PRE-ATTENTIVE DISCRIMINATION OF INTERESTINGNESS IN IMAGES

Harish Katti, Kwok Yang Bin, Tat Seng Chua, Mohan Kankanhalli

National University of Singapore, Singapore

ABSTRACT

Interestingness is an important aesthetic property, which literally means something that arouses curiosity and is a precursor to attention. Aesthetics is becoming more important as multimedia systems become more human and content centric as opposed to technology centric. In this paper, we use insights from cognitive science, neurophysiology of the early visual system and a mix of human experiments and computational modeling for the purpose of investigating interestingness. Categories in image interestingness and their computational realization are explored through a non-trivial dataset and a real-world problem.

Index Terms Pre-attentive vision, interestingness, categorization of interestingness.

1. INTERESTINGNESS: AN AESTHETIC ATTRIBUTE IN IMAGES

Interestingness is different from mere statistical similarity to a group of images representing a particular concept, this is illustrated in Figure 1 with image pairs related to the concepts "sunset" and "books", the lower image of the pair shows an image that is not only relevant to the concept, but also significantly interesting.

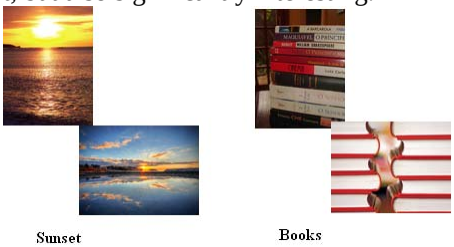

Figure 1: Sample images for "sunset" and "books" concepts from the flickr image pool. The images in lower row are found to be significantly interesting to the user community.

The explosion in digital image collections like Flickr®, Facebook®, etc are bringing out the importance of aesthetics. Aesthetic properties pose a big challenge for multimedia practitioners due to the higher levels of abstraction, subjectivity and computational cost. Understanding of human cognition and perception could give a richer and meaningful insight to tackle aesthetics and similar properties which represent a much higher level in the semantic hierarchy. Different amounts of cognitive processing and prior knowledge might be required to determine if an image is interesting. This is illustrated in Figure 2 where images become more abstract from left to right and need more computation, real-world knowledge and abstract thinking.

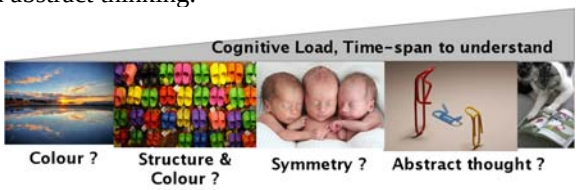

Figure 2: Images with different kinds of interestingness properties from the flickr image pool.

1.1. Problem Definition

- Are there categories in interestingness? Can interestingness be determined in pre-attentive time span for some of these categories?
- Can a computationally feasible model give comparable results to humans for interestingness discrimination in a non-trivial dataset?

2. RELATED WORK: AESTHETICS, PRE-ATTENTIVE VISION AND GLOBAL PROPERTIES

Closely related work to interestingness discrimination is that of image categorization (indoor-outdoor, natural-manmade, etc), object-recognition, etc. Of particular interest is [5] where the authors explored aesthetics scores similar to interestingness in the Photonet community using 56 statistical measures and a classification model to obtain moderate accuracy for aesthetics ranking. Though similar in spirit, our paper focuses on the study of human perception and ties it meaningfully to a computational model for interestingness and involves a very large real world corpus in the process. The visual perceptual process involves initial pre-attentive processing of images and subsequent fixation over points in the image as the attention mechanism sets in. The pre-attentive vision is significant because of the short time spans of 30-50 milliseconds involved and that robust object recognition, segmentation, etc are yet to be performed [3]. The authors in [3] showed that basic categorization of scene type is possible within such a time span for categories such as "indoor", "outdoor", "natural", etc. Their experiments also showed that description of visual input becomes richer and comprehensive as the presentation time for the stimuli is increased. Table 1

illustrates a sample result from [3] in their experiment to guage the scene understanding possible in pre-attentive time span and shows increase in description detail even with an increase of 10's of milliseconds in presentation time.

Presentation time	
27 millisecond	Mostly dark, some square things, maybe furniture
40 millisecond	Indoor shot, large framed object, white background
67 millisecond	Interior of room, picture to right & black, table in center

Table 1: Sample result from [3] where subjects attempt to describe a visual stimulus which is shown over different presentation times.

Another attractive feature of pre-attentive vision is the possibility of mainly feed-forward architecture of processing [7], which could make robust and useful computational models possible. Salient features of the pre-attentive stages are: faithful reproduction of retinal image on the cortex and separate processing of the intensity and colour information present in the visual stimulus. Global properties of a scene like its overall structure and the dominant orientations have been shown to be processed in this short time span [3]. These are helpful in capturing a 'gist' of the image. Low spatial frequency information can convey a good sense of this global information [1] and also generate the context which could then help improve subsequent segmentation, recognition [1] [7] and recall phases [1].

Global properties (Image-wide) and local properties (limited to a smaller region) have been shown to contribute to image categorization [8], where the authors selectively enabled global properties by blurring, local properties by dividing the image into 100 identical sized blocks and scrambling them in a categorization task. The impact of colour was investigated by using gray-scale versions of images.

3. OUR APPROACH

3.1. Data collection

Flickr implements an interestingness algorithm using human-activity and social-networking data [2], to compute interestingness scores, producing significantly engaging retrieval results. We queried more than 30,000 images with keywords belonging to one of 14 categories, 7 natural, 7 man-made as per [8] and then created the list of keywords by using a bag of words approach using synsets from WordNet. For example, the concept "forest" was expanded to (woods, timberland, woodland, timber, grove, jungle).

3.2. Experimental investigation of pre-attentive interestingness

The experiment is designed for answers to the first question in section 1.1. It involves presenting a pair consisting of an interesting image and a control image (each pair randomly selected from amongst 14 categories [6]) to the user over two time spans. The presentation time is varied between 16 to 1000 milliseconds in the first stage (using the MATLAB Psych Toolbox and it's APIs for stimulus presentation control) as shown in the upper panel in Figure 3. In the second stage, the same pair is presented simultaneously for a fresh decision on the interestingness (Figure 3, lower panel).

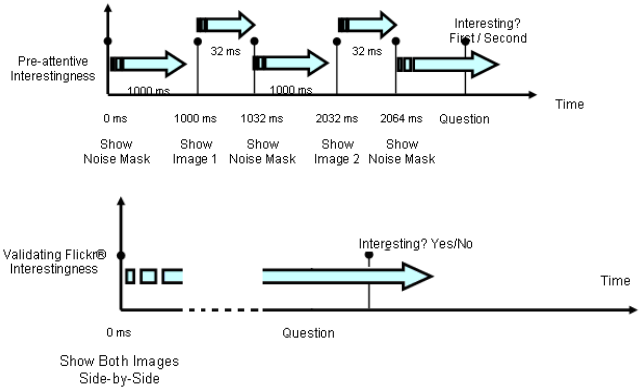

Figure 3: The experiment protocol for the first (top panel) and second stage (lower panel) of the experiment.

The first stage is a forced-choice experiment and a choice of rejecting the image pair is given in the second stage. This is done with two objectives; the first is to force a response from the user in the first stage where presentation time is too short to get elaborate idea of the image. The second stage focuses on validating Flickr's meta-data based interestingness algorithm and allows rejection of image pairs with similar interestingness.

4. RESULTS AND DISCUSSION

4.1. Experimental results

The agreement of the user's pre-attentive decision and long term decision is done based on the number of times the decision is consistent in the two stages. Trials in which user is undecided in the second stage are discarded. This agreement ratio is generated as,

$$pre\_attentive\_hit\_ratio = \frac{\sum_{i=1}^{k}((choice_{i,stage1} \equiv choice_{i,stage2})}{\sum_{i=1}^{k}((choice_{i,stage2} \neq -1)} \text{,……..…eqn 1}$$

$$(choice_{i,stage1} \equiv choice_{i,stage2}) = \begin{cases} 1, (choice_{i,stage1} = choice_{i,stage2}) \\ 0, (choice_{i,stage1} \neq choice_{i,stage2}) \end{cases}$$

For the indicator of users' agreement with the Flickr interestingness algorithm (User-to-Flickr agreement ratio), we used choices made in the second stage of the experiment,

$$user\_to\_flickr\_agreement = 1/k * (\sum_{i=1}^{k} C_i) \text{ .,…eqn 2}$$

where, C_i is the choice made in trial for the long term presentation of the i[th] image pair. The choices being (0-left image, 1-right image,-1-undecided). The goodness of pre-attentive decisions made by users is shown in Figure 4. Pre-attentive decisions made between 30-50 ms can be seen to be consistent and indicative with those made over a longer presentation times. The wide variation at 16 ms indicates lack of discrimination at very short time spans. High values at 50 ms are followed by a drop at ~100 ms before converging to the steady (high/higher) value beyond 500 ms. This could indicate different cognitive process responsible for short-term and longer-term discrimination.

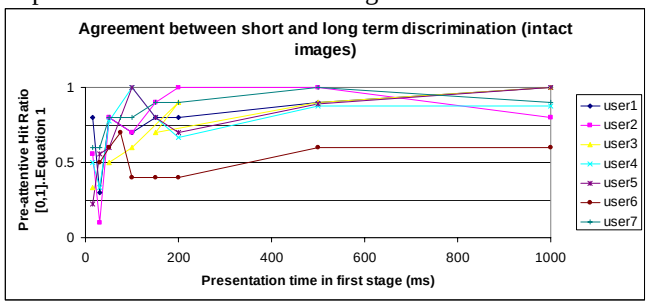

Figure 4: Agreement of decision made at short presentation times (< 100 milliseconds) with final decision on interestingness. Significant interestingness-discrimination is possible in pre-attentive time span.

A binomial test over the agreements between first and second stage showed that the agreements are significant from 33milli-seconds onwards. This could indicate that we make significant decisions about interestingness in very short time spans. A similar analysis showed significant agreement between user's notion of interestingness and that of Flickr's algorithm which supports the use of the dataset for this study. The influence of different kinds of visual information on the User's interestingness in Flickr content is shown in Figure 5.

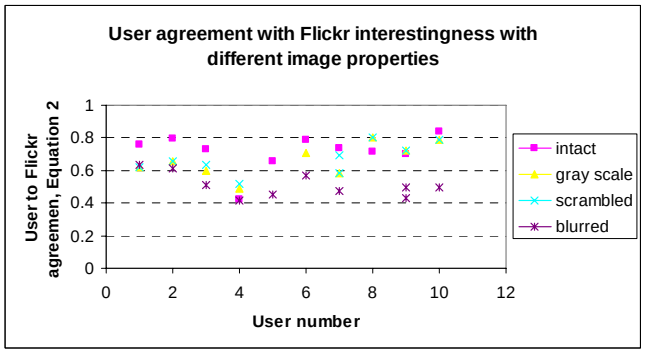

Figure 5: The figure illustrates how closely Flickr's notion of interestingness matches with that of different users when colour, local features and global features are selectively enabled

Intact image information allows for maximum discrimination followed by selective enabling of local-properties, gray-scale information and global information. This can be used to weigh the features extracted for discrimination in a computational model. In another experiment, user rating of randomly chosen images was performed by 8 users on a scale of 0 (least)- 9 (most interesting) and the time to score images was also recorded and standardized to obtain Z scores for a total of 640 images as shown in Figure 6.

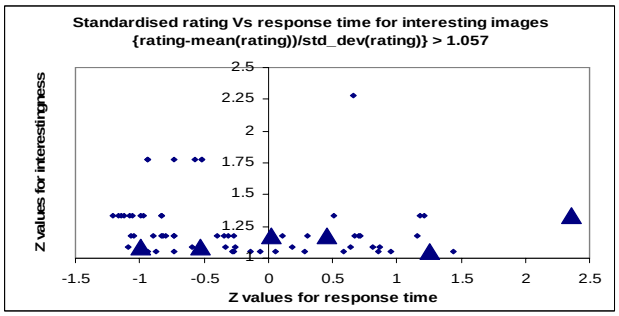

Figure 6: Grouping of interesting images according to user response times. Z scores for interestingness rating are plotted against Z scores for response times. $Z(Var) = \frac{Var - mean(Var)}{\sigma(Var)}$

Further analysis brought up different kinds of images that seem to represent categories of interestingness based on the dominance of visual features. Table 2 groups the high-interestingness images according to clusters from Figure 6. Representative images are chosen from these response time groups and their complexity is analyzed in Table 2.

Representative images	**Response time Z scores**	**Salient features from literature that are dominant in identified categories**
	-1.1	Low depth of field (dof) familiarity(left).Shape , Colour, low dof (right)
	-0.5	Shape, Form, Lines, Colour (left). Symmetry, Lines (right)
	0	Medium dof, familiarity, natural scene (left) Symmetry, dof, symmetry (right)
	0.5	Colour, Shape, form, 1/3rd rule
	2.4	High Symmetry, high dof, pattern, colour, shape, form

Table 2: Features from highly interesting images that are dominant at different user response times.

Higher response time seems to indicate higher complexity (e.g.; increased symmetry). Further analysis of

these results will yield better insight into the nature of these categories.

5. APPLICATION TO REAL WORLD PROBLEMS

The notion of interestingness in digital images can be used in more than one application that addresses the second problem in section 1.1. Possible scenarios include,

-Filters for image retrieval results, to improve the perceptual quality and make them more engaging to the user.

-Identifying community and individual preferences in image collections for more effective browsing.

5.1. Personalized, Intelligent agents for interaction with digital image collections

An important challenge is that of finding features that correlate well with image aesthetics and realizing them as scores for model training. As many features like depth-of-field and navigability are computationally difficult, a novel approach is employed by approximating such perceptually meaningful visual features with Exchangeable File format(Exif) header fields as described in Table 3.

Property and computational realization
Line, colour distribution, (realized using edge histogram, colour histogram)[4] symmetry, shape (2D) and form (3D) , texture ,pattern (need to be approximated in future work)[4]
Depth(Exif-DOF), expanse(Exif-Focal length, zoom), openness, temperature(Exif-Light, Exposure), navigability (Exif-Zoom) [3]
Other Exif Information-Aperture value, Exposure time, F-Number, Focal Length, ISO speed, etc.

Table 3: The table shows different perceptually relevant properties of images that can be helpful for computational modeling of aesthetics.

One individual agent per user and one community agent is trained with data selected as shown in Figure 7.

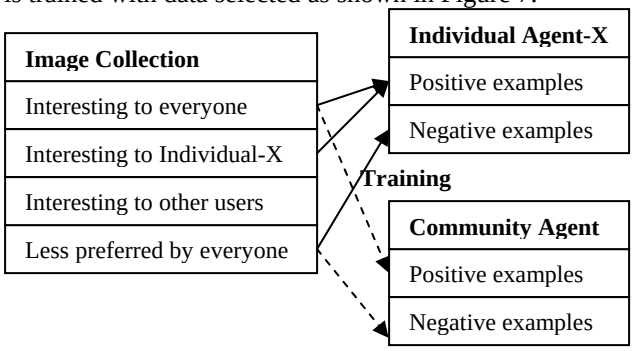

Figure 7: Selection of training data for the community agent and one individual agent.

The effectiveness of the agents is verified by performing SVM regression in the Weka environment (SVMreg, polynomial kernel, exponent=1) with 10/1 cross-validation over two personal image datasets containing close to 1000 ranked (multi-user ranking), un-manipulated images with Exif data. Almost all images fall in the three low response time categories in Table 2 to avoid dominance of complex attributes like symmetry which are not in the scope of this paper. Individual agents capture user's bias and yield moderate correlation values in the range [0.4, 0.65] to the corresponding individual's rating. The community agent was trained on images having extreme interestingness across majority of users, i.e.; consistently high and low scored images as positive and negative instances respectively. This avoids the effect of individual bias in the community agent. The community agent models community preferences with correlation in the range [0.6, 0.7] over average interestingness scores, highlighting that the community agent is capable of modeling community preference better than individual agent's modeling of an individual preference. Attribute selection showed Aperture, ISO-speed, focal-length as influencing interestingness strongly. Future work aims to expand the range of features and use metadata and visual data from the larger flickr pool.

6. CONCLUSION

We have shown using user-studies and real-world image database from that there is significant evidence showing that people can discriminate interestingness in pre-attentive time spans. Also that such an approach can lead to fruitful real-world applications. The experiments also affirm that activity and social network analysis has significant merit.

7. REFERENCES

[1] Bar M., Visual Objects in Context, Nature Reviews: Neuroscience, 5, 617-629, 2004

[2] Butterfield D S, Costello E, Fake C, Media Object Metadata Association and Ranking, United States Patent Application, 20060242178, 2006

[3] L. Fei-Fei, Natural Scene Categorization, from humans to computers, Scene Understanding Symposium, 2007

[4] Peterson Bryan, Learning to See Creatively: Design, Color & Composition in Photography, Amphoto Books, 2003

[5] R. Datta, D. Joshi, J. Li, and J. Z. Wang, Studying Aesthetics in Photographic Images Using a Computational Approach, *Proc. ECCV*, Graz, Austria, 2006

[6] Torralba A., Oliva A., Statistics of natural image categories, Network: Computation in Neural Systems, Vol. 14, 391-412. 2003.

[7] Serre T., Oliva A., Poggio T., A Feedforward Architecture Accounts for Rapid Categorization, Proceedings of the National Academy of Sciences PNAS, 104(15): 6424 – 6429, April 10, 2007.

[8] Vogel J, Schwaninger A, Wallraven C, Categorization of Natural Scenes: Local versus Global Information and the Role of Color, ACM transactions in applied Perception, 2007