A Social Signal Processing Perspective on Computational Aesthetics: Theories and Applications

Candidate: Cristina Segalin
Supervisor: Prof. Marco Cristani

May 31st, 2016 – XXVIII Cycle
Overview

Producer

Consumer

- Objectively beautiful
- Subjectively beautiful

Computational Aesthetics

Learn aesthetics

Human-Human Interaction

Social Signal Processing

Consumer

Like

Comment

Share

Overview 2

Producer

Consumer

- Objectively beautiful
- Subjectively beautiful

Computational Aesthetics

Learn aesthetics

Human-Human Interaction

Social Signal Processing

Overview 2
Goals and Contributions

**Goals:**
1. Enrich the computational aesthetics field, by accounting for the personal sphere
2. Consider an aesthetical preference as a social signal, that can be captured and interpret by others

**Contributions:**

Personal Aesthetics:
- A new soft biometric trait
- A new social signal for Personality Computing
  - Collection of a personality-augmented image dataset of aesthetical preferences: *PsychoFlickr*
Outline

**Part I:**
- Personal Aesthetics for Soft Biometrics
  - Computational Aesthetics (features)
  - A new soft biometric trait
    - Lasso Regression approach
    - Multiresolution Regression approach

**Part II:**
- Personal Aesthetics for Personality Computing
  - Social Signal Processing
  - A new social signal for Personality Computing
    - Multiple Instance Regression approach
    - Convolutional Neural Network

**Conclusions and Future Works**
Computational Aesthetics

**Definition**

**Computational Aesthetics:** aims at developing «computational methods that can make applicable aesthetic decisions in a similar fashion as humans can»

- Aesthetics: study of beauty and taste [Honeig ’05]
- Computational Aesthetics focuses mostly on capturing and modeling a *general and shared* sense of beauty... [Adams ‘03]
- ... automatically select high aesthetic quality images from large image collections [Dhar et al. ‘11]
- ... but the beauty is in the *eye of the beholder!* [Beiderman & Vessel ‘03]

→ **personal aesthetics!**
State of the art

- **Aesthetics - Psychology**
  - Applied mainly to paintings [Rawlings ‘98]
  - It studies how individual preferences can be modeled, depending on the subject at hand [Furham ‘01]
  - Guidelines of appreciation

- **Computational Aesthetics - Computer Vision**
  - Applied to digital images
  - It aims at finding a general sense of beauty
  - Many applications: prediction of aesthetic score, recommender systems, features engineering
  - It studies features which capture perceptual and content based features [Biederman ’06, Furham ’01, Girshick ‘10]
  - Recently, it exploit the wisdom of crowds for learning common preferences (Flickr, etc) [Bacukhage’13, Dhar ’11, Murray ‘12]
## Computational Aesthetics Features:

### Low level representation

<table>
<thead>
<tr>
<th>Category</th>
<th>Name</th>
<th>L</th>
<th>Short Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td>Use of light</td>
<td>1</td>
<td>Average pixel intensity of V channel</td>
</tr>
<tr>
<td></td>
<td>HSV statistics</td>
<td>3</td>
<td>Mean of S channel and standard deviation of S, V channels</td>
</tr>
<tr>
<td></td>
<td>Emotion-based</td>
<td>3</td>
<td>Amount of <em>Pleasure, Arousal, Dominance</em></td>
</tr>
<tr>
<td></td>
<td>Circular Variance</td>
<td>1</td>
<td><em>Circular variance</em> of the H channel in the IHLS color space</td>
</tr>
<tr>
<td></td>
<td>Colorfulness</td>
<td>1</td>
<td>Colorfulness measure based on Earth Mover’s Distance (EMD)</td>
</tr>
<tr>
<td></td>
<td>Color Name</td>
<td>11</td>
<td>Amount of <em>Black, Blue, Brown, Green, Gray, Orange, Pink, Purple, Red, White, Yellow</em></td>
</tr>
<tr>
<td>Composition</td>
<td>Edges</td>
<td>1</td>
<td>Total number of edge points, extracted with Canny</td>
</tr>
<tr>
<td></td>
<td>Level of detail</td>
<td>1</td>
<td>Number of regions (after mean shift segmentation)</td>
</tr>
<tr>
<td></td>
<td>Regions</td>
<td>1</td>
<td>Average size of the regions (after mean shift segmentation)</td>
</tr>
<tr>
<td></td>
<td>Low depth of field (DOF)</td>
<td>3</td>
<td>Amount of focus sharpness in the inner part of the image w.r.t. the overall focus</td>
</tr>
<tr>
<td></td>
<td>Rule of thirds</td>
<td>2</td>
<td>Mean of S,V channels in the inner rectangle of the image</td>
</tr>
<tr>
<td></td>
<td>Image parameters</td>
<td>1</td>
<td>Size of the image</td>
</tr>
<tr>
<td>Texture</td>
<td>Entropy</td>
<td>1</td>
<td>Image entropy</td>
</tr>
<tr>
<td></td>
<td>Wavelet textures</td>
<td>12</td>
<td>Level of spatial graininess measured with a three-level (L1,L2,L3) Daubechies wavelet transform on the HSV channels</td>
</tr>
<tr>
<td></td>
<td>Tamura</td>
<td>3</td>
<td>Amount of <em>Coarseness, Contrast, Directionality</em></td>
</tr>
<tr>
<td></td>
<td>GLCM-features</td>
<td>12</td>
<td>Amount of <em>Contrast, Correlation, Energy, Homogeneity</em> for each HSV channel</td>
</tr>
<tr>
<td>Content</td>
<td>Objects</td>
<td>28</td>
<td>Objects detectors : in particular, here are the objects for which detectors are available: <em>people, plane, bike, bird, boat, bottle, bus, car, cat, dog, table, horse, motorbike, chair</em>. In all the cases we kept the number of instances and their average bounding box size</td>
</tr>
<tr>
<td></td>
<td>Faces</td>
<td>2</td>
<td>Number and size of faces after Viola-Jones face detection algorithm</td>
</tr>
</tbody>
</table>
Personal Aesthetics for Soft Biometrics
Soft Biometrics

**Soft Biometrics**: physical, behavioral or HCI human characteristics, classifiable in pre-defined human compliant categories, for establishing the identity of an individual.

Traits which accept this definition include, but are not limited to:

- **Physical**: skin color, eye color, hair color, presence of beard, presence of moustache, height, weight.
- **Behavioral**: gait, postures, gestures.
- **Adhered human characteristics**: clothes color, tattoos, accessories.
- **HCI-based**: use of Internet applications, chatting, browsing histories, mouse dynamics  \[\rightarrow personal aesthetics^{new}\]
A New Soft Biometric Trait based on Personal Aesthetics

MIR approach

Goal: discriminate a single user from all the other ones

Personal aesthetics is exploited into a biometric recognition/authentication system:

- Enrollment stage: the “preference model” of a user is learned from a set of preferred images
- Verification/recognition phase, the user model is tested with an unseen set of favorites preferred by a probe subject.

\[
X(i) = \{x_1(i), x_2(i), x_3(i), \ldots, x_N(i)\}
\]

\[
Y = \begin{cases} 
  +1 & \text{when the image is preferred by the user} \\
  -1 & \text{when the image is not preferred by the user} 
\end{cases}
\]
A New Soft Biometric Trait based on Personal Aesthetics
MIR approach – Examples of users’ preferred pictures
A New Soft Biometric Trait based on Personal Aesthetics

MIR approach - Enrollment

Learning of the loading is performed by LASSO regression as a binary problem on all the training set

\[ y_n = w^{(u)T} x_n \]

the feature vector

the coefficients vector

where

\[ E(w^{(u)}) = \sum_{n=1}^{N_{TR}} (y_n - w^{(u)} x_n)^2 \]

subject to the sparsity constraint

\[ \sum_{d=1}^{D} |w_d| \leq t \]
A New Soft Biometric Trait based on Personal Aesthetics
MIR approach - Result of enrollment
A New Soft Biometric Trait based on Personal Aesthetics

MIR approach - Matching score

- Given an image $n$ of the unknown subject $v$, the goal is to evaluate how probably it is preferred by a subject $u$ (to check if $u$ and $v$ do match!)

- Intuitively, the expressivity of a single image is limited, so multiple test images ($N_{TE}$) belonging to the same subject are taken into account, and the final matching score is

$$\beta^{(u,v)}_n = w^{(u)^T}x^{(v)}_n$$

$$\beta^{(u,v)} = \frac{1}{N_{TE}} \sum_{n=1}^{N_{TE}} \beta^{(u,v)}_n$$
A New Soft Biometric Trait based on Personal Aesthetics

MIR approach - Results of Authentication (Verification)
A New Soft Biometric Trait based on Personal Aesthetics

MIR approach - Results of Recognition (Identification)

![Graph showing recognition rates and cumulative distribution functions (CDFs) for different probe image protocols. The graph compares 1 probe image (nAUC = 0.760 ± 0.001), 5 probe images (nAUC = 0.898 ± 0.001), 20 probe images (nAUC = 0.960 ± 0.001), and 100 probe images (nAUC = 0.983 ± 0.001).]

<table>
<thead>
<tr>
<th>Protocol</th>
<th>rank 1</th>
<th>rank 5</th>
<th>rank 20</th>
<th>rank 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 probe image</td>
<td>0.063</td>
<td>0.188</td>
<td>0.408</td>
<td>0.829</td>
</tr>
<tr>
<td>5 probe images</td>
<td>0.143</td>
<td>0.399</td>
<td>0.688</td>
<td>0.966</td>
</tr>
<tr>
<td>20 probe images</td>
<td>0.254</td>
<td>0.629</td>
<td>0.883</td>
<td>0.998</td>
</tr>
<tr>
<td>100 probe images</td>
<td>0.359</td>
<td>0.796</td>
<td>0.970</td>
<td>0.999</td>
</tr>
</tbody>
</table>
Mapping Image Preferences on a Counting Grid
Organizing and making sense of Bag of Words

- Generative model for feature extraction and information visualization [Perina & Jojic CVPR 2011, under patenting]
- An image is modeled as an histogram of features
Mapping Image Preferences on a Counting Grid
Embedding map of user aesthetic preferences

- Inferences have been used to extract where the images of a user are, creating user maps

\[ \gamma(i) = \frac{\sum_{t} \sum_{k|i \in W_k} q_k^t \cdot y^t}{\sum_{t} \sum_{k|i \in W_k} q_k^t} \]

- People’s preferences are clusters in the grid!
Mapping Image Preferences on a Counting Grid
Embedding map of user aesthetic preferences
Statistical Generative Multiresolution Approach

MRCG approach - Initialization step

Training images → Bags of features → Multi-resolution counting grid
Statistical Generative Multiresolution Approach

MRCG approach – Visualization of multiresolution CGs
Statistical Generative Multiresolution Approach

MRCG approach - Enrollment

\[ \gamma_u^{(r)} = \sum_{t \in T_u} p(k^t | \{c_z^t\}, \pi^{(r)}) \]

Gallery user $x_u$

Embedding maps

Multiple SVM training
Statistical Generative Multiresolution Approach
MRCG approach - Visualization of embedding maps
Statistical Generative Multiresolution Approach

MRCG approach - Identification/Verification

Probe user $\chi_v$

Embedding maps

Multiple SVM classification

$\pi^{(1)}, \pi^{(2)}, \ldots, \pi^{(R)}$

$\gamma_u^{(1)}, \gamma_u^{(2)}, \ldots, \gamma_u^{(R)}$

$\lambda_u^{(1)}, \lambda_u^{(2)}, \ldots, \lambda_u^{(R)}$

$C_{u,v}^{(1)}, C_{u,v}^{(2)}, \ldots, C_{u,v}^{(R)}, C_{u,v}$
Statistical Generative Multiresolution Approach

MRCG approach – Identification/Verification results

Statistical Generative Multiresolution Approach

MRCG approach - Feature Analysis

Statistical Generative Multiresolution Approach

MRCG approach - Limitations

- Robustness of features
- Missed disliked aesthetics
- Lack of distinctiveness
- Variability of aesthetics
- Number of images

Feature Learning

Future work

Personality Computing
Computing Personality from Aesthetic Preferences
**Statement: Social Signal Processing**

**Motivation and definition**

**Social Signal Processing**: aims at understanding «social signals» i.e.,

- implicit, unconscious cues that communicate something, outside the verbal content
- that depend on implicit cognitive processes [Vinciarelli et al. 2009]

<table>
<thead>
<tr>
<th>Social cues</th>
<th>Example social behaviours</th>
<th>Tech.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Emotion</td>
<td>Personality</td>
</tr>
<tr>
<td>Physical appearance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Attractiveness</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Body shape</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Gesture and posture</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Hand gestures</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Posture</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Walking</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Face and eyes behaviour</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Facial expressions</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Gaze behaviour</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Focus of attention</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Vocal behaviour</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Prosody</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Turn taking</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Vocal outbursts</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Silence</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Space and environment</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Distance</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Seating arrangement</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

[Social Signal Processing Survey of an Emerging Domain. Vinciarelli, Pantic, Bourlard, JIVC]
The Big Five

- **Extraversion**: Active, Assertive, Energetic, Outgoing
- **Agreeableness**: Appreciative, Forgiving, Generous, Kind, Sympathetic, Trusting
- **Conscientiousness**: Efficient, Organized, Reliable, Responsible, Thorough
- **Neuroticism**: Anxious, Self-pitying, Tense, Touchy, Unstable, Worrying
- **Openness**: Artistic, Curious, Imaginative, Insightful

State of the art

- **Implicit cognitive processes:**
  - Study the perception of profile pictures on social media [Fitzgerald ‘09]
  - Perception of profile from all elements that can appear in an online profile [Gosling ‘08]
  - Prediction of favorite images [Lovato ‘12]
  - Emotions through the characteristics of paintings [Sebe ‘08]

- **Main result:**
  - Confirm the action of the implicit cognitive processes when using multimedia data
  - Identification of correlations between aesthetic preferences and personality
State of the art

- **Aesthetics – Psychology**
  - Big Five personality traits are taken as individual characterization [Rammstedt ‘07]
  - High Openness correlates to liking pictures with few elements [Rawling ‘98]
  - High Openness correlates to liking pictures with "complex" and "meaningful" polygons [Rawling ‘98]
  - Extrovert people prefer humanized landscape [Abello ‘86]
The Brunswick Lens Model
Personality from aesthetic preferences

MIR approach – Dataset Collection

- 300 Flickr professional users (the data producer)
- For each user:
  - Take 200 random faved images, from which we extract
    - 15 computational aesthetic features [Datta ‘06, Machajdick’10]
    - 14 objects, scenes [Felzenszwalb et al. ‘10, Oliva et al. ‘01]
  - Let him fill a personality questionnaire (the Big Five Inventory 10)
    - It gives 5 scores (-4...4) for the personality traits of Openness, Consciousness, Extraversion, Agreeableness, Neuroticism

Personality from aesthetic preferences
MIR approach– Dataset collection

- 12 assessors (the data consumer)
- For each assessor:
  - Check the 200 favorites of each user
  - For each user, fill the BFI questionnaire, inferring how the user could be!
  - We check homogeneity in the scores and the average of the test was computed.

→ Perceived State
Personality from aesthetic preferences

MIR approach - Goals

- At this point we want to:
  1. Measure correlations state/perc. state, state/feats, perc. state/feats
  2. Given the faved pictures of a user
     - infer the state
     - infer the perceived state
Corr. state/perceived state $\leq 0.26$
Personality from aesthetic preferences
MIR approach—Visualization of correlations

**Extraversion**

(perceived) High

(perceived) Low

0.12 #People
0.12 Size People
0.16 # Faces

0.52
0.40
0.46
Personality from aesthetic preferences

MIR approach – Regression approach

- It is an interesting problem:
  - We have multiple images connected with only a user
  - Not all of them are equally important
  - Seems suitable the Multiple Instance Learning paradigm
  - So far, simple regression on a low dimensional image representation has been performed [Babenko ‘08]
Personality from aesthetic preferences

MIR approach - Prediction results

Attributed Traits

Self-Assessed Traits

Correlation Actual-Predicted Scores

Naive MIR  cit-kNN  Clust-Reg  Topic-Sum  Gen-MoG  Gen-LDA  CG

Ope  Con  Ext  Agr  Neu

49
## Discussion

### Approach Weaknesses and Breakthrough with CNN

**Hand-crafted features:**

- Computational Aesthetics hand-crafted features focus on designing features explaining how a particular image has been captured, discarding the content of the images.
- Standard object recognition and feature extraction techniques might not be sufficient to capture significant dependences between pictures and personality traits.
- Non linear relation between image features - user likes.
- Time consuming.
- No disentangle of all explanatory factors of data.

**Deep/Feature Learning:**

- Automatic discovery of new high level representation.
- Non-linear transformation of data.
- Disantangle factors of observed data in multiple level of abstractions.
- Distributed representation.
Feature Learning via Convolutional Neural Networks

[Y. Jia et al., Caffe: Convolutional architecture for fast feature embedding.]
Social Profiling through Image Understanding:
CNN approach

Personality traits

Fine-tuning → DATASET [PsychoFlickr]

Pre-training → DATASET [Deng’09]

ARCHITECTURE
CaffeNet

Model

SOFTWARE
Caffe
[Jia’14]
Social Profiling through Image Understanding:
CNN approach – fine-tuning process

![Diagram of CNN layers and pre-trained model](image-url)
### Social Profiling through Image Understanding:
**CNN approach - Classification results**

<table>
<thead>
<tr>
<th></th>
<th>SELF</th>
<th></th>
<th>ATTRIBUTED</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TRAIN</td>
<td>TEST</td>
<td>TRAIN</td>
<td>TEST</td>
</tr>
<tr>
<td>O</td>
<td>0.57</td>
<td>0.53</td>
<td>0.73</td>
<td>0.61</td>
</tr>
<tr>
<td>C</td>
<td>0.59</td>
<td>0.54</td>
<td>0.81</td>
<td>0.66</td>
</tr>
<tr>
<td>E</td>
<td>0.60</td>
<td>0.54</td>
<td>0.76</td>
<td>0.64</td>
</tr>
<tr>
<td>A</td>
<td>0.57</td>
<td>0.54</td>
<td>0.76</td>
<td>0.64</td>
</tr>
<tr>
<td>N</td>
<td>0.55</td>
<td>0.52</td>
<td>0.81</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Accuracy on training and testing set

<table>
<thead>
<tr>
<th></th>
<th>SELF</th>
<th></th>
<th>ATTRIBUTED</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ieeeac</td>
<td>our</td>
<td>Ieeeac</td>
<td>our</td>
</tr>
<tr>
<td>O</td>
<td>0.49</td>
<td>0.53</td>
<td>0.59</td>
<td>0.62</td>
</tr>
<tr>
<td>C</td>
<td>0.50</td>
<td>0.54</td>
<td>0.51</td>
<td>0.66</td>
</tr>
<tr>
<td>E</td>
<td>0.52</td>
<td>0.54</td>
<td>0.64</td>
<td>0.65</td>
</tr>
<tr>
<td>A</td>
<td>0.47</td>
<td>0.54</td>
<td>0.56</td>
<td>0.64</td>
</tr>
<tr>
<td>N</td>
<td>0.51</td>
<td>0.53</td>
<td>0.75</td>
<td>0.69</td>
</tr>
</tbody>
</table>

Comparison with previous approach

---

Social Profiling through Image Understanding:
CNN approach - Learned Attributes

Extraversion

(perceived) Low          (perceived) High
Social Profiling through Image Understanding:
CNN approach - Learned Attributes

Neuroticism

(perceived) Low

Agreeableness

(perceived) High
Social Profiling through Image Understanding:
CNN approach - Learned Attributes

Conscientiousness

(perceived) Low

Openness

(perceived) High
Social Profiling through Image Understanding:
Demo: http://psychoflickr.di.univr.it:8000/demo/
Social Profiling through Image Understanding:
Demo: http://psychoflickr.di.univr.it:8000/demo/

VIPS PsycoFlickr Personality Demo

Predictions:
- Openness: Low
- Conscientiousness: Low
- Extraversion: High
- Agreeableness: Low
- Neuroticism: High
Conclusions

1. **Key Idea**: Multimedia content as social signal
   - Multimedia role has changed with Social Networks and Big Data
   - Images associated to people are spread through social platforms
   - Image not just as a message, a social signal: personal aesthetics preferences, that reveal something about its author
   - Implicit cognitive processes of image appreciation are unique and personal and can capture/produce this social signal to understand the state of a person

2. **Soft biometric trait**
   - We considered images tagged as favorites by a person
   - Proposed Computational Aesthetic features capturing the aesthetic of images
   - Built Hybrid model personal aesthetic preferences for Re-identification

3. **Personality Computing**
   - PyshcoFlickr dataset
   - Multiple Instance Regression and Deep Learning approaches to map aesthetic preferences into personality traits, to model producer and consumer state
   - Demo application
Future Perspective

1. **Social network**: Infer demographic information from aesthetics preferences
2. **Marketing**: Evaluation of impact of set of images for hypothetical customers
3. **Viral Marketing**: Implicit cognitive processes might contribute to explain and enhance virality
4. **Big Data**: Making sense of large amount of data
5. **Virtual agents**: Improve machines that exhibit human-like features and behavior like robots, animated characters, embodied conversation agents
6. **Neuroscience/Phototherapy**: Understand the role/activation of neurons when appreciating images
7. **Online Games**: Customization using user preferences and profile
PhD Publications

- **Journal**

- **Conference**
What’s next?

February-April 2016
Internship

August 2016-2017
Postdoc

Disney
Research, Pittsburgh

Caltech
Thanks to all the VIPS people and ... 

While writing the thesis ... 

... to you for your attention!!