

# 基于分级显著信息的空间编码方法

赵宏伟<sup>1,2</sup>, 李清亮<sup>1</sup>, 刘萍萍<sup>1,2</sup>

(1. 吉林大学计算机科学与技术学院, 吉林长春 130012; 2. 符号计算与知识工程教育部重点实验室, 吉林长春 130012)

**摘 要:** 传统的 Bag of Words 模型检索方法并不具备局部特征间的空间关系, 因此影响检索性能. 本文提出了基于分级显著信息的空间编码方法. 通过分层次的提取显著区域并对每个显著区域内的特征点进行空间编码. 目的是探索特征间的空间关系, 并根据分级显著信息提高特征间的相关性. 在几何验证过程中, 本文通过任意三点间的角度编码和位移编码构成的空间编码方法完成图像对之间的空间关系匹配, 同时根据图像各个区域间的显著程度赋予该区域空间关系匹配得分相应权重, 得到最终的几何得分, 重新排列检索结果. 实验结果表明本文提出的方法既改善了最终检索结果的精确度又降低了几何验证阶段的计算时间.

**关键词:** 图像检索; Bag of Words 模型; 分级显著性信息; 几何角度编码; 几何位移编码

**中图分类号:** TP391.41      **文献标识码:** A      **文章编号:** 0372-2112 (2014)09-1863-05

**电子学报 URL:** <http://www.ejournal.org.cn>      **DOI:** 10.3969/j.issn.0372-2112.2014.09.031

## Spatial Encoding Based on Hierarchical Salient Information

ZHAO Hong-wei<sup>1,2</sup>, LI Qing-liang<sup>1</sup>, LIU Ping-ping<sup>1,2</sup>

(1. School of Computer Science and Technology, Jilin University, Changchun, Jilin 130012, China;

2. Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Changchun, Jilin 130012, China)

**Abstract:** The traditional model “Bag of Words” does not capture the spatial relationship among local features, thus affecting the retrieval performance. Hence, the spatial encoding method based on hierarchical salient information is proposed, which aims at fully exploring the geometric context of all visual words in images and increasing the discriminative power of the features based on hierarchical salient information. We propose a new encoding method in the geometric verification step. The spatial layout of every 3 points within a certain salient area will be represented by angle encoding and location encoding, meanwhile we sum all spatial matching scores with weights based on hierarchical salient information to generate the final ranking list. Experimental results prove that our scheme improves the retrieval accuracy significantly and reduces the computing time during the geometric verification step.

**Key words:** image retrieval; bag of words model; hierarchical salient information; angle encoding; location encoding

## 1 引言

随着多媒体信息的迅速发展, 图像检索技术已成为研究热点<sup>[1-3]</sup>, 然而大多数图像检索系统都是基于 Bag Of Words (BOW) 模型<sup>[4]</sup>实现的, 将图像中的局部特征<sup>[5]</sup>量化为视觉词并建立直方图检索图像. 然而, 量化后的视觉词降低了局部特征的辨别能力, 并且也不具备特征间的空间关系, 因此会产生错误的匹配结果并影响检索性能. 为了验证由于量化错误或特征间几何关系的缺失产生的特征误匹配, 有多种方法试图改进最初的 BOW 检索模型. 这些方法可以分为两大类: 第一种, 在量化阶段加入几何信息<sup>[6-8]</sup>; 第二种, 在搜索后引入几何验证方法, 重新验证初始检索结果中排序靠前的若干幅图像

与查询图像的几何关系<sup>[9-14]</sup>. 使用搜索后的几何验证步骤是改进检索精确度的常用方法. 几何验证没有改变图像表示与图像匹配方法, 而是得到特征匹配关系后根据特征点间的几何一致性关系过滤错误的特征匹配. 最近几年, 出现了很多几何验证方法用以优化图像检索的性能: Perdoch 等人通过离散化空间关系的局部几何表示维持了局部区域的几何一致性<sup>[9]</sup>, 但是如果重力矢量不确定, 对于旋转与仿射变换不具备不变性. Weak Geometric Consistency (WGC) 方法<sup>[10]</sup>, 假设在不同的尺度和旋转角度变换下, 会出现明显的峰值, 以此过滤错误的匹配特征. 然而 WGC 方法需要保证查询图像与候选图像的变换一致, 如果图像发生视角变化或者产生非刚性的形变, 效果改进有限. Local Geometric Consistency (LGC)

方法<sup>[11]</sup>假设图像发生形变时,基于局部区域特征相似的特点,在其中考虑了匹配特征相邻区域特征点的信息,精确计算了匹配特征间的变换矩阵,使 LGC 算法改善了图像发生视角变化与非刚性形变时的检索性能. Ransac<sup>[12]</sup>是一种全局几何验证方法,在整幅图像中计算所有特征的空间关系,耗时较大,仅适用于初始检索结果的前几幅图像. Zhou<sup>[13]</sup>等人随后又提出了新的几何编码方法,在原有的轴线编码基础上,结合了方形编码. Location Geometric Similarity Scoring (LGSS) 方法<sup>[14]</sup>使用任意两点特征的欧式距离比,验证几何一致性,有效地改善了运算效率. 由于仅用两点编码方法验证特征的几何关系,匹配特征出现错误的概率较大一些. 受 LGSS 算法的<sup>[14]</sup>启发,本文探索使用更多的特征点构成的空间关系验证查询图像和结果图像的几何一致性. 同时,由于多点编码的方法具有较高的计算复杂性,为了提高计算速度,因此进一步考虑了图像的分级显著机制,不仅提高了特征间的相关性,改善了检索精确度,同时降低了后验证算法的运算量,提高了检索速

度.

## 2 基于分级显著信息的空间编码的检索系统 (SEHSI)

本文提出了基于分级显著信息的空间编码方法 (Spatial Encoding Based on Hierarchical Salient Information, SEHSI), 如图 1 所示. 图 1(a) 为查询图像, 图中 (b1) 是经过 SIFT 特征提取后的特征图像, (b2) 表示基于 GBVS 方法得到的分级显著区域. 通过 Spatial Pyramid Matching (SPM) 方法<sup>[6]</sup>得到的初始检索结果如图 1(c) 所示. 同时匹配查询图像与结果图像的特征, 并查看查询图像中的匹配特征点的显著性级别, 将得到匹配特征后的位置信息保存到各个层次的匹配列表中, 如图 1(d) 最后对各个匹配列表中的位置信息进行几何编码并验证匹配关系, 记录匹配得分, 将不同显著性级别的空间编码匹配得分加权汇总得到最终的几何得分, 最终的检索结果如图 1(e) 所示.

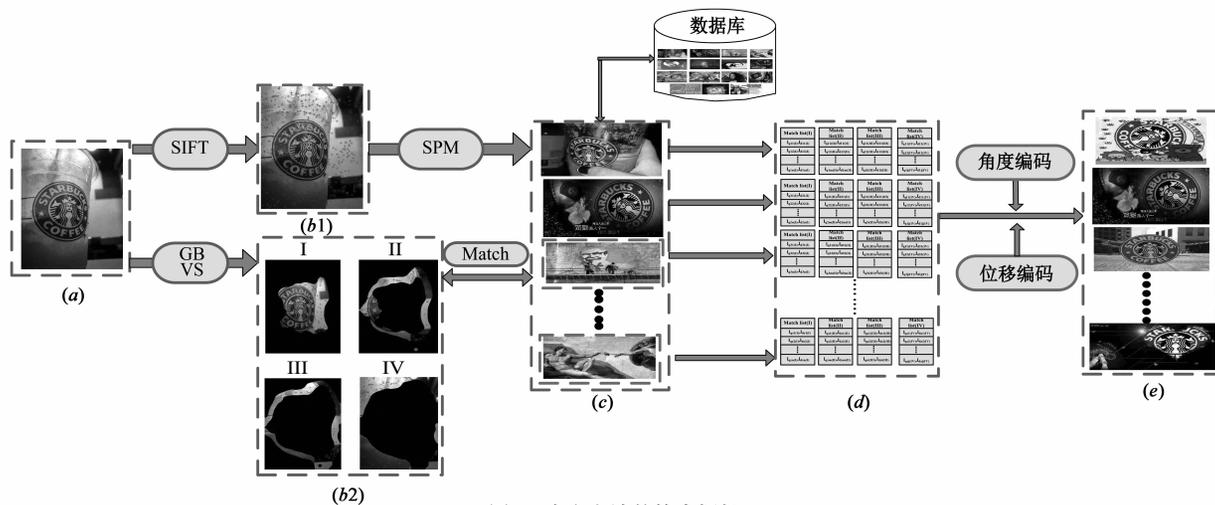


图 1 本文方法的检索框架

## 3 基于分级显著信息的空间编码

### 3.1 分级匹配列表

本文基于 GBVS 算法<sup>[15]</sup>生成层次化显著区域, 根据突出显著区域内最大显著值的百分比, 计算该显著级别的显著区域上限值  $pA_z$  与下限值  $pB_z$ , 并生成分级显著图像, 实验中, 设定显著度级别为 4. 如下式:

$$pA_z = \max(A) * prA_z \quad (1)$$

$$pB_z = \max(A) * prB_z \quad (2)$$

式中,  $\max(A)$  表示显著图像的最大显著值,  $z \in [1, 2, 3, 4]$ , 表示显著图像的显著级别,  $prA_z, prB_z$  分别表示  $z$  级显著区域内最大显著值的上限与下限的百分比.

分级匹配列表用于存储匹配特征的位置信息, 基

于各个匹配特征点的位置与图像中各个显著级别的显著区域关系, 将落入该显著级别内的匹配特征对的位置信息存入该级别的匹配列表中, 如式(3)所示

$$\text{Matchlist}_z = \text{QS} \Theta \text{DS}, \text{QS} \subseteq G_z \quad (3)$$

其中 QS, DS 分别指查询图像与数据库图像的 SIFT 特征,  $\Theta$  表示特征间的欧式匹配关系,  $G_z$  表示  $z$  级显著区域,  $\subseteq$  表示特征点位于该级别的显著区域内, Matchlist<sub>z</sub> 指各个显著级别中生成的匹配列表, 存放的是满足匹配关系的特征点的位置信息.

### 3.2 分级显著信息的空间编码

本文的几何验证方法只利用了特征间的位置信息计算几何得分, 由于仅用两点编码时, 判断匹配特征出现错误的概率很高, 因此本文引入另一个特征点, 增多

制约条件,降低错误判断的概率.基于三对特征点可能出现的空间位置关系提出了由角度编码和距离编码组成的几何编码方法,通过分级匹配列表任取三对匹配特征点,若它们三点不共线,则生成一对三角形,并判断某一角度是否足够相似;若三点共线,则只计算三点中任意两点的距离.

**角度编码** 选取匹配特征对中任意三点,若其不共线便能构成三角形,本文根据几何角度编码计算该三角形任意角度的余弦值,通过角度的相似程度判断特征间的空间一致性,最后根据图像间三角形中相似角度的数量,确定查询图像与结果图像的相似度,几何角度编码如下式:

$$\text{Ang}(y, x, z) = \frac{\text{dist}(x, y)^2 + \text{dist}(x, z)^2 - \text{dist}(y, z)^2}{2 * \text{dist}(x, y) * \text{dist}(x, z)} \quad (4)$$

$(x, y, z) \in \text{notline}$

式(4)中  $\text{dist}(\cdot)$  关于两个特征点的欧氏距离.  $(x, y, z) \in \text{notline}$  表示三个特征点  $x, y, z$  不在一条直线上,此时  $\text{Ang}(y, x, z) \in (-1, 1)$ .

**位移编码** 当三个匹配点在一条直线上,由于没有形成三角形便不能使用角度编码.在这种情况下本文提出位移编码,即  $\text{dist}(\cdot)$ ,通过特征点距离的相似程度去衡量几何一致性.

**基于分级显著信息的几何验证** 基于分级显著信息的几何验证方法通过查询图像与结果图像几何编码的几何得分重新排列结果图像,分别计算了各个级别显著区域内的几何角度编码与几何位移编码的几何得分,最终赋予相应权重和得到最后得分.

假设在某一级别匹配列表,查询图像中存在任意三个特征点  $(x, y, z)$  且这三点不共线,对应的在结果图像中与  $(x, y, z)$  匹配的三点分别为  $(x', y', z')$ . 则三点间的角度匹配值计算过程为:

$$S_{\text{Ang}} = \begin{cases} \frac{\text{Ang}(x, y, z)}{\text{Ang}(x', y', z')} & ; \text{cond1} \\ 0 & ; \text{cond2} \\ 1 & ; \text{cond3} \end{cases} \quad (5)$$

当结果图像三点的角度编码值不为 0 时,即  $\text{Ang}(x', y', z') \neq 0$ ,  $S_{\text{Ang}}$  就可以用  $\frac{\text{Ang}(x, y, z)}{\text{Ang}(x', y', z')}$  表示,即 cond1; 若结果图像中三点角度编码值为 0,即  $\text{Ang}(x', y', z') = 0$ ,便判断分子  $\text{Ang}(x, y, z)$  的值,若分子不为 0,那么  $S_{\text{Ang}}$  值为 0,即 cond2; 否则  $S_{\text{Ang}}$  值为 1,即 cond3. 当三点共线时,取任意两点构成的欧式距离比作为位移编码匹配值:  $S_{\text{dist}} = \frac{\text{dist}(y, z)}{\text{dist}(y', z')}$ .

为了保证角度编码特征的相似性,本文定义了角度编码的阈值,如式(6):

$$A_z = \{ S_{\text{Ang}} | (x, x'), (y, y'), (z, z') \in M_z, S_{\text{Ang}} \rightarrow \epsilon \} \quad (6)$$

式中  $M_z$  表示  $z$  级显著级别的显著匹配列表,  $S_{\text{Ang}} \rightarrow \epsilon$  表示角度编码比值  $S_{\text{Ang}}$  趋近于阈值  $\epsilon$  的条件,  $A_z$  表示  $z$  级显著级别匹配列表中所有匹配特征对的角度编码比值的集合,这些比值均满足阈值条件.

根据同样的原理,本文定义了距离编码的阈值,如式(6):

$$D_z = \{ S_{\text{dist}} | (x_a, x_b), (y_m, y_n), (z_i, z_j) \in M_z, S_{\text{dist}} \rightarrow \epsilon \} \quad (7)$$

$D_z$  表示  $z$  级显著级别匹配列表中所有匹配特征对的位移编码比值的集合,这些比值均满足阈值条件.

基于集合  $A_z$ , 计算  $z$  级显著级别的显著区域内的几何角度编码得分,如下式:

$$\text{Score. } A_z = \sum_{S_a \in A_z} S_a \quad (8)$$

同理,位移编码得分如下式:

$$\text{Score. } D_z = \sum_{S_d \in D_z} S_d \quad (9)$$

本文对两个编码方法的几何得分求和,得到各个显著级别显著信息背景下的几何得分,如下:

$$\text{Score}_z = \text{Score. } A_z + \text{Score. } D_z \quad (10)$$

根据显著级别,对相应的几何得分赋予相应权重,最后求和得到最终查询图像与结果图像间的显著性几何得分.

$$\text{FinalScore} = \sum_{z=1}^4 \omega_z * \text{Score}_z \quad (11)$$

其中  $\omega_z$  为相应显著级别的权重值.

## 4 实验结果及讨论

本文通过图像检索的两个重要因素验证 SEHSI 的检索性能:检索精确度与检索时间,并与目前几种普遍的几何验证方法<sup>[12,14]</sup>作比较.实验中计算初始检索后的前 100 幅图像的几何得分并重排序.分别在 DupImage<sup>[16]</sup>与 Ukbenhc<sup>[17]</sup>测试集中,各自选取其中 100 幅图像作为查询图像并测试检索性能.在 CPU 为 Core i3, 2.27GHZ, 2GRAM, 32 位的 WIN7 操作系统, Matlab R2010a 中进行测试,以 mAP<sup>[18]</sup>与运行时间作为衡量标准.

本文分析分级显著信息的引入与多点编码方法对检索性能的影响,比较了六种检索方法的 mAP,如表 1 所示,同时表 2 也列出了各个方法的计算时间.每个方法的相关参数均与原始文献一致,视觉词数量为 400,空间金字塔检索模型(SPM)中,  $L = 3$ ,提取分级显著图像过程时,式(1)和(2)中,根据  $z$  取值从 1 到 4,  $prA_z$  与  $prB_z$  分别为 (100%, 80%, 50%, 30%) 与 (80%, 50%, 30%, 0),几何编码过程的阈值  $\epsilon$  为 0.99,并且四个显著级别的权重  $\omega_z$  分别为 2, 1, 1/2, 1/4,同时设置每幅图像尺寸均小于  $500 * 500$ .

表 1 在两种测试集中,各种检索方法的 mAP

方法	检索方法	后验证方法	mAP (DupImage)	mAP (Ukbenhc)
1	BOW	-	0.4215	0.4689
2	SPM	-	0.5420	0.5759
3	SPM	Ransac	0.6173	0.6389
4	SPM	LGSS	0.6882	0.6998
5	SPM	SE	0.7489	0.7306
6	SPM	SEHSI	0.7758	0.7673

表 2 各种方法的计算时间

方法	特征处理 (s)	检索时间 (s)	每幅图像的后 验证时间(s)	每幅图像总的 查询时间 (s)
1	0.1664	0.1482	-	0.3146
2	0.1664	0.9334	-	1.0998
3	0.1664	0.9334	0.5709	58.1898
4	0.1664	0.9334	0.0394	5.0398
5	0.1664	0.9334	0.2474	25.8398
6	0.2421	0.9334	0.0451	5.6855

实验对比中,第一种方法是传统的 BOW 检索模型<sup>[4]</sup>,第二种使用 SPM 作为检索方法<sup>[6]</sup>,第三种与第四种都是通过几何验证对图像重排序的方法,分别是使用 VLFeat 工具箱的基于仿射变换的 Ransac 方法<sup>[12]</sup>与 LGSS 方法<sup>[14]</sup>,第五种则是本文提出的空间编码(SE)方法,此方法并未引入分级显著信息,最后一个本文提出的 SEHSI 方法。

首先分析多点空间编码(方法 5)对检索性能的影响,从表 1 明显看出,本文在 SPM 基础上提出多点分布的几何验证方法进一步改善检索性能,SE 方法的精确度相比传统的几何验证方法(方法 3,4)有所提高,由于引入多点编码方法,增多判断错误匹配特征的约束条件,相比 Ransac 的随机性与 LGSS 两点编码时验证特征匹配的不稳定性,降低了错误匹配出现的概率,因此精确度 mAP 有所提升。

由表 2 可知,Ransac 算法在  $N$  个特征点构成的数据集中每次采样其中的 4 个数据点,时间复杂度为  $O(N^4)$ ,LGSS 算法则在  $N$  个匹配对中任取 2 点做距离匹配,时间复杂度为  $O(N^2)$ ,本文 SE 算法任取 3 点做角度余弦值比,时间复杂度为  $O(N^3)$ ,因此相比 Ransac 算法节约了计算时间,但比 LGSS 方法消耗了较多的计算时间.因此考虑了分级显著信息,查询一幅图像时,本文提出的方法由于在特征处理阶段提取图像的分级显著性信息,增加了计算时间,但是基于分级显著性的几何验证算法由于降低了几何验证阶段的计算次数,增加的计算时间要远小于几何验证阶段节省的计算时间,因此减少了每幅图像的总计算时间.在理想情况下,当存在  $M$  对匹配特征时,由于本文将图像分为 4 个

显著区域,当特征数量平均落入每个显著区域时,相比 SE 方法,计算次数( $C_M^3 = M(M-1)(M-2)/6$ ),SEHSI 方法计算次数减为( $4 * C_{M/4}^3 = (M(M-4)(M-8))/(6 * 16)$ ).同时在计算几何得分时没有使用 LGSS 算法的构建软直方图的方式而是利用阈值  $\epsilon$  约束特征对的选取,更为简单,提高了精确度。

## 5 结论

在复杂背景的目标检索的研究中,由于传统检索模型(BOW)忽略了特征间几何关系的缺点,影响检索效率.本文提出的算法首先分级提取显著物体定位显著信息,并在后验证处理过程中,对相应的编码得分赋予不同权重,可以在不丢失有用检索信息的基础上,对查询图像根据与结果图像之间的相关性进行排序,同时也减少了检索时间,然后将各个级别显著区域内特征点的位置信息转换为角度编码与位移编码,最后,基于两幅图像的几何相似得分重新排列结果图像。

实验中,利用 DupImage 和 Ukbenhc 测试集,与其他常用的几何验证方法作比较,结果表明,本文提出的方法可以在保证检索速度的前提下,提高检索的精确度。

## 参考文献

- [1] 陈慧中,陈永光,景宁,陈萃,王钧.基于显著区域的月球影像内容特征研究[J].电子学报,2012,40(5):911-919.  
Chen Hui-zhong, Chen Yong-guan, Jing Ning, Chen Luo, Wang Jun. Research of content feature descriptors for lunar images based on saliency regions[J]. Acta Electronica Sinica, 2012, 40(5):911-919. (in Chinese)
- [2] 黄霞.基于本体和奇异值分解的图像数据索引查询算法[J].电子学报,2014,42(2):288-291.  
Huang Xia. An image data query algorithm based on ontology and singular value decomposition[J]. Acta Electronica Sinica, 2014, 42(2):288-291. (in Chinese)
- [3] Zheng L, Wang S, Zhou W, Tian Q. Bayes merging of multiple vocabularies for scalable image retrieval[A]. Proceedings of the 2014 IEEE International Conference on Computer Vision and Pattern Recognition [C]. Columbus, OH, USA: IEEE, 2014. 4321-4328.
- [4] Sivic J, Zisserman A. Video Google: A text retrieval approach to object matching in videos[A]. Proceedings of the 2003 9th IEEE International Conference on Computer Vision[C]. Nice, France: IEEE, 2003. 1470-1477.
- [5] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.
- [6] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories

- [A]. Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops [C]. New York, NY, USA: IEEE, 2006. 2169 – 2178.
- [7] Wu Z, Ke Q, Isard M, Sun J. Bundling features for large scale partial-duplicate web image search [A]. Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops [C]. Miami, FL, USA: IEEE, 2009. 25 – 32.
- [8] Hu S. Region-based partial-duplicate image retrieval [A]. Proceedings of the 2012 International Conference on Industrial Control and Electronics Engineering [C]. Xi'an, China: IEEE, 2012. 1521 – 1524.
- [9] Perd'och M, Chum O, Matas J. Efficient representation of local geometry for large scale object retrieval [A]. Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops [C]. Miami, FL, USA: IEEE, 2009. 9 – 16.
- [10] Zhou W, Li H, Lu Y, Tian Q. Large scale image search with geometric coding [A]. Proceedings of the 2011 ACM Multimedia Conference and Co-located Workshops [C]. Scottsdale, AZ, USA: ACM, 2011. 1349 – 1352.
- [11] Xie H, Gao K, Zhang Y, Li J. Local geometric consistency constraint for image retrieval [A]. Proceedings of the 2011 18th IEEE International Conference on Image Processing [C]. Brussels, Belgium: IEEE, 2011. 101 – 104.
- [12] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography [J]. Communications of the ACM, 1981, 24(6): 381 – 395.
- [13] Zhou W, Li H, Lu Y, Tian Q. Sift match verification by geometric coding for large-scale partial-duplicate web image search [J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), 2013, 9(1): 1 – 18.
- [14] Tsai S S, Chen D, Takacs G, Chandrasekhar V, Vedantham R, Grzeszczuk R, Girod B. Fast geometric re-ranking for image-based retrieval [A]. Proceedings of the 2010 IEEE International Conference on Image Processing [C]. Hong Kong: IEEE, 2010. 1029 – 1032.
- [15] Harel J, Koch C, Perona P. Graph-based visual saliency [A]. Proceedings of the 20th Annual Conference on Neural Information Processing Systems [C]. Vancouver, British Columbia, Canada: MIT Press, 2006. 545 – 552.
- [16] Zhou W. Dupimage [DB/OL]. <https://dl.dropboxusercontent.com/u/42311725/DupGroundTruthDataset.rar>, 2014 – 08 – 14.
- [17] Nister D, Stewenius H. Scalable recognition with a vocabulary tree [A]. Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C]. New York, NY, USA: IEEE, 2006. 2161 – 2168.
- [18] Philbin J, Chum O, Isard M, Sivic J, Zisserman A. Object retrieval with large vocabularies and fast spatial matching [A]. Proceedings of the 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C]. Minneapolis, MN, USA: IEEE, 2007. 1 – 8.

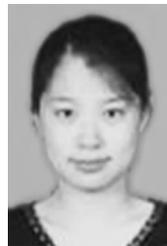
#### 作者简介



**赵宏伟** 男, 出生于 1962 年。现为吉林大学计算机科学与技术学院教授、博士生导师。主要从事智能信息系统、嵌入式技术、计算图像处理与可视化方面的研究。



**李清亮** 男, 出生于 1988 年。现为吉林大学计算机科学与技术学院博士。主要从事机器视觉和图像处理方面的研究。



**刘萍萍(通信作者)** 女, 出生于 1979, 现为吉林大学计算机科学与技术学院讲师。2009 年毕业于吉林大学计算机科学与技术学院计算机应用技术专业, 获得工学博士学位, 近年来主要从事机器视觉, 模式识别方面的研究工作。

E-mail: liupp@jlu.edu.cn