# 利用游客受雇拍摄法和云图标注应用程序接口探究均质化景观的识别与分类
## ——以日本东京桥区河流景观为例

# Recognition and Classification of Homogeneous Landscape With Visitor–Employed Photography and Cloud Image Annotation API
## —An Example of the Riverscape in Nihonbashi, Tokyo, Japan

**施佳颖**
东南大学建筑学院景观学系讲师

**本條毅**
日本千叶大学园艺学研究科环境科学与绿地造园学系教授

**矢泽优里子**
日本千叶大学园艺学研究科环境科学与绿地造园学系博士后研究员

**古谷胜则**
日本千叶大学园艺学研究科环境科学与绿地造园学系教授

**SHI Jiaying**
Lecturer, Department of Landscape Architecture, School of Architecture, Southeast University

**Tsuyoshi HONJO***
Professor, Department of Environmental Science and Landscape Architecture, Graduate School of Horticulture, Chiba University

**Yuriko YAZAWA**
Postdoctoral Researcher, Department of Environmental Science and Landscape Architecture, Graduate School of Horticulture, Chiba University

**Katsunori FURUYA**
Professor, Department of Environmental Science and Landscape Architecture, Graduate School of Horticulture, Chiba University

*Corresponding Author
Address: Department of Environmental Science and Landscape Architecture, Graduate School of Horticulture, Chiba University, Chiba 271-8510, Japan
Email: honjo@faculty.chiba-u.jp

**编辑** | 王颖、田乐
**翻译** | 闫露、申瑞琪、王颖、张晨希
**EDITED BY** | WANG Ying, Tina TIAN
**TRANSLATED BY** | YAN Lu, SHEN Ruiqi, WANG Ying, ZHANG Chenxi

**摘要**

对景观照片进行有效的分类是数据处理和环境分析中至关重要的一步。随着地理信息的收集逐渐采用众包模式，越来越多的研究开始利用带有地理标记的照片，将人们对场所的感知与互动可视化，并探究场所的美学、文化和游憩价值。近年来，图像识别机器学习算法的应用极大提高了关键词匹配的效率，并实现了大批量照片的自动分类。然而，这类方法在景观分类实践中——尤其是针对具有相似特点的均质化景观——应用仍显不足。本研究利用谷歌云视觉API和多层次聚类法，研发了一种半自动化的分类器来识别均质化景观照片，并将其应用于日本东京桥区城市河流均质化景观照片的分类中。分类器将所有河流景观划分为9个特征组，这些特征组的视觉印象与人们的直观感知一致。研究中所应用的混淆矩阵显示，分类器分类结果总体上的准确性达82.61%，表明机器分类与人工分类的结果十分相近。因此，该分类器对均质化河流景观照片的分类切实有效。这种方法可大力推动评估过程中的公众参与及城市旅游管理。

**关键词**

均质化景观；景观特征；图像标注；照片分类；城市河流景观；机器学习；聚类

**ABSTRACT**

Effective classification of landscape photographs is a vital step in data processing and environment analysis. With the popularity of crowdsourcing geo-information, an increasing number of studies have used geotagged photographs to visualize how people perceive and interact with destinations and explore the aesthetic, cultural, and recreational value of the areas. In recent years, machine-learning algorithms for image recognition have dramatically improved the efficiency of the assignment of keywords and provide possibilities for the automatic classification of numerous photographs. However, the applicability of such methods for the practical landscape classification is still not clear, especially for the photographs presenting a homogeneous landscape that has similar characteristics. This study developed a semi-automatic classifier for homogeneous landscape photographs by using Google Cloud Vision API and multi-level hierarchical clustering. The classifier was applied to the classification of urban riverscape photographs, which is a typical example of homogeneous landscapes in Nihonbashi, Tokyo, Japan. The riverscapes can be classified into 9 characteristic groups by the classifier and the visual impression of these groups matches well with our intuitive feeling. A confusion matrix showed that the overall accuracy was 82.61%, indicating a strong agreement between the classifier and manual classification. Therefore, the classifier is practical for classifying homogeneous riverscape photographs. Such methodology also provides the possibility of public participation in the assessing process, which, in turn, contributes to urban tourism management.

**KEYWORDS**

Homogeneous Landscape; Landscape Characteristics; Image Annotation; Photograph Classification; Urban Riverscape; Machine Learning; Clusters

**基金项目**
日本学术振兴会Kakenhi基金项目“作为私人和非正式绿地的绿色基础设施：论参与式维护政策”（编号：JP 17K08179）

**RESEARCH FUND**
Private and Informal Green Space as Green Infrastructure: Towards Participatory Maintenance Policies, Japan Society for the Promotion of Science (JSPS) Kakenhi (No. JP 17K08179)

https://doi.org/10.15302/J-LAF-1-020054
收稿时间 RECEIVED DATE | 2021-07-20

中图分类号 | TP181, TU984
文献标识码 | A

## 1 引言

景观分类为相关交流提供了参考框架，有利于推动景观的管理与研究[1]。景观照片有助于理解空间数据及呈现景观特征与公众偏好[2]，将景观照片有效分类是数据处理和分析过程中至关重要的一步；而游客受雇拍摄法（visitor-employed photography，VEP）亦有助于完善旅游规划[3]。随着众包模式在地理信息采集中的普及，越来越多的研究开始利用包含地理标记的照片将人们对场所的感知与互动可视化[4]~[6]，并探索了场地的美学、文化和游憩价值[7]~[12]。尼卡·巴洛梅诺和布莱恩·加德罗指出，照片资料可以弥合社会科学层面上认知词汇和视觉词汇之间的差异[13]。然而大多数与景观相关的研究仅聚焦于照片中包含的元数据（如地点、时间、标题等）[5][14]，只有少数研究探究了照片本身的内容。近些年，越来越多的证据表明，在环境与土地利用管理中，景观的视觉品质越发重要[15]~[17]。景观的特征、美学价值和可见性等已被纳入生态系统服务的相关政策导向与评估范畴，与地方及区域发展息息相关；特别是公众偏好和景观的视觉品质可能对空间规划产生极大影响[18]。因此，为了探究某场地的景观意象与公众偏好之间的关系，需要对相关的照片内容进行详细解析，而以往的人工分类方法难以满足大批量照片的客观分类需求。

多温·P·卡特莱特认为，需要对这类照片批量进行系统性和结构性的量化分类[19]。阿蒂亚·维拉亚等人研发了一种分层分类器，通过设定基础特征及手动添加关键词的方式来区分城市和景观的意象[20]。有关照片分类的研究大多聚焦于图像内容的分析[8][11][21]~[23]。基于以往研究，伊莉莎·奥特罗斯—罗萨斯等人根据景观特征和文化生态系统服务，开发了一系列有助于进一步照片精细化分类的指标[14]。但上述研究均需人为指定关键词，费时费力。

近年来，针对图像识别的机器学习算法的应用大大提高了研究中关键词的匹配效率，谷歌云视觉、微软Azure计算机视觉、IBM Watson视觉识别和亚马逊Rekognition图像识别等云图标注工具，无需通过前期训练即可批量标注照片[24]。具体而言，谷歌云视觉应用程序接口（API）可为用户提供的大量预设的图像标签，以辅助目标图像的快速标注[25]。有研究表明，虽然谷歌云视觉API提供的关键词能够非常贴切地描述照

## 1 Introduction

Landscape classification provides a frame of reference for communication and is conducive to the promotion of landscape management and research[1]. Effective classification of landscape photographs—beneficial not only for understanding spatial data but also for landscape characteristics and preferences[2]—is a vital step in data processing and analysis. Research also proved that visitor-employed photography (VEP) can enhance tourism planning[3]. With the popularity of crowdsourcing geo-information, an increasing number of studies have used geotagged photographs to visualize how people perceive and interact with destinations[4]~[6] and explore the aesthetic, cultural, and recreational value of the area[7]~[12]. Nika Balomenou and Brian Garrod suggested that photographic data is an important means to bridge the gap between cognitive vocabulary and visual vocabulary in social sciences[13]. However, most landscape-related studies have focused on metadata embedded in the photographs (location, time, title, etc.)[5][14], while only a small proportion addressed the visual content of the photographs themselves. There is growing evidence that the visual quality of landscape has become more important in environmental and territorial management in recent years[15]~[17]. The character, aesthetics, and visibility of landscapes are included in political agendas and evaluations of ecosystem services, and are taken into account in local and regional development. The public's preference and value of the visual quality of landscape may have a considerable impact on spatial planning[18], so a more detailed analysis of photographs' content is necessary to interpret the relationship between destination image and public preferences. Manual classification was once the only method, but it is difficult to classify numerous photographs objectively.

According to Dorwin P. Cartwright, such photographs need to be quantified in a systematic and structured form for classification in batches[19]. Aditya Vailaya et al. developed a hierarchical classifier to distinguish urban and landscape images using low-level features and manually added keywords[20]. Much of the literature on photograph classification focuses on image content analysis[8][11][21]~[23]. Based on previous research, Elisa Oteros-Rozas et al. developed a set of indicators from landscape characteristics and cultural ecosystem services to categorize photographs in more detail[14]. In these studies, keywords were still assigned by humans, which required considerable time and human resources.

In recent years, machine-learning algorithms for image recognition have dramatically improved the efficiency of the assignment of keywords. Cloud image annotation tools, such as Google Cloud Vision API, Computer Vision in Microsoft Azure, IBM Watson Visual Recognition, and Amazon Rekognition, have the advantages of photograph annotation in batches without training[24]. For instance, the Google Cloud Vision API enables users to annotate any requested images quickly using numerous predefined image-level labels[25]. It has been shown that keywords annotated by Google Cloud Vision API describe the photograph contents satisfactorily[26]. However, the average number of

片内容[26]，但每张照片平均对应约20个关键词，因而当图片数量较大时，较难辨别出特定主题。

已有部分研究通过谷歌云视觉API匹配的关键词来归类或描述具有地理标记的照片。费鲁莉娜·M·沃特曼等人发现，基于谷歌云视觉API的分类器比人工分类更加精确[27]。丹尼尔·R·理查兹等人首次依据层次聚类法，将在新加坡拍摄的25 000张照片分为7类[28]。宋晓平等人利用同样的方法，将94 890张摄于公园中的照片分为10类[29]。另有学者通过这种方法分析了将照片上传至媒体平台的摄影者。例如，维罗妮卡·阿兰皮·索蒂尼等人利用9 304张葡萄酒庄景观照片的拍摄者分为4组[30]。阿尔然·S·戈萨尔等人采用相似的方法，将1 292名用户归为6组[31]。在这些研究中，研究人员利用基于关键词的分类器，将大量照片归类为相互独立且易区分的不同类别，如“动物”“植物”和“人”。然而，这种方法在景观特征的分类实践中——尤其是针对呈现了均质化景观的照片——应用仍显不足。在本研究中，“均质化景观”（homogeneous landscape）指具有相似地形、土地覆被、水文条件和聚落形态的土地单元[32]，如河流、森林、农田、城市区域等。

本研究研发了一套半自动分类器，尝试应用机器学习技术，以类似人工分类的方式处理VEP模式下拍摄的大量均质化景观照片。研究的目的在于：1）确定该分类器是否适用于均质化景观的识别和分类；2）评估该分类器相较于人工分类的准确性。研究运用谷歌云视觉API技术来标注一系列城市河流景观照片，照片所呈现的是典型的均质化景观。此外，研究还基于分类成果生成了研究区域的热力图。

## 2 研究资料与方法

### 2.1 研究区域

本研究以日本东京日本桥区的日本桥川、亀岛川和隅田川三条河流流经区域为研究对象（图1）。日本桥区是东京中央区的商业和文化中心，拥有典型的都市河流景观，包括城市运河、商业建筑、桥梁、墙式护岸和露天平台等。区域中的地形、土地覆被、水文条件和聚落形态也较为相似，是理想的调研场地。此外，从日本桥登船码头出发的游船是热点观光项目，每年接待游客约5~6万人次。

为缓解城市发展的压力，首都高速公路于1964年东京奥运会期间落成。它覆盖了日本桥川的部分河道，改变了原来的河流景观。许多评论人士曾就该高速路的建设对河流的冲击提出质疑，认为其影响了日本桥区的文化交流和经济发展。为回应这些质疑，“日本桥区复兴计划”提出将首都高速公路的部分路段移至地下。鉴于日本桥川即将面临新的变化，掌握游客对相关河段景观的偏好和感知非常重要。

### 2.2 照片收集

研究采用VEP方法收集现场照片。该方法由G·J·切雷姆首次提出[33]，主要利用游客而非研究者拍摄的照片来评估公众对景观的感知[34][35]。VEP方法可最大程度地减少研究者对游客体验的干预，是快速有效地分析景观特征与游客偏好和感知的重要方法。与社交媒体上标记

keywords per photograph is around 20, so distinguishing specific themes among the contents of numerous photographs remains elusive.

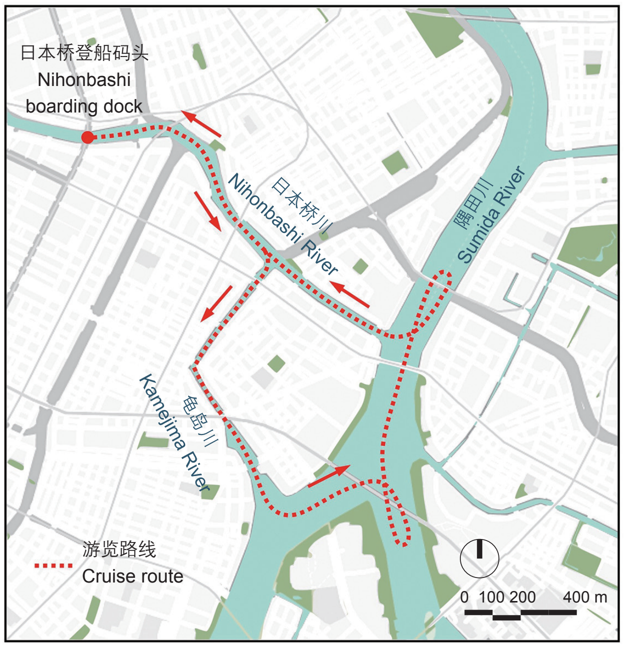
Some studies have used keywords assigned by the Google Cloud Vision API to classify or characterize geotagged photographs. Flurina M. Wartmann et al. showed that a classifier based on the Google Cloud Vision API showed a relatively high level of accuracy compared to manual classification[27]. Daniel R. Richards et al. first classified 25,000 photographs taken in Singapore into 7 categories by hierarchical clustering[28]. Song Xiaoping et al. used the same methods but focused on park photographs and classified 94,890 photographs into 10 categories[29]. Based on this approach, some scholars have analyzed photographers who uploaded photographs to social media. Veronica Alampi Sottini et al. grouped winescape users into 4 clusters from 9,304 photographs[30]. Arjan S. Gosal et al. used similar methods to classify 1,292 users into 6 groups[31]. In these studies, a classifier based on keywords was used for classifying numerous photographs into independent and distinguishable categories, e.g., “animal,” “plants,” and “people.” However, the applicability of such methods for landscape characteristic classification is still not clear, especially for photographs presenting a homogeneous landscape. The term “homogeneous landscape” in this study refers to a unit of land with relatively the same topography, land cover, hydrology, and settlement pattern[32], such as the landscape of rivers, forests, farmlands, metropolitan areas, etc.

This study applied machine-learning techniques to human-like classification for a large number of photographs. We developed a semi-automatic classifier for VEP photographs of the homogeneous landscape. The purposes of this investigation were 1) to determine whether the classifier is practical for recognition and classification of the homogeneous landscape, and 2) to evaluate the accuracy of the classifier by comparing it with manual classification. The Google Cloud Vision API was used to annotate a set of urban riverscape photographs, which is a typical example of a homogeneous landscape. We also created heat maps of the study area based on the classification.

## 2 Materials and Methods

### 2.1 Study Area

The study area covers the Nihonbashi River, Kamejima River, and the Sumida River flowing in Nihonbashi, Tokyo, Japan (Fig. 1). Nihonbashi is an urban district of business and culture in Chūō-ku with a typical metropolitan riverscape, characterized by city canals, commercial buildings, bridges, waterfront revetment walls, and terraces. The site is perfect for the investigation purposes of this study as it is a unit of land with relatively the same topography, land cover, hydrology, and settlement pattern. Cruises starting from Nihonbashi Boarding Dock are popular, with approximately 50,000 to 60,000 visitors a year.






1

1. 研究区域与游览路线
1. Location map of the study area and the cruise route

了地理信息的照片不同，VEP方法获取的照片更具针对性，也更适用于小尺度的研究区域。已有研究证明该方法尤其适用于调研河流和步道等连续性景观[36][37]。

现场照片拍摄于2019年7月30日和8月5日的同一时间段，这两日均为晴天且天气情况相似。研究共招募31名就读于日本千叶大学的志愿者（平均年龄22.4岁，标准差为3.2），他们来自景观系的不同年级（平均入学年数为4.1年，标准差为3.1）。在这次实验中，他们需要乘坐指定游船，并在游船上拍摄10张以上的照片，照片内容为游览过程中任何影响他们观光体验的意象或物体。照片由带有全球定位系统的手机拍摄，以确保所有照片都包含经纬度信息。研究中的游览线路（从日本桥码头出发，途径日本桥川、亀岛川和隅田川，最终回到码头）全长6.8km，共耗时45分钟（图1）。

### 2.3 照片标注

两次乘船游览共收集到508张包含有效元数据的照片。这些照片经谷歌云视觉API处理，被标注为相应的关键词（每张照片不超过50个①）。同时，该API也注明了每个关键词相应的置信度，分数范围为0.5~1，分数越高，置信度越高[38]。

① 选择以50为上限的主要依据为：1）谷歌云视觉API针对每张图片反馈的关键词数量最高阈值为50；2）针对均质化景观，应当尽可能收集更多的关键词，以提高分类结果的准确性。

In response to urban development pressures, the Metropolitan Expressway was built during the 1964 Tokyo Olympics. It covers part of the Nihonbashi River and alters the original riverscape. Many critics questioned its impingement on the rivers, which led to the cultural and commercial decline of the Nihonbashi area. In response to these concerns, the Metropolitan Expressway will be partially moved underground as part of the "Nihonbashi Revitalization Project." Because the Nihonbashi River is again facing changes, it is vital to understand visitors' preferences and perceptions of the related riverscape.

### 2.2 Photograph Collection

We used VEP to collect on-site photographs. The method was first proposed by G. J. Cherem[33], using photographs taken by visitors rather than by researchers to measure the public perception of the landscape[34][35]. VEP minimizes researchers' intervention in visitors' experience as much as possible, and was used to analyze the landscape characteristics and visitors' preferences and perceptions quickly and effectively. Compared with geotagged photographs from social media, VEP photographs are more targeted and applicable for small-scale study areas. It has been proven particularly suitable to investigate sequential landscapes, such as rivers and trails[36][37].

The on-site photographs were collected in the same time period on July 30 and August 5, 2019, two sunny days with similar weather. A total of 31 respondents (mean age = 22.4, SD = 3.2) from Chiba University were recruited by posters. They were students with varying years of postsecondary education of landscape (mean year = 4.1, SD = 3.1). In the experiment, they were required to board a designated sightseeing ship and take 10 or more photographs of any images or objects that influenced their viewing experience during the cruise. The photographs were taken by cell phones with a GPS function to ensure that all the photographs recorded latitude and longitude information. The entire 6.8-kilometer cruise route took 45 minutes, starting from the Nihonbashi Boarding Dock, traveling through the Nihonbashi River, Kamejima River, and Sumida River, and returning to the dock (Fig. 1).

### 2.3 Photograph Annotation

A total of 508 photographs with valid metadata were collected during the two cruises. They were sent to the Google Cloud Vision API to obtain photograph annotation keywords (at most 50 keywords per photograph①). The API also provided the corresponding confidence scores of each keyword, which ranged from 0.5 (low confidence) to 1 (very high confidence)[38].

① The largest number of keywords set as 50 was due to two reasons: 1) Google Cloud Vision API can annotate as many as 50 keywords for each photograph, and 2) for homogeneous landscapes, the more keywords it is annotated, the higher the classification accuracy will be.

## 2.4 照片分类

与理查兹等人使用的方法类似[28]，本研究创建了一套数据集，包含照片编码和每张照片的所有关键词及其置信度分数等信息。在此基础上，对数据集进行自下而上的层次聚类，样本距离采用欧式距离计算，簇（聚类）间距离采用全连接算法。该层次聚类由R语言中的hclust函数实现。

而后，研究团队改进了聚类算法，即对聚类的大小（即聚类中照片的数量）设定上限，并将此方法命名为“多层次聚类法”（multi-level clustering method）。将任意聚类的最大值设定为$N_{max}$，若聚类$i$的大小$N_{ci}$大于$N_{max}$，将被切分为多个子集，子集的数量为$N_{ci}$ / $M$的取整值（$M$为常数）。在原集合上不断执行递归操作，直至所有聚类子集的大小均小于$N_{max}$。在此过程中如何确定这些参数的最优值最具挑战性。经过多次组合尝试后，研究团队发现当每个聚类中的照片数量小于60时，聚类模型运作最佳。因此，研究将层次聚类系统树图中一级切分的聚类数设置为8，$N_{max}$和$M$的取值分别为60和40。

研究团队从整体性视角[39]确认了通过上述分类器得到的聚类分组，并确保一个聚类中的所有照片都属于同一主题。随后，将主题高度相似的聚类合并为同一特征组，最终得到9类具有不同河流景观特征的组别。

## 2.5 照片内容的空间分布

研究利用地理信息系统获得了照片的空间分布情况，并通过热力图将每组中地理坐标点的密度以彩色渐变图的形式体现，密度最高的区域处显示为黄色，密度最低的区域处显示为蓝色。

## 2.6 准确性检验

为评估该分类器的性能，研究团队从每个特征组中随机选取10%（张数取整）的照片作为样本照片进行检验。共邀请8名千叶大学景观规划专业的学生参与，他们需观察样本照片和由分类器分类得到的照片集合（不包含样本照片），而后将样本照片分类至最合适的照片集合中。该过程无时间限制，且为确保检验的独立性，参与者不允许相互交流。最后，8名测试者选择频率最高的组即为每一样本照片的人工分类结果。

表1：混淆矩阵
Table 1: Confusion matrix

| 样本照片是否属于特定特征组？<br>Is the sample photograph a member of the specific group? | 实际为真<br>Actual: True | 实际为假<br>Actual: False |
|---|---|---|
| 预测为正<br>Predicted: Positive | 真正例<br>TP | 假正例<br>FP |
| 预测为负<br>Predicted: Negative | 假负例<br>FN | 真负例<br>TN |

## 2.4 Photograph Classification

We created a dataset consisting of the coded photographs, and keywords and their confidence scores assigned to each photograph. Subsequently, hierarchical clustering was executed by applying the “hclust” function in R to the matrix with the Euclidean distance metric and complete linkage agglomeration. Richards et al. used similar approaches[28].

After that, a clustering algorithm was improved to make the size of each cluster (i.e., the number of photographs in the cluster) smaller than a certain value, and we named it the “multi-level clustering method.” The maximum size of a cluster was set to $N_{max}$. If $N_{ci}$, the size of cluster $i$, was bigger than $N_{max}$, cluster $i$ should be divided into subgroups or nodes. The number of subgroups was set to the rounded-up value of $N_{ci}$ / $M$, where $M$ is a constant. This procedure was performed recursively on the pruned sets until the size of all clusters was smaller than $N_{max}$. It was challenging to determine the optimal value of these parameters. Based on many attempts of combinations, the clustering model could best serve the needs when the number of photographs in each cluster is smaller than 60. So, for the first-level cut in the hierarchical dendrogram, the number of clusters was set to 8, and the values of $N_{max}$ and $M$ were 60 and 40, respectively.

We confirmed the clustering results through a holistic view[39] and ensured that the photographs of each cluster had a common theme. Clusters with highly similar contents were merged into the same characteristic group. The result was that 9 groups with different riverscape characteristics were identified.

## 2.5 Spatial Distribution of Photographs’ Contents

We observed the spatial distribution of the photographs using GIS. Heat maps were used to display the density of the geographic points in each group as colored gradient layers, with the highest density shown as yellow, the lowest density as blue.

## 2.6 Accuracy Test

To assess the performance of the classifier, we randomly selected 10% (rounding to the nearest integer) of the photographs from each characteristic group as the sample photographs for the test. There were 8 students majoring in landscape planning at Chiba University participating in the test. They were asked to observe the sample photographs and the categorized photograph collections (the photograph groups obtained from the classifier, without the sample photographs), and then assigned the sample photographs into the photograph collections they considered most suitable. There was no time limit for the test process. To ensure the independence of the test, participants were not allowed to communicate with each other. The result of the manual classification of each sample photograph was determined as the group with the highest frequency among the 8 testers’ selections.

A multiclass confusion matrix was used to compute the agreement between the classifier and manual classification. According to the conventional definitions,

此外，研究使用了多分类混淆矩阵来计算分类器和人工分类之间的一致性。真正例（TP）、真负例（TN）、假正例（FP）和假负例（FN）的常见定义如表1所示。在本研究中，真正例代表准确将样本照片分类为属于某组；真负例代表准确将样本照片分类为不属于某组；假正例代表错误将样本照片分类为属于某组；假负例代表错误将样本照片分类为不属于某组。

召回率指特定组的样本照片被成功分配到该组的比例。精确率指特定组的样本照片真正属于该组的比例。此外，研究引入了$F_1$值（即召回率和精确率的加权平均值[40]）来衡量检验的准确性。$F_1$值为1时表示准确性最高。所涉及方程包括：

$$Recall = \frac{TP}{TP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$F_1 = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (3)$$

## 3 研究结果

### 3.1 关键词提取

研究通过谷歌云视觉API为508张照片匹配了共9 946个关键词（平均每张照片匹配19.58个关键词，标准差为8.99）。其中包含关键词最多（43个）的照片有1张，包含关键词最少（1个）的照片有4张。此外，共检测到290个较为独特的关键词。图2的双轴图展示了25个最常见关键词的频率，以及每个关键词的总置信度与平均置信度分数。

true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) are described in Table 1. TP means samples which classified as belonging to the group correctly; TN means samples which classified as not belonging to the group correctly; FP means samples which classified as belonging to the group incorrectly; and FN means samples which classified as not belonging to the group incorrectly.

Recall was defined as the fraction of sample photographs of a specific group that were successfully assigned to that group. Precision was defined as the fraction of sample photographs assigned to a specific group that truly belong in that group. We also used the $F_1$ score, which is the weighted average of recall and precision[40], to measure the test's accuracy. The $F_1$ score reaches maximum accuracy at 1. These calculations are presented in the following equations:

$$Recall = \frac{TP}{TP + FN} \quad (1)$$

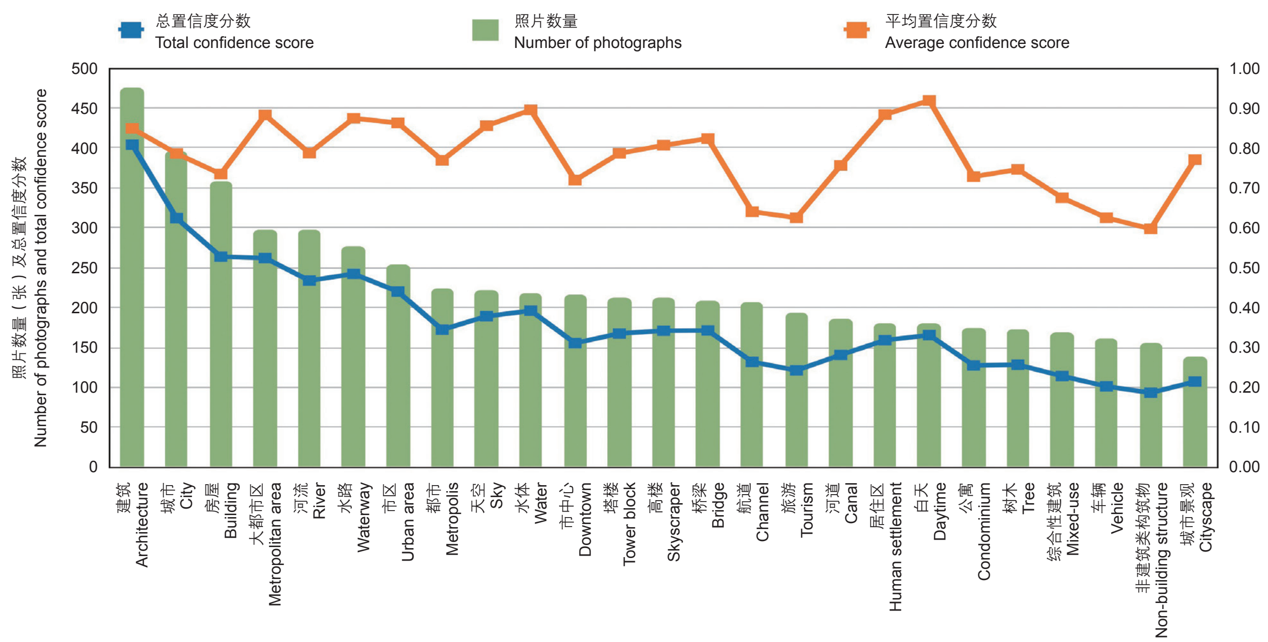
$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$F_1 = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (3)$$

## 3 Results

### 3.1 Keyword Extraction

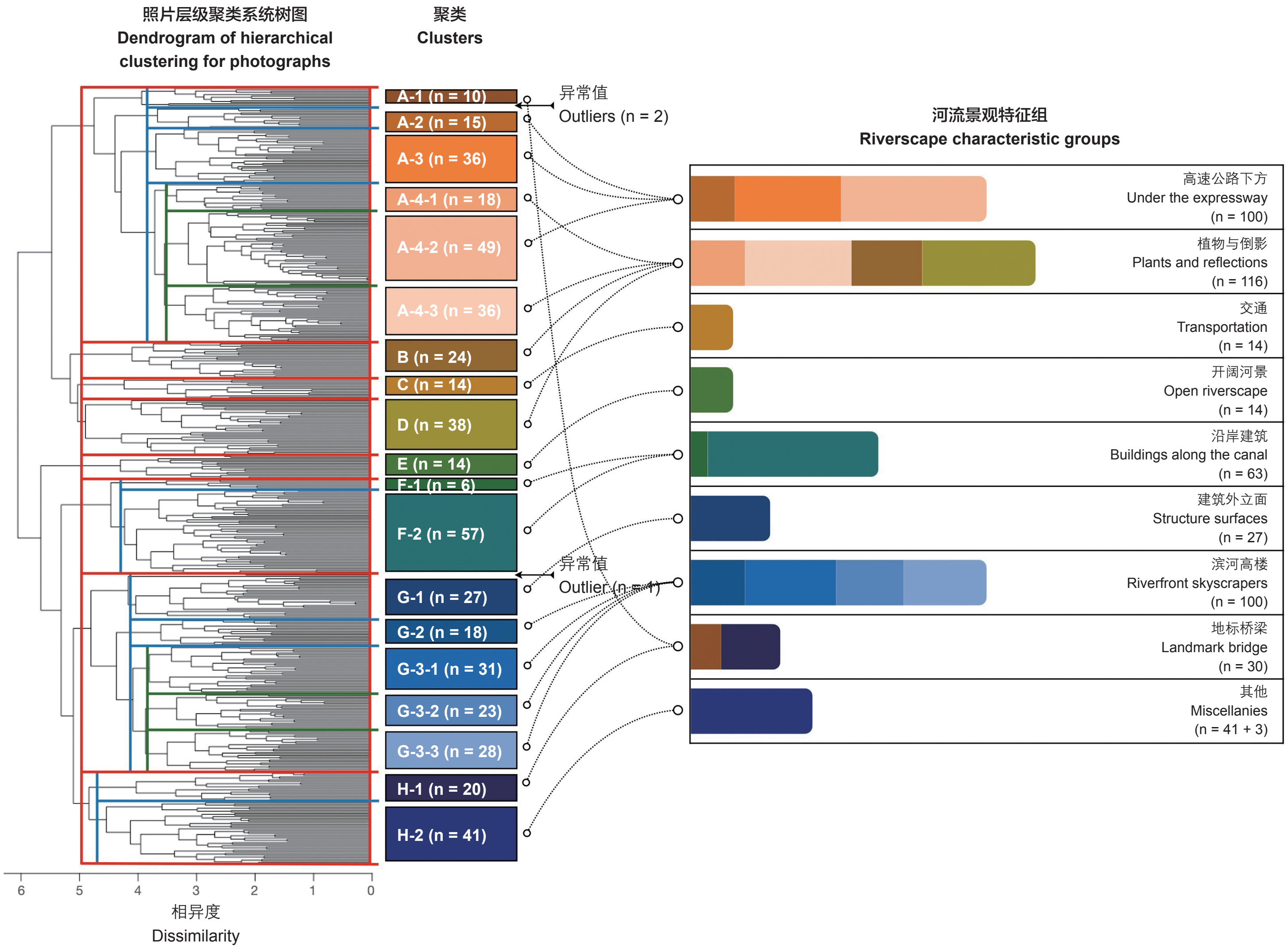
For the 508 photographs, the Google Cloud Vision API assigned a total of 9,946 keywords (mean = 19.58, SD = 8.99) with a maximum of 43 keywords (n = 1) and a minimum of 1 keyword (n = 4) per photograph. A total of 290 unique keywords were detected. Figure 2 displays a 2-axis chart of frequency, as well as the total and average confidence scores for the 25 most common keywords.




2

2. 出现频率最高的25个关键词及其对应的平均置信度和总置信度分数

2. The 25 most frequent keywords with average and total confidence scores






3. 由多层次聚类法生成的系统树图及将照片聚类匹配至相应的特征组。图中的距离矩阵基于关键词的置信度分数生成，红色、蓝色和绿色分割线分别代表三级聚类的相似性阈值。

3. Dendrogram generated by the multi-level clustering method and assignment of photograph clusters to characteristic groups. The distance matrix is assigned by the confidence scores of the keywords. The red, blue, and green cutting lines represent three levels of similarity thresholds.

总体而言，每个关键词的总置信度分数与照片数量呈现一致的趋势。虽然平均置信度分数在0.6~0.9之间波动，但这些关键词都十分贴切地描述了照片中的实体对象。同时，它们之中有些含义非常相近，如“建筑”和“房屋”，以及“城市”“市区”“市中心”和“城市景观”。相对地，诸如“天空”“旅游”“树木”和“车辆”等特定关键词尽管出现频率较低，相关关键词也较少，但它们有望在照片分类过程中发挥重要作用。

### 3.2 层次聚类结果

图3所示层次聚类系统树图展示了多层级聚类分析结果中不同聚类的排布，其中每一叶节点代表一张照片，而每一子树代表一个照片聚类[41]。将一级切分后获得的8个子树分别命名为A~H，而后对聚类中照

The trend of the total confidence score for each keyword is generally the same as the number of photographs. Although the average confidence scores fluctuated between 0.6 and 0.9, all these keywords could well describe the entity objects in the photographs, and some of them had similar meanings, such as “Architecture” and “Building” as well as “City,” “Urban area,” “Downtown,” and “Cityscape.” Comparatively, specific keywords, such as “Sky,” “Tourism,” “Tree,” and “Vehicles,” were detected less often and had fewer related keywords, but they were expected to play an important role in distinguishing photographs.

### 3.2 Hierarchical Clustering Results

As the output of the multi-level clustering analysis, the dendrogram in Figure 3 illustrates the arrangement of the clusters, where a leaf represents an individual photograph, and a subtree represents a cluster of photographs[41]. The 8 subtrees obtained from the first-level cut were named from A to H. Then, the second-level cut was conducted in A, F, G, and H because the sizes of these clusters were larger

片数量超过60的子树A、F、G和H聚类进行二级切分，直到所有聚类包含照片的数量均小于60——即完成三级分类后，结束递归多层级聚类。

如图3所示，系统树图的三个层级的切分以不同颜色标注——红色、蓝色和绿色的分割线分别代表第一、第二和第三级聚类的相似性阈值。它将树图切分为了19个互不重叠的聚类。

### 3.3 河流景观特征匹配

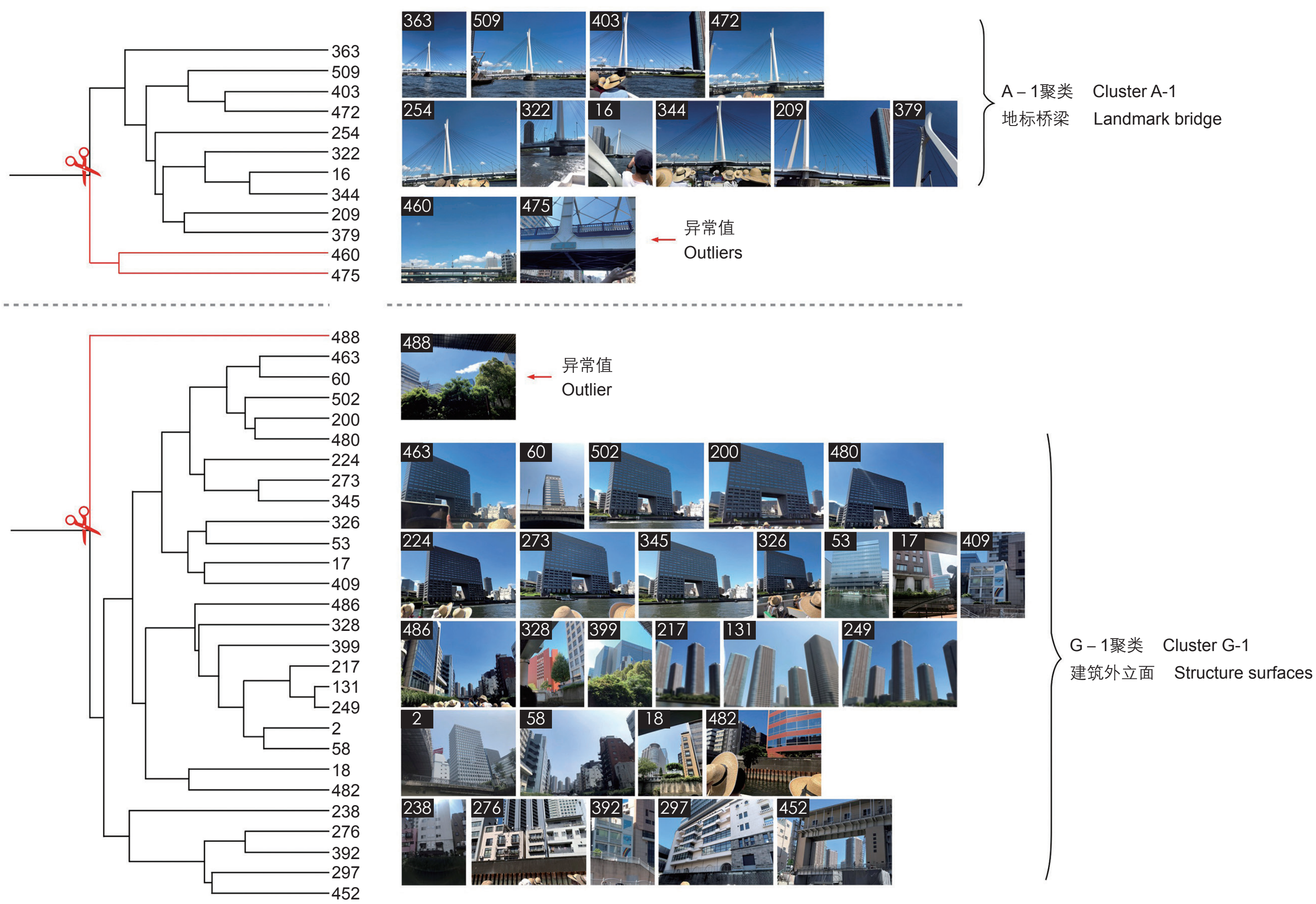
在确认了聚类分组并确保每个聚类中的照片都属于相同的主题后，研究团队依据特征的相似性，将内容高度相似的聚类合为一组。最终，确定了河流景观特征各不相同的9个组："高速公路下方""植物与倒影""交通""开阔河景""沿岸建筑""建筑外立面""滨河高楼""地标桥梁"和"其他"（图3）。其中，H—2聚类因照片内容包含多个对象，无法简单汇总，被归入"其他"组。此外，在针对A聚类和G聚类进行二级切分时，发现了3个异常图片，因明显有别于相邻的聚类，也被归入"其他"组（图4）。

than 60. The recursive multi-level clustering stopped at the third level because all clusters consisted of less than 60 photographs.

Three-level cuts in the dendrogram are emphasized with different colors in Figure 3. The red, blue, and green cutting lines indicate the similarity thresholds for the first, second, and third levels of clustering, respectively, which divides the dendrogram into 19 non-overlapping clusters.

### 3.3 Riverscape Characteristic Assignment

We confirmed the clustering results and ensured that the photographs in each cluster had a common visual theme. Clusters with highly similar contents were then combined into a single group based on their common characters. Finally, the researchers identified 9 groups with different riverscape characteristics: "Under the expressway," "Plants and reflections," "Transportation," "Open riverscape," "Buildings along the canal," "Structure surfaces," "Riverfront skyscrapers," "Landmark bridge" and "Miscellanies" (Fig. 3). Cluster H-2 was assigned to "Miscellanies" as the photographs in it contained a variety of subjects that could not be simply summarized. Three outliers were found in the second-level cut of Clusters A and G and also assigned to "Miscellanies" because they significantly differed from their adjacent clusters (Fig. 4).




© Shi Jiaying, Tsuyoshi Honjo, Yuriko Yazawa, Katsunori Furuya
4

4. 系统树图中检测到的异常值和对应照片

4. Outliers detected from hierarchical dendrogram and corresponding photos

如图3所示，4个较大规模的组——“植物与倒影”（116张）、“高速公路下方”（100张）、“滨河高楼”（100张）和“沿岸建筑”（63张）——呈现了研究区域内的主要景观特征；较小规模的组包括“交通”（14张）、“开阔河景”（14张）、“建筑外立面”（27张）和“地标桥梁”（30张），这些组的照片可能隐含着某些当地独特的景观特征。

图5为各聚类中最具代表性的照片，它们呈现了不同的河流景观特征。同时，图片右侧列出了每个聚类中出现频率最高的关键词及其平均置信度分数。人们对不同组别照片的视觉印象与直观感知一致，即此研究中的分类器在依据特征对照片进行分类的结果与人工分类结果相似度较高。实验中发现，同一组聚类中的关键词大多相似，且除“其他”组外，每个聚类都包含有与河流景观特征类型高度匹配的关键词。此外，尽管部分分属不同组别的照片看起来非常相似——例如，“开阔河景”中的E聚类和“滨河高楼”中的G—3—3聚类——但结合了关键词置信度分数的层级聚类可以准确将它们区分。

鉴于系统树图中的相邻子树大多归属于同一个景观特征组的聚类，从某种程度上讲，在前一阶段研究中对树图一级切分是有效的。尽管A—1和H—1聚类在系统树图中差异较大，但仍被归为了同一个特征组——“地标桥梁”组。相比之下，A—4—1、A—4—2和A—4—3聚类虽是树图中的相邻子树，但却分属两个不同的特征组。若此研究未采用多层级聚类算法，这些具有混合特征的聚类可能较难提取，使最终的聚类分组准确性降低。

如上所述，A—1和H—1聚类均包含相同且易辨识的元素——地标桥梁。这两个聚类尽管关键词的词组搭配不尽相同，但仍被归为同一特征组。研究团队在仔细观察后发现了其中的差异：高层建筑出现在H—1聚类的照片中，而未在A—1聚类中出现。这一有趣的发现表明，人们更易聚焦于照片中的主体元素，而本研究应用的分类器则关注于照片中的所有元素。再者，上述相邻的A—4—1、A—4—2和A—4—3聚类被划分至“高速公路下方”和“植物与倒影”两组，表明关键词“立交桥”“植物”和“倒影”在照片分类中起着重要的作用。此外，“植物”与“倒影”这两个关键词密切相关。

尽管“其他”组的照片内容杂糅、主题众多，但从H—2聚类中仍可以发现相互之间的共通点。具体来说，部分照片的主体为构筑物的细节，而部分照片则以游客为前景。如果对系统树图继续切分，那么可以此为标准进行。

From Figure 3, it can be seen that there are 4 large size groups: “Plants and reflections” (n = 116), “Under the expressway” (n = 100), “Riverfront Skyscrapers” (n = 100), and “Buildings along the canal” (n = 63). These groups reflected the main landscape characteristics of the study area. Small size groups were: “Transportation” (n = 14), “Open riverscape” (n = 14), “Structure surfaces” (n = 27), and “Landmark bridge” (n = 30). These groups hinted at some unique characteristics of the study area.

Figure 5 shows representative photographs of each cluster, which present different groups of riverscape characteristics. The most common keywords of each cluster with their average confidence scores are set out in the sidebar. The visual impression of these groups matches well with people’s intuitive feeling. The classifier’s performance in categorizing photographs by characteristics was similar to that of humans. We can find that patterns of keywords for clusters in the same group were similar, and the highly-related keywords of each cluster highly matched the types of riverscape characteristics, except in the Miscellanies group. Although some photographs look very similar between groups, e.g., Cluster E in “Open riverscape” and Cluster G-3-3 in “Riverfront skyscrapers,” the hierarchical clustering based on the confidence scores of keywords could still accurately distinguish them.

Clusters belonging to the same landscape characteristics were mostly adjacent branches of the dendrogram, proving that clustering by a single cut in the dendrogram in prior studies makes sense somehow. Although with the highest level of dissimilarity in the hierarchical structure, Cluster A-1 and Cluster H-1 were suited to the same characteristic group, “Landmark bridge.” In contrast, Cluster A-4-1, Cluster A-4-2, and Cluster A-4-3, though they were adjacent subtrees in the dendrogram, belonged to two different characteristic groups. If not for the multi-level clustering algorithm, these clusters with mixed characteristics may not have been extracted, but  divided into another characteristic group with their adjacent clusters.

Cluster A-1 and Cluster H-1, which were mentioned above, contained the same apparent element, the landmark bridge. Although their keyword patterns were different, they were still classified into the same group. By closer inspection of the figure, we can find the difference: tall buildings appeared in the photographs of Cluster H-1, but not in those of Cluster A-1. This intriguing finding indicates that humans prefer to focus on the unique elements in the picture, while the classifier in this study considers all factors in the view. Also mentioned above, the neighbors Cluster A-4-1, Cluster A-4-2, and Cluster A-4-3 were grouped into two groups of “Under the expressway” and “Plants and reflections,” which illustrates that the keywords “overpass,” “plant,” and “reflection” play an essential role in classification. Besides, the keyword “plant” has a strong correlation with the keyword “reflection.”

Although the content of the photographs in the “Miscellanies” group was multi-themed, we still find some patterns in Cluster H-2. In detail, some photographs presented the details of the structures, and some contained tourists as the foreground. If we continue to “cut” the dendrogram, it is possible to distinguish within this group.


5. 分属特定景观特征组的所有聚类照片示例，以及每个聚类中出现最频繁的关键词及其平均置信度分数。其中，在视觉上与景观特征高度相关的关键词用橙色突出显示。

5. Sample photographs of each cluster belonging to the specific landscape characteristic groups and most frequent keywords of each cluster with their average confidence scores. Keywords most visually high-related to landscape characteristics are highlighted in orange.

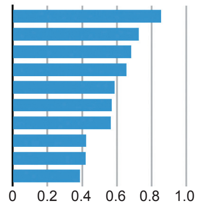
## 交通 Transportation

C聚类 Cluster C

出现频率最高的10个关键词
**Top 10 keywords**

车辆 Vehicle
水路 Waterway
小船 Boat
水路运输 Water transportation
建筑 Architecture
航道 Channel
船只 Watercraft
城市 City
河道 Canal
水体 Water

平均置信度分数
**Average confidence score**



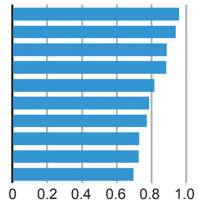

## 开阔河景 Open riverscape

E聚类 Cluster E

出现频率最高的10个关键词
**Top 10 keywords**

天空 Sky
白天 Daytime
城市 City
水体 Water
云朵 Cloud
河流 River
海洋 Sea
建筑 Architecture
水路 Waterway
大都市区 Metropolitan area

平均置信度分数
**Average confidence score**



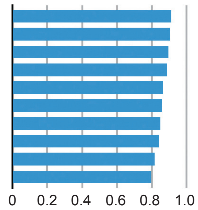

## 沿岸建筑 Buildings along the canal

F－1聚类 Cluster F-1

出现频率最高的10个关键词
**Top 10 keywords**

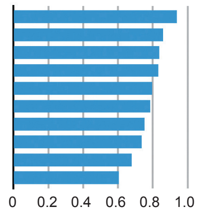
水路 Waterway
建筑 Architecture
公寓 Condominium
不动产 Property
城市 City
市区 Urban area
大都市区 Metropolitan area
房屋 Building
房地产 Real estate
公寓 Apartment

平均置信度分数
**Average confidence score**




F－2聚类 Cluster F-2

水路 Waterway
建筑 Architecture
城市 City
河流 River
水体 Water
河道 Canal
大都市区 Metropolitan area
房屋 Building
航道 Channel
市区 Urban area



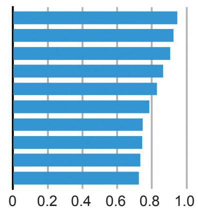

## 建筑外立面 Structure surfaces

G－1聚类 Cluster G-1

出现频率最高的10个关键词
**Top 10 keywords**

建筑 Architecture
大都市区 Metropolitan area
房屋 Building
城市 City
塔楼 Tower block
白天 Daytime
综合性建筑 Mixed-use
商业建筑 Commercial building
公寓 Condominium
都市 Metropolis

平均置信度分数
**Average confidence score**



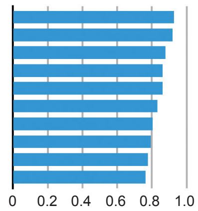

## 滨河高楼 Riverfront skyscrapers

G－2聚类 Cluster G-2

出现频率最高的10个关键词
**Top 10 keywords**

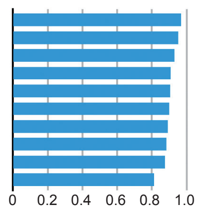
城市 City
大都市区 Metropolitan area
市区 Urban area
建筑 Architecture
河流 River
居住区 Human settlement
水路 Waterway
都市 Metropolis
高楼 Skyscraper
水路运输 Water transportation

平均置信度分数
**Average confidence score**




G－3－1聚类 Cluster G-3-1

大都市区 Metropolitan area
城市 City
高楼 Skyscraper
市区 Urban area
居住区 Human settlement
白天 Daytime
塔楼 Tower block
建筑 Architecture
都市 Metropolis
市中心 Downtown






5-2

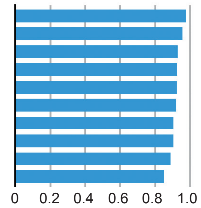
## 滨河高楼 Riverfront skyscrapers

G－3－2聚类 Cluster G-3-2

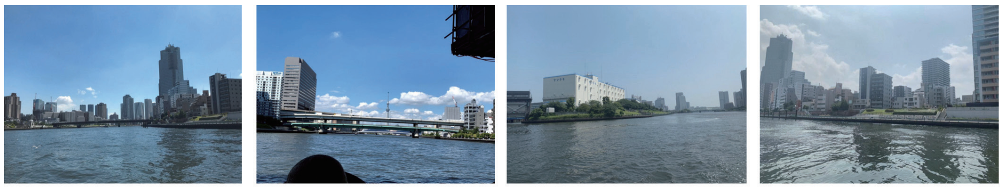
**出现频率最高的10个关键词**
**Top 10 keywords**

大都市区 Metropolitan area
城市 City
市区 Urban area
白天 Daytime
高楼 Skyscraper
塔楼 Tower block
建筑 Architecture
居住区 Human settlement
都市 Metropolis
水路 Waterway

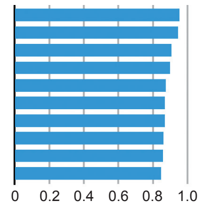
**平均置信度分数**
**Average confidence score**




G－3－3聚类 Cluster G-3-3



城市 City
大都市区 Metropolitan area
市区 Urban area
高楼 Skyscraper
天际线 Skyline
居住区 Human settlement
白天 Daytime
天空 Sky
河流 River
城市景观 Cityscape



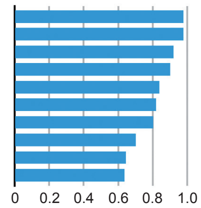

## 地标桥梁 Landmark bridge

A－1聚类 Cluster A-1

**出现频率最高的10个关键词**
**Top 10 keywords**

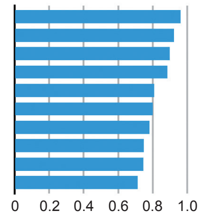
桥梁 Bridge
斜拉桥 Cable-stayed bridge
吊桥 Suspension bridge
矮塔斜拉桥 Extra dosed bridge
建筑 Architecture
天空 Sky
固定杆 Fixed link
非建筑类构筑物 Non-building structure
混凝土桥 Concrete bridge
地标 Landmark

**平均置信度分数**
**Average confidence score**



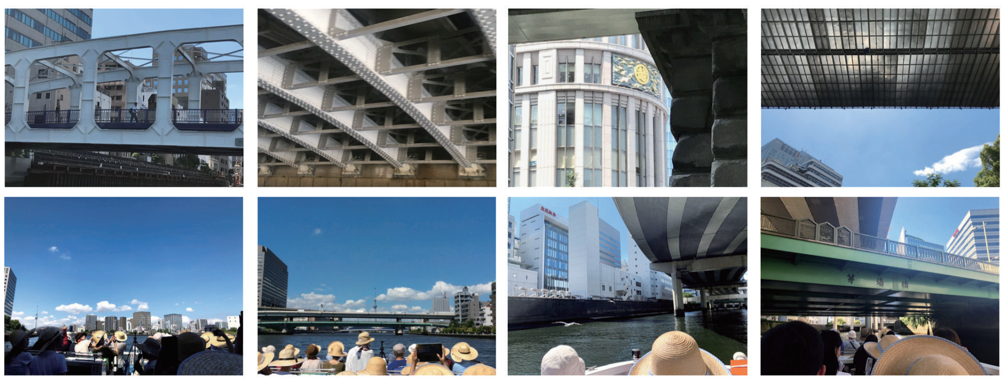

H－1聚类 Cluster H-1

桥梁 Bridge
大都市区 Metropolitan area
建筑 Architecture
城市 City
市区 Urban area
高楼 Skyscraper
河流 River
塔楼 Tower block
高架公路 Skyway
都市 Metropolis




## 其他 Miscellanies

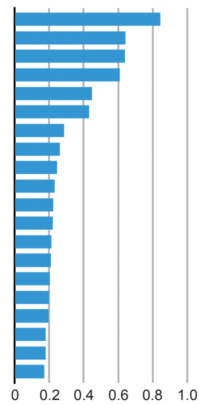
H－2聚类 Cluster H-2



**出现频率最高的20个关键词**
**Top 20 keywords**

建筑 Architecture
房屋 Building
城市 City
大都市区 Metropolitan area
市区 Urban area
天空 Sky
白天 Daytime
金属制品 Metal
都市 Metropolis
高楼 Skyscraper
旅游 Tourism
市中心 Downtown
综合性建筑 Mixed-use
立面 Facade
人群 Crowd
云朵 Cloud
日光 Daylighting
塔楼 Tower block
树木 Tree
窗户 Window

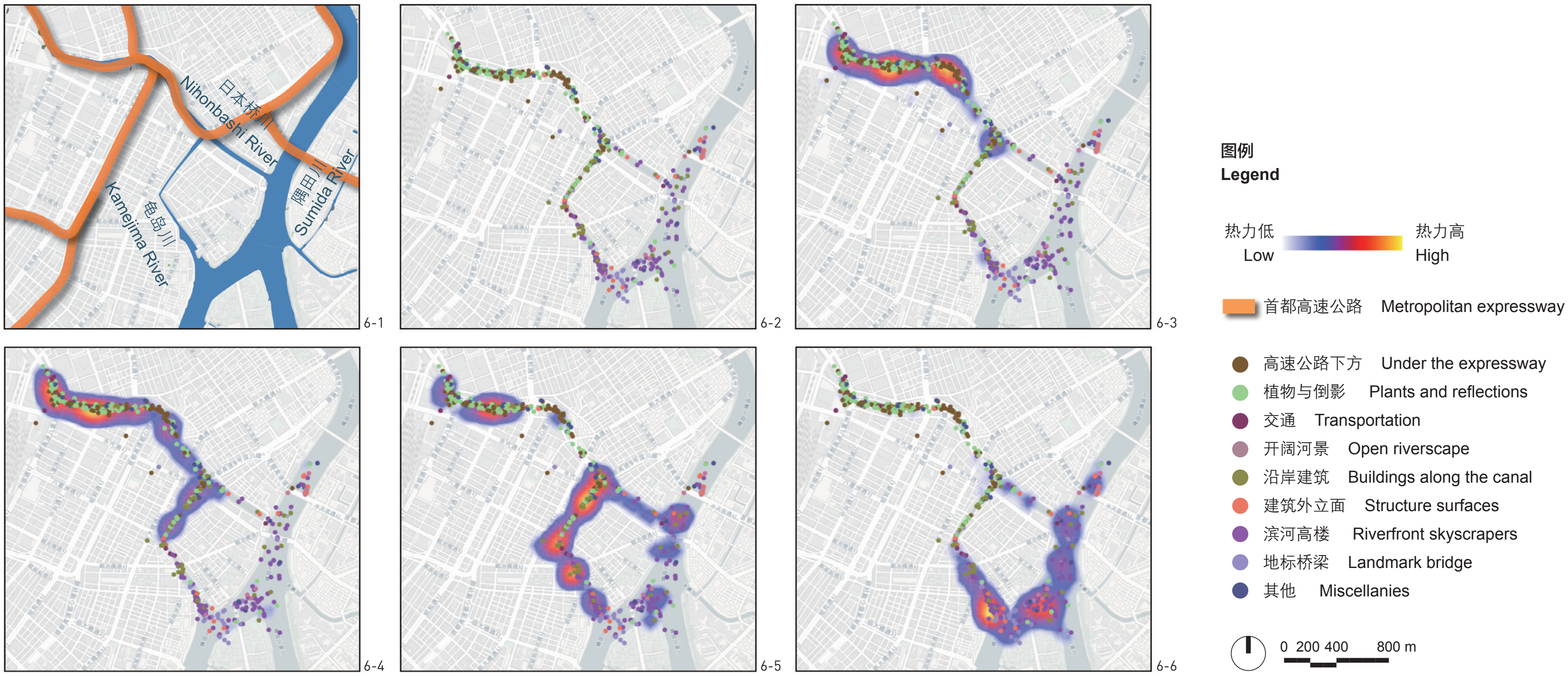
**平均置信度分数**
**Average confidence score**




异常值 Outliers



5-3




© Shi Jiaying, Tsuyoshi Honjo, Yuriko Yazawa, Katsunori Furuya
6

### 3.4 特征组的空间分布

研究还通过地图绘制分析了不同类型河流景观的空间特征。图6展示了首都高速公路地图、所有照片拍摄地点，以及反映了4个最大规模特征组（“高速公路下方”“植物与倒影”“沿岸建筑”和“滨河高楼”）拍摄地点密度的热力图。因其余组所含照片相对较少，较难形成热力图，本次研究不予讨论。

根据每组照片的密度，可推断可能影响游客观光体验的景观特征。如图所示，归类为“高速公路下方”和“植物与倒影”两组照片多拍摄于日本桥川段（图6—3，6—4）。“沿岸建筑”的照片主要拍摄于在游览初始阶段和龟岛川沿岸（图6—5）。“滨河高楼”的照片主要拍摄于隅田川沿岸，以及龟岛川和隅田川的下游交汇处（图6—6）。此外，在游览之初，距码头约650m的区域为了第一个热点，游客在此处拍摄了大量照片，照片主题多为植物、水中倒影、

6-1. 研究区域与首都高速公路地图
6-2. 所有照片的分布（不同颜色代表不同的特征组）
6-3. “高速公路下方”组照片的热力图
6-4. “植物与倒影”组照片的热力图
6-5. “沿岸建筑”组照片的热力图
6-6. “滨河高楼”组照片的热力图

6-1. Map of the study area with the Metropolitan Expressway
6-2. All photographs (classification indicated with different colors)
6-3. Heat map of "Under the expressway"
6-4. Heat map of "Plants and reflections"
6-5. Heat map of "Buildings along the canal"
6-6. Heat map of "Riverfront skyscrapers"

### 3.4 Mapping of the Characteristic Groups

Spatial characteristic analysis was performed by mapping different types of riverscapes. Figure 6 illustrates the position of the Metropolitan Expressway and all collected photographs, as well as heat maps of the 4 largest photograph groups, "Under the expressway," "Plants and reflections," "Buildings along the canal," and "Riverfront skyscrapers." The rest groups would not be discussed in this research because they contained relatively a small number of photographs and hardly formed a heat map.

Landscape characteristics that influenced visitors' viewing experience can lead to inferences based on the density of photographs in each group. It is apparent that photographs of both "Under the expressway" and "Plants and reflections" were concentrated in the Nihonbashi River section (Fig. 6-3, 6-4). "Buildings along the canal" mainly appeared at the beginning of the cruise and along the Kamejima River (Fig. 6-5). Photographs of "Riverfront skyscrapers" were primarily located along the Sumida River and the intersection of the lower reaches of the Kamejima River and Sumida River (Fig. 6-6). Visitors tended to take many pictures at the beginning of the cruise and formed a hotspot area around 650 m away from the dock. These photographs are on the themes relating to plants, water reflections, the overpass, and buildings. Although still under the canopy of the expressway, tourists seemed to lose interest in photographing the highway in the second half of the Nihonbashi River. A new hotspot can be observed when the ship sailed into the Kamejima River—Along the narrow canal, visitors took many urban valley-style photographs. With

立交桥和建筑物。在去程中当游览至日本桥川段的后半部分时，尽管仍然位于高速公路下方，游客已无兴趣再拍摄速公路。当游船驶入龟岛川经过狭窄的河道时，游客拍摄了许多类似“峡谷”的城市景观照片，地图上由此产生了一个新的热点。随着游览的进行，另一个重要的热点区域出现在龟岛川和隅田川的交界处——这里视野突然开阔，可以看到许多摩天大楼。

### 3.5 人工分类与分类器的一致性

研究者从除“其他”组以外的特征组（共464张照片）中，随机挑选出46张样本照片（图7）来评估分类器的效果。基于80%的效力和0.05的$\alpha$值，样本照片的数量符合kappa一致性检验在8 × 8列联表的最小样本量[42]。Kappa系数越高，则分类准确性越高。

如图8所示，混淆矩阵导入了不同分类途径对应的每一特征组照片的统计信息。矩阵横坐标和纵坐标分别对应人工分类和分类器分类的结

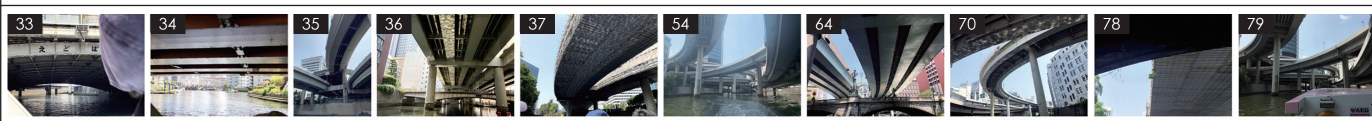
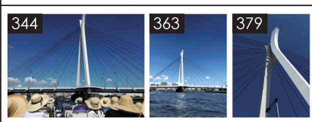
the cruise going, a striking hotspot appeared at the junction of the Kamejima River and the Sumida River, where visitors' vision suddenly widened, and many skyscrapers came into their sight.

### 3.5 Agreement Between Manual Classification and the Classifier

A total of 46 sample photographs (Fig. 7) were randomly selected from the 8 groups of 464 photographs, excluding the "Miscellanies" group, for assessing the performance of the classifier. The number of sample photographs met the minimum sample size for kappa agreement test of an 8-by-8 table based on power at 80% and alpha of 0.05[42]. The higher the kappa coefficient, the higher the classification accuracy.

As shown in Figure 8, a confusion matrix yielded the identification statistics for each group. The abscissa corresponds to the manual classification data, and the ordinate corresponds to the classifier data. The numbers appearing on

7. 用于准确性检验的样本照片
7. Sample photographs for the accuracy test

| 照片特征组 Photograph characteristic group | 照片总数（张） n (total) | 检验照片数量（张） n (test) | 000 分类正确的样本照片 Sample photograph classified correctly; 000 分类错误的样本照片 Sample photograph classified incorrectly |
|---|---|---|---|
| 高速公路下方 Under the expressway | 100 | 10 |  33, 34, 35, 36, 37, 54, 64, 70, 78, 79 |
| 植物与倒影 Plants and reflections | 116 | 12 |  355, 357, 359, 365, 367, 368, 378, 388, 390, 391, 395, 422 |
| 交通 Transportation | 14 | 1 |  161 |
| 开阔河景 Open riverscape | 14 | 1 |  413 |
| 沿岸建筑 Buildings along the canal | 63 | 6 |  96, 98, 145, 167, 278, 280 |
| 建筑外立面 Structure surfaces | 28 | 3 |  276, 273, 480 |
| 滨河高楼 Riverfront skyscrapers | 100 | 10 |  51, 111, 149, 327, 382, 431, 477, 494, 229, 248 |
| 地标桥梁 Landmark bridge | 32 | 3 |  344, 363, 379 |


7

真值集（人工分类）中相应类型照片数量（张）
Number of photographs in the truth set (manual)

预测集（分类器分类）中相应类型照片数量（张）
Number of photographs in the predicted set (classifier)

| | 高速公路下方<br>Under the expressway | 植物与倒影<br>Plants and reflections | 交通<br>Transportation | 开阔河景<br>Open riverscape | 沿岸建筑<br>Buildings along the canal | 建筑外立面<br>Structure surfaces | 滨河高楼<br>Riverfront skyscrapers | 地标桥梁<br>Landmark bridge | 总数（分类器分类）<br>Total number (classifier) |
|---|---|---|---|---|---|---|---|---|---|
| 高速公路下方<br>Under the expressway | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 |
| 植物与倒影<br>Plants and reflections | 1 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| 交通<br>Transportation | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 开阔河景<br>Open riverscape | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 沿岸建筑<br>Buildings along the canal | 0 | 0 | 0 | 0 | 4 | 0 | 2 | 0 | 6 |
| 建筑外立面<br>Structure surfaces | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 3 |
| 滨河高楼<br>Riverfront skyscrapers | 0 | 0 | 0 | 1 | 0 | 4 | 5 | 0 | 10 |
| 地标桥梁<br>Landmark bridge | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 3 |
| 总数（人工分类）<br>Total number (manual) | 11 | 11 | 1 | 2 | 4 | 7 | 7 | 3 | 46 |



8

8. 多分类混淆矩阵
8. Multiclass confusion matrix

**表2：多分类混淆矩阵结果**
**Table 2: Results of multiclass confusion matrix**

| 特征组<br>Group | 精确率<br>Precision | 召回率<br>Recall | $F_1$值<br>$F_1$ score |
|---|---|---|---|
| 高速公路下方<br>Under the expressway | 1.00 | 0.91 | 0.95 |
| 植物与倒影<br>Plants and reflections | 0.92 | 1.00 | 0.96 |
| 交通<br>Transportation | 1.00 | 1.00 | 1.00 |
| 开阔河景<br>Open riverscape | 1.00 | 0.50 | 0.67 |
| 沿岸建筑<br>Buildings along the canal | 0.67 | 1.00 | 0.80 |
| 建筑外立面<br>Structure surfaces | 1.00 | 0.43 | 0.60 |
| 滨河高楼<br>Riverfront skyscrapers | 0.50 | 0.71 | 0.59 |
| 地标桥梁<br>Landmark bridge | 1.00 | 1.00 | 1.00 |

the diagonal line indicate where the classifier was consistent with the manual classification, which can be considered as the number of samples correctly classified, and the data appearing outside the diagonal line are the numbers of samples poorly classified. Obviously, the number of true positive cases (photographs that were classified correctly) was 38, the overall accuracy was 82.61% (38/46), and Cohen's Kappa was 0.79, indicating a strong agreement between manual classification and the classifier[43]. The results generated from the matrix are listed in Table 2. Thus, we believe that the method of this study was fast and efficient for homogeneous landscape classification.

Overall, the classifier achieved very good precision and recall outcomes. The categories "Landmark bridge" and "Transportation" achieved the best accuracy, followed closely by "Under the expressway," "Plants and reflections," and "Buildings along the canal." According to the $F_1$ score, confusion occurred primarily in "Open riverscape," "Structure surfaces," and "Riverfront skyscrapers." Despite their low recall, "Open riverscape" and "Structure surfaces" obtained very high precision statistics. However, the classifier performed poorly on "Riverfront skyscrapers," with relatively lower precision and recall.

果。对角线上的数字表示分类器与人工分类结果一致的照片数量，可视之为正确分类的样本数；对角线外的数据为分类不当的样本数。由结果可知，真正例（正确分类的照片）的数量为38例，总体准确率为82.61%（38／46），kappa系数为0.79，表明人工分类和分类器之间存在高度的一致性[43]。表2所示混淆矩阵得出的结果表明，本研究的方法对于均质化景观的分类是快速有效的。

总体而言，该分类器实现了较高的精确率和召回率。“地标桥梁”和“交通”组别的准确性最高，“高速公路下方”“植物与倒影”和“沿岸建筑”三组次之。根据$F_1$值结果，“开阔河景”“建筑外立面”和“滨河高楼”三组常会发生混淆。尽管这三组的召回率均较低，但“开阔河景”和“建筑外立面”两组的精确率较高。然而，该分类器在“滨河高楼”组的分类上表现不佳，精确率和召回率均相对较低。

## 4 讨论

### 4.1 均质化照片分类的实用算法

用于层次分析的聚类数量是照片分类过程的重要参数之一。以往研究涉及的照片内容较为多样，提取的关键词差异性也较大，如动物、食物、花卉和人[28][29]。因此，设定单个相似性阈值——只进行一级切分形成树图——即已足够。在这种情况下，常用的分类标准（如贝叶斯信息准则）可帮助确定聚类的数量。

但本研究中的照片内容均质化程度高，关键词也较为分散，难以通过常用的分类标准获取最佳的聚类数量。因此，研究采用了多层次聚类法，更加切实有效地将相似的照片自动分类到设有数量上限的聚类中，具有相似特征的照片均被归为一组（图5）。此外，研究所使用分类器的分类结果与人工分类的一致性较强（图8）。综上所述，该方法能够对大部分拍摄了均质化景观的旅游照片进行有效分类，但未来仍需进一步探索更合理的方法来确定聚类数量的最佳值。

### 4.2 人工分类与机器分类的差异

图9为被错误分类的照片示例。在图9—1所示的高精确率、低召回率的情况下，该分类器的筛选过于精细，可能误选了实际属于“开阔河景”或“建筑外立面”组的照片。低精确率但高召回率的情况则意味着该分类器将部分在人工分类时被视为其他组的照片，归入了“沿岸建筑”组（图9—2）。

由图9可得出一个重要结论，即“滨河高楼”“建筑外立面”和“沿岸建筑”三个特征组的照片较易混淆。在进行准确性检验时，参与人员会将某些高度相似的样本照片归入一个特定的组。尽管研究团队是根据相应组别中大多数照片的共同特征为该组命名，但仍有照片与其他组照片更为相似。

机器错误分类的原因有多种：首先，较易混淆的照片常包含相似元素，因此它们匹配的关键词也较相似；其次，云视觉API标注的关键词会出现偏差或遗漏；再者，关键词的置信度分数也会影响分类结果，一些置信度分数较低的关键词也可发挥重要作用。

## 4 Discussions

### 4.1 Practical Algorithms for Homogeneous Photograph Classification

The number of clusters for hierarchical analysis is one of the critical parameters during the classification process. Previous studies involved a wide variety of photographs and achieved keywords that are easy to distinguish between, such as animals, food, flowers, and people[28][29]. In this case, a single similarity threshold is usually adequate, which means cutting the dendrogram at a single height. Standard criteria methods, such as the Bayesian information criterion, are useful for determining the number of clusters.

However, photographs in this current study were homogeneous, and keywords were not very clustered. It is difficult to find the optimal number of clusters using the criteria method. Hence, a more practical method, multi-level hierarchical clustering, was developed to automatically classifying similar photographs into groups with an upper limit size. Figure 5 shows the result that photographs with similar characteristics appear to be grouped together holistically. In addition, the classifier achieved a good agreement with manual classification (Fig. 8). These findings indicate that the methods used in this study are practical for classifying tourist photographs of largely homogeneous landscapes. Further studies should explore more rational ways to determine the optimal size of the cluster.

### 4.2 Differences Between Manual and Machine Classification

Figure 9 shows the sample photographs that were classified incorrectly. The cases with high-precision but low-recall indicate that the classifier is too selective, which may miss photographs actually belong to “Open riverscape” or “Structure surfaces” as Figure 9-1 shows. In contrast, the low-precision but high-recall cases mean the classifier categorizes some photographs as “Buildings along the canal,” however, they were considered as other groups in manual classification (Fig. 9-2).

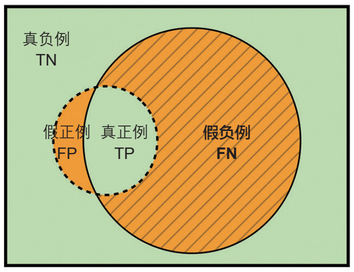
A significant outcome from Figure 9 is that photographs are easily confused between the groups of “Riverfront skyscrapers,” “Structure surfaces,” and “Buildings along the river.” Testers classified the sample photographs into a specific group because they saw highly similar photographs in that group. Although we named a group depending on the common characteristics of most photographs in that group, some photographs looked very similar to other groups.

The reasons for the misclassification are, first, elements of the confused photographs are similar, thus the keywords between them are similar; second, there are some errors or omissions of keywords when annotating from the Cloud Vision API; third, the confidence scores of keywords influence the classification because some keywords with low confidence scores are actually significant.

### 4.3 Clustering by Continuous Confidence Scores or Binary Confidence Scores

The conclusions above are all based on the dataset, which consists of keywords assigned by continuous scores from 0.5 to 1. Through the accuracy test between manual classification and the classifier, we found that one of the factors that

—— 真值集（人工分类） Truth set (manual)
········ 预测集（分类器分类） Predicted set (classifier)



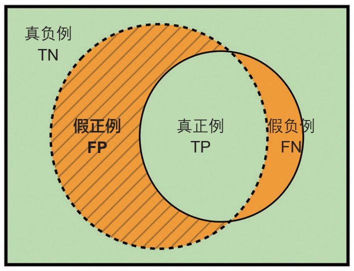

**高精确率、低召回率**
**High-precision, low-recall**

人工分类结果：开阔河景
分类器分类结果：滨河高楼
Manual: Open riverscape
Classifier: Riverfront skyscraper

人工分类结果：建筑外立面
分类器分类结果：滨河高楼
Manual: Structure surfaces
Classifier: Riverfront skyscrapers

人工分类结果：建筑外立面
分类器分类结果：滨河高楼
Manual: Structure surfaces
Classifier: Riverfront skyscrapers

© Shi Jiaying, Tsuyoshi Honjo, Yuriko Yazawa, Katsunori Furuya
9-1




**低精确率、高召回率**
**Low-precision, high-recall**

人工分类结果：滨河高楼
分类器分类结果：沿岸建筑
Manual: Riverfront skyscrapers
Classifier: Buildings along the canal

人工分类结果：滨河高楼
分类器分类结果：沿岸建筑
Manual: Riverfront skyscrapers
Classifier: Buildings along the canal

© Shi Jiaying, Tsuyoshi Honjo, Yuriko Yazawa, Katsunori Furuya
9-2

### 4.3 依据连续置信度分数或二进制置信度分数分类

以上结论均基于由关键词（置信度分数为0.5~1）组成的数据集得出。分类器相较人工分类的准确性测试的结果表明，关键词的置信度分数是影响分类准确性的因素之一。但在实验中，某些置信度分数较低的关键词同样可以呈现出照片的重要特征。基于此，研究团队探究了另一种方法：通过将原置信度分数取整——转换为二进制置信度分数（0或1）——组成关键词数据集。新的数据集不采纳关键词的实际置信度分数，每个被检测的关键词的权重均相同。随后通过与2.4小节中描述的方法进行分类。

在多层级聚类分类后，研究中出现了两个与其他特征组相似度低，但相互具有共同特征（游客出现在了前景中）的聚类。于是，这两个聚类被归为一个新的组别，即“游客”。其余组别的命名方式与上述连续置信度分数数据集中的方式相同。

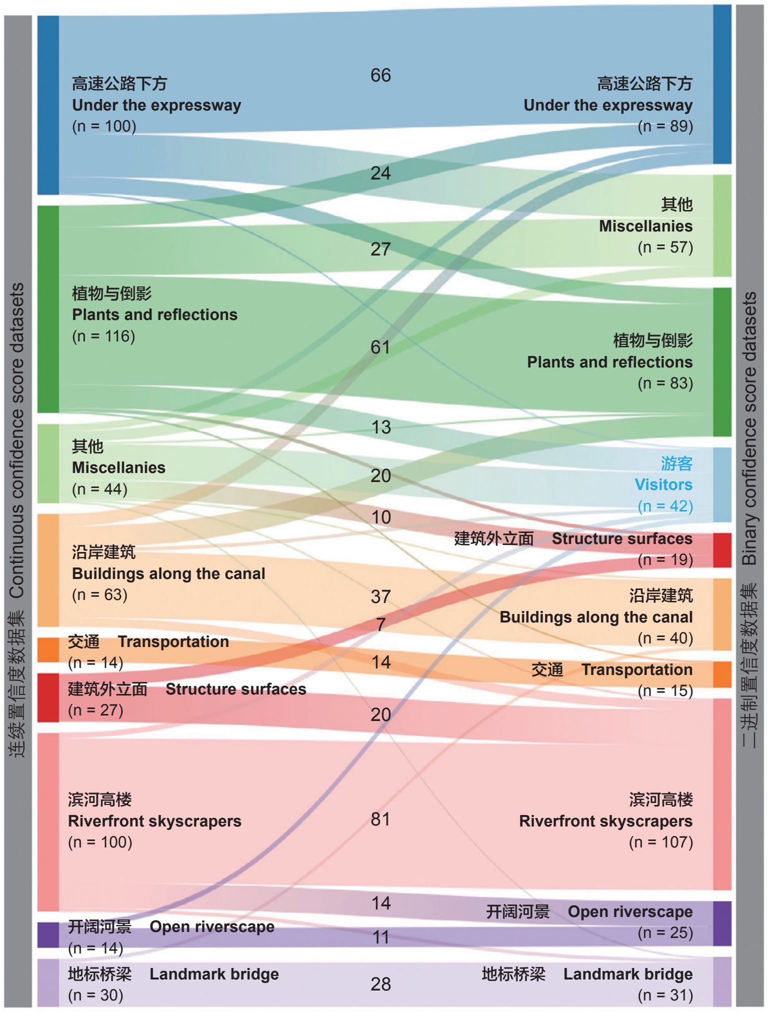
结果表明，二进制置信度分数数据集的分类情况与连续置信度分数数据集的一致性中等（kappa系数为0.55）。图10所示桑基图呈现了两个数据集之间的分组差异。如图，在大多数组别中，最宽的分支为两个数据集在同一组别下相同的部分。具体而言，“高速公路下方”“植物与倒影”“沿岸建筑”“交通”“滨河高楼”和“地标桥梁”这6组的分类结果并未发生太大变化，且两个数据集中的“交通”和“地标桥梁”两组的一致性非常高。此外，这两个特征组在人工分类中也具有更高的一致性（图8）。

influences accuracy is the confidence score. Some keywords with low scores represent significant characteristics of the photograph. Therefore, we tested another method to express the dataset of keywords with a binary confidence score (0 or 1) by rounding the original value instead of assigning continuous values. In the dataset, the weight of each detected keyword was the same, regardless of its actual confidence score. The subsequent classification process was consistent with the methods described in Section 2.4.

After multi-level clustering, we found two clusters that cannot be classified as any other group. Photographs in these two clusters have the same characteristics: tourists appear in the foreground. Therefore, we categorized these two clusters

9-1. 准确性模式维恩图和错误分类的照片示例：具有高精确率和低召回率的假负例照片（预测为负，实际为真）。
9-2. 准确性模式维恩图和错误分类的照片示例：具有低精确率和高召回率的假正例照片（预测为正，实际为假）。

9-1. Venn diagrams of accuracy patterns and incorrectly classified photographs: false-negative photographs (predicted "negative" and actually "true") in the groups with high-precision but low-recall.
9-2. Venn diagrams of accuracy patterns and incorrectly classified photographs: false-positive photographs (predicted "positive" but actually "false") in the groups with low-precision but high-recall.






10

10. 用于比较连续和二进制置信度分数数据集分类结果的桑基图

10. Sankey charts comparing the classification results between the continuous and binary datasets

两个数据集的分类结果中差异最大的是“其他”和“游客”两组。连续置信度分数数据集中“其他”组的照片主要被归入二进制置信度分数数据集的“游客”和“建筑外立面”两组，而二进制置信度分数数据集中“游客”的照片则主要来自连续置信度分数数据集的“植物与倒影”与“其他”两组。由于关键词“旅游”与照片中游客的出现高度相关，上述分类结果可能很大程度上受到该关键词的影响。如图2所示，关键词“旅游”的出现频率较高，但平均置信度较低。由于二进制置信度的分类方式增加了其在聚类分析中的权重，因此在此数据集出现了“游客”这一新的特征组。基于此，本研究建议，在未来实验中，当连续置信度分数数据集难以获得理想的分类结果时，可尝试使用二进制置信度分数数据集进行聚类分析。若后者分类结果更优，则那些置信度较低的关键词较重要。这一论述还需通过未来的研究进一步证实。

### 4.4 未来应用与改进

尽管通过绘制照片拍摄地点的地图进行空间特征分析并非本次研究的主要目的，这些热力图反映出的规律一方面证实了该分类器的有

as a new group, “Visitors.” The other groups were named the same as in the continuous dataset.

The agreement between the classification results of the continuous datasets and the binary datasets was moderate, with a Cohen’s kappa of 0.55. Figure 10 shows the Sankey charts displaying the differences between the two datasets. From the diagram, we can see in most groups that the thickest band shows pair-correspondence between the two datasets, which reflects the proportion of the same photographs in the same group. So, we see that the classification results for “Under the expressway,” “Plants and reflections,” “Buildings along the canal,” “Transportation,” “Riverfront skyscrapers,” and “Landmark bridge” do not change much. Especially “Landmark bridge” and “Transportation” have a high consistency between the two datasets. Interestingly, these groups also received a higher agreement with the manual classification (Fig. 8).

The most changes occurred in the groups “Miscellanies” and “Visitors.” Photographs of “Miscellanies” in the continuous dataset mainly flowed to “Visitors” and “Structure surfaces” in the binary dataset. Photographs of “Visitors” in the binary datasets were mostly from “Plants and reflections” and “Miscellanies” in the continuous dataset. These findings could be attributed to the change in the keyword “tourism,” which is highly related to the appearance of visitors in the photograph. Figure 2 shows that “tourism” is detected at a high frequency but low average score. Because the binary dataset increases its weight in the clustering, a new group, “Visitors” appears. Therefore, we put forward a suggestion that if researchers use the continuous dataset and find that the classification result not ideal, they can use the binary dataset to perform cluster analysis instead. If the classification result based on binary dataset is better than the continuous dataset, it means that those keywords with low confidence scores are important. This discussion needs to be further confirmed by future research.

### 4.4 Application and Improvements in the Future

It should be noted that although spatial analysis through the mapping of characteristic groups is not the main purpose of this research, the law reflected by the heat maps on the one hand confirms the effectiveness of the classifier, and on the other hand shows the potential of spatial analysis through the classifier. According to the heat maps, it is evident that the abrupt change in space evoked visitors’ higher interest. Visitors can easily get tired of repetitive or constant scenery, especially in a sequential landscape. Therefore, through the preliminary analysis, the landscape quality at the lower reaches of the Nihonbashi River needs to be improved. In addition to the initial section of the tour, the number of photographs related to plants is not large, and their distribution is also very average. This shows that the soft landscape in the study area does not attract tourists. Such knowledge will be useful in creating an attractive landscape for the “Nihonbashi Revitalization Project.”

效性，另一方面也展示了它应用于空间分析的潜力。由热力图可明显看出，游客极易对重复或持续出现的景观感到厌倦，突然变化的空间会激发游客强烈的拍照兴趣。因此，基于上文的初步分析可知，日本桥川下游的景观品质有待提升。除游览初始段外，其他阶段拍摄的植物照片数量较少，且拍摄位置分布也非常均匀。这表明研究区域的软质景观对游客的吸引力较弱。如上研究成果可为“日本桥区复兴计划”提供参考，以构建更具吸引力的景观。

本研究亦显示了该分类器在未来旅游管理和景观评估（或基于旅游业的景观评估）等研究中的应用潜力。例如，英国的“城市景观特征评估”方法要求规划者识别城区的景观特征，并依据其对应的土地单元类型进行归类[44]。此外，该分类器也可以分析特定的景观类型或元素，例如，通过追踪由API自动标注了“地标”标签照片的地理信息，决策者可以规划最适宜观赏“地标”的位置。

未来研究需进一步提升量化场所景观特征结果的准确性。方法之一是调整在照片分类时起重要作用的关键词的权重和对应关系；在这一方法中，照片的地理信息（经度和纬度）是影响聚类分析的另一因素。在本研究中，照片的关键词大多由谷歌云视觉API标注，这些标签均为客观名词，不包含“美丽”“强烈”等形容词或描述颜色及情绪的短语。由此推测，本研究中河流景观的美学质量无从可知。随着科技的发展，机器学习已可辨别更多新的特征，如主体色调、不同对象和人物面部表情等。基于这些成果的未来研究将探索更多发现，并实现景观照片识别的优化。正如凯文·林奇[45]所说，“……图像研究非常有必要，它将有助于引导决策。”照片识别和分类准确性的提高，将帮助人们认识景观特征与公众偏好之间的关系。

## 5 结语

本研究利用谷歌云视觉API标注的关键词，开发了一个基于多层级聚类的分类器，以对大量均质化景观照片进行分类。结果表明，该分类器实用且高效，可实现对河流景观照片的半自动化分类，且结果与人的直观感受大体一致。针对不同特征组照片拍摄地点绘制的热力图也可帮助评估游客的偏好与感知。此外，准确性检验结果表明，该分类器的分类结果与人工分类相符。最后，研究团队对比了连续置信度分数数据集和二进制置信度分数数据集的分类结果，发现当重要关键词的置信度分数较低时，后一种数据集将更加有效。本研究中的方法可应用于旅游管理和景观评估中，尤其是那些需要对大量均质化景观照片进行分类的情况。基于照片自动化分类这项技术，上述方法不仅可以提高对不同图片局部差异的敏感度，也有助于推动评估过程的公众参与及旅游管理的优化。LAF

The results of this study are encouraging future application of the classifier in tourism management and landscape assessment. The classifier can also contribute to landscape evaluation for tourism. For example, the Townscape Character Assessment in the UK requires planners to identify the landscape character types in urban areas and divide them into corresponding land units[44]. In addition, the classifier can analyze the specific landscape type or element, for example, by tracking the photographs with the “landmark” label (fed back by the annotation API automatically), decision-makers can determine the most attractive place for sightseeing the “landmark.”

Further studies are needed to improve the accuracy of the destination landscape characteristics being quantified. Adjusting the weights and relations of keywords that play a significant role in classification is one of the approaches. The location of photographs (longitude and latitude) can also be considered as a factor in the clustering analysis. In addition, the annotation keywords in this study highly depend on the Google Cloud Vision API. They are all entity labels and do not include adjectives like “beautiful,” “strong,” or phrases describing colors or moods. Thus the aesthetic quality of the landscape is unknown. Recently, machine learning has added more characteristics, such as detecting the dominant colors, multiple objects, and facial expressions. A greater focus on these achievements could produce interesting findings that improve the identification of landscape photographs. As Kevin Lynch stated[45], “… image research is absolutely necessary, and policy recommendations can be easily made.” The more accurate the identification and classification of photographs becomes, the better the understanding of the relationship between landscape characteristics and public preferences will be.

## 5 Conclusions

This study used keywords annotated by the Google Cloud Vision API and developed a classifier based on multi-level hierarchical clustering, which was then used to classify a large number of photographs of homogeneous landscape. The results indicated that the classifier is practical and efficient; it can semi-automatically classify photographs of riverscapes, and the results match human intuition well. The mapping of the photograph groups provides credentials for evaluating the preferences and perceptions of visitors. The accuracy test established that the classifier agreed well with manual classification. We compared the results of clustering by datasets of continuous confidence scores and binary confidence scores. The binary dataset was useful if important keywords sometimes had low confidence scores. Methods in this study can be applied for practical tourism management and landscape evaluation, especially in situations where large numbers of photographs of a homogeneous landscape need to be classified. Within automated image classification, these methods help improve sensitivity to local variability. It also provides the possibility of public participation in the assessing process, which, in turn, contributes to tourism management. LAF

## REFERENCES


[1] Brabyn, L. (2009). Classifying Landscape Character. *Landscape Research*, 34(3), 299–321. https://doi.org/10.1080/01426390802371202

[2] Jacobsen, J. K. S. (2007). Use of Landscape Perception Methods in Tourism Studies: A Review of Photo-Based Research Approaches. *Tourism Geographies*, 9(3), 234–253. https://doi.org/10.1080/14616680701422871

[3] Balomenou, N., & Garrod, B. (2014). Using volunteer-employed photography to inform tourism planning decisions: A study of St David's Peninsula, Wales. *Tourism Management*, (44), 126–139. https://doi.org/10.1016/j.tourman.2014.02.015

[4] Bubalo, M., van Zanten, B. T., & Verburg, P. H. (2019). Crowdsourcing geo-information on landscape perceptions and preferences: A review. *Landscape and Urban Planning*, (184), 101–111. https://doi.org/10.1016/j.landurbplan.2019.01.001

[5] Deng, N., & Li, X. R. (2018). Feeling a destination through the "right" photos: A machine learning model for DMOs' photo selection. *Tourism Management*, (65), 267–278. https://doi.org/10.1016/j.tourman.2017.09.010

[6] Dunkel, A. (2015). Visualizing the perceived environment using crowdsourced photo geodata. *Landscape and Urban Planning*, (142), 173–186. https://doi.org/10.1016/j.landurbplan.2015.02.022

[7] Casalegno, S., Inger, R., DeSilvey, C., & Gaston, K. J. (2013). Spatial Covariance between Aesthetic Value & Other Ecosystem Services. *PLOS ONE*, 8(6), 1–5. https://doi.org/10.1371/journal.pone.0068437

[8] Depellegrin, D., Blažauskas, N., & Vigl, L. E. (2012). Aesthetic value characterization of landscapes in coastal zones. *2012 IEEE/OES Baltic International Symposium (BALTIC)*, 1–6. https://doi.org/10.1109/BALTIC.2012.6249166

[9] Gliozzo, G., Pettorelli, N., & Haklay, M. M. (2016). Using crowdsourced imagery to detect cultural ecosystem services: A case study in South Wales, UK. *Ecology and Society*, 21(3), 6. https://doi.org/10.5751/ES-08436-210306

[10] Tenerelli, P., Demšar, U., & Luque, S. (2016). Crowdsourcing indicators for cultural ecosystem services: A geographically weighted approach for mountain landscapes. *Ecological Indicators*, (64), 237–248. https://doi.org/10.1016/j.ecolind.2015.12.042

[11] van Zanten, B. T., Van Berkel, D. B., Meentemeyer, R. K., Smith, J. W., Tieskens, K. F., & Verburg, P. H. (2016). Continental-scale quantification of landscape values using social media data. *Proceedings of the National Academy of Sciences*, 113(46), 12974–12979. https://doi.org/10.1073/pnas.1614158113

[12] Zhai, X., & Lange, E. (2020). Using Social Media to Explore Perceptions of Ecosystem Services by Nature-Based Solution Projects. *Landscape Architecture Frontiers*, 8(3), 58–77. https://doi.org/10.15302/J-LAF-1-020030

[13] Balomenou, N., & Garrod, B. (2019). Photographs in tourism research: Prejudice, power, performance and participant-generated images. *Tourism Management*, (70), 201–217. https://doi.org/10.1016/j.tourman.2018.08.014

[14] Oteros-Rozas, E., Martín-López, B., Fagerholm, N., Bieling, C., & Plieninger, T. (2018). Using social media photos to explore the relation between cultural ecosystem services and landscape features across five European sites. *Ecological Indicators*, (94), 74–86. https://doi.org/10.1016/j.ecolind.2017.02.009

[15] Gobster, P. H., Ribe, R. G., & Palmer, J. F. (2019). Themes and trends in visual assessment research: Introduction to the Landscape and Urban Planning special collection on the visual assessment of landscapes. *Landscape and Urban Planning*, (191), 1–7. https://doi.org/10.1016/j.landurbplan.2019.103635

[16] Ives (Prieskorn), E., Burley, J. B., Russcher, K., Schutzki, R., & Zhou, J. (2021). Visual Metrics for the Maxton Plains Alvars in Michigan, USA. *Landscape Architecture Frontiers*, 9(2), 26–36. https://doi.org/10.15302/J-LAF-1-020043

[17] Liu, Y., Mou, T., Huang, Z., & Ha, R. (2019). Research on Visually Perceived Image and Strategies for Vista View System Improvement of the Jinsitao Waterfront of Shichahai in Beijing. *Landscape Architecture Frontiers*, 7(2), 121–131. https://doi.org/10.15302/J-LAF-20190212

[18] Robert, S. (2018). Assessing the visual landscape potential of coastal territories for spatial planning. A case study in the French Mediterranean. *Land Use Policy*, (72), 138–151. https://doi.org/10.1016/j.landusepol.2017.12.037

[19] Cartwright, D. P. (1988). Analiza Kvalitativnog Materijala [Analysis of Qualitative Material]. *Revija za Sociologiju*, 19(1–2), 87–112.

[20] Vailaya, A., Jain, A., & Zhang, H. J. (1998). On Image Classification: City Images vs. Landscapes. *Pattern Recognition*, 31(12), 1921–1935. https://doi.org/10.1016/S0031-3203(98)00079-X

[21] Barry, S. J. (2014). Using social media to discover public values, interests, and perceptions about cattle grazing on park lands. *Environmental Management*, 53(2), 454–464. https://doi.org/10.1007/s00267-013-0216-4

[22] Pastur, G. M., Peri, P. L., Lencinas, M. V., García-Llorente, M., & Martín-López, B. (2016). Spatial patterns of cultural ecosystem services provision in Southern Patagonia. *Landscape Ecology*, 31(2), 383–399. https://doi.org/10.1007/s10980-015-0254-9

[23] Rose, G. (2006). *Visual methodologies: An introduction to the interpretation of visual materials* (2nd ed.). Thousand Oaks, CA: Sage Publications.

[24] Alzu'bi, A., Amira, A., & Ramzan, N. (2015). Semantic content-based image retrieval: A comprehensive study. *Journal of Visual Communication and Image Representation*, (32), 20–54. https://doi.org/10.1016/j.jvcir.2015.07.012

[25] Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., ... Ferrari, V. (2020). The Open Images Dataset V4: Unified Image Classification, Object Detection, and Visual Relationship Detection at Scale. *International Journal of Computer Vision*, (128), 1956–1981. https://doi.org/10.1007/s11263-020-01316-z

[26] Chen, S. -H., & Chen, Y. -H. (2017). A Content-Based Image Retrieval Method Based on the Google Cloud Vision API and WordNet. In N. T. Nguyen, S. Tojo, L. M. Nguyen, & B. Trawiński (Eds.), *Intelligent Information and Database Systems* (pp. 651–662). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-54472-4_61

[27] Wartmann, F. M., Tieskens, K. F., van Zanten, B. T., & Verburg, P. H. (2019). Exploring tranquillity experienced in landscapes based on social media. *Applied Geography*, (113), 1–10. https://doi.org/10.1016/j.apgeog.2019.102112

[28] Richards, D. R., & Tunçer, B. (2018). Using image recognition to automate assessment of cultural ecosystem services from social media photographs. *Ecosystem Services*, (31), 318–325. https://doi.org/10.1016/j.ecoser.2017.09.004

[29] Song, X. P., Richards, D. R., & Tan, P. Y. (2020). Using social media user attributes to understand human–environment interactions at urban parks. *Scientific Reports*, (10), 1–11. https://doi.org/10.1038/s41598-020-57864-4

[30] Sottini, V. A., Barbierato, E., Bernetti, I., Capecchi, I., Fabbrizzi, S., & Menghini, S. (2019). Winescape perception and big data analysis: An assessment through social media photographs in the Chianti Classico region. *Wine Economics and Policy*, 8(2), 127–140. https://doi.org/10.1016/j.wep.2019.07.001

[31] Gosal, A. S., Geijzendorffer, I. R., Václavík, T., Poulin, B., & Ziv, G. (2019). Using social media, machine learning and natural language processing to map multiple recreational beneficiaries. *Ecosystem Services*, (38), 100958. https://doi.org/10.1016/j.ecoser.2019.100958

[32] Warnock, S., & Griffiths, G. (2015). Landscape Characterisation: The Living Landscapes Approach in the UK. *Landscape Research*, 40(3), 261–278. https://doi.org/10.1080/01426397.2013.870541

[33] Cherem, G. J. (1972). Looking through the eyes of the public or public images as social indicators of aesthetic opportunity. In P. J. Brown (Ed.), *Proceedings of Aesthetics Opportunity Colloquium* (pp. 52–64). Logan, UT: Utah State University.

[34] Dakin, S. (2003). There's more to landscape than meets the eye: Towards inclusive landscape assessment in resource and environmental management. *The Canadian Geographer / Le Géographe Canadien*, 47(2), 185–200. https://doi.org/10.1111/1541-0064.t01-1-00003

[35] Haywood, K. M. (1990). Visitor-Employed Photography: An Urban Visit Assessment. *Journal of Travel Research*, 29(1), 25–29. https://doi.org/10.1177/004728759002900106

[36] Dorwart, C. E., Moore, R. L., & Leung, Y. -F. (2009). Visitors' Perceptions of a Trail Environment and Effects on Experiences: A Model for Nature-Based Recreation Experiences. *Leisure Sciences*, 32(1), 33–54. https://doi.org/10.1080/01490400903430863

[37] Gou, S., & Shibata, S. (2017). Using visitor-employed photography to study the visitor experience on a pilgrimage route—A case study of the Nakahechi Route on the Kumano Kodo pilgrimage network in Japan. *Journal of Outdoor Recreation and Tourism*, (18), 22–33. https://doi.org/10.1016/j.jort.2017.01.006

[38] *Detect Labels | Cloud Vision API*. (n.d.). Google Cloud. Retrieved April 15, 2020, from https://cloud.google.com/vision/docs/labels

[39] Simensen, T., Halvorsen, R., & Erikstad, L. (2018). Methods for landscape characterisation and mapping: A systematic review. *Land Use Policy*, (75), 557–569. https://doi.org/10.1016/j.landusepol.2018.04.022

[40] van Rijsbergen, C. J. (1979). *Information Retrieval* (2nd ed.). London, England: Butterworth-Heinemann.

[41] Everitt, B. S., & Skrondal, A. (2010). *The Cambridge Dictionary of Statistics* (4th ed.). Cambridge, England: Cambridge University Press.

[42] Bujang, M. A., & Baharum, N. (2017). Guidelines of the minimum sample size requirements for Cohen's Kappa. *Epidemiology Biostatistics and Public Health*, 14(2), 1–10. https://doi.org/10.2427/12267

[43] McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282.

[44] Townscape character assessment. (n.d.). *Landscape Institute*. Retrieved from https://www.landscapeinstitute.org/technical-resource/townscape/

[45] Lynch, K. (1990). *City Sense and City Design*. Cambridge, MA: The MIT Press.