Towards a personalized computational framework for predicting aesthetic quality of photographs

Tom Runia
Delft University of Technology
Multimedia Information Retrieval Lab, Delft University of Technology
Delft, The Netherlands
tomrunia@gmail.com

ABSTRACT

The study of aesthetics has always been an intensively studied branch of philosophy. In recent years there has been an increasing interest in the design of computational frameworks for predicting aesthetics from natural images. Although judging beauty is a highly subjective task, certain visual features are considered important to the aesthetic quality. Working prototypes have already been developed for predicting the aesthetic quality of photographs and paintings. These systems yield promising results, however they lack a component of personalization. We argue that adding an individual component to these predictive models is crucial and will be of great value to image search and feedback systems. In this survey paper we present the current state of research on computational frameworks for inferring beauty. Furthermore we define the open challenges and propose future research directions on the personalization aspect of aesthetics prediction.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; I.4.9 [**Image Processing and Computer Vision**]: Applications; I.5.4 [**Pattern Recognition**]: Applications

General Terms

Theory, Algorithms

Keywords

Aesthetics, Image quality assessment, Computational frameworks, Photography, Score prediction, Personalization

1. INTRODUCTION

Aesthetics is the branch of philosophy that deals with questions on beauty and artistic taste [9]. Ancient Greek philosophers such as Plato and Aristotle already thought about the fascinating question "*what is beauty?*". We all seem to possess an inner need for beauty, everyone wants to experience beauty and to be beautiful. However giving a formal definition of the word beauty is impossible due to its subjective nature [21]. In the context of this paper we define *aesthetics* as an abstract notion of quality influenced by various visual image features such as hue, saturation, contrast, colorfulness and composition [17].

Interest in the research community on predicting the aesthetic quality of images has increased dramatically over the past few years [49]. It was established in Datta *et al.* that photo aesthetics, though being subjective, can be estimated using a set of images with *general consensus* on their aesthetic quality. Using average aesthetic ratings, mathematical models can be developed that are able to predict the aesthetic quality of an image based on visual image features.

In the process of predicting the aesthetic value of an image the main problem is correlating low-level image features (pixels) to high level semantics (perception of beauty). Bridging this *semantic gap* turns out to be a very hard problem mainly because low-level image properties are often insufficient for characterizing the high-level perception of aesthetics [5, 44]. Even simple images contain a complex arrangement of objects that can be described using colors, textures, shapes and composition [41]. Another major challenge is the loose and highly *subjective* nature of semantics associated with emotions and aesthetics. Aesthetic quality is always considered as subjective, so there are no absolute standards for measuring the aesthetic quality of an image [27]. The ultimate goal is to develop a user-adaptive system for providing users with images or feedback that is tailored to their individual taste [32].

Better understanding of aesthetics can aid many applications, like summarization of photo collections or providing aesthetics feedback [36]. In recent years extensive work has been done on improving search algorithms to find relevant content in the vast amounts of multimedia data [33]. Popular online image sharing websites (e.g. Flickr) are constantly developing new ways of searching and ranking images to improve user-experience [1]. Historically image search systems focused on search *relevance*, however this might not be sufficient anymore with the growing amount of data. Given that a large number of images are known to be equivalent in relevance, one way to rank them is by their aesthetic quality. This particular area of image retrieval has recently began to generate interest [20].

We also envision a future where consumer cameras or smartphones are equipped with an automated personal assistant that can help capture beautiful pictures. While this may seem unlikely in the near future, literature [4, 22] suggests that (impersonal) aesthetic judgment models can be embedded into cameras without requiring too much computational resources. These technological developments come with some controversial ethical questions that need to be addressed, however this discussion is beyond the scope of the current paper.

The remainder of this paper is organized as follows. In Section 2 we present an overview of existing methods and computational frameworks for inferring aesthetics from photographs without including personalization. We then describe, in Section 3, attempts that have been made on including a personalized component into these computational frameworks. After addressing the available literature we continue by pointing out open challenges. Finally, we conclude and discuss future work in Section 5.

2. COMPUTATIONAL FRAMEWORKS

When a photograph is rated on aesthetics by a group of people, the *average* aesthetic rating can be regarded as an estimator for its intrinsic aesthetic quality. Given a sufficient amount of images with corresponding average ratings (*labeled data*) a computational system can be built to find correlations between certain (visual) image features and aesthetic quality. Here we can make a distinction between a *regression* approach, in which the system does score prediction, versus *classification* in which the system classifies new images as low or high quality. Although score prediction is more feasible, current literature often changes the problem to one of classification since score prediction is highly challenging, mainly due to the large variance among user ratings on aesthetics [20, 27].

Using a simple threshold the average aesthetic rating can be converted into two classes, either a photo is of low or high aesthetic quality. Recent works focus on predicting aesthetic quality of images (low versus high quality) often follow a traditional *machine learning* methodology. In such systems the first step is feature extraction: calculating various image features by applying image processing algorithms. This is followed by a training phase where a classifier is trained using the extracted features together with the labeled data [45]. The trained model can then be used to predict class labels and aesthetic ratings of new unlabeled images based on its features. The remaining part of this section is devoted to addressing these steps and presenting an overview of existing literature.

2.1 Feature Extraction

Obtaining an appropriate set of image descriptors that is best capable of separating objects from each class, usually is the beginning of a machine learning problem. In the context of training a classifier for aesthetics prediction we make a distinction between *two* different categories of features. First we discuss *visual image features* that are contained in the raw pixels and can be extracted using various image processing techniques. The second category consists of features that are outside the image e.g. image descriptions, tags or the view count. We indicate these as *non-visual features*.

2.1.1 Visual Features

With computer vision and image processing techniques becoming increasingly popular over the last decade, there have been significant contributions to the field of feature extraction and image representation for semantics and image understanding [5]. The quality of predictions by a system based on machine learning, depends on the features of images, therefore feature extraction is the foundation of every computational framework for aesthetics prediction based on visual image features. The subject of correlating various visual image features to aesthetic response is widely studied. For example, colorfulness, composition, saturation, depth of field and the presence of humans in images are often related to the aesthetic quality. Although the goal of this paper is not to provide an extensive overview of visual image features that are correlated to beauty, we will discuss some important usage patterns.

Datta *et al.* select visual image features based on artistic intuition to predict aesthetic and emotional quality of images [4, 5]. Tong *et al.* approach the problem of image quality assessment by measures related to image distortion [46]. A more traditional approach is given by Ke *et al.* – they select low-level features such as colorfulness, hue count and the distribution of edges that may be related to high-level attributes such as color preferences or simplicity [22]. Luo *et al.* have developed a method that focuses on the subject region in the image. Based on the subject region, various high-level semantic features are extracted and used for aesthetics prediction [30]. In a work from Bhattacharya *et al.* [1] the problem of image quality assessment is approached by complementing saliency information extracted from the images by high-level segmentation that is associated with the geometric context. SIFT descriptors are a commonly used technique for encoding information about saliency points [31]. Still another approach is by using photographic rules about composition, such as the rule of thirds. As described in [19], compositions that follow the rule of thirds and visual weight balance are strong indicators for high image quality. Image descriptors for encoding texture are also frequently applied, one way to encode the smoothness of an image is by using the Daubechies wavelet transform which is a common method to characterize texture [7].

For in-depth literature on visual image features we refer readers to Torralba *et al.* [15] or Dhar *et al.* [8], both papers contain an excellent overview of common visual image features used in the context of image quality assessment and aesthetics prediction.

2.1.2 Non-visual Features

We now turn our attention to non-visual image features that are suitable as image descriptor (metadata). Additional features such as image descriptions, tags, view count, user-comments or other *metadata* can contain valuable information in relation to aesthetic quality.

Opinion Mining

Analyzing user-comments to recognize opinions on multimedia has become increasingly popular over the last years when user-ratings are limited. Such *opinion mining* or *sentiment analysis* techniques can also be applied for labeling images in online communities [41, 35]. Subjectivity analysis involves various methods that originate from artificial intelligence and natural language processing [47].

In a recent work by Pedro *et al.* [40] textual metadata associated with images is used with the purpose of image search re-ranking. For example words like ‘beautiful’, ‘great’ and ‘awful’ are strong indicators on the aesthetics quality of the photograph. Additionally metadata such as tags, view count and descriptions can also be included as features. These data-driven approaches by means of opinion mining are relatively fast and cheap compared to extracting visual image features using image processing. Therefore it offers an interesting opportunity for large-scale data labeling. Note that opinions can be used both to obtain labeled data and as indicator for image quality in the prediction phase.

2.2 Data Resources

The next step in the process of building the predictive model is obtaining labeled data for training the classifier. The data we need, generally consists of a set of images together with corresponding class label or aesthetics score. In this section we discuss three different methods for collecting labeled data.

2.2.1 Controlled User-Studies

Controlled experiments are the traditional approach for obtaining data in many psychological experiments. In the context of obtaining data on the aesthetic responses of a population, a typical experiment might look as follows.

A group of participants is invited to the study and are asked to give ratings to images simply by looking to them one by one. Additionally the subjects might be asked to write down their impressions or give verbal judgments. Given a sufficient sample size, this method enables researchers to draw conclusions about correlations between certain visual image features and the perception of beauty. For the design of a research experiment with the goal of finding suitable image features, we refer to the work by Li *et al.* [27].

Over recent years various new methods have been developed for measuring the participants stimuli when looking at an image. Researchers now have the ability to include measurements on the heart rhythm, pupillary reactions and eye movement. Even studying neural activity using functional MRI is becoming increasingly common for measuring aesthetic response [24]. Recording eye movements is also a valuable technique, in particular for examining the relationship between image composition and aesthetic quality [34].

Controlled studies can be of great help in understanding the effects of certain image features because the researchers have full control over the images, participants, survey questions and measurements. However, full control of these variables comes with a number of disadvantages – most importantly these studies are generally expensive, time consuming and do not scale very well.

2.2.2 Community-Based Photo Ratings

Obtaining data from controlled user-studies, as we have discussed in the previous section, is a time consuming business. Researchers are increasingly turning their attention towards *data mining* on the web to obtain large amounts of labeled data. In practice this is usually done by downloading a large image database together with a large amount of user ratings associated with each image [20]. One good data source is the large online photo sharing community, Photo.net[1]. This community has more than $400,000$ users and attracts both amateur and professional photographers. Since both photos and rating are publicly available and the website attracts a large number of users, Photo.net is often chosen as dataset for statistical learning on aesthetics [4, 20]. Quite similar and also frequently used is the online photo community DPChallenge[2]. Ke *et al.* have crawled the website to construct a dataset of 60.000 photos each having at least 100 user ratings [22]. In addition to the two discussed data sources, many alternatives are possible. Joshi *et al.* [20] provide an excellent overview of data sources with respect to image quality assessment.

An alternative approach is to exploit the system of 'like' and 'share' on which Facebook is based. In an ongoing experiment Galetta [12] submitted more thousands of images to the social network site with a reach of more than 10.000 followers. He did this by opening three different Facebook profiles, which now have several thousands of followers each. Every day, each of these profiles post a particular image that display some image features. By analyzing the metrics on liking and sharing it was observed that people are inclined to react the same way towards certain visual stimuli coming from images. This approach can reach an extremely large audience for determining average reactions to certain image features, however such datasets are often biased and do not purely measure 'aesthetics'.

2.2.3 Crowd-sourcing

Recently, crowd-sourcing platforms such as Amazon's Mechanical Turk[3] have emerged as powerful tools for efficient and relatively inexpensive experiments that require human intelligence [39]. MTurk is an online platform that contains all major ingredients required to conduct scientific research: a large participant pool, a streamlined process of designing an experiment, participant recruitment and data collection [3]. These features make MTurk an excellent platform for designing experiments on human perception, e.g. aesthetics. Researchers use the platform on an extensive scale to analyze users' perception towards certain image features. Based on large-scale user tests researchers can setup experiments which can give better insight in what people perceive as beautiful. For example Rudinac *et al.* [39] provide new insights in aesthetic appeal towards images by not only describing the images on their properties but also in the context of semantically related images. Such large-scale experiments can be designed in all possible ways and increase our general understanding of beauty.

2.3 Classification and Score Prediction

With a large amount of labeled data available, the next step is to train a computational model. Generally this is done by thresholding the labeled data set into two classes: low and high aesthetic quality. Given this two-class dataset traditional machine learning algorithms can be trained such as support vector machines (SVM), Naive Bayes classifiers or a neural networks [49]. The trained model can then predict the aesthetic class of unlabeled images by feeding its features.

Turning a two-class prediction into score prediction (e.g. on a scale from 1 to 100) for the aesthetic quality can be

[1] `http://www.photo.net`

[2] `http://www.dpchallenge.com`

[3] `https://www.mturk.com`

achieved by mapping the output the binary output of the classifiers described below to a certain numeric scale. Datta *et al.* [6] achieve this by using a *sigmoid function* that maps the distance from the optimal hyperplane (SVM) for a certain image to a numeric scale.

3. PERSONALIZATION

In the previous section we have considered computational frameworks for predicting the aesthetic qualities of images using the general consensus as *ground truth*. These models all have in common that the average rating of an image is used in the learning phase. Although images can be assessed based on aesthetic rules to some extend, these 'universal' rules do not capture personal taste. The highly subjective nature of aesthetics therefore requires computational systems from the previous section to be extended with some form of personalization. For example, someone may prefer images with specific color profile or have a strong preference towards scenic photographs as opposed to portraits [50].

There are two different levels in which we can explore personalization. First, one can consider preference groups, i.e. groups of people who share similar tastes within a social or cultural setting. We call this *generic* personalization. And second, personalization on individual level by including the users' individual taste into the framework (*individual* personalization). The rest of this section discusses these two levels separately and provides an overview of existing literature with regard to personalization.

3.1 Generic Personalization

The first level of including a personalized component in the models is by considering preference groups. As we have seen, computational models commonly use the average rating of the entire user group as consensus of aesthetic quality. Thus the entire population of people that rate the image is regarded as one entity. The first level of personalization would be to divide the population into smaller preference groups that share a similar taste. Online communities generally contain a large variety in *demographic* and *social* context [20]. If we can cluster the population in a number of distinct preference groups that share equal taste, then we can decrease the inner-class variability of aesthetic ratings and we might be able to improve the prediction algorithms. However this is a largely unexplored field, to our best knowledge no literature has been written on finding online preference groups with the goal of personalization in an image quality assessment system.

3.2 Individual Personalization

The field of personalization on individual level is at least as hard and also remains largely unexplored. However there are some attempts on adjusting image quality assessment systems to individual taste. In a personalized system for detecting highlights of multimedia content, Joho *et al.* [18] have developed a system that gives recommendations based on analysis of the facial activity of the viewer. Another relevant work is established on the *personalized affective content analysis method* suggested by Wang *et al.* [48].

Available literature on online personalization often faces the problem of obtaining sufficient and accurate data on individuals. User-ratings and likes associated with multimedia (e.g. images) are good candidates at first glance, however this data is often scarce and can be hard to collect. The idea that ratings and recommendations can be obtained from traces people leave on web is discussed in Liu *et al.* [28]. Opinion mining, sentiment analysis or social data mining are research topics on its own, but can be of help in the context of extracting labeled data on aesthetics. An approach of applying preference mining for personalized applications is given by Holland *et al.* [13]. Alternatively over the last decade there have been written many papers on analyzing *click behavior* in online systems. For example Jain *et al.* have developed a system for query-dependent image re-ranking using click data [16]. However these examples of personalization have not been studied in context of aesthetics prediction.

3.3 Recommender Systems

The majority of literature on online personalization has been written in the form of recommendation systems, such as the ones used by large companies e.g. Netflix, IMDb and Amazon. Over the last couple of years online *recommender systems* have received much attention in the scientific community [43, 25]. Such systems can also be of interest to aesthetic recommendations and approaches the problem from a different perspective compared to machine learning on image features. For example, an approach might be to design a personalized system by drawing inspiration from the collaborative- and content-based filtering paradigms that are frequently applied in recommender systems.

3.3.1 Collaborative Filtering

In particular *collaborative filtering* has proven to be a very effective method – it is known to be the most successful recommendation technique [42]. Collaborative filtering is a method of making recommendations by finding correlations among users of a system. It presents an uniform approach to finding items of potential interest and predicting the rating that the current user would give to an item. Collaborative filtering systems use statistical learning techniques to find a set users (neighbors) that are very similar based on user-profile or past ratings [23]. The main advantage of CF is that there is no need for knowing aesthetic preferences of a user explicitly, since the prediction is based on a set of nearest neighbors [32]. This fundamental difference compared to the machine learning approach makes collaborative filtering an interesting research direction.

Collaborative filtering can be applied in an *item-item similarity* [38] design to provide personalization. Often collaborative filtering approach use aggregate statistics, but ideally we want to predict a aesthetic rating specifically tailed to the individual. If someone likes images of certain paintings of Vincent van Gogh, then he or she might also like other images from Van Gogh. In recommender systems this personalization can be achieved using item-item similarity. In the context of image search re-ranking a system that incorporates item-item similarity, the system would compute the similarity between the images that must be re-ranked and the images that were previously rated by the user. Based on the similarity between the image features the system can make a prediction about the personalized aesthetic quality.

4. CHALLENGES

After reviewing available literature on both impersonal and personal computational frameworks, we now address the open challenges that can be the topic of future research on predicting aesthetic quality.

4.1 High-Level Beauty Semantics

One of the central challenges that always emerges in the discussion on personalized aesthetics is the inherently difficult concept of beauty. For example when comparing aesthetic prediction to recommender systems, the former is much harder due to its highly subjective nature. Also, it is extremely hard to completely separate aesthetic perception from other aspects within the humans' feeling when looking at an image. Interestingness or the photographic theme are factors that can influence aesthetic ratings [27]. Measuring pure aesthetic response is not a trivial challenge.

4.2 User Feedback

The highly subjective nature of aesthetics perception requires personal data in order to design personalized systems at individual level [20]. Traditional machine learning frameworks discussed in Section 2 can help us to understand what visual features are aesthetically attractive in a *general sense*, however they provide no useful information for personalization. The missing piece in the puzzle for making these systems adapt to individual preference is more personal information and aesthetic preference. We make a distinction between to types of data: *personal information* (e.g. age, gender, demographic) and *labeled data* (e.g. ratings) from individuals. Both of these challenges are discussed separately in the following paragraphs.

4.2.1 Personal Information

Generic personalization can be approached from a perspective of clustering users into preference groups (Section 3.1). Incorporating cultural, social and demographic information to find suitable preference groups in online populations is an open research question that has not received much attention. Novel insights in what preference groups are effective could help us to understand what personal attributes are important towards aesthetics perception.

Another important challenge is the context and time dependency of beauty perception. Our perception of aesthetics is not just a snapshot but constantly changes over time. We grow older and develop new interests and notions about what is visually pleasing. On the other hand aesthetic appeal to a users also depends on emotional state. Therefore algorithms that can adapt over time and also take into account the context of the user are desirable. In this context better understanding of retraining classifiers to take into account changes in preferences can result in better performance over time [41].

4.2.2 Labeled Data

Giving personalized recommendations or feedback requires information on the person's aesthetic preferences. However we do not expect the user to constantly give aesthetic ratings or other forms of direct feedback. Therefore we are concerned with the challenge to obtain user-ratings from indirect indicators (e.g. comments) – the underlying challenge of opinion mining is a major challenge in multimedia computing. More specifically relating indirect user-feedback to aesthetic preferences is a research area that can be improved significantly by better understanding of natural language and efficiently correlating comments to each other. Can indirect measures on the user's behavior such as click behavior, query history or attention time be related to aesthetic preference? With better understanding of such relationships we can possibly overcome the problem of insufficient personal data.

4.3 Visual Image Features

The quality of predictions on aesthetics relies heavily on visual image features. In this section we point out some important challenges related to feature extraction.

4.3.1 Understanding and Benchmarking

As we have seen many research has shown that certain visual image features are important to aesthetic quality. Using large scale (crowd-sourcing) experiments we can determine image features which on average are perceived as beautiful. However our understanding of why these features are aesthetically pleasing is very limited. Understanding complex relationships of how and why image features are important with respect to aesthetic quality is an open challenge. Analyzing such relationships might show new breakthroughs in our way towards personalized computational frameworks for inferring beauty.

We also argue that the development of a standardized data set of images for testing visual image features is important. Currently it is hard to compare the effectiveness of various image features because authors use different data sets and testing methodologies. Reliable benchmarking of the effectiveness for image features can be achieved by developing and agreeing on a standardized data set. Addressing this challenge could be very advantageous in determining the best feature set for predicting aesthetic quality.

4.3.2 Computational Cost

Image processing and computer vision algorithms are relatively computational expensive. Large scale applications of beauty prediction or image quality assessment can require feature extraction for thousands of images. In online systems such as Flickr or Facebook that contain huge amounts of images, feature extraction is not yet feasible. Decreasing the cost of running image processing algorithms for millions of images is also a challenge related to feature extraction. An interesting study might be to determine the performance versus computational cost ratio for image features to define a set of features that can yield good results without being highly CPU intensive.

4.3.3 Understanding Content

Content is often a large contributor to human aesthetic perception. Estimating complete and accurate content is beyond current image recognition systems. One of the open challenges in computer vision is understanding the visible content displayed in the image. Examples of content recognition methods are face-detectors, object detection and extracting of scene attributes. Using the available literature on *content-based image retrieval* (CBIR) content related images features can also be included in systems for predicting aesthetic judgments [8, 29, 44]. However the relationship between content and aesthetic quality has received to attention in literature.

4.4 Explaining Predictions

An important research theme in systems that provide predictions or recommendations to users is providing *explanations* that justify the recommendations given by the system. Explaining the predictions on aesthetic quality is an open research direction that has not received attention before. However when personalized aesthetic prediction systems are deployed, integrating a component that explains the choices of the system is important since it aids in maintaining a higher degree in *user-confidence* in the results given by the system [2].

An excellent survey of explanation styles is reviewed by Papadimitriou *et al.* [37]. In the context of aesthetic prediction systems we would argue that *feature style explanations* are the most promising in achieving good results. In a system that takes into account various preference groups *human style explanations* could also be a feasible approach. For example a prediction might be explained by: "our system predicts that you think this image is beautiful *because* in general people from The Netherlands have rated this image as beautiful". However the question whether such explanations are useful and attractive remains open.

5. FUTURE WORK

In the last section of this paper we suggest some future research directions. Our intent is not to provide recommendations for each of the addressed challenges but rather identify the future research that we believe is most important.

5.1 Obtaining Personal Information

We envision a future in which computational systems give truly personalized suggestions and feedback on aesthetics on the level of individual taste. Such systems that can adapt to individuals, require user-feedback on personal preference. An important starting point for future research would be to determine efficient and reliable ways to obtain (indirect) user-ratings in online photo communities.

During the registration steps for an online photo community website the system can ask the user to assign an aesthetics rating to a set of diverse images. This initial set of ratings can be related to image features and integrated into the computational model for aesthetics predicting. The existing frameworks discussed in Section 2 show promising results on classification performance used as underlying foundation for a personalized framework. An interesting research attempt would be to develop a *feature-weighted* neural network based on individual image ratings [14]. For example, if someone generally tends to like colorful photos, then increase the weight on the features associated to (high) colorfulness.

5.1.1 Implicit User Feedback

For accurate predictions on aesthetic quality, asking the users only once (during registration) to provide their perception of beauty on a set of image, clearly is not enough in such complex systems. Our perception of beauty changes constantly over time posing the need for *adaptive* systems. Therefore new methods of obtaining user-ratings for images need to be considered. One approach would be to analyze users' click behavior when browsing through an online photo community. Dou *et al.* [10, 16] show that straight-forward click-based personalization strategies perform considerably well in search strategies and can also be a compelling idea for future work in the context of aesthetics prediction. Fox *et al.* [11] have made a comprehensive study to infer user preference from implicit user feedback in online systems.

5.1.2 Wearable Technology

Another unexplored research area is the application of wearable monitoring devices that can measure heart rate, blood pressure and other signals. With the announcements of Android Wear and Apple Watch it is certain that we will see many new applications based on monitoring and analyzing our body using these 'wearables'. The next decade will witness pioneering advances and many interesting new applications. One interesting application would be to use these signals for indirect aesthetic preference inferrence. To our best knowledge there has been done no work on developing methods of using these personal signals in recommender systems or aesthetic prediction. A interesting work by Lang *et al.* [26] has shown that heart rate can be used to infer aesthetics perception. We believe that the use of such personal signals can be of great importance to personalized aesthetics prediction.

5.2 Additional Recommendations

Additional suggestions for future research include decreasing the computational complexity of visual feature extraction, identifying online preference groups by unsupervised learning and explaining predictions to the user. Also studies on including context-aware aspects into computational frameworks for predicting aesthetic quality might result in new insights. Properties of the context such as the weather outside, time of the day and social context might influence our perception of beauty at a particular time.

6. CONCLUSION

In this survey paper we have provided the reader with a comprehensive overview of the available literature in the field of computational frameworks on aesthetics prediction. We have discussed general themes of visual image features and addressed frequently used data sources. The section on personalization shows that this field is relatively unexplored and many open challenges are available for future research. While very limited work has been published so far, we hope that this exposition will encourage more contributions.

7. ACKNOWLEDGMENTS

This work has been written as part the graduate course *Selected Topics in Multimedia Computing* given at Delft University of Technology. Thanks to Alan Hanjalic and Martha Larson for their feedback and suggestions regarding this paper. Also the in-class discussions with the participants of the 2014 course were very useful and provided new ideas.

8. REFERENCES

[1] S. Bhattacharya, R. Sukthankar, and M. Shah. A framework for photo-quality assessment and enhancement based on visual aesthetics. In *Proceedings of the international conference on Multimedia*, pages 271–280. ACM, 2010.

[2] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez. Recommender systems survey. *Knowledge-Based Systems*, 46:109–132, 2013.

[3] M. Buhrmester, T. Kwang, and S. D. Gosling. Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1):3–5, 2011.

[4] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Studying aesthetics in photographic images using a computational approach. In *Computer Vision–ECCV 2006*, pages 288–301. Springer, 2006.

[5] R. Datta, J. Li, and J. Z. Wang. Algorithmic inferencing of aesthetics and emotion in natural images: An exposition. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 105–108. IEEE, 2008.

[6] R. Datta and J. Z. Wang. Acquine: aesthetic quality inference engine-real-time automatic rating of photo aesthetics. In *Proceedings of the international conference on Multimedia information retrieval*, pages 421–424. ACM, 2010.

[7] I. Daubechies et al. *Ten lectures on wavelets*, volume 61. SIAM, 1992.

[8] S. Dhar, V. Ordonez, and T. L. Berg. High level describable attributes for predicting aesthetics and interestingness. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1657–1664. IEEE, 2011.

[9] O. E. Dictionary. Oxford: Oxford university press, 1989.

[10] Z. Dou, R. Song, and J.-R. Wen. A large-scale evaluation and analysis of personalized search strategies. In *Proceedings of the 16th international conference on World Wide Web*, pages 581–590. ACM, 2007.

[11] S. Fox, K. Karnawat, M. Mydland, S. Dumais, and T. White. Evaluating implicit measures to improve web search. *ACM Transactions on Information Systems (TOIS)*, 23(2):147–168, 2005.

[12] G. Galetta. An introduction to the aesthetic precognition: Threat or opportunity for contemporary art? 2014.

[13] S. Holland, M. Ester, and W. Kießling. Preference mining: A novel approach on mining user preferences for personalized applications. In *Knowledge Discovery in Databases: PKDD 2003*, pages 204–216. Springer, 2003.

[14] K. H. Im and S. C. Park. Case-based reasoning and neural network based expert system for personalization. *Expert Systems with Applications*, 32(1):77–85, 2007.

[15] P. Isola, D. Parikh, A. Torralba, and A. Oliva. Understanding the intrinsic memorability of images. In *Advances in Neural Information Processing Systems*, pages 2429–2437, 2011.

[16] V. Jain and M. Varma. Learning to re-rank: query-dependent image re-ranking using click data. In *Proceedings of the 20th international conference on World wide web*, pages 277–286. ACM, 2011.

[17] W. Jiang, A. C. Loui, and C. D. Cerosaletti. Automatic aesthetic value assessment in photographic images. In *Multimedia and Expo (ICME), 2010 IEEE International Conference on*, pages 920–925. IEEE, 2010.

[18] H. Joho, J. Staiano, N. Sebe, and J. M. Jose. Looking at the viewer: analysing facial activity to detect personal highlights of multimedia contents. *Multimedia Tools and Applications*, 51(2):505–523, 2011.

[19] P. Jonas. *Photographic composition simplified*. Amphoto, 1976.

[20] D. Joshi, R. Datta, E. Fedorovskaya, Q.-T. Luong, J. Z. Wang, J. Li, and J. Luo. Aesthetics and emotions in images. *Signal Processing Magazine, IEEE*, 28(5):94–115, 2011.

[21] K. Karvonen. The beauty of simplicity. In *Proceedings on the 2000 conference on Universal Usability*, pages 85–90. ACM, 2000.

[22] Y. Ke, X. Tang, and F. Jing. The design of high-level features for photo quality assessment. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 419–426. IEEE, 2006.

[23] J. K. Kim, H. K. Kim, H. Y. Oh, and Y. U. Ryu. A group recommendation system for online communities. *International Journal of Information Management*, 30(3):212–219, 2010.

[24] U. Kirk, M. Skov, O. Hulme, M. S. Christensen, and S. Zeki. Modulation of aesthetic value by semantic context: An fmri study. *Neuroimage*, 44(3):1125–1132, 2009.

[25] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl. Grouplens: applying collaborative filtering to usenet news. *Communications of the ACM*, 40(3):77–87, 1997.

[26] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm. Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30(3):261–273, 1993.

[27] C. Li and T. Chen. Aesthetic visual quality assessment of paintings. *Selected Topics in Signal Processing, IEEE Journal of*, 3(2):236–252, 2009.

[28] H. Liu and P. Maes. Interestmap: Harvesting social network profiles for recommendations. *Beyond Personalization-IUI*, 2005.

[29] J. Luo, M. Boutell, and C. Brown. Pictures are not taken in a vacuum-an overview of exploiting context for semantic scene content understanding. *Signal Processing Magazine, IEEE*, 23(2):101–114, 2006.

[30] Y. Luo and X. Tang. Photo and video quality evaluation: Focusing on the subject. In *Computer Vision–ECCV 2008*, pages 386–399. Springer, 2008.

[31] L. Marchesotti, F. Perronnin, and F. Meylan. Learning beautiful (and ugly) attributes. *BMVC. IEEE*, 2013.

[32] B. Mobasher. Data mining for web personalization. In *The adaptive web*, pages 90–135. Springer, 2007.

[33] A. K. Moorthy, P. Obrador, and N. Oliver. Towards computational models of the visual aesthetic appeal of

consumer videos. In *Computer Vision–ECCV 2010*, pages 1–14. Springer, 2010.

[34] C. F. Nodine, P. J. Locher, and E. A. Krupinski. The role of formal art training on perception and aesthetic judgment of art compositions. *Leonardo*, pages 219–227, 1993.

[35] S. Nowak and S. Rüger. How reliable are annotations via crowdsourcing: A study about inter-annotator agreement for multi-label image annotation. In *Proceedings of the International Conference on Multimedia Information Retrieval*, MIR '10, pages 557–566, New York, NY, USA, 2010. ACM.

[36] P. Obrador, R. De Oliveira, and N. Oliver. Supporting personal photo storytelling for social albums. In *Proceedings of the international conference on Multimedia*, pages 561–570. ACM, 2010.

[37] A. Papadimitriou, P. Symeonidis, and Y. Manolopoulos. A generalized taxonomy of explanations styles for traditional and social recommender systems. *Data Mining and Knowledge Discovery*, 24(3):555–583, 2012.

[38] A. M. Rashid, I. Albert, D. Cosley, S. K. Lam, S. M. McNee, J. A. Konstan, and J. Riedl. Getting to know you: learning new user preferences in recommender systems. In *Proceedings of the 7th international conference on Intelligent user interfaces*, pages 127–134. ACM, 2002.

[39] S. Rudinac, M. Larson, and A. Hanjalic. Learning crowdsourced user preferences for visual summarization of image collections. *Multimedia, IEEE Transactions on*, 15(6):1231–1243, 2013.

[40] J. San Pedro, T. Yeh, and N. Oliver. Leveraging user comments for aesthetic aware image search reranking. In *Proceedings of the 21st international conference on World Wide Web*, pages 439–448. ACM, 2012.

[41] N. Sawant, J. Li, and J. Z. Wang. Automatic image semantic interpretation using social action and tagging data. *Multimedia Tools and Applications*, 51(1):213–246, 2011.

[42] J. B. Schafer, J. Konstan, and J. Riedl. Recommender systems in e-commerce. In *Proceedings of the 1st ACM conference on Electronic commerce*, pages 158–166. ACM, 1999.

[43] Y. Shi, M. Larson, and A. Hanjalic. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Computing Surveys (CSUR)*, 47(1):3, 2014.

[44] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(12):1349–1380, Dec 2000.

[45] X. Sun, H. Yao, R. Ji, and S. Liu. Photo assessment based on computational visual attention model. In *Proceedings of the 17th ACM international conference on Multimedia*, pages 541–544. ACM, 2009.

[46] H. Tong, M. Li, H.-J. Zhang, C. Zhang, J. He, and W.-Y. Ma. Learning no-reference quality metric by examples. In *Multimedia Modelling Conference, 2005. MMM 2005. Proceedings of the 11th International*, pages 247–254. IEEE, 2005.

[47] M. Tsytsarau and T. Palpanas. Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, 24(3):478–514, 2012.

[48] H. L. Wang and L.-F. Cheong. Affective understanding in film. *Circuits and Systems for Video Technology, IEEE Transactions on*, 16(6):689–704, 2006.

[49] L. Yao, P. Suryanarayan, M. Qiao, J. Z. Wang, and J. Li. Oscar: On-site composition and aesthetics feedback through exemplars for photographers. *International Journal of Computer Vision*, 96(3):353–383, 2012.

[50] C.-H. Yeh, Y.-C. Ho, B. A. Barsky, and M. Ouhyoung. Personalized photograph ranking and selection system. In *Proceedings of the international conference on Multimedia*, pages 211–220. ACM, 2010.