行实验,证明文中算法不仅提高了匹配正确率,而且缩短了匹配时间,对于视频中基于特征匹配的跟踪、识别等都有很大的实际作用。下一步可以在特征的提取和匹配上使用 GPU 加速进一步提高匹配效率。

参考文献:

- [1] 余道明. 图像配准技术研究及应用[D]. 成都:西南交通大学,2014.
- [2] Lowe D G. Object recognition from local scale-invariant fea-

ture[C]//Proceedings of the seventh international conference on computer vision. [s. l.]:IEEE,1999:1150-1157.

- [3] Lowe D G. Distinctive image features from scale-invariant features points[J]. International Journal of Computer Vision, 2004,60(2):91-110.
- [4] Bay H. SURF:speeded up robust features[J]. Computer Vision and Image Understanding,2008,110(3):346-359.
- [5] 葛盼盼,陈 强. 基于 SURF 特征提取的遥感图像自动配准[J]. 计算机系统应用,2014,23(3):16-24.
- [6] 张开玉,梁凤梅. 基于改进 SURF 的图像配准关键算法研究[J]. 科学技术与工程,2013,13(10):2875-2879.
- [7] 崔振兴,曾 威,杨明强,等. 一种改进的 SURF 快速匹配方法[J]. 江苏师范大学学报:自然科学版,2014,32(3):41-46.
- [8] Yang Gui. Point-pattern matching method using SURF and shape context[J]. OPTIK,2013,124:1869-1873.
- [9] 佟爱华. 一种改进的高精度视频帧间匹配算法[C]//第二届仪表、自动化与先进集成技术大会论文集. 出版地不详:出版者不详,2008:146-150.
- [10] 卓武汉,严京旗. 基于 SURF 的连续帧图像配准及高光去除[J]. 微型电脑应用,2011,27(1):37-39.
- [11] Yu J, Jeon M, Pedrycz W. Weighted feature trajectories and concatenated bag-of-features for action recognition[J]. Neurocomputing,2014,131(7):200-207.
- [12] 赵璐璐. 基于 SURF 和快速近似最近邻搜索的图像匹配算法[J]. 计算机应用研究,2013,30(3):921-923.
- [13] Feng Qi, Xu W, Zhang W, et al. Research of image matching based on improved SURF[J]. Telkomnika Indonesian Journal of Electrical Engineering,2014,12(2):1395-1402.
- [14] 宋卫艳. RANSAC 算法及其在遥感图像处理中的应用[D]. 北京:华北电力大学,2011.

(上接第 19 页)

查询提供另一种方式。但是语义规则不具有通用性和易读性,规则可能造成本体关系冲突,维护成本较高。在实际应用中,如何提升查询的准确率和效率,简化 SQWRL 语言,解决冲突等问题是今后的研究方向。

参考文献:

- [1] O'Connor M J, Das A. SQWRL: a query language for OWL [C]//Fifth international workshop on OWL: experiences and directions. Chantilly, VA:[s. n.],2009.
- [2] 张艳涛,陈俊杰,相 洁. 基于 SWRL 本体推理研究[J]. 微计算机信息,2010,26(3-3):182-183.
- [3] 张宗仁,杨天奇. 基于自然语言理解的 SPARQL 本体查询[J]. 计算机应用,2010,30(12):3397-3400.
- [4] 王 海,范 琳,李增智. 基于 SQWRL 的语义 Web 服务发现[J]. 微电子学与计算机,2010,29(9):76-79.
- [5] Zhang Yingzhong, Luo Xiaofang, Zhao Yong, et al. An ontology-based knowledge framework for engineering material selec-

tion[J]. Advanced Engineering Informatics,2015,29(4):985-1000.

- [6] 周 亮,黄志球,倪 川. 基于 SWRL 规则的本体推理研究[J]. 计算机技术与发展,2015,25(10):67-70.
- [7] O'Connor M J, Das A. A pair of OWL 2 RL reasoners[C]//Proceedings of OWL experiences on directions workshop. [s. l.]:[s. n.],2012:1-10.
- [8] Horridge M, Bechhofer S. The OWL API: a java API for OWL ontologies[J]. Semantic Web,2011,2(1):11-21.
- [9] Horrocks I, Patel-Schneider P F, Boley H, et al. SWRL: a semantic web rule language combining OWL and RuleML[C]//World wide web consortium. [s. l.]:[s. n.],2004.
- [10] 张新华,张 飞. “知识”概念及其涵义研究[J]. 图书情报工作,2013,57(6):49-58.
- [11] Protégéteam. SWRLTab build project[EB/OL]. 2014. <https://github.com/protegeproject/swrltab-project>.
- [12] 刘 伟. Java 规则引擎-Drools 的介绍及应用[J]. 微计算机应用,2005,26(6):717-721.