Interestingness Discrimination in Images

Harish Katti, Kwok Yang Bin, Chua Tat Seng, Mohan Kankanhalli, *Member, IEEE*

***Abstract*— Interestingness in images can be described as a property of its content that arouses curiosity and is a precursor to attention. The study and modeling of such aesthetic properties is becoming more important as people are increasingly becoming comfortable with capturing, authoring and consuming multimedia content. In this paper, we use insights from cognitive science and insights into the early visual system with a mix of human experiments to study interestingness discrimination. We also perform computational modeling from our studies on global colour properties of the images and study its capability to perform semantic categorization and interestingness discrimination is studied. Automating the process of interestingness discrimination is explored through real-world applications. A large dataset from the Flickr pool and real-world classification problems are also explored. Our studies show the significant difference between semantic relevance of images and that of user-preference and appeal.**

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

I. 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 “Matsumoto Castle” and “Sunrise”, the lower image of the pair shows an image that is not only relevant to the concept, but also significantly interesting. The explosion in digital image collections like Flickr®, Facebook®, etc are bringing out the importance of aesthetics. A variety of statistical methods based on image content, related text content, tags are used to provide most relevant answers to a users query to such collections. Image content being sensory in nature, makes it very difficult to evaluate the appeal for the intended user. This has made text content a very powerful in such tasks, as any form of text including tags, text description, titles, etc are composed by human users and capture higher level semantics as compared to the image itself. Aesthetic properties pose a big challenge for multimedia practitioners due to the higher levels of abstraction, subjectivity and computational cost and even intractability.

Manuscript received April 15, 2008. Harish Katti, Chua Tat-Seng and Mohan Kankanhalli are with the Department of Computer Science, School of Computing, National University of Singapore, Computing 1, Law Link, Singapore, 117590. E-mail: {harishk,chuats,mohan}@comp.nus.edu.sg. Telephone:+65 6516 2727, Fax: +65 6779 4580

Kwok Yang Bin is with ZopIM, LLP, 49 Grove Drive, Singapore, 279089 E-mail: yangbin@zopim.com. Telephone:+65 9049 3060.

A. Aesthetics in Images

Attempts have been made to define aesthetic properties like colorfulness [11] colour harmony [10] using the HSV colour space. Often these rules model heuristics followed by photographers or features that are popular in image processing to perform classification and categorization and are similar that way to Multimedia Information retrieval techniques.

Figure 1: Sample images for “Matsumoto Castle” (left) and “Sunrise” (right) concepts from the flickr image pool. The images in lower row are found to be significantly interesting to the user community.

An example is [6], where the authors explore a combination of such rules and content derived features to perform aesthetics classification.

B. Role of the user’s cognition in aesthetics discrimination

Intuitive methods have been used like Relevance feedback using user feedback to incrementally improve the quality of retrieved result set and active learning which is a similar approach where user feedback is taken to generate positive or negative examples for further learning in the system. These methods can be seen as an attempt to involve the cognitive processes of the user into the system implementation. The problem of ensuring that the algorithms developed for search, retrieval and browsing are not only efficient, but also perceptually meaningful for the user is becoming important. This is inevitable as multimedia systems become more human and content centric as opposed to technology centric. It could be useful at this point to consider how modeling of psycho-physics into computationally feasible definitions and implementations has given us insights such as perceptually relevant colour spaces and compression methods.

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 are ordered according to the complexity of features that makes the images interesting. The image become more abstract from left to right and need more computation, real-world knowledge and abstract thinking.

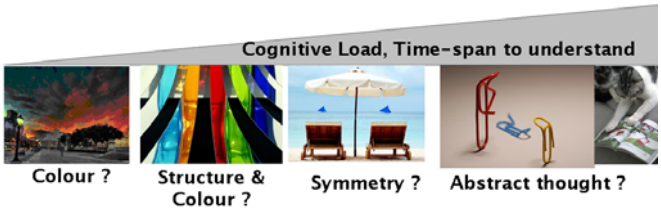


Figure 2: Images with different kinds of interestingness properties from the flickr image pool. The cognitive load appears to increase as the images get more complex and abstract from left to right.

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. To make the work relevant to the nature and scale of a real-world problem, a non-trivial collection of more than 30,000 images across 14 categories is chosen from the popular Flickr® collection for the experiments and analysis.

C. *Problem Definition*

The capability of humans to discriminate interestingness and the possibility of categories and ways to automatically recognize them are investigated through the following problems,

- 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?

It is important if this can be done, as it would demonstrate that publicly available social network information can give deeper cognitive and behavioral insights. It will also go on to reinforce that akin to text-processing and natural language processing, analysis of large corpora of human expression is useful, not only in textual domain, but also in non-textual media.

D. *Overview*

This paper explores interestingness discrimination done by humans viewers when presented with different images from the same semantic category. This work is extended from [3] where some preliminary results are presented. In [3], different properties of the image content are explored, as well as the discrimination capability variation with presentation time. The authors also show in [3] that humans have a significant capability to discriminate interestingness in images in a pre-attentive time scale. This paper includes computational experiments to find the impact of global colour properties on categorization of images and interestingness discrimination and demonstrates its discrimination capability. The paper also explores the possibility of making use of utilizing some of the experimental findings into practical, real-world applications.

II. Related work: Aesthetics, Pre-attentive vision and Global properties

Closely related work to this paper on interestingness discrimination is that of image categorization (indoor-outdoor, natural-manmade, etc), object-recognition, etc. Of particular interest is [7] 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.

A. *Pre-attentive Vision*

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 [4]. The authors in [4] 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. The experiments also point out to the rich and comprehensive description of visual input as the presentation time for the stimuli is increased. Another attractive feature of pre-attentive vision is the possibility of mainly feed-forward architecture of processing [8]. Table 1 illustrates a sample result from [4] in their experiment to gauge the scene understanding possible in pre-attentive time span. It can be seen that the description detail increases even with an increase of 10's of milliseconds in presentation time.

The upper panel illustrates the stimulus image and the lower panel illustrates outcome of a trial in which the subject was presented the same image with different presentation times (at different times). It can be seen that the description of the image improves as the presentation time is increased even in 10's of milliseconds. The same results are re-tabulated in Table 1. It can be seen that small increments in presentation time result in remarkable refinement in scene description.
Another attractive feature of pre-attentive vision is the possibility of mainly feed-forward architecture of processing [8], which could make robust and useful computational models possible. Salient features of the pre-attentive stages are: faithful reproduction of retinal image on the Lateral geniculate nucleus (LGN) and cortical retino-topic maps and

separate processing of the intensity and colour information present in the visual stimulus. The LGN is a part of the brain which is the primary processor of visual information coming in from the retina in the human central nervous system.

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 [4] where subjects attempt to describe a visual stimulus which is shown over different presentation times.

B. Global properties in images

Global properties of a scene like its overall structure and the dominant orientations have been shown to be processed in this short time span [4]. 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] [8] and recall phases [1].

Global properties (Image-wide) and local properties (limited to a smaller region) have been shown to contribute to image categorization [9], 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.

Some of the salient features of the pre-attentive stages involving the retina, Lateral Geniculate Nucleus (LGN) and V1 region in the human brain cortical region are:

- Existence of retinotopic maps in LGN and cortical regions
- Faithful reproduction of retinal image on the LGN and cortical retinotopic maps
- Separate processing of the intensity and colour information present in the visual stimulus
- Trichromatic colour in the retina and colour opponency in the retina and LGN regions

Global properties of a scene like its overall structure and the dominant orientations have been shown to be processed in this short time span [7]. Global properties of images are helpful in capturing a 'gist' of the image. These features could be related to the structure (orientation) or colour (saturation of colours, colour composition), etc. Low spatial frequency information can convey a good sense of this global information [1] and also generate the context which can then help improve subsequent segmentation, recognition [1] [8] and recall phases [1]. Figure 3 illustrates properties of man made image categories as they show up in the global, spectral signature computed using a Fast Fourier transform (FFT) over images from three categories in our dataset following the approach in [7]. The contour plot representing the spectral components in which the energy (information) of the image is concentrated is shown.

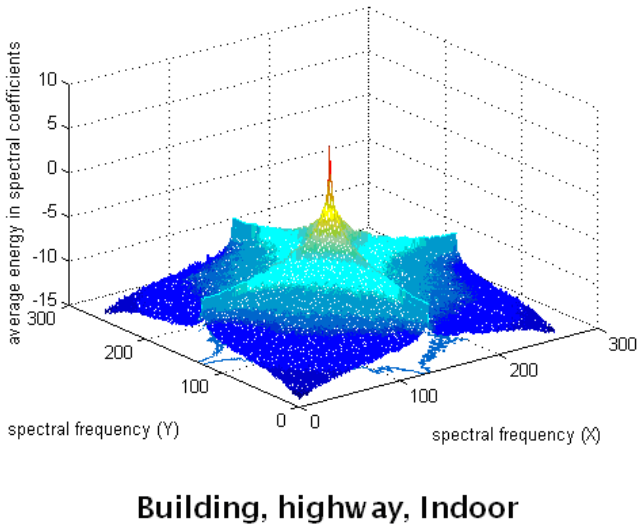


Figure 3: The figure illustrates FFT based spectral signature computation for natural scenes for the dataset being used in the experiments. The strong vertical and horizontal orientations in man-made scenes come out as peaks in the spectral signature.

The strong vertical and horizontal orientations in man-made scenes come out as peaks in the spectral signature.

III. Our Approach

A. Data collection

Using Flickr's public API, we queried images with keywords belonging to one of 14 categories, 7 natural, 7 man-made as per [7]. We created the list of keywords by using a bag of words approach using synsets from WordNet. The control set of images was downloaded using "relevance" as the sorting priority, and the interesting set of images was downloaded using "interesting-descending" as the sorting priority. In total, we downloaded 9,137 interesting and 16,244 relevant images. Table 2 shows the bag-of-words approach with a few example categories, and the number of images retrieved as well as the number of images that had to be purged.

It was found that subjectivity of users introduced lot of noise in publicly available social networked media as the words relating to a concept are often used in different ways, for example "woods" is frequently used as a name for people, buildings, etc. These images were filtered out manually. During the experiment it was also found that the nature of

images is also of concern as it could be offensive or unpleasant to users. Manual effort is required for such cleaning and was a laborious process in a large dataset such as the one used in this project. The description of various image categories and the details of the bag-of-words used for each category is tabulated in the following table 2. This is for 5 representative categories of the 14 for which images were extracted namely -beach, city-view, coast, field, forest, high-building, highway, indoor scene, man-made object, mountain, natural object, portrait, street.

Category	Forest	Mountain	Field	Beach	Indoor Scene
Bag of words	woods timberland woodland timber grove jungle	mount highland hill ridge alp volcano peak	clearing grassland crop harvest paddy cultivation	shore plage sand seaside	interior bedroom office dining kitchen library
Images	531	679	623	500	753
Noise	63	110	76	76	115

Table 2: The table illustrates the bag-of-words approach used to expand the query concept for image retrieval. The second and third rows are the number of images collected in the category and irrelevant images that were manually filtered out. The experimental section of the project is devoted to finding out whether interestingness in images can be discriminated at a pre-attentive time scale. At the same time, we attempted to validate Flickr's algorithm by asking subjects (given enough time for attention) whether they thought certain images were interesting, and correlating that with what Flickr returned.

B. Experiment to capture the pre-attentive discrimination of 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 [7]) 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 its APIs for stimulus presentation control) as shown in Figure 4.

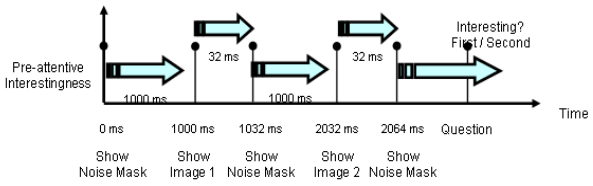


Figure 4: The experiment protocol for the first stage of the experiment.

In the first stage, we presented the image pair one after the other, interspersed with noise masks. The presentation time of images is varied between 16 to 1000 milliseconds. The user was forced to pick either the first or second image as the more interesting one in the first stage. In this stage the presentation is momentary and the user may not have a strong urge to select either image as interesting, so by forcing the user we bring out any influence the presented stimulus had. If there really was no difference in interestingness between the two images, the trial would be discarded as we processed data from the next 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 and otherwise the users would reject most pre-attentive presentations. This decision is effective and can be seen from the significance in discrimination in first stage for presentation time greater than 50 milli seconds.

C. Experiment to capture interestingness discrimination over long term presentation of Images

The same pair is then presented simultaneously for a fresh decision on the interestingness as shown in Figure 5.

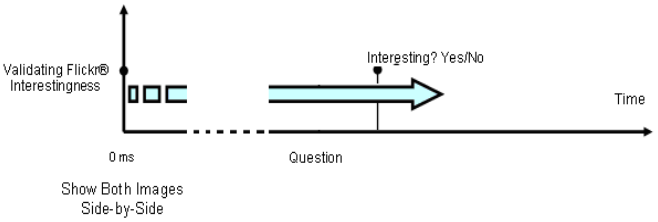


Figure 5: The experiment protocol for the second stage of the experiment.

The second stage allows rejection of image pairs with similar interestingness. In the second stage, each pair of images presented earlier is displayed side-by-side for a longer duration. The subject has as long as he/she wants to decide which image is more interesting. In addition, the subject is allowed a third option to reject the image pair if it is difficult to discriminate based on interestingness. Trials in which users reject the image pair are not considered for calculating agreement of the user's pre-attentive and attentive decision, but are still used for checking agreement with Flickr. We randomize the order in which the images are shown, both temporally as in the first stage and spatially in the second. Decisions made at both stages are recorded and analyzed. Figure 6 shows representative examples of our noise mask (screen capture at 1024×768 pixels) and the stimulus presentation in stage 2.

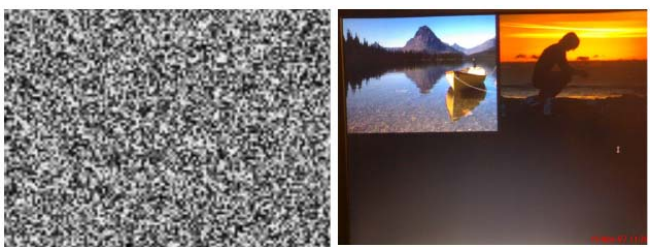


Figure 6: Shows noise mask generated to avoid image persistence in the first stage (left panel) and simultaneous presentation of control and interesting images (right panel) to user in the second stage of the experiment.

D. Experiment to investigate the effectiveness of noise masks in destroying image persistence

Noise masks used in the two experiments ensure that image persistence on the eye does not influence the results for short presentation of images. This can significantly alter the results for short presentation spans. This is illustrated in Figure 7 for a user across different presentation time spans. An overall increase in agreement can be seen when the noise mask is absent. This is made possible by the persistence of images in the visual system after the presentation stimulus is removed.

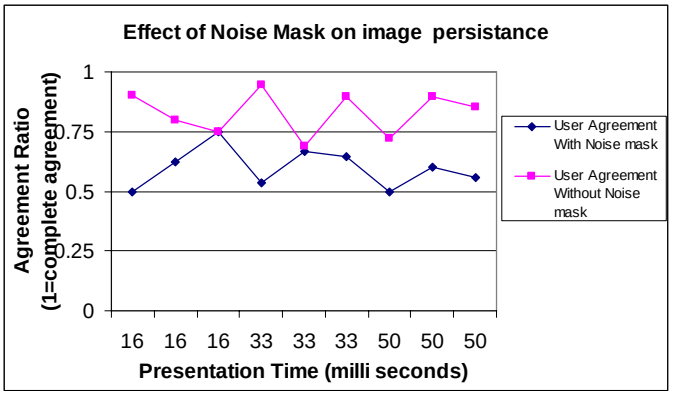


Figure 7: The figure illustrates the impact of noise masks in reducing the effect of persistence of visual stimulus. The two plots are measures of consistency of user-decisions made for stage 1 and stage 2 as defined in equation 1 earlier. It can be seen that persistence in the absence of the noise mask significantly increases the overall discrimination capability of the user.

E. Experiment to investigate the role of global, local properties and colour information for the discrimination of interestingness

To investigate the contribution of global and local intensity based information and that of colour on interestingness discrimination, we conducted experiments similar to those in [9] with different modalities of images as presented in Figure 8. Global colour and intensity information information in images has been shown to be processed in the short pre-attentive time spans [1].

The influence of global intensity information is established by asking subjects to perform interesting versus non-interesting discrimination between colour-drained versions of images from our corpus. Influence of local features is investigated by dividing colour drained images into 100 blocks and scrambling them randomly. Global feature information influence is evaluated by blurring images to destroy local information. Once the appropriately manipulated images are generated as shown in Figure 8, the experiment protocol is the same as the earlier one for pre-attentive discrimination. Images are first manipulated to get the appropriate global or local properties and then presented to the user.

Figure 8: The figure illustrates different modes of presenting the images. The four modes represent the following manipulations, intact (top left), scrambled (top-right), grayscale (bottom left), blurred (bottom right).

F. Computational experiment to capture global colour properties

An attempt is made to model the global colour properties of images using a biologically inspired approach. The YCbCr space was explored as it ties very well with pre-attentive vision models. Previous work [7] based on Fourier Transforms captured scene structure but in Y space, discarding colour information. We believe colour to have an important pre-attentive role in conveying scene context. To capture global colour information, we created a colour histogram based on values in CrCb space.

This histogram has values on a 2D grid, and is shown in projected 3D. Firstly, the image is converted to CrCb space. We quantize the CrCb space into a 16-by-16 grid of bins. Next, for each bin, we count the number of pixels in the image with a CrCb value that falls into that particular bin. For example, an image with very low saturation or little colour will have a histogram with values mainly near the centre, while an image with only saturated reds and greens will have values along the red-green axis at the extremes. Figure 9 shows the side- and top-view of histograms of five images, where each image is completely filled with red, yellow, green, blue or gray. These images have their histograms coloured with the colour their images contain. As can be seen from the diagram, most of the 'color energy' for a colour-filled image is concentrated around the CrCb representation of that particular colour.

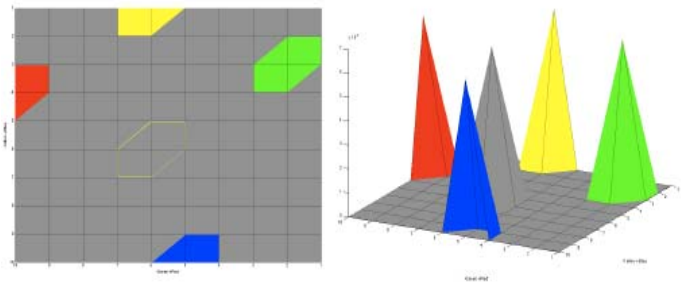


Figure 9: Showing the colour histogram for 5 images, each with pure red, yellow, green, blue and gray respectively. The points are plotted according to their intensity of Cr and Cb values.

The red / green images are not completely neutral about the yellow-blue axis: this is due to colorspace differences between RGB and YCrCb space. We now have a 128 element vector that describes the colour information of each image. The bin resolution can be adjusted, and preliminary results indicated a 16-by-16 grid gave a reasonable compromise between histogram resolution and a descriptor vector that is not too long. This method was used to explore whether effective discrimination of image categories could be done using the biologically relevant YCbCr colour space.

IV. Results and Discussion

A. 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,

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

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

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,

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

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 10. 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.

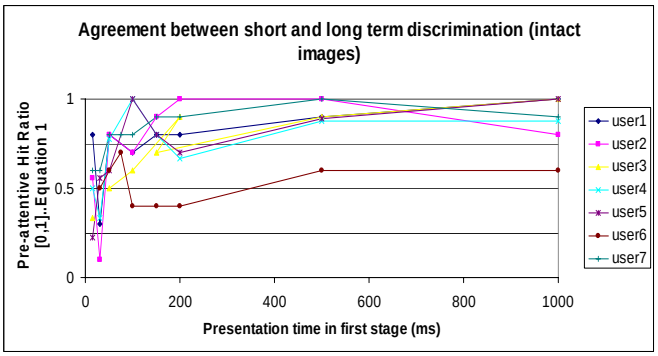


Figure 10: Agreement of decision made at short presentation times (< 100 milliseconds) with final decision on interestingness. Significant interestingness-discrimination is possible in pre-attentive time span.

A statistical test using the 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 statistical analysis showed significant agreement between user's notion of interestingness and that of Flickr's algorithm as shown in Figure 11 when images were shown to the user for sufficiently long time span.

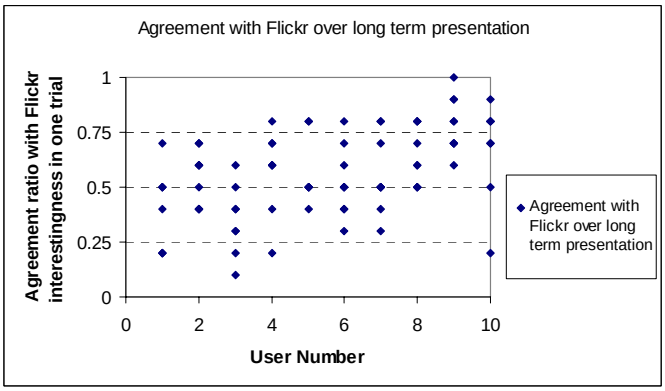


Figure 11: The figure illustrates how closely Flickr's notion of interestingness matches with that of different users over long duration presentation.

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. These clusters could represent some notion of categories within interestingness based on visual content of images. Response time is used to perform this grouping as shown in Figure 12. The triangles represent such groups and are elaborated further in Table 3.

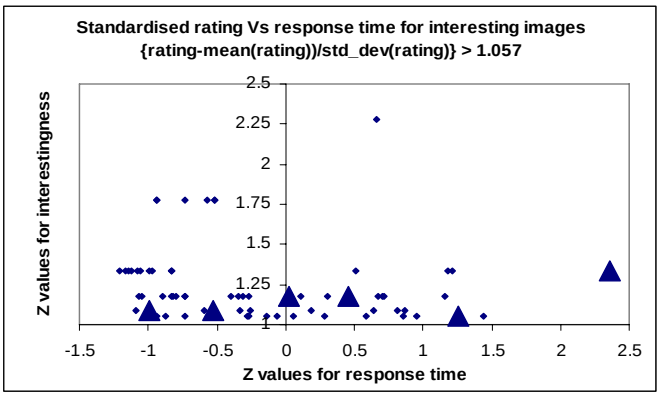


Figure 12: Grouping of interesting images according to user response times. Z scores for interestingness rating are plotted against Z scores for response times. $Z(Var) = \frac{Var - mean(Var)}{\sigma(Var)}$

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 12. Representative images are chosen from these response time groups and their complexity is analyzed in Table 2.

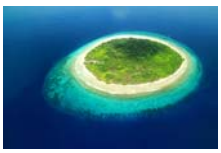
Representative images	Response time Z scores	Salient features from literature that are dominant in identified categories
	-1.1	Low depth of field (dof) familiarity, Colour, low dof (
	-0.5	Shape, Form, Lines, Colour
	0	Medium dof, familiarity, natural scene , Symmetry, depth-of-field, symmetry
	0.5	Colour,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

Table 3: 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).

B. The role of global, local properties and colour information for the discrimination of interestingness

We found that colour influences our capability significantly for interestingness discrimination over short time spans significantly. This is shown in Figure 13 (intact versus gray-scale). The global and local intensity information is found to contribute almost equally as can be seen in the results for blurred and scrambled image types. Colour seems to be an inherently global property, as it has been found necessary to completely remove colour from the scrambled images to ensure that only local properties are used for discrimination. The spatial layout of colour in the image appears to be a strong indicator of semantic category, interestingness is also significantly influenced by colour as seen from the difference in discrimination for intact and grayscale versions of images. 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.

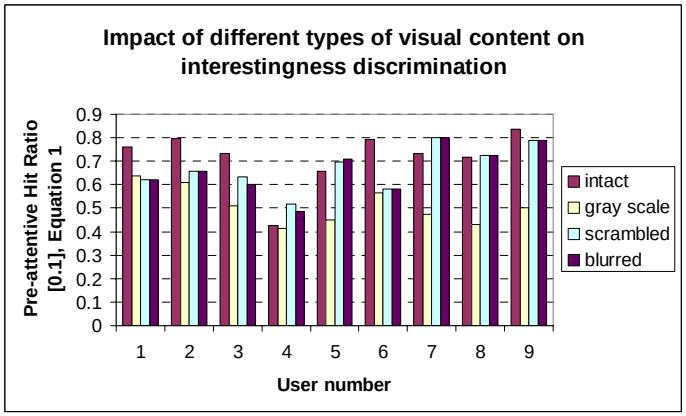


Figure 13: The figure illustrates the contribution of colour, local and global properties of images for discrimination of interestingness. The averaged pre-attentive hit ratio for each user over different presentation times is shown for each of the 4 different modes.

C. Investigation of global colour properties

Using the binning method suggested earlier, images from the dataset from seven categories were processed and the capability of the measure to distinguish the categories is explored, a representative result is shown in Figure 14. The categories are colour coded. The figure 14 illustrates the results of binning images from 7 categories; the result suggests that category discrimination can be done using the proposed global colour based signature. The next question to be explored is that of interestingness discrimination using this method.

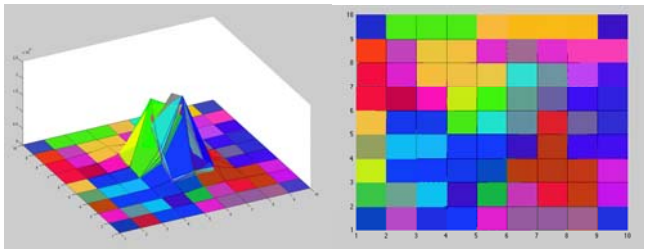


Figure 14: The figure illustrates the results from applying the global colour feature over seven categories in the dataset (beach, coast, highway, mountain, forest, high building). Each category is coded by a different colour. The right panel illustrates the dominant category falling in each bin and the panel to the left shows the actual proportion of categories in every bin.

The following figure 15 illustrates the results of inter-category discrimination and interestingness discrimination within the same semantic category.

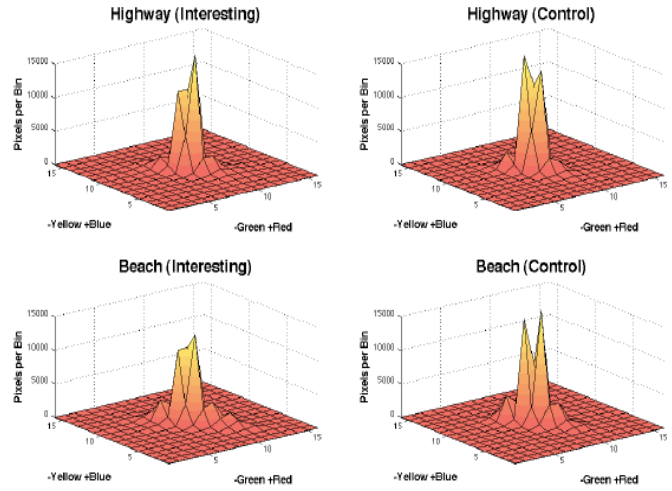


Figure 15: The panel row shows four global colour signatures showing discrimination of categories as well as interestingness. 200 images were used for each histogram.

The colour signature can be seen to provide distinct features both for images from different concept categories as well as those from same category but being interesting and non-interesting. Figure 15 illustrates differences between interesting and non-interesting images between the same category in the upper panel and those between different semantic categories (Highway-interesting versus Beach-interesting or Highway-control versus Beach-control). The possibility of categorization of images based on these cues has been explored in [7] with structural analysis using FFT and PCA over the intensity information (Y plane of the YCrCb representation) to perform automatic categorization with good accuracy for discrimination between natural and man-made images. This can be extended and augmented with these colour based features and other meta-information that approximates the visual system. The meta-data aspect is investigated in the following section.

V. 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.

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

Property and computational realization
Line, colour distribution, (realized using edge histogram, colour histogram)[6] symmetry, shape (2D) and form (3D) , texture ,pattern (need to be approximated in future work)[5]
Depth(Exif-DOF), expanse(Exif-Focal length, zoom), openness, temperature(Exif-Light, Exposure), navigability (Exif-Zoom) [4]
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 are trained with data selected as shown in Figure 16.

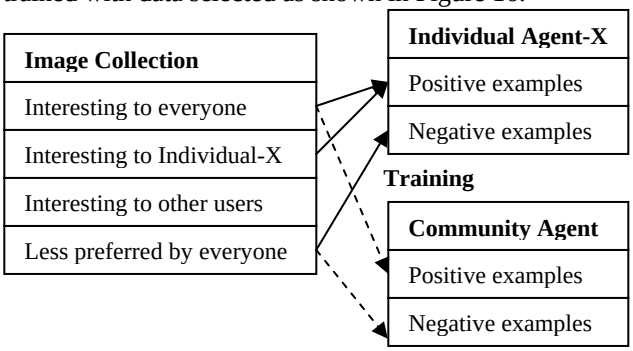


Figure 16: Selection of training data for the community agent and one individual agent.

B. Data selection

The data for training the agents consists of positive examples made up of highly interesting images and negative examples from semantically relevant but less interesting images, both types of images are selected such that they are accompanied with EXIF information. The positive examples are taken from the top of the interesting list and the negative examples are

taken from images which are low in the interesting list, but high on the relevance list. This is done to ensure that the negative examples are still semantically relevant. The search query is popular concept "Beach", which is quite extensively photographed

C. Observed on Flickr Data

The search results obtained for the two modes representing "relevance" and "interestingness" are found to be un-correlated, implying that text description and tag-based measures for semantics relevance do not give any idea of the interestingness of an image as can be expected. This is illustrated in Figure 17.

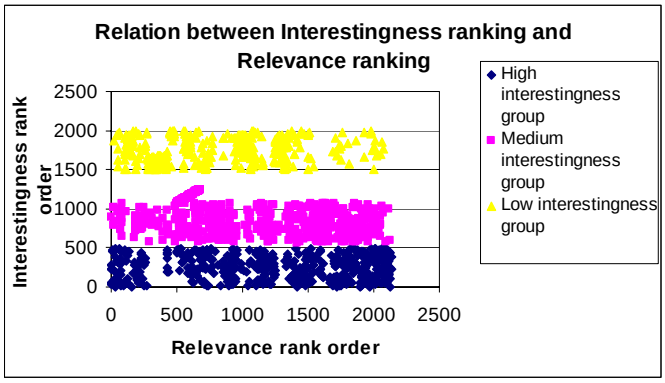


Figure 17: The figure illustrates the significant difference in ordering of the same list of images on Flickr for the concept "Beach". The semantic relevance based ordering of 2132 images is plotted against their ranks on interestingness. The images are grouped into high, medium and low ranked images according to interestingness. It can be seen that ordering based on user preference and hence interestingness of images is quite different from that based on mere semantic relevance.

EXIF penetration was found to be significant in Flickr data with up to 60% images having EXIF information in the top ranked images from the interesting list.

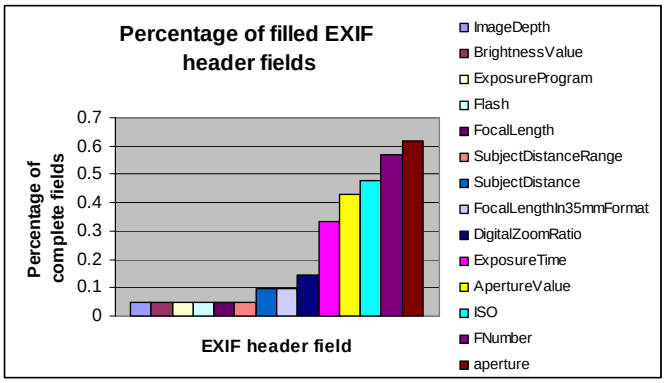


Figure 18: The figure illustrates the proportion of complete EXIF fields in top ranked interesting images.

This is a good indicator of the quick adoption of the standard and popularity with users. Not all EXIF fields are used by users though, and even amongst them, some tend to be more complete. From our dataset we found that Aperture, ISO speed, Exposure, F-Number, Flash status, Focal Length and Digital Zoom Ratio tend to be filled with more reliable values. This is illustrated in the following Figure 18.

D. Community agent for interestingness discrimination

The community agent is trained using positive and negative examples as described above. The Exif parameters chosen for classification are (aperture, ISO, Image Depth, Value, Brightness Value, Exposure Time, F-Number, Program, Subject Distance, Flash, Focal Length, Digital Zoom Ratio, Focal Length In 35mm Format, Subject Distance Range).

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 and 1/3 split with over 2100 images containing EXIF data for the concept "Beach". The regression is performed using EXIF fields and interestingness order to find the effectiveness of EXIF information for discrimination. The community agent yields a correlation of 0.38 with the combined (High interestingness + Low interestingness) rank list and accuracy of 0.65% (Root mean square error 35%) on classification between High/Low interestingness groups. Training the community agent with images from different users avoids the excessive influence from any particular user.

E. Personal agent for interestingness discrimination

The personal agent is trained from portions of the community training data belonging to the same user. This ensures that the positive and negative samples represent the preferences of a single user. Results for three user agents trained for Flickr members who have significant contribution in Flickr for images relating to "Beach" are presented in Table 4.

Flickr User ID	Total Images	Range of ranks	Accuracy, Ranking Correlation
25056484@N00	130	185-2090	60%, 0.1
37985559@N00	123	560-2036	53%, 0.2
78779687@N00	157	529-1684	55%, 0.25

Table 4: The table shows details of the user agents built to do interestingness discrimination based on EXIF information from each user's (Flickr member) contributed images in the dataset.

The correlation value and accuracy obtained is limited compared to earlier work like [6] where authors use extensive content based features. The motivation of these experiments is to demonstrate that EXIF based classification adds value to any classification scheme. Future work aims to study the contribution of the EXIF using search over different concepts in Flickr and use of content-based features that can compute higher level properties of images reliably.

VI. Conclusion

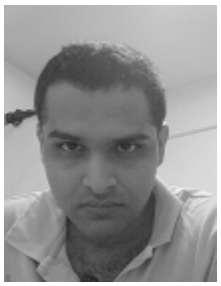
We have shown that there is significant evidence that people can discriminate interestingness in pre-attentive time spans. The investigation of global structure and colour in images indicates that colour is an important cue for interestingness discrimination. Also, that such an approach can lead to fruitful computational modeling and real-world applications. Activity analysis and social network analysis when captured in the system can capture the cognitive process of users and build on knowledge of user interaction with the system. The study of user preferences in Flickr data shows that users notion of interestingness varies significantly from statistical relevance of the content as computed from accompanying text. This makes it all the more important to explore media properties such as aesthetics.

Acknowledgment

The authors would like to thank all the participants in the user studies and Prof. Shih-Cheng Yen of the Department of Electrical Engineering, National University of Singapore for his valuable insights.

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] Katti Harish, Kwok Yang Bin, Chua Tat-Seng, Kankanhalli Mohan, "Pre-attentive discrimination of interestingness in images", International conference on Multimedia and Expo, ICME, 2008
[4] L. Fei-Fei, Natural Scene Categorization, from humans to computers, Scene Understanding Symposium, 2007
[5] Peterson Bryan, Learning to See Creatively: Design, Color & Composition in Photography, Amphoto Books, 2003
[6] R. Datta, D. Joshi, J. Li, and J. Z. Wang, Studying Aesthetics in Photographic Images Using a Computational Approach, *Proc. ECCV*, Graz, Austria, 2006
[7] Torralba A., Oliva A., Statistics of natural image categories, Network: Computation in Neural Systems, Vol. 14, 391-412. 2003.
[8] 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.
[9] 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
[10] Cohen-Or Daniel, Sorkine Olga, Gal Ran, Leyvand Tommer, Qing Xu Ying, Color Harmonization, SIGGRAPH 2006
[11] Xiaodi Hou, Zhang Liquing, Colour conceptualization, Proceedings of the fifteenth ACM international conference on Multimedia (ACM MM), 2007



Harish Katti received the B.Engg. degree in Computer science and Engineering from Karnatak University and M.Tech degree in Bio-Medical Engineering from IIT Bombay. He is a PhD candidate in the Department of Computer Science, School of Computing, National University of Singapore. He has worked the area of multimedia systems in Sasken Communications Pvt Ltd and Emuzed India Pvt Ltd. His current research interests are in multimedia information systems (content processing, multimedia semantics).

Kwok Yang Bin received the B.Engg. degree in Electrical Engineering from Department of Electrical Engineering, National University of Singapore. He has recently co-founded ZopIM LLP (http://zopim.com), Singapore, and is currently exploring computational linguistics.

Tat-Seng Chua received the PhD degree from the University of Leeds, UK. He is a professor in the School of Computing, National University of Singapore. He was the Acting and Founding Dean of the School of Computing from 1998-2000. He spent three years as a research staff member at the Institute of Systems Science (now I2R) in late 1980s. Dr Chua's main research interest is in multimedia information processing, in particular, on indexing, retrieval, and extraction of information in video and text. His current projects include: news video retrieval, question answering, video QA, and information extraction on the Web. Dr Chua is active in the international research community. He has organized and served as program committee member of numerous international conferences in the areas of computer graphics and multimedia, including ACM Multimedia, ACM SIGIR, etc. He is the co-chair of ACM Multimedia 2005 and ACM SIGIR 2008. He serves in the editorial boards of the IEEE Transactions of Multimedia, The Visual Computer, and Multimedia Tools and Applications.

Mohan S. Kankanhalli is a Professor at the Department of Computer Science of the School of Computing at the National University of Singapore. He obtained his BTech (Electrical Engineering) from the Indian Institute of Technology, Kharagpur, and his MS and PhD (Computer and Systems Engineering) from the Rensselaer Polytechnic Institute. He has worked at the Institute of Systems Science in Singapore and at the Department of Electrical Engineering of the Indian Institute of Science, Bangalore. His current research interests are in Multimedia Signal Processing (sensing, content analysis, retrieval) and Multimedia Security (surveillance, digital rights management and forensics). He is on the editorial board of several journals including the IEEE Transactions on Multimedia and the IEEE Transactions on Information Forensics and Security.