

SIFT 算法在不同维数下的图像匹配效率

袁勃¹, 王殿君², 姜永成³, 张秀华^{3,4}

1. 东北石油大学化学化工学院, 黑龙江大庆 163318
2. 北京石油化工学院机械工程学院, 北京 102617
3. 佳木斯大学机械工程学院, 黑龙江佳木斯 154007
4. 哈尔滨理工大学机械动力工程学院, 哈尔滨 150080

摘要 在移动机器人视觉定位领域中, SIFT (Scale Invariant Feature Transform) 算法由于具有尺度、旋转和光照不变等特性被得到广泛应用。图像匹配效率与 SIFT 算法中所取维数有关, 使用多少维数来表示一个特征点, 满足不同的图像匹配精度和实时性的要求, 还没有明确定义。为解决这个问题, 针对不同维数的 SIFT 算法的匹配效率做了实验分析, 得到了不同应用场合下使用维数的指导范围。以室内场馆导航机器人视觉定位为例, 选取了 16 维作为匹配维数, 实验结果能够满足移动机器人实时定位的需要。

关键词 SIFT 算法; 移动机器人; 匹配定位; 维数

中图分类号 TP242.6

文献标识码 A

文章编号 1000-7857(2010)14-0085-05

Efficiency of Sift Image Matching Algorithm of Different Dimensions

YUAN Bo¹, WANG Dianjun², JIANG Yongcheng³, ZHANG Xiuhua^{3,4}

1. College of Chemistry and Chemical Engineering, Northeast Petroleum University, Daqing 163318, Heilongjiang Province, China
2. College of Mechanical Engineering, Beijing Institute for Petrochemical Technology, Beijing 102617, China
3. College of Mechanical Engineering, Jiamusi University, Jiamusi 154007, Heilongjiang Province, China
4. College of Mechanical & Power Engineering, Harbin University of Science and Technology, Harbin 150080, China

Abstract For mobile robot visual positioning, SIFT (Scale Invariant Feature Transform) algorithm is widely used because of its features of scale, rotation and illumination invariability. The efficiency of the SIFT image matching algorithm is related to the number of dimensions, which is taken as the characteristic point to meet different image matching precision and real-time requirements. To solve this problem, for different dimensions of the SIFT algorithm, experimental analysis was carried out for efficient matching to obtain a range of dimensions with a certain degree of practical significance. The test results show that, for image position, the characteristic point dimension functions are different. For a high performance of real-time image positioning, the function of $1 \times 1 \times N$ is selected, and for the mobile robot image positioning, a high real-time but normal position precision is needed, the function of $2 \times 2 \times N$ is a best choice. Taking the robot navigation of visual indoor positioning as an example, selecting 16 dimensions in a dimension matching of experimental results of the mobile robot, the needs of real-time location can be satisfied.

Keywords SIFT algorithm; mobile robot; matching location; dimension

0 引言

在室内环境中, 移动机器人的视觉定位需要很强的实时

性, 而在实际情况下采用基于经典的 SIFT 算法很难达到定位匹配效率的要求, 为此研究 SIFT 算法在不同维数下对匹配效

收稿日期: 2010-05-15

基金项目: 国家高技术研究发展计划(863 计划)项目(2007AA04Z255)

作者简介: 袁勃, 研究方向为自动控制技术, 电子信箱: dqsyuanbo@126.com; 王殿君(通信作者), 副教授, 研究方向为机器人技术, 电子信箱: wdjkevin@126.com

率的影响具有非常重要的意义。

1999 年 Lowe 等^[1]提出用 128 维特征向量表示一个特征点的经典 SIFT 算法,该算法具有很好的稳定性和匹配性。多尺度空间特征点形成以后,为了得到不同的匹配效率,可对同一幅图像采用不同的维数来处理。张春美等^[2]提出改进的 SIFT 算法,对每个特征点只采用 64 维特征向量表示,该算法减少了算法的复杂度,缩短了匹配时间。刘立等^[3]提出简化的 SIFT 算法,对每个特征点只采用 12 维特征向量表示,一定程度上提高了图像匹配的效率。但在针对不同的应用场合如何选择维数方面,很少见到相关文献报道。

鉴于此,本文基于 SIFT 算法,对不同维数下的图像匹配效率进行研究。进行了不同维数下的图像匹配实验,得到了维数和匹配效率的关系,为在不同场合使用 SIFT 算法,提高系统实时性提供了有益的参考。

1 SIFT 算法不同维数的描述

1.1 SIFT 算法简介

SIFT 算法主要包括两个部分:特征点提取过程和特征点匹配过程。其中,SIFT 特征提取流程如图 1 所示。

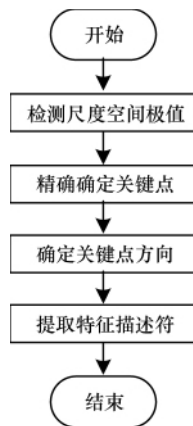


图 1 SIFT 特征提取流程图

Fig. 1 Diagram of SIFT feature extraction

1.2 SIFT 算法分析

1.2.1 尺度空间极值检测

“尺度空间”概念是著名的“图像金字塔”概念的补充,Koenderink 和 Lindeberg 等证明高斯核是唯一可能线性尺度核^[4]。

$$L(x, y, \sigma) = G(x, y, \sigma)I(x, y) \quad (1)$$

公式(1)中, (x, y) 为像素位置坐标; σ 为尺度空间因子,其大小决定图像的平滑程度; L 为图像尺度空间。 $G(x, y, \sigma)$ 为二维高斯函数,其定义如下:

$$G(x, y, \sigma) = e^{-\frac{(x^2 + y^2)}{2\sigma^2}} / 2\pi\sigma^2 \quad (2)$$

为了使特征具备良好的独特性和稳定性,提出高斯差分尺度空间 DOG(Difference of Gaussian),建立图像的多尺度空间,如图 2 所示。DOG 算子为

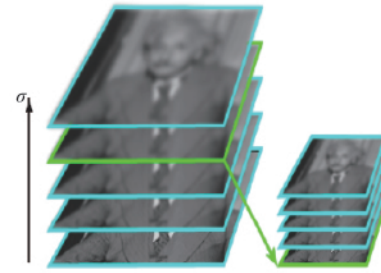


图 2 DOG 尺度空间的构建

Fig. 2 Diagram of DOG dimension space

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma))I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (3)$$

如图 3 所示,要获得某个尺度空间中的极值点,除了将该点和其周围的 8 个点进行比较外,还将它与相邻尺度对应的 9×2 个点,共 26 个点进行大小比较,以确保在尺度空间和二维图像空间都能检测到极值点。

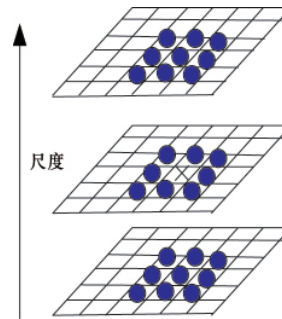


图 3 尺度空间极值检测

Fig. 3 Diagram of DOG dimension space limiting value detection

1.2.2 精确确定特征点位置

在找到极值点后,需要去除低对比度的点和不稳定的边缘响应点。一个不恰当的高斯差分函数会在边缘造成较大的主曲率。主曲率可表示为一个 2×2 的 Hess 矩阵

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (4)$$

令 σ 为最大特征值, β 为最小的特征值,则

$$\text{Tr}(H) = D_{xx} + D_{yy} = \alpha + \beta \quad (5)$$

$$\text{Det}(H) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta \quad (6)$$

令 $\alpha = r\beta$, 则

$$\frac{\text{Tr}(H)^2}{\text{Det}(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r + 1)^2}{r} \quad (7)$$

为了检测主曲率是否在某域值 r 下,只需检测

$$\frac{\text{Tr}(H)^2}{\text{Det}(H)} < \frac{(r + 1)^2}{r} \quad (8)$$

Lowe 在文章中^[1]取 $r = 10$ 。如果主曲率间的比值大于 10,则认为该点是位于边缘而被过滤掉。

1.2.3 指定算子的方向参数

以关键点邻域像素的梯度方向分布特征为每个关键点

指定主方向,各像素梯度的模和方向为

$$m(x, y) = \left[(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2 \right]^{\frac{1}{2}} \quad (9)$$

$$\theta(x, y) = \frac{\arctan(L(x, y+1) - L(x, y-1))}{(L(x+1, y) - L(x-1, y))} \quad (10)$$

1.2.4 提取特征点描述符

检测完图像的关键点,可以确定一个 SIFT 特征区域。图 4 (a) 中每个小格代表关键点邻域所在尺度空间的一个像素点,图 4(b)中一个关键点由 4 个种子点组成,每个种子点包含 8 个方向的信息。为了增强算法抗噪声的能力,同时对含有定位误差的特征匹配提供较好的兼容性,可以将邻域设为 4x4 大小(图 4 只表示其中一部分)。这样,关键点由 128 维特征向量(4x4 个种子点)表示。此时的 SIFT 特征向量已经去除了尺度变化、旋转等几何变形因素的影响。为减少光照对特征点描述符的影响,需要将特征点描述符向量标准化为单位长度,即进行向量长度归一化。

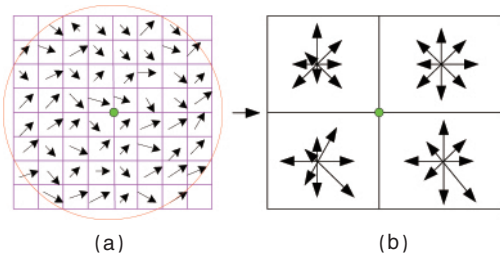


图 4 特征点邻域梯度信息生成特征向量

Fig. 4 Feature vector generation by neighboring feature gradient information

1.2.5 图像向量长度归一化

为了减少光照对特征点描述符的影响,需要将特征点描述符向量标准化为单位长度,即进行向量长度归一化。其中,

1) 当图像对比度改变时,每个像素的梯度值都被乘以一个相同的常量,因此,对比度对像素的影响会通过向量的标准化而被消除掉。

2) 当亮度改变时,每个像素都会被增加一个常数,由于梯度是通过像素差计算得到的,特征点描述符不受图像光线仿射变化的任何影响。

3) 照相机的饱和度和光线的改变会在物体表面不同地方引起梯度值的较大改变,但对梯度方向影响较小。为了减少由于梯度值较大变化而产生的影响,规定如果某个单位特征向量小于一个阈值(例如 0.2),将被重新标准化为单位长度。

2 SIFT 算法匹配效率实验

2.1 SIFT 算法与维数的关系

提取特征点描述符占整个算法计算时间的 70% 以上。在这个过程中,维数的确定至关重要,采用不同维数可以得到不同的特征点数和计算时间,决定 SIFT 算法效率的关键也在

于此。本文综合考虑移动机器人的定位精度和实时性要求,对 SIFT 算法进行了改进,对不同维数下的图像匹配性能进行测试分析。对同一幅图像分别以 12 维和 128 维为例进行对比计算测试,特征点数和时间关系如图 5 所示。

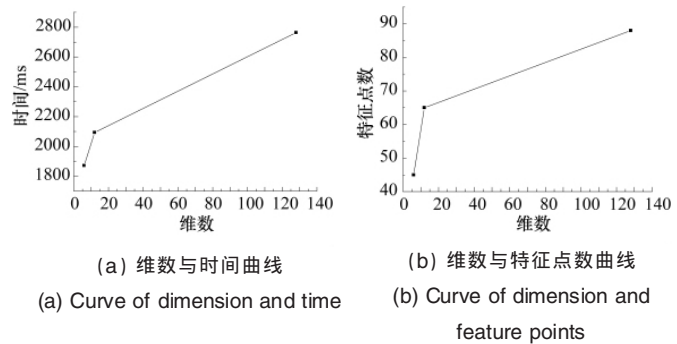


图 5 维数和特征点数、时间关系图

Fig. 5 Relationship between dimension and feature points

由图 5(a)可以看出,随着维数的增加,计算时间明显提高;由图 5(b)可知,随着维数的增加,特征点数也相应增加。

2.2 不同维数对匹配效率的影响分析

图像匹配效率是指正确的匹配率与计算时间的比值。根据移动机器人视觉定位的要求,为了找到适合移动机器人视觉定位的最佳匹配效率,对 30 幅图像进行了图像匹配分析,分别对每幅图像提取 10 次特征点。实验结果表明,随着维数的增加,图像提取特征点的数目增加且匹配精度提高,但匹配时间也大大增加;随着维数的减少,特征点的数目减少,精度下降,但大大减少了匹配时间。如何在匹配精度和匹配时间之间找到一个适合移动机器人实时定位要求的平衡点,是本文研究的一个重点。

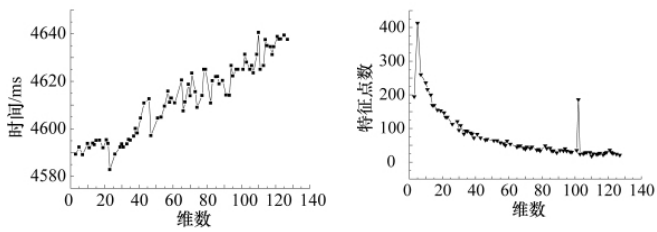
2.3 实验平台简介

实验平台为 Windows XP 操作系统、CPU 主频 1.66GHz、512MB 内存的个人计算机,编程环境为 VC++ 6.0 和 OpenCV 视觉库。其中,OpenCV 是由 Intel 公司资助的开源计算机视觉库,包括计算机视觉方面的多种通用算法,来实现图像处理^[5]。基于以上实验条件设计了一套图像匹配软件,该软件能实现图像的匹配,并自动记录匹配时间和获得的特征点个数。

2.4 对比实验

根据研制的图像匹配系统,利用 SIFT 算法分别对图像从 1 维至 128 维做了 10 组匹配实验,获得了 1280 组数据,并进行了统计绘图。在对图像进行 SIFT 算法处理的过程中,128 维向量的计算表达式为: $M \times M \times N$ (M 为行或列数, $M=1, 2, 3, 4, 5$,行数和列数保持相等; N 为关键点邻域内的像素点个数, $1 \leq N \leq 128$)。

图 6 为 $1 \times 1 \times N$ 的时间、特征点数和维数的关系图。由图 6 (a)可知,随着维数的增加,匹配计算时间也在增加,时间变化范围在 4580~4650ms 之间;由图 6(b)可看出,随着维数的增加,特征点数逐渐变少,且特征点数的变化幅度较大。

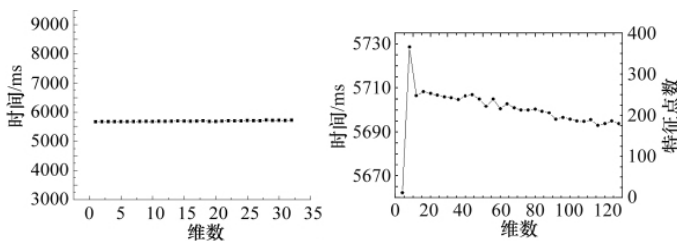


(a) 维数与时间曲线 (b) 维数与特征点数曲线
(a) Curve of dimension and time (b) Curve of dimension and feature points

图 6 $1 \times 1 \times N$ 的时间、特征点数和维数的关系
Fig. 6 Relationship of $1 \times 1 \times N$ type among time, dimension and feature points

综上所述,此 $1 \times 1 \times N$ 维数类型的表达方式不满足图像定位的要求。

图 7 为 $2 \times 2 \times N$ 的时间、特征点数和维数的关系图。由图 7(a)可知,随着维数的增加,匹配计算时间线变化幅度较小,基本保持在一个时间水平上,表明在确定行列数后,邻域周围像素点个数的多少对匹配时间的影响不大;由图 7(b)可知,在匹配过程中,在 N 为 0~3 这段区间内存在一些虚匹配,随着邻域周围像素点个数的增加,错配点数也减少,而且匹配的特征点也减少,变化范围在 200~250 之间。维数在 4~32 这段区间内,特征点数变化比较稳定,匹配时间较少,满足移动机器人图像匹配定位的需要。

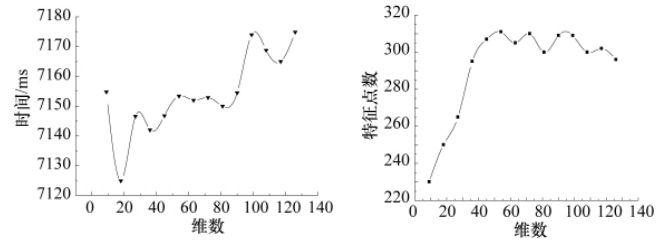


(a) 维数与时间曲线 (b) 维数与特征点数曲线
(a) Curve of dimension and time (b) Curve of dimension and feature points

图 7 $2 \times 2 \times N$ 的时间、特征点数和维数的关系
Fig. 7 Relationship of $2 \times 2 \times N$ type among time, dimension and feature points

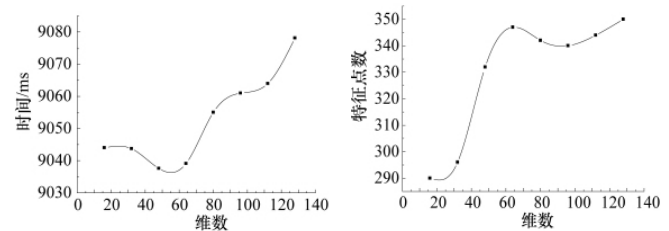
图 8 为 $3 \times 3 \times N$ 的时间、特征点数和维数的关系图。由图 8(a)可知,满足 3×3 行列的特征向量个数随着维数的增加,匹配时间也在增加,且变化幅度较大,时间变化范围在 7120~7180ms 内;由图 8(b)可看出,随着维数的增加,特征点数逐渐变小,而且特征点数的变化幅度较大,变化范围在 230~320 之间,不适合移动机器人的定位要求。

图 9 为 $4 \times 4 \times N$ 的时间、特征点数和维数的关系图。由图 9(a)可知,满足 4×4 行列的特征向量个数,随着维数的增加,匹配时间大幅增加,时间变化范围在 9030~9080ms 之内;由图 9



(a) 维数与时间曲线 (b) 维数与特征点数曲线
(a) Curve of dimension and time (b) Curve of dimension and feature points

图 8 $3 \times 3 \times N$ 的时间、特征点数和维数的关系
Fig. 8 Relationship of $3 \times 3 \times N$ type among time, dimension and feature points



(a) 维数与时间曲线 (b) 维数与特征点数曲线
(a) Curve of dimension and time (b) Curve of dimension and feature points

图 9 $4 \times 4 \times N$ 的时间、特征点数和维数的关系图
Fig. 9 Relationship of $4 \times 4 \times N$ type among time, dimension and feature points

(b)可看出,随着维数的增加,特征点数逐渐增多,而且特征点数的变化幅度较大,变化范围在 290~350 之间。综上所述,由于系统的匹配时间较长,而且特征点数变化较大,不利于提高系统定位的实时性,不适合移动机器人的定位使用。

对于 $5 \times 5 \times N$ 模式,实验同样证明随着维数的增加,特征点数变化不稳定,而且变化幅度较大。不利于移动机器人定位。

图 10 为匹配效率与行列数的关系图。每个特征点可以用不同数目的行列数和像素点数的乘积表示,且每个特征点包含的梯度的模值、方向和像素点的数量也不同。

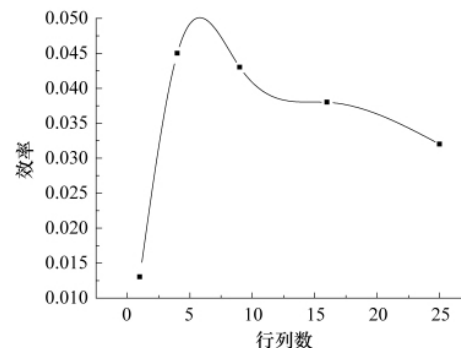


图 10 匹配效率与行列数的关系图
Fig. 10 Relationship of match efficiency and the number of columns and rows

通过对上述不同维数实验的对比分析,可以得出以下结论,如表 1 所示。

表 1 维数不同表达式对图像匹配的影响
Table 1 Influence of image matching by deferent dimension functions

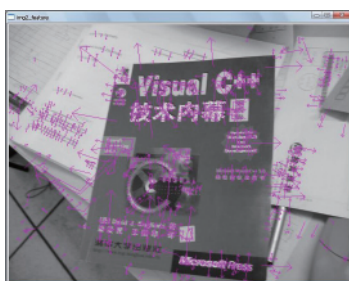
维数表达式	匹配精度	实时性
1×1×N	不高	好
2×2×N	稍高	较好
4×4×N	高	弱

对于实时性要求高的图像定位,可以选择 1×1×N 维数表示一个特征点;对于实时性要求较高、精度要求不太高的基于视觉的机器人定位,选择 2×2×N 维数表示一个特征点的方式作为最优选择;对于精度要求高、实时性要求不高的图像匹配定位,选择 4×4×N 维数表示一个特征点。

SIFT 特征匹配时,采用最近邻方法^[8]寻找特征点的最近邻点和次近邻点,因此寻找最近邻点和次近邻点的效率将直接影响整个特征匹配的效率。本文采用 BBF^[6]近似算法,该算法能够以较大概率找到最近邻点,效率相对较高。



(a) 目标图
(a) Objective diagram



(b) 背景图
(b) Background diagram



(c) SIFT 匹配图
(c) Match diagram with SIFT

图 11 图像匹配测试图

Fig. 11 Matching test of image

由此得出维数为 16 时, 移动机器人图像匹配可以获得较好的实时性, 满足移动机器人定位需要。

3 结论

- 1) 通过对 SIFT 算法维数的分析及实验分析, 得出随着维数的增加, 匹配效率逐渐下降但匹配点个数逐渐增多的趋势。
- 2) 通过大量对比实验, 得到了匹配效率和维数的指导性范围, 为不同场合应用提供了指导性参考。
- 3) 根据指导范围, 得到了移动机器人视觉定位的合适维数, 为移动机器人获得很好的实时定位效果奠定了基础。

参考文献 (References)

[1] Lowe D G. Object recognition from local scale invariant features [C]// International Conference on Computer Vision. Corfu, Greece. 1999: 1150-1157.

2.5 实验结论

通过对实验总结分析, 对 1~128 维向量进行分类, 分别对行列数、时间、特征点数和特征点与时间的比值做了对比分析, 如表 2 所示。

表 2 不同维数的参数比较

Table 2 Parameter comparison of different dimension

行列数	特征点数	平均时间/ms	特征点数与时间比
1×1	20~200	4613	0.023
2×2	200~300	5650	0.045
3×3	230~320	7400	0.037
4×4	290~350	9106	0.035
5×5	310~330	11416	0.028

由表 2 可知, 行列数为 1×1 时得到的特征点数很少, 不适合机器人图像匹配。而行列数由 2×2 变为 3×3 和 4×4 时, 特征点数变化不大而查找特征点时间增加很多。因此, 行列数为 2×2 是合理的选择。根据维数表达式而确定的维数为 16 维 (2×2×4)。相应的图像匹配测试图如图 11 所示。匹配过程中测试的图像匹配参数为: 特征点数 258, 计算所用时间 5000ms, 特征点数与时间的比值为 0.045。

[2] 张春美, 龚志辉, 孙雷. 改进 SIFT 特征在图像匹配中的应用[J]. 计算机工程与应用, 2008, 44(2): 95-97.
Zhang Chunmei, Gong Zhihui, Sun Lei. *Computer Engineering and Applications*, 2008, 44(2): 95-97.

[3] 张凯良, 杨丽, 张铁中. 草莓采摘位置机器视觉与激光辅助定位方法[J]. 农业机械学报, 2010, 41(4): 152-156.
Zhang Kailiang, Yang Li, Zhang Tiezhong. *Transaction of the Chinese Society for Agricultural Machinery*, 2010, 41(4): 152-156.

[4] Lindeberg T. Scale space theory: A basic tool for analyzing structures at different scales [J]. *Journal of Applied Statistics*, 1994, 21: 224-270.

[5] 刘瑞祯, 于仕琪. OpenCV 教程——基础篇[M]. 北京: 北京航空航天大学出版社, 2008.
Liu Ruizhen, Yu Shiqi. *Basic tutorials of OpenCV* [M]. Beijing: Beihang University Press, 2008.

[6] Beis J, Lowe D G. Shape indexing using approximate nearest-neighbour search in high dimensional spaces [C]//Conference on Computer Vision and Pattern Recognition, Puerto Rico, 1997: 1000-1006.

(责任编辑 齐志红)