Inferring Affective Experience from the Big Picture Metaphor: A Two-dimensional Visual Breadth Model

Song Tong [1,3*], Jingyi Duan [1*], Xuefeng Liang [2†], Takatsune Kumada [3],
Kaiping Peng [1†], Ryoichi Nakashima [3]
1. Department of Psychology, Tsinghua University, Beijing, China
2. School of Artificial Intelligence, Xidian University, Shaanxi, China
3. Graduate School of Informatics, Kyoto University, Kyoto, Japan
* equal technical contribution
† xliang@xidian.edu.cn; pengkp@tsinghua.edu.cn

Abstract

This study explores the psychological significance of the commonly used visual metaphor 'seeing the big picture' and examines whether and how it leads to positive experiences in real-life situations. To elucidate this phenomenon, a two-dimensional model of visual breadth is proposed, then respectively operationalized by two computer vision approaches. Our approaches are evaluated on a collected data set with 29,216 photos. The results revealed that physical and contextual breadth are two essential visual structures that compose a 'big picture'. Furthermore, these two visual breadths interactively shape people's affective experiences. This study provides insight into the psychological implications of the 'big picture' metaphor and sheds light on its practical potential for computer vision approaches and affective computing in the wild.

1. Introduction

Visual perception has a significant impact on cognition and emotion [35], especially in the current era of heightened visual media. Across cultures, the visual metaphor of 'seeing the big picture' is frequently used to uplift people or alleviate their distress, e.g., 'Kan Kai Dian' in Chinese. Numerous psychological studies have demonstrated a bidirectional causal relationship between positive affection and widen attentional scopes [5, 14, 28]. However, it remains uncertain whether and how the whole structure of 'seeing the big picture' plays a role in real-life affective experience.

Limited laboratory settings, small sample size and constrained data estimation capacity of conventional psychological experiments once constrained the academic emphasis to emotional phenomena in brief or transient periods [23, 29]. As a result, the holistic, macroscopic affective experience was mainly left to be a theoretical construct without much empirical exploration, making it challenging to incorporate it into tangible estimation methods.

Big data and machine learning pose a way to breakthrough [29]. While studying behavior in the wild brings in its essence too many confounders and therefore rends it too difficult to isolate the specific phenomena of interest [23], it suits well these new computational paradigms.

Firstly, the internet provides a natural repository of behavioral records from diverse scenarios with high ecological validity [22], neutralizing distractors and confounding variables with sufficiently large and diverse samples [9]. Secondly, with the idea of crowdsensing, the analytical scope can be lifted beyond individual tracking. Crowdsensing utilizes group experience by integrating sensory data of similar kinds uploaded by a large number of ordinary users dispersed in a large area [7]. In this way, studies become less dependent on individual information, and theoretically, shared experience is focused on a macroscopic view. Finally, these methods enable the understanding of cognition and emotion from the perspective of the environment. As the affordance theory suggested, the agency and the environment constitute an impartible interactive system for the constructions and perceptions of the world [8]. In regard to affective experience, incorporating such a perspective enables a more comprehensive understanding of the macroscopic affective experience of people in the context as a whole.

Thus, the present study endeavors to employ computer vision methods to investigate the appeal of the 'big picture' through photo-analysis from internet visual data and explore its potential impact on general affective and cognitive experiences in the context of tourism. Specifically, visitor-generated photographs obtained from the internet were detected as spontaneous visual products, and a 2-dimensional

visual breadth model was proposed to delineate the visual experience. Moreover, this study shifts the focus from an individual-centric perspective to a holistic approach that comprehends the macro-emotional experience of individuals in the context as a whole, with the aim of estimating the potential impact of visual experience on general affective and cognitive experiences.

Figure 1 illustrates the theoretical framework of the two-dimensional visual breadth model, which provides a foundation for comprehending affective experience from the interaction between environment and visual experience. The introduction will be arranged in two aspects. Firstly, the theoretical construction of the two-dimensional visual breadth model theory will be introduced in Section 1.1-1.2. Section 1.1 outlines the broaden-and-build theory and the theory of affordance, which lays a foundation for a trilateral interactive network between environment, perception (and cognition), and affection. Section 1.2 proposed a two-dimensional conceptualization of the 'big picture' in the interactive visual structure, which encompasses the physical breadth and the contextual breadth. Secondly, the verification of our proposed model using destination photos will be introduced in Section 1.3-1.4. Section 1.3 introduces photography in tourism research, which highlighted its significant duality in reflecting visual experience. Section 1.4 briefly sums up the idea to validate the proposed two-dimensional model using computer vision methods with an internet visual data-set and presents some of the important results.

1.1. The Impact of Big Picture on Affective Experience

The 'big picture' metaphor encourages individuals to adopt a broader perspective, pay attention to the larger context and appreciate the interconnections of things [10]. This notion resonates with the broaden-and-build theory of positive emotions [5], which posits that attention and emotion are bi-directionally causally related. Specifically, positive emotion facilitates exploratory cognition and behaviors, while negative cognition restricts people's resources on limited goals [6]. For example, induced positive emotion was found to significantly widened participants' attention from the central target to the peripheral stimuli [38] and the irrelevant surroundings [4, 24, 25]. That is, positive emotion broadens the area of perceptual and attentional processing.

On the other hand, this process also holds in reverse. Broadening attention can enhance positive emotions and diminish negative emotions while narrowing attention can decrease positive emotions and amplify negative emotions [10, 12]. For instance, Gu *et al.* [10] demonstrated significantly reduced negative emotions and lifted positive emotions in depressed participants after 8 weeks of training to watch distant scenes.

Beyond the relationship between emotion and thought-action repertoire, the role of the environment was less discussed. In fact, this relationship cannot be clarified outside the context as it is the background and the essential premise where it takes place. As the embodied theory suggests, there is an integral relationship between the mind and the body experience. Thus, the environment and individual's perception of it also shapes higher cognition and consciousness [15, 36].

Furthermore, the affordance theory posits that the environment provides organisms with action possibilities where their concepts and behaviors stem from and interact with [8]. Empirical evidence supports these theories and found that the cognitive framework can be shaped by the visual context, e.g., a complex and ambiguous street scene promotes a holistic perception for people from both eastern and western cultures [18].

By considering cognitive-emotional and environmental-cognitive interrelationships, a trilateral environment-cognitive-affection network of experience shows up. This framework offers a potential model to understand the environmental influences on affective experience. In particular, a 'big picture' cannot be understood solely as the characteristic of the environment, nor is it a pure abstract feeling of the individual. Instead, it is an interactive visual structure that lies between the agent and the environment.

1.2. What is Big: a Two-dimensional Visual Breadth Model

After clarifying the visual interactive structure of the 'big picture', another issue is the definition of 'big'. Despite its common use, the 'big' in the phrase 'big picture' is actually a crucial premise that has not been extensively explored. The definition of 'big' varies in previous studies, with some studies defining it as 'zooming out' the picture [10], others as paying more attention to peripheral stimuli [38], or a global processing tendency facing a global-local stimulus [26].

This study defines 'big' in a proposed two-dimensional visual breadth model. The two-dimensional idea integrates the understanding of the 'big picture' from previous studies, the connotation of the 'big picture' in everyday language, and the theoretical approach of analytical-holistic analysis.

'Big' was supposed to present wide visual breadth, which was supposed to consist of physical breadth and contextual breadth. The physical breadth refers to the degree of how vast and wide a scene is. It goes back to the literal meaning of 'big picture', which refers to a widened sight, or an enlarged angle of vision. Contextual breadth refers to the diversity of visual elements in an area. It responds to the metaphoric meaning of the 'big picture', which highlighted the integration of multiple aspects [17].

To put it in the academic framework, the physical

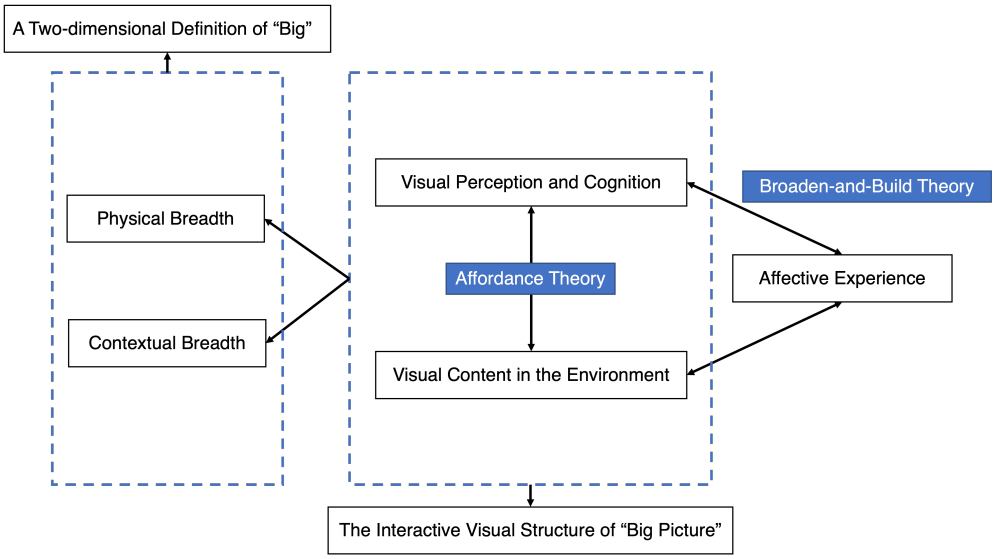

Figure 1. The two-dimensional visual breadth model of ‘Big Picture’ metaphor for understanding affective experience. The two cues of visual breadth are the interactive visual structure that lies between the human perception and the environment, which is hypothesized to impact the affective experience according to the broaden-and-build and affordance theories.

breadth is in line with the most definition in previous studies, while contextual breadth fills the gap that was formerly neglected. The two-dimensional model of visual breadth provides an operationalizable theoretical framework for analyzing the visual ‘big’ and makes it possible to bring the ‘big picture’ into concrete affective experience estimations.

1.3. Photo as a bridge

Photos have a duality that maps the interactive structure of the ‘big picture’, which posits its priority in studying this phenomenon. According to Bandura [2, 20], photographs reflect the interplay between the environment, cognition, and behavior. By selectively capturing and attending to meaningful visual content, photography can convey both the objective reality of the environment and the subjective interpretation of the photographer [34]. From a personal perspective, tourist-employed photographs can provide insight into a tourist’s perception of a destination, including their attention and preferences [20]. From a spot-focused perspective, photographs reflect the visual offerings of tourist attractions.

Among different types of photos, tourist-employed destination photos are of great value in examining the interaction between individuals and their environment [20, 39, 40]. On one hand, the tourist-employed photos are usual containers of important visual experience in real life [1]. On the other hand, there are many diverse travel photos available publicly, as photography serves as a ubiquitous recording method for tourists of varied ages and cultural backgrounds [40].

Therefore, in this paper, visitor-employed destination photos were chosen to validate the proposed two-dimensional visual breadth model.

1.4. The Present Study

According to the broaden-and-build theory and the theory of embodied cognition, tourist attractions with high visual breadth should offer visitors a more positive experience and receive higher overall experiences, i.e., the general affective experience. Thus, the purpose of this study is to investigate whether visual breadth can be used to infer the overall experience rating of tourist attractions.

To achieve this, we utilized visitor-employed photos from the Internet to test our two-dimensional visual breadth model, which combines physical breadth and contextual breadth measures. Concretely, we collected 29,216 photos and employed computer vision methods to calculate the physical and contextual breadth, traditional statistical methods to assess these measures for each point of interest. The experimental results showed that physical and contextual breadth significantly and interactively influence the overall rating of a tourist attraction. Therefore, we conclude that our proposed two-dimensional visual breadth model can provide a meaningful framework for analyzing visual breadth and pave a pathway for future researchers to integrate person-environment information in affective assessments. The findings of this study have practical implications for wild affective computing, e.g., by utilizing computer vision methods, as well as tourist attractions can enhance the general affective experience of visitors by considering the physical and contextual breadth of the attractions.

Figure 2. The geo-locations and representative photographs of 20 tourist attractions.

Table 1. The details about the photo number of 20 scenic attractions. R and N_p represent the overall experience rating and photo number of each tourist attraction, respectively.

No. (a)	Spot name	Country	R^a	N_p^a
1	Manneken Pis	Belgium	3.260	998
2	The Little Mermaid	Denmark	3.418	1,997
3	Jungle Island	USA	3.769	982
4	Marina Beach	India	3.774	640
5	Khaosan Road	Thailand	3.810	1,006
6	Las Ramblas	Spain	3.901	990
7	Ho Chi Minh Mausoleum	Vietnam	3.914	1,015
8	Television Tower	Germany	4.003	884
9	Tiananmen Square	China	4.027	997
10	National Monument	Indonesia	4.113	1,033
11	Kiyomizu dera Temple	Japan	4.407	988
12	Tower Bridge	England	4.587	1,998
13	Chichen Itza	Mexico	4.589	1,753
14	Eiffel Tower	France	4.594	1,982
15	Colosseum	Italy	4.674	1,970
16	Sydney Opera House	Australia	4.674	1,994
17	Parthenon	Greece	4.687	1,998
18	Golden Gate Bridge	USA	4.697	1,993
19	Taj Mahal	India	4.790	2,000
20	Bryce Canyon Park	USA	4.908	1,998

2. Materials and Methods

2.1. Datasets

Selection of Tourist Attractions. We selected 20 of the most popular tourist attractions worldwide for this study based on rigorous criteria that aimed to ensure diversity, objectivity, and independence. Specifically, the criteria were: (1) Popularity: the most frequently recommended tourist attractions recommended by top search engines - TripAdvisor, National Geographic, and Travel + Leisure; (2) Objectivity: having at least 1.5K votes for each site regardless of language, age, gender, nationality, *etc*; (3) Generality: located across in Asia, Europe, and Americas to ensure geographic diversity; (4) Diversity: cover a broad range of site types while avoiding religious places; (5) Independence: sufficiently distant from other sites to prevent cross-rating. Figure 2 illustrates the geographical location of these 20 tourist attractions included in our study [29].

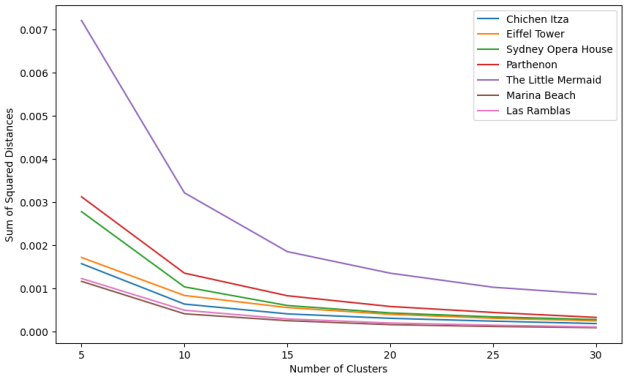

Figure 3. The Elbow Method for determining the optimal number of clusters for k-Means.

Collection of Photos. To obtain a representative sample of photos for each tourist attraction, we collected ratings and photos from two sources: TripAdvisor and Flickr. For photos, we used geotags to ensure that each photo corresponded to the correct tourist attraction [37]. We collected a total of 245,000 photos from the specified zones of attractions, but to address the issue of unbalanced photo distribution, we randomly selected a subset of 29,216 photos that corresponded to the 20 tourist attractions to include in our study as shown in Table 1.

To provide a more comprehensive analysis, we employed the k-means clustering algorithm to segment each tourist attraction into several distinct spots based on its geographical distribution. The elbow method was used to determine the appropriate number of clusters for the k-means algorithm. As shown in Figure 3, we randomly selected seven tourist attractions and calculated the sum of squared distances (SSD) for each of the different k values. The plot demonstrates that the optimal value for k is 10. A visual representation of this segmentation is presented in Figure 4, which illustrates the distribution of photographs taken at a

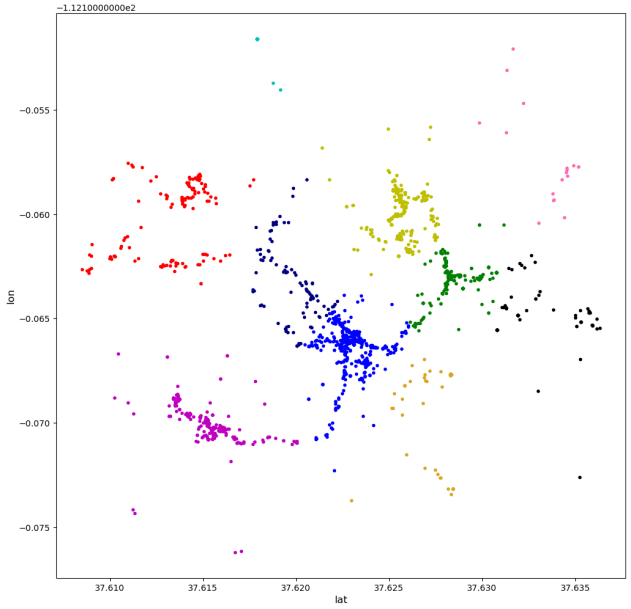

Figure 4. The distribution of photographs of the Bryce Canyon Park.

specific tourist attraction. Each dot in the figure represents a photograph, with its location mapped from its geographical coordinates. By segmenting each tourist attraction into 10 distinct spots, we were able to capture the spatial distribution of tourist activity and identify the areas of highest photographic activity within each attraction. While there may be some discrepancy between the physical boundaries of the tourist attractions and the spots identified by our algorithm, such a boundary is still in line with personal experience, as a tourist attraction is perceived as a spatial entity with significant tourist footfall and photographic activity [27, 40, 41]. Ultimately, our analysis resulted in the collection of 200 spots from renowned tourist attractions worldwide.

2.2. Physical Breadth Evaluation

In this study, we operationalize the concept of physical breadth as the classification of photos' view type as wide or narrow, e.g., wide views are panoramic images that present a vast and wide scene. Our argument draws from the theory of affordance, which suggests that perception interacts with the environment. Visitor-employed photographs, as a record of people's perception of the environment, can represent this interactive system with its duality [18, 19].

It is important to note that view type, which refers to wide and narrow views is distinct from other aspects of view in the field of computer vision, such as angle of view (e.g., wide and narrow angle) or depth of view (e.g., close-up and distant photos) [3, 33]. While the depth of view focuses on "the absolute distance between the photographer and a scene" [33], and angle of view is defined as the way the camera lens is used [1], view-type represents photographers' attentional scopes and is the feeling of the interaction between depth and angle of view. Thus, view-type classification is a relatively new and challenging problem in the field of computer vision since the photographers' attentional scope is *a subjective property of photography*. Unlike objective tasks such as face recognition and object detection, it is difficult for algorithms to identify specific patterns for this subjective task due to the wide variety of contexts present in natural images [32]. To address this challenge, we propose a cumulative feature convolutional neural network (CFCNN) [29] that mimics the breadth of attention in humans. Specifically, the CFCNN is trained to dichotomously classify a given photo as either a narrow or wide view.

To train our CFCNN, we first collected 46,906 travel photos of tourist attractions. We then used the visual attention-inspired model proposed in [32] to automatically detect which photos were wide view and narrow view. To ensure accuracy, two independent individuals (a male aged 24 and a female aged 28) also classified the photos. For training, we used 43,906 photos, with 25,776 classified as wide view and 18,130 as narrow view. The remaining 3000 photos were used for testing, and five participants (3 male and 2 female, mean age = 23.2 ± 1.2) were asked to dichotomously classify each photo as wide or narrow view. Based on the consistency of decisions among the participants, each photo was assigned a tag of the view type, ranging from obvious wide view (i.e., all five participants rated the picture as wide) to the obvious narrow view, with other tags in between including normal wide view, ambiguous wide view, normal narrow view, and ambiguous narrow view.

Our CFCNN achieved testing results of 93.67% [30], outperforming existing methods such as reference [32] at 89.59% and reference [31] at 79.85% for obvious sets. Moreover, there was a significant correlation between human and CFCNN view-type cognition, even considering different levels of ambiguity. Therefore, the CFCNN was used to further estimate whether a photo in the attraction set contains a narrow view or a wide view, corresponding to the two situations of the physical breadth [30].

2.3. Contextual Breadth Evaluation

The contextual breadth of a scene refers to the extent of visual heterogeneity within a given area. In accordance with the interactive visual structure, visual heterogeneity describes more than a pure objective characteristic of the environment. Rather, it emphasizes whether what is visually attended to and perceived is heterogeneous. A scene with broad contextual breadth is characterized by a diverse range of visual elements that span across multiple cate-

[1] https://inst.eecs.berkeley.edu//~ee198-4/fa07/week11_assignment.shtml

gories being attended, while in a scene with narrow contextual breadth, only a high concentration of similar visual elements are attended. In a contextually wide scenic spot, individuals are afforded the opportunity to appreciate and photograph visual elements of different kinds (e.g. food, plants, buildings, *etc*.). Conversely, in a contextually narrow scene, the visual stimuli are more homogeneous, potentially limiting the diversity of visual experiences available to tourists.

Two steps were taken to computationally define contextual breadth in this study.

Firstly, we recorded attended visual domains in each scenic spot. For this purpose, categories of visual elements in each photo were defined using the SUN attribute model, a computational tool for visual understanding [21, 42]. The SUN attribute database encompasses 14,340 images from 102 domains of visual elements, including human interaction (e.g., reading, socializing), natural composites (e.g., rocks, ocean), abstract concepts (e.g., warm, open area), *etc*. To train the scene attribute model, a deep convolutional network pre-trained model was utilized to extract a 4096-dimensional feature vector from the images in the SUN attribute database, which was then employed to train a support vector machine (SVM) classifier for each of the 102 scene attributes [16, 42].

As a result, all 29,216 pictures in our dataset were identified using the SUN attribute model. The raw model outcomes indicate the likelihood of each attribute being recognized in the photos. To ensure precision, two researchers(a 32-year-old male and a 23-year-old female) meticulously reviewed the outcomes of the sun attributes of these randomly selected 300 photos and determined to take the top five attributes of each photo for the final analysis. The photo-based sun-attribute results then converged into spots. Specifically, after the visual domains were recognized for each photo, a spot-based cumulative sum was done. Thus, we attained the distribution of the 102 visual categories for each of the 200 scenic spots.

Secondly, we calculated the concentration of attended visual domains. Attribute kurtosis was utilized to measure this concentration. The attribute kurtosis (ak) for a given location is computed using the formula:

$$ak_i = (1/N_a) \times \sum_{j=1}^{N_a} ((n_i^j - \mu_i)/\sigma_i)^4 - 3, \qquad (1)$$

where n_i^j represents the frequency of the j-th attribute for the i-th location, $N^a = 102$ is the number of attributes, ak_i represents the attribute kurtosis of the i-th location, and μ_i, σ_i are the mean and standard deviation of the frequency of attributes for the i-th location. A higher attribute kurtosis value indicates a narrower contextual breadth, whereas a lower kurtosis value indicates a wider contextual breadth. Specifically, a peak distribution of attributes indicates that the visual domains being photographed are highly concentrated, suggesting a contextually narrow environment where tourists tend to attend to limited attributes. Conversely, a flat distribution indicates attention to diverse categories of visual attributes, indicating a contextually broad environment.

3. Results and Analysis

In this study, we utilized logistic regression analysis to investigate the influence of physical breadth (i.e., view type) and contextual breadth (i.e., attribute kurtosis) on the overall tourism experience rating. To ensure comparability between the variables, we normalized both view type and attribute kurtosis using z-scores. Descriptive statistics revealed that attribute kurtosis followed a normal distribution (Shapiro-Wilk $W > .99$, $p = .77$), and was treated as a continuous variable in subsequent analysis. In contrast, the view type exhibited a bimodal distribution, and we transformed it into a dichotomous variable using 0 as a boundary as followed,

$$vt_i = \begin{cases} 1, & \text{if } f_{vt_i} > 0 \\ 0, & \text{if } f_{vt_i} \leq 0 \end{cases}, \qquad (2)$$

where f_{vt_i} represents the original continuous variable that had a bimodal distribution, and vt_i represents the transformed dichotomous variable for the i-th location, where 1 represents a wide view and 0 represents a narrow view.

Using a logistic regression model with view type and attribute kurtosis as the independent variables, and rating (dichotomous, high, or low) as the dependent variable, we found no significant problem of collinearity (tolerance $> .1$). The model was statistically significant, X^2 (2) = 52.735, $p < .001$, and explained 32.3% (Nagelkerke R^2) of the total variance in experience rating. The logistic regression analysis revealed a significant positive main effect of view type on rating (Odds Ratio, OR = 2.56, 95% Confidence Interval, CI [4.569, 18.650]) as shown in Table 2. Lower attribute kurtosis was shown to associate with a higher probability of high experience rating (OR = 0.526, 95% CI [0.013, 0.888]), indicating that contextual breadth also plays a role in shaping the overall tourism experience.

Furthermore, a significant interaction was observed between view type and attribute kurtosis on rating (OR = 2.669, 95% CI [1.342, 5.311]). It suggested that physical breadth and contextual breadth are closely intertwined and work together to shape the overall tourism experience. Table 3 contains further information. Simple effect analysis was conducted to further investigate interaction effects. The results showed that attribute kurtosis significantly predicted experience rating when the spot contained a narrow view (OR = 0.503, 95% CI [0.287, 0.881], p = .012), but not

Table 2. The effect of View Type (vt) and Attribute Kurtosis (ak) on Destination Rating. Odds Ratios (OR) represents the odds of an event occurring in one group compared to the odds of the same event occurring in another group, while the confidence interval (CI) provides a range of values that is likely to include the true value of the population parameter.

			95 % CI (OR)	
	OR	p	Lower	Upper
ak	0.526	0.016	0.312	0.888
vt	9.231	$< .001$	4.569	18.650
$ak \times vt$	2.669	0.005	1.342	5.311

Table 3. Simple effects of ak.

Moderator Levels			95 % CI (OR)	
vt	OR	p	Lower	Upper
Narrow	0.526	0.016	0.312	0.888
Wide	1.405	0.135	0.899	2.196

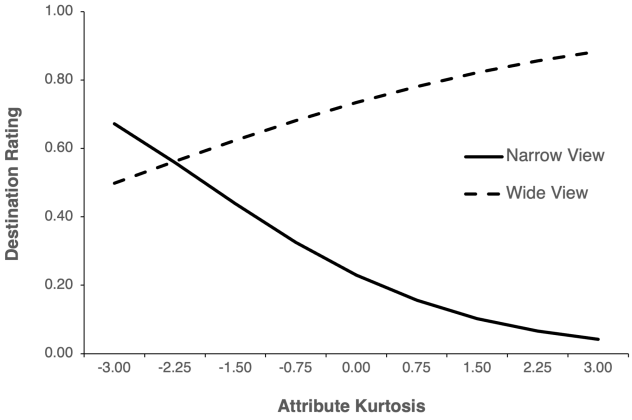

Figure 5. The interaction between vt and ak.

a wide view (OR = 1.37, 95% CI [0.905, 0.090], p = .133). Figure 5 shows the interaction effect. The finding suggested that the contextual breadth only exhibits an impact on the overall experience when physical breadth is low, indicating a complementary influence.

In addition, we also performed an analysis with a traditional approach predicting tourism exprience using visual categories as a baseline model [16, 40]. We selected five types of elements (traffic, social, building, water, green) calculated via SUN attributes as independent variables, categorized based on previous research [16]. After discarding insignificant predictors, the findings revealed that these elements significantly predicted the rating. However, the proportion of explained variance was lower(R^2 = .218) compared to the former model(Nagelkerke R^2 = .323).

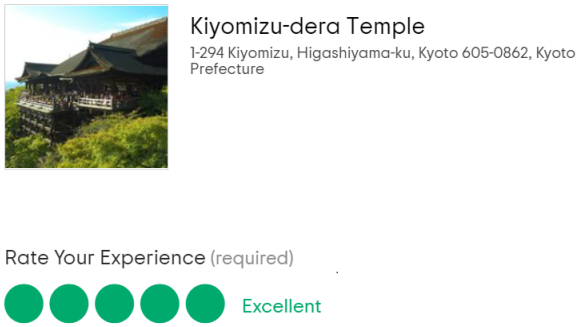

Figure 6. The questionnaire survey provided by TripAdvisor, i.e., the affective experience of 'Excellent, 'Very good', 'Average', 'Poor', and 'Terrible' can be selected.

4. Discussion and Conclusions

The present study investigates the impact of visual breadth on the attractiveness of tourism destinations and examines how high visual breadth can provide people with a positive emotional experience. Our findings indicate that physical and contextual breadth interactively influence the experience rating of tourist attractions (see the rating questionnaire in Figure 6), highlighting their complementary relationship and demonstrating better predictive performance compared to related studies that only considered visual elements. These findings align with the broaden-and-build theory and support the proposed two-dimensional visual breadth model for the affective experience.

Prior cognitive studies often treated visual contents as simple combinations of pure, independent elements or categories [40]. While such studies were sensitive to apparent categorical differences, they might overlook the congruent phenomena that are more general, abstract, and universal [11]. To address this limitation, this study adopts a holistic approach to establish a more stable visual prediction of the general affective experience in tourism. The visual breadth proposed in this study tests a more unified visual experience and may establish a theory with cross-environment and cross-cultural consistency. In addition, the visual breadth model also addresses the duality of breadth, highlighting the multiple structures in visual breadth. Instead of suggesting that one kind of breadth is of higher aesthetic value, the study emphasizes the need to elicit a balance between physical and contextual breadth to design site-specific visual solutions that consider cultural and geographical features. Lastly, the computer vision approaches of the visual breadth measurements can help further research in cognitive and affective computing with high ecological validity. This study also provides theoretical support for the visual construction of tourism scenes and environmental-oriented emotional interventions.

Our two-dimensional visual breadth model can also be

used to promote computer vision models in recognizing daily behaviors and facial emotion (valence and arousal) [13]. As it links the environment and the affective experience with a holistic view, it may help to comprehend facial emotions and behavioral cues in the interaction between individuals and their environment. Thus, our model offers a valuable theoretical foundation and practical approach to improve emotion recognition for computer vision models.

However, the study has limitations, such as focusing only on high-rating spots and a limited assessment of the general affective experience. Future studies should validate the proposed methods on both high- and low-rating spots, refine the algorithm of contextual breadth based on qualitative research, and search for more direct indexes for affective experience.

5. Acknowledge

The authors would like to thank Dr. YuanPeng Loh (Multimedia University), Guanyu Chen (Kyoto University), and Dr. Yang Liu (Kyoto University) for their efforts in the data collection and comments on the earlier version of this manuscript; Dr. Lixin Fan (Nokia Technologies) and Dr. Chee Seng Chan (University of Malaya) for their great comments for this project. This work was supported by the China Postdoctoral International Exchange Program under Grant No. YJ20210266.

References

[1] Mona Afshardoost and Mohammad Sadegh Eshaghi. Destination image and tourist behavioural intentions: A meta-analysis. *Tourism Management*, 81:104154, 2020. 3

[2] Albert Bandura. Social Cognitive Theory of Mass Communication. *Media Psychology*, 3(3):265–299, Aug. 2001. 3

[3] Yanpeng Cao and Kay O'Halloran. Learning human photo shooting patterns from large-scale community photo collections. *Multimedia Tools and Applications*, 74(24):11499–11516, 2015. 5

[4] Mark J Fenske and Jane E Raymond. Affective influences of selective attention. *Current Directions in Psychological Science*, 15(6):312–316, 2006. 2

[5] Barbara L Fredrickson. The broaden–and–build theory of positive emotions. *Philosophical Transactions of The Royal Society of London, Series B: Biological Sciences*, 359(1449):1367–1377, 2004. 1, 2

[6] Barbara L Fredrickson and Christine Branigan. Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition & Emotion*, 19(3):313–332, 2005. 2

[7] Raghu K Ganti, Fan Ye, and Hui Lei. Mobile crowdsensing: current state and future challenges. *IEEE Communications Magazine*, 49(11):32–39, 2011. 1

[8] James J Gibson. The theory of affordances. *Hilldale, USA*, 1(2):67–82, 1977. 1, 2

[9] Robert L Goldstone and Gary Lupyan. Discovering psychological principles by mining naturally occurring data sets. *Topics in Cognitive Science*, 8(3):548–568, 2016. 1

[10] Li Gu, Xueling Yang, Liman Man Wai Li, Xinyue Zhou, and Ding-Guo Gao. Seeing the big picture: Broadening attention relieves sadness and depressed mood. *Scandinavian Journal of Psychology*, 58(4):324–332, 2017. 2

[11] Martin Heidegger and Friedrich-Wilhelm von Herrmann. *Sein und zeit*, volume 2. M. Niemeyer Tübingen, 1977. 7

[12] Li-Jun Ji, Suhui Yap, Michael W Best, and Kayla McGeorge. Global processing makes people happier than local processing. *Frontiers in Psychology*, 10:67001–67010, 2019. 2

[13] Dimitrios Kollias and Stefanos Zafeiriou. Analysing affective behavior in the second abaw2 competition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3652–3660, 2021. 8

[14] Christof Kuhbandner, Stephanie Lichtenfeld, and Reinhard Pekrun. Always look on the broad side of life: Happiness increases the breadth of sensory memory. *Emotion*, 11(4):958, 2011. 1

[15] George Lakoff and Mark Johnson. *Philosophy in the flesh: the embodied mind and its challenge to Western thought*. Basic Books, New York, 1999. 2

[16] Liu Liu, Bolei Zhou, Jinhua Zhao, and Brent D Ryan. C-image: city cognitive mapping through geo-tagged photos. *GeoJournal*, 81:817–861, 2016. 6, 7

[17] Inc Merriam-Webster, editor. *The Merriam-Webster dictionary*. Merriam-Webster, Incorporated, Springfield, Massachusetts, 2019. 2

[18] Yuri Miyamoto, Richard E. Nisbett, and Takahiko Masuda. Culture and the Physical Environment: Holistic Versus Analytic Perceptual Affordances. *Psychological Science*, 17(2):113–119, Feb. 2006. 2, 5

[19] Donald A. Norman. Affordance, conventions, and design. *Interactions*, 6(3):38–43, May 1999. 5

[20] Steve Pan, Jinsoo Lee, and Henry Tsai. Travel photos: Motivations, image dimensions, and affective qualities of places. *Tourism Management*, 40:59–69, 2014. 3

[21] Genevieve Patterson, Chen Xu, Hang Su, and James Hays. The sun attribute database: Beyond categories for deeper scene understanding. *International Journal of Computer Vision*, 108(1):59–81, 2014. Publisher: Springer. 6

[22] Alexandra Paxton and Thomas L Griffiths. Finding the traces of behavioral and cognitive processes in big data and naturally occurring datasets. *Behavior research methods*, 49:1630–1638, 2017. 1

[23] Robert W Proctor and Aiping Xiong. *From small-scale experiments to big data: Challenges and opportunities for experimental psychologists.*, chapter 2, pages 35–58. American Psychological Association, 2020. 1

[24] Gillian Rowe, Jacob B Hirsh, and Adam K Anderson. Positive affect increases the breadth of attentional selection. *Proceedings of the National Academy of Sciences*, 104(1):383–388, 2007. 2

[25] Taylor W Schmitz, Eve De Rosa, and Adam K Anderson. Opposing influences of affective state valence on visual cortical encoding. *Journal of Neuroscience*, 29(22):7199–7207, 2009. 2

[26] Narayanan Srinivasan and Asma Hanif. Global-happy and local-sad: Perceptual processing affects emotion identifica-

tion. *Cognition & Emotion*, 24(6):1062–1069, Sept. 2010. 2

[27] Jens K Steen Jacobsen. Use of landscape perception methods in tourism studies: A review of photo-based research approaches. *Tourism Geographies*, 9(3):234–253, 2007. 5

[28] Maya Tamir and Michael D Robinson. The happy spotlight: Positive mood and selective attention to rewarding information. *Personality and Social Psychology Bulletin*, 33(8):1124–1136, 2007. 1

[29] Song Tong. *Informatics Approaches for Understanding Human Facial Attractiveness Perception and Visual Attention*. PhD thesis, Kyoto Univeristy, 2021. 1, 4, 5

[30] Song Tong, Xuefeng Liang, Takatsune Kumada, Peng Zhang, and Kaiping Peng. Detecting the attention scopes from travel photos. In *Proceedings of the 2022 International Conference Information Technology and Biomedical Engineering*, pages 213–217. IEEE, 2022. 5

[31] Song Tong, Yuen Peng Loh, Xuefeng Liang, and Takatsune Kumada. Visual attention inspired distant view and close-up view classification. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 2787–2791. IEEE, 2016. 5

[32] Song Tong, Yuen Peng Loh, Xuefeng Liang, and Takatsune Kumada. Wide or narrow? a visual attention inspired model for view-type classification. *IEEE Access*, 7:48725–48738, 2019. 5

[33] Antonio Torralba and Aude Oliva. Depth estimation from image structure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1226–1238, 2002. 5

[34] John Urry and Jonas Larsen. *The tourist gaze 3.0*. Sage, 2011. 3

[35] Rufin VanRullen and Simon J. Thorpe. The Time Course of Visual Processing: From Early Perception to Decision-Making. *Journal of Cognitive Neuroscience*, 13(4):454–461, 2001. 1

[36] Francisco J. Varela, Evan Thompson, and Eleanor Rosch. *The embodied mind: cognitive science and human experience*. MIT Press, Cambridge, Massachusetts ; London England, revised edition edition, 2016. 2

[37] Huy Quan Vu, Gang Li, Rob Law, and Ben Haobin Ye. Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management*, 46:222–232, 2015. 4

[38] Heather A. Wadlinger and Derek M. Isaacowitz. Positive mood broadens visual attention to positive stimuli. *Motivation and Emotion*, 30(1):87–99, 2006. 2

[39] Kun Zhang, Dongzhi Chen, and Chunlin Li. How are tourists different?-reading geo-tagged photos through a deep learning model. *Journal of Quality Assurance in Hospitality & Tourism*, 21(2):234–243, 2020. 3

[40] Kun Zhang, Ye Chen, and Chunlin Li. Discovering the tourists' behaviors and perceptions in a tourism destination by analyzing photos' visual content with a computer deep learning model: The case of beijing. *Tourism Management*, 75:595–608, 2019. 3, 5, 7

[41] Yan-Tao Zheng, Zheng-Jun Zha, and Tat-Seng Chua. Mining Travel Patterns from Geotagged Photos. *ACM Transactions on Intelligent Systems and Technology*, 3(3):1–18, 2012. 5

[42] Bolei Zhou, Liu Liu, Aude Oliva, and Antonio Torralba. Recognizing city identity via attribute analysis of geo-tagged images. In *Proceedings of the European Conference on Computer Vision*, pages 519–534. Springer, 2014. 6