

# 基于卷积神经网络的 SIFT 特征描述子降维方法

周宏浩<sup>1,2</sup>, 易维宁<sup>2</sup>, 杜丽丽<sup>2\*</sup>, 乔延利<sup>1,2\*\*</sup>

<sup>1</sup>中国科学技术大学环境科学与光电技术学院, 安徽 合肥 230031;

<sup>2</sup>中国科学院通用光学与表征技术重点实验室, 安徽 合肥 230031

**摘要** 针对 128 维尺度不变特征变换(SIFT)特征描述子进行图像局部特征点提取时匹配时间过长,以及三维重建进行特征点配准时的应用局限性,结合深度学习方法,提出一种基于卷积神经网络的 SIFT 特征描述子降维方法。该方法利用卷积神经网络强大的学习能力实现了 SIFT 特征描述子降维,同时保留了良好的仿射变换不变性。实验结果表明,经过训练的卷积神经网络将 SIFT 特征描述子降至 32 维时,新的特征描述子在旋转、尺度、视点以及光照等仿射变换下均具有良好的匹配效果,匹配效率比传统 SIFT 特征描述子效率提高了 5 倍。

**关键词** 图像处理; 神经网络; 图像局部特征点提取; 降维

中图分类号 TP391.4

文献标识码 A

doi: 10.3788/LOP56.141008

## Convolutional Neural Network-Based Dimensionality Reduction Method for Image Feature Descriptors Extracted Using Scale-Invariant Feature Transform

Zhou Honghao<sup>1,2</sup>, Yi Weining<sup>2</sup>, Du Lili<sup>2\*</sup>, Qiao Yanli<sup>1,2\*\*</sup>

<sup>1</sup>*School of Environmental Science and Optoelectronic Technology, University of Science and Technology of China, Hefei, Anhui 230031, China;*

<sup>2</sup>*Key Laboratory of Optical Calibration and Characterization, Chinese Academy of Sciences, Hefei, Anhui 230031, China*

**Abstract** Since local feature descriptors extracted from an image using the traditional scale-invariant feature transform (SIFT) method are 128-dimensional vectors, the matching time is too long, which limits their applicability in some cases such as feature point matching based on the three-dimensional reconstruction. To tackle this problem, a SIFT feature descriptor dimensionality reduction method based on a convolutional neural network is proposed. The powerful learning ability of the convolutional neural network is used to realize the dimensionality reduction of SIFT feature descriptors while maintaining their good affine transformation invariance. The experimental results demonstrate that the new feature descriptors obtained using the proposed method generalize well against affine transformations, such as rotation, scale, viewpoint, and illumination, after reducing their dimensionality to 32. Furthermore, the matching speed of the feature descriptors obtained using the proposed method is nearly five times faster than that of the SIFT feature descriptors.

**Key words** image processing; neural network; image local feature point extraction; dimensionality reduction

**OCIS codes** 100.4996; 100.2000; 100.2960

## 1 引言

尺度不变特征变换(SIFT)图像特征提取算法具有良好的旋转和尺度不变性,在图像局部特征点

提取领域应用广泛。在许多领域需要提取不同视点拍摄图像的局部特征点并进行匹配,如运动恢复结构(SFM)<sup>[1]</sup>和文献[2]中基于多幅图像的三维重建过程都需要提取不同角度拍摄的图像特征点并进行

收稿日期: 2019-01-14; 修回日期: 2019-01-20; 录用日期: 2019-02-21

基金项目: 国家自然科学基金(41601379)、国家 973 计划(61322504)

\* E-mail: yljiao@aiofm.ac.cn; \*\* E-mail: lilydu@aiofm.ac.cn

配准。SIFT 算法提取的局部特征点描述子是一个 128 维的向量,其提取特征点的匹配效率较低,因为在特征点配准中,一般根据两个局部特征点描述子之间的欧氏距离度量特征点之间的相似性,距离越近,相似度越高。因此,对于三维重建等应用场景,探索具有良好仿射不变性和高匹配效率的图像局部特征描述子十分重要。

图像局部特征点提取一直是计算机视觉领域的一个研究热点,图像特征点的提取和配准在许多领域也具有广泛应用。最开始,Lowe<sup>[3]</sup>提出的 SIFT 局部特征点提取算法利用高斯金字塔和 DoG 金字塔得到具有旋转与尺度不变性的关键点,然后根据关键点周围点的梯度信息形成特征点描述子;SIFT 算法的改进算法——SURF(speeded up robust features)<sup>[4]</sup>,其局部特征点提取算法利用海森矩阵行列式的值检测特征点,使用积分图进行加速运算,改进了特征点描述子的生成方式;随后,基于主成分分析(PCA)的 SIFT(PCA-SIFT)<sup>[5]</sup>算法出现,该算法在生成特征点描述子时利用 PCA 对特征描述子进行降维,旨在提高特征点匹配的速率;BRISK(binary robust invariant scalable keypoints)<sup>[6]</sup>特征描述子具有良好的旋转尺度不变性;ORB(oriented FAST and rotated BRIEF)<sup>[7]</sup>算法利用 FAST<sup>[8]</sup>算法完成特征点检测,使用 BRIEF(binary robust independent elementary features)<sup>[9]</sup>算法获取特征点的描述子;FREAK(fast retina keypoint)<sup>[10]</sup>在 BRISK 的基础上改变了采样方式,采用的是一种更接近人眼获取图像信息的采样方式。上述算法的匹配效率较传统 SIFT 算法有一定程度的提高,但旋转、尺度以及光照不变性要远弱于传统的 SIFT 算法。上述算法的匹配效率较传统 SIFT 算法有一定程度的提高,但旋转、尺度以及光照不变性要远弱于传统的 SIFT 算法。另外,针对无色差光滑曲面特征点难以提取的问题,文献[11]中提出了一种利用均匀激光网格创造曲面上特征以及特征提取的方法。在 ORB 算法的基础上,文献[12]中提出了一种融合彩色不变量与基于 SURB 和对象请求代理检测的优化匹配方法,该方法提高了匹配效率,同时获得了较高的匹配精度。

上述局部特征点提取算法都基于传统图像特征提取算法。近年来,深度学习方法在计算机视觉领域取得了巨大的成功,在图像识别以及目标检测等领域具有良好的应用前景,例如,文献[13]中提出的基于深度学习的空-谱联合特征提取算法,有效提取了数据中的空-谱特征;文献[14]中使用卷积神经网

络(CNN)完成了点云配准;文献[15]中利用深度学习学习方法学习天气和季节变化下重复出现的特征点,从而检测出环境变化时存在的稳定特征点;文献[16]中利用 Siamese 网络学习 SIFT 特征点周围一定区域的小块图像进而得到特征点描述子;文献[17]中提出了一种使用深度学习计算方法计算特征点方向的方法;在文献[15-17]的基础上,文献[18]中提出的 LIFT(learned invariant feature transform)局部特征点提取算法使用深度学习计算方法完成图像特征点定位、特征点方向估计以及特征点描述子计算,从而端对端地完成局部特征点提取,相对于 SIFT 特征提取方法,其可以提取更多的特征点,但是匹配效率和仿射变换不变性较差。

受到深度学习方法在计算机视觉领域应用的启发,为了使特征描述子同时具有良好的仿射变换不变性和匹配效率,本文结合传统 SIFT 方法和卷积神经网络,提出一种利用卷积神经网络降低 SIFT 特征描述子维度的方法,以提高特征点匹配效率,与自编码器等无监督学习完成数据压缩的方法不同,该方法使用一种监督学习方法完成数据降维。最终建立了一个 Siamese 网络来降低 SIFT 特征点描述子的维度,将所得的新特征描述子称为 CNN-SIFT,并使用从旋转、尺度、视点以及光照变化图像提取的 SIFT 特征点描述子正确匹配对和相应的错误匹配对完成网络训练。

## 2 基本原理

### 2.1 训练数据集创建

使用 ETH3D<sup>[19]</sup>数据集创建训练数据集,该数据集包含 1676 幅不同视点拍摄的不同场景图像,对该数据集图像进行旋转、尺度以及亮度变换以完成数据增强,最终获得 6704 幅图像用于产生训练数据。创建训练数据集的过程如下:1) 利用 SIFT 特征点提取算法提取每张图像的特征点;2) 使用最近邻方法初步匹配 2 幅对应图像的特征点,并使用 RANSAC(random sample consensus)方法消除初步匹配结果中的误匹配,经过这 2 步之后,得到的匹配点基本准确;3) 对于一个图像对,可以得到若干个特征点匹配对,匹配的特征点所对应的特征描述子对被称为一个正样本对,同时,对于特征点对中的一个特征点,也可以在其对应仿射变换的图像中找到一个不匹配的点对,这一对不匹配的特征点对应的特征描述子对被称为一个负样本对。

在训练的过程中,将一对正样本对及其对应的

负样本对视为一个训练样本,实验产生的训练数据集共包含从 3352 幅图像对中提取的 146 万个训练样本。

### 2.2 数据预处理

SIFT 算法对特征点进行定位后,将关键点周围  $16 \times 16$  区域的点划为 16 个  $4 \times 4$  的子区域,梯度划为 8 个方向,统计每个  $4 \times 4$  子区域在各个梯度方向的分布,使用一个 8 维向量描述该子区域,最终可以得到描述这个关键点的  $128(16 \times 8)$  维向量。

将 SIFT 特征提取算法得到的 128 维特征描述子直接输入神经网络时,训练效果不佳,主要原因是直接输入 128 维向量训练神经网络时,整个网络为全连接网络,网络参数多,训练效率低,且在参数过多的情况下出现过拟合的概率非常大。因此,对输入数据进行形状变化,即将 128 维特征描述子转化为  $16 \times 8$  的形式,其中每一行表示特征点周围每一个子区域的描述子,共 16 个子区域,最终得到一个  $16 \times 8$  的二维矩阵,过程如图 1 所示。

图 2 所示为经过形状变化后的特征描述子,其中图 2(a)、(b)为一对匹配特征点的描述子,图 2(a)、(c)为一对不匹配特征点的描述子,



图 2 形状变化之后的 SIFT 特征描述子。(a)特征描述子;(b)与(a)匹配的特征描述子;(c)与(a)不匹配的特征描述子

Fig. 2 SIFT feature descriptors after changing shape. (a) Feature descriptor; (b) feature descriptor matched with Fig. 2(a); (c) feature descriptor unmatched with Fig. 2(a)

### 2.3 网络训练

一般通过不同图像特征点的特征描述子之间的欧氏距离衡量 2 个特征点的相似性,即一对匹配点描述子之间的距离要尽可能小;相反地,一对不匹配点描述子之间的欧氏距离要尽可能大,只有这样才能更好地区分 2 个特征点是否匹配。

训练卷积神经网络时使用小批量梯度下降方法,该方法能够弥补梯度下降方法训练时间过长,以及随机梯度下降方法容易陷入局部最优的缺点,并且在训练的过程中加入动量(momentum)方法以加速网络的收敛。由于训练数据集过于庞大,很难保证网络训练的结果对每一个样本都最优,因此,应该更多地关注那些难以训练的样本。

针对上述情况,在网络训练过程中,使用困难样本挖掘(hard mining)训练方法<sup>[15]</sup>加强一些难以学习样本的训练,该方法的核心思想是:输入  $M$  个样本至神经网络时,使用损失函数值较大的  $m$  个样本

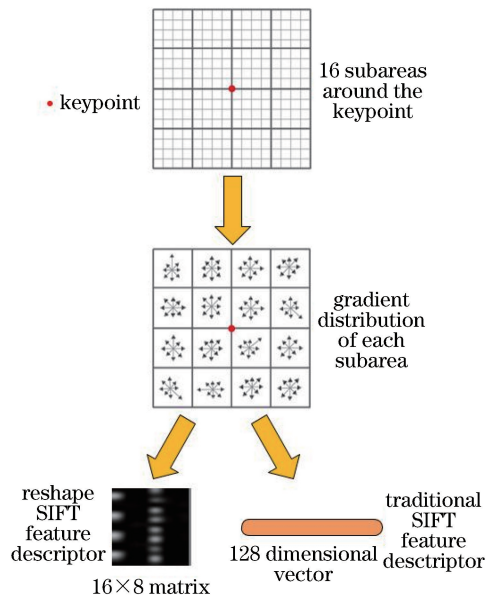


图 1 SIFT 特征描述子形式变换示意图

Fig. 1 Reshapeschematic of SIFT feature descriptor

图 2(a)、(c)为一对不匹配特征点的描述子。从图 2 可以看到,图 2(a)、(b)的梯度信息的分布很相似,相反地,图 2(a)、(c)的梯度信息分布差别较大。

计算最终的损失函数,并且进行反向传播完成参数更新。本文将输入网络样本数和用于网络参数更新的样本数之比  $M/m$  称为困难样本挖掘比(mining-ratio),在训练开始阶段,困难样本挖掘比为 1,然后每 5000 轮训练后比例翻倍。

搭建一个 Siamese 网络完成 SIFT 特征描述子的降维,该网络结构如图 3 所示。从图 3 可以看到,网络一共包含 3 个分支, $D_1$ 、 $D_2$  和  $D_3$  分别表示形态变化后的 SIFT 特征描述子,其中  $D_1$  和  $D_2$  为对应图像中匹配的特征描述子; $D_1$  和  $D_3$  为对应图像中不匹配的特征描述子。网络输出结果为  $d_1$ 、 $d_2$  和  $d_3$ ,分别为  $D_1$ 、 $D_2$ 、 $D_3$  经过卷积神经网络后得到的低维特征描述子。

网络的具体结构参数如表 1 所示,为了加快网络的计算速度,使用 5 层的神经网络,前 3 层中每一层都包含卷积、最大池化以及非线性变换过程,后两层为全连接层, $N$  为 SIFT 特征描述子经过卷积神

经网络降维后的维度。

训练中,将 SIFT 特征描述子的维度降低至 32 维,利用欧氏距离衡量 2 个特征描述子之间的相似性。匹配特征描述子( $d_1, d_2$ )之间的损失函数以及不匹配特征描述子( $d_1, d_3$ )之间的损失函数为

$$L(d_i, d_j) = \begin{cases} \|d_i - d_j\|_2, & \text{when } d_i \text{ and } d_j \text{ are corresponding} \\ \max(0, C - \|d_i - d_j\|_2), & \text{when } d_i \text{ and } d_j \text{ are non-corresponding} \end{cases}, \quad (1)$$

式中; $d_i$  和  $d_j$  为经过卷积神经网络降维后的特征描述子; $\|\cdot\|$  表示欧氏距离; $i, j$  为降维后的特征描述子编号。

通过减少匹配特征描述子( $d_1, d_2$ )之间的损失和不匹配特征描述子( $d_1, d_3$ )之间的损失之和训练网络,最终用于网络训练的损失函数为

$$L_{\text{all}} = L(d_i, d_j)_{\text{corresponding}} + L(d_i, d_j)_{\text{non-corresponding}} \quad (2)$$

### 3 实验方案和结果分析

#### 3.1 测试数据集和测试方法

SIFT<sup>[3]</sup>、SURF<sup>[4]</sup>、BRISK<sup>[6]</sup>、ORB<sup>[7]</sup> 以及 FREAK<sup>[10]</sup> 等特征提取方法都是为了获得具有良好仿射不变性的特征点,其中 BRISK、ORB 和 FREAK 具有较高的匹配效率。为了评估所提方法的性能,将新特征描述子 CNN-SIFT 与上述方法进行对比,测试时使用的 2 个数据集分别为:

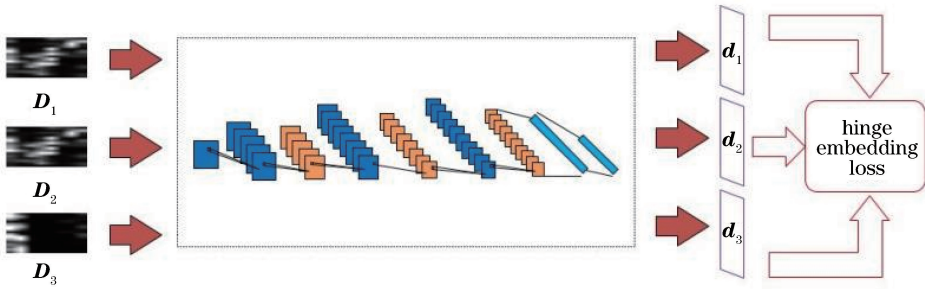


图 3 基于卷积神经网络的 SIFT 特征描述子降维方法的网络结构

Fig. 3 Diagram of convolutional neural network for SIFT feature descriptor dimensionality reduction

表 1 网络结构参数表

Table 1 Structural parameters of network

Layer	1	2	3	4	5
Input size	16×8	8×4	4×2	1×128	1×96
Filter size	3×3	3×3	3×3	-	-
Number of output channel	32	48	64	1	1
Max-pooling size	2×2	2×2	2×2	-	-
Nonlinearity	tanh	tanh	tanh	tanh	tanh
Out size	8×4	4×2	2×1	1×96	1×N

1) 牛津大学仿射变换标准数据集,该数据集包含多组旋转、尺度和视点变化下的图像对,用于测试各个特征描述子在旋转、尺度和视点变化情况下的匹配性能;

2) Webcam<sup>[15]</sup> 数据集,该数据集包含不同光照情况下的图像对,用于测试各个特征描述子在光照变化情况下的匹配性能。

部分用于测试的图像如图 4 所示,图 4 中第 1 行为 2 组旋转与尺度变换图像对,第 2 行为 2 组视点变换图像对,均来自牛津大学标准仿射变换数据集;第 3 行为 2 组光照变换图像对,来自 Webcam 数据集。使用文献[20]中的最近邻方法进行特征描述

子配准,对于每张图像,提取最好的 1000 个特征点测试匹配性能。对于某类特征描述子,使用该特征描述子正确匹配的特征点数和所有特征点提取方法匹配的特征点数之比<sup>[18]</sup>(match score, 匹配分数)来评估匹配性能。本文同时评估各个特征描述子在旋转、尺度、视点、光照变化下的匹配性能和各个特征描述子的匹配效率。部分匹配结果如图 5 所示,图 5 左边所示为 CNN-SIFT 特征描述子的匹配效果图;右边所示为 SIFT 特征的匹配效果图。图 5 第 1 行为部分旋转和尺度变换匹配结果图;第 2 行为部分视点变换匹配结果图;第 3 行和第 4 行为部分自然光照变换匹配结果图。



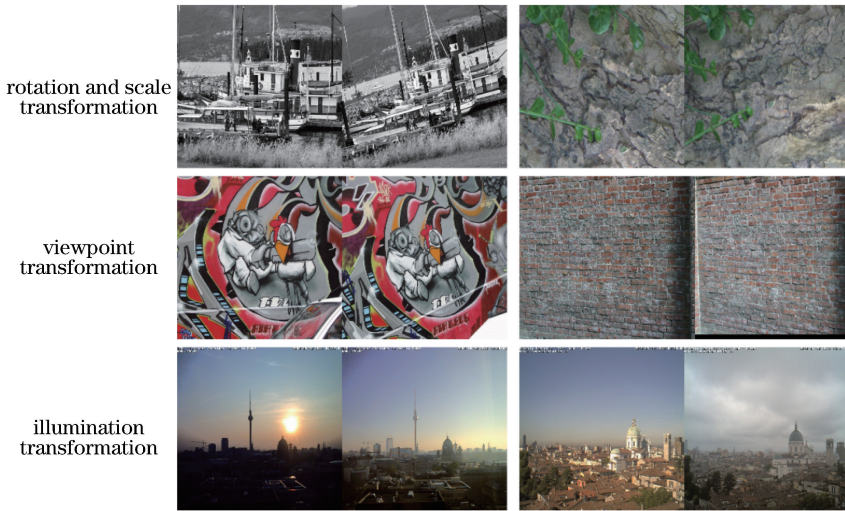


图 4 部分用于测试特征描述子匹配性能的图像对

Fig. 4 Partial atlases used to test matching performance of feature descriptor

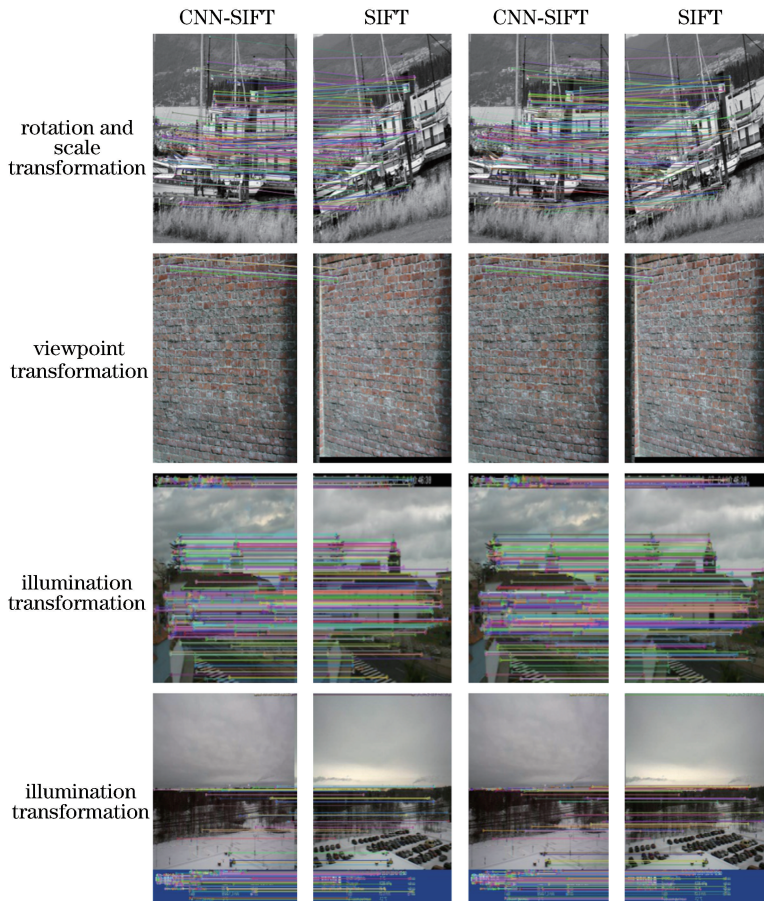


图 5 CNN-SIFT 特征描述子和 SIFT 特征描述子在仿射变换情况下的部分匹配结果对比图

Fig. 5 Comparison of partial matching results between CNN-SIFT feature descriptor and SIFT feature descriptor in affine transformation

### 3.2 旋转和尺度变换

利用牛津大学仿射变换标准数据集测试各类特征描述子在旋转、尺度变换情况下的匹配性能,结果

如图 6 所示。从图 6 可以看到,FREAK 特征描述子在旋转和尺度变换情况下表现较差,其他几个特征描述子的匹配效果十分接近,SIFT 特征描述子在旋转

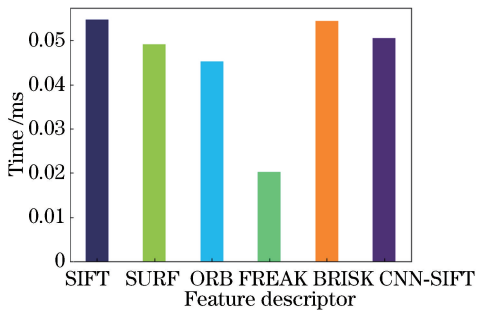


图6 各特征描述子在旋转和尺度变换下的匹配效果  
Fig. 6 Matching performance of feature descriptors in rotation and scale transformations

和尺度变换情况下匹配性能最优,其次是 BRISK 特征描述子, CNN-SIFT 特征描述子的匹配性能与 SURF 特征描述子十分接近,略弱于 SIFT 和 BRISK 特征描述子,但仍具有良好的旋转和尺度不变性。这主要是因为 CNN-SIFT 特征描述子仍然使用 SIFT 特征点提取算法,而 SIFT 特征点提取算法具有良好的旋转和尺度不变性,并且用于特征描述子降维的卷积神经网络能够保持 SIFT 特征描述子的优势。

### 3.3 视点变换

同样使用牛津大学仿射变换标准数据集测试视点变换情况下各个特征描述子的匹配效果,结果如图7所示。从图7可以看到, SIFT 特征描述子的匹配效果最好,而 BRISK 特征描述子在视点变换情况下表现很差, ORB 特征描述子效果最差。该结果充分说明,对于不同的仿射变换,各个特征描述子的匹配效果差别很大,对于不同仿射变化,有些特征描述子的稳健性较差,而 CNN-SIFT 特征描述子在视点变换的情况下仍然具有良好的匹配性能,匹配效果和旋转、尺度变换下匹配效果类似,虽然略逊色于 SIFT 特征描述子,但是远优于其他特征描述子。

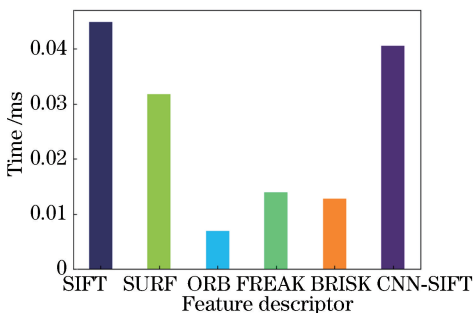


图7 各特征描述子在视点变换下的匹配效果  
Fig. 7 Matching performance of feature descriptors in viewpoint transformation

### 3.4 光照变换

在许多实际的匹配场景下,很难保证拍摄的照

片都保持相同的光照条件,因此,在实际应用中,特征点在光照变换情况下的匹配效果十分重要。Webcam 数据集包含许多光照变化下的图像对,使用该数据集测试特征描述子在光照变化情况下的匹配效果,结果如图8所示。从图8可以看到,各个特征描述子在光照变化情况下的匹配效果差距比较明显, SIFT 特征描述子依然表现最优, CNN-SIFT 特征描述子紧随其后,且显著优于其他特征描述子, ORB 特征描述子在光照变换下的匹配性最差。出现这种情况的原因主要有: 1) 卷积神经网络强大的学习能力使降维后新特征描述子的匹配性能十分接近 SIFT 特征描述子; 2) 在创建训练数据集时,为了增强数据集图像,对训练图像进行了大量不同程度的光照变换,这使得更多光照变换下的匹配点对参与卷积神经网络的训练,从而使得产生的降维特征描述子具有良好的光照变换不变性。

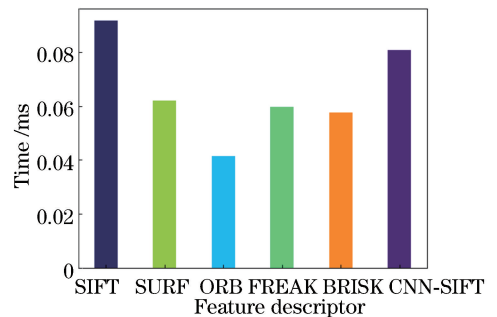


图8 各特征描述子在光照变换下的匹配效果  
Fig. 8 Matching performance of feature descriptors in light transformation

### 3.5 匹配效率

采用 Webcam 数据集测试各个特征描述子的匹配效率。使用各个特征提取方法对每幅图像提取 1000 个关键点后,采用最近邻方法完成特征点匹配,并计算各个特征描述子在 Webcam 数据集的平均匹配时间,结果如图9所示。从图9可以看到,虽然 SIFT 特征描述子在旋转、尺度以及光照变化情况下的匹配效果优于其他特征描述子,但是其匹配用时较长,不适用于一些注重匹配实时性的场景,相反, CNN-SIFT 的平均匹配时间仅为 SIFT 特征描述子的 1/6,并且优于其他特征描述子。另外,图9给出的各个特征描述子匹配所需要的时间,但没有包含 CNN-SIFT 卷积神经网络计算部分所需的时间。包含卷积神经网络计算时间之后, CNN-SIFT 在 Webcam 数据集上的测试结果为 2.4 ms,相对于 SIFT (7.3 ms)、SURF (5.4 ms)、ORB (3.7 ms)、Freak (3.9 ms) 和 BRISK (2.6 ms),该结果仍然最优。

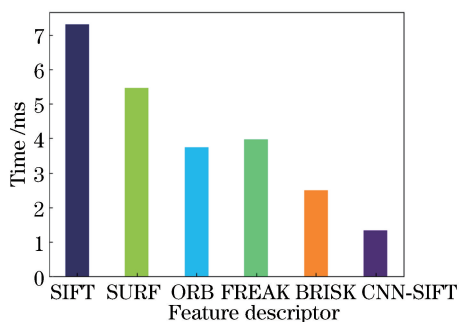


图9 各个特征描述子在 Webcam 数据集上的平均匹配时间

Fig. 9 Average matching time of feature descriptors in Webcam dataset

## 4 结 论

针对传统 SIFT 特征点描述子匹配效率低的问题,结合深度学习技术,提出了一种基于监督学习的深度学习降维模型,并详细介绍了训练数据集创建和网络训练的过程。利用所提模型对 SIFT 特征提取方法得到的描述子进行降维,得到了一种新的描述子——CNN-SIFT。使用牛津大学仿射变换数据集和 Webcam 数据集分别测试 CNN-SIFT 特征描述子的旋转和尺度变换、视点变换、光照变换下的匹配性能以及匹配效率,并与 SIFT<sup>[3]</sup>、SURF<sup>[4]</sup>、BRISK<sup>[6]</sup>、ORB<sup>[7]</sup>以及 FREAK<sup>[10]</sup>等特征描述子进行对比。结果表明,当维度降至 32 维时,CNN-SIFT 特征描述子在旋转、尺度、视点、光照变换下的匹配效果比较稳定,在各个仿射变换下均具有良好的匹配性能和稳健性,且匹配效率优于其他特征点描述子。本文方法不仅获得良好的仿射变换匹配性能,同时大幅提高了匹配效率,在 SFM<sup>[1]</sup> 三维重建等需要大量进行特征点配准的场景下具有良好的应用前景。

## 参 考 文 献

- [1] Wu C C. Towards linear-time incremental structure from motion [C] // 2013 International Conference on 3D Vision, June 29-July 1, 2013, Seattle, WA, USA. New York: IEEE, 2013: 127-134.
- [2] Guo M, Hu L L, Li J. Local point cloud reconstruction of ceramic-bowl-surface defect based on multi-image sequences [J]. Acta Optica Sinica, 2017, 37(12): 1215002.  
郭萌, 胡辽林, 李捷. 基于多幅图像的陶瓷碗表面缺陷的局部点云重建 [J]. 光学学报, 2017, 37(12): 1215002.
- [3] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [4] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features (SURF) [J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.
- [5] Ke Y, Sukthankar R. PCA-SIFT: a more distinctive representation for local image descriptors [C] // 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 27-July 2, 2004, Washington, DC, USA. New York: IEEE, 2004: 8161522.
- [6] Leutenegger S, Chli M, Siegwart R Y. BRISK: binary robust invariant scalable keypoints [C] // 2011 International Conference on Computer Vision, November 6-13, 2011, Barcelona, Spain. New York: IEEE, 2011: 2548-2555.
- [7] Rublee E, Rabaud V, Konolige K, et al. ORB: an efficient alternative to SIFT or SURF [C] // 2011 International Conference on Computer Vision, November 6-13, 2011, Barcelona, Spain. New York: IEEE, 2011: 2564-2571.
- [8] Rosten E, Drummond T. Machine learning for high-speed corner detection [M] // Leonardis A, Bischof H, Pinz A. Lecture notes in computer science. Berlin, Heidelberg: Springer, 2006, 3951: 430-443.
- [9] Calonder M, Lepetit V, Strecha C, et al. BRIEF: binary robust independent elementary features [M] // Daniilidis K, Maragos P, Paragios N. Lecture notes in computer science. Berlin, Heidelberg: Springer, 2010, 6314: 778-792.
- [10] Alahi A, Ortiz R, Vanderghenst P. FREAK: fast retina keypoint [C] // 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 16-21, 2012, Providence, RI, USA. New York: IEEE, 2012: 510-517.
- [11] Xu P F, Jiang Z L, Zhao Y, et al. Feature point extraction and matching algorithm of smooth surfaces without chromatic aberration [J]. Acta Optica Sinica, 2018, 38(4): 0451002.  
许鹏飞, 姜兆亮, 赵阳, 等. 无色差光滑曲面特征点的提取及匹配算法 [J]. 光学学报, 2018, 38(4): 0451002.
- [12] Chen S, Yang T, Sun S Y. Feature point matching algorithm for fusion of color invariants and SURB detection [J]. Laser & Optoelectronics Progress, 2018, 55(5): 051007.  
陈树, 杨天, 孙顺远. 融合彩色不变量和 SURB 检测

- 的特征点匹配算法[J]. 激光与光电子学进展, 2018, 55(5): 051007.
- [13] Huang H, He K, Zheng X L, *et al.* Spatial-spectral feature extraction of hyperspectral image based on deep learning[J]. *Laser & Optoelectronics Progress*, 2017, 54(10): 101001.  
黄鸿, 何凯, 郑新磊, 等. 基于深度学习的高光谱图像空-谱联合特征提取[J]. 激光与光电子学进展, 2017, 54(10): 101001.
- [14] Shu C X, He Y T, Sun Q K. Point cloud registration based on convolutional neural network[J]. *Laser & Optoelectronics Progress*, 2017, 54(3): 031001.  
舒程珣, 何云涛, 孙庆科. 基于卷积神经网络的点云配准方法[J]. 激光与光电子学进展, 2017, 54(3): 031001.
- [15] Verdie Y, Yi K M, Fua P, *et al.* TILDE: a temporally invariant learned detector[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 5279-5288.
- [16] Simo-Serra E, Trulls E, Ferraz L, *et al.* Discriminative learning of deep convolutional feature point descriptors [C] // 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 118-126.
- [17] Yi K M, Verdie Y, Fua P, *et al.* Learning to assign orientations to feature points [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 107-116.
- [18] Yi K M, Trulls E, Lepetit V, *et al.* LIFT: learned invariant feature transform[M]//Leibe B, Matas J, Sebe N, *et al.* Lecture notes in computer science. Cham: Springer, 2016, 9910: 467-483.
- [19] Schöps T, Schönberger J L, Galliani S, *et al.* A multi-view stereo benchmark with high-resolution images and multi-camera videos [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2538-2547.
- [20] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(10): 1615-1630.