The Interestingness of Images

Michael Gygli, Helmut Grabner, Hayko Riemenschneider, Fabian Nater, Luc Van Gool

(ICCV), 2013

Cemil ZALLUHOĞLU

Outline

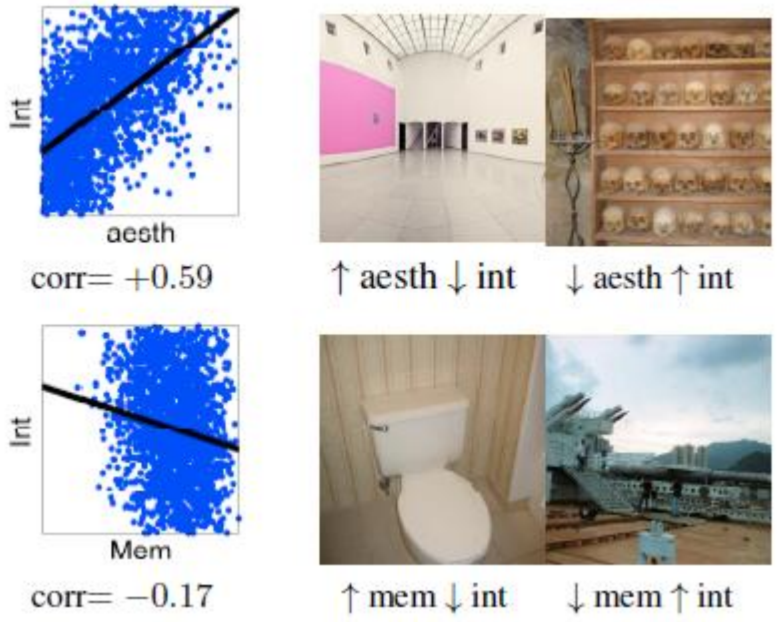
Introduction

Problem Statements

- What makes an image interesting?
- Can we build a model to predict it?

According to psychological experiments

- **Interestingness** related to **aesthetic** and **memorability**



Outline

1.Introduction
2.Related Works
3.Algorithm
4.Experiments
5.Conclusion

Related Works

Berlyne(1960)

- Interest is influenced by
  - Novelty
  - Conflict
  - Uncertainty
  - Complexity

Biederman and Vessel(2006)

- Model based on perceptual pleasure
  - Novel
  - Comprehensive
  - Natural scenes rather than man made

Figure 6. Scenes used in fMRI experiments were independently rated by subjects as "highly preferred" (*top row*) or "not preferred" (*bottom row*).

Methods

Keeping work with psychology

Decide three groups which has a high influence

- novelty/unusualness (attributes: unusual, is strange, mysterious)
- aesthetics (attributes: is aesthetic, pleasant, expert photography)
- general preferences for certain scene types (attributes: outdoor-natural vs. indoor and enclosed spaces).

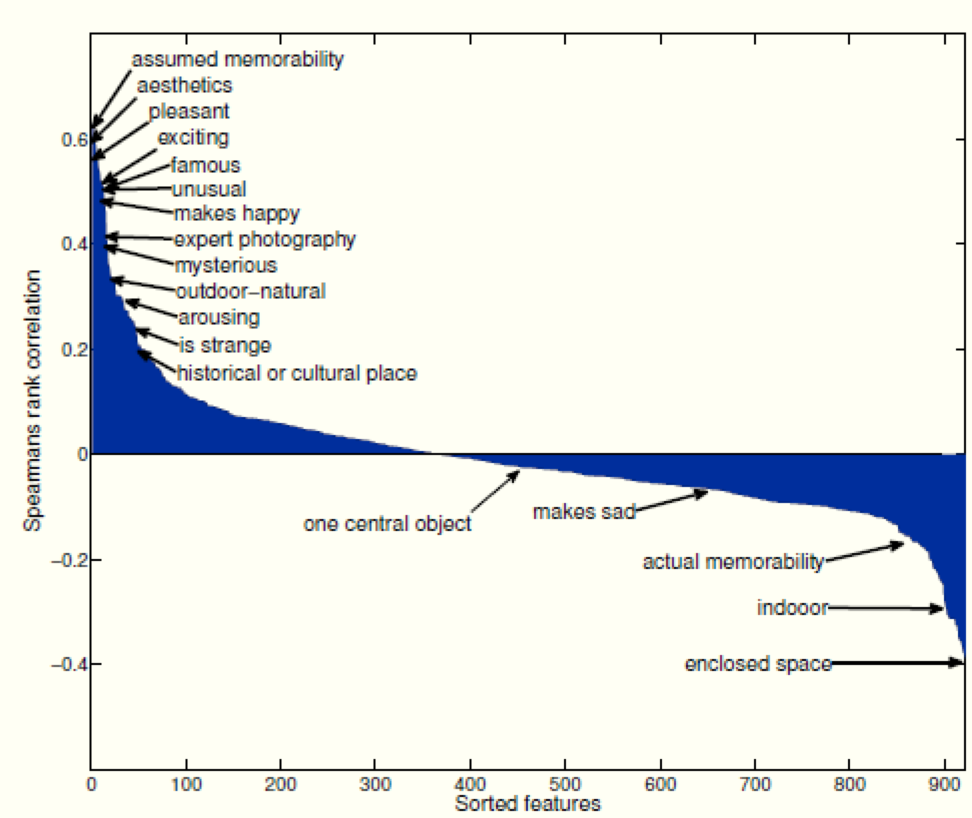
Aim: computationally predict interestingness based on the above cues

Outline

Algorithm

Propose features that computationally capture the aspects/cues of interestingness which we found most important and are implementable:

- unusualness
- aesthetics
- general preferences.



1)Unusualness

Single image from **arbitrary** scene
Proposed two methods

1.Global Outliers:

- Use Local Outlier Factor (LOF) algorithm to global image descriptors to detect global outliers in the dataset.
- Outlier factor is calculated wrt. the density of its closest cluster

- All experiments use 10-distance neighbourhood and as features

i.The raw RGB pixel values = s_{pixel}
ii.GIST= s_{gist}
iii.Spatial Pyramids on SIFT histograms= s_{pyr}

2. Composition of Parts
   1. Model the image as a graph with superpixels as nodes

$$E(\mathbf{L}) = \sum_{i \in \mathcal{S}} D_i(l_i) + \lambda \sum_{\{i,j\} \in \mathcal{N}} V(l_i, l_j)$$

◦S: the set of Superpixels
◦N: the set of superpixel neighbours
◦$D_i(l_i)$: the unary cost of assigning label *l* to the superpixel i.
◦$V_i(l_i,l_J)$:the cost of two neighboring nodes taking labels l_i and l_j
◦λ: 0.02

$$s_{compose}^{unusual} := E(\mathbf{L})/|\mathcal{S}|$$

2)Aesthetics

- Use content preferences
  - The presence of people
  - The presence of Animals
  - The preference for certain types

- Focus on capturing visually pleasing images, without semantic interpretation

○**Colorfulness**: $s^{aesth}_{colorful} := -\mathrm{EMD}(H_I, H_{uni})$

○**Arousal:** Extracted emotion scores from raw pixels.

$$s^{aesth}_{arousal} := \sum_p -0.31 \text{ brightness}(p) + 0.60 \text{ saturation}(p)$$

○**Complexity:** compare its size after JPEG compression against its uncompressed size.

$$s^{aesth}_{complex} := \frac{\text{bytes}(\text{compress}(I))}{\text{bytes}(I)}$$

○**Contrast:** $s^{aesth}_{contrast}$

○**Edge Distribution:** $s^{aesth}_{edges} := 1 - w_x w_y$ w_x and w_y being the box's normalized width and height.

3)General Preferences

- Certain scene types
- Propose to learn such fetures from global image descriptors.
- Train a Support Vector Regressor on following features
    - Raw RGB-pixels
    - GIST
    - Spatial Pyramids of SIFT histograms
    - Color histograms

4)Combination

- The scores obtained from the respective features are
  - First normalized with respect to their mean and variance.
  - Second, they are mapped into the interval [0; 1] using a sigmoid functior $\bar{s} = \frac{\exp(\mu s)}{1+\exp(\mu s)}$

- Simple linear combination $\bar{s}_{comb} = \mathbf{w}^{\mathrm{T}} \bar{s}_{sel}$

- Also applied whitening to deccorelate the features

$$\bar{s}_{decorr} = \Sigma^{-1/2} \bar{s}$$

Outline

1.Introduction
2.Related Works
3.Algorithm
4.Experiments
5.Conclusion

Experiments

- Parameter Selection:
  - Features based on raw pixels, used downscaled images 32x32 pixels.
  - For each data set use training/validation/test split.
  - For general preferences, trained v-SVR on the training set

- Evaluation:
  - Use multiple measures to evaluate feature performance
    - Recall-Precision(RP)
    - Average Precision(AP)
    - Spearman's correlation(ρ)
    - Top_N Score:

$$Top_N := \frac{\sum_{i \in P_N} s_i^*}{\sum_{i \in S_N} s_i^*}$$

Strong Context: Webcam dataset

❑This dataset consists of 20 different webcam streams, with 159 images each. It is annotated with interestingness ground truth, acquired in a psychological study

❑Mean interestingness score of 0,15

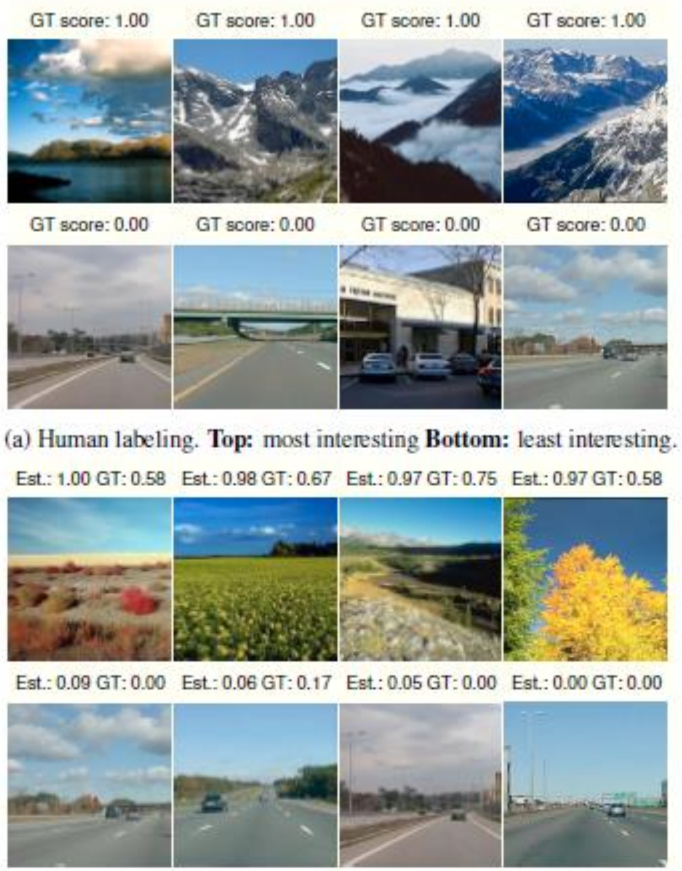
❖use different thresholds for RP calculation: s > 0:5 as positive s < 0:25 as negative samples

(a) Human labeling. **Top:** most interesting **Bottom:** least interesting.

(c) Predicted interestingness.
Top: most interesting **Bottom:** least interesting.

Weak Context: Scene Categories Dataset

- The 8 scene categories dataset of Oliva and Torralba
- consists of 2’688 images with a fixed size of 256x256 pixels.
- The images are typical scenes from one of the 8 categories
- (coast, mountain, forest, open country, street,
- inside city, tall buildings and highways)



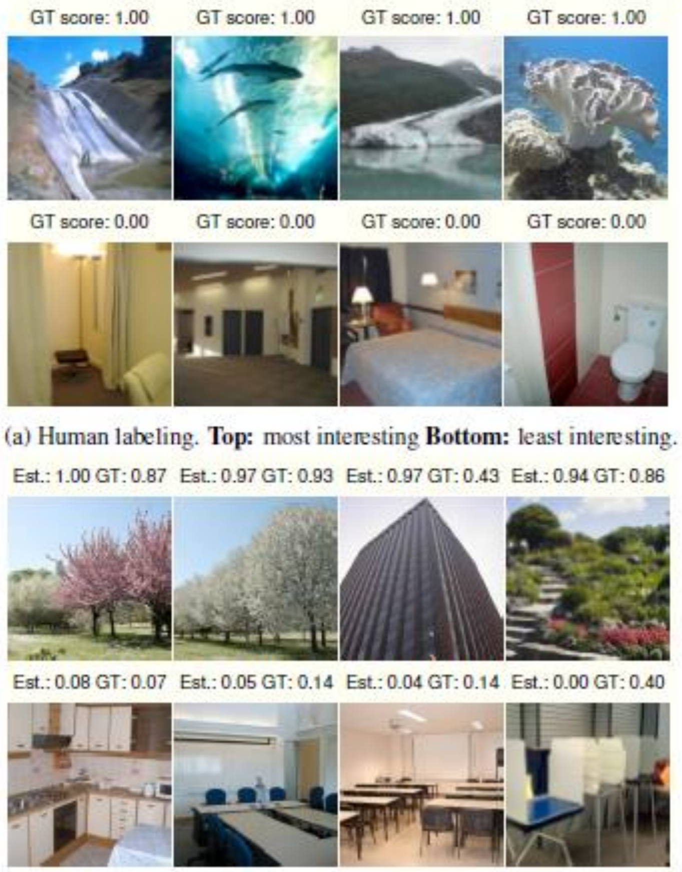
(a) Human labeling. **Top:** most interesting **Bottom:** least interesting.

(c) Predicted interestingness.
Top: most interesting **Bottom:** least interesting.

Arbitrary photos: Memorability dataset

- The memorability dataset consists of 2’222 images with
- a fixed size of 256 256 pixels.

asked a user to classify an image as interesting/non-interesting.



(c) Predicted interestingness.
Top: most interesting **Bottom:** least interesting.

Strong Context: Webcam dataset

Context	Cue	Feature	ρ	AP	Top_5
Strong Webcams [12] Static camera: 20 different outdoor sequences	Unusual	compose	**0.29**	**0.35**	<u>0.51</u>
		pixel	<u>0.23</u>	0.22	**0.53**
		pyr	0.01	0.10	0.31
		gist	0.03	0.12	0.28
	Aesthetic	arousal	0.13	0.24	0.41
		complex	0.09	<u>0.26</u>	0.48
		colorful	-0.06	0.06	0.26
		edges	-0.04	0.07	0.34
		contrast	0.10	0.15	0.41
	Pref.	pixel	0.04	0.11	0.35
		pyr	0.05	0.10	0.31
		gist	0.16	0.18	0.39
		colorhist	0.05	0.12	0.36
		combined	**0.32**	0.39	0.57
		comb. decorr.	0.31	**0.42**	**0.61**
		chance	0	0.04	0.25

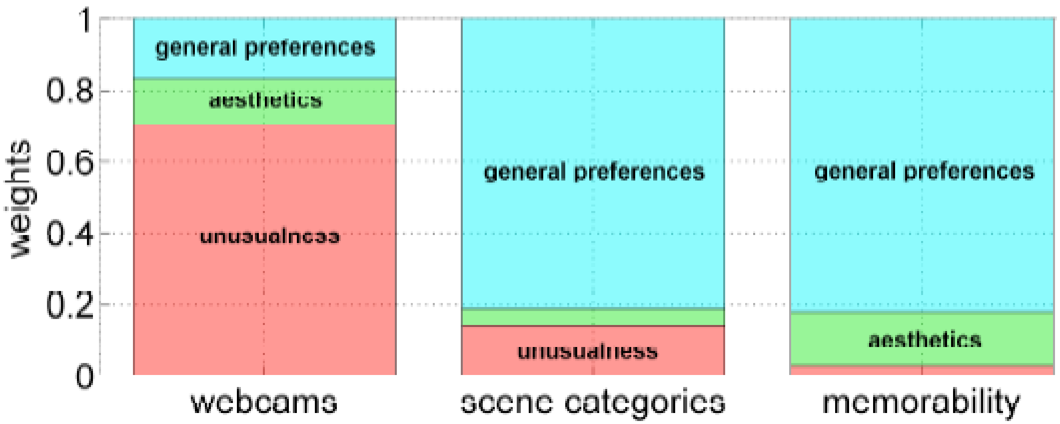
Weak Context: Scene Categories Dataset

Context	Cue	Feature	ρ	AP	Top_5
Weak Scene categories [18] 8 scenes types: coast, mountain, forest, open country, street, inside city, tall building, highway	Unusual	compose	0.18	0.28	0.38
		pixel	0.23	0.32	0.32
		pyr	0.17	0.27	0.66
		gist	0.19	0.23	0.47
	Aesthetic	arousal	0.43	0.45	0.65
		complex	0.19	0.31	0.53
		colorful	0.24	0.33	0.67
		edges	0.30	0.34	0.51
		contrast	0.19	0.34	0.62
	Pref.	pixel	0.43	0.40	0.62
		pyr	0.64	**0.78**	0.70
		gist	**0.67**	0.75	0.76
		colorhist	0.54	0.69	**0.83**
		combined	**0.71**	**0.83**	**0.68**
		comb. decorr.	0.70	**0.83**	**0.68**
		chance	0	0.26	0.48

Arbitrary photos: Memorability dataset

Context	Cue	Feature	ρ	AP	Top_5
None Memorability[14] Arbitrary photos: Indoor, Outdoor, man-made, natural, people, animals	Unusual	compose	0.10	0.35	0.46
		pixel	0.01	0.31	0.65
		pyr	-0.11	0.29	0.60
		gist	-0.01	0.30	0.45
	Aesthetic	arousal	-0.03	0.31	0.47
		complex	0.27	0.42	0.63
		colorful	0.03	0.34	0.61
		edges	0.11	0.42	0.55
		contrast	0.05	0.33	0.67
	Pref.	pixel	0.25	0.51	0.67
		pyr	0.52	0.66	**0.78**
		gist	**0.58**	**0.69**	0.77
		colorhist	0.33	0.55	0.64
		combined	**0.60**	0.73	**0.82**
		comb. decorr.	**0.60**	**0.77**	0.80
		chance	0	0.26	0.47

The normalized weights for the feature combinations



Outline

Conclusion

- ❑Proposed a set of features able to capture interestingness in varying contexts.
- ❑With strong context, such as for static webcams, **unusualness** is the most important cue for interestingness.
- ❑In single, context-free images, **general preferences** for certain scene types are more important
- ❑To overcome the current limitations of interestingness prediction, one would need:
  - (i) an extensive knowledge of what is known to most people,
  - (ii) algorithms able to capture unusualness at the semantic level and
  - (iii) knowledge about personal preferences of the observer.