CLASS: Contemplative Landscape Automated Scoring System*

Lukas Navickas[1] and Agnieszka Olszewska[2] and Theofrastos Mantadelis[3]

Abstract— This paper presents an interdisciplinary study joining insights of landscape architecture and computer vision. In this work we used a dataset of contemplative landscape images that was collected and evaluated by experts in landscape architecture. We used the dataset to develop nine k-means clustering and one K-nearest neighbors models that are able to score landscape images based on seven different landscape image features (layers, landform, vegetation, color and light, compatibility, archetypal elements, character of peace and silence) that were identified as contributing to the overall contemplativeness of a landscape. Finally, we chose the combination of models that would produce the best combined contemplativeness score and created CLASS a scoring system that can evaluate the contemplativeness of landscape images with scores similar to those of experts.

I. INTRODUCTION

In the field of landscape architecture and urban planning classifying landscape scenes (settings) according to their visual quality or scenic beauty plays an important role in the land management, and contributes to the continuous endeavor to improve the quality of the living environment. The development of the empirical studies, gave researchers a big palette of methods for evaluating the landscape quality. However, there has been a contest between two opposite research models: (1) expert-based, where the evaluation of the visual quality is performed by an expert; (2) perception-based, where the quality of visual landscape is assessed by the group of independent subjects, or in the other words, the general public. Experts usually work alone and their scores are based on the expert's knowledge, experience and the assumption that an expert is trained to "perceive more" details about the landscape [4]. One big limitation of the expert approach is the low level of precision, reliability and validity. Meaning, that the differences between responses of different raters (experts) about the same landscape are sometimes as big as differences between two different landscape setting [5]. In the perception-based approach, data is collected from multiple raters (general public) and for that reason it

achieves higher levels of reliability as it is shown in [9], [11], [22], [24], [26], [34]. This approach emphasize sensory, emotional, cognitive and personal relationship factors, which is not a case in the expert based approach. However, with the perception-based approach, there is a validity problem [4]. Researchers never know whether the raters assess the "true" factor of interest, or something else, such as personal tastes of groups, recent social trends etc.

Recently, the concept of contemplative green spaces was defined and used to evaluate landscape scenes [19]. The contemplativeness of space until recently was not a technical but rather an artistic term. It was important to make this concept operational because it joins together visual and ecological quality with psychological benefits of the designed green spaces, including: high aesthetics, promoting biodiversity, attention restoration, stress reduction and recovery from mental fatigue that most of city inhabitants encounter on every day basis [12]. Finally, it was important to identify the attributes of contemplative landscapes in order to identify them in our cities and introduce more of them both for public and private purposes.

It is clear that public institutions, professionals as well as the general public would benefit greatly from automated approaches that could combine rating information from both approaches (expert-based, perception-based) of existing landscape evaluations in order to evaluate new landscapes. For example, landscape architects could use an automated evaluation approach in order to evaluate images of existing sites or the visualizations of their proposals. This could result in improving the quality of the new living environments. Also, private or public mental-health retreats could use the tool to boost their therapeutic capacities by spotting or introducing contemplative green spaces within their properties. Finally, addressing our research to the general public would offer a unique tool to evaluate their living space for those who are interested in improving the quality of day to day life.

The rapid development of technology has already established the use of digital imaging for architects and has familiarized them with automated design techniques. In this work we started an interdisciplinary dialogue between arts and computer science in order to build a model that can automatically evaluate landscapes based on their contemplativeness.

For each of the seven landscape features we designed models (nine k-means and one K-nearest neighbors) to compute similarity scores with the experts' annotated landscape images. We use these scores to computed a similar score with that of the experts by computing K-nearest neighbors weighted average of the K (from one (1) to twenty (20))

most similar images. Finally, we compared ten (10) different models and choose the best performing model for each of our features (one for each different landscape feature) and combined them in order to make CLASS: a scoring system that computes the contemplativeness score of landscapes. In order to choose the best model we computed the mean square error for each model over the seven landscape features. Finally, we computed the mean square error of the final contemplativeness scores and presented our results.

The rest of the paper is structured as follows: at Section II we present related work, at Section III we explain what contemplative green space are and how the dataset was collected and annotated, at Section IV we present our methodology, Section V is dedicated to our experiments and finally we conclude our paper at Section VI.

II. RELATED WORK

Automatic image classification is a task that has a growing interest both in academic and industry societies. Proposed solutions to it have existed as long as computer processors themselves have. Since then, researchers are creating and testing various algorithms that help to extract and compare features.

Picard and Szummer [29], performed classification over a dataset of 1343 indoor/outdoor images and propose a color and texture feature (MSAR or DCT-based) combination for best results. They achieved a classification accuracy of 90,2% and proved that their algorithm can easily detect features such as blue sky or green plants in order to facilitate the recognition process.

A further study of image classification was done by Vailaya et al. [33]. They concluded that it is an easier task to classify images into two different classes, instead of extracting features and trying to "teach" the computer to recognize them. They delimit their research to image classification between landscape and city images and used a database of 2716 images. By assuming that their dataset only contains images of landscapes and cities, they successfully achieved an accuracy of 93%. Furthermore, they showed that histograms, coherence vectors and texture features are not suited for city-landscape classification.

An innovative image classification algorithm was presented by Orlow et al. [21]. The researchers created a WND-CHARM (weighted neighbor distances using a compound hierarchy of algorithms representing morphology) algorithm which is intended to use in the field of biology. In order to evaluate their classifier they used three different datasets (HeLa, Pollen and CHO) and compared their results against the following methods: M-Spyd [17], C-SPyd [27] and Gabor Wavelets [7] over the same datasets. They achieved an average accuracy of 90% which out performed all other algorithms that where used. The downside of this approach is its computational complexity. The algorithm requires high-end hardware making to expensive to use in a home environment.

Vailaya et al. [32] grouped vacation images into meaningful categories and trained the computer using clustering techniques. They used a dataset of 6931 vacation images. Images were divided as indoor/outdoor, then were divided into city/landscape and finally a subset of landscape images were divided into sunset/forest/mountain. Their approach achieved an average accuracy of 96.6% while classifying landscape images into sunset/forest/mountain subclass.

III. CONTEMPLATIVE GREEN SPACE DATASET

For this study we use an image dataset consisting of 70 digital images of designed green spaces (parks and gardens) divided in two groups.

- **Group 1:** A group of 40 designed landscape photographs evaluated by ten (10) experts, according to the Contemplative Landscape Questionnaire (CLQ) [20]. This group of images constitutes the primary source of contemplative score information.
- **Group 2:** A group of 30 designed landscape photographs evaluated by one (1) expert, according to the CLQ. This group constitutes the secondary source of contemplative score information.

All images were acquired personally by the researcher (Olszewska) between years 2011 and 2014, in various parks and gardens in Europe and North America. Some of the parks and gardens reviews that can be found in the literature suggested that they were very contemplative, for example the Salk Institute in California [30]. There were also parks which were designed with the designer's intention to be contemplative, such as Porto City Park [23].

A. Contemplative Landscape Questionnaire (CLQ) and Seven Landscape Features

In order to evaluate the contemplativeness of the landscapes in our dataset we used the CLQ and ten experts.

The development of the questionnaire was based on the Delphi experts' technique, which is a commonly used tool to establish a professional, objective judgment about a complex problem with a wide range of scenarios. Evaluation of landscapes according to a set criteria fits well into that concept [10]. The questionnaire consists of seven different landscape features asking for a score from 1 (worst) to 6 (best) for each feature. The mean score of the seven different feature scores represents the contemplativeness score of a landscape. The features used are.

- **Layers:** This feature describes the depth of the contemplative landscape, it is directly connected to the visibility of three planes and the comfort of long distance views. Landscapes with views in several distances have a higher contemplativeness score.
- **Landform:** Landform is based on the smoothness of the ground and the manipulation of the skyline such as openings, closings to the view and elements that attract attention to the skyline. This suggests that the subtle hills, mounds and diversified skylines would be the most desired for contemplation, thus flat or rugged landforms fall weaker in contemplativeness.
- **Vegetation:** Vegetation includes such characteristics as seasonal changes of vegetation form, a high degree of

wilderness and the biodiversity of the vegetation. For example, tree species that turn orange in autumn or trees that drop their leaves have better contemplativeness score compared with trees that remain the same throughout the year. The degree of wilderness describes how natural/native the vegetation appears in the scenery and improves the contemplativeness score similarly. Finally, the more biodiversity a scenery has the contemplativeness score improves with large mono-cultures scoring low. Experts where especially challenged in this part of the questionnaire, because they needed to balance between the three characteristics of the vegetation and choose the most appropriate score (in their opinion).

- **Color and Light:** Color and light, evaluates how light/shade influence the quality of the setting. An important factor for this feature is that broken natural colors are better opposed to vivid contrasting colors. Important is also the visibility of sunlight, or shade movements. For this feature the observer is assumed to not be directly exposed to the sun as shady locations are better for contemplation.
- **Compatibility:** This feature is identical to the Adjacent Scenery feature of the VRM model [31]. In short, it is about the quality of the design in terms of balance and harmony. It includes features such as: spatial order, absence of disturbing elements, working out the openings and closings of views, as well as the physical and visual relations between the elements, and the hierarchical relation between signage or communication system elements. In other words, experts here are asked to evaluate if the design is clear, balanced, well composed and not overwhelming or confusing.
- **Archetypal Elements:** This feature includes the evaluation of how strong the archetypal elements influence the overall perception of the scene. The fact that some archetypal element is present does not guarantee that it will be acknowledged by the viewer, therefore the experts here were asked not only to identify archetypal elements, but also to evaluate how strongly they influence the overall perception of the observer. Archetypal elements to consider included: path, clearing, single old tree, forest, still water (water mirror), waterfall, circle, grave, and boulder.
- **Character of Peace and Silence:** This feature is difficult to evaluate by only looking at the photographs. The experts for this feature had to recognize design strategies that invite rest and relaxation, give a sense of solitude to the observer, or create a visual enclosure and an environment contrasting with the urban chaos. In short, this feature evaluates the tranquility and serenity of the presented setting.

Figure 1 summarizes the contemplativeness scores for each feature and presents the score scale for each feature. Figure 2 presents the four most contemplative and four least contemplative landscape images according to the experts evaluation.

There is no consensus on the optimal number of subjects

Fig. 1. Scoring chart with a 1 to 6-point Likert scale, used by experts to evaluate landscape photographs according to 7 categories.

in the Delphi expert evaluation [10]. If the experts represent different areas then up to fifty respondents is recommended. If the background of the Delphi subjects is consistent, which is the case, ten to fifteen would be an optimal number of experts [6]. We invited thirty experts from around the world to participate in our investigation. We closed the data collection when the number of respondents reached ten. All participant experts (four women and six men) are academics in the area of landscape architecture, with a PhD, 6 to 31 years of professional practice, and have at least one peer-reviewed publication on: landscape design strategies, perception of the landscape, theory of design, or environmental psychology. Their countries of origin are: Portugal (n = 2), Poland (n = 3), United States of America (n = 2), Scotland (n = 1), Germany (n = 1), and Italy (n = 1).

The methodology of scoring the landscapes was evaluated with psychometric measures and revealed the satisfactory level of statistical reliability and validity as shown in Table I.

IV. METHODOLOGY

In this work, we build a regression model that evaluates landscapes based on their contemplativeness. The contemplativeness of the landscape scene is not a technical term, but it have been found that it depends on the seven components that are easier to commonly address [20]. Therefore, we separated our analyses into these seven contemplative features. We built seven different regression models and then combined their results in order to compute the final contemplativeness score. This final contemplative score could be then compared with the final contemplative score obtained from experts, which will verify the validity of our invention.

A. Models using k-means clustering

In this initial approach, we used k-means clustering in order to compute a similarity score of the landscape under evaluation compared with our $40 + 30$ expert annotated landscapes. Furthermore, we used the weighted average of the k-Nearest Neighbors (kNN) in order to combine the similarity scores from the k-means clustering and produce the contemplativeness scores of the landscape. We use as weight the contemplativeness score annotated from the experts and we average with the normalization of the Euclidian distances returned by the k-means clustering.

Fig. 2. Examples of landscape images with the (a) highest scores (> 4.7 points); (b) lowest scores (< 3.2 points) according to CLQ (photos by A. Olszewska and A. Chadala).

Statistical index	Name of the test	Result	Meaning
Reliability	Cronbach's Alpha	Alpha = 0.817	Good
	Gutmanns's split-half	Alpha = 0.854	Very good
	Inter-rater agreement	ICC (10 raters) = 0.81	Almost perfect agreement
Validity	Pearson's correlation coefficient	r = .772 for the mean score	Strong positive correlation

TABLE I
RELIABILITY AND VALIDITY INDEXES OF THE CLQ AND THEIR MEANING.

In order to apply k-means over our images we used Bag of Features (BoF) to extract image features. The BoF approach is usually used for image category classification. It is also called as bag of visual words and it is a technique adapted to computer vision from natural language processing. Since images do not contain discrete words, BoF first requires to construct a vocabulary of words representative of each image category. As experience has shown, every task has a different better approach and for our project we actually had seven different tasks to compute. For that reason we tested BoF with eight (7) different word extraction options: Harris [8]; MinEigen [28]; MSER [15]; SIFT [14], FAST [25]; SURF [1]; and finally, BRISK [13].

- **Harris:** The feature extraction used by Harris, focuses on locating edges, corners and flat areas based on image contrast. For more details about Harris approach look at [8].
- **MinEigen:** This approach uses minimum eigenvalues to detect affine image changes, image texturedness and image dissimilarities. For more details about MinEigen look at [28].
- **MSER:** MSER focuses in detecting exterior regions by intense changes in pixel intensity. For more details about MSER look at [15].
- **SIFT:** SIFT uses cascade filtering in order to detect scale and space invariant keypoints. For more details about SIFT look at [14].
- **FAST:** FAST detects corner points by locating contiguous pixels with the same intensity. For our purposes we used the FAST method in two different settings, using the extracted features directly or by grouping them in histograms like in [3]. For more details about FAST look at [25].
- **SURF:** SURF focuses on locating repeatable image features such as corners, blobs, and T-junctions. We also used a segmented SURF approach as presented at [2]. For more details about SURF look at [1].
- **BRISK:** BRISK focuses on locating Scale-space keypoints. For more details about BRISK look at [13].

After the extraction of the BoF we initialized our k-means cluster with the $40 + 30$ landscapes we have annotated creating clusters.

B. Using K-Nearest Neighbors

We also developed one model with K-nearest neighbors in order to compute the similarity scores. The K-nearest neighbor model we used was inspired from [18] and uses the SIFT features. This model has the benefit of being computationally less demanding than k-means.

C. Combining Scores

Finally, in order to combine the similarity scores produced by our models (either the k-means clustering or the one K-nearest neighbors model) we used a weighted averaged of the K-nearest neighbors (the K best similarity scores in our case) as shown in the following equation:

$$Feat\_Score(I, K) = \sum_{i=1}^{K} \left(\frac{S_i(I)}{\sum_{j=1}^{K} S_j(I)} * Feat\_Score(I_i) \right), \quad (1)$$

where $Feat\_Score(I_i)$ is the experts' score for the annotated image used as the cluster node i; $S_i(I)$ and $S_j(I)$ are the produced similarity scores of the input image I against the cluster nodes i, j respectively.

V. EXPERIMENTS

In order to compare the candidate models, we chose ten random images out of Group 1 to be our test dataset and used the rest 60 images as our training dataset. For our implementation we used Matlab 2015b [16]. Figure 3,
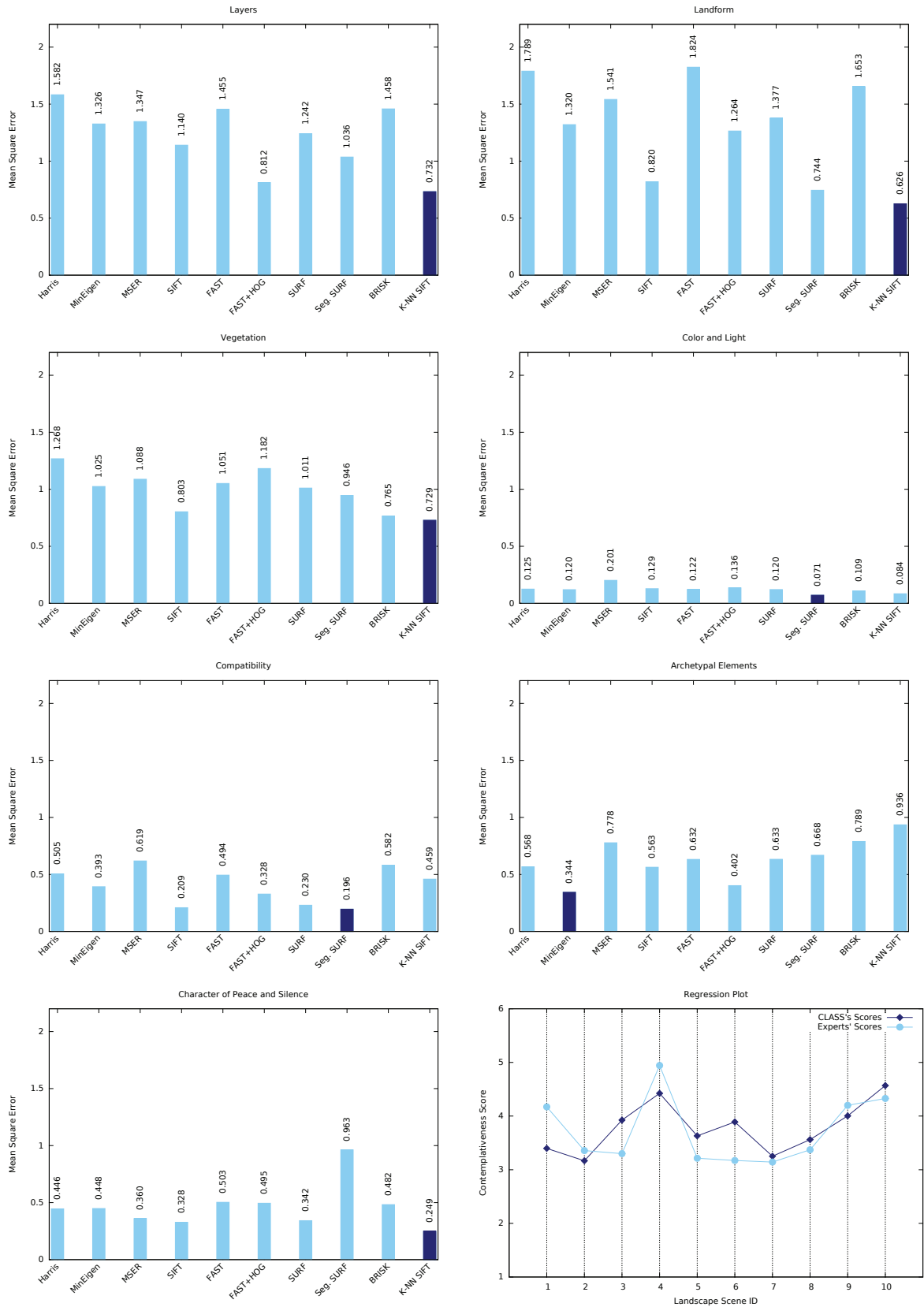
Fig. 3. Comparison of the MSE of each model, with dark blue we annotated the best model. Finally, the regression plot of the CLASS.

presents our results. First we compare the best MSE for each landscape feature in order to choose the best model. As it is shown for the landscape features: layers, landform, vegetation and character of peace and silence the best model was K-nearest neighbor using SIFT feature extraction; for the landscape features: color and light and compatibility the best model was k-means using Segmented SURF for feature extraction and finally, for archetypal elements the best model was k-means using MinEigen feature extraction. Furthermore, for combining the similarity scores we found that the best Ks are $9, 9, 2, 16, 2, 2, 15$ respectively for each feature. Finally, in Figure 3 (Regression Plot) we present the contemplativeness scores of CLASS for the test set compared with the contemplativeness scores given by experts. The MSE of CLASS over the test set is 0.213.

VI. Conclusion & Future Work

This paper presented the first (to the best of our knowledge) model that can automatically evaluate the contemplativeness score of landscapes. We trained ten different models and compared them over seven different landscape features. We used the best model (by MSE) for each feature and combined them in order to compute the contemplativeness score of landscapes. As the regression plot of Figure 3 illustrates, CLASS generates fairly comparable results with experts.

The created framework can be used by public institutions, professionals as well as the general public for various evidence-based design approaches, helping them to better understand our living environments. It is easy to use and provides an inexpensive, less time consuming approach than hiring qualified experts, while still providing reliable information about potential mental-health and well-being benefits of investigated landscapes.

As this is the very first model ever developed to evaluate the contemplativeness score we are sure that we can further improve the generated scores by using different regression models such as models based on Artificial Neural Networks and further investigating different feature extraction methods.

References

[1] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-Up Robust Features (SURF)," *Journal of Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346 – 359, 2008.

[2] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Workshop on Statistical Learning in Computer Vision, ECCV*, 2004, pp. 1–22.

[3] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conference*, vol. 1, June 2005, pp. 886–893 vol. 1.

[4] T. C. Daniel, "Whither scenic beauty? Visual landscape quality assessment in the 21st century," *Journal of Landscape and Urban Planning*, vol. 54, no. 1-4, pp. 267–281, may 2001.

[5] T. C. Daniel and J. Vining, "Methodological Issues in the Assessment of Landscape Quality," in *Behavior and the Natural Environment*. Springer Science + Business Media, 1983, pp. 39–84.

[6] A. Delbecq, A. Van de Ven, and D. Gustafson, *Group techniques for program planning: a guide to nominal group and Delphi processes*, ser. Management applications series. Scott, Foresman, 1975.

[7] J. Han and K.-K. Ma, "Rotation-invariant and scale-invariant gabor features for texture image retrieval," *Journal of Image and Vision Computing*, vol. 25, no. 9, pp. 1474 – 1481, 2007.

[8] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," in *Alvey Vision*. British Machine Vision Association and Society for Pattern Recognition, 1988.

[9] J. Hetherington, T. C. Daniel, and T. C. Brown, "Is motion more important than it sounds?: The medium of presentation in environment perception research," *Journal of Environmental Psychology*, vol. 13, no. 4, pp. 283–291, dec 1993.

[10] C.-C. Hsu and B. A. Sandford, "The Delphi technique: making sense of consensus," *Practical assessment, research & evaluation*, vol. 12, no. 10, pp. 1–8, 2007.

[11] R. Hull and W. Stewart, "Validity of photo-based scenic beauty judgments," *Journal of Environmental Psychology*, vol. 12, no. 2, pp. 101–114, jun 1992.

[12] R. Kaplan and S. Kaplan, *The experience of nature: A psychological perspective*. CUP Archive, 1989.

[13] S. Leutenegger, M. Chli, and R. Siegwart, "BRISK: Binary Robust invariant scalable keypoints," in *Computer Vision (ICCV), IEEE Conference*, Nov 2011, pp. 2548–2555.

[14] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[15] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Journal of Image and Vision Computing*, vol. 22, no. 10, pp. 761 – 767, 2004.

[16] MATLAB, *version 7.10.0 (R2010a)*. Natick, Massachusetts: The MathWorks Inc., 2010.

[17] J. Montoya-Zegarra, N. Leite, and R. da S.Torres, "Rotation-invariant and scale-invariant steerable pyramid decomposition for texture image retrieval," in *Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI)*, Oct 2007, pp. 121–128.

[18] M.-E. Nilsback and A. Zisserman, "A Visual Vocabulary for Flower Classification," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conference*, vol. 2, 2006, pp. 1447–1454.

[19] A. A. Olszewska, P. F. Marques, and F. Barbosa, "Enhancing Urban Landscape with Neuroscience Tools: Lessons from the Human Brain," *CITYGREEN*, vol. 01, no. 11, p. 60, 2015.

[20] A. A. Olszewska, P. F. Marques, and F. B. F. Barbosa, "Urban Planning, Neurosciences and Contemplation for improving well-being in our cities," in *ARCC Conference Repository*, 2014.

[21] N. Orlov, L. Shamir, T. Macura, J. Johnston, D. M. Eckley, and I. G. Goldberg, "WND-CHARM: Multi-purpose image classification using compound image transforms," *Journal of Pattern recognition letters*, vol. 29, no. 11, pp. 1684–1693, 2008.

[22] J. F. Palmer, "Stability of landscape perceptions in the face of landscape change," *Journal of Landscape and Urban Planning*, vol. 37, no. 1-2, pp. 109–113, jun 1997.

[23] S. C. Pardal, *Parque da Cidade Porto*, 2006.

[24] R. G. Ribe, "Scenic Beauty Perceptions Along the ROS," *Journal of Environmental Management*, vol. 42, no. 3, pp. 199–221, nov 1994.

[25] E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," in *International Conference on Computer Vision (ICCV), IEEE Conference*. IEEE, 2005.

[26] H. Schroeder, "Estimating park tree densities to maximize landscape esthetics," *Journal of Environmental Management*, vol. 23, no. 4, pp. 325–333, 1986.

[27] J. Sevie and M. Steer, "Analysis of GaAs MESFET spectrum regeneration driven by a $\pi/4$-DQPSK modulated source," in *Microwave Symposium Digest, IEEE conference*, May 1995, pp. 1375–1378 vol.3.

[28] J. Shi and Tomasi, "Good features to track," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conference*. IEEE, 1994.

[29] M. Szummer and R. W. Picard, "Indoor-outdoor image classification," in *Content-Based Access of Image and Video Database, IEEE Workshop*. IEEE, 1998, pp. 42–51.

[30] M. Treib, "Attending," in *Contemporary Landscapes of Contemplation*, R. Krinke, Ed. Informa UK Limited, 2005, ch. 1, pp. 13–35.

[31] USDA Forest Service, "National forest landscape management," 1973.

[32] A. Vailaya, M. Figueiredo, A. Jain, and H. Zhang, "Content-based hierarchical classification of vacation images," in *Multimedia Computing and Systems, IEEE Conference*, vol. 1. IEEE, 1999, pp. 518–523.

[33] A. Vailaya, A. Jain, and H. J. Zhang, "On image classification: city vs. landscape," in *Content-Based Access of Image and Video Libraries, IEEE Workshop*. IEEE, 1998, pp. 3–8.

[34] E. H. Zube, D. G. Pitt, and T. W. Anderson, "Perception and measurement of scenic resources in the Southern Connecticut River Valley publication no r-74–1," *Journal of Landscape Research*, vol. 1, no. 8, pp. 10–11, jun 1974.