University of Dayton
eCommons

Computer Science Faculty Publications | Department of Computer Science

10-2015

Sense Beyond Expressions: Cuteness

Kang Wang
Rensselaer Polytechnic Institute

Tam Nguyen
University of Dayton, tnguyen1@udayton.edu

Jiashi Feng
University of California - Berkeley

Jose Sepulveda
National University of Singapore

Follow this and additional works at: http://ecommons.udayton.edu/cps_fac_pub

 Part of the Graphics and Human Computer Interfaces Commons, and the Other Computer Sciences Commons

eCommons Citation

Wang, Kang; Nguyen, Tam; Feng, Jiashi; and Sepulveda, Jose, "Sense Beyond Expressions: Cuteness" (2015). *Computer Science Faculty Publications.* Paper 71.
http://ecommons.udayton.edu/cps_fac_pub/71

This Conference Paper is brought to you for free and open access by the Department of Computer Science at eCommons. It has been accepted for inclusion in Computer Science Faculty Publications by an authorized administrator of eCommons. For more information, please contact frice1@udayton.edu, mschlangen1@udayton.edu.

Sense Beyond Expressions: Cuteness

Kang Wang*
Rensselaer Polytechnic Institute
wangk10@rpi.edu

Tam V. Nguyen*
Singapore Polytechnic
nguyen_van_tam@sp.edu.sg

Jiashi Feng
University of California, Berkeley
jshfeng@gmail.com

Jose Sepulveda
Singapore Polytechnic
sepulveda_jose@sp.edu.sg

ABSTRACT

With the development of Internet culture, *cute* has become a popular concept. Many people are curious about what factors making a person look cute. However, there is rare research to answer this interesting question. In this work, we construct a dataset of personal images with comprehensively annotated cuteness scores and facial attributes to investigate this high-level concept in depth. Based on this dataset, through an automatic attributes mining process, we find several critical attributes determining the cuteness of a person. We also develop a novel Continuous Latent Support Vector Machine (C-LSVM) method to predict the cuteness score of one person given only his image. Extensive evaluations validate the effectiveness of the proposed method for cuteness prediction.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Algorithms, Experimentation

Keywords

Cuteness, Attribute, Continuous Latent SVM

1. INTRODUCTION

Cuteness describes a type of attractiveness commonly associated with youth and appearance, which activates in others the motivation to care [6]. Recent studies suggest that cute images stimulate the pleasure centers of the brain which is closely related with the positive emotion of human [3]. This explains why everybody prefers *cute* persons or stuff in social network, shopping, browsing images/videos on the web and so on. For example, some survey shows that women's fashions opt for the cute even over the sensible or glamorous[1]. This makes cuteness be a quite important factor to consider in product design, advertisement and so on.

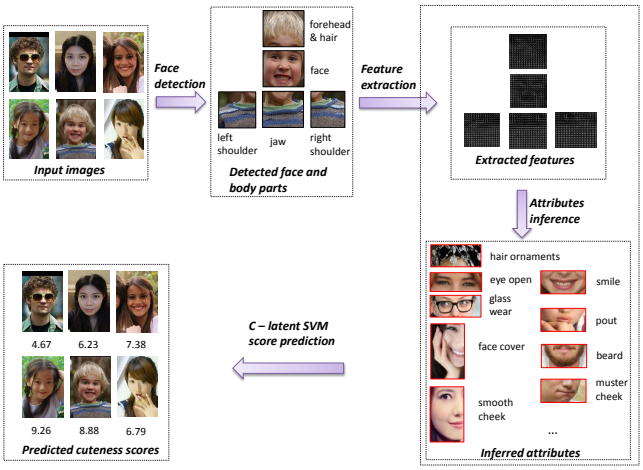


Figure 1: The framework of cuteness prediction. For each input image, we first detect the bounding box of the human faces and neighboring body parts, and then extract the appearance features accordingly. Based on the appearance features, we infer the underlying middle-level attributes and the cuteness score simultaneously.

Cuteness has received the attention of psychologists and neuroscientists for several years. For example, Kim *et al.* [4] conducted studies on why humans think certain animals are cute using functional magnetic resonance imaging (fMRI) to measure changes in human brain activity. Some evidence from this work suggests the brain activity is greater when the stimulus has juvenile characteristics – a button nose, big eyes, a large wobbly head, fat cheeks, *etc.*

Though it has been investigated by psychologists and neuroscientists, cuteness has not roused the attention of computer scientists yet. In computer vision and multimedia files, there are plenty of works focusing on recognizing ordinary expressions, *i.e.*, happiness and sadness. However, cuteness, which is beyond these traditional expressions and has higher-level semantics, is far more difficult to recognize. In this work, we explore the secrets of cuteness through the application of machine learning techniques. This is the first

*indicates the co-first authorship

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
MM '15, October 26-30, 2015, Brisbane, Australia
Copyright 2015 ACM ISBN 978-1-4503-3459-4/15/10$15.00
DOI: http://dx.doi.org/10.1145/2733373.2806283 ...$15.00.

[1] http://www.nytimes.com/2006/01/03/science/03cute.html?pagewanted=all

research attempt of computational analysis on what factors determine the cuteness, a high-level concept. We construct a model that learns from human images and their respective cuteness ratings to produce human-like evaluation of cuteness. Our work is based on the underlying theory that there are objective regularities in cuteness to be analyzed and learned. And in this work, we also provide a general framework for investigating and analyzing other high-level expressions, such as "funny" and "scary".

To investigate the factors determining the cuteness of a person, we construct a large dataset of human images with comprehensively annotated attributes and cuteness scores. In this work, the attributes are defined empirically and used as the middle-level descriptors of certain characteristic of the persons. For example, some of the attributes describe the facial appearance such as *skin smoothness* and *age*. And others describe the pose or expression of the human, such as *smile* and *face cover*. Based on this dataset, we propose a novel model to automatically learn which features and attributes determine the cuteness of persons. And we train the predictors on these features and attributes for predicting the cuteness score of a new person image. In previous works, latent SVM is a widely adopted method for the attributes mining and prediction [5, 10]. However, for the cuteness prediction problem, the annotations of the samples (cuteness scores) are continuous values. Thus, the traditional latent SVM, which can only handle discrete annotations, cannot be applied here. In this work, we propose a novel Continuous Latent SVM (C-LSVM) method, which can handle the continuous labels of the samples, to solve this issue. And we show that C-LSVM is a more general method than standard Latent SVM and has a great potential for solving many other problems involving predicting continuous variables.

Studying what makes person looks cute and how to predict cuteness from person images alone may have many useful applications. These applications include choosing a collection of cute clips from a video to generate the attractive video summarization, organizing photos in an album according to the cuteness, automatically retrieve the cute images in a large image set. Cuteness prediction and generation also benefit greatly the advertisement and production design.

2. DATASET CONSTRUCTION

Since none datasets exist for the cuteness research, in order to investigate this problem well, we collect a new dataset of personal images by ourselves. The images are crawled from the web and $4,800$ images are collected in total. Most of these images contain the frontal face of the persons.

We invite 40 subjects to participate the annotation of the images' scores and attributes. To relieve the burden of the cuteness score annotation, in this work, we adopt a k-wise comparison to estimate the rank of the images and then automatically infer their absolute scores [7, 8]. In each round of annotation, the subjects are required to rank k images in a descend order of their scores. After they finish all the annotations, we estimate the absolute score of the cuteness based a rank SVM method, which maintains the rank of the photos annotated in each k-wise annotation. For the details please refer to [7, 8]. The cuteness degree of a person heavily depends on his appearance and pose. For example, a young and pretty girl pouting her mouth will look quite cute. In this work, we define following 19 attributes to comprehensively describe the appearance and pose of a person in a middle semantic level. The adopted attributes include Gender (Male, Female), Age (Young, Teen, Middle, Old), Eye (Open, Close), Mouth Variation, Mouth (Open, Close), Teeth (Visible, Invisible), Smile, Wearing Glasses, Beard, Skin Color (Bright, Dark), Hair Color (Black, Blonde, Other), Hair Ornaments, Face Cover, Skin Smoothness. For the attribute of mouth variation, we consider the poses such as pout, muster cheek having the value of 1. For the attribute of face cover, it includes finger touches lips, hand holds jaw and so on. Some examples of the above attributes are shown in Figure 1 and Figure 2. Most of the attributes admit two values and are represented by a binary variable.

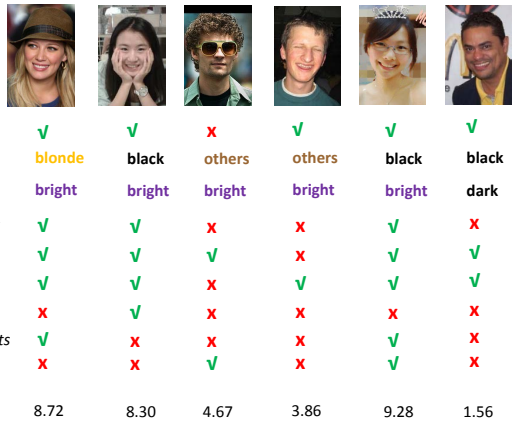


Figure 2: Examples of the collected photos with annotated attributes and cuteness scores.

3. C-LSVM FOR CUTENESS PREDICTION

3.1 Features

We first run Viola-Jones face detector [9] on the collected images to obtain bounding boxes for the human faces. Then we extract the appearance features from the face bounding box and the four spatially neighboring bounding boxes with the same size as shown in Figure 1. Here all of the bounding boxes are resized to 128×128 pixels. We extract Gabor, LBP [1], HOG [2] features within the bounding boxes to describe the pose and facial texture of persons. More specifically, for the Gabor feature, we extract the filter responses in 5 scales and 8 orientations and form a $128 \times 128 \times 5 \times 8$ feature vector for each bounding box. Then its feature dimension is reduced to 200 dimensions via PCA. Similarly, we extract HOG feature and reduce its dimensionality from $8,000$ to 200 and LBP feature from 256 to 40. We experimentally find that reducing the feature dimensionality can enhance both the prediction performance and efficiency.

3.2 Prediction Model

We propose a continuous latent support vector machine (C-LSVM) for modeling the relationship between the raw feature with the attributes and cuteness score. In particular, we use three linear models to describe the relationship between raw feature $\mathbf{x}$ and attribute $\mathbf{a}$, raw feature $\mathbf{x}$ and cuteness score y, attributes $\mathbf{a}$ and cuteness score y respectively. For an image I_i, we extract the aforementioned features $\mathbf{x}_i$. The relationship of the attributes and cuteness

score with the raw feature are modeled as follows:

$$\begin{aligned} y_i &= \mathbf{w}_{x,y}^T \mathbf{x}_i + b_{x,y}, \\ \mathbf{a}_i &= W_{x,a}^T \mathbf{x}_i + b_{x,a}, \\ y_i &= \mathbf{w}_{a,y}^T \mathbf{a}_i + b_{a,y}. \end{aligned}$$

Here $\mathbf{w}_{x,y}, b_{x,y}, W_{x,a}, b_{x,a}, \mathbf{w}_{a,y}, b_{a,y}$ are the parameters of the linear prediction models, which will be determined in the model learning process. In the proposed C-LSVM method, the cuteness score of an image is inferred by maximizing the following fitness function,

$$\begin{aligned} \mathcal{F} = & -\beta_1(\mathbf{w}_{x,y}^T \mathbf{x}_i + b_{x,y} - y)^2 - \beta_2(\mathbf{w}_{a,y}^T \mathbf{a} + b_{a,y} - y)^2 \\ & -\|\Lambda(W_{x,a}^T \mathbf{x}_i + \mathbf{b}_{x,a} - \mathbf{a})\|^2 + \mathbf{a}^T (P \otimes M)\mathbf{a}. \end{aligned} \quad (1)$$

The fitness function is a linear combination of the following three types of fitness, *i.e.*, score prediction from raw feature, score prediction from attributes, the attributes prediction from raw feature. And the last term accounts for the attributes correlation, which encourages the correlated attributes to be predicted simultaneously and can help improve the attributes prediction accuracy. The matrix M, constructed based on the statistics from the training data, is of size $n \times n$ and accounts for the attributes co-occurrence.

Besides the three linear models, there are also parameters trading-off the cost terms including $\beta_1, \beta_2, \Lambda, P$. Λ is a diagonal matrix with the size of $n \times n$. The ith element in the diagonal weights the prediction cost for the ith attribute, which in fact reflects the importance of the ith attribute in the cuteness score prediction. And P is a matrix with the same size as M, which weights the co-occurrence of each pair of attributes. All of these parameters, including β_1, β_2, Λ and P need to be determined in the learning process. Here we also employ a latent max-margin framework for the parameter learning and the details are provided in the following subsections.

3.3 Model Learning

In the learning process, we construct the following three prediction functions: the first one predicts the cuteness score from the raw feature, $\mathbf{w}_{x,y} : \mathbf{x} \to y$; the second one predicts the attributes from the raw feature, $W_{x,a} : \mathbf{x} \to \mathbf{a}$; the third one predicts the cuteness score from the attributes, $\mathbf{w}_{a,y} : \mathbf{a} \to y$. We adopt a max-margin regression scheme to learn the above three prediction functions individually.

$$\begin{aligned} & \min \frac{1}{2}\|\mathbf{w}_{x,y}\|^2 + C_{x,y}\sum_{i=1}^{m}(\xi_i + \xi_i^*) \\ \text{s.t.} \quad & y_i - \langle \mathbf{w}_{x,y}, x_i \rangle - b_{x,y} \le \varepsilon + \xi_i \\ & \langle \mathbf{w}_{x,y}, x_i \rangle + b_{x,y} - y_i \le \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \ge 0 \end{aligned} \quad (2)$$

After optimizing the above objective function via off-the-shelf solvers, we can obtain the parameters $\mathbf{w}_{x,y}, b_{x,y}$ of the prediction function $\phi_{x,y}$. The other two prediction functions, $W_{x,a}, b_{x,a}, \mathbf{w}_{a,y}, b_{a,y}$, can be obtained in a similar way.

For the parameters optimization, including $\beta_1, \beta_2, \lambda, P$, we propose the following C-LSVM method. The objective function is defined as:

$$\begin{aligned} & \min \frac{\gamma}{2}\|\mathbf{z}\|^2 + \sum_{i=1}^{m}\zeta_i \\ & \text{s.t.} \ \max_{a \in \mathcal{S}} \mathbf{z}^T \phi(\mathbf{x}_i, \mathbf{a}, y_i) \ge \max_{\substack{a \in \mathcal{S} \\ \bar{y} \in \mathcal{Y}_i}} \mathbf{z}^T \phi(\mathbf{x}_i, \mathbf{a}, \bar{y}) + \Delta(y_i, \bar{y}) - \zeta_i, \end{aligned}$$

where the set $\mathcal{Y}_i$ is defined as $\mathcal{Y}_i = \{y | y < y_i - \epsilon\} \cup \{y | y > y_i + \epsilon\}$ containing all of the incorrect cuteness scores. ϵ is a parameter to control the tolerance of the prediction error and is set as 0.5 throughout the experiments. And the set $\mathcal{A}$ is defined as $\mathcal{A} = \{\mathbf{a} | 0 \le a_i \le 1, \forall i = 1, \ldots, n\}$. Here the parameter $\mathbf{z}$ is the concatenation of $\beta_1, \beta_2, \lambda, P$. The potential function $\phi(\mathbf{x}_i, \mathbf{a}, y)$ is defined as:

$$\phi(\mathbf{x}_i, \mathbf{a}, y) = [\phi_{x,y}(\mathbf{x}_i, y); \phi_{a,y}(\mathbf{a}, y); \phi_{x,a}(\mathbf{x}_i, \mathbf{a}); \phi_{a,a}(\mathbf{a}, \mathbf{a})].$$

In particular, the contained four potential functions are defined as follows:

$$\begin{aligned} \phi_{x,y}(\mathbf{x}_i, y) &= -(\mathbf{w}_{x,y}^T \mathbf{x}_i + b_{x,y} - y)^2, \\ \phi_{a,y}(\mathbf{a}, y) &= -(\mathbf{w}_{a,y}^T \mathbf{a} + b_{a,y} - y)^2, \\ \phi_{x,a}(\mathbf{x}_i, \mathbf{a}) &= -(W_{x,a}^T \mathbf{x}_i + \mathbf{b}_{x,a} - \mathbf{a}) \otimes (W_{x,a}^T \mathbf{x}_i + \mathbf{b}_{x,a} - \mathbf{a}), \\ \phi_{a,a}(\mathbf{a}, \mathbf{a}) &= \mathbf{a}^T P \otimes M \mathbf{a}. \end{aligned}$$

The above optimization problem is equivalent to minimizing the following loss function:

$$\mathcal{L}(\mathbf{z}) = \frac{\gamma}{2}\|\mathbf{z}\|^2 + R(\mathbf{z}), \quad (3)$$

where

$$R(\mathbf{z}) = \max_{\substack{a \in \mathcal{S} \\ \bar{y} \in \mathcal{Y}_i}} \mathbf{z}^T \phi(\mathbf{x}_i, \mathbf{a}, \bar{y}) + \Delta(y_i, \bar{y}) - \max_{a \in \mathcal{S}} \mathbf{z}^T \phi(\mathbf{x}_i, \mathbf{a}, y_i) \quad (4)$$

The subgradient for the above function can be calculated as follows,

$$\begin{aligned} \partial L(\mathbf{z}) = & \ \mathbf{z} + [\phi_{x,y}(\mathbf{x}, y); \phi_{a,y}(\mathbf{a}^*, y); \phi_{x,a}(\mathbf{x}, \mathbf{a}^*); \mathbf{m} \otimes \tilde{\mathbf{a}}] \\ & -[\phi_{x,y}(\mathbf{x}, y); \phi_{a,y}(\mathbf{a}^\star, y); \phi_{x,a}(\mathbf{x}, \mathbf{a}^\star); \mathbf{m} \otimes \bar{\mathbf{a}}]. \end{aligned}$$

Here $\mathbf{a}^*$ and $\mathbf{a}^\star$ are obtained from solving the first and second maximizing problems in (4) respectively. Note that the optimization problems are standard quadratic programming and can be solved efficiently. Here $\mathbf{m}$ is the long vector formed by stacking the column vectors of matrix M. $\tilde{\mathbf{a}}$ and $\bar{\mathbf{a}}$ are formed by stacking the column of matrices $\mathbf{a}^*\mathbf{a}^{*T}$ and $\mathbf{a}^\star\mathbf{a}^{\star T}$. In the optimization, we alternatively solve the problems (2) and (3). In particular, the problem (2) can be solved via standard SVR solver and we can obtain the individual prediction model. Then the problem (3) is solved via standard quadratic programming and the estimation of the underlying attributes are updated. This procedure is repeated until convergence.

3.4 Inference

After learning the prediction functions and the model parameters, for a new image, its cuteness score can be inferred as follows,

$$\{\mathbf{a}, y\} = \arg\max_{\mathbf{a}, y} \mathbf{z}^T \phi(\mathbf{x}, \mathbf{a}, y).$$

In solving the above optimization problem, we first infer the attributes confidence vector $\mathbf{a}$. Then we binarize the elements in $\mathbf{a}$ via a fixed threshold 0.5[2]. After determining the

[2]It is widely used in applying regression for classification.

Table 1: MAE of the cuteness score prediction.

Method	NN	F-S	F-A-S	C-LSVM
MAE	1.92	1.49	1.35	**1.27**

attributes, we then infer the cuteness score y via solving a standard quadratic programming problem. This optimization procedure is inspired by the fact that a sparse attribute vector $\mathbf{a}$ generally produces better prediction result of the cuteness score. The rationale lies on that in the learning process, the attributes annotation associated with each sample are binary values. Thus the learned model prefers such sparse input attribute vectors.

4. EXPERIMENTS

In this section, we evaluate the proposed C-LSVM cuteness score prediction on the constructed dataset. In the experiments, 3,000 images from the dataset are used for training. We compare the C-LSVM method with the following three methods: the first one is based on constructing a feed forward neural network, which has been successfully applied in the beauty prediction [7]. The second one is to directly predict the cuteness score from the raw feature, where the prediction is based on a support vector regression model; and the third one is to predict the attributes at first and then predict the score from the estimated attributes based on two individual SVR models.

The accuracy of the cuteness score prediction is measured by the mean absolute error (MAE) in the evaluation. Note that the groundtruth score is ranged from 0 to 10. The evaluation results are presented in the Table 1. From the results, we can observe that the neural network method performs worst. Introducing the attributes will improve the results over only using raw feature by 0.14. And our proposed C-LSVM can further reduce the MAE by 0.08 and achieves the best result. Note that the MAE for the last three methods are quite small and such improvement is in fact significant. We also present some attributes inference results for the test samples in Figure 3. We can see that most of the attributes can be correctly inferred. While for the *cheek smoothness*, the accuracy is relatively low. The reason may be that the low-level feature we adopted are describing the whole face region, instead of only describing the cheek region. Thus some noise may contaminate the prediction for the smoothness. To more intuitively show how the defined attributes determine the cuteness of one person, we visualize the learned inference model $\mathbf{w}_{a,y}$ in Figure 4. We observe that the attribute *cheek smoothness* is most important for the cuteness of one person. And the attributes *age (young)* and *skin color* are also important. Meanwhile, the *age (old)* and *gender* attributes are least important for the cuteness.

5. CONCLUSIONS

In this work, we present the first research attempt of computational analysis on the cuteness, which is beyond the ordinary expressions such as happiness or sadness. We construct a large dataset of persons' images with well annotated cuteness scores and attributes. We propose the novel C-LSVM method which automatically mines the important features and attributes determining the cuteness of a person. Extensive evaluations show that our method can better capture the relationship between the raw feature, attributes and the cuteness score, over the traditional linear predictors.

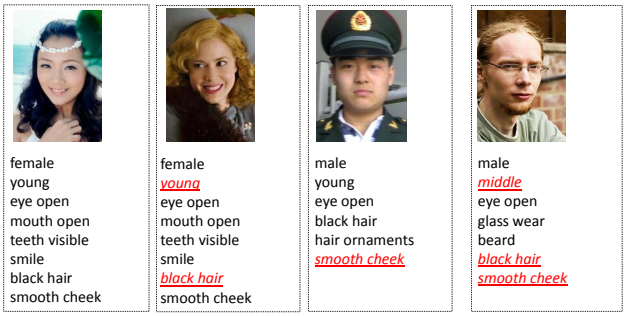


Figure 3: Examples of the inferred attributes. Most of the attributes can be inferred correctly. And the incorrect inferred attributes are highlighted by red color.

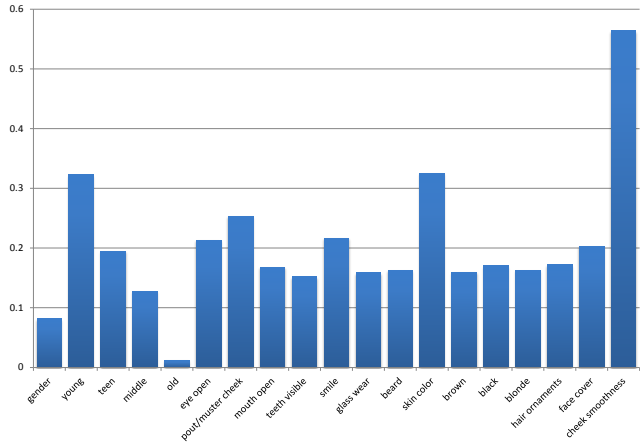


Figure 4: The mined weights of the attributes for cuteness prediction. The horizontal axis displays the attributes and the vertical axis shows the learned weights.

6. ACKNOWLEDGMENTS

This work was supported by Singapore Ministry of Education under research Grant MOE2014-TIF-1-G-007.

7. REFERENCES

[1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. In *ECCV*, pages 469–481, 2004.

[2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, pages 886–893, 2005.

[3] M. L. Glocker, D. D. Langleben, K. Ruparel, J. W. Loughead, J. N. Valdez, M. D. Griffin, N. Sachser, and R. C. Gur. Baby schema modulates the brain reward system in nulliparous women. *Ethology*, pages 9115–9119, 2009.

[4] S. H. Kim and S. Hamann. Neural correlates of positive and negative emotion regulation. *J. Cognitve Neuroscience*, pages 776–798, 2007.

[5] S. Liu, T. Nguyen, J. Feng, M. Wang, and S. Yan. Hi, magic closet, tell me what to wear! In *ACM MM*, pages 1333–1334, 2012.

[6] K. Lorenz. Part and parcel in animal and human societies. *Studies in animal and human behavior*, 1971.

[7] T. Nguyen, S. Liu, B. Ni, J. Tan, Y. Rui, and S. Yan. Sense beauty via face, dressing, and/or voice. In *ACM MM*, 2012.

[8] T. Nguyen, S. Liu, B. Ni, J. Tan, Y. Rui, and S. Yan. Towards decrypting attractiveness via multi-modality cues. *TOMCCAP*, 9(4):28:1–28:20, 2013.

[9] P. Viola and M. J. Jones. Robust real-time face detection. *IJCV*, pages 137–154, 2004.

[10] Y. Wang and G. Mori. A discriminative latent model of object classes and attributes. *ECCV*, pages 155–168, 2010.