Think again about my picture: different approaches investigating factors of influence in profile images context perception

Filippo Mazza, Matthieu Perreira da Silva, Patrick Le Callet

Think again about my picture: different approaches investigating factors of influence in profile images context perception

Filippo Mazza[1], Matthieu Perreira Da Silva[2], Patrick Le Callet[2]

[1]*LUNAM, Ecole Centrale de Nantes, IRCCyN UMR CNRS 6597, FRANCE*
filippo.mazza@irccyn.ec-nantes.fr

[2]*LUNAM, Université de Nantes, IRCCyN UMR CNRS 6597, FRANCE*
(matthieu.perreiradasilva, patrick.lecallet)@univ-nantes.fr

***Abstract*—Online interactions have nowadays a huge impact on everyday life. People use social networks daily both for personal and work relationships. These interactions are inevitably influenced by the few elements present on online profiles. The adagio 'an image speaks a thousand words' is true also in this context: online profile portraits can have a great influence on interactions. Recently multimedia quality assessment opened the field to broader aspects like image aesthetics, interestingness and memorability. Higher level quality dimensions have been indicated as important and proposed in literature under the concept of 'image psychology'. This is particularly true for portraiture images. How to find and measure such factors is still an ongoing research subject. In this paper we refine some results on our previous research work, where we focused on face digital portraits context perception. Those results underlined the importance of some high level content features, e.g. the dress of the portrayed person and scene setting, in categorizing image. Here we consider different statistical approaches to factors of influence analysis, underlining positive and negative points of each one. Our reference model is the one of our previous work: logistic regression, adopted to model category fit based on images features.**

***Index Terms*—portrait images, content perception, high level features, social bias, social networks**

I. INTRODUCTION

Social networks are adopted worldwide on a daily basis. The use of these tools is varied: they can be adopted to meet new people, for dating purposes, for work but also to stay in touch with friends. People are present on social networks through their online profiles, usually under the form of a web page showing their personal informations - including uploaded pictures. These images will be visible to our list of contacts under privacy restrictions we select. However, our main profile picture will usually be shown at every interaction. Very likely, this image can influence profile perception, as demonstrated by a previous work[1]. Online profiles show a big variety of profile pictures both in terms of objective quality (i.e. high quality shots VS simpler consumer produced ones) and content. The impact that these pictures can produce is also very different. In particular, different photos of the same person can deliver very different messages. This point is often overlooked by people: our first impressions do not only influence online interactions, but also real life communications. For example job recruiters may look for online profiles to complement their opinions, as demonstrated by Manant et Al.[2]. It is then important to carefully consider our online profile pictures as they can influence overall personal profile perception. However, it is not easy to define and evaluate image "quality" in this context. What are the elements that can bias profile perception? How to address the problem? Many elements in an image may come into play. As stated by authors of [3], it is important to consider image semantics where "human subjectivity" plays an important role. In this respect, current research related to image quality assessment seems to agree that considering only low level technical features is simply not enough, and so, we have to shift the attention towards high level features, more focused on "the domain of psychology". The latter concept is still broad and vague, but first attempts to outline it have been done by Fedorovskaya[4]. Portrait images have been studied in different fields. Computer vision research focused on face features as input in machine learning algorithms to predict aesthetic assessment (as done by Li[5], or more recently by Khan[6]). Research in cognitive sciences tried to understand and modify image memorability[7]. Our previous work [8] investigated which image features influence portraits' perception, addressing the problem through assessing the context to which a portrait is perceived to belong to. Our assumption was that portraits can suit different needs; for example, a portrait can fit well as a social network profile picture, but then may not fit a resume picture at all. Our analysis considered some classical low level features as well as higher level features related to image interpretation. We approached the problem considering features we supposed mostly important and we showed how multiple features can

SIFWICT2015, 12 June 2015, Nantes, France

be addressed simultaneously. Many data analysis approaches adopted in literature consider black box methods (i.e. ANN, SVM) and/or mathematical descriptors (i.e. HOG, GIST). However, these strategies lose interpretability of the results[9]. We used instead a white box approach since our aim is to understand which interpretable features are influential. While the proposed analysis is not able to explain all variability in the subjective assessments, some statistically relevant features have been underlined and greatly helped to discriminate between portrait contexts. Our first results are consistent with expectations from empirical experience. But other approaches would have been possible to asses features contribution. In this work we go through data analysis again, adopting different statistical analysis. In particular, we adopt Decision Tree as already done in image aesthetics research by Datta[10]. Results underline that classification accuracies of different approaches do not change significantly. The only great advantage of some approaches is indeed interpretability of results.

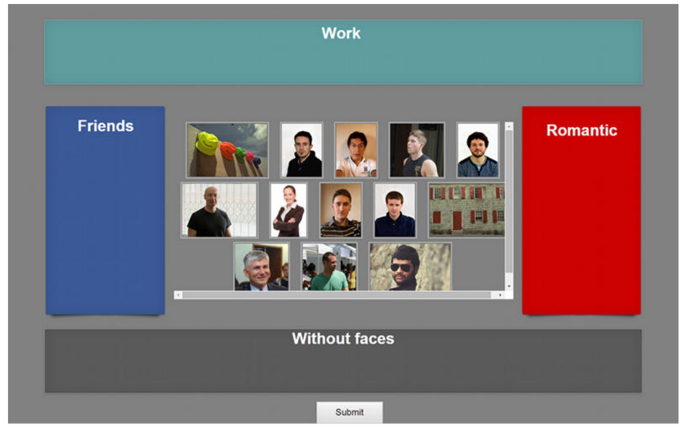
II. DATASET

We provide here a small description of originally adopted dataset, related to our previous work[8]. In that work we collected real online portraits related to three different contexts, namely to friends, work or dating purposes. As ground truth, we collected subjective assessments of the perceived context of each portrait. Successively we extracted both low and high level features from these portraits. The former ones were directly computed from pixel intensities (e.g., contrast), whereas the latter related to content interpretation were assessed subjectively. A statistical analysis on many factors at the same time usually requires a conspicuous number of data points. To have enough subjective evaluations of both content category and high level features, we opted for a large scale subjective campaign via crowdsourcing.

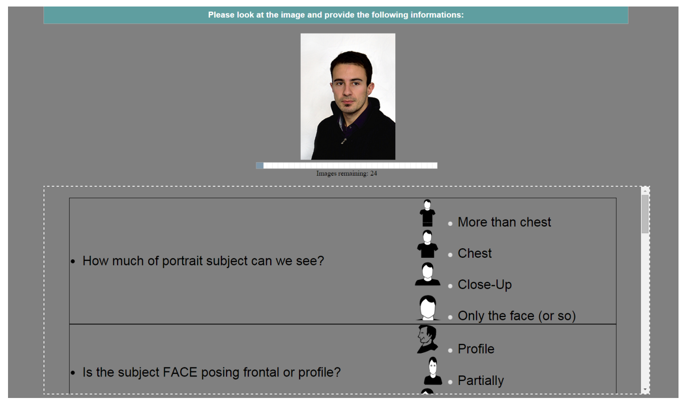
A. Portrait images

Dataset consists of 216 different portraits gathered on online networks and image sharing sites, collecting only publicly accessible images and taking care about licenses' restrictions [1]. Images are mainly amateur or semi-professional pictures, related to the three chosen context categories: friends, work and dating purposes. We based this choice on current trends in social networks, nowadays focused on these three kinds of interactions (e.g., Facebook, Linkedin and Meetic). We chose where possible less formal and "posed" portraits - a characteristic that we supposed influence subjective context perception. We looked for pictures including also a part of a person's torso and partly showing a background, as such portraits convey complementary information regarding the context.

[1]CC attribution licensed images have been taken, provided as found, citing sources.



(a)



(b)

Fig. 1. Interfaces adopted for evaluating context (a) and high level features (b) of portraits.

B. Evaluation of perceived context

Gathered portraits express in our opinion different online interaction purposes. However, as this perception can be highly subjective, we conducted a subjective experiment to collect ground truth. In order to have many evaluations and to reduce cost at the same time, we adopted crowdsourcing. This technique exploits the power of the web in order to outsource small tasks to an online crowd of participants. In our experiment, participations consisted in small ten minutes sessions, each one providing 25 images at a time, paid 0.50$. More informations are provided in a previous work we made, from which we adopted the same platform and software framework[1]. In the subjective experiment, participants were asked to express which context fits best each portrait. They were given three options, referring to friends, work and dating purposes, as shown in Figure II-A. Participants have also been asked to indicate the second best category of their choice. This approach allowed us to have a ranking of the three contexts for each image. However in the analysis carried out in the current paper, we will consider only the first chosen context. Future works will focus on deeper statistical analysis considering second and third choices. Crowdsourcing unfortunately is

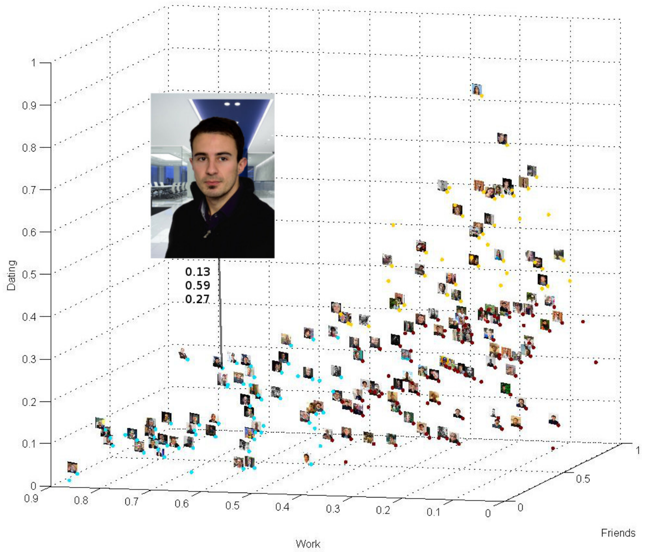


Fig. 2. Subjective context evaluation results in Friends-Work-Dating purposes space for a subset of portraits.

usually affected by a higher outlier rate. Strategies to correct this issue have been adopted, as explained later in section II-D. After removing outlier participants from our experiment, we ended up having each image evaluated by 28 people in average.

C. High level features evaluation

As previously said, we extracted from portraits both low level and high level features; these latter focus on complex features related to image content. The adoption of high level features is not new in literature. Different researches address the analysis of features to describe image content. In particular, the authors of[11] investigate semantic features, mainly related to the setting of the scene (i.e., indoor/outdoor and which scene). Particular focus regarding features of humans is given in[12], where visibility, clothing and appearance are considered. Based on this literature we chose a subset of those features that we believed to be important in portrait context evaluation. We also added some high level features focused on the depicted context; in particular we included the depicted person's gender, the presence of glasses, his/her gaze direction and whether he/she was smiling in the picture. We didn't add emotion as a feature because we considered that smiling was a sufficient predictor to discriminate between emotions as in[13]. All features adopted in this research are summarized in 3. Regarding low level features and image quality assessment, a huge amount of research is reported in literature, and reviewed in[6]. The authors of [6] also adopted some of the proposed features to assess human portraits.

Computer vision based high level features extraction can be time consuming and error prone. As first step, we prefer to focus on features importance and therefore, we adopted crowdsourcing to manually label them. The advantage of this approach was twofold: (1) it greatly speeded up the process and avoided errors, and (2) it offered a subjective opinion for some cognitive features that might be perceived differently between people (e.g., where a portrait has been shot). For this purpose, we followed the same crowdsourcing strategy, changing only the task for the participants. They now were provided with a web interface to evaluate each feature for each picture (as shown in Figure II-A(b)). For some features (e.g., subject profile side) we gave guidelines in the form of icons near the answer options or the possibility to answer 'don't know' when in doubt. In this work we use a majority strategy to define the value of the high level features for each portrait picture. So, if a picture was reported to be shot outdoor by 90% of the participants, it was considered as such for the labelling of the high level features. However a more fine grained quantization to better express human subjectivity can be valuable for our analysis. We will consider this approach in future works.

D. Crowdsourcing reliability issues

Crowdsourcing quickly provided many subjective evaluations (around 450 and 1000 participants, respectively for evaluating subjective context and high level features). While useful, this technique has however some disadvantages. The most important is that participants can provide untruthful subjective opinions, either in good or bad faith (i.e. misunderstanding instructions VS cheating to earn easy money). To have some control on that, we included reliability checks - a.k.a. honeypots. In case of failure in any of the honeypots participants were excluded from analysis. The first honeypot was in the demographic questionnaire provided to participants before starting the actual experiment. The honeypot consisted in the birth year, set to 1880 by default. Users that did not pay attention to the questions, then reported an impossible age. The second honeypot was part of the picture categorization process as two images not showing any person were included in each session. The instructions clearly stated to mark such pictures as 'non faces' through the apposite "without faces" category in the interface. Many participants instead labelled them as portraits and we considered them as outliers. Regarding high level evaluations, some features were used jointly as honeypots to detect outliers - notably depicted person's gender and beard.

III. DATA ANALYSIS AND RESULTS

The main goal of our research is now to link portraits' perceived context with the extracted features. Different approaches are possible, providing different insights on the influential factors. It is important to underline that many of our features are categorical variables - as head tilt and orientation (left,center,right) or scene setting (don't know/indoor/outdoor) - and some of these are ordinal - as subject size (from small to big). This point limits the kind of possible analysis as not all independent variables are normally distributed. Adopted models are Logistic Regression, used in previous work, and Decision Trees. These models allow also to classify new data samples. We then measured classification accuracy as a quantitative measure of model fit. For every classifier we used a leave-one-out (LOO) cross validation, due to the small size of our dataset, providing mean classification accuracy and confidence intervals. For these last we did similarly to [14] considering a Binomial distribution but with a Wald method. Two class labelling strategies have been considered for our ground truth. A first one adopting a majority strategy: every portrait has been labelled with the most selected context. A second strategy took into account the fact that portraits may not present a clear context choice, i.e. when a portrait has been evaluated 50% of times for working purposes and 50% for friends. To find these cases we checked if portrait context choice was stastically significant with Barnard tests on subjective context evaluations. We discovered that the majority of evaluations (71%) do not present a clear context choice. We then added other three classes to the ground truth - Friends/Work, Work/Romantic and Romantic/Friends - to better model our dataset. We will refer to these different class labellings as majority strategy and detailed strategy respectively.

A. *Logistic Regression*

A possible analysis is a Multivariate Logistic Regression, that even if less accurate than a black box approach, allows easier interpretation of the results. Portraits' features represent our regressors and the context ranks as observations to fit. Similarly to Linear Discriminant Analysis, this method attempts to explain the dependent variable as combination of independent variables. However this model is suitable in our case as some independent variables do not have a normal distribution (i.e. dress typology). The model, in case of a single observation can be written as:

$$y_n = \beta_0 + \sum_{k=1}^{K} x_{nk}\beta_k + \epsilon_n \qquad (1)$$

where y is the dependent variable - the context rank for a particular category - x are our predictor values, β are the coefficients to be estimated and ϵ indicates the error term. We fitted two models considering all the samples, one adopting the majority strategy and one adopting the detailed strategy previously described. All features were normalized to the same mean and range, so that the computed coefficients could reflect actual relevance weights. Logistic Regression can also be used to classify new data. First model achieved a classification accuracy of 66% (c.i. 6.3%) while regarding the second one, coefficients regression did not converge. This is due to the low number of samples for the classes in the detailed strategy. As result, its accuracy did not exceed 40% (c.i. 6.5%). Features' coefficients of the first logistic regression model are shown in Figure 3. The Friends context was selected as reference in our model; each computed coefficient expresses then the influence of a feature on the relative chance that a portrait picture is perceived in another context (i.e., Work or Dating) than the reference. This chance is expressed in log odds. So

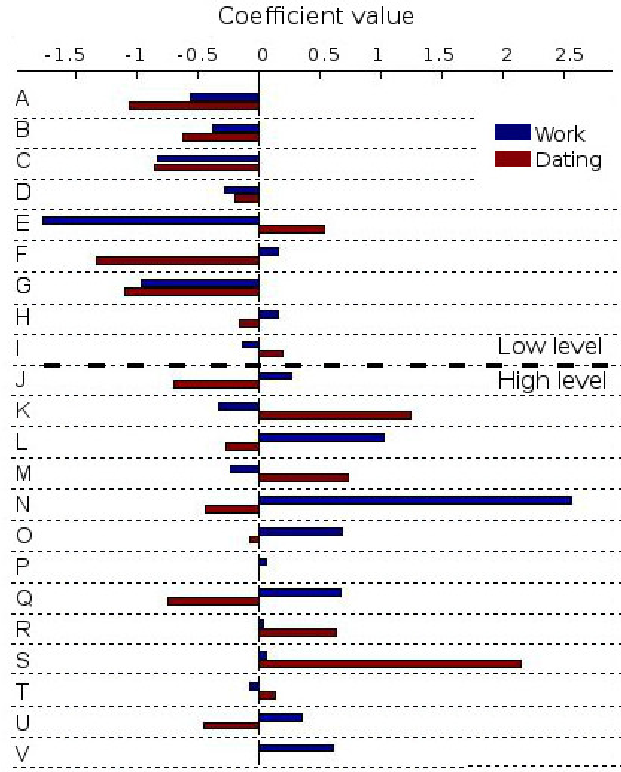


Fig. 3. Logistic regression coefficients. Friend context has been taken as reference.

for example, an increment of one unit in the feature "Dress" increases β_k times the relative log odds of a portrait picture being perceived as Work context, where β_k is the coefficient related to the feature "Dress" in the model of the Work context - assuming everything else being equal.

This model offers many advantages. First of all results are interpretable, as we link directly each feature contribution on context probabilities. Moreover coefficients give us a quantitative measure of each contribution. Secondly, we can also compute a p-value for each feature in the model, indicating the statistical influence of each regressor. Statistical significance of results is shown in table 2. Outcome illustrates that the prediction model for a portrait picture having a Work context is significantly affected by the dress (N) that the person in the picture wears, as well as by the portrait setting (L) and by the low level feature of mean saturation (E) in the picture. As expected, a formal dress increases the appropriateness of the portrait for working purposes, while instead an increase in saturation decreases it. For predicting the appropriateness of a portrait for dating purposes instead the gender of the portrayed person, and his/her face orientation are statistically significant contributors. In particular portraits depicting female persons are reported as more appropriate for dating purposes.

Lastly, the model is relatively easy and computationally inexpensive. However, this lower complexity comes also with the price of considering only linear relationships.

Feature	Work	Dating
A: Aspect ratio		
B: Resolution		
C: Hue μ		
D: Hue σ		
E: Saturation μ	*	
F: Saturation σ		
G: Value μ		
H: Value σ		
I: Contrast		
J: Face size		
K: Face orientation		**
L: Indoor/outdoor	**	
M: Beard		
N: Dress	***	
O: Glasses		
P: Smile		
Q: Sight		
R: Background		
S: Gender		***
T: Head tilt		
U: Background blur		
V: Eyes opened		

Fig. 4. Features significance. *=$p < 0.05$, **=$p < 0.01$, ***=$p < 10^{-3}$

B. Decision Tree

Another possible approach is to adopt a decision tree modelling. This approach has already been employed in image aesthetics assessments, obtaining positive results adopting low level image features[10]. The model builds a tree of decisions to classify the input data. Leaves are classification outcomes (i.e. responses to inputs), and each input corresponds to a path on the tree, starting from the root. Each internal node is labelled with a feature, on which successive splits are made (i.e. decision). Originated branches differentiate by the value on this feature. The learning algorithm then tries to infer the best features and thresholds to construct the tree. At each step, the algorithm examines all possible data splits for every predictor variable, applying the one that maximises a certain criterion. In our case we adopted binary splits and Gini's diversity index as criterion. Our stopping criterion is instead the requirement that all leaves must correspond to a class observation. We fitted two decision trees, one for each class labelling strategy. For the majority strategy this approach gives on our dataset a 58% classification accuracy (c.i. 6.5%, leave-one-out crossvalidation). Part of the obtained decision tree is visualized in figure 5. For the detailed strategy accuracy significantly decreases as for logistic regression, achieving only 34% (c.i. 6%).

IV. DISCUSSION

In this work we adopted different mathematical approaches to addressed the problem of determining influential factors for perceived context of portrait pictures. Differently from our previous work, we considered also context choice significance in our subjective evaluations dataset, representing our ground truth. We discovered that the majority of portraits do not present a statistically significant choice. This points out that inherently our problem is highly subjective, as same portraits are perceived for different purposes from different people. In statistical modelling, we considered both interpretability of results as well as classification accuracy predicting the context of new samples. Both adopted models achieved classification accuracy significantly higher than chance. Features' significance results are the same for of our previous experiment: dress and gender of the portrayed person are shown to be discriminative for the likelihood of a portrait to be perceived respectively as work or dating related. Background interpretation - especially if picture was taken indoor or outdoor - marginally influences context perception. Future research will focus on two points. First, we will consider nonlinear approaches such as Neural Networks and Support Vector Machines. While these are black-box approaches and features contribution will be hard to check, we expect an improvement in prediction accuracy. Second, we will focus on dataset expansion - both in terms of stimuli number and variety - to carry out more accurate statistical analysis.

REFERENCES

[1] F. Mazza, M. Perreira Da Silva, P. Le Callet, and M. P. Da Silva, "Would you hire me? Selfie portrait images perception in a recruitment context," in *Proc. SPIE 9014, Human Vision and Electronic Imaging XIX, 90140X*, B. E. Rogowitz, T. N. Pappas, and H. de Ridder, Eds., no. February, Feb. 2014, pp. 2–6. [Online]. Available: http://proceedings.spiedigitallibrary.org/proceeding.aspx? articleid=1838230

[2] M. Manant, S. Pajak, and N. Soulié, "Do recruiters 'like' it? Online social networks and privacy in hiring: a pseudo-randomized experiment," Munich Personal RePEc Archive (MPRA), Tech. Rep. 56845, 2014. [Online]. Available: http://mpra.ub.uni-muenchen.de/56845/

[3] D. Joshi and R. Datta, "Aesthetics and emotions in images," *IEEE Signal Processing Magazine*, vol. 28-5, no. SEPTEMBER 2011, pp. 94–115, 2011. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5999579

[4] E. a. Fedorovskaya and H. De Ridder, "Subjective matters: from image quality to image psychology," in *SPIE Proceedings Vol. 8651*, B. E. Rogowitz, T. N. Pappas, and H. de Ridder, Eds., vol. 8651, Mar. 2013, pp. 86 510O–86 510O–11. [Online]. Available: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2008681

[5] C. Li, A. Gallagher, A. C. Loui, and T. Chen, "Aesthetic quality assessment of consumer photos with faces," in *Proceedings of 17th IEEE International Conference on Image Processing (ICIP)*. IEEE, Sep. 2010, pp. 3–6. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5651833

[6] S. Khan and D. Vogel, "Evaluating visual aesthetics in photographic portraiture," *Proceedings of the Eighth Annual Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging (CAe '12)*, pp. 55–62, 2012. [Online]. Available: http://dl.acm.org/citation.cfm?id=2328898

[7] A. Khosla, W. a. Bainbridge, A. Torralba, and A. Oliva, "Modifying the Memorability of Face Photographs," in *International Conference on Computer Vision (ICCV)*, 2013. [Online]. Available: http://www.cv-foundation.org/openaccess/content_iccv_2013/papers/Khosla_Modifying_the_Memorability_2013_ICCV_paper.pdf

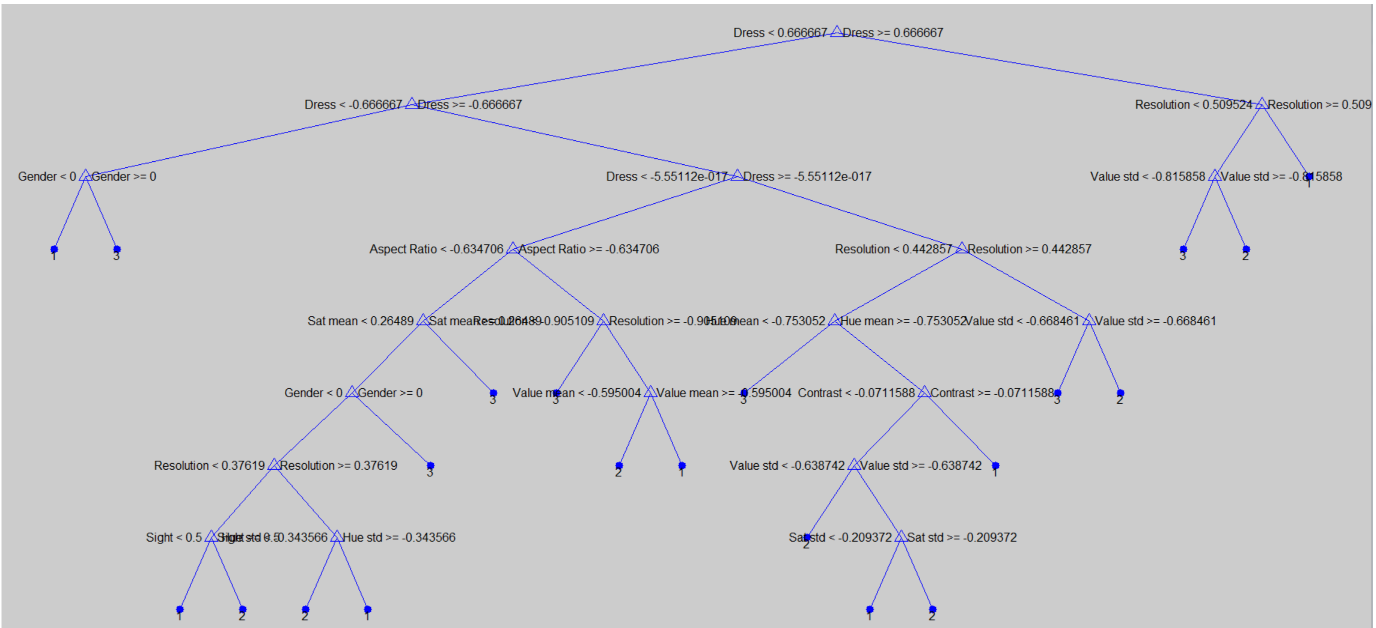


Fig. 5. Excerpt of trained decision tree. Although regressors are normalized, it can be seen that 'Dress' and 'Gender' features are important features influencing portraits classification.

[8] F. Mazza, M. Perreira Da Silva, P. Le Callet, and I. E. J. Heynderickx, "What do you think of my picture? Investigating factors of influence in profile images context perception," in *Human Vision and Electronic Imaging XX*, 2015.

[9] L. Marchesotti, N. Murray, and F. Perronnin, "Discovering Beautiful Attributes for Aesthetic Image Analysis," *International Journal of Computer Vision*, no. November, 2014. [Online]. Available: http://link.springer.com/article/10.1007/s11263-014-0789-2

[10] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach," *Computer Vision-ECCV 2006*, 2006. [Online]. Available: http://www.springerlink.com/index/J062Q28222K34382.pdf

[11] L. Totti, F. Costa, S. Avila, E. Valle, W. J. Meira, and V. Almeida, "The Impact of Visual Attributes on Online Image Diffusion," in *ACM Web Science Conference (WebSci)*, 2014. [Online]. Available: http://files.figshare.com/1518804/webs146_totti.pdf

[12] P. Isola, D. Parikh, A. Torralba, and A. Oliva, "Understanding the Intrinsic Memorability of Images." *NIPS*, pp. 1–9, 2011. [Online]. Available: https://papers.nips.cc/paper/4451-understanding-the-intrinsic-memorability-of-images.pdf

[13] S.-f. Xue, H. Tang, D. Tretter, Q. Lin, and J. Allebach, "Feature design for aesthetic inference on photos with faces," in *2013 IEEE International Conference on Image Processing*. IEEE, Sep. 2013, pp. 2689–2693. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6738554

[14] W. Jiang, S. Varma, and R. Simon, "Calculating confidence intervals for prediction error in microarray classification using resampling." *Statistical applications in genetics and molecular biology*, vol. 7, no. 1, p. Article8, 2008.