

基于改进的图像局部区域相似度学习架构的 图像特征匹配技术研究

惠国保¹⁾ 童一飞²⁾ 李东波²⁾

¹⁾(中国电子科技集团公司第二十研究所电子信息网络实验室 西安 710068)

²⁾(南京理工大学机械工程学院 南京 210094)

摘 要 该文在 AdaBoost 算法的基础上提出了一种图像局部区域相似度的学习架构,利用该架构训练图像局部特征来获得低维数、独特的特征描述子,以实现图像局部区域高精度地匹配。所提学习架构通过学习图像局部区域相似性得到一组非线性弱学习器对图像局部特征进行描述;同时,在响应函数组合形式和弱学习器权重优化配置方面,针对浮点描述子和二值描述子分别提出了新的补丁相似性度量函数作为目标函数的核函数,提高了图像特征相似性匹配效果。该学习架构不会受限于任何预定义的图像特征信息采集模式,能产生基于灰度信息或方向梯度信息的特征描述子。实验结果表明采用这种学习架构获得的特征描述子,在所有对比描述子中图像局部匹配查准率是最好的。所提学习框架能有效地配置优化描述子弱学习器,能提高图像特征描述子对图像尺度和视角变化的鲁棒性。

关键词 AdaBoost;特征描述子;图像特征匹配;图像处理;机器学习

中图法分类号 TP391 **DOI号** 10.11897/SP.J.1016.2015.01148

Image Features Matching Based on Improved Patch Similarity Learning Framework

HUI Guo-Bao¹⁾ TONG Yi-Fei²⁾ Li Dong-Bo²⁾

¹⁾(Laboratory of Electronic Information Network, No.20 Research Institute CETC, Xi'an 710068)

²⁾(Department of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210094)

Abstract A learning framework of image patch similarity to learn low-dimensionality and high-discriminative descriptor based on AdaBoost is proposed. In the framework, the representations of image patches are modeled by non-linear weak learners which are trained by AdaBoost algorithm. Meantime, in the aspect of the response function combination and the weak learners' weight optimization allocation, two new similarity functions for float point descriptor and binary descriptor respectively are proposed to use as kernel function of optimized object function to learn a similarity embedding for improving image feature matching effect. The proposed learning framework is more generalizing than others as it won't be restricted to any predefined feature-sampling model and it could encompasses intensity and director-gradient information. The results show that the proposed feature descriptor outperforms overall comparing descriptors on recall-precision performance of image local area matching. The proposed learning framework is able to effectively optimize over the descriptor filter configuration leads to boost robustness to image scale and perspective changes.

Keywords AdaBoost; feature descriptor; feature matching; image processing; machine learning

1 引言

图像局部区域匹配是判定两幅图像局部区域是否对应的一种手段. 两个图像局部区域的对应关系取决于这两块区域对同一对象投影的相似性. 利用这种对应关系可用来解决许多计算机视觉问题, 比如, 图像配准、相机标定和三维重建等. 在目标或场景识别领域, 局部匹配要比全局匹配效果更好, 因为局部匹配对阻塞 (Occlusions) 和视角变化更加鲁棒^[1]. 尽管局部区域匹配已被用来进行对象分类和目标识别, 但其精度不高、稳定性差并且对一些图像转换缺乏鲁棒性, 致使图像局部区域匹配不能被广泛应用. 所以如何构建更好的局部区域匹配算法是研究的热点.

图像局部区域匹配过程有两个步骤^[2]: 第 1 步是检测图像特征点 (又称为兴趣点, Interest Point), 一般是根据兴趣函数 (Interest Function) 或显著性函数 (Saliency Function) 的极值来确定, 通常位于显著性区域的中心位置. 第 2 步是描述特征点. 以特征点为中心的图像局部区域 (又称图像补丁, Image Patch) 里的像素灰度信息通过一定方式能被描述成该特征点独有的特征描述子, 就是表示该特征点的向量描述符. 特征描述子既能在同一幅图像上将所描述的特征点独一无二地表达的同时, 又能重复地在不同转换的图像上识别出该特征点. 所以, 特征描述既要设计地能描述特征的独特性, 又要设计地对几何和视角等变化具有不变性, 这样的特征描述子可以代表特征点进行后续视觉处理^[3]. 描述子其实就是一串向量, 从多维描述子向量空间来看, 对应的两个局部特征描述子的距离最近, 反之亦然, 据此可以用来进行目标的匹配和识别.

由于图像转换非线性的特点, 使得 SIFT^[4] 和 SURF^[5] 等局部特征描述子通过手工调节过滤器对图像局部区域响应优化. 虽然这种整合能带来性能上的提升, 但是手动调节参数很难达到性能上的最优化, 于是后来采用机器学习方法学习图像特征描述子的不变性特征. 不变性特征是指图像上某局部区域的特征在各种图像转换条件下, 具有不变性、低冗余性和独特性, 无需预先对图像分割. 不变性特征再现学习可以看作是找一个恰当的相似性度量方法, 通过该方法度量出的相似性在各种的图像转换下保持不变性^[6]. 尽管学者们^[7-9] 提出了许多的学习特征描述子的方法, 但是他们的目标都是找一个能

映射到内核特征空间的线性特征. 于是, 在构建非线性内核特征模型时就要求选择一个恰当的内核函数, 将输入的线性特征映射到高维的特征空间里. 选择一个恰当的内核函数是一个复杂而困难的工作, 却是算法里重要的一步.

随着移动手持设备对图像配准和目标识别功能的迫切需求, 同时随着图像数据规模的不断增加, 基于局部特征图像的索引和匹配不仅要求高精度, 而且也要求高效率. 为了替代运算复杂的浮点描述子, 学者们尝试采用二值描述子来描述补丁特征. 二值描述子不仅运算速度快, 而且存储空间小, 能直接使用哈希表技术进行最近邻搜索^[10-11] (Nearest Neighbor Searching), 并且通过汉明距离能很快地判断是否匹配. 二值描述子通常有两种构建方法: 一种是对浮点描述子二值化; 另一种是在原图像特征局部区域上映射采样模式后, 随机提取采样点对灰度值来构建二值描述子. 无论是哪种方法, 虽然都大大提升了描述子的计算速度和匹配速度, 但是这些二值描述子的性能往往不及浮点描述子.

本文提出了一种对图像补丁描述子相似性学习的架构, 通过机器学习得到了低维数、高独特性的描述子. 该学习架构的具体思路是通过 AdaBoost 学习得到一组非线性弱学习器, 对局部图像局部特征建模, 其中非线性弱学习器是通过 AdaBoost 方法得到的; 同时, 在响应函数组合形式和弱学习器权重配置优化方面, 提出了新的补丁相似性度量函数作为目标函数的核函数, 改进优化效果. 该学习方法不会受限于任何一种预定义的采样模式, 并且比其他的学习框架更加广义, 因而该学习框架能产生基于灰度信息和基于梯度信息的特征描述子. 训练的规模根据训练样例的数量线性增加, 使之易于进行大规模的训练来提高描述子匹配的精度. 鉴于二值描述子占用内存少、计算速度快的特点, 所提学习架构可扩展到二值情形, 以此来学习独特的、紧凑的二值描述子. 对于二值描述子的每一维, 通过学习得到一个与之对应的、与 Adaboost 强分类器形式相同的哈希函数, 也就是一个弱分类器线性组合的符号函数. 所得的二值描述子在可与浮点描述子或量化描述子性能媲美的情况下, 减少了存储空间和计算成本.

2 相关研究

近些年提出了许多图像局部区域的描述方法. 如 SIFT^[4]、SURF^[5]、GLOH^[12]、DAISY^[13]、基于拉

普拉斯的局部特征描述子^[14]、颜色仿射变换下的局部特征描述子^[15]等. 在一份关于描述子算法性能调查报告^[1]中指出 SIFT 描述子在各种图像转换条件下鲁棒性是最好的, 该算法已被应用于解决许多计算机视觉问题. SIFT 描述子是由 128 维浮点数构成, 占用内存空间较大, 且计算速度较慢, 很难应用于一般场合中.

为了解决这个问题, 许多学者提出了改进 SIFT 算法的方法, 这些方法大部分需要手动调节参数, 但是它们通常适用于关键点的匹配, 而且不能产生大尺度关键点匹配所要求的短小描述子^[8]. Athitsos^[16]从学习图像特征的角度优化改进 SIFT 类描述子, 提出了从一个训练图像集中学习图像相似性的方法, 这启发学者们采用机器学习来构建图像特征描述子. 值得注意的是, 一些基于相似敏感性哈希技术 (Similarity-Sensitive Hashing) 和局部敏感性哈希技术 (Locality-Sensitive Hashing)^[17-18] 的算法都试图找到一组有效的二值表征高维特征数据, 并且使这些高维特征数据的相似性能保留在汉明空间里. 已经有学者采用这些方法来构建全局图像特征描述子和特征袋 (Bag-of-Feature)^[19-20].

除此之外, 还有许多方法采用其他方式构建特征二值描述子, 如采用像素灰度比较来构建二值描述子, 包括 BRISK^[21]、ORB^[22]、FREAK^[23] 和 BRIEF^[24] 等, 还有采用胡夫曼 (Huffman) 译码^[25] 和乘积量化^[26] 来建立有向梯度二值描述子, 也有基于归一化梯度的直方图构建特征边缘二值描述子^[27]. 尽管这些二值描述子匹配效率提高了, 但是因为手工设计的, 构建的二值描述子特征描述性不高, 也不如浮点描述子匹配精度高. 为了解决这个问题, 最近有学者采用机器学习来构建快速、鲁棒的特征描述子. 用机器学习来学习图像的相似性, 能改进图像特征描述子匹配的效率和精度. 采用机器学习方法构建特征描述子的方式主要包括无监督学习和监督学习.

无监督学习 (Unsupervised Learning) 能得到紧凑的二值描述子, 是因为该描述子的汉明距离反映的是输入图像的相似性, 相似性度量大多在最初所输入图像的量化空间内进行. 首先采用语义哈希 (Semantic Hashing)^[28] 训练一个多层的神经网络, 然后在相似训练样本中采用谱哈希 (Spectral Hashing)^[29] 最小化汉明距离的期望值和对样本关系的优化. Kulis 与 Darrell^[30] 及 Norouzi 与 Fleet^[31] 都是通过无监督学习方式构建特征的二值描述子, 而且所得描述子的汉明距离很好地逼近到最初欧几

里得的距离.

监督学习 (Supervised Learning) 得到的特征空间能用来描述图像特征, 它采用带标记样本对学习邻近图像特征的关系信息. Jain 等人^[32] 使用一对带标记样本的距离作约束条件来学习优化一个马哈朗诺比斯距离 (Mahalanobis Distance), 以此来判断邻近特征的相似度; Strecha 等人^[8] 使用特征学习器线性独特性来学习邻近特征组合耦合信息; Wang 等人^[33] 通过迭代和顺序优化策略过滤掉最误差相似的邻近图像特征, 然后采用半监督 (Semi-supervised) 学习邻近特征的相似性.

最近学者们注意到学习特征描述子不光要注重权重优化, 也要注重目标函数选择和弱分类器的组合配置^[7,34]. Brown 等人^[7] 从不同特征选择和梯度特征的权重配置进行优化, 但优化难度大, 于是 Simonyan 等人^[34] 提出了凸集优化策略对其优化, 但优化条件比较苛刻, 过程较复杂. 对描述子权重和响应函数组合形式联合优化是个困难的问题, 因为响应数在图像补丁上的组合形式有很多种. 实际上, 即使对于非常小的图像补丁, 响应函数在其上的作用次数都会达到上百万, 但这个问题很适合通过 Boosting 和 Schapire^[35]. 尽管 Boosting 用到贪婪优化, 但仍然是构建一个高精度分类器的有效方法, 它能把输入数据映射到一个高维的特征空间.

本文介绍的使用 AdaBoost 训练方法学习得到独特的紧凑的二值描述子的方法, 是受到 Boosted SSC^[18]、L-BGM^[36]、BinBoost^[37] 等描述子算法的启发. 最初是 Boosted SSC 使用 Boosting 学习图像的相似性度量, 但 Boosted SSC 只考虑了输入图像的线性投影, 所以会产生维数很高的描述子. L-BGM 对 Boosted SSC 进行了扩展, 使用 Boosting 对图像显著区域学习得到复杂的非线性图像局部特征描述子; BinBoost 是对 L-BGM 或 Boosted SSC 的二值化, 因为目标优化过程复杂而难以适用, 而且这两个学习算法的泛化性差, 所得描述子在训练集和测试集一致性要求很高的情况下才可用. 针对这些问题, 本文分别对浮点描述子和二值描述子重新设计相似性度量函数和优化目标函数, 在简化目标函数计算和提高相似度计算效率的同时, 提高了学习算法的泛化性能.

3 学习图像特征描述子

给定一个灰度图像补丁 x , 可以用一个 D 维的

向量独特地描述它,即将图像补丁特征数据映射到 D 维的向量空间 Y :

$$Y: x \in X \rightarrow [c_1(x), \dots, c_D(x)] \quad (1)$$

其中 X 是图像补丁的集合(包括训练补丁和测试补丁),向量 $C(x)=[c_1(x), \dots, c_D(x)]$ 又称为补丁 x 的描述子. 可以用一组响应函数 $\{h_m(x)\}_{m=1}^M$ (其中 $h_m \in \{0,1\}$) 作用于补丁来获得描述子,则补丁 x 的描述子可以表示为

$$C(x) = A^T H^T(x) \quad (2)$$

其中 A^T 是 $D \times M$ 矩阵,记 $A=[\bar{a}_1, \dots, \bar{a}_D]$, \bar{a}_d 为描述子第 d 维对应的弱分类器系数向量; $H(x)=[h_1(x), \dots, h_M(x)]$ 是由响应函数构成的向量. 所以描述子第 d 维的值为 $c_d(x) = \bar{a}_d H^T(x) = \sum_{m=1}^M a_{d,m} h_m(x)$. 描述子的每一维数对应补丁的一个特征,增加描述子的维数能增加自身的独特性. 本文利用 AdaBoost 方法学习到与描述子每一维数相关的图像特征分类器(弱学习器)的最佳选择和权重系数. 下面介绍如何通过 AdaBoost 学习算法获得分类器 $H(x)$ 和系数矩阵 A .

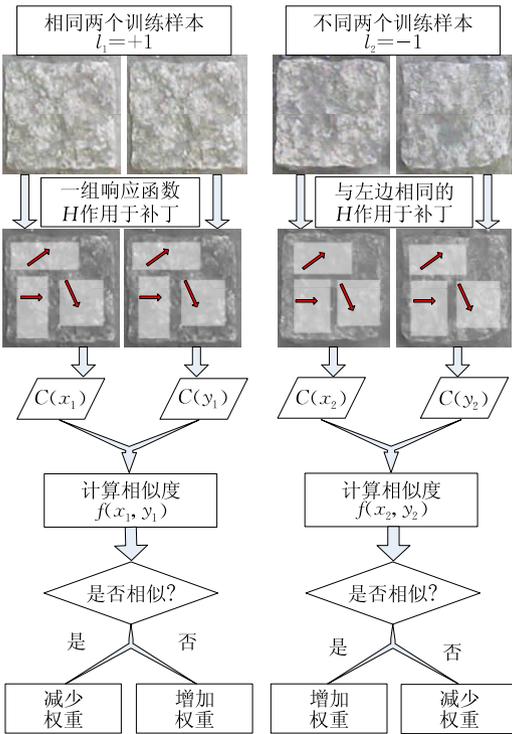


图 1 相似性度量流程示意

使用 AdaBoost 算法,首先需要确定训练样本. 这里训练样本是带标记的成对补丁对 $\{(x_j, y_j), l_j\}_{j=1}^N$, 其中 $l_j \in \{-1, 1\}$, $l_j = -1$ 表示补丁对是两个不同的补丁, $l_j = 1$ 表示补丁对的两个补丁是相似的. Friedman 等人^[38]从统计学的角度对 AdaBoost

算法进行了研究,指出 AdaBoost 算法实际上是一种对特殊的指数型损失函数优化的方法. 所以可以从训练样本对相似性指数损失最小化的优化过程中提炼出分类器和相关系数,于是学习目标函数为

$$O = \sum_{i=1}^N \exp(-l_i f(x_i, y_i)) \quad (3)$$

其中 $f(x_i, y_i)$ 是补丁对相似性函数. 本文采用描述子表示图像补丁特征,所以两个图像补丁之间的相似性度量就是对应的两个描述子之间的相似性度量. 相似性度量流程示意图如图 1 所示. 该图反映的是学习 M 个响应函数中一轮循环中的对前两个不同样本对的学习过程(这里假设训练样本对前两个样本相似性是不同的).

要度量两个图像补丁之间的相似性,要经历两个阶段:一个是由图像补丁到描述子,这是在图像补丁上操作弱学习器的响应函数得到的;另一个是两个描述子相似性比较,这是根据相似性度量函数实现的. 所以这样得到的描述可以被视为一个二层神经网络^[28]. 因为描述子是对图像局部特征的间接描述,所以每个阶段都会产生误差. 为了将误差减少到最小,选择合适的响应函数和相似性函数很重要.

3.1 特征响应函数

与 Trzcinski 等人^[37]类似,这里选择基于梯度的响应函数作为弱学习器,它考虑了整个图像区域的灰度梯度的方向. 基于梯度的弱学习器由 3 个参数构成:在整个补丁 x 的任意位置上可确定的一个矩形区域 R 、一个方向 e 和一个阈值 T . 所以响应函数定义为

$$h(x; R, e, T) = \begin{cases} 1, & \phi_{R,e}(x) \leq T \\ 0, & \text{否则} \end{cases} \quad (4)$$

其中 $\phi_{R,e}(x) = \sum_{m \in R} \xi_e(x, m) / \sum_{e' \in \Phi, m \in R} \xi_{e'}(x, m)$, 并且 $\xi_e(x, m) = \max(0, \cos(e - o(x, m)))$, 其中 $o(x, m)$ 是补丁 x 在 m 处的像素梯度方向. 方向 $e \in \Phi = \{0, \frac{2\pi}{q}, \frac{4\pi}{q}, \dots, (q-1)\frac{2\pi}{q}\}$, q 是梯度方向 Bin 的个数. Trzcinski 提到利用积分图像计算的效率很高. 下面针对浮点描述子和二值描述子分别提出新的图像补丁相似性度量函数和学习目标函数.

3.2 学习特征的浮点描述子

Boosted SSC 方法^[18]根据响应函数在两个图像局部特征上作用结果乘积的权重和,提出了一种相似性函数:

$$f_{\text{BSSC}}(x, y) = \sum_{d=1}^D \partial_d h_d(x) h_d(y) \quad (5)$$

其中 ∂_d 表示描述子的第 d 维的权重. 由此得到的描述子为浮点型, 而且描述子的每一维对应一个响应函数.

将式(5)代入式(3)得到目标函数:

$$O_{\text{BSSC}} = \sum_{i=1}^N \exp\left(-l_i \sum_{d=1}^D \partial_d h_d(x_i) h_d(y_i)\right) \quad (6)$$

实际上响应函数空间可能是无限大的, 这就给优化 O_{BSSC} 带来了困难, 不过这个问题非常适合用 Boosting 解决, 但是 Boosting 是个贪婪算法, 而它又能有效地构建一组高度精确的弱学习器. 由于 Boosting 贪婪的特性, 使得构建的弱学习器因为过度冗余而效率低下. 这里通过修改 Boosted SSC 的相似性函数, 采用 AdaBoost 算法学习出低维数、独特的图像特征浮点描述子, 将该浮点描述子构建方法标记为 AdaBoost-Float-point.

本文也是使用式(2)计算描述子, 因为系数矩阵 \mathbf{A} 一般为浮点数, 所以得到的描述子也为浮点描述子. 要度量两个图像补丁 (x, y) 的相似性, 可以计算两个补丁分别对应的描述子的关联度, 于是相似性函数定义为

$$f_F(x, y) = \frac{\sum_{d=1}^D (c_d(x) - \bar{C}(x))(c_d(y) - \bar{C}(y))}{\sigma_x \sigma_y} \quad (7)$$

其中 $\bar{C}(x)$ 和 σ_x 分别为补丁 x 描述子维数上的均值和方差, 即

$$\bar{C}(x) = \sum_{d=1}^D c_d(x) / D, \sigma_x = \sqrt{\sum_{d=1}^D (c_d(x) - \bar{C}(x))^2},$$

其中补丁 y 的相关表达式与 x 表达类似. 由于不是直接对描述子的某一维操作, 使得该相似性度量方法比基于维数值度量具有更高的鲁棒性.

将式(7)代入式(3)得到目标函数为

$$O_F = \sum_{i=1}^N \exp\left[-l_i \frac{\sum_{d=1}^D (c_d(x_i) - \bar{C}(x_i))(c_d(y_i) - \bar{C}(y_i))}{\sigma_{x_i} \sigma_{y_i}}\right] \quad (8)$$

该目标函数是凸函数, 所以在优化过程中能找到最优解. 在优化 O_F 的过程中, 使用两步学习策略^[36]: 首先在训练样本上, 使用 AdaBoost 最小化式(6)以获得 M 个弱学习器 $\{h_m\}_{m=1}^M$; 然后使用梯度下降法最小化 O_F 以获得每个弱学习器的权重. 通过算法 1 对目标函数的不断优化, 最终会学习出配置优化的特征弱学习器非线性组合. 算法 1 中, 使用 AdaBoost 算法学习得到一组响应函数和相应的权

重系数, 以此来构建补丁的特征描述子. 本算法是对 Boosted SSC 算法的扩展, 将其直接根据补丁像素水平来预估相似性改进为用补丁描述子这个间接向量来预估相似性. 算法中 ThresholdRate 函数是文献[18]中的算法 4, TP 和 FP 分别表示相似的样本预测为相似和不相似的期望.

算法 1. 描述子学习算法.

输入: 一组补丁对样本 $P\{(x_i, y_i), l_i\}_{i=1}^N, l_i \in \{-1, +1\}$.

$l_i = +1$ 表示补丁对两个补丁是相似的, $l_i = -1$ 表示补丁对两个补丁是不同的

输出: 一组响应函数 $h_m: H \rightarrow \{0, a_m\}, m = 1, \dots, M$. H 为响应函数空间

1. 初始化权重, $w_{1,i} = 1/N$

2. FOR $d=1, \dots, D$.

3. 记 $W^p := \sum_{i: l_i = +1} w_{d,i}, W^n := \sum_{i: l_i = -1} w_{d,i}$.

4. FOR $m=1, \dots, M$.

5. 设 $f_m(x_i, y_i) \leftarrow \{f_F \parallel f_B\}$.

6. 关于 f_m 每一个可行的阈值 T_m , 使用 ThresholdRate($P, f_m, w_{d,m}$) 计算出 FP_d, TP_d .

7. 记 $r_d^{(AB)}(T_m) := 2(TP_d - FP_d) + W^n - W^p$.

8. FOR $m=1, \dots, M$ 循环结束. 选择最佳的阈值 $T_d := \arg \max_{d,j} r_d^{(AB)}(T_m)$.

9. 选择最佳的弱分类器系数向量 \bar{a}_d , 使目标函数 O_F 或 O_B 最小化.

10. 如果 $\{a_{d,m}\}_{m=1}^M < 0$, 跳出 FOR $d=1, \dots, D$ 循环.

11. 更新权重:

$$w_{d+1,i} := \frac{w_{d,i} \exp\left(-l_i \sum_{m=1}^M (a_{d,m} f_{d,m}(x_i, y_i))\right)}{\sum_{i=1}^N \left(w_{d,i} \exp\left(-l_i \sum_{m=1}^M (a_{d,m} f_{d,m}(x_i, y_i))\right)\right)}$$

3.3 学习特征的二值描述子

上面得到的是浮点描述子, 为了学习获得二值描述子, 需要设计新的相似性函数以适应 D 维汉明距离空间. 这里采用两个二值描述子的汉明距离来衡量它们的相似性:

$$f_B(x, y) = \sum_{d=1}^D |c_d(x) - c_d(y)| \quad (9)$$

其中 $c_d(x) = \text{sign}(\bar{a}_d \mathbf{H}_d^\top(x))$, 并且 $H_d(x) = [h_{d,1}(x), \dots, h_{d,M}(x)]$ 是 M 个弱学习子, \bar{a}_d 为权重系数向量. 目标函数定义为

$$O_B = \sum_{i=1}^N \exp\left(-\gamma l_i \sum_{d=1}^D |c_d(x_i) - c_d(y_i)|\right) \quad (10)$$

该目标函数类似于文献[37], 但相似性度量方法不同. 文献[37]对相似性的计算是先计算两个相比较的二值描述子的相应位值乘积, 再将每个乘积相加得到的. 这样操作将对应两个位的值为零的位

归为不相似,从而降低了正确相似的衡量,而且构造的目标函数因为是不连续且非凸而难以优化.这里根据两个二值描述子的汉明距离评估两个描述子的相似性是比较客观的,而且该目标函数简单且易于优化学习,学习算法伪代码如算法 1 所示.

因为 AdaBoost 是一个指数损失最小、样本对响应边际值最大化的有约束的梯度下降搜索过程^[39],所以在使用 AdaBoost 算法优化该目标过程中,在迭代到 d 步时,为了使 O_B 最小,需要最大化每个弱学习器的权重关系.实际上,迭代到 d 步时, \bar{a}_d 和 H_d 都可以通过下式获得:

$$\arg \max_{a_d, H_d} \sum_{i=1}^N l_i W_d(i) |c_d(x_i) - c_d(y_i)| \quad (11)$$

其中 $W_d(i) = \exp\left(-\gamma \sum_{d'=1}^{d-1} |c_{d'}(x_i) - c_{d'}(y_i)|\right)$ 是一个样本权重系数.这表示此前迭代错误分类的样本在本次迭代中将获得更高的权重,而正确分类的样本在本次迭代中权重降低.因为式(11)包含 $\sum_{d=1}^D |c_d(x) - c_d(y)|$ 不方便计算,又因为 $c_d(x_i), c_d(y_i)$ 都是二值元素,所以使用 $\sum_{d=1}^D (c_d(x) - c_d(y))^2$ 来替换它.因为该目标函数是凸函数,所以有最优解.学习算法迭代一次后需要提供 γ 的值,因为根据 AdaBoost 规则,搜索到 $c_1(x)$ 和 $c_1(y)$ 后需要设置该参数,这里与文献[37]类似,定义 γ 为 $\gamma = \nu \cdot 0.5 \log(1 + r_1) / (1 - r_1)$,其中 $r_1 = \sum_{i=1}^N W_1(i) l_i |c_1(x_i) - c_2(y_i)|$, ν 是一个收缩参数,用来调整优化过程,取 $\nu = 0.4$.基于补丁对的相似性学习,得到构建图像特征二值描述子的优化配置的弱学习器非线性组合,采用该学习成果能对任一图像补丁构建能表征特征的二值描述子,这里将该二值描述子构建方法标记为 AdaBoost-Binary.

4 描述子评估方法建立

为了衡量所提学习架构产生的描述子的性能,这里使用对比实验方法对其进行评估.首先,简单介绍一下图像特征描述子的评估方法;然后,通过测试选择适当的参数来初始化学习算法程序.本文方法旨在改进二值描述子性能,所以,后面的实验中主要评估本文方法得到的二值描述子的性能.

4.1 描述子评估方法介绍

本文在两个数据集上对特征描述子的性能进行

评估.第 1 个是经典的描述子性能评估数据集 Mikolajczyk,其方法的介绍请见文献[1].第 2 种评估方法用到了 Brown 数据集^[7].下面简单地介绍一下如何在 Brown 数据集上评估描述子. Brown 数据集包含 3 组取自不同目标的图像补丁集: Liberty、Notre Dame、Yosemite,每组包括 40 万个 64×64 补丁,这些补丁是以 DoG 检测子检测出的兴趣点为中心.这些补丁包含尺度和视角变化,他们都是取自兴趣点周围不同高斯平滑后的区域,他们之间的对应关系是根据多视立体(Multi-View Stereo)算法得到的,这个数据集里补丁包含了各种视角和灯光变化.每个数据集里实际对应的补丁数分别是 10 万、20 万、50 万,其中 50% 能匹配成对.实验里,使用 32×32 的降采样补丁,数据里的 20 万补丁用来训练得到描述子,剩余的补丁里选出 10 万作为测试集.评估的结果是用 ROC 曲线和 95% 误差率(95% error rate)进行表述. ROC 曲线描述的是对应补丁对认为是相似的比率(True Positive Rate)与对应补丁对认为是不相似的比率(False Positive Rate)之间的关系^[12]. 95% 误差率指当 95% 正确匹配下不正确匹配所占的比率^[7].

4.2 学习算法参数选择

不像其他的图像特征学习算法^[18,36-37],本文基于 AdaBoost 描述子学习算法是基于一般优化方法定义的,因此有几个参数需要设定.所提二值描述子有 3 个主要参数:响应函数梯度方向 Bin 的个数 q , 响应函数的个数 M 和最终描述子的维数 D .这 3 个参数都与最终所得到的描述子的性能和复杂度有关.为了对这 3 个变量选择恰当的值,且目标最优值具有普遍性,采用几种不同的组合形式进行测试.选择的训练集和测试集组合与文献[7]类似.训练集和测试集可以组合为 4 种情况: Yosemite-Notre Dame、Yosemite-Liberty、Notre Dame-Yosemite、Notre Dame-Liberty.每个组合的第 1 个数据集是训练集,第 2 个是测试集.因为本文主要目标是提升二值描述子性能,所以这里参数选择对描述子性能影响评估测试是对所提二值描述子(AdaBoost-Binary)进行的.其实这些参数只是对学习架构的 AdaBoost 算法和优化过程进行配置,所以对由此学习架构产生的浮点描述子也适用.

因为本文所选择的参数与文献[37]相近,所以这里根据文献[37]中提到的 3 个参数最优值($q=8$, $M=128$, $D=64$)来设计几种情形进行测试.除了文献[37]中取的最优的 3 个参数组合作为一种情形外,

还需根据每个变量最优值的前后两个值,组合出新的测试情形,新的一组测试情形中必须保证有两个变量是最优值.例如,在文献[37]中, q 的最优值前后两个值分别是 4 和 12,于是就得到两组参数设置,分别为 $q=4, M=128, D=64$ 和 $q=12, M=128, D=64$ 两种测试条件.类似的方式选择 M 和 D ,总共可得到 7 种测试情形.本次测试使用 ROC 曲线评测结果,测试结果如图 2 所示.图 2 每幅子图对应到不同的训练集和测试集组合.这里根据 Brown^[10]提出了 4 种训练集和测试集组合分别进行测试的.另外,3 个参数共产生组合有 7 种,所以总共可以得到 28 个描述子的测试结果.从 28 个测试情况来看, $q=12, M=128, D=64$ 这种参数组合测试结果比较好,而且比

较稳定,所以本文算法以此参数设置进行初始化.

从图 2 中可以看出,无论在何种训练集-测试集组合下, $q=12, M=128, D=64, q=8, M=256, D=64, q=8, M=128, D=72$ 和 $q=8, M=128, D=64$ 这 4 种情形测试结果比较接近,而且都好于其余几个. $q=12, M=128, D=64$ 这种参数组合测试结果比较稳定,而且测试效果比其他要好.在文献[37]中梯度方向 Bin 的最优个数 q 为 8,那是因为它只和描述子维数 D 组合进行测试,而且是在 95% 误差率评估标准下得到的,与本文测试情形不同.所以,本文后续的对比性实验就采用响应函数梯度方向 Bin 的个数 $q=8$ 、响应函数的个数 $M=128$ 和最终描述子的维数 $D=64$ 的参数设置.

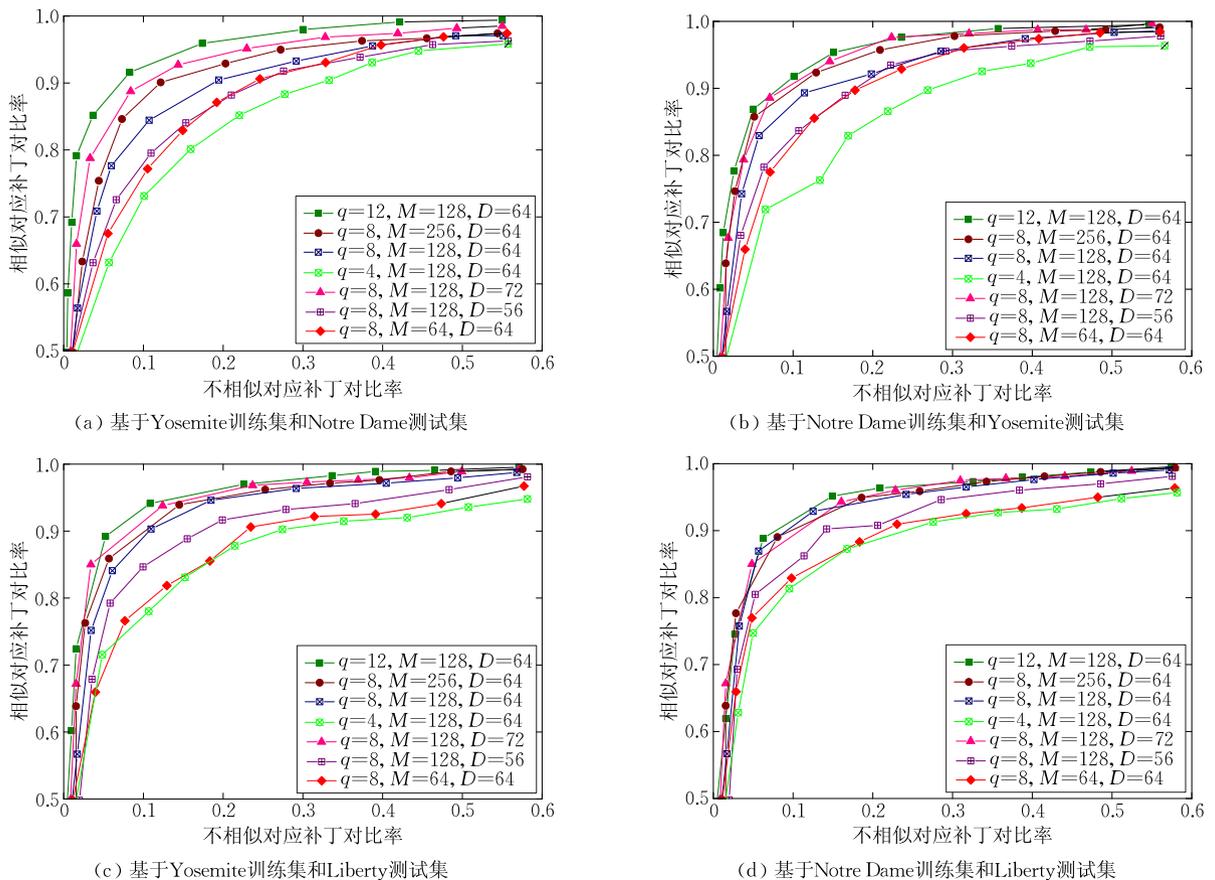


图 2 在不同参数配置下得到的描述子性能评估结果——ROC 曲线图

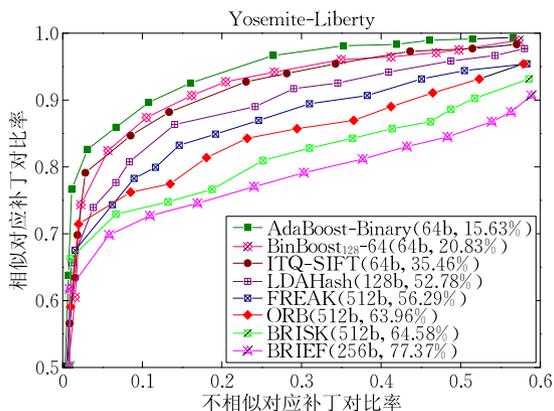
5 描述子性能对比实验

在这节里,本文将所提的描述子方法与对比描述子方法分别在 Brown 数据集^[7]和 Mikolajczyk 数据集^[12]上作性能对比实验.参与对比的描述子包括:SIFT^[4]、SURF^[5]、GLOH^[12]、DAISY^[13]、二值 LDAHash 描述子^[8]、二值化 SIFT 得到的 ITQ 描

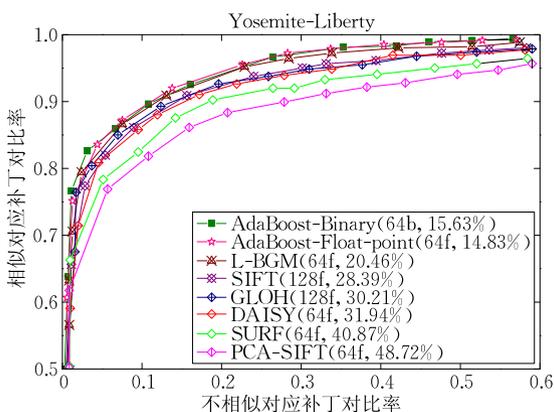
述子^[40]、Boosting 学习描述子 L-BGM^[36]、快速二值描述子 BRIEF^[24]、ORB^[22]、FREAK^[23]、BRISK^[21]和 BinBoost^[37].这些描述子大多在 OpenCV 上有算法代码,可以直接编译运行测试;BRIEF、SIFT 和 LDAHash 等可以从作者或者作者所在单位的主页上获得代码;BinBoost 和 ITQ 则是引用了原文作者介绍的实验结果.实验中所提描述子的学习架构参数变量按照 4.2 节介绍设置.

5.1 Brown 数据集上对比实验

根据 4.1 节介绍的描述子评估方法,在 Brown 的 3 个数据集上对本文所提的描述子进行性能评估.为了直观地观察到对比效果,将评估结果分成二进制描述子和浮点描述子.下面简单的以 Yosemite 作为训练集,以 Liberty 作为测试集,分别对不同描述子进行评估.根据第 4.2 节里参数选择测试发现,对于不同训练集-测试集组合的评估效果应该是类似的.图 3 展示了本文的 AdaBoost-Binary 和 AdaBoost-Float-point 描述子与先进的描述子方法的对比的 ROC 曲线.图 3 中二值描述子对比对象有 7 个,浮点描述子对比对象有 6 个.在浮点描述子的 ROC 曲线图里,保留了 AdaBoost-Binary 的 ROC 曲线作为参照,来反映所提的二值描述子性能提升.图 3 中每个描述子后面圆括号里,第 1 个参数是描述子的维数,第 2 个参数是 95% 误差率.



(a) AdaBoost-Binary 二值描述子评估



(b) AdaBoost-Float-point 浮点描述子评估

图 3 所提描述子与其他描述子对比实验结果——ROC 曲线

从图 3 中可以看出所提的描述子在查准率方面要好于其他描述子.所提的二值描述子比 SIFT 描述子要好,所以可以用 64 位二值描述子取代 128 位浮点描述子,而且节省内存,运算速度比浮点描述子快很多.从图中还可以看出,通过所提的 AdaBoost

学习架构得到的浮点描述子性能得到了提升,比其他浮点描述子要好,而且所提二值描述子和浮点描述子性能很接近.分析原因,这与学习方式的选择,学习架构参数的优化配置和训练集等因素有关.

由于描述子的性能会受到描述子维数变化的影响,所以本文对在不同维数下各描述子的 95% 误差率进行了对比,对比结果如图 4 所示.这里主要是对二值描述子的 95% 误差率评估,为了与浮点描述子形成对比,于是加入了 SIFT 和 SURF 两个描述子.在 6 个二值描述子中,除了 AdaBoost-Binary 和 BinBoost 比 SIFT/SURF 好以外,其他 4 个二值描述子的 95% 误差率都比浮点描述子高.从图 4 中可以看出每个描述子的维数对描述子的性能是有影响的,除了 BRISK 相对稳定外,其他描述子在低维数区性能变化较大,高维数区趋于稳定.本文所提的描述子在 95% 正确匹配下,错误率在对比描述子中是最低的.还可以发现在低维数区所提描述子优势比其他描述子明显,而在高维数区与 BinBoost 和 SIFT 趋于相同.造成这种现象主要是因为描述子训练形式和算法架构的不同.

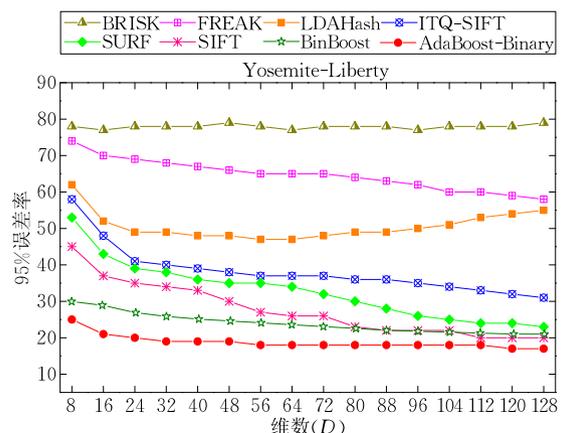


图 4 描述子不同维数下的描述子 95% 误差率实验结果

所提二值描述子不仅比一般浮点描述子要节省 4~6 倍的内存空间,而且匹配速度更快,这要得益于计算机擅长计算像汉明距离这样的简单运算,而一些浮点描述子运算却要复杂些,一般计算欧几里得距离耗时是计算汉明距离的两个数量级^[37].即使是量化后的浮点描述子也不能达到二值描述子运算速度.图 5 是几个描述子的每个补丁对匹配时间评估结果,这里的匹配时间包含了弱分类器组合在图像补丁上作用产生描述子的时间,即图像补丁到描述子生成时间.描述子的匹配时间其实就是搜索时间,即一个图像补丁在另一幅图像众多的补丁里寻找对应的补丁的过程.图 5 中的描述子匹配时间是

10 万对补丁进行匹配的平均值. 测试环境是 Lenovo Inter Core(TM)2 Duo CPU@2.2 GHz. 其中, 对浮点描述子的变量采用无符号(unsigned)运算.

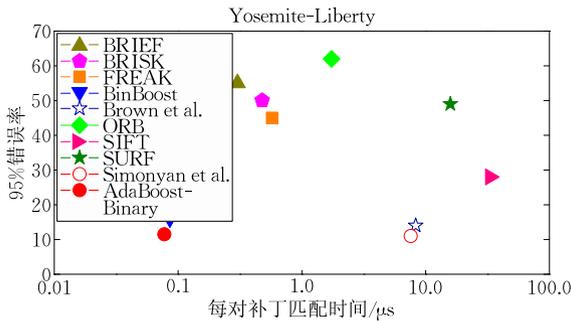


图 5 补丁对匹配时间对比评估结果

从图 5 中可以看到, 所提的二值描述子是最快的, 是浮点描述子的两个数量级. 从图 5 中还可以看到, 所提二值描述子也比其他二值描述子稍快, 如 BinBoost, 这是因为弱分类器所选用的响应函数组合形式不同, 这是由 4.2 节介绍的参数选择决定的. 所以即使对同一个图像补丁描述、产生同维数的二进制描述子, 耗用的时间也不尽相同.

另外, 本文的描述子匹配时间与其他文献, 如文献[37]的描述子匹配时间也有所不同, 分析原因有以下几个方面, 一方面是因为本文采样积分图像, 提高了弱分类器在图像补丁上的响应效率; 另外, 这里采用的计算机 CPU 能运行 POPCNT SSE4.2 指令, 调高了计算汉明距离的效率; 还有是因为训练集和测试集不同, 就是说在同样算法架构下, 使用不同的训练集获得的描述子匹配时间也不相同.

5.2 Mikolajczyk 数据集对比实验

为了测试本文所提描述子的泛化性能, 先在 Brown 数据集上进行训练, 然后在 Mikolajczyk 数据集^[12]上进行测试评估. Mikolajczyk 数据集包括结构场景图片和纹理场景图片, 而且这些图片带有不同程度的几何和光度转换. 这 6 个转换包括: 视角变化、尺度变化、图像旋转、图像模糊、亮度变化和 JPEG 压缩. 因为 Mikolajczyk 数据集里的每个图像集都各具特点, 而且许多学者^[12-13, 21-24]都用它来评估描述子的性能, 所以对比参照性很强. 这里实验对象除了本文所提的两个描述子外, 还包括公认性能很好的浮点描述子(SIFT、SURF)、最新基于手工设计的二进制描述子(BRISK、FREAK)、最新基于 Boosting 学习的二值描述子 BinBoost 以及 Mikolajczyk 提出的 GLOH, 这些描述子要么应用比较广泛要么自称效果是最好的.

这里依据 Mikolajczyk^[1]评估方法, 使用单个关键点比较描述子性能, 并使用 OpenCV 里的 SURF Hessian-based 方法来检测关键点. 每个图像对里的每幅图像检测到 1000 个关键点, 接着对每个关键点进行描述得到描述子, 然后使用穷尽搜索法来寻找匹配对. 例如, 为了找到一幅图像里的一个关键点在另一幅图像的与其匹配的关键点, 需首先根据所提的特征描述子构建方法描述两幅图像里的每个关键点, 在后续的操作中以这些描述子代替相应关键点完成特征匹配识别. 接着计算出一幅图像的特征描述子与另一幅图像里所有特征描述子的汉明距离, 最短汉明距离比上次最短汉明距离的比值小于 0.8, 则认为最短距离对应的两个描述子是匹配的, 否则认为是不匹配的. 这样得到的匹配结果根据实际特征是否是对应关系判断匹配的正确性.

本文根据识别率评估描述子算法. 识别率就是正确匹配(即匹配的是两个对应关键点)数与总匹配数的比值. 所提描述子与对比描述子关于 Mikolajczyk 8 个数据集测试结果如图 6 所示. 从图中可以看出, 所提的 AdaBoost-Binary 和 AdaBoost-Float-point 的查准率要好于其他描述子. 尽管它们的维数并不比其他描述子大, 但它们能把图像补丁特征描述为具有较高独特性描述子. 尽管如此, 在某些测试环境(亮度变化和结构图像视角变化)下, 相比对比描述子来说, 所提描述子表现不是很好, 原因是由于在不同训练条件下对描述子描述性的抵消造成的. 虽然提高学习描述子泛化性能的工作仍然具有挑战性, 但是这个测试结果表明了学习描述子在不同训练集环境下测试也能表现出很好的特征描述性.

这里在 Mikolajczyk 数据集上对所提描述子的算法效率问题再次进行探讨. 这里所说的算法效率是指在测试集上, 对测试图像补丁描述、匹配效率. 由 5.1 节可以知道, 对图像补丁描述生成描述子的过程就是将集成学习得到弱学习器组合在图像补丁上作用产生一串向量的过程. 当然图像集里的一个图像可能检索到几千个关键点, 这样可能就有相应数量的对应图像补丁, 所以对这么多的图像补丁进行描述是要消耗一定时间的. 描述完一幅图像的所有图像补丁后, 就得到了大量的描述子. 匹配是发生在图像集里的一幅图像与另一幅图像对应关键点的搜索, 是根据一幅图像里的关键点标识符(描述子)在另一幅图像里找到与其对应的关键点标识符. 这里采样的是二值型的描述子, 所以搜索过程主要是计算两个描述子的汉明距离.

这里取自 Mikolajczyk 数据集里的 Graffiti-2

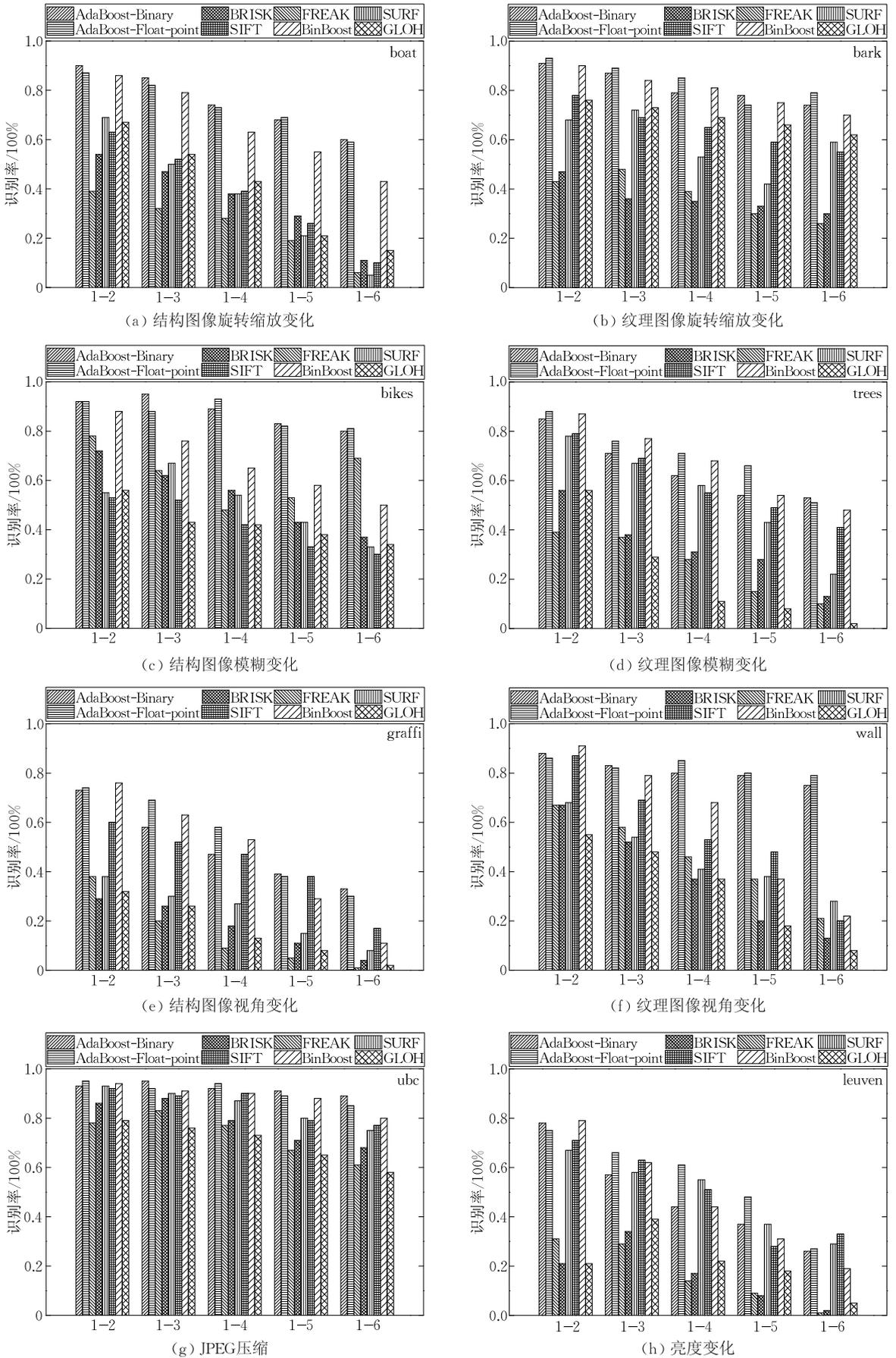


图 6 各种描述子在 Mikolajczyk 数据集上识别率测试结果

与 Graffiti-4 两幅图像进行测试. 首先检测出这两幅图像的关键点, 这里采用 SIFT 方法里的 DoG 检测子检测图像的关键点, Graffiti-2 和 Graffiti-4 分别得到 1233 和 1432 个关键点. 这里将所提方法与其他方法进行对比, 选择 BRIEF、BRISK、FREAK 和 BinBoost 作为对比方法. 然后使用每种方法分别对两幅图像的关键点进行描述, 记录每个方法下的两个图像关键点描述的平均时间. 最后对每种方法下

的两幅图像的关键点对应的描述子进行匹配, 匹配完后得到每个匹配对平均耗用的时间, 如表 1 所示. 从表 1 中可以看出, 在 Graffiti-2 和 Graffiti-4 上进行的测试, 所提的方法在补丁描述和描述子匹配上都较其他方法快, 具体原因同 5.1 节分析, 即因为弱分类器所选用的响应函数组合形式不同, 这是由参数选择决定的. 所以即使对同一个图像补丁描述、产生同维数的二进制描述子, 耗用的时间也不尽相同.

表 1 Graffiti-2(1233)/Graffiti-4(1432) 图像补丁二值描述匹配时间比较

方法	参数	描述时间/ns	匹配时间/ns
BRIEF-64	512 维, 采样点对提取规则: $(x, y) \sim G\left(0, \frac{1}{25}S^2\right)$, $S=25\sigma$, $\sigma=3.05$	33.1/34.2	48.4
BRISK	512 维, 采样模式 60 个采样点, 高斯平滑参数 $\sigma=13.67$	37.6/32.8	59.6
FREAK	512 维, 采样模式 43 个采样点, 高斯平滑参数 $\sigma=10$	38.1/38.4	62.2
BinBoost-64	64 维, 64×64 补丁块, 128 个弱学习器, 响应函数梯度方向 Bin 的个数 8	27.3/25.4	53.1
AdaBoost-Binary	64 维, 64×64 补丁块, 128 个弱学习器, 响应函数梯度方向 Bin 的个数 8	23.5/24.4	49.8

5.3 所提方法的应用效果

这节里, 用几个实例图片来展现本文方法的应用效果. 首先, 在一组补丁上应用本文方法进行匹配, 观察匹配正确率. 本文方法首先在 Yosemite 上进行训练, 得到一定组合的弱学习器. 然后在图 7 所示的图像补丁集上进行描述匹配. 匹配的结果可分为 4 种情况: 对应补丁对匹配(true positive pairs)、对应补丁对不匹配(false positive pairs)、不对应补丁对匹配(true negative pairs)和不对应补丁对不匹配(false negative pairs). 为了提高识别率, 需要提高补丁被正确分类的比重. 所谓正确分类是指对应补丁对匹配和不对应补丁对不匹配. 匹配分类结果如图 7 所示, 图 7 中左上和右下是正确分类的补丁对, 右上和左下是错误分类的补丁对. 所提算法对视角变化和运动模式表现出很强的鲁棒性, 如图 7 中左上所示; 某些补丁被错误分类了, 这主要是由于遮挡和非常大的视角变化造成, 如图 7 中右上所示; 不对应的补丁能被匹配上与训练内容的差别有关, 使其在匹配时有极小的相似度都认为是对应的, 如图 7 中左下所示; 不对应补丁对不匹配是体现描述子的独特性, 如图 7 中右下所示. 通过这种测试还可以说明所提方法具有一定的泛化性能, 因为这里训练集选择的是 Yosemite, 测试是在不同训练集的图像补丁上进行的, 能将图像补丁正确分类.

其次, 在两幅有日规的图像采用本文方法对日规上的关键点进行匹配. 测试图像图 8 所示, 图 8 中两幅图像是从不同的角度对日规拍摄得到的, 相互之间有尺度、旋转变化, 右边图像有压缩. 这里本文的描述算法也是在 Yosemite 数据集上训练的, 然后

对取得的补丁(日规上小方框)进行描述, 然后进行匹配. 图 8 中左边图像提取的补丁, 每个都与右边图像上提取的补丁一一进行匹配. 若是匹配的, 就用线条将两个补丁块连接起来. 图 8 展示两幅图像目标识别效果. 实验结果表明匹配效果是好的, 在 26 个匹配对中只有一个匹配对是错误的, 识别率达到 96.2%. 说明所提的二值描述子对视角、旋转、尺度和压缩变化具有很好的鲁棒性. 此外, 这是使用 Yosemite 数据集训练出的描述子进行测试的, 而且测试图像与训练集差别较大, 所以说明了描述子的泛化性能比较好.

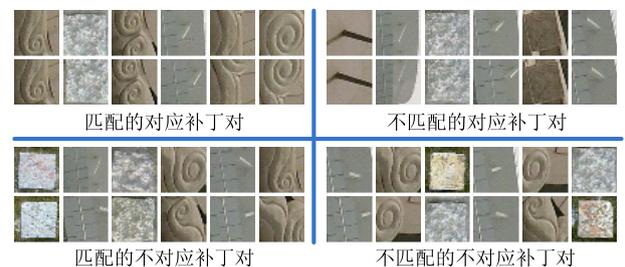


图 7 使用 AdaBoost-Binary 在不同形式的补丁上的匹配结果

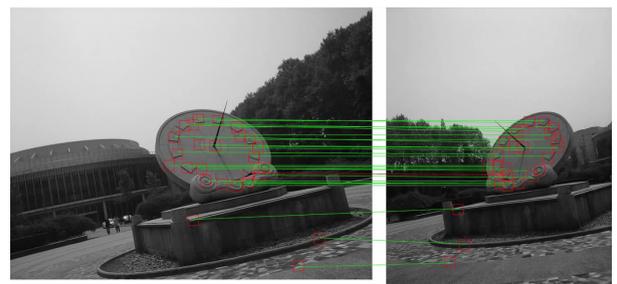


图 8 本文所提的 AdaBoost-Binary 描述子在实例上应用效果

6 结 论

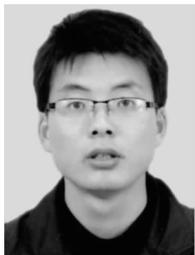
从集成学习的角度,提出了一种基于 AdaBoost 算法的学习架构,通过学习图像补丁的相似性来构建维数低、高独特性的特征描述子. 在该架构里,通过 Adaboost 学习得到一组非线性图像特征弱学习器,组合学习图像补丁局部特征信息,优化响应函数组合形式和弱学习器权重配置;而且提出了新的图像补丁相似性度量函数作为目标函数的核函数,基于 AdaBoost 算法设计了图像补丁相似性的整个学习算法流程,以改进特征弱学习器组合优化的效果. 该学习架构不会受限于任何预定义的图像特征信息采集模式,并能产生基于灰度信息和方向梯度信息的特征描述子. 该学习架构能有效地优化配置图像特征的弱学习器,得到的特征描述子具有较高的鲁棒性和较强的泛化性.

致 谢 审稿专家和编辑为本文提出了宝贵的意见和建议,作者在此表示衷心感谢!

参 考 文 献

- [1] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(10): 1615-1630
- [2] Babenko B, Dollár P, Belongie S. Task specific local region matching//*Proceedings of the IEEE International Conference on Computer Vision*. Rio de Janeiro, Brazil, 2007: 1-8
- [3] Schmid C, Mohr R. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, 19(5): 530-535
- [4] Lowe D G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, 60(2): 91-110
- [5] Bay H, Ess A, Tuytelaars T, et al. SURF: Speeded up robust features. *Computer Vision and Image Understanding*, 2008, 110(3): 346-359
- [6] Trzcinski T, Lepetit V. Efficient discriminative projections for compact binary descriptors//*Proceedings of the European Conference on Computer Vision*. Florence, Italy, 2012: 228-242
- [7] Brown M, Hua G, Winder S. Discriminative learning of local image descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(1): 43-57
- [8] Strecha C, Bronstein A, Bronstein M, et al. LDAHash: Improved matching with smaller descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(1): 66-78
- [9] Cai H P, Mikolajczyk K, Matas J. Learning linear discriminant projections for dimensionality reduction of image descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(2): 338-352
- [10] Norouzi M, Punjani A, Fleet D. Fast search in hamming space with multi-index hashing//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2012: 3108-3115
- [11] Malekesmaeili M, Ward R, Fatourech M. A fast approximate nearest neighbor search algorithm in the hamming space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(12): 2481-2488
- [12] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2003: 257-263
- [13] Tola E, Lepetit V, Fua P. DAISY: An efficient dense descriptor applied to wide baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(5): 815-830
- [14] Tang Yong-He, Lu Huan-Zhang, Hu Mou-Fa. Local feature description algorithm based on Laplacian. *Optics and Precision Engineering*, 2011, 19(12): 2999-3006(in Chinese)
(唐永鹤, 卢焕章, 胡谋法. 基于 Laplacian 的局部特征描述算法. *光学精密工程*, 2011, 19(12): 2999-3006)
- [15] Ge Juan, Cao Wei-Guo, Zhou Wei, et al. A local feature descriptor under color affine transformation. *Journal of Computer-Aided Design & Computer Graphics*, 2013, 25(1): 26-33(in Chinese)
(葛娟, 曹伟国, 周炜等. 一种颜色仿射变换下的局部特征描述子. *计算机辅助设计与图形学学报*, 2013, 25(1): 26-33)
- [16] Athitsos V, Alon J, Sclaroff S, Kollios G. Boostmap: A method for efficient approximate similarity ranking//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, USA, 2004: 268-275
- [17] Gionis A, Indik P, Motwani R. Similarity search in high dimensions via hashing//*Proceedings of the 25th International Conference on Very Large Databases*. Edinburgh, Scotland, 2004: 518-529
- [18] Shakhnarovich G. Learning Task-Specific Similarity [Ph. D. dissertation]. Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, 2005
- [19] Jain P, Kulis B, Grauman K. Fast image search for learned metrics//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Anchorage, USA, 2008: 1-8
- [20] Jegou H, Douze M, Schmid C. Packing bag-of-features//*Proceedings of the International Conference on Computer Vision*. Kyoto, Japan, 2009: 2357-2364
- [21] Leutenegger S, Chli M, Siegwart R. BRISK: Binary robust Invariant scalable keypoints//*Proceedings of the IEEE International Conference on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2011: 2548-2555

- [22] Rublee E, Rabaud V, Konolidge K, et al. ORB: An efficient alternative to SIFT or SURF//Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2011: 2564-2571
- [23] Alahi A, Ortiz R, Vanderghyest P. FREAK: Fast retina keypoint//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2012: 2069-2076
- [24] Calonder M, Lepetit V, Ozuysal M, et al. BRIEF: Computing a local binary descriptor very fast. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(7): 1281-1298
- [25] Chandrasekhar V, Takacs G, Chen D M, et al. CHoG: Compressed histogram of gradients a low bit-rate feature descriptor//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA, 2009: 2504-2511
- [26] Jégou H, Douze M, Schmid C, Pérez P. Aggregating local descriptors into a compact image representation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA, 2010: 3304-3313
- [27] Zitnick C. Binary coherent edge descriptors//Proceedings of the European Conference on Computer Vision, Heraklion, Crete, Greece, 2010: 170-182
- [28] Salakhutdinov R, Hinton G. Semantic hashing. *International Journal of Approximate Reasoning*, 2009, 50(7): 345-353
- [29] Weiss Y, Fergus R, Torralba A. Multidimensional spectral hashing//Proceedings of the European Conference on Computer Vision. Firenze, Italy, 2012: 340-353
- [30] Kulis B, Darrell T. Learning to hash with binary reconstructive embeddings//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2009: 1042-1050
- [31] Norouzi M, Fleet D. Minimal loss hashing for compact binary codes//Proceedings of the International Conference on Machine Learning. Bellevue, Washington, USA, 2011: 353-360
- [32] Jain P, Kulis B, Davis J, Dhillon I. Metric and kernel learning using a linear transformation. *Journal of Machine Learning Research*, 2012, 13(3): 519-547
- [33] Wang J, Kumar S, Chang S F. Sequential projection learning for hashing with compact codes//Proceedings of the International Conference on Machine Learning. Haifa, Israel, 2010: 1127-1134
- [34] Simonyan K, Vedaldi A, Zisserman A. Descriptor learning using convex optimisation//Proceedings of the European Conference on Computer Vision. Firenze, Italy, 2012: 243-256
- [35] Freund Y, Schapire R. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 1997, 55(1): 119-139
- [36] Trzcinski T, Christoudias M, Lepetit V, Fua P. Learning image descriptors with the boosting-trick//Proceeding of the Advances in Neural Information Processing Systems. Lake Tahoe, Nevada, USA, 2012: 278-286
- [37] Trzcinski T, Christoudias M, Fua P, Lepetit V. Boosting binary keypoint descriptors//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Portland, Oregon, USA, 2013: 2874-2881
- [38] Friedman J, Hastie T, Tibshirani R. Additive logistic regression: A statistical view of boosting. *The Annals of Statistics*, 2000, 28(2): 337-407
- [39] Zhang Yan-Ping, Zhang Ling, et al. *The Theory and Algorithm of Machine Learning*. Beijing: Science Press, 2012 (in Chinese)
(张燕平, 张铃等. *机器学习理论与算法*. 北京: 科学出版社, 2012)
- [40] Gong Y, Lazebnik S, Gordo A, Perronnin F. Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2011: 817-824



HUI Guo-Bao, born in 1982, Ph.D., engineer. His research interests include computer vision and pattern recognition.

TONG Yi-Fei, born in 1981, Ph. D., associate research fellow. His research interest include grid manufacture technology and manufacturing information technology.

LI Dong-Bo, born in 1957, Ph. D., professor. His research include advanced manufacture technology and opto-mechatronics technology.

Background

Local image feature descriptors have been vastly used in many computer vision applications such as image retrieval, pose estimation and 3D reconstruction. Their main goal is to represent a salient image region while remaining invariant to various illumination and viewpoint changes. In practice,

however, obtaining such a robust representation becomes a challenging task, especially for embedded device with limited resources.

The performances of descriptors structured in different way are various. To obtain better performance of descriptor

under various image transformations, numerous descriptors emerged over the past several years. They could be divided into two main classes according to their presenting styles, i. e. floating point style and binary style. It is notable that floating point descriptors reduce speed of matching and consume more memories though their performances are approbatory.

With increasing requirement for computer vision applications that can be used in embedded hand-holding devices, it starves for a substitute of floating point descriptors to apply to the device with limited computational capabilities. One way to speed up matching and reduce memory consumption is to work with binary descriptors. The main advantage of binary descriptor is that its similarity can be measured fast by the Hamming distance. This is very fast because the Hamming distance can be computed very efficiently via a bitwise XOR operation followed by a bit count.

However, as our experiments show, lots of binary descriptors often perform worse than their floating-point competitors: some are built on top of existing representation such as SIFT or GIST by relying on training data, and are limited by the performance of the intermediate representation. Other start from raw image intensity patches, but focus on computation speed and reply on fast-to-compute image features, which limit their accuracy.

To address these shortcomings, we proposed a learning

framework of image patch similarity to learn low-dimensionality and high-discriminative descriptor based on AdaBoost in this paper. In the proposed framework, we model the representations of image patches by non-linear weak learners which are trained by AdaBoost algorithm. Furthermore, in the aspect of the response function combination and optimal allocation of the weak learners' weight, we propose two new similarity functions for float point descriptor and binary descriptor respectively to use as kernel function of optimized object function to learn a similarity embedding for improving image feature matching effect.

We also performed a set of experiments to show that the proposed learning framework is more generalizing than others as it won't be restricted to any predefined feature-sampling model and it could encompasses intensity and director-gradient information. The empirical results show that the proposed feature descriptor outperforms overall comparing descriptors on recall-precision performance of image local area matching. In Conclusion, we develop an effective and efficient learning framework that is able to effectively optimize over the descriptor filter configuration leads to boost robustness to image scale and perspective changes.

This research is funded by the National Natural Science Foundation of China (Grant No. 61104171).