

# Using eye-tracking to evaluate the viewing behavior on tourist landscapes

Gianpaolo Zammarchi, Giulia Contu, Luca Frigau

## 1. Introduction

According to World Travel & Tourism Council (WTTC), tourism's direct and indirect impact accounted for 10.3% of global GDP, and one over ten jobs around the World are tourism-related (WTTC, 2020). In the last years, a sheer number of people started to use Internet as a primary source to search for travel information and choose their travel destination (Garín-Muñoz et al., 2011). In this sense, digital media now exert a relevant influence on tourism management. Several hotels, travel agencies, or other entities (e.g., municipalities, cultural sites, or leisure destinations) use websites, social media accounts, or pages on travel fare aggregators/search engines to attract clients. All these resources make use of a high number of images to transmit the attractiveness of their destinations (Ruhanen et al., 2013). The image can influence travel choice and behavioral intention (Wang & Sparks, 2016). The effectiveness of these tools might be enhanced by exploiting information on user viewing behavior, which can be provided by eye-tracking technology (Scott et al., 2019). Eye-tracking allows measuring the exact position of the eyes during the visualization of images, texts, or other visual stimuli. Consequently, eye-tracking data can be used to compute quantitative measures of viewing behavior that can provide information useful for many applications, such as improving the effectiveness of a website or consumer segmentation.

The first aim of this study is to analyze viewing behavior on images depicting natural and city landscapes. The visual processing of tourism image is investigated in order to evaluate the tourists' perceived destination image and the capacity to impact on the tourist decision making process (Li et al., 2016). The second goal is to compare performances of different widely used supervised and unsupervised models in the classification of these two classes of images.

## 2. Materials

The dataset used in this study comprises 1003 images (779 in landscape mode and 228 in portrait mode), mostly depicting natural indoor or outdoor scenes, obtained from the MIT saliency benchmark repository (freely available online) (Judd, 2009). Data were collected from a group of 15 participants (ages: 18-35). Each participant looked at each image for 3 seconds in free viewing (no specific instruction given to the subjects prior to the experiment) with a 1 second pause (gray screen) between images. Viewers were seated in a dark room two feet apart from the screen (19" and 1280x1024 resolution), and a chin rest was used to stabilize the head (to limit the range of motion). The eye-tracker used for the study was an ETL 400 ISCAN 240Hz model. Data do not contain the first fixation (point observed) of each participant on each image to correct for the central fixation bias (Busswell, 1935; Mannan et al., 1996; Parkhurst & Niebur, 2003; Itti, 2004). The images were collected from two online repositories: Flickr and LabelMe are very different in nature (e.g., people, animals, objects, buildings, mountains, and so on). In this study, we assigned each image to one of three possible classes: (i) natural landscapes, (ii) city landscapes, (iii) other. To assign each image to one of these three classes, we have taken into account the main element of the image. Since

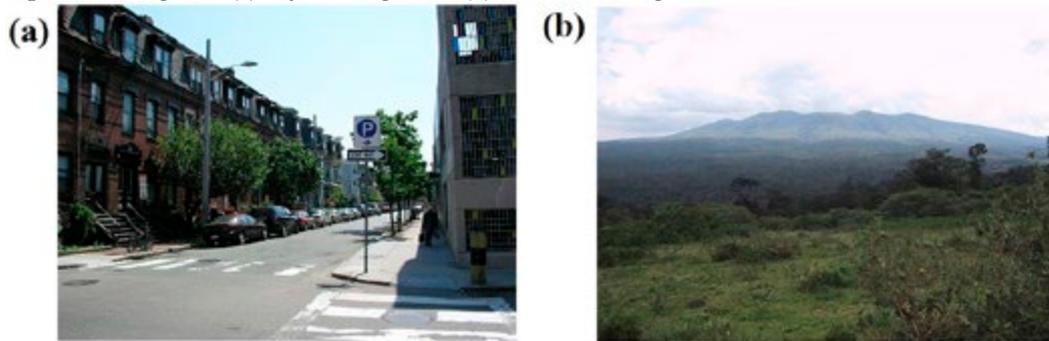
Gianpaolo Zammarchi, University of Cagliari, Italy, gp.zammarchi@unica.it, 0000-0002-9733-380X  
Giulia Contu, University of Cagliari, Italy, giulia.contu@unica.it, 0000-0001-9750-1896  
Luca Frigau, University of Cagliari, Italy, frigau@unica.it, 0000-0002-6316-4040

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Gianpaolo Zammarchi, Giulia Contu, Luca Frigau, *Using eye-tracking to evaluate the viewing behavior on tourist landscapes*, pp. 141-146, © 2021 Author(s), CC BY 4.0 International, DOI 10.36253/978-88-5518-304-8.28, in Bruno Bertaccini, Luigi Fabbris, Alessandra Petrucci, *ASA 2021 Statistics and Information Systems for Policy Evaluation. Book of short papers of the opening conference*, © 2021 Author(s), content CC BY 4.0 International, metadata CC0 1.0 Universal, published by Firenze University Press (www.fupress.com), ISSN 2704-5846 (online), ISBN 978-88-5518-304-8 (PDF), DOI 10.36253/978-88-5518-304-8

our focus was the behavior of people looking at natural or city landscapes, we selected only images where the main element depicted on the scene was a natural landscape or a city landscape. For example, if the image depicts a valley or a desert, it would be classified as “natural landscape”. Conversely, if the whole image was focused on a single flower, even if flowers are typical elements of natural environments, that image would be classified as “other”. At the end of the manual labelling, we removed every image classified as “other” (591 images), and the remaining 412 images (187 classified as “city landscape” and 225 classified as “natural landscape”) were used for subsequent analyses. Figure 1 represents an example of each of the two classes: (a) city landscapes and (b) natural landscapes.

**Figure 1. Examples of (a) city landscapes and (b) natural landscapes**



The landscape is considered as a “factor of attraction and development for tourism” (Jiménez-García et al., 2020). Our hypothesis was that an average user (e.g., a visitor of a touristic website) tends to look at a city landscape shifting from one object to another (e.g., from a car to a building to a road sign), while a natural environment might represent a more homogenous picture with fewer different stimuli to focus on. In accordance, if we measure the path followed by the observer’s eye on a picture, we should expect a longer path in city landscapes than in natural environment pictures.

For each image, we calculated two metrics reflecting the viewing behavior of participants: number of fixations and path length covered by the eye gaze of each participant during observation of each image (computed for each image, using X and Y coordinates of each fixation, as the sum of the Euclidean distances between fixations). The normality of distribution for both variables was assessed using Shapiro-Wilk test. Homogeneity of variance was assessed using Levene’s test. Based on results from these tests, Mann Whitney’s U test and Welch’s t-test were used to compare the number of fixations and the path length between the two classes of images, respectively.

Next, we used a classification approach using the path length and the number of fixations as predictors and the image class as the outcome. We applied supervised and unsupervised methods and compared the results for logistic regression (LR) with a decision rule, linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and K-nearest neighbours (KNN). The four models are trained using 80% ( $n = 330$ ) of the images and tested over the remaining 20% ( $n = 82$ ) using k-fold cross-validation ( $k = 5$ ). We also compared the hard clustering performed using K-Means Clustering algorithm (K-means) with the soft clustering performed using Gaussian Mixture Model clustering method (GMM) to show which one provides better visualization. K-means and GMM are both popular clustering methods which work following an iterative procedure, but the former is non-probabilistic and performs hard assignments, that is, each point can only belong to one class while the latter is a probabilistic algorithm based on multivariate Gaussian distributions as in eq. (1)

$$N(x | \mu, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right\} \quad (1)$$

so that, when the EM (expectation-maximization) algorithm converges, each point is assigned to a class with a certain probability. GMM is more flexible than K-means because it allows decision boundaries to assume an elliptical shape while K-means only a circular shape. All analyses were carried out with R (v. 3.6.3, R Core Team, 2020) using the packages `mclust` (Scrucca et al., 2016), `MASS` (Venables & Ripley, 2002), `class` (Venables & Ripley, 2002), `factoextra`, and `ggplot2` (Wickham, 2009).

### 3. Results

We observed a significant difference in both path length and number of fixations between natural and city images. Namely, we observed shorter path length ( $p < 0.001$ ) and number of fixations ( $p < 0.001$ ) in natural compared to city landscapes (Table 1).

**Table 1. Summary statistics for path length and number of fixations**

	Path length (pixel)		Number of fixations	
	Natural (n=187)	City (n=225)	Natural (n=187)	City (n=225)
Min	4668	8011	70	79
Q1	14267	18522	103	116
Median	17504	21317	112	123
Mean ( $\pm$ SD)	17766 ( $\pm$ 4942)	21431 ( $\pm$ 4322)	111.2 ( $\pm$ 12.84)	123 ( $\pm$ 12.67)
Q3	21287	24298	120	131
Max	31938	32020	148	160

Next, we applied several widely used classification methods to assess if path length and number of fixations could be used to automatically separate pictures of natural and city landscapes. The results of LR, LDA, QDA, and KNN are showed in Table 2.

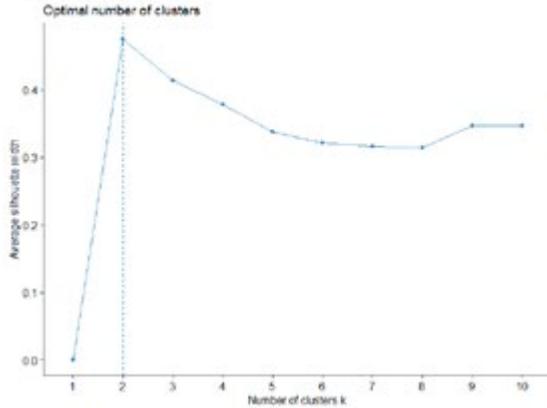
**Table 2. Performance of four models (LR, LDA, QDA, and KNN) in the classification of landscapes**

	LR	LDA	QDA	KNN
Sensitivity	<b>0.743</b>	0.724	0.719	0.662
Specificity	0.608	0.616	0.642	<b>0.662</b>
Accuracy	<b>0.680</b>	0.672	0.680	0.621
F1-score	<b>0.714</b>	0.704	0.707	0.653

Best performances are reported in bold.

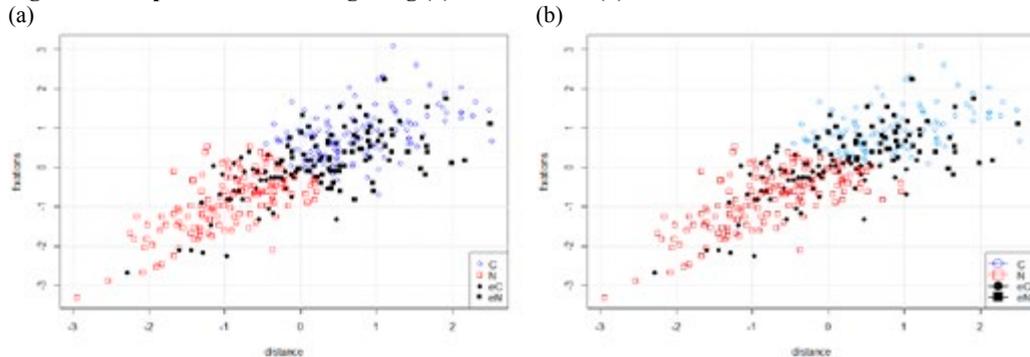
As shown in Table 2, the four classification methods showed very similar results. In particular, sensitivity ranged from slightly above 66% to 74%, and specificity had the lowest values (with best performance achieved by KNN with 66%). This means that most misclassification errors are made when we try to predict the “city landscapes” class. The accuracy ranged from 62% to 68% and that means that, overall, we make many errors when we try to assign images to one of the classes. The results show that the highest accuracy was obtained by logistic regression, which also reached the highest sensitivity and F1-score, so overall can be considered as the best classification method for this task. Finally, we compared the results of two unsupervised classification methods. Since we have two classes of images, we set the number of clusters equal to two. This number was confirmed to be the optimal number of clusters by the plot shown in Figure 2, obtained using the silhouette method.

**Figure 2. Optimum number of clusters based on the silhouette method**



K-means and GMM provided very similar results, as we can see from Figure 3. Both in K-means clustering and GMM plots, the “city landscapes” class is colored in blue and the “natural landscapes” class in red. We used different symbols for correctly classified points (an empty circle for city and an empty square for nature) and misclassified points (a filled circle for city and a filled square for nature). If we compare the two plots from panel (a) and panel (b) we can see that the two methods produce very similar results as regards to misclassification errors.

**Figure 3. Comparison of clustering using (a) K-means and (b) GMM**



Legend: C: city landscapes, N: natural landscapes, eC: fixations erroneously classified as city landscapes, eN: fixations erroneously classified as natural landscapes

## 4. Discussion

In our study we showed that, given a set of images depicting a city or natural environment, it is possible to perform an automatic classification in the two classes using only path distance and number of fixations. To do this we used a subset (412 images) of the MIT dataset (1003 images depicting a large variety of subjects) available online on a public repository, selecting only those images manually labelled as “natural landscapes” or “city landscapes”. We used the path length and the number of fixations in our preliminary statistical analysis showing that both metrics were significantly lower in natural compared to city landscapes. This result is in accordance with our hypothesis that natural landscapes are easier to visually explore, possibly due to a generally lower number of objects of interest and a more homogeneous background compared to city images. This result is in line with Wang & Sparks (2016), who have underlined how nature images are easier to comprehend, and with Dupont

et al (2013) who have discovered that a panoramic photograph may be easier to recognize and memorize.

We also compared four widely used classification methods (LR, LDA, QDA and KNN) in the classification of images in natural and city landscapes. Performances were very similar, but logistic regression proved to be the best method based on the highest sensitivity, accuracy and F1-score and a slightly lower specificity compared to KNN. Our results can be useful for example, for stakeholders involved in tourism management who have to decide whether to insert images depicting “city landscapes” or “natural landscapes” in their web portals. The choice could fall on images of “natural landscapes” as these can be observed with a lower number of fixations (therefore leaving more time for the user to explore a higher number of pictures or other parts of the website), or on images of the city with a reduced number of elements, in order to simplify their perception. In general, the results suggest the necessity to simplify the communication through images which should be clear, simple and with few elements that can attract the viewers’ attention.

## **5. Conclusions**

In the last two decades, tourism promotion is deeply changed and the use of images through websites and travel aggregators for the travel and tourism industry has become crucial to promote travel destinations. Particular attention has been posed on the literature to identify the best images to insert in websites. In this paper, we have investigated the different viewing behavior on images depicting natural and city landscapes. The aim was to evaluate how different classes of images are observed and which images can be easily processed by our brain, thus being potentially more effective in the engagement of viewers. In order to reach this aim, we analyzed eye-tracking data focusing on two metrics: number of fixations and path length. The results showed significant differences in viewing behavior between images picturing natural and city landscapes. The natural images were perceived as easier to visually explore. Moreover, the results have highlighted a relevant utility of the analysis of eye-tracking data to gain insights into the use of images in tourism promotion. The comparison of the performances of different supervised models showed similar performances in the classification of the two classes of images with logistic regression achieving slightly better results. Finally, two commonly used unsupervised methods produced very similar results as regards to misclassification errors when dividing the observations in two clusters. The main limitations of our study include the small number of participants for which viewing behavior data were available as well as the limited number of metrics that we were able to analyze. For instance, as time of observation was fixed to 3 seconds for each image, it was not possible to use this variable as a predictor. Additionally, removal of images not depicting city or natural landscapes resulted in a relatively small dataset (especially when we divided it into training and test set). However, this limitation was partially addressed using a k-fold cross-validation approach, that allows to exploit the entire dataset. Nonetheless, our results should be confirmed in larger and independent datasets. Future developments of this study will involve the analysis of images from different datasets to assess whether other variables (e.g., time of observation) might be helpful to reduce the misclassification errors.

## References

- Buswell, G.T. (1935). *How people look at pictures: A study of the psychology of perception in art*. University of Chicago Press, Chicago, (US).
- Dupont, L., Antrop, M., Van Eetvelde, V. (2013). Eye-tracking analysis in landscape perception research: influence of photograph properties and landscape characteristics. *Landscape Research*, **39**(4), pp. 1-18.
- Garín-Muñoz, T., Amaral, T. (2011). Internet Usage for Travel and Tourism. The Case of Spain. *Tourism Economics*, **17**, pp. 1071-1085.
- Itti, L. (2004). Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Transactions on Image Processing*, **13**, pp. 1304–1318.
- Jiménez-García, M., Ruiz-Chico, J., Peña-Sánchez, A.R. (2020). Landscape and Tourism: Evolution of Research Topics. *Land*, **9**(12), pp. 1-17.
- Judd, T., Ehinger, K., Durand, F., Torralba, A. (2009). Learning to Predict where Humans Look. ICCV 2009.
- Li, Q., Huang, Z., Christianson, K. (2016). Visual attention toward tourism photographs with text: An eyetracking Study. *Tourism Management*, **54**, pp. 243-258.
- Mannan, S.K., Ruddock, K.H., Wooding, D.S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, **10**, pp. 165–188.
- Parkhurst, D.J., Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision*, **16**, pp. 125–154.
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, (AT). URL <https://www.R-project.org/>.
- Ruhanen, L.M., McLennan, C.-L. J., Moyle, B.D. (2013). Strategic issues in the Australian tourism industry: A 10-year analysis of national strategies and plans. *Asia Pacific Journal of Tourism Research*, **18**(3), pp. 220–240.
- Scott, N., Zhang, R., Le, D., Moyle, B. (2019) A review of eye-tracking research in tourism. *Current Issues in Tourism*, **22**(10), pp. 1244-1261.
- Scrucca, L., Fop, M., Murphy, T.B., Raftery, A.E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, **8**(1), pp. 289–317.
- Venables, W.N., Ripley, B.D. (2002). *Modern Applied Statistics with S*, Fourth edition. Springer, New York, (US). ISBN 0-387-95457-0.
- Wang, Y., Sparks, B.A. (2016). An Eye-Tracking Study of Tourism Photo Stimuli: Image Characteristics and Ethnicity. *Journal of Travel Research*, **55**(5), pp. 588-602.
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, New York, (US).
- WTTC, GLOBAL ECONOMIC IMPACT & TRENDS 2020. <https://wttc.org/Research/Economic-Impact/moduleId/1445/itemId/91/controller/DownloadRequest/action/QuickDownload>. Last access: 8 December 2020.