文章编号：1672-2892(2012)05-0566-05

无序多视图分类及其有序化技术

何周灿

(中国工程物理研究院 电子工程研究所，四川 绵阳 621900)

摘 要：无序图像分类问题是多视图匹配、三维重构、图像检索等应用的基础研究。提出一种新的相似度度量准则，准确划分场景相关图像和非相关图像；接着在通用匹配算法的基础上，提出试探匹配算法和种子扩散算法 2 种加速策略，显著提升了无序多图匹配及分类的效率，为三维重构等应用提供了更好的基础数据。

关键词：图像分类；相似度度量；试探匹配；种子扩散；三维重构

中图分类号：TN911.73；TP391 **文献标识码**：A

Grouping and organizing multiple unordered images

HE Zhou-can

(Institute of Electronic Engineering, China Academy of Engineering Physics, Mianyang Sichuan 621900, China)

Abstract: Grouping unordered images is considered a fundamental task in multi-view matching, 3D scene modeling, image retrieval and other visual computing applications. First a robust view-similarity measure is proposed and the images can be categorized accurately; then two speedup strategies, seed growing based grouping and tentative feature matching, are presented respectively to speed up calculating effectively, which can further provide nice data for 3D modeling.

Key words: image grouping; similarity measure; tentative match; seed growing; 3D modeling

无序图像分类是计算机视觉计算应用的一项基础研究，尤其当图像内容关联到多个场景且含有干扰图像时，准确分类场景相关图像变得更加困难。一般说来，对无序图像分类有 2 种思路：一种是早期提出的穷举匹配法[1-5]，依据所有图像两两匹配的结果对无序图像分类。Zisserman 等人[3]通过对全部视图进行两两粗匹配，以匹配数目作为度量对匹配结果进行分类，并辅以极线几何约束以确保关联准确。Yao 等人[2]提出一种复杂的相似度度量方式，并用视图生成树组织图像，Yao 只对相似度大于给定阈值但匹配数不到 50 的图像进行极线约束验证，取得了更高的计算效率和精确度。Zeng[4]从分组策略上提出一种基于模拟退火的分组算法，该算法可以大大减少粗匹配次数，但精确度较低。The Photo Tourism system[1]采用穷举匹配的分组方法，花了近 2 周的时间完成了 597 幅图像的三维重建。穷举匹配固然取得非常精确的结果，但是效率随着图像规模增加急剧降低；另外一种思路是近年来提出的基于机器学习的字典树方法[6-10]，它的思路是通过对一定量的图像特征数据聚类分解，训练出一棵字典树，然后对查询图像特征进行分散，字典树如果有 S 个叶子节点，查询图像就用一个为 S 维的高维向量来表示。Chum 等人[8-10]采用字典树组织大量无序图像，并采用 min-hash 算法获得场景相关联图像。Li 等人[7]采用类似的方法对成千上万的互联网图像进行聚类，并用极线约束方法排除干扰图像，实现了大规模三维场景重建。相比于传统的穷举匹配算法，机器学习的方法在效率上取得极大提升。但由于在机器学习的过程中会损失信息，字典树方法精确度较低，容易造成错误分类。

本文从传统思路角度出发，首先提出一种有效的相似度度量准则准确分类图像，接着应用试探匹配和种子扩散 2 种算法加速图像匹配，大大提升无序图像分类效率。

1 算法描述

首先，采用尺度不变特征变换(Scale Invariant Feature Transform，SIFT)[7,11]算法对所有无序图像提取局部特征，然后采用 K 近邻搜索算法—二分哈希(Dichotomy Based Hash，DBH)搜索算法[12]对 128 维的 SIFT 特征集进

收稿日期：2011-12-09；修回日期：2012-02-02

行匹配，并依据匹配结果计算各图像间的相似度值，最后依据相似度值分类并有序化无序图像。和传统匹配算法不同的是，不对所有图像特征集进行两两匹配，只选择种子图像进行相似度匹配(即种子扩散)，这样大大减少了匹配次数。同时，在种子图像进行扩散的过程中，加入试探匹配策略，快速鉴别不相关图像，进一步减少了大量不必要的计算。此外，新的相似度度量准则有效区分了关联图像和非关联图像，而不需要极线约束之类的辅助校验措施。

1.1 鲁棒的相似度度量准则

本质上，分类算法影响计算效率，而相似度度量决定了分类质量。Zisserman[3]首先提出了多视图分类问题，他采用匹配特征数目作为相似度度量，对城堡数据进行分类，结果不是很理想，把某些内容相关的图像分成了 2 组或多组。Yao 等人[2]提出了一种复杂的相似度准则，取得了更好的结果，也采用了随机抽样一致性(Random Sample Consensus，RANSAC)算法加以修正。

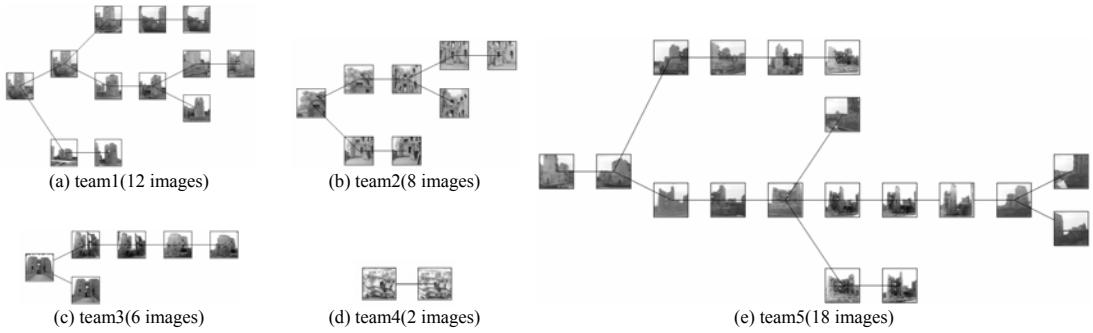

Fig.1 Grouping results of castle images based on the new measure
图 1 基于局部特征相似度度量准则的城堡数据分组结果

由于本文采用 SIFT 特征描述图像的局部不变特性，借鉴 SIFT 特征的匹配法则，提出一种基于局部特征的相似性度量准则，

$$S(A,B)=\frac{n_{\mathrm{m}}}{\min\{n_A,n_B\}}\sum_{i=1}^{n_{\mathrm{m}}}f(i) \tag{1}$$

式中：n_A,n_B 表示图像特征集合 A 和 B 的大小；n_{m} 表示匹配的特征数目；$f(i)$ 表示第 i 对匹配特征的匹配相似度值。$1/\min\{n_A,n_B\}$ 是一个自适应因子，用于应对不同大小但内容关联的图像。对于不同的局部特征描述子，$f(i)$ 会有不同的表现形式(比如城市距离、欧式距离、马氏距离等)。在 SIFT 特征空间中，$f(i)$ 被定义为：

$$f(i)=d_{i2}/d_{i1} \tag{2}$$

式中 d_{i1} 和 d_{i2} 分别表示第i个特征到其最近邻特征和其次近邻特征的欧式距离。

基于式(1)的度量准则，应用 DBH 匹配算法，采用穷举两两匹配的计算策略，用标准的城堡数据进行验证。对于相似度阈值 α，设为 1。图 1 的结果表明，基于局部特征的相似度度量准则，仍然能够获得和文献[2]一样的结果，且不需要任何辅助的验证策略。

1.2 种子扩散算法

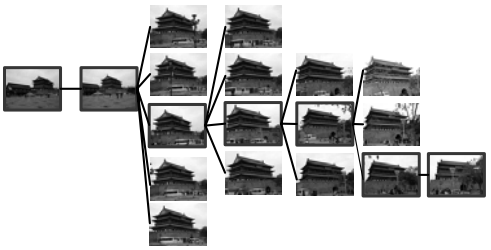

Fig.2 Seed growing process of drum-tower images and the square tagged ones are seeds
图 2 鼓楼图像的种子扩散过程，方框所标记的图像为种子图像

种子扩散的思路来源于图形学中的种子填充算法，但不同于种子填充。随机选择 1 幅图像作为种子图像建立索引结构，然后用剩余图像特征集去搜索，一般地，会得到 M 幅相似的图像，其中 $M<<N$，N 表示无序图像的规模。种子扩算的关键在于，并不采用所有 M 幅相似图像作为新的种子用于扩散，而只选其中的相似度值较小的图像作为新种子。一般地，M 幅相似图像以不同程度的相似度关联到种子图像，那些具有较大相似度的图像从某种程度上被看成是种子图像的拷贝，即便被用作新的种子进行扩散，检索到的相似图像也和原来的种子检索到的图像类似，不会增加太多相关联图像，反而增加不必要的计算代价。本文采用上一节提出的相似度度量准则，图像相似度阈值为 $\alpha=1$。自然地，可以引入另外

一个阈值 β，用来表示 2 幅图像是否高度相似，在本文中，$\beta=10\alpha$。种子扩散的过程中，只选择那些待扩散图像中和种子图像相似度值在 $[\alpha,\beta]$ 且彼此的相似度不大于 β 的图像用作新的种子进行种子扩散。种子扩散算法的思路见表 1。鼓楼数据的种子扩散过程见图 2。

表 1 种子扩散算法

Table1 Seed growing algorithm

No. of step	input:*N* unordered images of *S* scenes;output:*S* scene related teams
1	abstract SIFT features for *S* images
2	randomly select one image as seed image and insert into Que
3	while Que!=Null
4	use the head image of the seed Que to query and get *M* similar images
5	choose the images from the *M* similar images which are satisfying the follows: the similarity values with the seed images are between α and β, and the similarity values between each others are less than β
6	end
7	repeat step 2–6 until there are no images remained

1.3 试探匹配算法

上述种子扩散算法，能够提升 2~3 倍的计算效率。在本节中，在种子扩散算法的基础上，再提出另一个加速策略——试探匹配算法。种子扩散算法之所以高效，是因为相似度大于 β 的相关联图像并没有被用作种子去扩散。事实上，那些被当作种子的图像在扩散过程中，也进行了大量不必要的计算。为了得到 M 幅相似图像，种子图像需要和所有 L 幅剩余图像进行匹配，仍然有大量不必要的计算被执行。一种自然的想法是，探索一种合适方法，快速鉴别不相关图像，以减小不必要的计算量。试探匹配的核心思想是，通过随机采样的少部分特征进行匹配，计算图像的局部相似度，以快速鉴别图像是否相关。一般地，用 $Sim(\cdot)$ 表示 2 幅图像特征集的相似度。用部分特征进行匹配，计算出一个局部相似度 $Sim_{\mathrm{p}}(\cdot)$，如果能通过 $Sim_{\mathrm{p}}(\cdot)$ 推算出 $Sim(\cdot)$，那么问题便迎刃而解。

对于 2 幅图像的特征集 A 和 B，其大小分别为 n_A 和 n_B，用集合 A 建立索引结构 $H(A)$，然后用 B 中的特征去匹配(本文采用 DBH 算法)。建立索引的代价远远低于搜索匹配的代价，因此，仍然保持同常规搜索算法一样的索引结构，但在搜索时，仅随机采样部分特征去搜索。一般情况下，认为特征集 A 和 B 的大小 n_A 和 n_B 相当，于是用 n_B 代替 $\min\{n_A,n_B\}$，图像对 A 和 B 的相似度被重写如下：

$$S(A,B)=\frac{n_{\mathrm{m}}}{n_B}\sum_{i=1}^{n_{\mathrm{m}}}\frac{d_{i2}}{d_{i1}} \tag{3}$$

假设用所有 n_B 个特征去搜索，得到 n_{m} 个相匹配的特征对，其匹配率 n_{m}/n_B。理论上，如果随机选择 $n_{B\mathrm{p}}$ 个特征在 $H(A)$ 搜索，那么能够匹配的特征对数目 n_{mp} 的数学期望可以用式(4)表示，而其部分相似度 Sim_{p} 期望值可以用式(5)表示：

$$E(n_{\mathrm{mp}})=\sum_{i=1}^{n_{B\mathrm{p}}}iC_{n_{B\mathrm{p}}}^{i}\left(\frac{n_{\mathrm{m}}}{n_B}\right)^{i}\left(1-\frac{n_{\mathrm{m}}}{n_B}\right)^{n_{B\mathrm{p}}-i}=n_{\mathrm{m}}\frac{n_{B\mathrm{p}}}{n_B} \tag{4}$$

$$E(Sim_{\mathrm{p}})=\frac{(n_m\dfrac{n_{B\mathrm{p}}}{n_B})}{n_{B\mathrm{p}}}\sum_{i=1}^{n_{B\mathrm{p}}}\frac{d_{i2}}{d_{i1}}=\frac{n_{\mathrm{m}}}{n_B}\sum_{i=1}^{n_{B\mathrm{p}}}\frac{d_{i2}}{d_{i1}} \tag{5}$$

一般地，对于那些真正匹配的特征对，它们的最近邻特征的距离和次近邻特征的距离比 $d_{i2}/d_{i1}, i=1,2,\cdots,n_{B\mathrm{p}}$ 基本相似，因此把它看作一个常量，于是得到式(6)：

$$E(Sim_{\mathrm{p}})=\frac{n_{\mathrm{m}}}{n_B}\sum_{i=1}^{n_{B\mathrm{p}}}\frac{d_{i2}}{d_{i1}}=\frac{n_{B\mathrm{p}}}{n_B}\frac{n_{\mathrm{m}}}{n_B}\sum_{i=1}^{n_B}\frac{d_{i2}}{d_{i1}}=\frac{n_{B\mathrm{p}}}{n_B}Sim(A,B) \tag{6}$$

式(6)表明，可以从局部的相似度 Sim_{p} 推导出全局的特征相似度，即 $Sim(A,B)=Sim_{\mathrm{p}}\,n_B/n_{B\mathrm{p}}$。试探匹配的过程是自适应的，部分特征的数目 $n_{B\mathrm{p}}$ 在初始化时被设为 100。试探匹配的详细算法思路见表 2。试探匹配算法主要有两方面贡献，一是快速鉴别不相关图像，大大节省计算开销；二是对于真正的相关联图像，试探匹配执行全部匹配过程，保留了同穷举算法

表 2 试探匹配算法流程

Table2 Procedure of tentative matching

No.of step	procedure of tentative matching
1	generate a random feature queue Que_{rand} with all features in features set *B*
2	If Que_{rand}!=Null , pop n_{BP} features to search in *H*(*A*), else go to step 5
3	calculate Sim(*A*,*B*), if Sim(*A*,*B*)<Threshold, go to step 4, else go to step 2
4	return FALSE with the conclusion that the images *A* and *B* are unrelated
5	return TRUE with the conclusion that the images *A* and *B* are related

一样的有效结果，为后期组内关联图像的精细匹配准备了基础数据。

2　实验

2.1　分组及其有序化实验结果

互联网标准数据集。混合 6 个不同场景共 64 幅无序图像，这些图像组分别具有旋转、光照和仿射等变换。其分组及其有序化结果见图 3。从结果中可以看出，所有图像均被正确分类，同时也进一步验证了基于特征相似度度量的有效性。

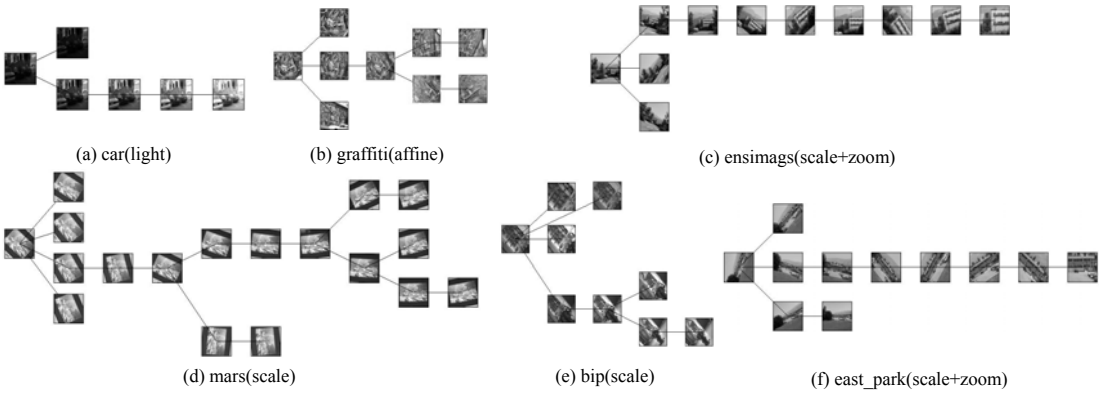

Fig.3 Grouping and organizing results of the web data
图 3　互联网数据的分组及其有序化结果

钟楼数据集合。用数码相机拍摄了西安钟楼、鼓楼等 6 处景点共 163 个图像。实验部分结果见图 4，所有图像被精确分类到相关组中。

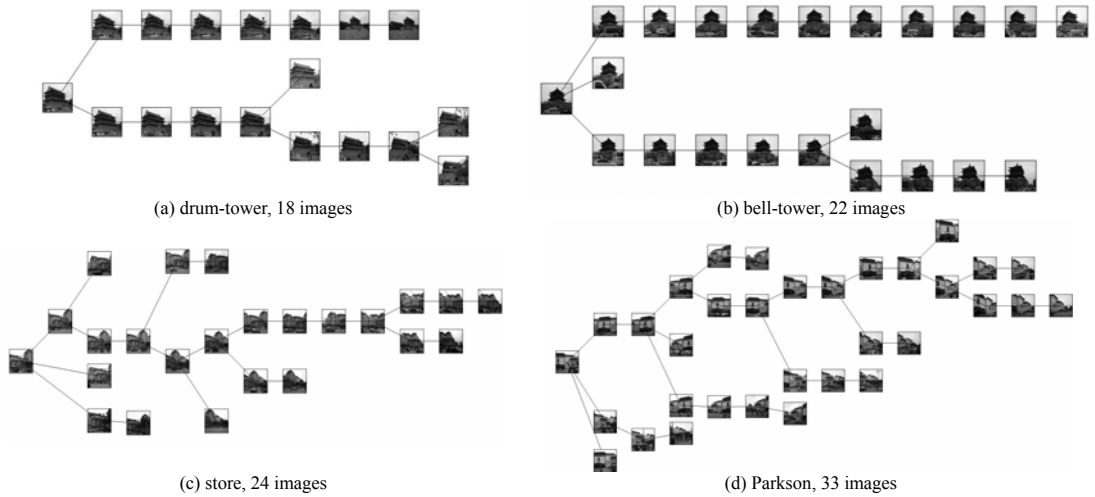

Fig.4 Grouping and organizing results of the bell-tower data
图 4　钟鼓楼广场附近图像的分组及有序化部分结果

2.2　计算效率分析

如果 2 幅图像真正内容关联，它们之间需要进行 1 次完整的匹配，把其计算量统计为 1。如果 2 幅图像不相关，并且它们只用了 1/10 特征参与搜索计算，则认为它们进行了 0.1 次计算。对于 n 幅输入图像，标准的穷举匹配算法需要进行 $n(n-1)/2$ 次计算。统计了上述 2 种数据的计算次数，结果见表 3。相比于标准穷举算法，提出的种子扩散和试探匹配策略在计算效率上提升了近 10 倍。

表 3　相关算法的计算次数统计
Table3 Efficiency comparison

	exhaustive	seed growing	tentative matching	seed growing+tentative matching
Web set	2 016	843	498.30	181.34
bell-tower set	13 203	5 977	2 629.35	1 293.81

3 结论

本文针对多视图匹配中的无序图像分类及有序化问题，提出了一种鲁棒的基于局部特征的相似度度量准则，有效识别场景相关图像和非相关图像；同时提出了种子扩散和试探匹配 2 种加速策略，在保证匹配和分类精确度的前提下提升 10 倍的计算效率。但针对更大规模(大于 1 000 幅)的三维重建等应用，分配和分类效率仍然是瓶颈问题，有待进一步探索更高效的分类算法。

参考文献：

[1] Snavely N,Seitz S M,Szeliski R. Photo tourism:exploring photo collections in 3D[J]. ACM Trans. Graph, 2006,25(3):835-846.
[2] Yao J,Cham W K. Robust multi-view feature matching from multiple unordered views[J]. Pattern Recognition, 2007,40:3081-3099.
[3] Schaffalitzky F,Zisserman A. Multi-view Matching for Unordered Image Sets, or How Do I Organize My Holiday Snaps?[C]// In ECCV. London:Springer-Verlag, 2002:414-431.
[4] Zeng X,Wang Q,Xu J. MAP Model for Large-scale 3D Reconstruction and Coarse Matching for Unordered Wide-baseline[C]// Photos,In:BMVC. London:[s.n], 2008:34-43.
[5] Hays J,Efros A A. Scene completion using millions of photographs[J]. ACM Transactions on Graphics(SIGGRAPH 2007), 2007,26(3):87-94.
[6] Berg T L,Forsyth D A. Automatic ranking of iconic images[R]. San Francisco:University of California,Berkeley, 2007.
[7] LI Xiaowei,WU Changchang,Christopher Zach,et al. Modeling and Recognition of Landmark Image Collections Using Iconic Scene Graphs[C]// ECCV. Heidelberg:Springer-Verlag Berlin, 2008:427-440.
[8] Philbin J,Chum O,Isard M,et al. Object retrieval with large vocabularies and fast spatial matching[C]// Proc. of CVPR, Minneapolis:[s.n], 2007:1-8.
[9] Chum O,Matas J. Web Scale Image Clustering[Z]. 2008.
[10] Chum O,Philbin J,Zisserman A. Near Duplicate Image Detection:min-Hash and tf-idf Weighting[C]// BMVC. London:[s.n.], 2008:493-502.
[11] David G. Distinctive Image Features from Scale-Invariant Key points[J]. International Journal of Computer Vision, 2004, 60(2):91-110.
[12] HE Zhoucan,WANG Qing. A fast and effective dichotomy based hash algorithm for image matching[C]// ISVC. Berlin: springer-verlag, 2008:328-337.

作者简介：

何周灿(1984-)，男，四川省射洪县人，硕士，工程师，主要研究方向为图像处理、模式识别、嵌入式软件可靠性设计.email:hezhoucan@163.com.