# Toward a Perceptual Space for Gloss

JOSH WILLS
Sony Pictures Imageworks
SAMEER AGARWAL
University of Washington, Seattle
and
DAVID KRIEGMAN and SERGE BELONGIE
University of California, San Diego

We design and implement a comprehensive study of the perception of gloss. This is the largest study of its kind to date, and the first to use real material measurements. In addition, we develop a novel multi-dimensional scaling (MDS) algorithm for analyzing pairwise comparisons. The data from the psychophysics study and the MDS algorithm is used to construct a low dimensional perceptual embedding of these bidirectional reflectance distribution functions (BRDFs). The embedding is validated by correlating it with nine gloss dimensions, fitted parameters of seven analytical BRDF models, and a perceptual parameterization of Ward's model. We also introduce a novel perceptual interpolation scheme that uses the embedding to provide the user with an intuitive interface for navigating the space of gloss and constructing new materials.

Categories and Subject Descriptors: I.3.7 [**Computing Methodologies**]: Three-Dimensional Graphics and Realism

General Terms: Algorithms, Experimentation, Human Factors, Measurement

Additional Key Words and Phrases: Rendering, reflectance models, human perception

**ACM Reference Format:**

Wills, J., Agarwal, S., Kriegman, D., and Belongie, S. 2009. Toward a perceptual space for gloss. ACM Trans. Graph. 28, 4, Article 103 (August 2009), 15 pages. DOI = 10.1145/1559755.1559760 http://doi.acm.org/10.1145/1559755.1559760



## 1. INTRODUCTION

The past two decades have seen a rapid increase in the use of computer generated imagery. With increased use has come the demand for increased physical realism in these images. One of the key factors that affects the photorealism of computer generated imagery besides the scene geometry and illumination is how the objects in the scene reflect light. Depending upon its material reflectance properties, a sphere under constant illumination can appear to be painted with glossy red paint, made out of polished steel, or cut out from a piece of blue sapphire.

A local model of material reflectance is the bidirectional reflectance distribution function (BRDF) [Nicodemus et al. 1977]. The BRDF is a four dimensional function that describes how light is scattered from a surface as a function of the incoming and outgoing directions. Given their central role in the process of image formation, BRDFs have been the subject of extensive study in both computer graphics and computer vision communities. This has included work on creating physics based analytic BRDF models [Torrance and Sparrow 1967; Cook and Torrance 1981; He et al. 1991; Ward 1992; Ashikhmin et al. 2000], empirical measurement of real world material reflectance [Marschner et al. 2000; Matusik et al. 2003a] and more recently measurement driven models [Matusik et al. 2003b]. However, the move towards more sophisticated reflectance models, both theoretical and measurement based, is not without its problems.

As it stands today, a digital artist has to develop a feel for the parameters of the various analytical reflectance models before he can use them to produce the desired effect. This is due to the complex relationship between model parameters and the resulting perceptual sensation. Learning this relationship is a complicated and error prone process based on repeated trial and error. The challenge is even more acute for data-driven BRDF models where the parameter space is particularly large and unintuitive. Imagine trying to select a desired color by specifying parameters that define a spectral density function. This has motivated the development of tools that allow

This work was partially supported by an NSF IGERT Grant (Vision and Learning in Humans and Machines #DGE-0333451), NSF CAREER #0448615 and the Alfred P. Sloan Research Fellowship and NSF Research Infrastructure Grant number NSF EIA-0303622.
Part of this work was done when Josh Wills and Sameer Agarwal were students at University of California, San Diego.
Authors' address: J. Wills, Sony Pictures Imageworks, 9050 Washington Blvd., Culver City, CA 90232; email: joshjwills@gmail.com.

DOI 10.1145/1559755.1559760 http://doi.acm.org/10.1145/1559755.1559760

Table I. Types of Gloss

| Type of Gloss | Incident Angle | Reflected Angle | Description |
|---|---|---|---|
| Specular Gloss 1 (ASTM D523) | 20° | 20° | Perceived brightness associated with the specular reflection from a surface |
| Specular Gloss 2 (ASTM C346) | 45° | 45° | |
| Specular Gloss 3 (ASTM D523) | 60° | 60° | |
| Sheen (ASTM D523) | 85° | 85° | Perceived shininess at grazing angles in otherwise matte surfaces |
| Distinctness of Image (DOI) (ASTM E430) | 30° | 30.3° | Perceived sharpness of images reflected in a surface |
| Bloom (2° Haze)(ASTM E430) | 30° | 32° | Perceived cloudiness in reflections near the specular direction |
| Haze (ASTM E430) | 30° | 35° | |
| Diffuseness (Hunter and Harold 1987 pg. 88) | 30° | 45° | Perceived brightness for diffusely reflecting areas |
| Contrast Gloss (Hunter and Harold 1987 pg. 285) | 45° | 0° and 45° | Perceived relative brightness of specularly and diffusely reflecting areas |

the user to navigate the space of BRDFs and to work with BRDFs in a manner similar to how they work with various color spaces [Ben-Artzi et al. 2006; Ngan et al. 2006].

Further, the continuing success of relatively simple phenomenological models of reflectance, for example, Ward's model, Torrance & Sparrow, and so on, indicates that the human visual system is not sensitive to every single variation in the BRDF. This observation lies at the heart of the recent focus on perceptually driven rendering [Ramanarayanan et al. 2007; Myszkowski 2002; Dumont et al. 2003; Gibson and Hubbold 1997; Luebke and Hallen 2001]. Knowing when and where in the scene certain variations in the BRDF can be ignored or simplified can lead to significant computational savings. Similarly, when approximating a complicated measured BRDF with an analytical model, instead of focusing on an algebraic measure of goodness of fit, we can instead build and use an error measure that only focuses on what is perceptually important [Ngan et al. 2005].

Thus, as we make progress in mathematically characterizing reflectance and come up with new and improved ways of measuring it, it is also important that we develop a better understanding of how the human visual system perceives the reflection of light. Such a development not only has implications for efficient image synthesis, but also for computer vision, where an understanding of reflectance perception will give us insight into the priors and constraints used by humans to solve various shading related problems, for example, shape from shading and object recognition with variable and unknown lighting.

In this article, we present a new study of the perception of the achromatic component of reflectance, also known as gloss. In a BRDF, gloss is responsible for changes in the magnitude and spread of the specular highlight as well as the change in reflectance that occurs as light moves away from the normal toward grazing angles. Gloss was first studied in the paper industry [Hunter and Harold 1987]. Since then there has been significant interest and work in various industries, including textiles and paint. The modern notion of gloss has been formalized by the American Society for Testing and Materials (ASTM). The ASTM defines gloss as "the angular selectivity of reflectance, involving surface-reflected light, responsible for the degree to which reflected highlights or images of objects may be seen as superimposed on a surface" [ASTM 2005a]. As no single quantitative measure captures the perception of gloss across different materials, a number of different gloss dimensions have been defined over the years [ASTM 1999, 2004, 2005b; Hunter and Harold 1987]. Table I describes a selection of nine Gloss dimensions. For each dimension we list the source of its definition, the incident and reflected directions at which the measurements are made, and the particular aspect of reflection that is captured.

We chose to consider only gloss because the largest publicly available database of reflectance measurements (the MIT-MERL database [Matusik et al. 2003b]) consists of only 55 usable isotropic BRDFs. This is a very small subset of the vast variety of reflectance functions. Color is such a strong perceptual cue that given the sparseness of our BRDF database, differences in color between two BRDFs will completely overwhelm differences due to Gloss. Thus, in the following, the term BRDF will refer to the achromatic aspects of reflectance; when we refer to the chromatic aspects, we will make specific note of it. The methods developed in this article, however, are not limited to the study of gloss and can be used to study the perception of chromatic as well as achromatic aspects of reflectance.

We contribute to the state of the art in perception research in computer graphics in three ways.

(1) We design and implement a comprehensive study of the perception of measured reflectance. We argue that our methodology is better suited for capturing the human perception of gloss and is less susceptible to experimental errors than previously used methods. This is the largest study of its kind to date, and the first to use real material measurements.
(2) Motivated by the design of our psychophysics study, we develop a novel Multi-Dimensional Scaling (MDS) algorithm for analyzing pairwise comparisons. This algorithm is a more general and efficient replacement for the widely used weighted non-metric MDS algorithm [Borg and Groenen 2005].
(3) We use the data from our psychophysics study to analyze the perception of gloss. As part of this analysis we estimate the dimensionality of the space of gloss perception and construct a perceptually meaningful embedding of these BRDFs. We perform an extensive validation of our results by comparing our embedding with ASTM gloss dimensions [ASTM 2005a], the parameters of seven analytical BRDF models fitting the MIT/MERL database [Ngan et al. 2005], and a perceptual parameterization of Ward's model [Pellacini et al. 2000]. We also introduce a novel perceptual interpolation scheme that uses the embedding obtained from human subject responses and the geometry of the space of BRDFs to provide the user with an intuitive interface for navigating the space of reflectances and constructing new ones.

The structure of the article, is as follows. We begin with a survey of related works in Section 2. In Section 3 we present our experimental framework for measuring the perception of gloss. Our new method for analyzing perception of gloss is described in Section 4. An analysis of the resulting embedding is presented in Section 5. Our method for perceptual interpolation is described in Section 6. We conclude with a discussion in Section 7.

## 2. RELATED WORK

With the increasing emphasis on photorealism, perception has become an active subject of research in a number of areas of computer

graphics including ray distribution for global illumination [Stokes et al. 2004], tone mapping [Tumblin and Rushmeier 1993; Ledda et al. 2005], evaluation of translucency [Fleming et al. 2004a], and perception of reflectance [Pellacini et al. 2000; Westlund and Meyer 2001]. In this section, we survey some of the recent work in computer graphics and vision science on the perception of reflectance. We refer the interested reader to Pellacini et al. [2000] and Westlund and Meyer [2001] for a more complete survey of the historical developments in this area.

In computer graphics the study of the perception of reflectance was pioneered by Pellacini et al. [2000], in which they present a perceptually meaningful reparameterization of the Ward reflectance model. The authors collected perceptual data by asking subjects to rate the similarity between images generated with varying model parameters. Multidimensional scaling was then used on these ratings to obtain a perceptual model with two parameters. The parameters roughly correspond to Hunter's contrast gloss and DOI gloss. The original Ward parameters for roughness and contrast gloss are independent, while contrast gloss depends on both specular and diffuse reflectance—most likely because the human visual system is sensitive to relative luminance. The study was entirely based on a single empirical BRDF model. Our work, while similar in spirit, is based entirely on measured reflectances and a much larger set of subjects. We will also argue in the next section that our psychometric study based on paired comparisons is a significant improvement over the ratings system employed by Pellacini et al.

Westlund and Meyer [2001] used appearance standards to build a new method for representing BRDFs. Each BRDF is represented by a set of two types of measurements: measurements for specular gloss and haze (ratio of the specular peak to the light a few degrees off specular), and measurements for flop (chromatic effects as seen in pearlescent and metallic paints). Color was measured both at specular peak and off specular followed by interpolation in CIELAB space. The model allows for simple representation and much simpler measurement of materials and inherits the psychophysics based qualities of the individual components (e.g., each step in the interval from 0 to 100 in the gloss dimension equals a uniform step in gloss space), though there is no attempt to capture the perceptual effects of different combinations of the dimensions. In more recent work Shimizu et al. [2003] have developed an interactive system for navigating the appearance space of metallic paints using face, flop, and travel as controls.

Obein et al. [2004] estimated the perceptual scaling of gloss using a series of 10 black plates[1] arranged into pairs of pairs, and users decided which of the two pairs were more similar. Since they were only interested in specular gloss, they assumed that the data is one-dimensional and used maximum likelihood difference scaling (MLDS) [Maloney and Yang 2003] to get an appropriate scaling of the data that obeyed the similarities observed by the subjects. They found that people are far more sensitive to small changes in low gloss samples and less sensitive in intermediate and high gloss samples.

A significant milestone in the availability of measured reflectance data was the work of Matusik et al. [2003a] who followed up on the work of Marschner et al. [2000]. The authors developed a gantry and used it to measure a number of isotropic materials. A part of this data set is now publicly available. Our work is based on this database of measurements. As part of the same work, the authors also developed a new data driven reflectance model. They explored both linear as well as nonlinear representations. Their model had 45 dimensions in the linear case and 14 dimensions in the nonlinear case. A user test was employed to classify the BRDFs into a number categories, for example, blueness, goldness, metalness; these were used to define trait vectors that were in turn used to navigate the space of BRDFs and assist a user in moving from one BRDF to another.

In terms of methodology, the work that is closest to ours is that of Ledda et al. [2005]. The authors examined the perceptual performance of various tone mapping operators by doing paired comparisons on pairs of tone mapped images displayed on two low dynamic range displays and a high dynamic range display showing the original high dynamic range image. While similar in the methodology of collecting the data, our analysis methods are significantly different, as they were only interested in questions of consistency and overall preferences.

The analysis of paired comparisons, the experimental paradigm used here, has a long history in statistics, psychometrics, and biometrics. This includes work on producing rankings, measuring consistency within and across subjects and MDS methods for ordinal data [David 1988; Kendall and Gibbons 1990; Borg and Groenen 2005]. In this study we are particularly interested in constructing an embedding from paired comparisons. The weighted non-metric MDS algorithm addresses this problem, however it has a number of shortcomings, the most significant of which is that it is based on an iterative majorization procedure that can only find a locally optimal solution to the stress minimization problem it solves. Our work addresses this problem by formulating the ordinal MDS problem as a semidefinite programming problem, which can be solved optimally in polynomial time. This approach has its roots in the work on semidefinite embeddings [Weinberger et al. 2004] and distance function learning from relative comparisons [Schultz and Joachims 2003].

## 3. EXPERIMENTAL FRAMEWORK

In this section we describe the design of our psychophysics study.

Each participant in the study was shown a series of triplets of rendered images with constant geometry and illumination, but with varying BRDFs, and was asked to indicate whether the center image was more similar to the image on the left or to the image on the right (Figure 1 shows a screenshot from one such test). We chose this method, known as *paired comparison*, over the rating method based on a continuous slider in Pellacini et al. [2000].

Rating methods using continuous intervals have been shown to have problems with validity and reliability, and subjects usually require a fair amount of training before the experiment [Kendall and Gibbons 1990]. In particular, each subject it seems has his or her own internal continuous scaling function that confounds the process of integrating responses across subjects; despite its precision, its accuracy is questionable. Paired comparisons on the other hand offer a much simpler task and enjoy far more intra- and inter-subject consistency.

The images used in the experiment all consist of the Stanford bunny [Turk and Levoy 1994] rendered under constant illumination and viewing direction with 55 BRDFs from the MIT/MERL BRDF database [Matusik et al. 2003b]. The database contains a large representative set of materials including metals, paints, fabrics, minerals, synthetics, and organic materials. Examples of some of these BRDFs appear in Figure 2.

We used natural illumination since it has been shown that subjects have more discriminative power under this type of illumination than under simple and/or synthetic lighting [Fleming et al.

[1] Black was chosen so the specular highlight dominates the diffuse component.


Fig. 1. Screen capture from the distance comparison test. The subject is asked to click on the appropriate button to indicate which pair appears more similar, *Left*: Left + Middle, or *Right*: Middle + Right. This mode of input has a number of advantages over the conventional approach of asking the subject to provide a continuous measure of similarity using a slider: (1) the paired comparison is a subjectively easier task, (2) the additional information content in a human specified continuous dissimilarity measure is of questionable value, (3) the mapping between different subjects' similarity scales is unknown a priori.

2003]. We used the illumination conditions that worked best in their experiment. We chose the bunny model because while it is simple, it provides a more varied distribution of surface normal/incident direction combinations than a sphere. Each image was rendered under the same high dynamic range illumination using structured importance sampling [Agarwal et al. 2003]. As in previous work [Fleming et al. 2003; Pellacini et al. 2000], we used Tumblin's rational sigmoid [Tumblin et al. 1999] to map the rendered high dynamic range images to our low dynamic range displays. The images were rendered in color and then converted to grayscale for our experiment. Our displays have a maximum brightness of 180 cd/m$^2$.

As there are over 78,000 possible triplets only a randomly sampled subset of comparisons could be performed. Our study has 75 subjects performing 200 comparisons for a total of 15,000 comparisons (there were a small number of repeated comparisons). None of the authors were subjects. All subjects were unaware of the aim of the experiment and all had normal or corrected to normal vision. The triplets were chosen at random for each subject. All the images and the subject response data will be made available online upon publication of this article.

## 4. ANALYZING PAIRED COMPARISONS

One of the aims of this study is to construct a Euclidean space in which the Euclidean distance between a pair of BRDFs corresponds to the perceptual distance between them. This is not to say that such a space necessarily exists. Indeed there is nothing that suggests a priori that human perception obeys the triangle law. However, the analytical, representational, and computational simplicity of a Euclidean space is attractive enough to warrant an attempt at discovering one that best fits the observations. This is an instance of the problem of multidimensional scaling.

### 4.1 Multidimensional Scaling

Multidimensional scaling (MDS) refers to the general task of assigning Euclidean coordinates to a set of objects such that, given a set of dissimilarity, similarity, or ordinal relations between the objects, the relations are obeyed as closely as possible by the embedded points. This assignment of coordinates is also known as Euclidean embedding. The most well known of the various MDS algorithms is classical multidimensional scaling, where the dissimilarities between points are assumed to be actual Euclidean distances. Appendix A gives a short introduction.

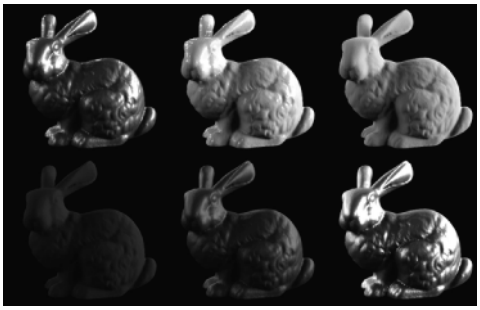

Fig. 2. Example BRDFs. Six of the 55 images used in our psychophysics study. While monochromatic, they have widely varying gloss properties. The BRDFs used include metals, paints, fabrics, minerals, synthetics, and organic materials.

Multidimensional scaling algorithms fall into two broad classes: metric algorithms, which seek an embedding with interpoint distances closely matching the input dissimilarities; and non-metric algorithms, which find an embedding respecting only the relative ordering of the input dissimilarities.

Non-metric MDS (NMDS) has been used extensively in the psychometrics and psychophysics communities to embed similarity and dissimilarity ratings derived from a variety of sources. Metric MDS is not appropriate in many of these applications since the magnitude of the input dissimilarities is unreliable, too difficult to measure, or simply unavailable. The problem of non-metric multidimensional scaling was first considered by Shepard [1962a, 1962b], but it was Kruskal who posed the problem as an optimization problem and introduced an alternating minimization procedure for solving it [Kruskal 1964a, 1964b; Cox and Cox 2000; Borg and Groenen 2005].

A curiosity of the the Shepard-Kruskal formulation of non-metric MDS is that it actually requires magnitudes of the dissimilarities as input, even though NMDS concerns only ordinal information. Indeed, one of the motivations for NMDS is to avoid the difficulties associated with collecting accurate magnitude information. Another quirk of the Shepard-Kruskal formulation is that it requires all order comparisons—one cannot be agnostic about the ordering of any pair of dissimilarities. Collecting all order comparisons may be difficult or impossible in some experiments, as is the case in the current study. Further, the algorithm has no stated time complexity or quality guarantees associated with it.

In the next section we present a new NMDS algorithm that utilizes modern convex optimization theory to solve for a Euclidean embedding in polynomial time. While the tools presented in this section were specifically developed for our study, we hope that they will find broader use in other psychometric and statistical studies.

### 4.2 Multidimensional Scaling for Paired Comparisons

We are now ready to present our method for performing multidimensional scaling on paired comparisons. Our presentation is brief and specific to the case of paired comparisons with one common object. A more general version of the algorithm, further analysis, and examples can be found in Agarwal et al. [2007].

We begin with some notation. We use lower case italicized roman symbols $i, j, k, \ldots$ to indicate scalars and to index into the set of BRDFs. Lower case bold faced symbols $\boldsymbol{x}$ indicate vectors. Upper case symbols P, Q, R, ... are used to denote matrices. The columns

of the matrix X are used to indicate the embedding coordinates for the BRDFs. The matrix K denotes the Gram matrix, $\mathrm{K} = \mathrm{X}^\top\mathrm{X}$. K is a symmetric positive semi-definite matrix, denoted by $\mathrm{K} \succeq 0$.

is the set of all collected observations consisting of 3-tuples $(i, j, k)$, where the subject indicated that perceptually, the image rendered using BRDF $j$ was more similar to the one rendered using BRDF $i$ than it was to the image rendered using the BRDF $k$. Let $\mathrm{D}_{ij}$ denote the perceptual distance between BRDFs $i$ and $j$, then:

$$= \{(i, j, k) | \mathrm{D}_{ij} < \mathrm{D}_{jk}\}. \tag{1}$$

Note that while our experiments do not provide an estimate of $\mathrm{D}_{ij}$, they do provide the inequality relation $\mathrm{D}_{ij} < \mathrm{D}_{jk}$. The set is allowed to have repetitions and inconsistencies.

As in classical MDS we convert the problem into one that can be stated in terms of the Gram matrix K.

$$\begin{aligned} D_{ij}^2 = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 &= \boldsymbol{x}_i^\top \boldsymbol{x}_i - 2\boldsymbol{x}_i^\top \boldsymbol{x}_j + \boldsymbol{x}_j^\top \boldsymbol{x}_j \\ &= \mathrm{K}_{ii} - 2\mathrm{K}_{ij} + \mathrm{K}_{jj}, \end{aligned}$$

where, $\mathrm{K}_{ij}$ is the $(i, j)$-th element of K. Since distances are by definition always non-negative, we can without loss of generality replace the constraint $\mathrm{D}_{ij} < \mathrm{D}_{jk}$ with $\mathrm{D}_{ij}^2 < \mathrm{D}_{jk}^2$, which we can then write in terms of the inner product matrix K as:

$$\mathrm{K}_{ii} - 2\mathrm{K}_{ij} + \mathrm{K}_{jj} < \mathrm{K}_{jj} - 2\mathrm{K}_{jk} + \mathrm{K}_{kk}. \tag{2}$$

Our aim now is to find a Gram matrix, K, that satisfies inequality constraints of this form for every triplet $(i, j, k)$ that is a member of . As we noted earlier, K is symmetric positive semidefinite. This is a necessary and a sufficient condition for K to be the inner product matrix for some set of points.

This set of inequality constraints is not sufficient to uniquely determine a positive semidefinite matrix K. This is because the relative comparison constraint has a scale, translation, and rotation ambiguity. Since Gram matrices are invariant to the rotation of the underlying points, the rotation ambiguity is not a problem. The Gram matrix $K$ however is not invariant to global translation and scaling of the point set. This can lead to numerical instability when solving for the embedding, so steps must be taken to better constrain the solution.

The translation ambiguity is eliminated by demanding that the embedding be centered at the origin, $\forall a = 1, \ldots, n,\ \sum_b \mathrm{X}_{ab} = 0$, which can be restated as:

$$\begin{aligned} \sum_a \left( \sum_b \mathrm{X}_{ab} \right)^2 &= 0, \\ \sum_{bc} \sum_a \mathrm{X}_{ab}\mathrm{X}_{ac} &= 0, \\ \sum_{bc} \mathrm{K}_{bc} &= 0. \end{aligned} \tag{3}$$

This is a linear equation in the entries of matrix K.

Handling the scale ambiguity is a bit more complicated. To prevent the embedding from collapsing into the origin, we constrain the scale of the embedding from below. We will demand that for a relative comparison to be valid the two distances should be different by at least 1 unit distance.

$$\mathrm{K}_{ii} - 2\mathrm{K}_{ij} + \mathrm{K}_{jj} + 1 \leq \mathrm{K}_{jj} - 2\mathrm{K}_{jk} + \mathrm{K}_{kk}. \tag{4}$$

Two things should be noted here. What was a strict inequality in Equation (2) has now been converted into a non-strict one. Secondly, the choice of 1 as the minimum difference between pairs of distances is arbitrary and does not affect the quality of the embedding. The choice of any other constant would result in a uniform scaling of the embedding.

An important consideration when performing MDS is the issue of dimensionality: how many dimensions should the embedding exist in? Ideally, we want the embedding of the smallest possible dimension. There are a number of reasons for this. An obvious one is computational complexity. A lower dimensional embedding is computationally easier to work with and to visualize. A more important reason, however, is that we want our embedding to not only explain the observed data but also to generalize well to unseen data. Statistical learning theory [Vapnik 1998] informs us that for the same training error a simpler model is expected to perform better than a more complex one and should be preferred. For our analysis, the rank of the embedding is its complexity and thus we prefer lower rank embeddings to higher rank ones. The dimensionality of the embedding is the same as the rank of the matrix X, which is in turn the same as the rank of the matrix K. Thus in the ideal case, where we have data that is completely noise free and there exists a Euclidean space in which it can be embedded, we would like to solve the following optimization problem:

$$\begin{aligned} &\arg\min_{\mathrm{K}} \ \mathrm{rank}(\mathrm{K}) \\ \forall(i, j, k) \qquad &\mathrm{K}_{kk} - \mathrm{K}_{ii} + 2\mathrm{K}_{ij} - 2\mathrm{K}_{jk} \geq 1 \\ &\sum_{ab} \mathrm{K}_{ab} = 0, \mathrm{K} \succeq 0. \end{aligned} \tag{E1}$$

This formulation has two problems. First, for the optimization problem to be feasible, there should be a positive semidefinite matrix that satisfies every relative comparison in the collected data. This is clearly not true in general. Second, the rank of a matrix is a non-convex function and therefore ours is a non-convex optimization problem. Indeed, minimizing the rank of a symmetric positive semi-definite matrix subject to linear inequality constraints is an NP-hard problem [Fazel et al. 2004].

To get around the first problem, we introduce slack variables $\xi_{ijk}$ in every inequality constraint, which allows for violations of the inequality and augments the objective function to minimize the total violation:

$$\begin{aligned} &\arg\min_{\mathrm{K},\xi} \sum_{(ijk)} \xi_{ijk} + \lambda\, \mathrm{rank}(\mathrm{K}) \\ \forall(i, j, k) \qquad &\mathrm{K}_{kk} - \mathrm{K}_{ii} + 2\mathrm{K}_{ij} - 2\mathrm{K}_{jk} \geq 1 - \xi_{ijk}, \\ &\xi_{ijk} \geq 0, \quad \sum_{ab} \mathrm{K}_{ab} = 0, \quad \mathrm{K} \succeq 0. \end{aligned} \tag{E2}$$

$\lambda$ is a positive scalar that controls the tradeoff between the violations and the rank of the matrix—the complexity of our model. To deal with the non-convexity of the objective function, we relax the rank function to its convex envelope, the trace. This relaxation is standard in the convex programming literature. Using this relaxation has the additional benefit of constraining the scale of the embedding from above.

$$\begin{aligned} &\arg\min_{\mathrm{K},\xi} \sum_{(ijk)} \xi_{ijk} + \lambda\, \mathrm{tr}(\mathrm{K}) \\ \forall(i, j, k) \qquad &\mathrm{K}_{kk} - \mathrm{K}_{ii} + 2\mathrm{K}_{ij} - 2\mathrm{K}_{jk} \geq 1 - \xi_{ijk}, \\ &\xi_{ijk} \geq 0, \quad \sum_{ab} \mathrm{K}_{ab} = 0, \quad \mathrm{K} \succeq 0. \end{aligned} \tag{E3}$$

There is an intuitive explanation for using the trace of the matrix K as the convex regularizer. The rank of a symmetric matrix can be restated as the number of non-zero eigenvalues, or the $L_0$ (counting) norm of the vector of its eigenvalues. A commonly used convex

relaxation for problems involving finding the sparsest vector is to replace the objective with the $L_1$ norm of this vector. For a symmetric positive semidefinite matrix the trace is exactly that, the $L_1$ norm of the vector of eigenvalues.

While solving problem (E3) is not exactly equivalent to solving problem (E2), as the optimal solution for (E3) can trade a lower trace value for a higher rank, practical experience indicates that this is not a significant problem and trace of a matrix is an excellent heuristic for reducing its rank [Fazel et al. 2004].

The optimization problem (E3) is a semidefinite program (SDP). SDPs are convex optimization problems that are generalizations of linear programming problems and can be solved efficiently and optimally using interior-point methods similar to the ones used for solving linear programs [Nesterov and Nemirousty 1994; Vandenberghe and Boyd 1996]. Efficient solvers exist for solving SDPs [Sturm 1999].

Once K is computed, the embedding itself can be recovered from the eigen-decomposition of K in the same manner as in classical multidimensional scaling, as shown in Appendix A.

## 5. EXPERIMENTS AND ANALYSIS

In this section we present our analysis of the human subject data. We first describe the various sources of error that can arise in a data set such as ours and the results of three independent experiments that we performed to estimate these errors. Next, we describe the perceptual space that results from performing MDS on the set of paired comparisons and discuss its properties. Finally, we show how the embedding correlates with standard measurements for gloss and the parameters of BRDF models fitting to the MIT/MERL database.

### 5.1 Sources of Human Error

As with any study involving human responses, our data is prone to errors and inconsistencies, both inter- and intra-subject. These errors are a function of not only the set of subjects used for our study, but also the particular experimental setup that we use to collect the data, including but not limited to the choice of the BRDFs, geometry, and illumination. In this section we describe three experiments: one that examines inter-subject inconsistencies and two that look at the two types of inconsistencies exhibited by a single subject. These experiments, each carried out with 12 different subjects, were independent of our main experiment, which used 75 subjects.

5.1.1 *Inter-Subject Consistency.* It is possible that not all subjects will agree on the response for any given triplet of images. To get an estimate of how often subjects came to the same conclusion on our data set, we performed a small pilot study in which 12 subjects evaluated the same set of 120 randomly chosen comparisons. We found significant agreement across subjects. The majority vote accounts for about 85% of our total data in this experiment.

5.1.2 *Repeated-Trial Consistency.* Similarly, it is possible that a given subject will exhibit variability in his or her responses when asked to make the same paired comparison multiple times. It is important to estimate this variability to get an idea of the repeatability of the experiment. High variability reduces our trust in the data and the conclusions we can draw from it.

We conducted a study in which 30 random comparisons were chosen and presented 4 times, randomizing both the order of the outer images as well as the order of presentation. Twelve subjects were used for this study. On average, 87% of the time subjects gave the same answer.

5.1.3 *Circular Preferences.* One of the aims of this study is to construct a Euclidean space in which distances correspond to perceptual dissimilarity. In a Euclidean space, every set of unique pairwise distances between a set of points can be ordered without ambiguity. If a subject expresses preference for three BRDFs as $D_{ij} < D_{jk} < D_{ki} < D_{ij}$, they are being circular in their preferences, and these preferences are not representable in a Euclidean space.

Since this type of violation can only be detected if a subject is given all three distance evaluations for a given set of 3 BRDFs, we conducted a study where for each of the 12 subjects a different random set of 40 triplets of BRDFs was randomly selected and the subjects evaluated each of the three distance comparisons. We found that the violation rate was 1.5% on average with a median of 1%.

### 5.2 Learning the Embedding

Despite the set of 55 BRDFs of the MIT-MERL database being a big step forward in terms of measured data availability, it is a only fraction of the space of BRDFs. It is therefore important that care is taken before making any inferences from it. The inferences we make should explain both the observed data points as well as unobserved portions of the space. Only then can we claim that our conclusions are not merely an artifact of our particular data set.

Given a set of subject responses to paired comparisons on the same 55 BRDFs, we measure the error of an embedding as the average number of paired comparisons that are violated if we use the pairwise distance between BRDFs in the embedded space as our estimate of the distance between them.

The expected error of an estimator over an independent test set is called the *test error* or *generalization error* [Hastie et al. 2001]. The training error is smaller than the testing error—in fact it can be arbitrarily smaller. Thus, when reporting the performance of our statistical estimates from the data, it is important to report an estimate of the test error in addition to reporting the training error.

Another issue that one faces in problems like the one we are solving is that of model selection. In the last section we argued that simpler models or lower complexity estimates are to be preferred to higher complexity ones. However, it is also the case that higher complexity models typically fit the training data better than lower complexity models. In our case the regularization parameter, $\lambda$, controls the complexity of our embedding. But how does one choose the optimal value of $\lambda$? If one could estimate the test error for the various choices of $\lambda$ then one could choose that $\lambda$ for which the test error was the lowest.

The most widely used method for estimating test error is cross-validation [Hastie et al. 2001]. In $k$-fold cross-validation the data set (in our case, the set of human responses) is split into $k$ roughly equal parts. At the $i$th iteration, the model is fitted (the embedding is learned) using $k-1$ parts of the data excluding the $i$th part, which is then used for measuring the test error. The final test error estimate is the mean of the $k$ error estimates obtained in this manner. Typical choices of $k$ are 5 or 10.

We ran our MDS algorithm on the data for varying values of $\lambda$ between 0 and 300, and performed 10-fold cross-validation for each value of $\lambda$. Figure 3(a) plots the training (red) and testing errors (green) as a function of the parameter $\lambda$. Figure 3(b) plots the average rank of the embedding as a function of $\lambda$. As expected, the rank of the embedding goes down as the regularization is increased.

As an additional check for the fact that our data does indeed contain structure, and that we are learning from it, we performed the following control experiment. We generated a new data set by taking each triplet $(i, j, k)$ in our data set and randomly swapping $i$

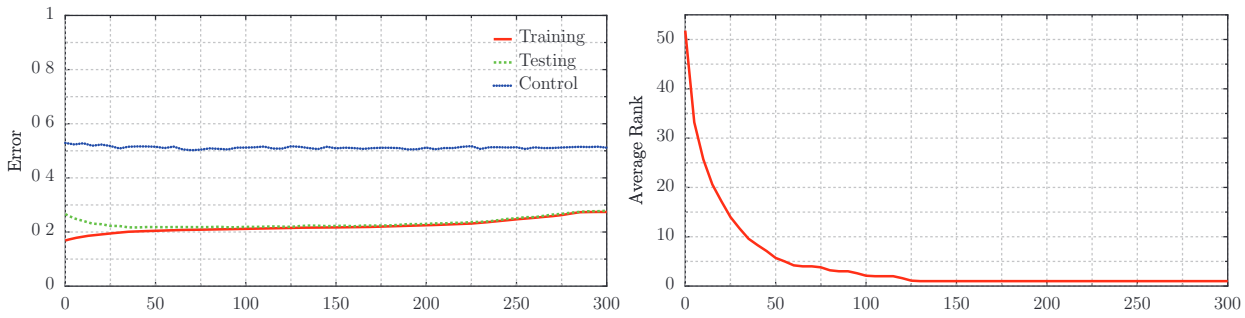


Fig. 3. Cross Validation and Rank. (a) Training (red) and testing (green) error curves for varying choices of the regularization parameter λ for our MDS algorithm. Testing error (blue) for the randomized control set. (b) Average rank as a function of the regularization parameter.

and $k$. This is equivalent to a random observer's response if he or she were shown exactly the same set of comparisons. We then learned an embedding for varying values of λ and measured the test error using cross-validation. In Figure 3(a) the blue curve plots this error. As can be seen, the test error never goes below 50%. The consistent and significant gap between the blue and the green curves indicates that our data set is far from purely random.

The choice of a non-zero λ indicates a tradeoff between the rank of the matrix K and the total amount of violation in the paired comparisons. Setting $\lambda = 0$ would focus the attention of the MDS algorithm entirely on reducing the violations. Doing so results in matrix K, which has a training error of 17%. The resulting embedding has 53 dimensions (which is only 2 less than the maximum). The test error for $\lambda = 0$ is 27%. This is a significant gap and indicates the poor generalization ability of this embedding. The algorithm without any regularization is allowed to come up with a complex model that over-fits to the noise in the training data, resulting in poor performance. As the regularization increases, the training error increases, but the testing error decreases at first and then starts to go back up again; as we increase the penalty for higher rank embeddings, the algorithm trades model complexity for training error. The simpler lower dimensional model does not over-fit to the noise, leading to an increase in generalization performance. As the regularization parameter continues increasing, however, the algorithm is biased too strongly towards choosing a low rank embedding, ultimately being restricted to one dimension. One dimension is not enough to explain the set of relative comparisons, as evidenced by the high test error.

The embedding with the best cross-validation error was obtained for $\lambda = 75$. It has a training error of 21.9% and a test error of 21.3%. The embedding has over 95% of the variance contained in the first two dimensions. Truncating the embedding at two dimensions increased the test error by 0.5%. We do not consider this significant.

To put these numbers in perspective, a trivial upper bound on the test error of an embedding is 50% since a purely random predictor—or even one that gives the same answer every time—will on average get half of the paired comparisons right. A better upper bound of 37.5% is obtained by using the $L_2$ distance between sampled BRDF vectors as the perceptual distance between them. A perceptual metric should perform at least as well as the $L_2$ norm. A lower bound on the minimum possible error is the intersubject error, which is 17%.

## 5.3 Stability of the Embedding

To test the dependence of the embedding on individual BRDFs we performed the following stability analysis. We constructed 55 different embeddings via a leave-one-out procedure, in which each BRDF is omitted in succession, and the remaining 54 BRDFs are used to compute an embedding. Each of the embeddings produced in this manner was then aligned with a similarity transformation[2] to the corresponding 54 points in the final embedding, and the average squared distortion was measured [Umeyama 1991]. To establish a scale for these errors, the average distance between pairs of points in the global embedding was calculated.

The root mean squared distortion was 0.027 and the average distance between points in the global embedding was 0.87. This is an error of 3%, or an order of magnitude difference, which indicates that the embedding is stable.

## 5.4 A Perceptual Space for Gloss

Figure 4 presents a visualization of the optimal 2-D embedding, with cropped windows of the BRDF images displayed in their embedded locations in the new space. Notice the clustering of the BRDFs into two distinct clumps and the similarity among the corresponding images. There are also two pronounced trends in the embedding, a vertical trend with the darker BRDFs at the top and a gradual transition to the brightest BRDFs at the bottom. The horizontal trend roughly breaks the BRDFs into two clusters: the primarily diffuse BRDFs and those that have a strong glossy or specular component.

It is also interesting that the metallic BRDFs are all in the lower left corner and the fabrics are in the upper right corner. This embedding is based entirely on the user preference data—no BRDF or image data was used—which points to the significant descriptive power contained in the paired comparison data.

The low rank of our embedding is interesting when compared to the system of trait vectors used by Matusik et al. [2003b]. Since their system classified full BRDF reflectance (as opposed to only considering gloss) it is difficult to guess how many of the 45 dimensions in the trait vectors would be required to classify the BRDFs according to their gloss properties, though it would likely be much more than the two found in our embedding. We suspect this is because a hand engineered space, while very useful, would likely contain a significant amount of redundancy when compared to an embedding that is constructed purely from psychophysics data.

Given the low fitting error, we know that that our embedding is consistent with the user preferences from the psychophysics experiment. But this says nothing about how our embedding is related to the actual BRDF measurements. In the next two sections, we

[2]Paired comparisons are invariant to similarity transformations.

Fig. 4. Perceptual embedding. The optimal 2-D embedding with cropped windows of the BRDF images displayed in the corresponding embedded coordinates in the new space.

consider the correlation between the coordinates of the perceptual embedding in Section 4 and two different ways of summarizing measured reflectance. In Section 5.4.1 we will consider the nine gloss dimensions described in Table I, and in Section 5.4.2 we will correlate the embedding with the parameters of seven different analytical BRDF models fitted to the measured reflectance. We also consider a perceptual parameterization of Ward's model. In each case we will report the coefficient of multiple correlation between the embedding coordinates and the quantity of interest.

The coefficient of multiple correlation is the multivariate generalization of the coefficient of correlation that measures the strength of the linear relationship between two scalar random variables. Given two scalar random variables, $x$ and $y$, with covariance $\sigma_{xy}$ and standard deviations $\sigma_x$ and $\sigma_y$, respectively, the coefficient of correlation, $r$, is defined as:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}. \tag{5}$$

When $x$ is a vector random variable with correlation matrix $C_x$, and $c$ is the vector of correlation between the coordinates of $x$ and $y$, then the coefficient of multiple correlation $R$ is defined as:

$$R = \left[c^\top C_x^{-1} c\right]^{1/2}. \tag{6}$$

In the interest of brevity, we will refer to $R$ as the correlation.

5.4.1 *Comparison with Gloss Standards.* Starting with the work of Hunter and Harold [1987], a number of different measures of gloss have been defined over the years [ASTM 1999, 2004, 2005b; Hunter and Harold 1987]. While we would not expect our embedding coordinates to correspond directly to any of the gloss measurements (especially considering the fact that our embedding is only defined up to an arbitrary rotation), if our embedding is capturing some of the perceptual qualities of gloss, we would expect there to be correlation between our embedding coordinates and some of the gloss measurements.

Figure 5 plots nine of the gloss measurements in our embedding space mentioned in Table I. The position of each circle corresponds to one of the BRDFs in the embedding space and the diameters correspond to the gloss measurements from Table I. We chose to plot these nine different gloss measurements as a combination of the measurements corresponding to the gloss dimensions of Hunter and Harold [1987] and the ASTM gloss dimensions mentioned as significant in previous work [Pellacini et al. 2000; Westlund and Meyer 2001]. Below each plot we also report the correlation between the embedding coordinates and the corresponding gloss measurement. For the the first eight gloss measurements in Table I we observed that the log of the measurement was significantly more correlated than the raw measurement. Therefore, Figures 5(a)–(h) are plotted on a logarithmic scale. The corresponding correlations are also calculated on the log of the measurements. The only exception is contrast gloss (or luster), which interestingly is the only measurement that is the ratio of reflected light along two directions.

Figure 5(a–d) shows the measurements of each BRDF for three types of specular gloss—the perceived brightness associated with the specular reflection from a surface at 20°, 45°, and 60°, and measurements for sheen—specular gloss measured at 85°. To calculate these values, we measured the reflectance for light that is incident at the appropriate angle (20°, 45°, 60°, and 85°) and viewed from the direction of specular reflection (20°, 45°, 60°, and 85°, respectively).

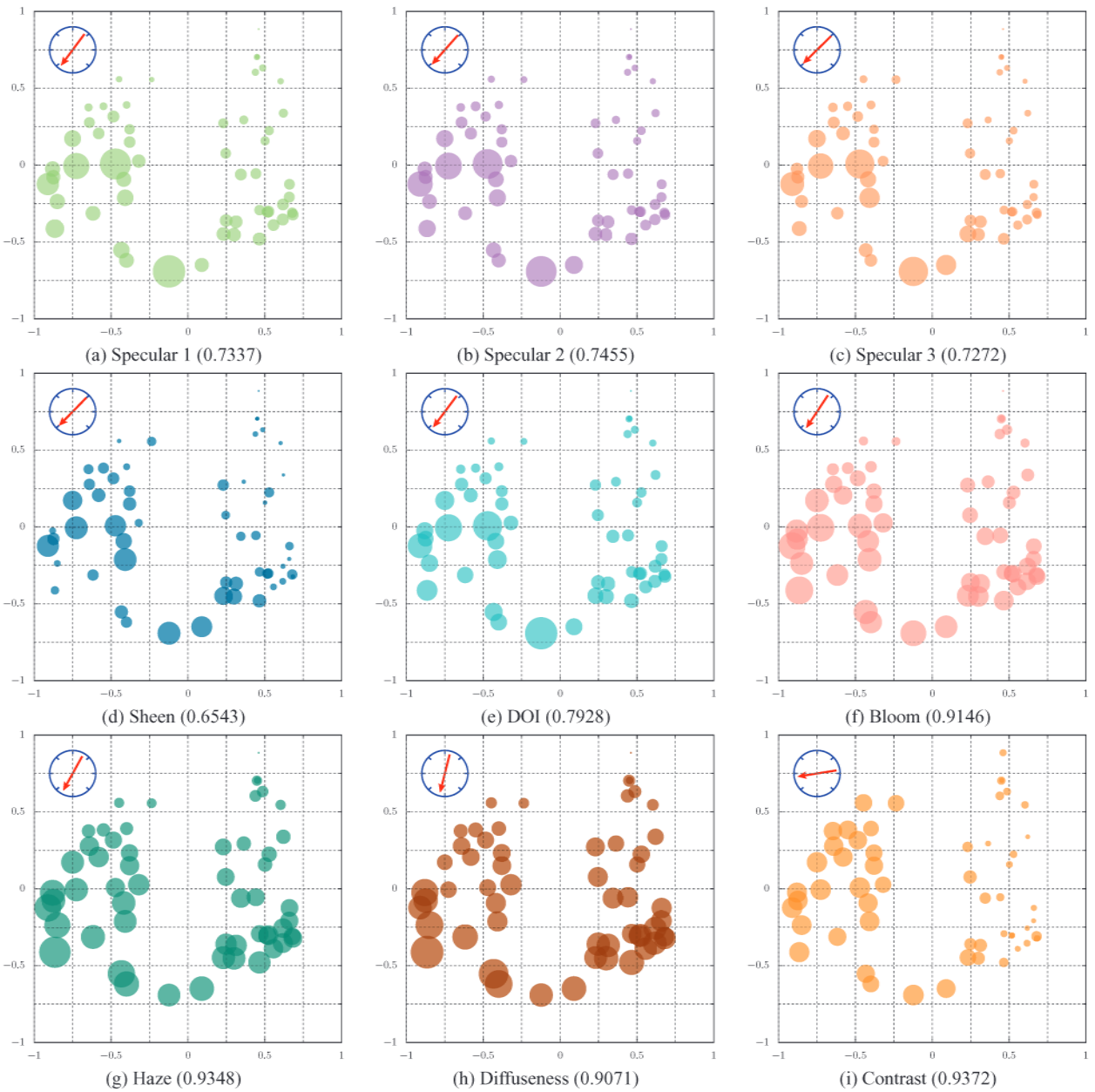


Fig. 5. Correlation between the embedding and formal gloss measurements. Figures (a)-(i) plot nine formal gloss measurements on the perceptual embedding. The diameter of the circles corresponds to the value for each property. The direction of the red arrow in the top left corner of each plot shoes the direction of maximum correlation. There are three pronounced trends that are visible in these plots: the diagonal trend in plots (a)–(d), the increasingly vertical trends in plots (e)–(h), and the horizontal trend in plot (i). These correspond to the trends observed in Figure 4 of increasing strength of the specular highlight moving from the upper right to the lower left, the increase in body color moving from top to bottom, and the increase in apparent shininess moving from right to left, respectively.

Table II. Correlation with BRDF Models
This Table Shows the Correlation Values for our Embedding with a Variety of BRDF Models. Each BRDF has a Set of Parameters Including a Diffuse and Specular Value and up to 3 Auxiliary Parameters (these are different for each BRDF). The BRDF-Specific Symbols for each Parameter are Shown in the Last Column

| BRDF Model | Diffuse | Specular | p0 | p1 | p2 | Parameter Symbols |
|---|---|---|---|---|---|---|
| Ashikhmin-Shirley | 0.9318 | 0.1673 | 0.1639 | 0.2677 | NA | $F_0$,$n$ |
| Blinn-Phong | 0.9646 | 0.6159 | 0.2575 | NA | NA | $n$ |
| Cook-Torrance | 0.9539 | 0.0869 | 0.2760 | 0.7250 | NA | $F_0$,$m$ |
| He et al. | 0.9558 | 0.0692 | 0.3049 | 0.5664 | 0.3442 | $\tau$,$\sigma$,$n$ |
| LaFortune et al. | 0.9556 | 0.3733 | 0.7651 | 0.6803 | 0.2933 | $C_{xy}$,$C_z$,$n$ |
| Ward-Duer | 0.9659 | 0.5167 | 0.6112 | NA | NA | $\alpha$ |
| Ward | 0.9574 | 0.5111 | 0.7391 | NA | NA | $\alpha$ |
| Pellacini et al. | NA | NA | 0.7391 | 0.6226 | NA | $d$,$c$ |

The measurements exhibit a trend increasing from the lower left corner to the upper right corner. This correlates with the trend we noticed before, with the glossy materials on the left and the metallic materials in the lower left corner. The correlation numbers are strong and similar for each of these measurements, with sheen having a lower correlation number; this may be due to the difficulty in measuring specular reflection near grazing angles [Matusik et al. 2003b].

Figure 5(e)–(h) shows the measurements of each BRDF for DIO gloss, bloom, haze, and diffuseness. To calculate these values, we measured the reflectance for light that is incident at 30° and viewed from a slightly off specular angle—30.3°, 32°, 35°, and 45° for DIO gloss, bloom, haze, and diffuseness, respectively. There is a strong trend in each plot that roughly increases from the lower left corner to the upper right corner. It is interesting to note that as the viewing angle used to compute the measurement moves further from the specular direction, the trend becomes increasingly vertical. As we move further from the specular direction, we get closer to measuring the diffuse reflectance, which gives rise to the perceived body color of the object and is exhibited by the vertical trend from dark to light. Also, since these measurements involve light that is reflected off specular, they may be less sensitive to noise; this may explain the smaller number of outliers as compared to the measurements taken on specular.

Figure 5(i) shows the measurements of each BRDF for contrast gloss—the perceived relative brightness of specularly and diffusely reflecting areas. To calculate this value, we measured the reflection of light that is incident at 45° in both the specular direction (45°) and off specular (0°) and compute their ratio; we then subtract this value from one to obtain the gloss value. Since contrast gloss is the ratio of light reflected far from the specular direction to the light reflected in the specular direction, it will be lower for matte materials and higher for materials with a strong specular component. Notice that there is a strong horizontal trend with contrast gloss increasing from right to left. It is also worth noting that if we consider contrast gloss and diffuseness together, we have two very strong correlations that point in almost orthogonal directions. It seems that much of what is captured by our embedding is described by these two measurements.

While it is interesting that we observe trends in our embedding that correlate well with the major gloss measurements, we expect that the redundancy in these trends is related to the sparsity of our BRDF database. Since the observers were offering judgements based on a limited set of materials, subtleties like the difference in specular reflection at 20° versus specular reflection at 85°—a difference often resulting from Fresnel effects—are ignored for higher level distinctions. We expect that as the size of the reflectance database increases, many of these subtleties would begin to affect the shape and arrangement of the embedding, and likely the dimensionality, as well.

5.4.2 *Relation to Fitted BRDF Parameters.* Ngan et al. [2005][3] have fitted seven different analytical models to the MIT/MERL BRDF database. We compare the embedding to these seven models as well as to the model of Pellacini et al. [2000], which is a function of the parameters of the Ward model (one of the seven models). While only Pellacini's model was explicitly designed to model material perception, we would expect that many of the parameters are related to perceptual qualities of the materials since they were designed to fit properties that the authors observed about the materials they were trying to model.

Table II shows the correlation results between the eight BRDF models and our embedding. Each BRDF model has up to 5 parameters and the last column in the table contains the symbols commonly associated with each parameter. We refer the reader to Ngan et al. [2005] and Pellacini et al. [2000] for more details about each of the parameters.

It is important to note that since correlation models linear relationships, low correlation values (below 0.5) can be due to two possible separate reasons. Low correlation can occur if two models are not related and one does not explain the other well. In can also occur if two models are related but that relationship is not linear. Since it can be difficult to distinguish between these two causes, we will focus on analyzing the parameters that have high correlation.

The first and most obvious source of correlation is the correlation with the diffuse component. This is a very high correlation and likely supports the presence of the approximately vertical trend we observed in our embedding going from light to dark.

Another source of strong correlation are the "roughness" parameters of the Ward, Ward-Duer, and Cook-Torrance models ($p0$,$p0$, and $p1$, respectively). Since these parameters approximately capture the spread of the specular lobes in their respective models, we would expect this to be a perceptually significant parameter. Pellacini et al. also found this parameter to be significant. The parameter $d$ in their model (found in $p0$) is actually $(1-\alpha)$, where $\alpha$ is the parameter of the Ward model in $p0$.

The other parameter of Pellacini's model ($c$, found in $p1$) has only moderate correlation with our model. Since the authors classify this parameter as approximately representing contrast gloss, this is somewhat surprising given the strong trend seen for contrast gloss in

[3]Ngan et al. provide fits for only 50 of the 55 BRDFs that were used in our study.

Figure 5(i). The correlation between our embedding and Pellacini's model is interesting, but given the differences in experiments and embedding techniques, perfect correlation between the two methods should not be expected.

## 6. PERCEPTUAL INTERPOLATION

As pointed out by Matusik in his thesis, the set of BRDFs is convex [Matusik 2003]: given any two BRDFs, $\boldsymbol{x}$ and $\boldsymbol{y}$, and a scalar $0 \leq \mu \leq 1$, $\mu\boldsymbol{x} + (1-\mu)\boldsymbol{y}$ is a mathematically valid BRDF. Thus, given a set of BRDFs, measured or otherwise, a simple way to generate new BRDFs is to compute all convex combinations of them. This approach, however, has two problems. First, arbitrary convex combinations, while mathematically correct, can result in physically implausible BRDFs [Matusik et al. 2003b]. That is, while the resulting BRDF may obey the mathematical properties of a valid BRDF (energy conservation, reciprocity, etc.), it may represent a material that is not likely to be found in nature. Second, one is typically interested in producing materials with properties close to some known collection of BRDFs; in such a case one would like the combination of weights to correlate with perceptual distance to the basis BRDFs, thus making the combination process intuitive and useful in practice. However, there is nothing to suggest that a linear algebraic combination of two BRDFs translates into a perceptual combination of their properties. Once again it is useful to make an analogy with color perception. There are colors that appear to be both red and blue (purples), both blue and green (blue-greens) and both yellow and green (yellow-greens). There is no color, however, that is subjectively the combination of red and green [Palmer 1999]. Thus we must be careful in our use of linearity when combining perceptual properties; in particular we should avoid linearly combining objects that are perceptually far apart.

Having a low dimensional perceptual space in which known BRDFs are embedded offers a solution to the problem of perceptual BRDF design. An artist wanting to design a new material can now easily move around in this space and indicate the desired perceptual position of the BRDF by indicating how close it is to the known BRDFs. Of course this requires the ability to generate a BRDF from its position in the perceptual embedding. Since we are only given a sparse sampling of the space of BRDFs, we must construct a perceptual interpolation scheme that uses the geometry of the embedding to interpolate over the measured BRDF data.

Given a point in the perceptual space, one naive solution would be to select the $k$-nearest neighbors of that point from among the set of BRDFs. The distance used for determining the neighbors is the Euclidean distance in the perceptual space. The point corresponding to the desired BRDF may or may not lie in the convex hull of its nearest neighbors and it is not clear what weighting scheme should be used to interpolate between the BRDFs.

Our solution to the problem is to start by first constructing a Delaunay triangulation of the space [Okabe et al. 1992] using the materials in our embedding as vertices. Delaunay triangulation constructs a natural neighborhood structure on the embedding by maximizing the minimum angle of all angles of the triangulation. This ensures that long thin triangles connecting far ends of the embedding are avoided. Now when the user specifies a point in the space, we select the Delaunay triangle containing the point and use its barycentric coordinates to linearly interpolate between the vertex BRDFs. Barycentric coordinates sum to one, thus the resulting interpolant is a convex combination of the vertex BRDF and thereby it is a mathematically valid BRDF. This means for any point inside the convex hull of the embedding, we need at most three BRDFs to generate a perceptual interpolation for it. Figure 6 illustrates this process.

An interesting consequence of this interpolation scheme is that even if a point in the perceptual space lies on the line joining two BRDFs, the interpolated value of a BRDF could be the result of three entirely different BRDFs. Figure 6(a) illustrates this phenomenon. Thus, even though our interpolation process is locally linear, the overall interpolation is a nonlinear process.

The validity of our interpolation procedure is a function of the density of our data. If we had a very dense set of BRDFs, the locally linear interpolation would give rise to very little deviation from the underlying nonlinear space. While our data set is the largest collection of measured materials, it is quite sparse in relation to the space of gloss measurements for all materials. As a result, there are portions of the embedding that contain gaps—linear interpolation within these gaps will likely lead to deviation from the true underlying nonlinear space and possibly result in interpolation artifacts. However, in practice we have seen that the interpolations are quite smooth. This is seen in Figure 7, which shows a collage of images obtained by uniformly sampling within the convex hull of the perceptual embedding. At each point we calculate the interpolated BRDF using the previously described perceptual interpolation procedure and then use it to render the corresponding image. Notice that overall there are smooth transitions from one portion of the embedding to another.

Figure 8 shows a comparison between our interpolation and simple linear interpolation. Note that the primary difference in this case is in the specular peak and the overall impression of glossy versus matte. It is obvious that the two interpolations are different, but what can we say about the perceptual quality of the two interpolations? To be a perceptually linear interpolation, we would expect that since the intermediate materials are equally spaced, the step size from one material to another would be approximately constant. Each of the steps in the perceptual interpolation are somewhat small and appear to be of fairly equal magnitude. The largest step appears to be between the sixth and seventh materials where there is an increase in the strength of the specular peak. The interpolation on the bottom, on the other hand, appears to begin with a very large step and finish with six very small steps. The transition from matte to specular appears to occur almost completely in the first step. This would seem to indicate that our interpolation is more perceptually uniform than the linear interpolation.

We note that if the computational overhead of using measured BRDFs is too much for a particular application, it is simple to replace each measured BRDF with the best fitting empirical model and return to the user the parameters of the best fitting BRDFs at the vertices of the chosen Delaunay triangle and the three combination weights. As our understanding of the space of perception improves and we construct more detailed and perhaps higher dimensional models, our perceptual interpolation scheme will extend naturally. In the higher dimensional case, one can replace the Delaunay triangulation in the plane with the n-space generalization [Okabe et al. 1992], though navigation of a space with more than three dimensions is a tricky user interface design problem.

### 6.1 Integration with Color

While we are aware that the perception of gloss is affected by the surface color, one can often decouple the chromatic and achromatic parts of the BRDF. Following Pellacini et al. [2000], we can integrate our perceptual model of surface gloss with color by assuming that gloss and chromaticity are approximately independent [Aida 1997; ASTM 2003]. We use our method to interpolate the $L$ channel in perceptual space and then choose two example BRDFs to use as the endpoints for color interpolation in the $a$ and $b$ channels of

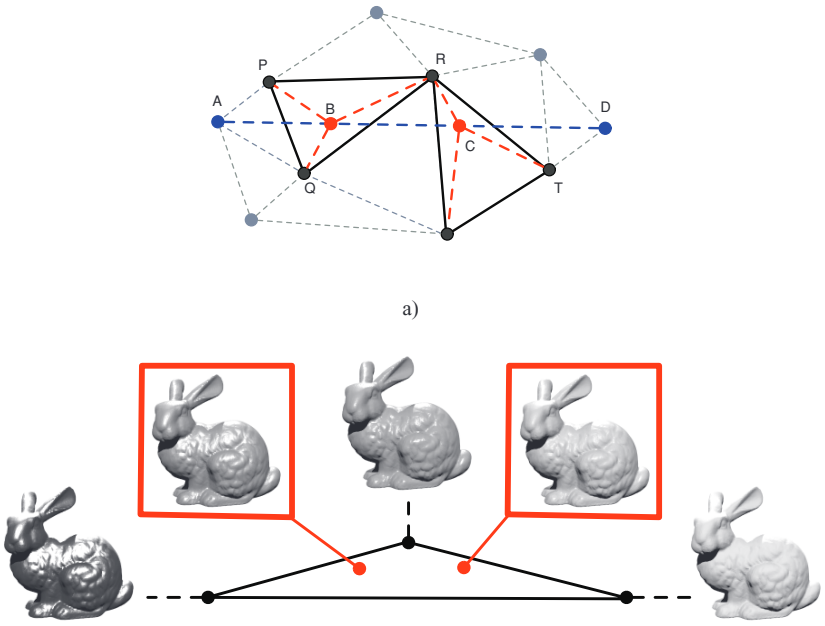


Fig. 6. Perceptual interpolation. Figure (a) illustrates the underlying geometrical method used for performing the perceptual interpolation between two BRDFs. Notice that in this case neither of the two end points (A and D) are used in the interpolation process and only one BRDF (R) is shared for the two interpolations (B and C). This illustrates the locally linear yet highly nonlinear nature of perceptual interpolation. Figure (b) illustrates the interpolation that occurs in a single triangle. The interpolated BRDFs (shown as red dots) are generated as linear combinations of the three end points (shown as black dots). Note that the selected BRDFs are further from each other than the BRDFs that form a typical triangle in our space and were chosen for illustration.

the CIELAB color space. Figure 9 shows an example of this interpolation. Note that while the color interpolation is a simple linear interpolation in color space based on the two endpoint BRDFs, the interpolation for gloss is based on interpolation between five intermediate BRDFs, as shown in Figure 6(b).

## 7. DISCUSSION

In this study we have presented the results of a psychophysical study of the perception of achromatic isotropic reflectance. The study uses the largest publicly available data set of measured reflectances. We introduced a novel MDS algorithm for analyzing the data we collected. This algorithm is an efficient and more general replacement for the widely used non-metric MDS algorithm.

Analysis of our data-set using this algorithm revealed a two-dimensional perceptual embedding. The embedding captures a large fraction of human subject responses indicating that at least the gross structure of the perceptual space of reflectance can be captured by approximating it with a low-dimensional Euclidean space. We compared and correlated this embedding with existing work on perceptual parameterization of the Ward model, the parameters of seven analytical BRDF models fitted to the MIT/MERL database, and nine gloss dimensions. In each case we observed a strong correlation with the perceptual embedding.

We also introduced a novel perceptual interpolation scheme that uses the geometric structure of this embedding to interpolate between BRDFs. This procedure is computationally efficient and locally linear. We showed how this scheme performs better than simple linear interpolation from one target BRDF to another.

We are aware that the small size of the BRDF database, and the use of a fixed geometry, illumination, and viewpoint to evaluate the various BRDFs limits the scope of our study and the strength of the conclusions we can draw from it. But for reasons of tractability, any study of this kind has to hold certain variables constant.

Despite the limited scope of our experiments, it is interesting that significant correlations with existing work can still be observed. The design and execution of experiments that investigate how the perception of a reflectance function varies as illumination, viewpoint, and geometry are allowed to change remains a subject of continuing research [Adelson 2001; Knill and Kersten 1991; Fleming et al. 2004b; Vangorp et al. 2007]. We hope that this work will set the stage for larger and more elaborate studies of this kind in the future.

There are a number of very interesting avenues for future work; we briefly mention some of them here. One interesting direction to explore is the effect of motion on the perception of reflectance [Hartung and Kersten 2002]. It is often the dynamic reflectance that makes it easier to spot fakes in rendered scenes. It may be interesting to show users rotating geometry and/or cameras to see the effect on discrimination by the user. Further, one can ask, how does the scale and magnification of the object affect the ability to discriminate between materials? This has important implications on level-of-detail research. Also in talking to the test subjects after the experiment, many of them mentioned that they became quite proficient at quickly examining certain portions of each of the images. It will be interesting to monitor the eye movement of the subjects and study how reflectance and image saliency are related.

The perceptual interpolation procedure described in this article is a first step towards constructing an easily navigable space for reflectance. Part of the ease of navigation of this space is due to its two

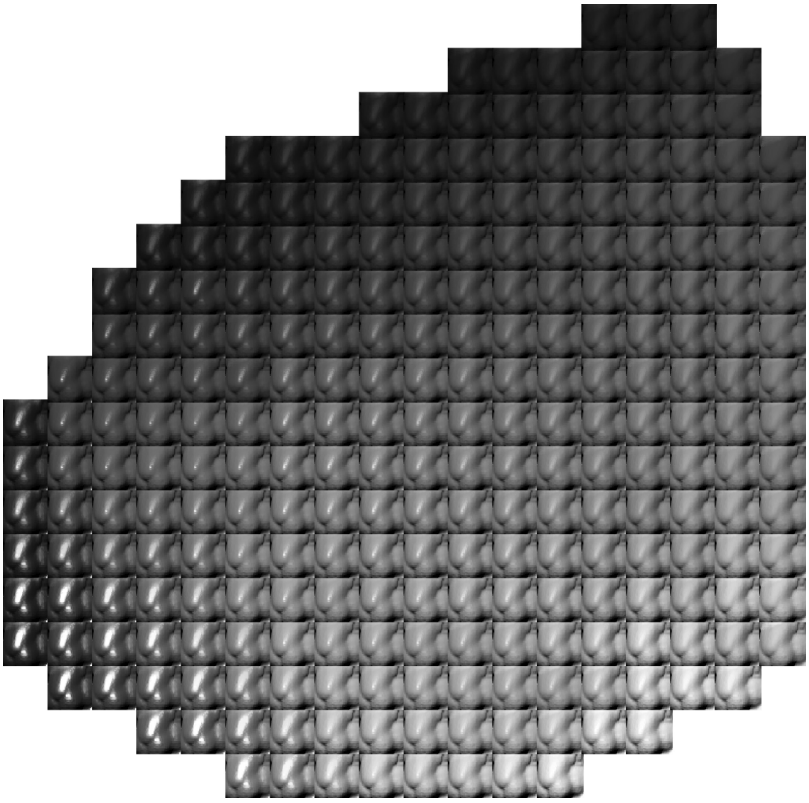

Fig. 7. Uniform perceptual sampling. The convex hull of the perceptual embedding was resampled, for each point a new BRDF was generated using the perceptual interpolation procedure that was then used to render the nose of the Stanford bunny.

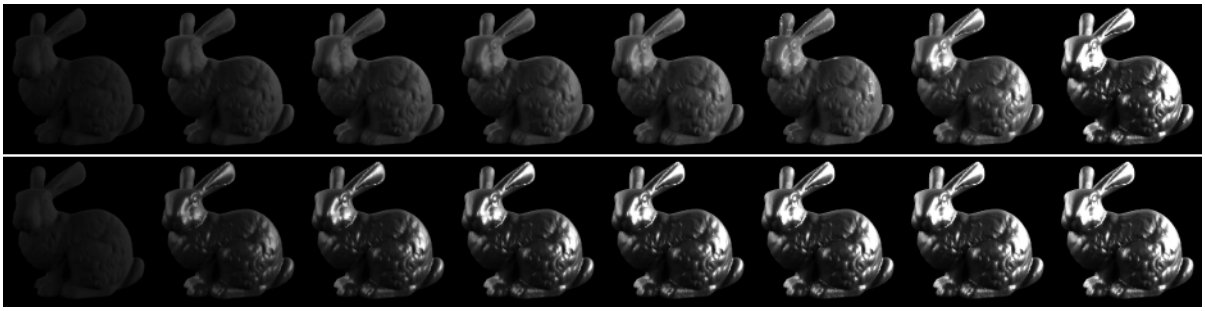

Fig. 8. Perceptual vs. linear Interpolation. The top row shows a perceptual interpolation (left to right) from a very matte material to one that is very specular. The bottom row shows the linear interpolation. The primary difference in this case is in the specular peak and the overall impression of glossy vs. matte. In the perceptual interpolation (top) the change in the specularity is more gradual whereas linear interpolation (bottom) jumps to a glossy material in just one step from the matte material.

dimensional structure. As we model finer scale structure variations, the dimensionality of this space will likely go up, which will necessitate novel user interfaces for navigating them.

The MDS algorithm we have proposed in this work is a novel and general tool that we expect to have applications in various fields including psychophysics, vision, and graphics. It is simple to extend it to other experimental setups like pairs of pairs and complete ranking [Agarwal et al. 2007]. A similar formulation can perhaps be used to learn empirical distance functions that will allow us to measure perceptual distance between two previously unseen BRDFs without performing an additional psychophysical study. Another direction of work is the so called out-of-sample extension; given an embedding with $n$ BRDFs, and a new BRDF with some paired comparisons relating it to these BRDFs, can we find its position in this embedding without calculating the entire embedding from scratch?

Fig. 9. Perceptual interpolation: four Buddhas rendered in the Galileo environment. The images on the far right and left are rendered from real measured BRDFs (aluminum bronze and teflon, repectively). The images in between are rendered using BRDFs that are constructed by perceptually interpolating the measured BRDFs in our perceptual space.

## APPENDIX

## A. CLASSICAL MULTIDIMENSIONAL SCALING

Let D be an $n \times n$ matrix of pairwise distances. The matrix $D$ is symmetric with a zero diagonal. We are interested in finding a $d \times n$ matrix X where each column $\boldsymbol{x}_i$ is the representation of the point $i$ in $R^d$ and $\mathrm{D}_{ij} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$. Denote the inner product (or Gram matrix) for this set of points by $\mathrm{K} = \mathrm{X}^\top \mathrm{X}$.

K is an $n \times n$ symmetric positive semidefinite matrix. Let us now abuse notation and use $\mathrm{D}^2$ to indicate the matrix of squared pairwise distances $\mathrm{K} = -\frac{1}{2}(\mathrm{I} - 11^\top)\mathrm{D}^2(\mathrm{I} - 11^\top)$. Here, I is the $n \times n$ identity matrix and 1 is the $n$-vector of all ones. In light of this, solution to the classical multidimensional scaling problem is straightforward. Given the eigenvalue decomposition $\mathrm{K} = \mathrm{U}\ \ \mathrm{U}^\top$, it follows that $\mathrm{X} = \mathrm{U}\ \ ^{1/2}$. The solution so obtained is ambiguous to a global rotation.

### ACKNOWLEDGMENTS

It is a pleasure to acknowledge our discussions with Prof. Gert Lanckriet, Prof. Don Macleod, Prof. Diana Deutsch, Lawrence Cayton and Kristin Branson.

### REFERENCES

ADELSON, E. H. 2001. On seeing stuff: The perception of materials by humans and machines. In *Proceedings of the International Conference on Human Vision and Electronic Imaging (SPIE)*. Vol. 4299. 1–12.

AGARWAL, S., RAMAMOORTHI, R., BELONGIE, S., AND JENSEN, H. W. 2003. Structured importance sampling of environment maps. *ACM Trans. Graph. 22*, 3, 605–612.

AGARWAL, S., WILLS, J., CAYTON, L., LANCKRIET, G., KRIEGMAN, D., AND CAYTON, L. 2007. Generalized non-metric multidimensional scaling. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*.

AIDA, T. 1997. Glossiness of colored papers and its application to specular glossiness measuring instruments. *Syst. Comput. Japan 28*, 1, 95–112.

ASHIKHMIN, M., PREMOZE, S., AND SHIRLEY, P. 2000. A microfacet-based brdf generator. In *Proceedings of the International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*. 65–74.

ASTM. 1999. *Standard Test Method for Specular Gloss* (E523-89(1999)). ASTM International.

ASTM. 2003. *Standard Practice for Establishing COLOR and GLOSS TOLERANCEs* (D3134-97(2003)). ASTM International.

ASTM. 2004. *Standard Test Method for 45-deg Specular Gloss of Ceramic Materials* (ASTM C346-87(2004)). ASTM International.

ASTM. 2005a. *Standard Termninology of Appearance* (E284-05a). ASTM International.

ASTM. 2005b. *Standard Test Methods for Measurement of Gloss and High-Gloss Surfaces by Abridged Goniophotometry* (E430-05). ASTM International.

BEN-ARTZI, A., OVERBECK, R., AND RAMAMOORTHI, R. 2006. Real-time BRDF editing in complex lighting. In *ACM Trans. Graph. 25*, 3, 945–954.

BORG, I. AND GROENEN, P. 2005. *Modern Multidimensional Scaling: Theory and Applications*. Springer Verlag.

COOK, R. L. AND TORRANCE, K. E. 1981. A reflectance model for computer graphics. *Comput. Graph. (SIGGRAPH) 15*, 4, 187–196.

COX, T. AND COX, M. 2000. *Multidimensional Scaling*. Chapman & Hall/CRC.

DAVID, H. A. 1988. *The Method of Paired Comparisons* 2nd Ed. Chapman and Hall, London.

DUMONT, R., PELLACINI, F., AND FERWERDA, J. 2003. Perceptually-driven decision theory for interactive realistic rendering. *ACM Trans. Graph. 22*, 2, 152–181.

FAZEL, M., HINDI, H., AND BOYD, S. 2004. Rank minimization and applications in system theory. In *Proceedings of the Americal Control Conference*.

FLEMING, R. W., DROR, R. O., AND ADELSON, E. H. 2003. Real-world illumination and the perception of surface reflectance properties. *J. Vis. 3*, 5, 347–368.

FLEMING, R. W., JENSEN, H. W., AND BULTHOFF, H. H. 2004a. Perceiving translucent materials. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization (APGV)*. 127–134.

FLEMING, R. W., TORRALBA, A., AND ADELSON, E. H. 2004b. Specular reflections and the perception of shape. *J. Vis. 4*, 9, 798–820.

GIBSON, S. AND HUBBOLD, R. 1997. Perceptually-driven radiosity. *Comput. Graph. Forum 16*, 2, 129–141.

HARTUNG, B. AND KERSTEN, D. 2002. Distinguishing shiny from matte. *J. Vis. 2*, 7, 551.

HASTIE, T., TIBSHIRANI, R., AND FRIEDMAN, J. 2001. *The Elements of Statistical Learning*. Springer Verlag.

HE, X. D., TORRANCE, K. E., SILLION, F. X., AND GREENBERG, D. P. 1991. A comprehensive physical model for light reflection. *Comput. Graph. (SIGGRAPH) 25*, 4, 175–186.

HUNTER, R. S. AND HAROLD, R. W. 1987. *The Measurement of Appearance*. Wiley.

KENDALL, M. AND GIBBONS, K. D. 1990. *Rank Correlation Methods*. Oxford University Press, UK.

KNILL, D. AND KERSTEN, D. 1991. Apparent surface curvature affects lightness perception. *Nature 351*, 228–230.

KRUSKAL, J. B. 1964a. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika 29*, 1–27.

KRUSKAL, J. B. 1964b. Nonmetric multidimensional scaling: A numerical method. *Psychometrika 29*, 115–129.

LEDDA, P., CHALMERS, A., TROSCIANKO, T., AND SEETZEN, H. 2005. Evaluation of tone mapping operators using a high dynamic range display. In *ACM Trans. Graph. 24*, 3, 640–648.

LUEBKE, D. AND HALLEN, B. 2001. Perceptually driven simplification for interactive rendering. In *Proceedings of the Eurographics Workshop on Rendering*.

MALONEY, L. T. AND YANG, J. N. 2003. Maximum likelihood difference scaling. *J. Vis. 3*, 8 (10), 573–585.

MARSCHNER, S. R., WESTIN, S. H., LAFORTUNE, E. P. F., AND TORRANCE, K. E. 2000. Image–based bidirectional reflectance distribution function measurement. *Applied Optics 39*, 16.

MATUSIK, W. 2003. A data-driven reflectance model. Ph.D. thesis, MIT.

MATUSIK, W., PFISTER, H., BRAND, M., AND MCMILLAN, L. 2003a. Efficient isotropic BRDF measurement. In *Proceedings of the 14th Eurographics Workshop on Rendering*. Eurographics Association. 241–247.

MATUSIK, W., PFISTER, H., BRAND, M., AND MCMILLIAN, L. 2003b. A data-driven reflectance model. In *ACM Trans. Graph. 22*, 3, 759–769.

MYSZKOWSKI, K. 2002. Perception-based global illumination, rendering, and animation techniques. In *Proceedings of the 18th Spring Conference on Computer Graphics (SCCG)*, A. Chalmers, Ed. 13–24.

NESTEROV, Y. E. AND NEMIROUSKY, A. S 1994. *Interior Point Polynomial Algorithms in Convex Programming: Theory and Algorithms*. Siam.

NGAN, A., DURAND, F., AND MATUSIK, W. 2005. Experimental analysis of BRDF models. In *Proceedings of the Eurographics Symposium on Rendering*. Eurographics Association. 117–226.

NGAN, A., DURAND, F., AND MATUSIK, W. 2006. Image-driven navigation of analytical BRDF models. In *Proceedings of the 17th Eurographics Workshop on Rendering*. 399–408.

NICODEMUS, F. E., RICHMOND, J. C., HSIA, J. J., GINSBERG, I. W., AND LIMPERIS, T. 1977. Geometric considerations and nomenclature for reflectance. Monograph 161, National Bureau of Standards.

OBEIN, G., KNOBLAUCH, K., AND VIENOT, F. 2004. Difference scaling of gloss: Nonlinearity, binocularity, and constancy. *J. Vis. 4*, 9 (8), 711–720.

OKABE, A., BOOTS, B., AND SUGIHARA, K. 1992. *Spatial Tesselations: Concepts and Applications of Voronoi Daigrams*. Wiley, New York.

PALMER, S. E. 1999. *Vision Science: Photons to Phenomenology*. MIT Press.

PELLACINI, F., FERWERDA, J. A., AND GREENBERG, D. P. 2000. Toward a psychophysically-based light reflection model for image synthesis. In *Proceedings of SIGGRAPH*. ACM Press, 55–64.

RAMANARAYANAN, G., FERWERDA, J., WALTER, B., AND BALA, K. 2007. Visual equivalence: Towards a new standard for image fidelity. In *ACM SIGGRAPH Papers*. ACM.

SCHULTZ, M. AND JOACHIMS, T. 2003. Learning a distance metric from relative comparisons. In *Proceedings of the Conference on Advance in Neural Information Processing Systems (NIPS)*.

SHEPARD, R. 1962a. The analysis of proximities: Multidimensional scaling with an unkn own distance function. I. *Psychometrika 27*, 2, 125–140.

SHEPARD, R. 1962b. The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika 27*, 219–246.

SHIMIZU, C., MEYER, G. W., AND WINGARD, J. P. 2003. Interactive goniochromatic color design. In *Proceedings of the Color Imaging Conference*. 16–22.

STOKES, W. A., FERWERDA, J. A., WALTER, B., AND GREENBERG, D. P. 2004. Perceptual illumination components: A new approach to efficient, high quality global illumination rendering. In *ACM Trans. Graph. 23*, 3, 742–749.

STURM, J. 1999. Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones. *Optim. Meth. Softw. 11-12*, 625–653.

TORRANCE, K. E. AND SPARROW, E. M. 1967. Theory for off-specular reflection from roughened surfaces. *J. Optic. Soc. Amer. 57*, 1105–1114.

TUMBLIN, J., HODGINS, J. K., AND GUENTER, B. K. 1999. Two methods for display of high contrast images. *ACM Trans. Graph. 18*, 1, 56–94.

TUMBLIN, J. AND RUSHMEIER, H. E. 1993. Tone reproduction for realistic images. *IEEE Comput. Graph. Appl. 13*, 6, 42–48.

TURK, G. AND LEVOY, M. 1994. Zippered polygon meshes from range images. In *Proceedings of SIGGRAPH '94*. 311–318.

UMEYAMA, S. 1991. Least-squares estimation of transformation parameters between two point patterns. *IEEE Trans. Pattern. Anal. Mech. Intell. 13*, 4, 376–380.

VANDENBERGHE, L. AND BOYD, S. 1996. Semidefinite programming. *SIAM Rev. 38*, 1, 49–95.

VANGORP, P., LAURIJSSEN, J., AND DUTRÉ, P. 2007. The influence of shape on the perception of material reflectance. *ACM Trans. Graph. 26*, 3, 77.

VAPNIK, V. 1998. *Statistical Learning Theory*. Wiley, New York.

WARD, G. J. 1992. Measuring and modelling anisotropic reflection. *Comput. Graph. (SIGGRAPH) 26*, 2, 265–272.

WEINBERGER, K. Q., SHA, F., AND SAUL, L. K. 2004. Learning a kernel matrix for nonlinear dimensionality reduction. In *Proceedings of the 21st International Conference on Machine Learning (ICML)*. 839–846.

WESTLUND, H. B. AND MEYER, G. W. 2001. Applying appearance standards to light reflection models. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*. 501–510.

Received April 2006; revised February 2008; accepted March 2009