

Pictorial Research: Objective versus Subjective Annotations on Avercamp and Rembrandt

M. W. A. Wijntjes¹ and M. J. P. van Zuijlen¹

¹Perceptual Intelligence Lab, Delft University of Technology, Landbergstraat 15, 2628CE Delft, The Netherlands
E-mail: M.W.A.Wijntjes@tudelft.nl

Abstract. Pictorial research can rely on computational or human annotations. Computational annotations offer scalability, facilitating so-called distant-viewing studies. On the other hand, human annotations provide insights into individual differences, judgments of subjective nature. In this study, we demonstrate the difference in objective and subjective human annotations in two pictorial research studies: one focusing on Avercamp's perspective choices and the other on Rembrandt's compositional choices. In the first experiment, we investigated perspective handling by the Dutch painter Hendrick Avercamp. Using visual annotations of human figures and horizons, we could reconstruct the virtual viewpoint from where Avercamp depicted his landscapes. Results revealed an interesting trend: with increasing age, Avercamp lowered his viewpoint. In the second experiment, we studied the compositional choice that Rembrandt van Rijn made in *Syndics of the Drapers' Guild*. Based on imaging studies it is known that Rembrandt doubted where to place the servant, and we let 100 annotators make the same choice. Subjective data was in line with evidence from imaging studies. Aside from having their own merit, the two experiments demonstrate two distinctive ways of performing pictorial research, one that concerns the picture alone (objective) and one that concerns the relation between the picture and the viewer (subjective).

Keywords: vision and art, perspective, composition, depth, annotations, distant viewing

© 2024 Society for Imaging Science and Technology.
[DOI: 10.2352/J.Percept.Imaging.2024.7.000501]

1. INTRODUCTION

A variety of scientific disciplines are involved in pictorial research, from perception psychology and computer science to (digital) art history. Recently, the art historical study of pictures has shifted from close viewing, the study of individual pictures, towards distant viewing [1, 2], where a large body of pictures are analyzed using computational methods. This computational approach affords cultural analytics [26]: quantifying visual trends and structures in large corpora of images. The data used for these analyses are computational annotations [2], which can range from image statistics [16] and head posture [9, 43] to many other levels of visual representation.

There is thus a paradigm shift from the human eye studying a single work to a computational eye studying (large) collections of works. This paradigm shift takes place along two dimensions, from humans to computers and from

close viewing to distant viewing. Although the convenience of computational scalability likely inspired distant viewing, the two dimensions are independent: it is possible to perform close viewing with computers [20] or distant viewing with humans [42]. However, there is one dimension to be added to draw a more complete picture of possible paradigms for pictorial research: the difference between subjective and objective annotations, as illustrated in Figure 1. While computational annotations are mostly deterministic and display noise that is only attributable to their performance accuracy, the variations in human annotations can be of various nature: aside from accuracy, it can also reflect individual differences.

To explore this distinction between objective and subjective annotations, we will first review existing work in these two fields and then present two case studies that exemplify an objective and a subjective annotation experiment.

1.1 Objective Annotations

The focus of an objective annotation study lies primarily in the picture. It uses the human as an (intelligent) measurement device, also known as human computation [14]. Many visual annotation studies come from the field of computer vision and aim at creating ground truth data for machine learning. A well-known (by now historical) example is LabelMe [36], which aimed at object labeling and polygonal segmentation (where the object is manually outlined). A crowdsourcing platform that has often been used is Amazon Mechanical Turk (AMT). Segmentation and labeling are now default options on AMT, but at the time the LabelMe platform facilitated many other segmentation and annotation studies. Besides plain segmentation, a considerable number of “richly annotated databases” have been created. These studies often involve not only the labeling/manipulation of 3D data such as position [11], 3D meaning of sketch lines [17], and 3D surface attitude [8] but also material reflectance [3] and shadows [24].

Computer vision and graphics are not the only fields making use of objective annotations. Museums annotate artifacts with information about artworks, such as artist name, year, medium, provenance, etc., namely, the metadata. This labeling was traditionally performed by art historians, but over the past two decades various digital tools have been developed [19, 37]. These tools can be used not only by

Received July 18, 2023; accepted for publication Apr. 1, 2024; published online June 28, 2024. Associate Editor: Michael A. Webster.

2575-8144/2024/7/000501/6/\$00.00

professionals but also through crowdsourcing such as the Art UK painting tagger [13]. Furthermore, playful interactions have been designed to motivate volunteers to participate in enriching metadata [45]. Although many projects aim at non-spatial tagging, for instance IconClass [4] labels, there is also a need for spatial (e.g. a bounding box) annotations [38]. Collecting bounding box data on a large scale can lead to interesting insights into digital art history such as spatial distributions of material depictions [42]. While computer vision also started with simple bounding boxes and evolved towards more complex annotations such as those in [14], the same trend has started to emerge in art history. For example, in [46], the authors added annotations of people, shadows, and highlights, which allowed for the analysis of stylistic trends such as perspective viewpoint and light direction.

Although there are obvious differences in rationale between computer vision/graphics and digital art history annotations, they share their focus on the picture. Insofar as the annotator plays a role, it concerns accuracy [25, 40] or expert versus crowd [31]. In the next section, we will review crowdsourced perception research where the picture has a totally different role, i.e. that of stimulus eliciting certain perceptions and appreciations.

1.2 Subjective Annotations

Collecting data through online crowdsourcing experiments has become the standard research practice for the behavioral sciences. There are various possible motivations to choose online over lab experiments, such as speed and efficiency, but also more specific reasons such as access to certain subject pools [6, 32]. Performance seems generally not to be degraded [5] although participants sometimes lack attention [15]. Most paradigms from experimental psychology give similar results when conducted online as compared to lab experiments [10, 18].

Without reviewing the whole perception literature, we will compare some annotation studies to perception research. Categorizations (labeling) in objective annotation research [36] often serve as “ground truth” and need to be unambiguous. In perception studies, object categorizations may rely on various independent variables such as background [12], and research is particularly interested in ambiguity [30]. Computer vision research on 3D inference from images as reviewed above [8, 11, 17] seems to ignore abundant evidence that these annotations are perceptually ambiguous [22, 23, 39]. The 3D inference relies on how the picture (or its actual representation) is viewed [21, 47, 48]. Furthermore, the material probe used by [3], which let observers manipulate gloss parameters, is well known in the perception literature [33]. It seems that many objective annotation studies assume perfect perceptual constancy (i.e. perceptual invariance with respect to extrinsic changes such as light, shape, or viewpoint). However, the aforementioned gloss perception can hardly be considered perceptually constant [7, 27, 29, 44].

One could argue that all perception experiments are subjective annotations of the stimuli used. Yet the goal of most perception research is to generalize across stimuli, which renders the annotation meaningless. Still, some

pictures have meaning beyond the perception experiment because they are also studied by other scientific disciplines such as media studies or art history. For example, Yarbus [50] measured the eye movements elicited by “Unexpected Visitors” by Ilya Repin. Yarbus’s goal was to measure the effect of different instructions (e.g. how wealthy the family is, the activity, describe clothes, etc.), but at the same time the eye movements tell us something about how that specific painting of Ilya Repin is viewed. This can be seen as a subjective annotation. To move forward the discussion of objective and subjective annotations, we now present two experiments and continue the discussion afterwards.

2. EXPERIMENT 1: OBJECTIVE ANNOTATION

One of the many elements of artistic style is perspective, i.e. how pictorial space is constructed. Radically different projection systems can be found across art history [49]. Also more subtle patterns have been found such as changes in viewpoint elevation [46]. In this experiment, we used a similar paradigm for tracing the horizon and human figures based on the principle of horizon ratio [35]. We choose to analyze the paintings of Hendrick Avercamp, a Dutch painter from the early 17th century. He is particularly famous for his winter landscapes, often depicting people ice-skating, which happen to be particularly suited for perspective reconstruction because ice is a horizontal plane.

2.1 Procedure

We used the platform Amazon Mechanical Turk to recruit and reimburse participants. We primarily used p5.js to program the visual interaction. The name p5 originates from “Proce55ing,” an alternative spelling of “Processing,” which is a widely used programming language [34] “in the context of visual arts” (<https://processing.org>). P5.js is a JavaScript library that shares much of the functionality of Processing [28]. Both Processing and p5.js aim to make code accessible to a wide audience and are thus relatively easy to use by beginners. A program created by a user is called a sketch, emphasizing the iterative design process with immediate visual feedback. We found it works well in many different kinds of annotation experiments.

Participants were instructed to first annotate where they saw the horizon by adjusting a red horizontal line. Then they had to draw lines between the feet and the head for about 10 to 15 people in the scene. Participants were encouraged to choose people who were at various distances, including those far away (and tiny on the screen).

2.2 Participants

The experiment was split into three blocks for which each 10 participants were recruited.

2.3 Data Analysis

A linear regression of the vertical distance between the feet and the horizon as a function of human sizes was performed. The slope of the model has a direct meaning: it is the height of the perspective viewpoint in terms of human sizes. For

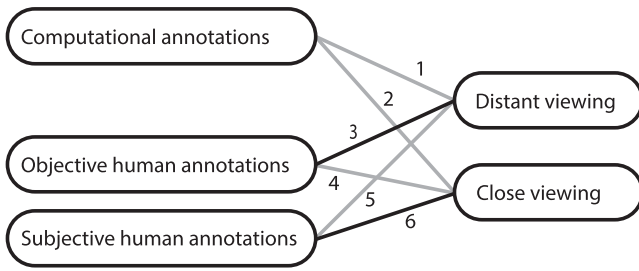
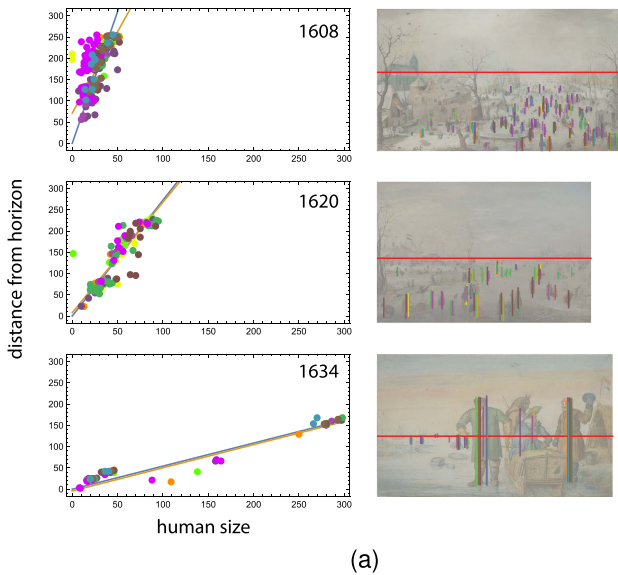


Figure 1. Various types of pictorial research. 1. Distant viewing with computers deals with analyzing connections within collections of images based on computational image analysis. 2. Close viewing with computational help is related to the study of a single picture, for example, an optical reconstruction. 3. Using objective human annotations for distant viewing analysis is related to the experiment on Avercamp in the current study. 4. Close viewing using objective human annotations could for example also be related to drawing perspective lines. 5. Subjective annotations for distant viewing could for example be beauty or preference ratings. 6. Subjective annotations for close viewing are exemplified by the study on Rembrandt in the current paper.

example, if the slope is 2, the viewpoint of the painter was (virtually) 2 human lengths (approximately 3.5 m) high. We not only set the offset to zero for the actual elevation data but also used it as a free parameter and found little difference (blue and orange lines in Figure 2).

2.4 Pictures

We first downloaded all Avercamp paintings we could find from WikiArt. Then we selected 33 that showed ice (ensuring a flat horizontal ground) and multiple people. We used the WikiArt metadata for production years.



2.5 Results and Discussion

In Figures 2 and 3, representative examples are displayed on the left and the overall perspective elevation over time is shown on the right. Before discussing these results, it should be reported that horizon estimation turned out to be rather fluctuating and seemed to be met with some confusion. Apparently, the concept of horizon is very clear to the authors, but not to the general AMT worker, or at least not in the way it was explained in the experiment. For further analysis, we choose one representative participant whose horizon data were used for further computations. Note that we did use the complete set of human figure annotations of all participants; only the horizon line annotation was taken from a single observer.

The negative slope (a decrease of 10% human lengths per year) was highly significant ($t(32) = -4.84405, p < 0.0001$). Furthermore, the coefficient of determination was moderate ($R^2 = 0.43$) indicating that although the trend itself is significant, it only explains 43% of the variance in perspective attitudes found in the 33 pictures.

The results imply that Avercamp lowered his viewpoint as he became older. The effect is rather strong, and we did not anticipate this effect. Interestingly, the same trend is found for Canaletto [46], an Italian painter who was active a century after Avercamp. As these two painters seem unrelated, it raises the question whether there is a general trend among artists when it comes to perspective elevation and age.

3. EXPERIMENT 2: SUBJECTIVE ANNOTATION

In the second experiment, we aim to answer a question of *artistic composition*. *Syndics of the Drapers' Guild* by Rembrandt is an intriguing painting; the unconventional viewpoint; the feeling of interrupting a meeting that just

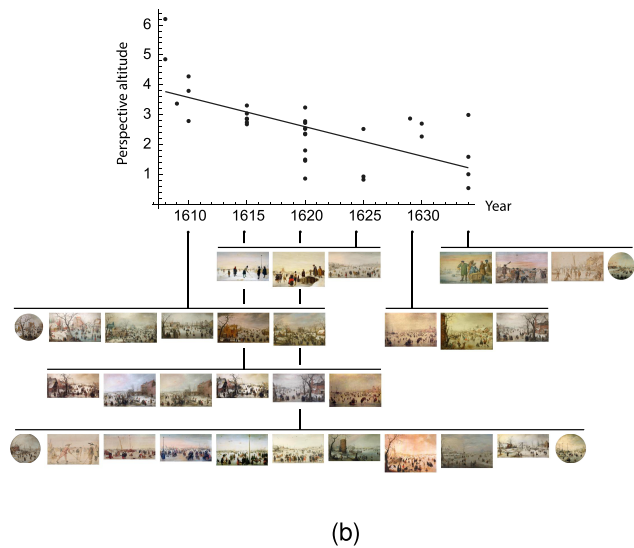


Figure 2. (a) Three examples of data. The (reduced contrast) paintings are overlaid with raw annotation data. Different colors denote different annotators. The slope in scatterplots indicates elevation of the viewpoint. (b) Here, viewpoint elevations are plotted over time. All images are shown under years. As can be seen, a negative trend is clearly visible.

Instructions: Rembrandt was in doubt where to put this sixth person. We are interested in where you think that person should be. Position him with the mouse to create the best composition. After clicking you can submit the HIT.

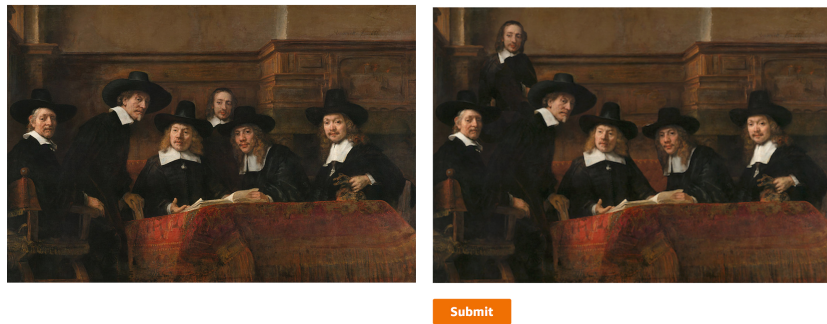


Figure 3. On the left, the original painting is shown: *Syndics of the Drapers' Guild* (1662) by Rembrandt. On the right, the experimental interface is shown. As can be seen, the servant is cut out of the original and can be freely positioned anywhere on the canvas. Painting reproduced from The Rijksmuseum, The Netherlands, under a CC 0.0 license.

started. Moreover, x-ray studies have shown [41] that Rembrandt had doubts on where to position the servant, the person in the middle behind the others. According to [41], Rembrandt initially planned to position this person at the far right of the painting. Informal conversations with art historians led us to an interesting question: Where would a naive viewer position the servant, and how does this compare to Rembrandt's choice and rejection?

3.1 Participants

A total of 100 participants were recruited on AMT.

3.2 Procedure

Participants were instructed to position the servant at the location that they believed resulted in the best composition. The experiment is shown in Fig. 3. The initial position of the servant depended on where the user's mouse was at that moment, and was thus not actively randomized.

3.3 Picture

On the left side of Fig. 3, the original (unedited) image of "Syndics of the Drapers" Guild is shown. We used photo-editing software to cut out the foreground scene and the servant. The place of the servant was filled with image elements of the remaining scene so as to not give away Rembrandt's choice.

3.4 Results and Discussion

The compositional preferences of 100 participants are visualized in Figure 4. It can be readily seen that four horizontal locations dominate the data. One of them is similar to the actual painting, and the far right alternative is similar to Rembrandt's underdrawing, i.e. the initial "sketch" [41]. Perhaps, the other two locations have also been considered by Rembrandt.

While the preferred horizontal positions seem to be discrete, the vertical preference is more continuous. The range of positions seems to be limited by what is anatomically and ergonomically plausible: somewhere between a sitting and standing position. It is interesting to consider what



Figure 4. Visualization of 100 responses where the servant was placed by participants. There are four dominant horizontal positions with more or less equal probability except the one on the left.

effect these horizontal and vertical variations have on the perception of the narrative.

4. GENERAL DISCUSSION

We presented two visual experiments that demonstrate the difference between objective and subjective annotations in the context of close and distant viewing. Both studies make use of spatial annotations, and both concern artistic choices of composition. Yet the perspective study relies on objective annotations while the composition study relies on subjectivity. What makes the two results categorically different, and how do they relate to existing pictorial research?

Individual differences are the core difference between data interpretation of objective versus subjective annotations. In both experiments, the data contains substantial variance: in Fig. 2, the linear regression plots show a certain noise level and in Fig. 4, we see continuous distributions along the vertical direction and discrete, multi-modal distributions along the horizontal direction. What do these

variances mean? In both cases, part of the noise emerges from human inaccuracies. This inherent inaccuracy of human judgment can be interesting in a purely behavioral context but is considered as irrelevant measurement noise in our context. However, this inherent noise does not seem responsible for all variances. In the first study, part of the variance can be attributed to the pictorial content. In Fig. 2(a), the residuals of the linear regressions denote how homogeneously the figure sizes diminished towards the horizon. Deviations from the regression line are either due to natural variations of human sizes or inaccurate perspective handling. Although we cannot dissociate between these two, we should realize that both explanations concern the depiction and not the viewer. In Fig. 2(b), the residuals refer to differences in viewpoint. We found that although the downward trend with time was significant, still 57% of the variance was unexplained by this trend. Again, this can be due to various reasons but they all concern the circumstances under which the picture was produced, e.g. artistic or practical choices for viewpoint difference. The variance in the data does not concern the observer.

On the other hand, Experiment 2 concerns the relation between the picture and the viewer. The variation in preferences shows clear clusters that arise from multiple individual judgments. The annotations are characterized by the individual differences of the observers. They reveal the different artistic choices that could have been made. Other than the viewpoint attitude of Avercamp, Syndics of the Drapers' Guild affords immediate visualization of a compositional choice; manually varying the perspective attitude would be rather complex. The possibilities for these subjective annotations thus depend on the type of annotations and the pictures.

A painting (collection) can be visually analyzed by humans and computers. Studying a single work can be called close viewing while the comparison between works is called distant viewing [2]. While many digital art historians use computer annotations, our study was based on human annotations. The advantage of computer annotations is the scalability. The advantage of humans is that they can be relatively easily instructed and that they display individual differences, which are not to be found in computational approaches. Unfortunately, the two experiments reported here span only a relatively small space of what is possible when we combine perception experiments with the study of pictures. Nearly every existing visual paradigm can be translated towards a depiction-centered instead of a vision-centered research question. It seems an exciting new research area that appears rather timely: online experiments are mainstream, and online art collections grow every year. Furthermore, we are convinced that meaningfully pictorial research cannot do without human annotations, however advanced computer annotations may become.

The two experiments we conducted investigate pictorial characteristics of paintings that are often cataloged in databases. The (computational) analysis of image corpora can make use of both image data and metadata. The latter

is created through annotations by art professionals and mostly concern objective, non-visual information such as the maker, medium, and provenance. Yet, many museum collections have been using their own archival methods, making comparison across collections challenging. This will change in the near future as initiatives like the International Image Interoperability Framework [38] start to gain traction. These new methods allow for a wider range of metadata such as user-generated annotations. In this perspective, any type of pictorial research related perception experiment can be added to these metadata. Being aware of other perception studies or annotations conducted on an image or image collection could result in unprecedentedly linked pictorial research.

ACKNOWLEDGMENT

This work is part of the research program Visual Communication of Material Properties with project number 276-54-001, which is financed by the Dutch Research Council (NWO). Furthermore, the authors would like to thank Prof. Dr. Joris Dik, who introduced them to many intriguing art historical questions of which the Drapers' Guild is merely one.

REFERENCES

- 1 T. Arnold and L. Tilton, "Distant viewing: analyzing large visual corpora," *Digit. Scholarship Humanit.* **34**, i3–i16 (2019).
- 2 T. Arnold and L. Tilton, *Distant Viewing: Computational Exploration of Digital Images* (MIT Press, Cambridge, MA, 2023).
- 3 S. Bell, P. Upchurch, N. Snavely, and K. Bala, "OpenSurfaces: A richly annotated catalog of surface appearance," *ACM Trans. Graph.* **32**, 1–17 (2013).
- 4 H. Brandhorst, "Aby Warburg's wildest dreams come true?," *Vis. Res.* **29**, 72–88 (2013).
- 5 M. Buhrmester, T. Kwang, and S. D. Gosling, "Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?," *Perspect. Psychol. Sci.* **6**, 3–5 (2011).
- 6 M. Buhrmester, T. Kwang, and S. D. Gosling, "Amazon's mechanical Turk: A new source of inexpensive, yet high-quality data?," *Perspect. Psychol. Sci.* **6**, 3–5 (2011).
- 7 A. C. Chadwick and R. Kentridge, "The perception of gloss: A review," *Vis. Res.* **109**, 221–235 (2015).
- 8 W. Chen, S. Qian, D. Fan, N. Kojima, M. Hamilton, and J. Deng, "Oasis: A large-scale dataset for single image 3D in the wild," *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Piscataway, NJ, 2020).
- 9 J.-P. Chou and D. G. Stork, "Computational tracking of head pose through 500 years of fine-art portraiture," *Electron. Imaging* **35**, 1–13 (2023).
- 10 M. J. Crump, J. V. McDonnell, and T. M. Gureckis, "Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research," *PLoS ONE* **8** (2013).
- 11 A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner, "ScanNet: Richly-annotated 3D reconstructions of indoor scenes," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2017), pp. 5828–5839.
- 12 J. L. Davenport and M. C. Potter, "Scene consistency in object and background perception," *Psychol. Sci.* **15**, 559–564 (2004).
- 13 K. Eccles and A. Greg, "Your paintings tagger: Crowdsourcing descriptive metadata for a national virtual collection," *Crowdsourcing our Cultural Heritage* (Routledge, 2016), pp. 185–208.
- 14 Y. Gingold, A. Shamir, and D. Cohen-Or, "Micro perceptual human computation for visual tasks," *ACM Trans. Graph.* **31**, 1–12 (2012).

- 15 J. K. Goodman, C. E. Cryder, and A. Cheema, "Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples," *J. Behav. Decis. Mak.* **26**, 213–224 (2013).
- 16 D. J. Graham and C. Redies, "Statistical regularities in art: Relations with visual coding and perception," *Vis. Res.* **50**, 1503–1509 (2010).
- 17 Y. Gryaditskaya, M. Sypsteyn, J. W. Hoftijzer, S. Pont, F. Durand, and A. Bousseau, "OpenSketch: A richly-annotated dataset of product design sketches," *ACM Trans. Graph.* **38**, 232 (2019).
- 18 S. Haghiri, P. Rubisch, R. Geirhos, F. Wichmann, and U. von Luxburg, "Comparison-based framework for psychophysics: Lab versus crowd-sourcing," Preprint, arXiv:1905.07234 (2019).
- 19 L. Hollink, A. T. Schreiber, J. Wielemaker, and B. J. Wielinga, "Semantic annotation of image collections," *Knowledge Capture 2003—Proc. Knowledge Markup and Semantic Annotation Workshop* (2003), Vol. 2, pp. 41–48.
- 20 M. K. Johnson, D. G. Stork, S. Biswas, and Y. Furuichi, "Inferring illumination direction estimated from disparate sources in paintings: an investigation into Jan Vermeer's girl with a pearl earring," *Proc. SPIE* **6810**, 164–175 (2008).
- 21 J. Koenderink, M. Wijntjes, and A. van Doorn, "Zograscopic viewing," *i-Perception* **4**, 192–206 (2013).
- 22 J. J. Koenderink, A. J. van Doorn, and A. M. L. Kappers, "Surface perception in pictures," *Perception Psychophysics* **52**, 487–496 (1992).
- 23 J. J. Koenderink, A. J. van Doorn, A. M. L. Kappers, and J. T. Todd, "Ambiguity and the 'mental eye' in pictorial relief," *Perception* **30**, 431–448 (2001).
- 24 B. Kovacs, S. Bell, N. Snavely, and K. Bala, "Shading annotations in the wild," *Computer Vision and Pattern Recognition (CVPR)* (IEEE, Piscataway, NJ, 2017).
- 25 S. Liu, C. Chen, Y. Lu, F. Ouyang, and B. Wang, "An interactive method to improve crowdsourced annotations," *IEEE Trans. Vis. Comput. Graphics* **25**, 235–245 (2018).
- 26 L. Manovich, *Cultural Analytics* (MIT Press, Cambridge, MA, 2020).
- 27 P. J. Marlow, J. Kim, and B. L. Anderson, "The perception and misperception of specular surface reflectance," *Curr. Biol.* **22**, 1909–1913 (2012).
- 28 L. McCarthy, C. Reas, and B. Fry, "Getting Started with P5.js: Making Interactive Graphics in JavaScript and Processing," (Maker Media, Inc., San Francisco, 2015).
- 29 I. Motoyoshi and H. Matoba, "Variability in constancy of the perceived surface reflectance across different illumination statistics," *Vis. Res.* **53**, 30–39 (2012).
- 30 C. Muth, V. M. Hesslinger, and C.-C. Carbon, "The appeal of challenge in the perception of art: How ambiguity, solvability of ambiguity, and the opportunity for insight affect appreciation," *Psychol. Aesthetics, Creat., Arts* **9**, 206 (2015).
- 31 J. Oosterman, A. Bozzon, G. J. Houben, A. Nottamkandath, C. Dijkshoorn, L. Aroyo, M. H. Leyssen, and M. C. Traub, "Crowd versus experts: nichesourcing for knowledge intensive tasks in cultural heritage," *Proc. 23rd Int'l. Conf. World Wide Web* (2014), pp. 567–568.
- 32 G. Paolacci, J. Chandler, and P. G. Ipeirotis, "Running experiments on Amazon Mechanical Turk," *Judgment Decis. Mak.* **5**, 411–419 (2010).
- 33 F. Pellacini, J. A. Ferwerda, and D. P. Greenberg, "Toward a psychophysically-based light reflection model for image synthesis," *Proc. 27th Annual Conf. on Computer Graphics and Interactive Techniques* (ACM Press/Addison-Wesley Publishing Co., USA, 2000), pp. 55–64.
- 34 C. Reas and B. Fry, *Processing (A Programming Handbook for Visual Designers and Artists)* (The MIT Press, Cambridge, MA, 2007).
- 35 S. Rogers, "The horizon-ratio relation as information for relative size in pictures," *Percept. & Psychophys.* **58**, 142–152 (1996).
- 36 B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: a database and web-based tool for image annotation," *Int. J. Comput. Vis.* **77**, 157–173 (2008).
- 37 G. Schreiber, A. Amin, L. Aroyo, M. van Assem, V. de Boer, L. Hardman, M. Hildebrand, B. Omelayenko, J. van Osenbruggen, A. Tordai, J. Wielemaker, and B. Wielinga, "Semantic annotation and search of cultural-heritage collections: The multimediaN E-culture demonstrator," *J. Web Semant.* **6**, 243–249 (2008).
- 38 S. Snyderman, R. Sanderson, and T. Cramer, "The international image interoperability framework (IIIF): A community & technology approach for web-based images," *Proc. IS&T Archiving 2015* (IS&T, Springfield, 2015), Vol. 2015, pp. 16–21.
- 39 J. T. Todd, A. H. Oomes, J. J. Koenderink, and A. M. Kappers, "On the affine structure of perceptual space," *Psychol. Sci.* **12**, 191–196 (2001).
- 40 P. Upchurch, D. Sedra, A. Mullen, H. Hirsh, and K. Bala, "Interactive consensus agreement games for labeling images," *Proc. AAAI Conf. on Human Computation and Crowdsourcing* (2016), Vol. 4, pp. 239–248.
- 41 A. Van Schendel, "De schimmen van de Staalmeesters: Een röntgenologisch onderzoek," *Oud Holland* **71**, 1–23 (1956).
- 42 M. J. van Zuijlen, H. Lin, K. Bala, S. C. Pont, and M. W. Wijntjes, "Materials in paintings (MIP): An interdisciplinary dataset for perception, art history, and computer vision," *Plos One* **16**, e0255109 (2021).
- 43 M. J. van Zuijlen, S. C. Pont, and M. W. Wijntjes, "Conventions and temporal differences in painted faces: A study of posture and color distribution," *Electron. Imaging* **2020**, 267-1 (2020).
- 44 G. Wendt, F. Faul, V. Ekroll, and R. Mausfeld, "Disparity, motion, and color information improve gloss constancy performance," *J. Vis.* **10**, 7–7 (2010).
- 45 C. Wieser, F. Bry, A. Bérard, and R. Lagrange, "ARTigo: building an artwork search engine with games and higher-order latent semantic analysis," *Proc. AAAI Conf. on Human Computation and Crowdsourcing* (2013), Vol. 1, pp. 15–20.
- 46 M. W. A. Wijntjes, "Shadows, highlights and faces: the contribution of a 'human in the loop' to digital art history," *Art & Perception* **9**, 66–89 (2021).
- 47 M. W. A. Wijntjes, "A new view through Alberti's window," *J. Exp. Psychol.: Hum. Perception Perform.* **40**, 488 (2014).
- 48 M. W. A. Wijntjes and S. C. Pont, "Pointing in pictorial space: Quantifying the perceived relative depth structure in mono and stereo images of natural scenes," *ACM Trans. Appl. Perception (TAP)* **7**, 1–8 (2010).
- 49 J. Willats, *Art and Representation: New Principles in the Analysis of Pictures* (Princeton University Press, Princeton, 1997).
- 50 A. L. Yarbus, *Eye Movements and Vision* (Springer, 1967/2013).