Computational Aesthetics in Graphics, Visualization, and Imaging (2008)
P. Brown, D. W. Cunningham, V. Interrante, and J. McCormack (Editors)

Perceptual and Computational Categories in Art

C. Wallraven[1] and D.W. Cunningham[2] and R. Fleming[1]

[1]Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[2]WSI-GRIS, University of Tübingen, Germany

Abstract

The categorization of art (paintings, literature) into distinct styles such as expressionism, or surrealism has had a profound influence on how art is presented, marketed, analyzed, and historicized. Here, we present results from several perceptual experiments with the goal of determining whether such categories also have a perceptual *foundation. Following experimental methods from perceptual psychology on category formation, naive, non-expert participants were asked to sort printouts of artworks from different art periods into categories. Converting these data into similarity data and running a multi-dimensional scaling (MDS) analysis, we found distinct perceptual categories which did in some cases correspond to canonical art periods. Initial results from a comparison with several computational algorithms for image analysis and scene categorization are also reported.*

Categories and Subject Descriptors (according to ACM CCS): J.4 [Computer Application]: Social and Behavioural Sciences Psychology; J.5 [Computer Application]: Arts and Humanities Fine arts

1. Introduction

Art historians commonly divide Western pictorial art into distinct stylistic classes, movements, or periods. In many cases, the attribution of an artist or artwork to a particular class or movement is uncertain, and the number of art periods, as well as their duration and definition is a matter of ongoing debate. Nevertheless, this categorization has had a profound influence on how art is presented, analyzed, and appreciated. How can one achieve such a categorization– how can different works of art and different painters be attributed to a specific art period? Given a quick glance at a painting, knowledgable experts can identify the art period (as well as the painter and the title of the work) based on several types of information which need to be integrated:

- Low-level pictorial information: technique, thickness of brush strokes, type of painting material (oil, acrylic, etc.), color composition of the scene
- Mid-level content information: specific objects or scenes that are depicted, type of painting or sujet (landscape painting, portrait, etc.)
- High-level background information: knowledge about specific historical events, knowledge about art periods

Based on these types of information, different clusters can be formed and the history of art can be structured by assessing similarities within a large selection of pictorial art. Following Little [Lit04], we can distinguish between periods defined by a trend within the visual arts, a broad cultural trend, an artist-defined movement, and a retrospectively applied label. These different types of art periods are usually found to be defined in quite abstract, art historical terms emphasizing high-level, expert concepts.

The first main question which this paper addresses is then: to what degree can art periods be determined based on low-level and mid-level information alone? The answer to this question will be therefore the degree with which *art periods are perceptually founded*. Similar to how art experts would assess the category of images by identifying similarities and dissimilarities, we asked naive participants to group several hundreds of images into categories based on the artistic style. The data which we gathered gives insights not only into perceptual qualities of art but also allows us to investigate the perceptual validity of the canonical art periods. Going beyond perceptual categories, the second main question was: can we also model the perceptual results computationally? In other words, can we find a computer algorithm which takes as input the same paintings given to our participants and which then clusters the paintings in a similar fashion? As today's computer vision algorithms are only beginning to be able to work on mid-level content information, this part of the study mostly focuses on how low-level, pictorial information determines (perceptual) art periods.

Art period	Artist				
Gothics (Goth)	Bondone	Bosch	di Buonisegna	Van Eyck	Weyden
Renaissance (Rena)	Altdorfer	del Sarto	Botticelli	Duerer	Signorelli
Baroque (Baro)	Caravaggio	Poussin	Rembrandt	Reni	Velazquez
Rokoko (Roko)	Fragonard	Nattier	Pesne	Rigaud	Watteau
Classicism (Clas)	David	Dietrich	Ingres	Kauffmann	Mengs
Romanticism (Roma)	Friedrich	Delacroix	Gericault	Spitzweg	von Blechen
Realism (Real)	Courbet	Menzel	Millet	Repin	Thoma
Impressionism (Impr)	Cezanne	Liebermann	Manet	Monet	Sisley
Expressionism (Expr)	Macke	Marc	Nolde	Pechstein	Schiele
Surrealism (Surr)	Chirico	Duchamp	Ernst	Magritte	Tanguy
Postmodern Art (Post)	de Maria	Hesse	Malevich	Newman	Stella

Table 1: *Art periods (and abbreviations) and representative artists chosen for our experiments.*

2. Related work

As this study is rooted both in perceptual and computational contexts, the following briefly reviews the most relevant work in these two areas.

2.1. Perception of art

The formation of categories is one of the most fundamental aspects of perception and perceptual learning [Har05] and has been studied extensively in cognitive and perceptual psychology, as well as in neuroscience. Recent work emphasizes especially the role of *similarity judgments* in creating categories [Slo03]. The study of art in perceptual terms also enjoys a long tradition in perception research. Perhaps one of the most influential accounts is the the book by Arnheim [Arn74] in which he develops a theory of art perception based on Gestalt principles - many of these Gestalt principles also had an influence on and were influenced by art and art theory. Going beyond representational work, recent research has also focussed on how we might perceive indeterminate art - art which was intentionally created not to allow a clear interpretation (e.g., [WKK*07]). Even in the neurosciences, perception of art has gained an interest, most notably in the recent book by Zeki [Zek99] which explores the connections between physiological structures of the visual system and art. In [Red07], natural image statistics—which the human visual system is highly tuned to—are used to analyze different works of art. As we cannot give an exhaustive overview of the many attempts to connect vision science and art, we refer the reader to two special issues of the Journal Spatial Vision [Spa06, Spa07], a short review paper by Spillmann [Spi07], as well as the book by Livingstone [Liv02].

2.2. Computational analysis of art

In the context of computational analysis of art, in [CHL05] the authors proposed a framework for distinguishing paintings from photographs based on color edges, spatial variation of colors, number of unique colors, and pixel saturation. No single feature could do the task, but a combination of features was shown to yield good discrimination performance. In [MTSI05] concepts based on color temperature (warm versus cold), color palette (primary, complimentary) and color contrasts (light-dark) derived from paintings are introduced. These resulted in good classification performance. By extending the feature concepts to include texture features [MCJ07], performance increased further as the system was also able to work with different brush stroke patterns. Based on this work, [LCR07] go beyond low-level feature analysis and propose a method which assigns visual concepts (such as the coloring or the brush stroke used) to parts of a painting. Using an ontology-based disambiguation method, their approach outperforms other techniques relying solely on classification of low-level features. In [DJLW06], 56 computational features were used to model human aesthetic scores on a database of photographs. These features comprised color, texture, as well as shape cues, and also included higher-level scores based on the photograph's similarity to a "standard database" of images. The latter turned out to be a crucial ingredient to modeling the aesthetic scores. All of the mentioned works mainly rely on lower-level features to analyze and categorize artworks. The book by Leyton [Ley07] proposes a mathematical framework for understanding the *structure* of paintings, i.e., the spatial organization and composition of the different elements and how they serve to create an aesthetic "whole" - although this is perhaps one of the most ambitious attempts at a quantitative description of art, so far no automatic algorithms exists which can perform the tasks required by the framework of [Ley07].

3. Stimuli

A total of 11 art periods were selected for investigation based on various sources. We tried to select art periods at a sufficiently broad level, which would also include larger periods of time in art history. Using this criterion, we came up with 11 art periods: Gothicism, Renaissance, Baroque, Rokoko, Classicism, Romanticism, Realism, Impressionism, Expressionism, Surrealism, Postmodernism.

For each of these art periods we identified 5 major artists whose work mainly can be attributed to the time period in question - these are listed in 1. Then, for each of the artists, we chose 5 representative paintings with which we tried to cover the artistic spectrum. As we were not interested in a familiarity task, we tried to choose less well-known paint-

Figure 1: *Example images for each artistic style.*

ings for the more famous artists in our database. All images were taken from `www.prometheus-bildarchiv.de` - an online library which provides access to a variety of large, online collections of art. Figure 1 shows example paintings from the 11 different art periods.

4. Free-Sorting Experiments

In order to assess "naive" judgments of art, we conducted two free-sorting experiments with similar experimental design and analyses, which are described in the following.

4.1. Experimental Design

15 participants took part in the first experiment and 10 participants in the second. We specifically selected participants who indicated to be non-experts in art, art history, or art practice. Before the main experiment, we first gathered familiarity ratings for all images from each participant. This was done not only to make sure that familiarity with a subset of the paintings would not interfere too much with the grouping but also to show all paintings once to the participant. For this task, participants viewed all 275 images in randomized order on a standard 21" CRT set to 1280x1024 pixel resolution. Images were shown centered on the screen and participants were instructed to rate the familiarity of each image on a scale of 1 (not familiar at all) to 7 (definitely familiar). In order to properly anchor the scale, every number was also explained with a short sentence. Participants could set their own pace and finished in ≈13 minutes on average.

The main experiments followed immediately after the rating task. Printouts of the paintings were shuffled and spread out on a large table. Participants were then instructed to group the printouts into clusters according to *painting style or art period*. They were explicitly asked *not* to group according to image content (still-life, portrait, etc.), if possible. For the first experiment, we did not constrain the number of clusters, whereas for the second experiment, we instructed participants to group all 275 images into 11 clusters. For both experiments, the number of paintings per cluster could be set freely (therefore leading to unevenly-sized clusters). After the experiment we gathered some more information in a questionnaire - these included specific strategies for solving the task as well as subjective ratings of the difficulty of the experiment, participants' expertise in art and familiarity with our chosen art periods. Overall, participants took ≈80 minutes on average for this task.

4.2. Analysis Methods

First of all, we analyzed the data of the familiarity rating experiments. Familiarity ratings were averaged across all paintings. We also looked for "outliers", i.e. paintings which received consistently higher ratings.

For the main experiments, the raw data for the first set of analyses consisted of determining which image was put into which cluster for each participant. From this data we determined three measures:

- Number of clusters for this participant (for the second experiment, this value was, of course, always 11)
- Consistency score for each artist c_{artist}: Regardless of our attribution of artists to art periods, the only "true" labels we have are the 5 images which belong to one artist. In order to determine how well participants did in the grouping task, this score therefore counts the number of clusters n_{artist} across which the paintings of one artist are spread. The higher the consistency score $c_{artist} = (4-(n_{artist}-1))/4$ (with $0 \leq c \leq 1$), the more paintings of one painter were put into the same cluster.
- Consistency score for each art period c_{period}: By averaging all 5 c_{artist} we can determine how consistently one particular art period was treated.

For the next step, we converted the data into a similarity matrix: for each image i we increased a counter in cell (i, j) if i was in the same cluster as j. Averaging all matrices across participants yields an average similarity matrix. Matrices for both experiments are shown in Figure 5. Such a matrix enables two types of analyses: (i) multi-dimensional scaling, which projects the data into a lower-dimensional space enabling us to investigate the criteria along which similarity was judged; and (ii) clustering analysis which tries to predict distinct, consistent clusters allowing us to see which items would be grouped together.

MDS refers to a family of algorithms which operate on proximity data taken between pairs of objects [BG05]. The output is a configuration of objects embedded in a multidimensional space. Psychologists have used MDS to explore perceptual representations of different visually (e.g., [SC73]) presented object sets. The technique has also found a large following in domains such as knowledge mapping and marketing because it allows for the identification of psychological dimensions of stimulus variation (e.g., dimensions along which buyers differentiate amongst competing products) and quantification of perceptual distances between stimuli (e.g., how closely related fields of research are).

Hierarchical clustering analysis was also applied to the similarity data. Briefly, the algorithm proceeds as follows: (1) initialize by assigning each item to one cluster, then (2) find the most similar pair of clusters and merge them, compute distances between this new cluster and the remaining clusters (based on averaging similarities); (3) repeat previous two steps until all items are clustered into one big cluster. The correlation between the cluster hierarchy and the similarity data provides a "goodness-of-fit" measure.

4.3. Results

4.3.1. Familiarity

Familiarity ratings for both experiments overall were very low - on average participants' familiarity with all paintings was only 1.5 (this is inbetween the lowest category ("Totally unfamiliar. You are sure that you have not seen this image before, and have no idea who might have created it.") and the second-lowest category ("Moderately unfamiliar. You feel that you probably have not seen this particular image before, although you may have seen similar ones.")). The maximum average value (averaged across all 25 participants) for a single painting was 2.2. Taken together this data indicates that - even though some paintings received the maximum rating of 5 from a participant - *overall* familiarity with the stimulus set was very low, such that we can exclude any influence of this factor on the data from the main experiment.

4.3.2. Raw data analysis

Number of clusters: Figure 2 plots the number of clusters for each participant in the first experiment. There is a large variety across participants ranging from 7 to 41 clusters. The average number of clusters was 17.8 - just looking at this number we already can tell that participants on average did not try to put each of the 55 artists into a separate cluster but rather tried to build larger groups. In order to go into more details, we need to look at the consistency scores, however.

Average consistency scores: Figure 3 shows averaged consistency scores for each participant for the first and second experiment. Only focusing on Experiment 1 and looking at Figure 2 it seems as if participants who made fewer clusters also scored higher on average. Intuitively, this might make

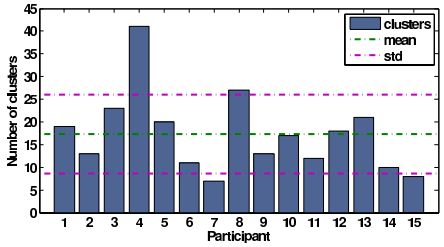

Figure 2: *Number of clusters for Experiment 1*

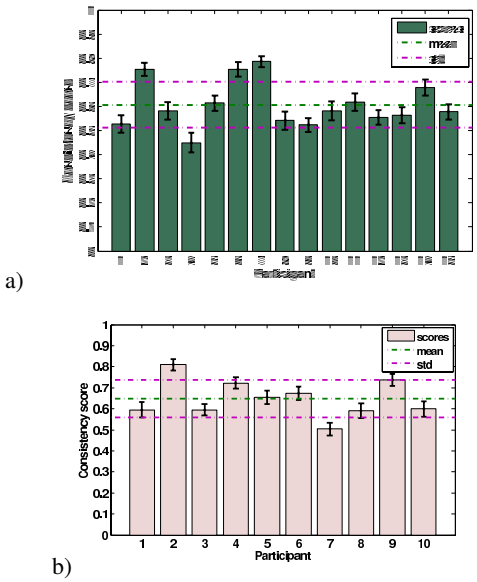

Figure 3: *Consistency scores for a) Experiment 1 b) Experiment 2*

sense at first glance, as the chances of putting all paintings of one artist into one cluster is higher the larger the cluster is. A correlation analysis between number of clusters and consistency score, however, revealed only a moderately strong (but statistically significant) correlation of $r^2 = 0.44$. Participant 14 had a larger number of clusters than participant 15, for example, but at the same time had a significantly *higher* consistency score showing that it cannot be the number of clusters alone which determines participants' consistency. Similarly, if this were the case then also the data from Experiment 2 should yield *similar* consistency scores for all participants. Figure 3, however, shows a different pattern: even though the average number of clusters is now 11 (as opposed to 17.8), on average consistency scores are only a little higher, showing a very similar spread. The average scores for the two experiments were 0.61 and 0.65, respectively.

Consistency scores for artists and art periods: Figures 4a,b plot c_{artist} for Experiment 1 and Experiment 2. The consistency matrices exhibit a large degree of variation - some artists are clearly treated more consistently than others.

For both experiments, Bondone, Di Buonisegna, Pechstein, Newman, and Stella were grouped most often into the same cluster, whereas Courbet and Thoma were the least consistently grouped artists. A closer look at the clusters shows that participants often had no trouble identifying older versus postmodern art (for example, Christian motifs versus fully abstract art) which explains this finding. The consistency scores in Figure 4c show that participants were least consistent for Realism and most consistent for Postmodernism, Expressionism, and Gothicism. Additionally, a post-hoc ANOVA comparing c_{artist}, c_{period} across experiments did *not* find significant differences due to the large individual variances. Overall, we thus found no differences between the two experiments both of which had modern art periods grouped more consistently than older ones. The full similarity matrix discussed in the following section makes this even clearer.

Similarity matrices: The similarity matrices in Figures 5 a,b both exhibit two large clusters as square patterns - these correspond to earlier and modern art periods. A closer look shows that a clear split between the two square patterns occurs at the Impressionism period. Interestingly, this corresponds to an important transition in the development of modern art, when artists moved away from realism, and experimented with new painting techniques resulting in thicker, broader brush strokes, vivid colorings, and abstracted shading patterns. On average, these patterns thus show that participants were well aware of the historic turn from realism to modern art - a non-trivial result given that all participants were non-experts! The consistent art periods show up as as small, reddish, square patterns in the matrix.

Figure 5 c also shows how the pattern should look like if participants were grouping by artist or by (our notion of) art period. Comparing these ideal patterns to the experimental similarity matrices clearly shows that participants grouped by period rather than by artist. Finally, the matrix also can give insights into the most common confusions that occur. Some examples of this include (a full analysis cannot be done here due to space constraints): some of Duchamp's paintings were frequently put into other categories; Van Eycks portraits were rarely in the earliest art clusters; paintings by Newman and Stella were often confused; etc.

MDS analyses: In order to find out how many dimensions can be used to approximate the similarity data well enough, the point where the Eigenvalues of the MDS solution taper off after a sharp, initial drop is commonly determined. This "elbow" can be set at roughly 25 dimensions which then capture $\approx$80% of the variance in the similarity data. The MDS solution can thus be used to localize each of the paintings in a 25-dimensional space. In order to determine what some of these dimensions might correspond to, we projected the paintings onto each dimension and plotted the images occupying the extreme ends of the scale. The results of this analysis are listed in Table 2 for the first 6 dimensions (corresponding to $\geq$60% of the variance) which had a clear perceptual interpretation.

First of all, the first dimensions for both experiments were very similar showing that participants in both experiments used similar criteria along which to group the images. The most important dimension separated the historically old art periods from the newer ones. The second dimension separated flat images (such as the abstract, single-color paintings from Newman, but also the perspective-free Gothic images) from images with more depth structure (such as landscapes, but also street scenes). The third dimension was used to distinguish between post-modern and expressionist art, whereas the fourth separates landscape paintings from portraits. The fifth and sixth dimension were used to differentiate between Surrealism and Postmodernism, and between whole-figure paintings versus portraits, respectively. On each of these six dimensions, neighboring paintings often share the same art period or even artist.

Whereas some dimensions are clearly genre-based, others separate whole art periods, which corresponds well to the consistent cluster structure seen in the similarity matrix in Figure 5. Participants thus used a mixture of historical and content-based clues to group the paintings. The reason for the success of content-based information - even though we explicitly asked participants not to group based on genre, for example - lies in the fact that many art periods do, indeed, have a clear content preference: For the period of Rokoko, for example, many painters were appointed to court and were expected to create often highly stylized and glorifying paintings of the absolutist monarchs - thus creating many portraits and full-figure paintings which are typical for this period.

Clustering analyses: Whereas the dimensionality analysis from MDS has revealed possible image properties that participants use to classify the paintings, clustering analysis allows us to look at which paintings on average will be grouped together based on the similarity matrix. First, we grouped all images into two clusters to find the most important distinction within the data. Second, we created a grouping based on 11 clusters to check what a perceptual grouping of the same size as our art period grouping would yield.

For both experiments, the two clusters split the 275 paintings into a larger group containing paintings up to the Impressionist period and a smaller group starting with the late impressionists. The clusters for both experiments are identical except for only two impressionist images which change cluster membership. This very stable result confirms the overall pattern discussed earlier and testifies to the ability of even non-experts to create theoretically meaningful structure in our sample of paintings. The results for the 11 clusters are summarized in Table 3. Again, results for both experiments are very similar. Additionally, we find period-based clusters, content-based clusters, as well as clusters based on painting-style (Cluster 7 for Experiment 1).

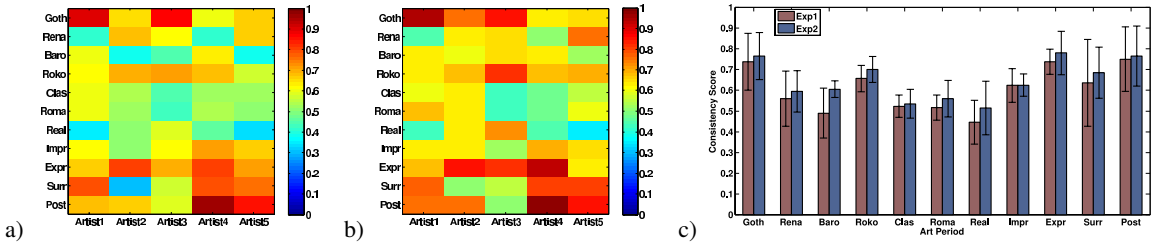

Figure 4: *Consistency matrix showing all c_{artist} for a) Experiment 1, b) Experiment 2; c) Consistency values for art periods c_{period} (error bars depict SEM). Each entry in the matrices corresponds to the entry in Table 1.*

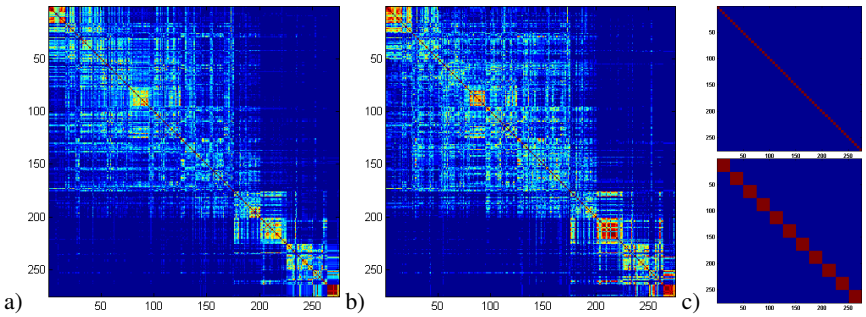

Figure 5: *Similarity matrices for a) Experiment 1, b) Experiment 2, c) ideal grouping according to artists or periods*

5. Computational Experiment

For the computational experiment, we were interested to find out to which degree low-level, pictorial cues would be able to explain the groupings in the previous, perceptual experiments. To this end, we implemented several image processing algorithms which yielded similarity values between all image pairs enabling us to conduct the same analyses (MDS, clustering) as for the perceptual experiments.

5.1. Algorithms

One of the most straightforward low-level similarity measure is to take the Euclidean distance (or L2 distance) between the pixel values of two color images. This is of course very far from being perceptually plausible as the brain does not assess similarity by subtracting two images, but it provides a good baseline for the further measures.

As color is one of the most prominent features in art, color histograms might provide a better correlation with the human data. In order to test this hypothesis, color histograms were extracted from both RGB-versions and CIELAB-versions of the paintings. Histograms always had 20 bins for each of the three color channels. CIELab is a perceptually plausible color space - the L-dimension corresponds to the luminance, the a-dimesion to green-red transitions, and b-dimension blue-yellow transitions. The color opponency scheme is derived from both perceptual and physiological experiments on how the human visual system processes color (among other things, color opponency explains the after-images appearing after fixating a colored surface for a long time).

Fourier analysis was done as the third feature class. The human visual system is highly tuned towards the statistics of the environment as measured by fourier statistics (see [SO01]). The slope of the amplitude spectrum (as a function of frequency) of natural images, for example, was found to be close to 2 ($a = \frac{1}{f^2}$) corresponding to the scale invariance or self-similarity usually found in nature. Perhaps the amplitude spectra of paintings could be used to tell different artists, or art periods apart given that fourier analysis is sensitive to spatial frequencies (for example, thicker brush strokes in later art periods, very fine brush strokes in the Realism period). For our experiments, we determined the amplitude spectrum of all images by binning across 100 frequency bins in the two-dimensional fourier spectrum yielding a 100-dimensional vector.

Several studies in scene perception have shown that humans are able to understand the general context of novel scenes even when presentation time is very short (<100 msec) [TFM96]. This overall meaning of a scene is often referred to as "gist" and most commonly refers to low-level, global features such as color, spatial frequencies and spatial organization. In computational studies [OT06] it was found that a very simple frequency analysis (based on the output of filters tuned to different orientations and scales) of the spectrum of an image can already be enough to classify a scene as indoors, outdoors, open, closed, - in short, the frequency spectrum allows to determine the *spatial content* of an image. In recent scene categorization experiments [OT06], this approach was also shown to have perceptual relevance as it was able to account for several experimental results on

Dim	Exp1 and Exp2	ImDist	RGB	CIE	FOUR	GIST
1	older realistic motifs↔Surr, Expr	dark↔light	dark↔light	dark↔light colorful	dark↔light	flat↔textured
2	perspective (flat↔open)	top light↔bottom dark	dark↔light	dark↔light	-	-
3	Post↔Expr	green↔red	light red↔dark blue	-	flat↔textured	flat↔depth
4	landscapes↔portraits	yellow↔blue	-	dark red↔yellow	-	fine↔-coarse-grained
5	Surr↔Post	blue↔green	-	-	-	-
6	figures↔portraits	blue↔red	-	-	-	-

Table 2: *Results of dimension analysis following MDS for perceptual grouping and five computational measures.*

Clus	Exp1	Exp2
1	Goth, Rena	Goth, Rena
2	Portraits (Goth,Rena,Baro,Roko,Clas)	Portraits (Goth,Rena,Baro)
3	Landscapes (Roma,Real,Clas)	Landscapes (Roma,Real,Clas)
4	Paintings with people (Rena,Baro,Clas,Roma)	Paintings with people (Rena,Baro,Clas,Roma)
5	Impr landscapes	Portraits (Roko,Clas,Real)
6	Outliers	Outliers
7	Fine brushstroke paintings	Outlier
8	Impr, Expr	Impr, Expr
9	Expr (Cubist)	Impr landscapes
10	Surr	Surr
11	Post	Post

Table 3: *Results of cluster analysis for perceptual grouping.*

scene categorization. For the following experiments, all parameters of the implementation based on [OT06] were set to the default values which yield a total of 320 values for a grayscale-image.

5.2. Analysis and Results

5.2.1. Similarity matrices

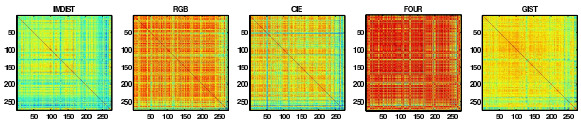

Figure 6: *Similarity matrices for the five computational measures.*

Figure 6 shows the similarity matrices for the five computational measures. It is very obvious that all features measure image properties that give rise to very different similarity patterns - see Figure 5 for comparison. We again performed MDS on the matrices in order to see how many dimensions would be necessary to explain the similarity data - these were IMDIST(12),RGB(12),CIE(18),FOUR(10),GIST(14). The overall number of dimensions needed is therefore much lower than that of the human data.

As Table 2 shows, the first 6 dimensions of most computational measures (as far as they were interpretable) extract low-level properties of the artworks such as overall brightness, or large color differences. Fourier analysis - as could be expected - also is sensitive to texture. GIST, however, differs from the other measures as it specifically focuses on scale properties of the images. More specifically, dimension 4 corresponds to a higher-level dimension as it separates images with flat appearance from those with more depth structure - nevertheless, the separation not as clear as for dimension 2 of the human data. Similarly, even though dimension 5 of the GIST measure distinguishes fine-grained from coarse-grained paintings, it does not do this fully consistently along art periods. Apart from this exception, none of the other computational measures thus correlate with the human data. This could have several reasons:

- our measures left out some important low-level correlate of human judgements
- no single measure can correlate with human data, multiple measures need to be taken into account
- humans also use higher-level properties of paintings or semantic knowledge to do the task

Although of course a host of algorithms measuring low-level image properties exist, the fact that neither texture nor color-based, nor scale-sensitive measures correlate at any dimension casts doubt on whether another measure will do much better. The second point also seems unlikely given that the dimensions for the human data clearly correspond to higher-level image properties - additionally, even higher dimensions in the computational data never explicitly capture such properties explicitly. In our opinion, participants therefore clearly used higher-level properties of the paintings which are currently beyond what our computational measures can extract - and most likely those developed by other authors as well.

5.2.2. Clustering analyses

Not surprisingly, the clustering analysis shows a similar result. None of the clusterings groups corresponding artists consistently together and correlations with human data are non-existent. We also tried different clustering approaches (such as different variations of k-means) directly on the computational measures, again with no better result.

6. Conclusion and Outlook

Our study has shown that non-experts were able to reliably group unfamiliar paintings of many artists into meaningful categories. Dimension and clustering analyses have shown that this was done based on both low-level, more perceptual (brush stroke, perspective) and mid-level, more cognitive information (genre, motif) resulting in historically correct categories. The most salient difference includes a clear category

break between pre-Impressionist and post-Impressionist art which corresponds well to the beginnings of modern art in art historical terms. Little difference was observed when restricting the number of categories versus free grouping showing that the underlying processes in both tasks were highly similar. The aesthetics of non-expert viewers thus leads to surprisingly "canon-conform" categorizations.

Of course we realize that our study is only the first step towards a full mapping of the perceptual and cognitive categories of art. More specifically, our particular choice of paintings - even though a random sampling from the online databases - might have introduced certain biases in the grouping. We are currently re-running the experiments with a different, disjoint set of images from the same artists and art periods to cross-validate the grouping results and also to get a better measure of the inter-subject variance for this task. Furthermore, as we asked participants to group based on the style of the image, it would be interesting also to group by different criteria such as early/late art, genre (stilllife, portrait) in order to get perceptual grouping results for these properties as well. Additionally, we are planning comparisons with expert-judgments of our stimulus set which also will give a better annotation and validation of our database. Finally, rapid categorization tasks as in [TFM96] will allow us to look at perceptual processes in isolation as the speeded task prevents cognitive influences.

Several computational measures sensitive to color, texture, and spatial composition were implemented in order to seek low-level correlates with the human data. Both in terms of dimension and clustering analyses results, none of the computational measures - with the notable exception of the GIST feature - came close to modelling the human data. In our opinion, this emphasizes the higher-level processing that even non-expert viewers make when viewing and interpreting works of art. Nevertheless, we want to stress that our computational studies were only the beginning of a larger set of experiments, for which we also plan to implement more complex types of measures. Given the success of the GIST analysis, we will test more sophisticated features based on automatic analysis of the depth structure [HSEH07] of paintings which might yield more insights.

References

[Arn74] ARNHEIM R.: *Art and visual perception: a psychology of the creative eye: the new version*. U of California Press, 1974.

[BG05] BORG I., GROENEN P.: *Modern multidimensional scaling*, 2nd ed. Springer, 2005.

[CHL05] CUTZU F., HAMMOUD R., LEYKIN A.: Distinguishing paintings from photographs. *Comput. Vis. Image Underst. 100*, 3 (2005), 249–273.

[DJLW06] DATTA R., JOSHI D., LI J., WANG J. Z.: Studying aesthetics in photographic images using a computational approach. In *ECCV (3)* (2006), pp. 288–301.

[Har05] HARNAD S.: Cognition is categorization, July 2005. Paper presented at UQˆM Summer Institute in Cognitive Sciences on Categorisation 2003.

[HSEH07] HOIEM D., STEIN A., EFROS A., HEBERT M.: Recovering occlusion boundaries from a single image. In *Proceedings ICCV* (2007).

[LCR07] LESLIE L., CHUA T.-S., RAMESH J.: Annotation of paintings with high-level semantic concepts using transductive inference and ontology-based concept disambiguation. In *MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia* (New York, NY, USA, 2007), ACM, pp. 443–452.

[Ley07] LEYTON M.: *The Structure of Paintings*. Springer, 2007.

[Lit04] LITTLE S.: *Isms: Understanding Art*. Universe Publishing, 2004.

[Liv02] LIVINGSTONE M.: *Vision and Art. The Biology of Seeing*. Harry N. Abrams, NY, 2002.

[MCJ07] MARCHENKO Y., CHUA T.-S., JAIN R.: Ontology-based annotation of paintings using transductive inference framework. In *13th International Multimedia Modeling Conference* (2007), pp. 13–23.

[MTSI05] MARCHENKO Y., TAT-SENG C., IRINA A.: Analysis and retrieval of paintings using artistic color concepts. *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on* (2005), 1246–1249.

[OT06] OLIVA A., TORRALBA A.: Building the gist of a scene: the role of global image features in recognition. *Visual Perception: Fundamentals of Awareness, Multi-Sensory Integration and High-Order Perception* (2006).

[Red07] REDIES C.: A universal model of esthetic perception based on the sensory coding of natural stimuli. *Spatial Vision 21* (2007), 97–117.

[SC73] SHEPARD R., CERMAK G.: Perceptual-cognitive explorations of a toroidal set of free-form stimuli. *Cognit Psychol 4* (1973), 351–377.

[Slo03] SLOUTSKY V.: The role of similarity in the development of categorization. *Trends Cogn. Sci. 7* (Jun 2003), 246–251.

[SO01] SIMONCELLI E., OLSHAUSEN B.: Natural image statistics and neural representation. *Annu. Rev. Neurosci. 24* (2001), 1193–1216.

[Spa06] Spatial vision - art and perception: Towards a visual science of art, part 1, 2006.

[Spa07] Spatial vision - art and perception: Towards a visual science of art, part 2, 2007.

[Spi07] SPILLMANN L.: Artists and vision scientists can learn a lot from each other - but do they? *Gestalt Theory - An International Multidisciplinary Journal 29* (2007).

[TFM96] THORPE S., FIZE D., MARLOT C.: Speed of processing in the human visual system. *Nature 381* (1996), 520Ŭ–522.

[WKK*07] WALLRAVEN C., KAULARD K., KÜRNER C., PEPPERELL R., BÜLTHOFF H. H.: In the eye of the beholder: Perception of indeterminate art. In *International Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging* (06 2007), pp. 121–128.

[Zek99] ZEKI S.: *Inner vision: an exploration of art and the brain*. Oxford University Press, 1999.