Vision Research xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Statistical correlates of perceived gloss in natural images

Christiane B. Wiebel, Matteo Toscani, Karl R. Gegenfurtner *

Abteilung Allgemeine Psychologie, Justus-Liebig-Universität Giessen, Otto-Behaghel-Str. 10F, 35394 Giessen, Germany

ARTICLE INFO

Article history:
Received 1 August 2014
Received in revised form 16 April 2015
Available online xxxx

Keywords:
Material perception
Gloss perception
Natural images
Image statistics

ABSTRACT

It is currently debated whether the perception of gloss is linked to the statistical parameters of the retinal image. In particular, it has been suggested that gloss is highly correlated with the skewness of the luminance histogram. However, other psychophysical work with artificial stimuli has shown that skewness alone is not enough to induce the perception of gloss. Here, we analyzed many images of natural surfaces to search for potential statistical correlates of perceived gloss. We found that skewness indeed correlates with gloss when using rendered stimuli, but that the standard deviation, a measure of contrast, correlates better with perceived gloss when using photographs of natural surfaces. We verified the important role of contrast by manipulating skewness and contrast within images. Changing the contrast in images significantly modulates perceived gloss, but manipulating the skewness of the luminance histogram had only a small effect.

1. Introduction

Human observers are remarkably good at perceiving the material qualities of objects (for review, see Adelson, 2001; Anderson, 2011; Fleming, 2014) and at making fast and accurate judgments of material categories (Sharan, 2009; Sharan, Rosenholtz, & Adelson, 2009, 2014; Wiebel, Valsecchi, & Gegenfurtner, 2013). This is true even though the underlying physical processes, including ray optics and differential geometry, can be quite complex (Blake & Bülthoff, 1990). Disentangling the different sources of information that define the appearance of a surface is not trivial at all, as many different combinations of the underlying factors can generate one and the same image. Therefore, material perception seems to be a highly challenging task for the visual system (see for example, Anderson, 2011; Fleming, 2014; Thompson et al., 2011). It is frequently assumed that certain statistical regularities of the visual input co-vary with certain surface properties, being diagnostic for material appearances. Thus, the extraction of simple image statistics has been suggested as a useful heuristic, resulting in appropriate information for the recovery of a surface's properties (see for example, Anderson, 2011; Fleming, 2012; Thompson et al., 2011; Toscani, Valsecchi, & Gegenfurtner, 2013a, 2013b).

The significance of image-based cues for the judgment of surface reflectance properties (e.g., glossiness) has been outlined in a series of previous studies on the topic (see for example Fleming, Dror, & Adelson, 2003; Motoyoshi et al., 2007; Nishida & Shinya, 1998; Sharan et al., 2008). Correlations between image-based cues and the appearance of glossiness received much attention following an article by Motoyoshi et al. (2007). The authors observed that the glossy appearance of a surface is often accompanied by a positively skewed luminance histogram across the surface. They showed that the skewness of the luminance histogram was directly related to observers' glossiness ratings, even when the skewness was digitally manipulated in the images. Several studies have since challenged this finding (Anderson & Kim, 2009; Kim & Anderson, 2010). Anderson and colleagues showed that skewness alone is not enough to induce a percept of glossiness (e.g. Anderson & Kim, 2009), even when dealing with surfaces that include highlights. For a surface to appear glossy, the specular reflections in an image need to be placed in appropriate positions (see also Beck & Prazdny, 1981). In a further study, Anderson and colleagues found that several perceptual measures, such as the perceived contrast, the sharpness of a highlight, and the size of a highlight, are better predictors of perceived gloss than skewness (Marlow, Kim, & Anderson, 2012).

Most of the previous work has been done with highly constrained sets of artificial surfaces that were mainly rendered by computer programs. The Motoyoshi et al. (2007) study did use photographs of real surfaces, but they used a very limited number of surfaces under restricted viewing conditions. Since real-world stimuli are quite cumbersome to control, most studies of gloss

* Corresponding author at: Justus-Liebig-Universität Gießen, Fachbereich 06, Psychologie und Sportwissenschaft, Abteilung Allgemeine Psychologie, Otto-Behaghel-Straße 10F, 35394 Gießen, Germany.
E-mail address: karl.gegenfurtner@psychol.uni-giessen.de (K.R. Gegenfurtner).

http://dx.doi.org/10.1016/j.visres.2015.04.010
0042-6989/

Please cite this article in press as: Wiebel, C. B., et al. Statistical correlates of perceived gloss in natural images. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.04.010

perception used stimuli rendered on a computer (but see Pont & Koenderink, 2005). In principle, modern rendering techniques are quite capable and allow generating images that are virtually indistinguishable from the real object. However, because setting up model parameters for new objects remains a time-consuming process, computer rendering has usually been done with a minimal set of stimuli that are simple to generate. Ferwerda, Pellacini, and Greenberg (2001) and Fleming, Dror, and Adelson (2003) used spheres, while others (e.g., Ho, Landy, & Maloney, 2008; Kim, Marlow, & Anderson, 2011) have used more general potato-like stimuli that look like bumpy spheres. In the real world, gloss can appear in a wide variety of different materials, such as metal or glass, and it can be due to a variety of different physical processes (see Hunter, 1937). Although these two different approaches – real physical stimuli and virtual, computer-rendered stimuli – have both proven valuable, we pursue a third approach here. Photographs of natural surfaces have the advantage that a large variety of surfaces is available and can be reasonably well controlled through image processing. Our goal was to investigate the role of simple image statistics – mainly skewness and standard deviation as a measure of contrast – in the perception of gloss, for photographs of a large variety of real surfaces.

We first ran an experiment to select groups of images appearing matte or glossy out of a large database of 1492 images. We employed computer classification techniques to investigate whether we could successfully separate these two groups of images using characteristic moments of the luminance distribution. We then applied the same approach to computer rendered objects. For the natural surfaces, the standard deviation had the highest discriminative power, while for the rendered objects skewness worked best. Finally, we ran two more psychophysical experiments where we systematically manipulated standard deviation and skewness. The results show that manipulations of the standard deviation of the images do have major effects on perceived glossiness. Changing the skewness had only a minor effect.

2. Experiment 1

2.1. Image selection

Studying the relationship between glossy appearances and image-based cues in real-world photographs requires choosing appropriate images. We used an empirical approach to determine the ground truth for the perception of gloss in natural surfaces, allowing us to collect a representative sample of images. We asked eight naïve observers to judge a collection of 1492 material images as glossy or matte. The images were obtained from different databases used in various earlier experiments on material perception. We used a rapid classification procedure that was successfully used in object categorization before (Thorpe, Fize & Marlot, 1996; Thorpe, Gegenfurtner, Fabre-Thorpe & BuÈlthoff, 2001). The brief presentations allow accurate classification, while at the same time encouraging fast and implicit judgments. The images consistently judged as glossy or matte were selected as the basis for this study.

2.2. Methods

2.2.1. Observers

Eight observers participated in the experiment. All observers had normal or corrected to normal visual acuity. All observers gave written informed consent in agreement with the local ethics committee and in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

2.2.2. Stimuli and Apparatus

1492 images were presented in the experiment. All of the images had been used in earlier studies on material perception, ensuring that the focus of the images was on the material surfaces being photographed. The images came from 10 different real-world material categories, were taken in different contexts and under different illumination conditions, making for a high variability of material appearances in our image set. 84 images were taken from the material samples used in the visuo-haptic studies of Baumgartner, Wiebel, and Gegenfurtner (2013, 2015), all of which were photographed under standardized laboratory lighting. 320 images were used in Wiebel, Valsecchi, and Gegenfurtner (2013) and were taken using a Nikon D70 camera (Nikon, Tokyo, Japan) under various indoor and outdoor illumination conditions. In addition, we used the 500 close-up images from the Flickr material database established by Sharan et al. (2009) and the 588 images collected from various internet sources for the fMRI study by Jacobs, Baumgartner, and Gegenfurtner (2014). Our goal was simply to evaluate as many images as possible from different sources. Images had a resolution 512 $\times$ 384 pixels and were presented on a Samsung Sync Master 2230R7 monitor with a refresh rate of 120 Hz.

2.3. Procedure

Color images were presented in randomized order for a duration of 33 ms per image. Observers were asked to indicate whether an image appeared unambiguously glossy or matte by pressing two buttons on a standard response box, respectively. They were told to withhold any response if the image could not be classified as either being glossy or matte. They were asked to respond as quickly as possible.

2.4. Results and discussion

Data were analyzed according to how many observers judged an image as matte or glossy. If observers decided to withhold their response, images were labeled as neutral. We were most interested in images that were consistently assigned the same label by all observers. For each image, the number of ratings across observers was counted. Since we wanted to be sure that our collection of images perceived as matte or glossy were as different as possible, we initially used only images that were rated as glossy or matte by all eight observers.

A total of 110 images rated "glossy" and 73 images rated "matte" were selected based on this criterion (Fig. 1). Fig. 1 also shows that the majority of images is considered as matte by the majority of observers. There are only few images ($N = 73$) that are considered as matte by all observers. This is quite different for the glossy images. Few images are considered as glossy by several observers, but there is good agreement on them ($N = 110$). The distributions in Fig. 1 indicate that it might be possible to relax the selection criterion a bit and use more images. We did this in some of our image classifications.

Fig. 2 shows the distribution of images rated matte or glossy with respect to the different material categories present in our image sampling pool. As expected, most images judged to be glossy belong to material categories that are highly associated with the property of glossiness, like metal, glass, or fluids. Images judged to be matte are mainly represented by fabric, wood, and stone images.

Our image sample does include a few images that show a significant portion of the background. We ran all of our analyses (explained below) with and without those images, but since the presence of the background did not systematically influence our results, we report results only on the full selection of images.

Please cite this article in press as: Wiebel, C. B., et al. Statistical correlates of perceived gloss in natural images. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.04.010

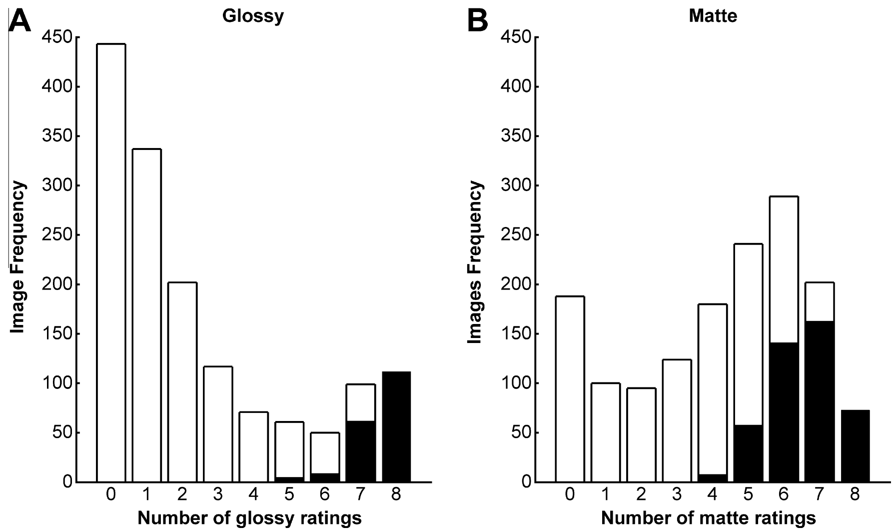

Fig. 1. Consistency across observers in glossy/matte judgments. (A) Glossy judgments (B) Matte judgments. The y-axis represents the number of images consistently judged as either matte or glossy by the corresponding number of observers specified on the x-axis. Filled histogram bars indicate those images that never received the opposite rating, e.g., glossy images that were not judged as matte by any of the other observers.

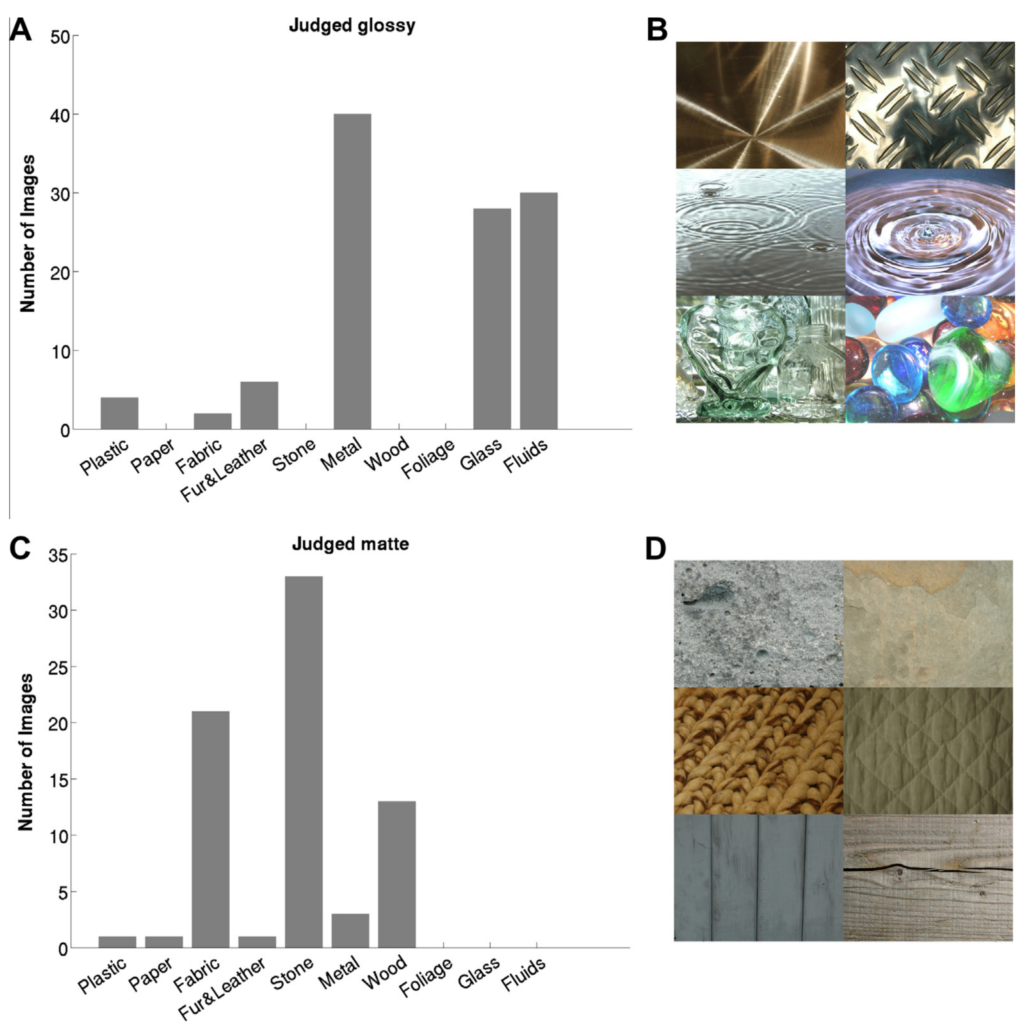

Fig. 2. Distribution of the selected images judged as glossy (A) or matte (C) in the different material categories of our database. To the right are two examples (columns) of the three most represented categories (rows) (B) Metal, fluids and glass for surfaces seen as glossy (D) Stone, fabric and wood for surfaces seen as matte.

Please cite this article in press as: Wiebel, C. B., et al. Statistical correlates of perceived gloss in natural images. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.04.010

3. Image classification

Our major interest was whether the perceptual categories for matte and glossy ratings could be recovered from the statistical features of the luminance histogram for each image. The importance of the luminance histogram for the characterization of surface textures has long been known. The moments of the luminance histogram, together with minimum and maximum, form the basis of the widely used texture analysis and synthesis algorithm put forward by Portilla and Simoncelli (2000). Here we use their "marginal statistics" as potential correlates of perceived gloss in images of natural surfaces.

3.1. Methods

We calculated the moments of the luminance distribution (mean, standard deviation, skewness, and kurtosis) as well as its minimum and maximum.

$$\text{Mean} = \bar{X} = \frac{\Sigma_{i=1}^{n} x_i}{n}$$

$$\text{Standard deviation} = \bar{S} = \sqrt{\frac{\Sigma_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

$$\text{Skewness} = \frac{n\sqrt{n-1}}{n-2} \frac{\Sigma_{i=1}^{n}(x_i - \bar{x})^3}{\left(\Sigma_{i=1}^{n}(x_i - \bar{x})^2\right)^{\frac{3}{2}}}$$

$$\text{Kurtosis} = \frac{n(n+1)(n-1)}{(n-2)(n-3)} \frac{\Sigma_{i=1}^{n}(x_i - \bar{x})^4}{\left(\Sigma_{i=1}^{n}(x_i - \bar{x})^2\right)^2}$$

while the mean is simply the average luminance within the image, we will use the standard deviation as our definition of the image contrast. Several other definitions of contrast in natural images are available, such as the rooted-mean-squared contrast or Michelson contrast, but we used the standard deviation because it is consistent with other moments of the luminance distribution. In particular, skewness describes the degree to which the luminance distribution of an image is asymmetric around the mean and kurtosis describes how peaked the data are. Thus, both skewness and kurtosis say something about the shape of the luminance distribution.

Luminance was extracted from the color images by converting the RGB images into the color-opponent DKL space (see Hansen & Gegenfurtner, 2012). Luminance in that space is defined in the standard way by the $V(\lambda)$ curve. Luminance for each pixel was calculated as a weighted sum of the red, green, and blue image components, with the weights determined by the luminance of the primaries of the monitor that was used to display these images. In particular, luminance in cd/m^2 was equal to 0.22R + 0.69G + 0.09B. We made sure that the results of our image classification did not depend on that particular choice. Close to identical results were obtained when using the G component of the images only, a frequently used shortcut, or with standard sRGB coordinates having weights 0.21R + 0.72G + 0.07B.

For the six statistics described above, we did a classification analysis to investigate how well they could differentiate between the matte and glossy image classes on their own and together. We also did *t*-tests for the two image groups to evaluate potential significant differences among the parameters.

Classification was done for each individual statistic and for all combinations of them. We trained the classifier on the whole set of images, leaving out one image at a time. The excluded image was then tested with the leave-one out classifier. All classification analyses were done with the linear discriminant analysis routines implemented in the "classify" function of the statistics toolbox for MATLAB (version R2014a, http://www.mathworks.com). We determined the contribution each one of the parameters made by comparing classification performance for the combinations including and excluding that parameter.

3.2. Results

Classification results are shown in Fig. 3. Overall classification with all six parameters achieved 90.16% correct. Using the standard deviation alone, 81.42% correct could be achieved, followed by 79.24% correct for the maximum and 77.24% correct for the mean. Skewness alone led to 66.67% correct classifications.

All of the differences in individual descriptive statistics were significant between glossy and matte image groups (see Table 1).

It is not quite straightforward to evaluate the contribution of individual statistics to the overall classification. We used two different approaches. First, when omitting individual statistics, leaving out the standard deviation had by far the biggest effect, reducing performance to 82.5%, suggesting a prominent role for

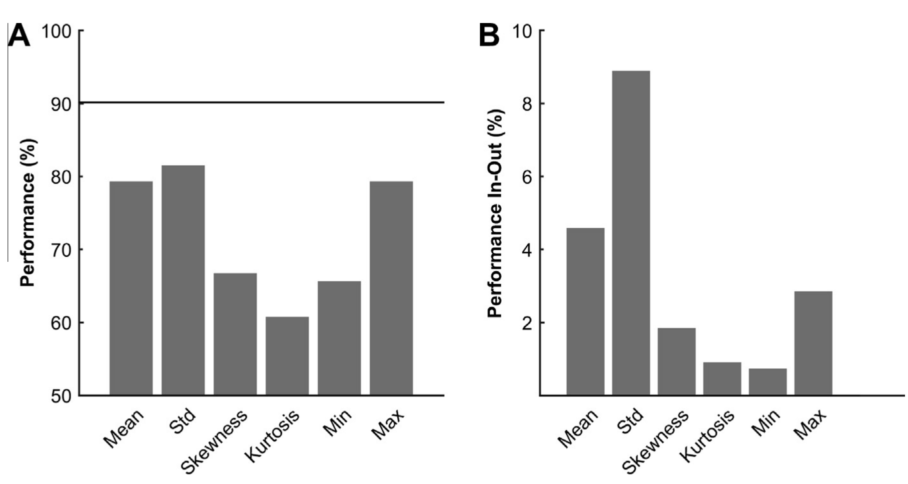

Fig. 3. Classification performance for the selected images. (A) Single statistics were used to classify the images judged as glossy and matte. The horizontal line specifies classification performance for all six image statistics together. (B) Contribution each one statistic made when looking at all possible combinations and comparing those including the statistic to those combinations excluding it.

Please cite this article in press as: Wiebel, C. B., et al. Statistical correlates of perceived gloss in natural images. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.04.010

Table 1
Descriptive statistics for the different parameters, together with t-tests on the statistical difference for each image parameter between glossy and image sample. All numbers rounded to two decimal places. All p-values were <0.001, except for kurtosis (<0.05).

	Mean gloss	Mean matte	Std gloss	Std matte	t_{181}	d'
Mean	0.59	0.46	0.12	0.11	7.61	1.26
Std	0.21	0.14	0.04	0.06	10.34	1.47
Skew	0.09	−0.28	0.70	0.76	3.47	0.60
Kurtosis	2.70	3.58	1.15	3.02	−3.06	0.67
Minimum	0.12	0.05	0.12	0.11	4.17	0.72
Maximum	0.99	0.90	0.01	0.14	8.33	1.25

this statistic. Excluding the mean (86.9%) or skewness (85.8%) had mild effects, while excluding the other parameters – kurtosis, minimum, and maximum – produced little deterioration in image classification.

Second, we analyzed the contribution of individual statistics by comparing classification accuracy between all those combinations of parameters that included and excluded a particular statistic. Overall, there are $2^6 - 1 = 63$ possible non-empty subsets of the six statistics. We calculated categorization performance for all of these 63 combinations, and then averaged across all the 32 subsets that included a particular statistic and across all 31 subsets that did not include that statistic. The difference between the two numbers specifies the contribution made by that particular statistic. The result of this analysis is shown in Fig. 3B. Standard deviation makes the biggest positive contribution, adding more than 8% performance improvement. The mean and the maximum lead to notable improvements of 3–4%, while skewness, kurtosis and minimum are negligible.

Fig. 4 shows an example of a two-dimensional classification using the standard deviation and skewness. In this case, a fairly good segregation of glossy and matte image classes is obtained. The two-dimensional distribution of the images indicates that the standard deviation has a higher weight for the classification.

The above classifications are for those images that were consistently categorized by all 8 observers. Presumably these are the images that give the strongest perceptual impressions of gloss and matte. This corresponded to our goal of maximizing our chances to successfully use image statistics for classification. To investigate the dependence of our result on this strategy, we determined classification accuracy and the importance of individual statistics also for less strict criteria, as shown in Fig. 5.

Two things can be seen. As the image selection gets stricter, classification performance also increases. A stricter image classification presumably selects images that are perceived to be more glossy or matte, at least to an average observer. Under that assumption, classification performance goes along with perceived gloss, which is in agreement with the idea that the image statistics determine perceived gloss at least partly. Fig. 5B shows the individual contributions of the six statistics we investigated. For some of them like standard deviation, the individual contribution also increases. For others, it remains at a low constant level (e.g., skewness) and for the mean it even decreases. This result indicates that the standard deviation of the luminance histogram is tightly linked to perceived gloss.

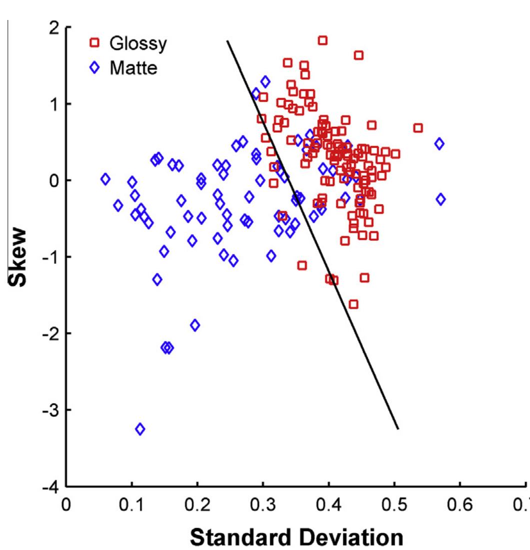

Fig. 4. Classification of images of natural surfaces into those judged glossy and matte, using standard deviation and skewness. A relatively high classification rate of 84% can be achieved in this case.

3.3. Discussion

Surfaces could be classified correctly based on skewness alone in 58% of all cases. This is in agreement with the results of Motoyoshi et al. (2007), who also found higher skewness in their photographs of glossy surfaces. Therefore, skewness is a correlate of perceived gloss, and this holds for the large variety of natural surfaces we investigated. However, we found that the standard deviation made for more accurate predictions than skewness, leading to an 82% correct classification of all surfaces. Combining information about the standard deviation with other statistics yielded classifications approaching the performance found with all predictors. All parameters together (mean, standard deviation, skewness, kurtosis, minimum, and maximum) could be used to correctly classify 91% of all surfaces. This is quite remarkable taking into consideration the large variety of images we had used.

These results suggest that the standard deviation is a strong cue to systematic variations in glossy appearance. Below, we will experimentally test the idea that the standard deviation is indeed used by human observers as a cue to perceived glossiness in photographs of natural surfaces. Before going into those experiments, we did the same classification analysis as above, but with rendered images. This allows us to evaluate whether classification performance arises from variations of parameter values linked to specular reflections in the (simulated) real world.

4. Classification with rendered objects

The photographs used above were not calibrated at all. They have gone through a photographic process, and due to the nature of our image database, this process could vary considerably. It is also quite clear that the dynamic range is compressed enormously, which affects nearly all of the statistics we investigated. In particular, large luminance values due to specular reflections typically get mapped to the largest possible output value, affecting skewness and the maximum disproportionally. Therefore, we ran simulations with rendered objects to see whether we could find similar differences for glossy and matte surfaces. By classifying the output of the rendering process, we can circumvent the step of actually displaying rendered images on a computer screen.

4.1. Methods

We rendered 83 different virtual objects each under 400 random conditions, using the software RADIANCE (Ward, 1994) through the MATLAB-based RenderToolbox3 (Lichtman, Xiao, & Brainard, 2007). RADIANCE performs a high-dynamic-range rendering allowing a higher precision than the 24 bits of the standard

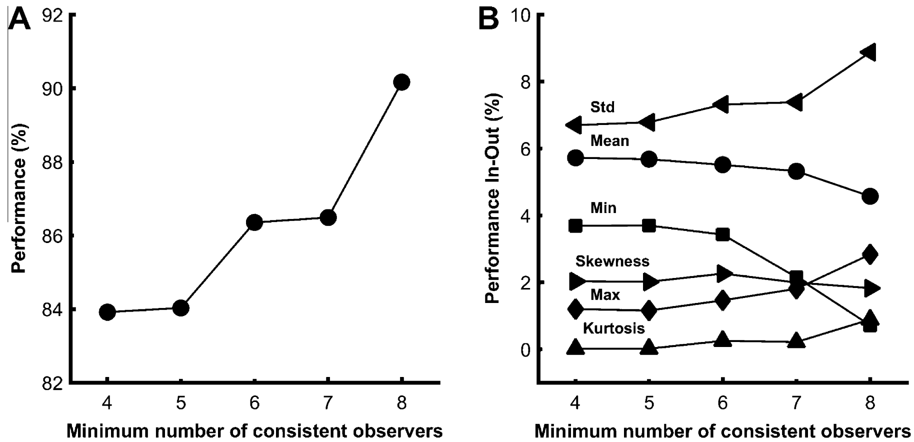

Fig. 5. Effects of making the image selection criterion more stringent on classification. (A) Overall classification performance using all statistics. (B) Contribution of individual statistics. Values for 8 observers are identical to the ones shown in Fig. 3.

RGB system. In every rendering only one virtual surface was present and the perspective was set to have the surface in the center of the image. We randomly rotated every object on its axis. The point from which the scene is viewed was randomly sampled from a sphere around the object. We rendered a total of 33,200 scenes, split into two groups of 16,600. The first group of scenes contained matte objects rendered with a Lambertian reflectance model. The second group contained glossy objects rendered with a Ward model (Ward, 1992). The diffuse reflectance components were randomly chosen for every virtual surface by sampling from a beta distribution, based on the results of Attewell and Baddeley (2007). These authors showed that samples of natural surface reflectance distributions were best approximated by a beta distribution with parameters $\alpha = 1.29$ and $\beta = 2.30$. The specular surfaces were defined by two parameters, one representing the magnitude of the specular component (specularity; set to .2 in the glossy surfaces we rendered) and the other the "spread" or blur of the specular reflection (roughness; set to .1). To simulate a naturalistic illuminant we used 9 of Debevec 's light probes (Debevec, 2008). For every scene one of these light probes was randomly chosen as the virtual light field (Fig. 6).

A perfectly dark surface was used to generate a mask for segmenting the surfaces from the background when calculating the luminance histograms. The three-dimensional models were sampled from three sources: the Turbosquid (http://www.turbosquid.com) license free repository, the Stanford Computer Graphics Laboratory (http://graphics.stanford.edu/data/3Dscanrep), and the AIM@SHAPE shape repository (http://vision-air.ge.imati.cnr.it/ontologies/shapes). We selected models that exhibited a large range of complexity, from simple geometrical shapes, such as a cube, to complex three-dimensional figures, like animals.

4.2. Results

Overall, the results show some similarities and some differences to the ones obtained with the photographic images (Fig. 7). Most importantly, matte and glossy rendered objects can be successfully categorized according to the statistics of their luminance histograms, as was the case for the images. Classification performance based on all of the parameters is much better than using any individual parameter and reaches a level of 85.68% correct. When using one parameter alone, the highest classification accuracy was achieved for the skewness (80.67%), followed by the maximum (74.29%), the standard deviation (68.26%) and then the kurtosis (65.3%). The mean reached 56.34% correct classification and the minimum basically performed at chance level (50.23%). When omitting each of the six parameters individually, the biggest drop

Fig. 6. Examples of rendered images. We randomly chose from a large number of shapes and different light fields, illustrated by the different backgrounds. The object was extracted from the background before calculating statistics. Matte surfaces on top, glossy surfaces in the bottom row.

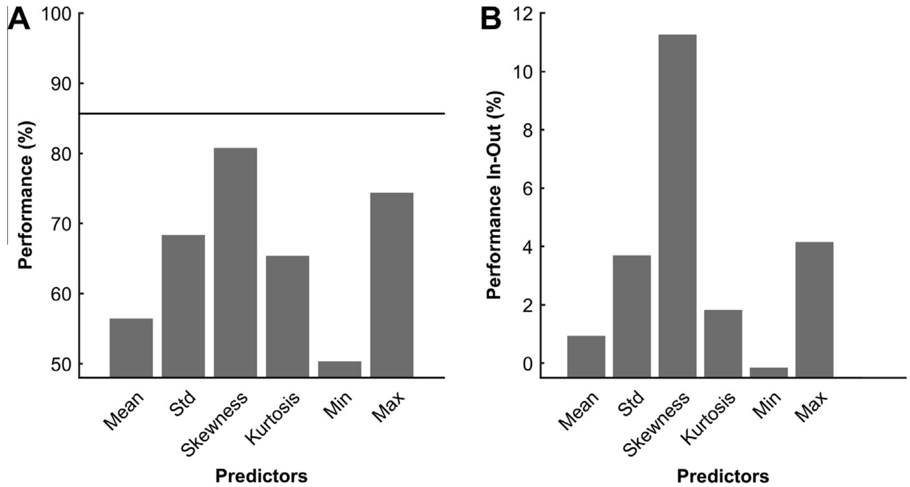

Fig. 7. Classification performance for the rendered objects. (A) Single statistics (B) Combinations of all parameters. The horizontal line specifies classification performance for all six image statistics.

Table 2
Descriptive statistics, together with t-tests on the statistical difference for each image parameter between glossy and matte rendered image samples. Image statistic values in linearly scaled RADIANCE units.

	Mean gloss	Mean matte	Std gloss	Std matte	t_{33198}	p	d'
Mean	0.428	0.313	0.412	0.371	26.58	<0.001	0.55
Std	0.404	0.129	0.366	0.153	89.10	<0.001	1.22
Skew	4.589	0.492	4.539	1.264	111.53	<0.001	1.71
Kurtosis	69.640	4.656	234.021	37.115	35.18	<0.001	1.71
Minimum	0.056	0.059	0.095	0.107	−2.45	0.0142	0.22
Maximum	6.157	0.678	6.764	0.702	103.42	<0.001	1.69

in classification accuracy down to 78.61% was found for the maximum. Omitting other parameters had little effect, due to a good degree of redundancy among the parameters. In all these cases, performance was still above 82.5% correct. Across the whole sample of rendered objects, the maximum was highly correlated with the mean ($\rho = 0.46$) and the standard deviation ($\rho = 0.64$). Skewness was correlated with the maximum ($\rho = 0.21$).

The analysis of individual contributions showed that the skewness made by far the largest contribution to classification accuracy (11.24%). The maximum (4.13%) and the standard deviation (3.68%) also made significant contributions on average, while mean, kurtosis and the minimum were negligible (Fig. 7B).

All of the differences in these parameters, with the exception of the minimum, were significant between the glossy and the matte object groups (see Table 2).

4.3. Discussion

The results show that both images of natural surfaces and rendered objects can be classified by the statistical properties of the luminance histogram. The degree to which this works for individual statistics is different for the two cases. For the rendered objects, skewness works best, while for the photographs the standard deviation makes the largest contribution.

There are at least two potential reasons for the differences. One reason is the photographic process, which, by necessity, reduces the dynamic range and severely affects the maximum, as well as other features. The second reason might be the much larger variety in illumination and material properties for the photographed surfaces. While we used many different shapes, light fields and diffuse reflectances for the rendered objects, all other surface properties of all the objects were identical, except for gloss.

Since the standard deviation was the most important predictor for the natural images and skewness worked best for the rendered objects, we will concentrate on these two statistics in the behavioral experiments that follow. The maximum works reasonably well for both kinds of scenes, but its individual contributions stayed far behind that of skewness and the standard deviation. Furthermore, the non-linearities present in the early visual system would most likely make it less useful for perceptual classification.

Of course, showing that a particular parameter correlates with perceived gloss does not prove that varying it causes corresponding changes in gloss perception. We therefore choose to actively manipulate the statistical properties of our images to determine whether there would be corresponding changes in perception.

5. Experiment 2

5.1. Effects of skewness and contrast on perceived gloss

We tested experimentally whether manipulations of image parameters would lead to changes in perception. For each image, we created six different versions resulting in different values of contrast and skewness. In addition to the original images, we created histogram-equalized images where all luminance levels have nearly equal frequency. These images have by definition zero skewness and their contrast is typically considerably larger than for the original images. On top of that, contrast manipulations were added to the original and histogram-equalized images. Furthermore, we created versions of the glossy images with a histogram that was matched to that of a typical image from the matte class. This way our image manipulations allowed a systematic and independent evaluation of the effects of skewness and contrast on perceived glossiness. In Experiment 2 observers viewed all six versions of the same image selected from those judged as glossy initially and ranked them according to gloss. Later on, in Experiment 3, images from both the glossy and matte classes will be displayed together.

5.2. Methods

5.2.1. Observers

12 observers participated in the study, nine were female and three were male. The mean age of the observers was 24. All observers gave written informed consent in agreement with the local ethics committee guidelines and in accordance with the Code of

Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

5.2.2. Stimuli and Apparatus

We used the consistently rated glossy images of Experiment 1 to produce the stimuli for this experiment. Six versions of the 110 glossy images were created. The different versions consisted of:

(1) Original images at their original contrast (*Original 100*), the way the images were obtained from the various databases, as described in Experiment 1. As above, contrast is defined here as the standard deviation of the luminance values.
(2) Images with overall contrast scaled to 80% of the original value (*Original 80*). During this step, the contrast *along all three color dimensions* was equally reduced to 80% of its original value. We converted images to DKL color space, as mentioned before, and then scaled the amplitude in each color direction by a factor of 0.8, before converting the image back to RGB. This contrast scaling was necessary for the image manipulations that we applied and where we selectively changed the luminance contrast. Because we used color images, there are some gamut constraints on the luminance changes that are possible. Reducing the contrast to 80% of its maximum greatly alleviated these problems. In the rare cases when there were isolated gamut over- or underflows, the individual pixels were scaled back equally along all three color dimensions to fit inside the gamut.
(3) Equalized luminance histogram without restrictions on the luminance contrast (*Equalized*). Here, histogram-equalization was applied to the original, contrast-scaled images (*Original 80*). No further scaling on contrast took place. Since histogram-equalization dramatically boosts image contrast, this class of images had the highest average contrast, even higher than the original images (*Original 100*) had.
(4) In a further step, we modified the luminance contrast of the histogram-equalized images (*Equalized*) to equal the luminance contrast of the original images (*Equalized 100*).
(5) The luminance contrast was even more reduced in the histogram-equalized images (*Equalized*) to equal the luminance contrast of the contrast-scaled original images (*Equalized 80*).
(6) Luminance histogram-matched to a typical "matte" image (*Histogram swapped*). For this, the luminance histogram of a "typical" image from the "matte" class was applied to the images from the "glossy" class. The image from the "matte" class with statistics closest to the median values of all the six statistics was selected as the "typical" matte image. It had a skewness of −0.268 and a contrast of 0.135. The images were contrast-scaled to 80% before applying these manipulations to avoid potential problems with over- or underflow of RGB-values.

All histogram manipulations were done using the "histeq" function implemented in the Matlab Image Processing Toolbox (The Math Works Inc., 2007, Natick, MA, USA). This function can be used to both *equalize* the frequencies of the luminance histogram and to *match* a given luminance histogram to another one, with equalization just being a special case of matching. In general, the "histeq" function finds a monotonic transformation to minimize the distance between the cumulative histogram of the image being equalized and the cumulative histogram of a reference image. Histogram equalization and matching can never be really perfect. However, the important aspect for our experiments was that the manipulation resulted in near zero skewness and a change in contrast. Stimuli were presented on the same monitor as before. The different image manipulation conditions are demonstrated in Fig. 8, where the effects on standard deviation and skewness are plotted as well.

5.3. Procedure

Observers were seated in a dimly lit room. Viewing distance to the monitor was approximately 107 cm. In each trial, all six versions of one image were presented in a grid (two rows, each with three images). The order of presentation was randomized for each observer. Observers ranked the images from most glossy to least glossy, by clicking on the images in order using a standard computer mouse. A colored frame that appeared around the image indicated selection of an image. The first frame was colored in red and each subsequent frame gradually changed to white for each successive selection. With this frame, observers could view the order of their glossiness ratings at all times. Observers could undo their last selection by pressing the right button of the computer mouse. After all the images were selected, observers had to press the left button of the mouse again to complete the trial. Observers were explicitly instructed to continue to the next trial only when they were satisfied with the order of the images.

5.4. Results and discussion

The average rank order for each image manipulation condition was calculated across all observers and trials. It is plotted as a function of either average skewness or average contrast (Fig. 9). If the skewness of the luminance profile drives the glossy appearance of the surface, than the images with an equalized luminance histogram (skewness = 0) should yield lower rank orders than the images in which the shape of the luminance profile was left untouched. Otherwise, the skewness of the luminance histogram would not be systematically linked to the glossy appearance of a surface. Alternatively, if contrast were related to the glossy appearance of a surface, then the ordering of contrast magnitude in the different image conditions would be systematically related to the rank order of the images.

Results are shown in Fig. 9 and indicate a clear relationship between our contrast measure and the degree of glossiness perception. The highest glossiness rank, averaged across the 12 observers, occurred for the histogram equalized image condition. Here, the luminance profile of the images was flattened, resulting in an overall skewness of zero, while the contrast in the images was increased compared with the original images. This result speaks against a dominant role of the luminance histogram skewness in the perception of glossiness. However, given identical amounts of contrast, skewness seems to play a role. For the direct comparison between the *Original 100* and *Equalized 100* and the *Original 80* and *Equalized 80* image condition, the skewed versions of the image (*Original 100* and *Original 80*) were ranked above the unskewed version (*Equalized 100* and *Equalized 80*). Overall, the effect of contrast is more pronounced: *Original 100* and *Equalized 100* are both ranked above *Original 80* and *Equalized 80*. In other words, if contrast remains constant, then a skewed image will look glossier than a non-skewed image. If instead skewness remains constant, then an image of more contrast will look more glossy than an image of less contrast. The two effects are difficult to compare, but overall, a contrast increase by 0.036 led to a higher rank, and a change in skewness by 0.26 increased the rank by one unit. The histogram swapped image condition was ranked last, since it had the lowest contrast and skewness. This also supports an image-based explanation, since the images originally rated as matte had lower contrast on average and were slightly negatively skewed.

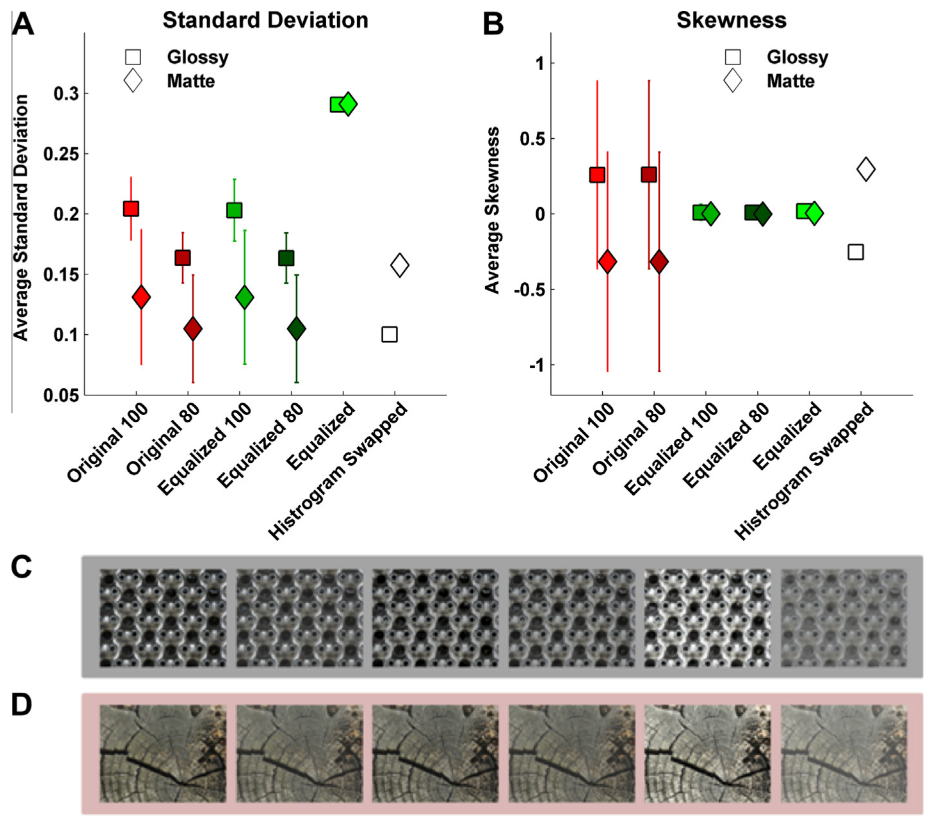

Fig. 8. Manipulation effect. (A) Average standard deviations and (B) Average skewness for each image condition, indicated on the x-axis. Square symbols represent images from the glossy category and diamonds represent images from the matte category. (C) Example "glossy" image (D) Example "matte" in the six image manipulation conditions. In Experiment 2, only "glossy" images were used. In Experiment 3, "glossy" and "matte" images were used.

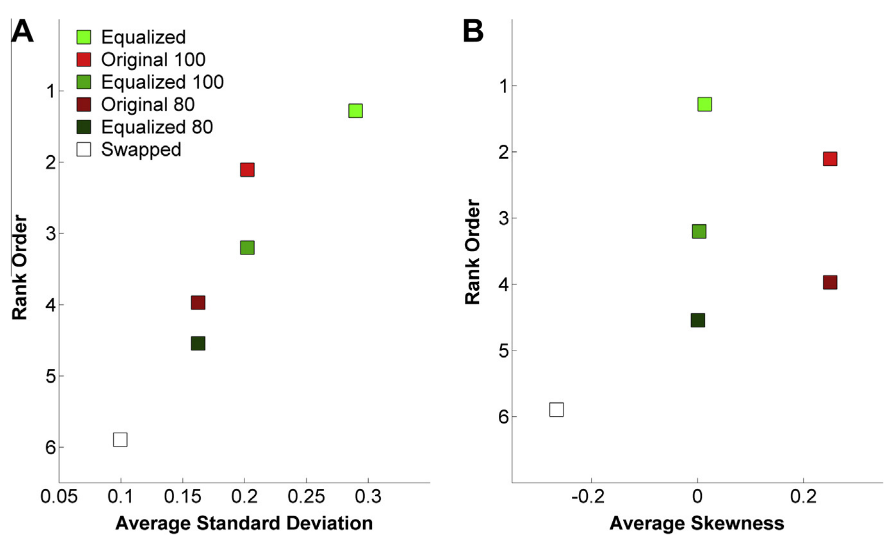

Fig. 9. Average rankings for original and manipulated glossy images. (A) Average rank for each glossy image version in Experiment 2 plotted against the average standard deviation. (B) Same data plotted against mean skewness. The color indicates the image manipulation.

One general problem with ranking or rating experiments is that it is never entirely clear what it is that observers are rating. In our case, we did give them a clear instruction to rate perceived gloss, and basically there is no a priori reason to assume they did not follow our instruction and rated something else. However, observers do have a tendency to use in their ratings features that do co-vary with the stimuli. Since contrast and skewness were the major attributes varying in this experiment, there is a small chance that observers choose to rank images along these dimensions, rather than relying on their perception of gloss.

We try to circumvent this potential problem in Experiment 3 by introducing further variability. First of all, six different images were presented on each trial, and three of the six images came from the matte class of images. If observers judge the images based

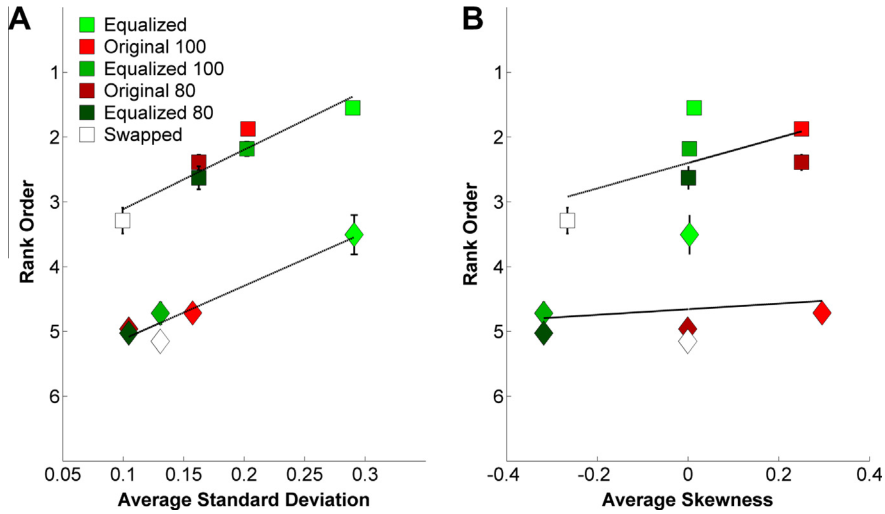

Fig. 10. Average rankings for the original and manipulated glossy and matte image groups. (A) Average rank order for the six image versions plotted against their average standard deviation. Square symbols represent images rated as glossy. Diamond symbols represent images rated as matte. The color indicates the image manipulation. (B) Same data plotted against average skewness. Lines show the result of a linear regression fit to the matte and glossy image data.

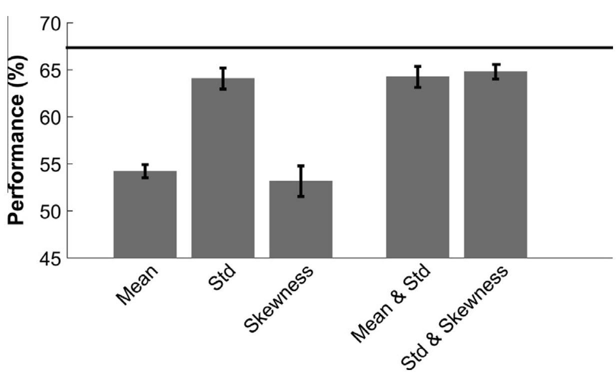

Fig. 11. Ranking prediction results. Considering the single predictors, the highest classification performance was reached by contrast, followed by the mean and then the skewness. The line represents performance using all predictors. Chance level was 50%.

on perceived gloss, we would expect them to consistently rank the images from the "glossy" class higher than the images from the "matte" class.

6. Experiment 3

6.1. Perceived gloss

We intermixed images initially rated as glossy and matte. The same image manipulations were used as in Experiment 2, but each trial contained three different "glossy" and three different "matte" images. This allowed us to assess the effect of different statistical parameters of the individual images across individual trials.

6.2. Methods

6.2.1. Observers

12 new observers participated in the experiment (10 female and two male). The mean age of our observers was 26. All observers gave written informed consent in agreement with the guidelines of the local ethics committee and in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

6.2.2. Stimuli and Apparatus

Images that were consistently rated as "matte" in the preliminary experiment were manipulated in the same way as the glossy images in Experiment 2 (see Fig. 8). Images in the histogram swapped image condition were matched to a typical gloss image with statistical parameters close to the median values. 73 of the "glossy" images were randomly selected for each observer in addition to the 73 "matte" images. Stimuli were presented on the same monitor as before.

6.2.3. Procedure

In each trial, three "matte" and three "glossy" images were presented in a grid (two rows; three images per row). The images were chosen so that all six images were different, and that all six image manipulations were on display. The same image was not presented more than once in a trial, regardless of the manipulation applied to it. The number of occurrences of each image manipulation was balanced across images from the matte and glossy image classes within the experiment. Each image manipulation condition was used 73 times in the glossy and 73 times in the matte image sample. 146 trials were presented in total.

6.2.4. Trial-by-trial ranking

We evaluated the trial-by-trial contribution of our image statistics to each observer's ranking using a leave-one-out linear classification approach like in Experiment 1. Only the three glossy images of each trial were used for this analysis. For each observer and for each trial, we trained the classifier with the rankings of all the images and their statistics except the ones of that particular trial. Then, the statistics and rankings of the left-out trial were used to test the performance of the classifier. The performance for each observer was averaged across trials. The rankings were recoded as binary decisions by comparing the rank of each image with all of the others, resulting in six comparisons for each scene. These comparisons indicated whether each image was judged as more or less glossy than the others.

6.3. Results and Discussion

The data were analyzed in the same way as in Experiment 2. Results in Fig. 10 show that the "glossy" images were rank ordered above the "matte" images, regardless of the image manipulations.

However, within "matte" and "glossy" images, we found the same systematic pattern of results as we found in Experiment 2. While skewness was only partially related to the induced glossiness percept, contrast was a better predictor for the observers' ordering. This was also captured by two linear regression fits, fitted separately to the two image groups. No significant fit was found when observer's ratings were fitted to the skewness data, neither for the glossy ($p > 0.05$) nor for the matte image class ($p > 0.05$). However, for contrast, both linear regression fits yielded significant results for the glossy ($R^2 = 0.904$, $p < 0.01$) and the matte image categories ($R^2 = 0.930$, $p < 0.05$). For the matte category, the effect mainly seemed to be caused by the histogram equalized images.

We found similar results by predicting the rankings with the image statistics on a trial-by-trial basis (Fig. 11). In this case, all variability due to factors based on image interpretation is simply ignored. This naturally leads to an overall low level of prediction. However, contrast gave the highest classification performance (64.08%). It outperformed the mean and the skewness, which were both just above the chance level, and approached the classification performance with all of the predictors taken together (67.35%). The classification performance with contrast alone is comparable to that reached by adding the mean (64.25%) or the skewness (64.8%) as predictors. Admittedly, these numbers are low, but they indicate that contrast differences do play a significant role when ranking images according to perceived gloss.

Taken together, the results show that contrast and gloss are closely related for our images, evidenced by the close fit of the regression line to the glossy ranking data in Fig. 10. For the images of surfaces perceived to be matte, contrast is rather low and uniform to begin with.

7. General discussion

The present study extends previous research on gloss perception to real world photographs. Our findings show that image statistics can modulate the perception of gloss, even in a large and varied stimulus set. Everything else being equal, contrast and skewness do have effects on the degree of perceived glossiness. Our data show that for our conditions, contrast might be a more informative cue than skewness. However, neither increasing contrast nor skewness was sufficient to consistently induce a strong percept of gloss in images originally rated as matte.

7.1. Image based cues to gloss

Presently, researchers are having a heated debate regarding the utility of image-based cues for material perception, especially with respect to gloss. Our results agree with findings on both sides of the discussion. On the one hand, we show that contrast and to a lesser degree skewness are correlates of gloss perception, confirming findings by Motoyoshi et al. (2007). On the other hand, the influence of these image based cues is limited compared to the variability introduced by the contents of the images.

To be useful for the visual system, image based cues that co-vary with surface reflectance properties must be somewhat stable across different settings of illumination, shape and viewpoint. One can study this relationship by comparing physically realistic renderings of material surfaces with different geometric properties and illumination configurations. A complementary strategy is to tackle this problem in real-world photographs that show a high variability in the way they portray a material surface, by using inside and outside illumination, different viewpoints, and large samples of different types of material, objects and surface structures. Here, we applied the second approach and systematically examined the relevance of the skewness and the contrast (standard deviation) of the luminance histogram for perceived surface gloss in photographs of natural surfaces.

Our image data set contained a selection of images taken from different databases. The images were taken by different photographers, under different illumination conditions, and in different contexts. This way we assured a high variability of different material appearances in our image database. Moreover, since the "glossy" and "matte" image samples were selected from a total of 1492 images and were based on the consistent judgments of eight naïve observers, we assume that they truly represent what human observers perceive as glossy or matte.

Using both the skewness and standard deviation of the luminance histogram, a linear classifier was able to reliably distinguish our empirically derived set of glossy images from our empirically derived set of matte images. For these images, the standard deviation was a better predictor than skewness, challenging the view that the skewness of the luminance histogram is "the" relevant image-based cue for surface gloss (Motoyoshi et al., 2007). In two subsequent experiments, we let human observers rank order the amount of glossiness induced by our images. The images were manipulated by varying the standard deviation and skewness independently, allowing us to test their contribution to a glossy appearance of a surface. In accordance with our image analysis, the standard deviation seemed to be more important for the perception of glossiness than skewness. This parallels other reports that showed no direct link between the skewness of the luminance histogram and the perception of glossiness (Anderson & Kim, 2009; Kim & Anderson, 2010). Nevertheless, skewness was not irrelevant in our experiments, as it still contributed to the amount of apparent glossiness, but to a lesser degree than contrast did.

In comparison with our results, a few other studies failed to find such an important role for contrast. Motoyoshi et al. (2007) not only investigated skewness as a cue to perceived gloss, but also tested the effect of contrast, computed as standard deviation divided by the mean. They did observe a correlation between contrast and glossiness rating, but it was lower ($r = 0.68$) than the one for skewness ($r = 0.89$). These results are quite similar to our results for rendered objects. This is no coincidence because their photographic techniques and use of coated surfaces might have shared more similarities with our computer-rendered objects than with the largely variable photographs of our database.

Olkkonen and Brainard, (2010, 2011) examined the role of luminance histogram statistics for predicting gloss matches under different illumination conditions and shapes. Their observers had to adjust the diffuse and the specular components of a test sphere to match the appearance of a reference sphere under different illumination conditions. While the diffuse component was matched veridically across the different conditions, the specular matches were influenced by the structure of the light fields used for the different illumination conditions. These differences in the specular matches could not be explained by equating the standard deviation, skewness or kurtosis of the test and the reference spheres under the different illuminations. This indicates that there are dimensions of gloss that are not captured by simple image statistics.

Using a different approach, Ferwerda, Pellacini, and Greenberg (2001) ran an experiment, aimed at investigating the perceptual dimensions of glossiness using multidimensional scaling techniques. They generated a set of spheres and varied their diffuse reflectance component, the energy of the specular reflectance component and the spread of the specular lobe, according to the physical reflectance model of Ward (1992). They found that a two-dimensional feature space was appropriate to represent their data. Most importantly, they observed that both axes closely corresponded to the distribution of apparent contrast and the distribution of the apparent sharpness in the reflected image. In a second experiment, they attempted to quantify the features that the axes represented. Based on these results, they built a model capable of predicting differences in apparent gloss for their – rather reduced – stimulus set. In agreement with our results, contrast plays an important role for determining perceived gloss according to their model.

Recently, Marlow, Kim, and Anderson (2012) presented a model combining four surface cues, contrast, sharpness, depth and coverage, to explain the perceived gloss of bumpy surfaces. The model was able to explain 94% of the variability in their data. Interestingly, they did not estimate their cues from the images. Instead, their observers estimated the cues by rating the images. They found major contributions of the perceived size of the area covered by the highlight, perceived depth, perceived contrast and perceived sharpness. The cue weights were determined by predicting the observers' gloss judgments of bumpy surfaces that had various relief heights and that were placed in different light fields. Skewness as a pure image statistic did not have any effect. It would have been worthwhile to see whether the standard deviation, calculated directly from the images, would have played a role.

7.2. Gloss constancy

In our third experiment, we found that when observers rank ordered images from the matte and glossy categories, "glossy" images were ranked higher on average (i.e., glossier) than "matte" images. This shows that observers did not simply order images according to the image cues that we evaluated. In fact, our observers did not order the images based solely on contrast, indicating that they did not base their decisions on a single simple image statistic. Our results show that image based cues can indeed be used as a proxy to the perception of gloss, especially when taking several statistics into account. However, as already pointed out by Motoyoshi et al. (2007) in their original study, image based statistics cannot explain the full range of surface gloss judgments for arbitrary surfaces. They can modulate the degree of gloss perception, but they cannot turn arbitrary surfaces into ones perceived as glossy. Other concepts must have been involved in the choices observers made. This parallels previous work with artificial stimuli that has outlined the importance of surface shape and illumination geometry for perceived glossiness (see for example Anderson & Kim, 2009; Beck & Pradzny, 1981; Berzhanskaya et al., 2005; Doerschner, Boyaci, & Maloney, 2010; Fleming, Dror, & Adelson 2003; Ho Landy, & Maloney, 2008; Kim & Anderson, 2010; Kim, Marlow, & Anderson, 2011; Marlow, Kim, & Anderson, 2012; Olkkonen & Brainard, 2010, 2011; Pont & te Pas, 2006; te Pas & Pont, 2005). It is an important next step to see how these properties relate to our findings for gloss perception in natural images.

Our present study makes an important contribution by extending previous research on gloss perception to real world photographs in a systematic and broad manner. We have shown that statistical correlates of perceived gloss are present in photographs of natural surfaces. Our data show that under these circumstances, contrast might be a more informative cue than skewness.

8. Conclusion

Our findings agree with the seemingly contradictory results of Motoyoshi et al. (2007) and Anderson and Kim (2009). Image statistics can modulate the perception of gloss, even in a large and varied stimulus set. Everything else being equal, contrast and skewness do have effects on the degree of perceived glossiness. However, they are not sufficient to consistently induce a strong percept of gloss in images originally rated as matte.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft Grant DFG GE 879/9. We are grateful to Roland Fleming for helpful discussions. Rob Ennis commented on a previous draft of this manuscript and helped us enormously to improve grammar and style. Christiane Wiebel is now with the Modeling of Cognitive Processes Group, Department of Software Engineering and Theoretical Computer Science, Technische Universität Berlin, Berlin, Germany.

References

Adelson, E. H. (2001). On seeing stuff: The perception of materials by humans and machines. In *Photonics West 2001 - Electronic Imaging* (pp. 1–12). International Society for Optics and Photonics.
Anderson, B. L. (2011). Visual perception of materials and surfaces. *Current Biology, 21*(24), R978–R983.
Anderson, B. L., & Kim, J. (2009). Image statistics do not explain the perception of gloss and lightness. *Journal of Vision, 9*(11), 1–17.
Attewell, D., & Baddeley, R. J. (2007). The distribution of reflectances within the visual environment. *Vision Research, 47*(4), 548–554.
Baumgartner, E., Wiebel, C., & Gegenfurtner, K. R. (2013). Visual and haptic representations of material properties. *Multisensory Research, 26*(5), 429–455.
Baumgartner, E., Wiebel, C., & Gegenfurtner, K. R. (2015). A comparison of haptic material perception in blind and sighted individuals. *Vision Research*. http://dx.doi.org/10.1016/j.visres.2015.02.006 (in this issue).
Beck, J., & Prazdny, S. (1981). Highlights and the perception of glossiness. *Perception and Psychophysics, 30*(4), 407–410.
Berzhanskaya, J., Swaminathan, G., Beck, J., & Mingolla, E. (2005). Remote effects of highlights on gloss perception. *Perception, 34*(5), 565–575.
Blake, A., & Bülthoff, H. (1990). Does the brain know the physics of specular reflection? *Nature, 343*(6254), 165–168.
Debevec, P. (2008). Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *ACM SIGGRAPH 2008 classes* (pp. 32). ACM.
Doerschner, K., Boyaci, H., & Maloney, L. T. (2010). Estimating the glossiness transfer function induced by illumination change and testing its transitivity. *Journal of Vision, 10*(4), 1–9.
Ferwerda, J. A., Pellacini, F., & Greenberg, D. P. (2001). Psychophysically based model of surface gloss perception. In *Photonics West - Electronic Imaging* (pp. 291–301). International Society for Optics and Photonics.
Fleming, R. W. (2012). Human perception: Visual heuristics in the perception of glossiness. *Current Biology, 22*(20), R865–R866.
Fleming, R. W. (2014). Visual Perception of materials and their properties. *Vision Research, 94*, 62–75.
Fleming Dror, R. O., & Adelson, E. H. (2003). Real-world illumination and the perception of surface reflectance properties. *Journal of Vision, 3*(5), 347–368.
Hansen, T., & Gegenfurtner, K. R. (2012). Higher-order color mechanisms: Evidence from noise masking experiments in cone contrast space. *Journal of Vision, 13*(1), 26.
Ho, Y. X., Landy, M. S., & Maloney, L. T. (2008). Conjoint measurement of gloss and surface texture. *Psychological Science, 19*(2), 196–204.
Hunter, R. S. (1937). Methods of determining gloss. *Journal of Research of the National Bureau of Standards, 18*(1), 19–39.
Jacobs, R. H. A. H., Baumgartner, E., & Gegenfurtner, K. R. (2014). The representation of material categories in the brain. *Frontiers in Psychology, 5*, 146. http://dx.doi.org/10.3389/fpsyg.2014.00146.
Kim, J., & Anderson, B. L. (2010). Image statistics and the perception of surface gloss and lightness. *Journal of Vision, 10*(9), 3.
Kim, J., Marlow, P., & Anderson, B. L. (2011). The perception of gloss depends on highlight congruence with surface shading. *Journal of Vision, 11*(9).
Lichtman, D. P., Xiao, B., Brainard, D. H. (2007). RenderToolbox Version 2. Available at <http://rendertoolbox.org>.
Marlow, P., Kim, J., & Anderson, B. L. (2012). The perception and misperception of specular surface reflectance. *Current Biology, 22*(20), 1909–1913.
Motoyoshi, I., Nishida, S., Sharan, L., & Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature, 447*(7141), 206–209.
Nishida, S., & Shinya, M. (1998). Use of image-based information in judgments of surface-reflectance properties. *Journal of the Optical Society of America A – Optics Image Science and Vision, 15*(12), 2951–2965.
Olkkonen, M., & Brainard, D. H. (2010). Perceived glossiness and lightness under real-world illumination. *Journal of Vision, 10*(9), 1–19.
Olkkonen, M., & Brainard, D. H. (2011). Joint effects of illumination geometry and object shape in the perception of surface reflectance. *i-Perception, 2*(9), 1014–1034.
Pont, S. C., & Koenderink, J. J. (2005). Reflectance from locally glossy thoroughly pitted surfaces. *Computer Vision and Image Understanding, 98*(2), 211–222.
Pont, S. C., & te Pas, S. F. (2006). Material illumination ambiguities and the perception of solid objects. *Perception, 35*(10), 1331–1350.

Please cite this article in press as: Wiebel, C. B., et al. Statistical correlates of perceived gloss in natural images. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.04.010

Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision, 40*(1), 49–70.

Sharan, L. (2009). *The perception of material qualities in real-world images* (Ph.D. thesis). Massachussets Institute of Technology.

Sharan, L., Li, Y. Z., Motoyoshi, I., Nishida, S., & Adelson, E. H. (2008). Image statistics for surface reflectance perception. *Journal of the Optical Society of America A –Optics Image Science and Vision, 25*(4), 846–865.

Sharan, L., Rosenholtz, R., & Adelson, E. H. (2009). Material perception: What can you see in a brief glance? *Journal of Vision, 9*(8), 784.

Sharan, L., Rosenholtz, R., & Adelson, E. H. (2014). Accuracy and speed of material categorization in real-world images. *Journal of Vision, 14*(9), 1–24.

te Pas, S. F., & Pont, S. C. (2005). A comparison of material and illumination discrimination performance for real rough, real smooth and computer generated smooth spheres. In *APGV'05: Proceedingsof the 2nd Symposium on Applied Perception in Graphics and Visualization* (pp. 75–81). New York: ACM.

Thompson, W. B., Fleming, R. W., Creem-Regehr, S. H., & Stefanucci, J. K. (2011). *Visual perception from a computer graphics perspective*. Wellesley, MA, USA: CRC Press.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520–522.

Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., & BuÈlthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *European Journal of Neuroscience, 14*(5), 869–876.

Toscani, M., Valsecchi, M., & Gegenfurtner, K. R. (2013a). Optimal sampling of visual information for lightness judgments. *Proceedings of the National Academy of Sciences, 110*(27), 11163–11168.

Toscani, M., Valsecchi, M., & Gegenfurtner, K. R. (2013b). Selection of visual information for lightness judgements by eye movements. *Philosophical Transactions of the Royal Society B: Biological Sciences, 368*(1628), 20130056.

Ward, G. J. (1992). Measuring and modeling anisotropic reflection. *ACM SIGGRAPH Computer Graphics, 26*(2), 265–272.

Ward, G. J. (1994). The RADIANCE lighting simulation and rendering system. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques* (pp. 459–472). ACM.

Wiebel, C. B., Valsecchi, M., & Gegenfurtner, K. R. (2013). The speed and accuracy of material recognition in natural images. *Attention, Perception & Psychophysics, 75*(5), 954–966.