Learning Perceptual Aesthetics of 3D Shapes from Multiple Views

Kapil Dev
School of Mathematics and Computer Science, Liverpool Hope University, UK

Manfred Lau
School of Creative Media, City University, Hong Kong

***Abstract*—The quantification of 3D shape aesthetics has so far focused on specific shape features and manually defined criteria such as the curvature and the rule of thirds respectively. In this paper, we build a model of 3D shape aesthetics directly from human aesthetics preference data and show it to be well aligned with human perception of aesthetics. To build this model, we first crowdsource a large number of human aesthetics preferences by showing shapes in pairs in an online study and then use the same to build a 3D shape multi-view based deep neural network architecture to allow us learn a measure of 3D shape aesthetics. In comparison to previous approaches, we do not use any pre-defined notions of aesthetics to build our model. Our algorithmically computed measure of shape aesthetics is beneficial to a range of applications in graphics such as search, visualization and scene composition.**

■ **IN OUR** day-to-day lives, we often encounter shapes that trigger universal pleasing emotions in our minds irrespective of our backgrounds and experiences. We call such shapes as "aesthetic shapes". The pleasing emotions aroused by aesthetic shapes impact us in different ways. For example, our decision to buy a consumer product could alone be influenced by how aesthetic it appears. The recent exponential growth of 3D data require us to develop new techniques to help navigate or visualize 3D data-sets for objects of interest (e.g. more beautiful) in a timely manner.

In this paper, we study perceptual aesthetics of 3D shapes. We consider those shapes as aesthetic shapes that are visually attractive or pleasing to the human mind. As we demonstrate, our 3D shape aesthetics measure could benefit a range of applications in computer graphics namely in search, scene composition, and visualization.

While the aesthetics of images has been extensively researched to date, work in 3D shape aesthetics is very limited, mainly exploring aesthetics with manually defined features such as curvature, symmetry and mathematical criteria such as bending energy and minimum variation surfaces [1], [2], [3]. In this work, we drop the notion of manually defined features as such features may be biased or incomplete to capture the aesthetics of a shape. Our approach works by deep learning a model of 3D shape aesthetics directly from human aesthetics preferences.

As crowdsourcing has recently become a popular approach for data collection for various graphics problems [4], we also use popular crowdsourcing platform Amazon Mechnical Turk to collect human aesthetics preference data. Specifically, we show participants shapes in pairs (i.e. a paired comparison test) as stimuli. To

Published by the IEEE Computer Society

which they respond by selecting one shape they think is more attractive or beautiful. We believe paired comparison based test is easier for humans to perform than the other methods such as providing an absolute score [4], [5], [6]. As a metaphor, we try to avoid comparing between "apples and oranges" but wish to compare apples with apples. For example, a chair is paired with a chair and a table is paired a table. The shapes are also made to spin along an up axis (i.e. y-axis) to show views from different directions.

Our deep neural architecture is an extension of multi-view based deep convolutional networks used for 3D shape recognition [7]. As described in Section 3, this architecture fits well with our paired data collection method. Our method is able to learn an aesthetics model directly from shape views than on any manually defined shape features. We use 3D models from the ShapeNet dataset [8] which are already classified into human-understandable categories of man-made objects such as club chairs, mugs, and lamps.

Contributions

The contributions of this paper are: (1) we explore perceptual aesthetics of 3D shapes with human aesthetics judgment data rather than using any pre-defined notion of shape aesthetics; (2) we extend multi-view based deep convolution network used for shape classification to the problem of learning and ranking 3D shape aesthetics from crowdsourced paired shape aesthetics comparison data; (3) we demonstrate that multiple views of 3D shapes is an appropriate representation for building a model of 3D shape aesthetics and (4) we demonstrate our learned measure of aesthetics is well aligned with human perception of aesthetics and can be used in applications such as aesthetics-based search, visualization, and scene composition.

Related Work

We briefly discuss the related works in the following areas: aesthetics using specific properties, aesthetics of images, crowdsourcing perceptual data, and deep learning aesthetics.

Aesthetic Properties

The visual aesthetics of objects has traditionally been examined in terms of specific handcrafted properties such as curvature and symmetry. Previous works have used mathematical tools and psychological experiments to study the relationship between shape properties and aesthetics. A product's geometry's influence on customer's emotional state is studied by Franca Giannini and Marina Monti [9]. In this work, authors specifically focus on stimuli from the automotive and household supplies and present a discussion on designer identified aesthetic properties. To facilitate aesthetic designs, fuzzy logic has been used to build a shape specification system that uses pre-defined aesthetic descriptors for designing shapes [10]. Similarly, several works have also used visual characteristics such as color, light, line, shape, texture, and space and movement to relate to aesthetic experience. Abstract mathematical criteria such as entropy, complexity, and deviation from normality have been defined for aesthetic 3D designs [2]. Geometry characteristics such as lines, curvatures of free surfaces, their deviation ratio etc. have been considered for modeling shapes for engineering product design [11].

Symmetry. The term "symmetry" has been associated with harmony and balance. Studies that have established symmetry as an important feature contributing to aesthetic experience [12]. For example, symmetric polygonal forms have been found to be more attractive by the observers [13]. An evaluation of jewelry designs in terms of aesthetic features such as golden ratio, mirror symmetry, and rotational symmetry has been done by Wannarumon et al. [14].

Curvature. Designs with rounded corners are perceived more aesthetic than designs with sharp edges. Consequently, customers tend to show strong preference for curvilinear products [15]. Researchers have also used mathematical criteria to describe aesthetic surfaces. Specifically, they define fairness metric, bending energy, and minimum variation surface a way to capture the relation between curvature and aesthetic surfaces [3].

The key difference between our work and the works reviewed above is that our technique does not use any pre-existing notions of esthetics as done in other works. Our learning mechanism uses human labeled data in the form of images and learns directly on it, without requiring us to explicitly specify shape feature descriptors such as curvature and symmetry etc. In the results section we show that our system can automatically learn to distinguish shape aesthetics based on their curvature and other features.

Image Aesthetics

There is much previous work in assessment of aesthetics of images, human faces, and paintings; which uses either handcrafted features with pre-selected image descriptors or machine learning directly on image pixels. The aesthetic quality of pictures based on their visual content has been explored by Datta et al. [16]. Specifically, they build a classifier to automate assessment of aesthetics using handcrafted image features such as measure of colorfulness, rule of thirds, saturation and hue, and familiarity measure etc. Use of textual comments to learn aesthetics attributes of images has also been studied [17]. Computational exploration of attractiveness of facial images has also seen a lot of research interest. For example, Eisenthal et al. [18] build a system to predict face attractiveness from facial images.

Crowdsourcing

Crowdsourcing has recently become very popular in data driven visual computing, especially for gathering data for perceptual studies. A perceptual model of style similarity of 2D clip art is proposed by Garces et al. [4], where they crowdsource data in the form of triplets. Similarly, style similarity models of 3D shapes have also been build on crowdsourced data to learn a measure of style compatibility for furniture models [5]. Our crowdsourcing setup as described later is inspired by these works. The key difference is that we display multi-view images to reveal shape details for helping make more informed decision and also the way we do the quality control by enabling the response to next shape pair only when enough time has been spent on previous shape pair and a response has also been provided.

Deep Learning for Perceptual Attributes

Previous works have used several variations of deep learning for computer vision problems. A convolutional deep learning based double column architecture exploiting local and global pixel information to predict image aesthetics is described by Lu et al. [19]. In another similar approach, deep metric learning has been applied to compare between facial images [20]. In case of 3D shapes, recent works in graphics research have used a variety of encodings to input to machine learning frameworks. These encodings include shape descriptors, depth or rendered images, point clouds, triangle meshes, and volumetric grids such as voxels etc. As aesthetics is a perceptual concept, we choose to have view based representation of shapes. In this work, we use the concept of deep learning to attain representational learning paradigm i.e. to automatically learn the features useful to 3D shape aesthetics using human preference data.

Method

Our method involves first collecting human aesthetics judgments on shape pairs and then using the same to build a model of visual aesthetics. In this section, we first discuss our aesthetics judgments data collection process followed by formulation of our deep learning mechanism (Figure 3), which extends previous work in multi-view based deep convolutional neural networks.

Figure 1. Example image pairs posted on Amazon Mechanical Turk online crowdsourcing platform for collecting human aesthetics preferences. For each pair, workers click on one shape they think is more aesthetic than the other. The shapes are initially shown front facing and then made to slowly spin along vertical axis to reveal more shape details.

Crowdsourcing Aesthetics Preference Data

Our data collection process is inspired by recent research in perceptual exploration of graphics content [4], [5]. For example, Garces et al. [4] use Amazon Mechnical Turk to collect style similarity data of 2D clip art by means of showing triplets of clip art (A, B, C) and asking which of the last two (B and C) is more similar in style to the first (A). In a similar fashion, we show participants pairs of 3D shapes and ask them to choose one they think is more aesthetic. In this section we describe the structure of shape dataset, the processes of data collection and testing for consistency.

We use 3D models from the ShapeNet online shape repository [8], which organizes the available 3D shapes into different manually verified categories such as 'chairs', 'benches' and 'cars' etc. Furthermore, same class models are already rotated and scaled relative to each other. In this work, we randomly choose 778, 40, 75, 88, and 277 models from clubs chairs, tables, mugs, lamps and dining chairs categories respectively.

Data Collection Process. In our study, workers on Amazon Mechanical Turk platform are shown shapes in pairs (Figure 1). To which they respond by selecting one shape they think is more aesthetic. We choose to have 30 shape pairs in each experiment (called Human Intelligence Task (HIT) in Amazon Turk terminology). In our data collection process workers must select one shape over the other. They are paid \$0.10 for each completed HIT i.e. for providing aesthetics judgments to 30 shape pairs. We collected 8000, 2875, 825, 2500, and 5100 data samples (or responses) for clubs chairs, tables, mugs, lamps and dining chairs respectively. We consider a *"data sample"* as a pair of shapes (of the same class) and a human selection of the more aesthetic shape in the pair. We split the collected samples randomly into training ($\mathcal{I}_{train}$), validation ($\mathcal{I}_{validation}$) and test set ($\mathcal{I}_{test}$), and report prediction accuracy on the test set.

Quality Control. In an attempt to collect as good data as possible, we use various methods during the data collection process. First, we provide clear instructions to tell the potential participants that dishonest workers can be blocked from doing future work. We classify a worker as dishonest if he repeatedly answers control questions incorrectly. Second, we force participants to spend enough time to view the shapes and then provide their responses. To do this, after a user has provided a response, we have a 4 second delay before they can click on the next response. Third, we include control questions as in previous work [4] where one shape from the pair is intentionally made to look ugly (see Figure 6 for some examples). For each HIT, we have five control questions and the user must correctly answer all of them for us to accept the tasks in the HIT.

Participant Demographics. We collect gender, age group and geographic region related demographics data from the participants. For gender, we had 403 males and 360 females (no gender information provided by 12 workers). The HIT acceptance rates for males and females were 87.1% and 82.8% respectively. We divided the age into the following groups: 0-20, 21-30, 31-40, 41-50, 51-60, 60-100. The percentages of Turkers (or workers) in each group respectively were: 1.6%, 36.0%, 37.1%, 14.3%, 9.6%, 1.3%. The HIT acceptance rates for five geographic regions: Africa, Asia, Australia, Europe, North America, South America respectively were: 0% (no workers), 85.1%, 100%, 87.9%, 85.0%, 77.3%.

Data Consistency and Subjectivity. In this paper, we perform consistency analysis on two aspects of the data in a separate online study.

First we check if there are any differences between the data from paid vs volunteer participants. For example, Redi and Povao [21] find differences between volunteer and paid participants. In our case, we compare between data (responses to 25 shape pairs) from 15 unpaid participants recruited on Facebook and from 15 paid participants recruited on Amazon Mechanical Turk. We use Fishers test for comparison and found that at 0.05 significance level it does not reject the null hypothesis of non-random association between the data collected from two platforms. We use this result to motivate us to use paid crowdsourcing to collect human preferences on shape aesthetics as it is relatively easy to recruit low cost participants.

Second, we look at population level consistency of the data. We do this by performing Fishers test to compare between data provided by male and female participants. As in the first test, we found non-rejection of null hypothesis. This result shows that the collected data is consistent based on the population. In Figure 2, we show how often humans agree on the pairwise comparisons i.e. the number of pairs with the vote split of 50%, 60%, 70%, 80%, 90%, and 100%. We consider those pairs for which we receive responses from ten or more participants. Although, we can observe that there exist pairs with 50% vote split, however their number is very small.

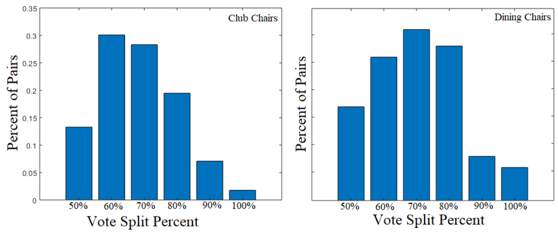

Figure 2. Plots of how often humans agree on pairwise comparisons for clubs chairs and dining chairs. We consider only those pairs for which we receive ten or more responses. Along y-axis we show percentage of pairs for which the vote was split 50%, 60%, 70%, 80%, 90%, and 100% (along x-axis).

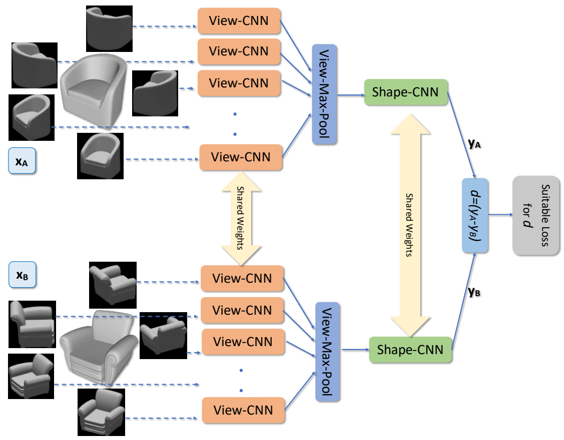

Figure 3. Illustration of our multi-view deep ranking architecture for shape aesthetics. We have two CNNs, View-CNN accepts rendered views as input and Shape-CNN takes as input the max-pooled output of View-CNN. Given rendered views of (x_A, x_B) shape pair with inequality constraints (i.e. x_A is rated higher than x_B), the output y_A should have a higher value than y_B. The difference between y_A and y_B is used to design the loss function as described in Equation 2.

Multi-view Deep Convolutional Ranking

Since we collect human aesthetics judgment data in pairs, we use pairwise technique to learn to rank and develop a measure of 3D shape aesthetics. Our multi-view based deep ranking formulation is inspired by previous work [6], [7], [20] and fits well with our collected data and problem. As shown in Figure 3, rendered views (12 views) of each shape in a pair (x_A, x_B) are first fed to a convolutional neural network (we call it View-CNN in Figure 3). The output of which for each view is then combined using an element wise max pooling operation (View-Max Pool in Figure 3). Then another convolutional neural network (called Shape-CNN as in Figure 3) takes the output of view-max pool to produce y_A and y_B.

In a nutshell, the learning proceeds as follows. Given a pair of shapes (x_A, x_B), where shape x_A is ranked higher than x_B (i.e. human finds first shape more attractive than the second shape), our network should output higher score y_A for shape x_A than the score y_B output for shape x_B. If this is not the case, we adjust the network parameters $\mathbf{W}$ and $\mathbf{b}$. Our cost function is designed to *"minimize the misclassification of pairwise difference"* in aesthetics scores (y_A and y_B) predicted by the network. Specifically, assuming our network models $h_{\mathbf{W},\mathbf{b}}(\mathbf{x})$, we aim to learn $\mathbf{W}$ (i.e. weights) and $\mathbf{b}$ (i.e. bias) such that:

$$h_{\mathbf{W},\mathbf{b}}(\mathbf{x_A}) > h_{\mathbf{W},\mathbf{b}}(\mathbf{x_B}) \qquad \forall(\mathbf{x}_A, \mathbf{x}_B) \in \mathcal{I}_{train} \tag{1}$$

Our training data $\mathcal{I}_{train}$ contains inequality constraints i.e. if $(\mathbf{x}_A, \mathbf{x}_B) \in \mathcal{I}_{train}$, where $\mathbf{x}_A$ is rated higher than $\mathbf{x}_B$ by human participants, our neural network should give a higher aesthetics score to shape $\mathbf{x}_A$ than to shape $\mathbf{x}_B$ (i.e. $h(\mathbf{x}_A)$ should be greater than $h(\mathbf{x}_B)$).

We learn $\mathbf{W}$ and $\mathbf{b}$ to minimize the following hinge loss function with a margin:

$$L(\mathbf{W}, \mathbf{b}) = \sum_{(\mathbf{x}_A, \mathbf{x}_B) \in \mathcal{I}_{train}} max(0, m - (y_A - y_B)) \tag{2}$$

where margin $m = 1$, $y_A = h_{\mathbf{W},\mathbf{b}}(\mathbf{x}_A)$ and $y_B = h_{\mathbf{W},\mathbf{b}}(\mathbf{x}_B)$. To minimize $L(\mathbf{W}, \mathbf{b})$, we perform an end-to-end backpropagation with batch gradient descent. Our algorithm takes the set $\mathcal{I}_{train}$ and learns a deep neural network that maps the rendered views of a 3D shape $\mathbf{x_A}$ to the shape's aesthetics score $y_A = h_{\mathbf{W},\mathbf{b}}(\mathbf{x_A})$ (Figure 3).

Network Implementation. We implemented our multi-view deep neural network (Figure 3) in Keras deep learning framework. View-CNN contains four convolution (i.e. Conv2D) layers each followed by MaxPooling2D layers. We also add BatchNormalization layers twice in the architecture. View-Max-Pool (Figure 3 is implemented as an element wise max operation (i.e. using Maximum layer in Keras). Shape-CNN is implemented using three convolution layers followed by two fully connected layers. We also use MaxPooling2D after convolution layers and Dropout after fully connected layers. Output of Shape-CNN for two shape pairs is combined using the Subtract layer in Keras. We train our neural network for 30-50 epochs with batch size set to 100. We prefer to learn a single function that maps multi-view images of different shape categories to their relative aesthetics scores. Specifically, although, a category specific network (i.e. a separate network) can be trained for each shape category or a class label can be added to the current network in addition to the predicted scores, we train a single network for all categories, which takes longer to train and optimize, however, builds an understanding of shape aesthetics from multiple shape categories, thus making it more general.

Results and Analysis

In this section we discuss the prediction performance of our network, qualitative patterns in visual aesthetics based shape rankings, specific shape features and the concept of "aesthetics duality".

Table 1. Prediction results for different categories of shapes (rows). We measure the prediction accuracy using three different criterias. "$\mathcal{I}_t$" is the percentage of correctly predicted pairs in $\mathcal{I}_{test}$. "*Sub* $\mathcal{I}_t$" is the percentage of correctly predicted subset of pairs in $\mathcal{I}_{test}$ that have a high difference (i.e. 80% or more) in the percentage of votes to first and second shapes. "*Rank Correlation* $\mathcal{R}_c$" is the correlation between ranking of test shapes by humans (i.e. averaged) and on the scores predicted by the network.

Shape Category	Accuracy		
	$\mathcal{I}_t$	*Sub* $\mathcal{I}_t$	$\mathcal{R}_c$
Club Chairs	69.3%	78.6%	0.76
Pedestal Tables	71.7%	81.3%	0.72
Mugs	68.0%	77.4%	0.79
Lamps	72.8%	80.0%	0.81
Dining Chairs	64.4%	76.3%	0.73

Performance Evaluation

Table 1 above shows the prediction results of our method. We use the test data-sets $\mathcal{I}_{test}$ as ground truth data to report accuracy in three ways. First ($\mathcal{I}_t$ column in the table), as the percentage of pairs (A, B) from $\mathcal{I}_{test}$ for which network produces higher score for shape A than for shape B i.e. it agrees with human aesthetics preference of A having higher rank than B. Second (*Sub* $\mathcal{I}_t$ column in the table), we repeat the first method above but with those shape pairs from $\mathcal{I}_{test}$ that have at lest 10 responses (i.e. each pair rated by at least 10 humans) and have a high (80% or more) difference between the percentages of votes to first and second shape. For example, the shape pair (A, B) when rated by 10 workers with 9 votes for A and 1 vote for B will be included in this evaluation. We prefer to use "*Sub* $\mathcal{I}_t$" for the evaluation of accuracy due to the fact that there are pairs where human labels can give uncertain answers (e.g. half choose A and half choose B). Thus, we argue that it is not useful to consider these to measure accuracy as the ground truth itself would be uncertain leading to lower accuracy. Third ($\mathcal{R}_c$ column in the table), we report accuracy as a rank correlation for a set of test shapes. Specifically, for a set of test shapes (Figure 6), we compute the rank correlation (R_C) between the ranking produced using the scores given by the network and the ranking produced by 10 human participants in a study. We found a strong positive correlation as shown in the table above.

We perform further analysis on the predicted aesthetics scores and human votes received by each shape in pairs. Specifically, for shapes in pairs: (1) humans either vote equally (50%-50%) i.e. no difference in percentage of votes, or (2) all votes go to only one shape (100%-0%) i.e. 100% difference in percentage of votes or (3) have different combinations of votes (e.g. 80%-20% et.). We see how these differences in percentage of votes above relate to the predicted aesthetics scores. To this end, for shape pairs with 10 or more votes, we plot (Figure 4) the difference in percentages in votes for each shape pair along the x-axis and the mean of difference in aesthetics scores along the y-axis. In the plot, looking along the x-axis, as the difference in percentage of votes increases, we also see an increase in the mean of difference in aesthetics scores. This implies that the shapes in pairs that are rated as having similar aesthetic value are predicted to have similar scores and thus have low mean value for difference in their scores. On the other hand, shape pairs that receive a clear majority (e.g. 100% participants choose one shape over the other), they show relatively high difference in their predicted aesthetics scores. This also explains why we have higher accuracy in the third column of the accuracy Table 1.

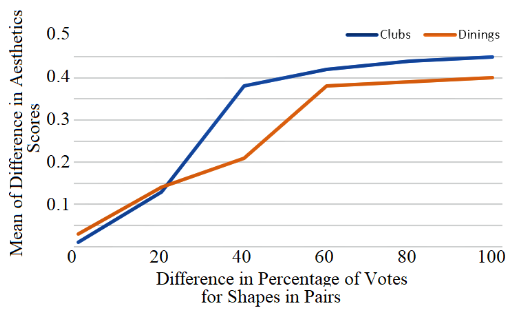

Figure 4. We consider shape pairs (A, B) rated by 10 or more participants with response (V_A, V_B), where V_A and V_B are the number of votes received by shape A and shape B respectively. Along x-axis we have the difference in percentage of votes received by each shape. For example, shape pairs for which $V_A = V_B$, the percentage votes will be same (i.e. a difference of 0). Along y-axis, we take the mean of difference (i.e. for pairs) in predicted aesthetics scores by our method.

Qualitative Patterns in Rankings

Figure 5 shows visual rankings for three categories of shapes: lamps, chairs, and sofas (or club chairs). In these results, we can easily observe certain elements of visual aesthetics, which we discuss below.

Smoothness of Surfaces. For the three categories of shapes in Figure 5, comparing the top and bottom shapes, we see that there is a sharp contrast in overall shape curvature. The highest ranked models tend to have more curved surfaces while the lowest ranked models show an opposite trend. Also at shape part level, we can observe this difference between top and bottom shapes. For example, for lamps, the *bottom parts* in high ranked models are more curved while in low ranked models they are more flat. As highlighted before, this aspect of aesthetics is learned automatically from the human aesthetics data i.e. we do not explicitly use features related to curvature.

Design and Surface Patterns. Aesthetic designs are believed to include interesting patterns. In our case, the top models for chairs category show some design patterns at the part level i.e in their backrests. Similarly, some top ranked sofas also exhibit surface patterns on their backs. These patterns provide a unique identity to these models which can't be perceived in models ranked at the bottom of the list. Our aesthetics model is able to capture this aspect of aesthetics from human aesthetics judgment data.

Visual Balance. Visual balance in design terminology refers to positioning and proportions of parts in relation to each other. In our rankings, some low ranked chairs exhibit imbalance in proportions of legs (i.e. short, very wide, or very thin etc.) compared to other parts and overall design. This is also true for low ranked sofas, for example, they have low backs and relatively high arm rests. Moreover, design symmetry is missing in one or more dimensions for both the chairs and sofas.

Test Data Rankings. We also use our network predict aesthetics scores of a set of test shapes and produce visual ranking of some top and bottom shapes in Figure 6. With this test set, we can see a similar pattern as described for the the rankings in Figure 5. Moreover, this test set also includes manually created ugly looking shapes, which are ranked at the bottom of list. We argue that the low scores provided by the network to these distorted shapes suggest that these can be considered ugly and are a good choice for designing the control questions with, as described in data collection process in the section above.

Analysis of Specific Features

In this section, we compute various quantitative shape features and study their relation with aesthetics scores. Our analysis shows that some of these features find a strong correlation with shape aesthetics. We look at quantitative features such as volume, area and bounding box dimensions and qualitative features related to curvature and structure. We also briefly touch upon the topic of "perceptual functional aesthetics" which relates to the concept of how well a design serves it's function.

Simple Shape Features. Can simple shape feature predict the aesthetics of a shape? To this end, we compute the bounding box volume, surface area, and intrinsic volume and then plot these against the sorted aesthetics scores of shapes (Fig-

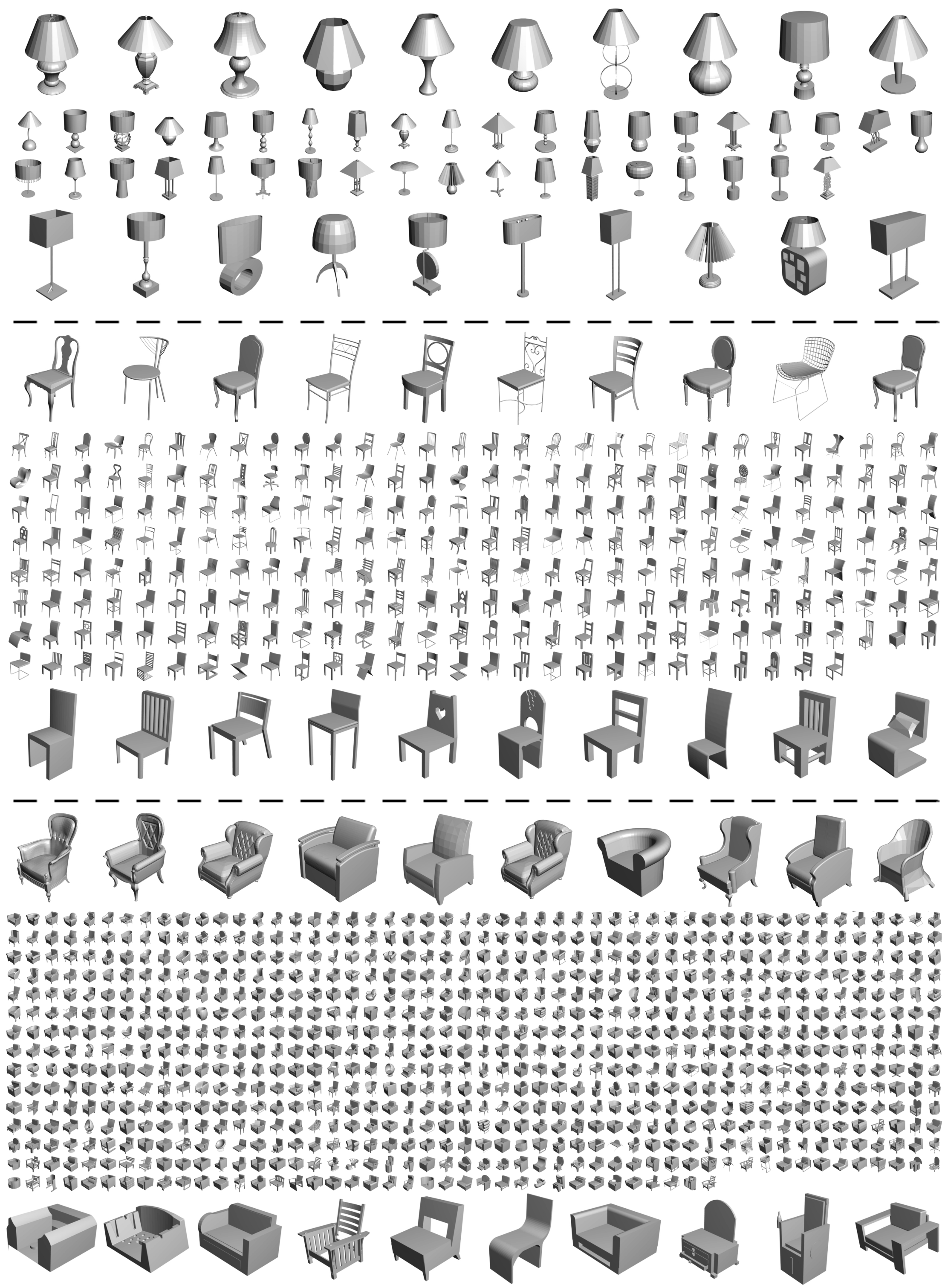

Figure 5. Shapes ranked (from top to bottom and left to right) according to our aesthetics measure (for each category, please zoom in to see details of shapes in the middle, only top and bottom ten shapes are shown in large size). There are 78 lamps, 267 dining chairs, and 778 club chairs. We show more results in the applications section.

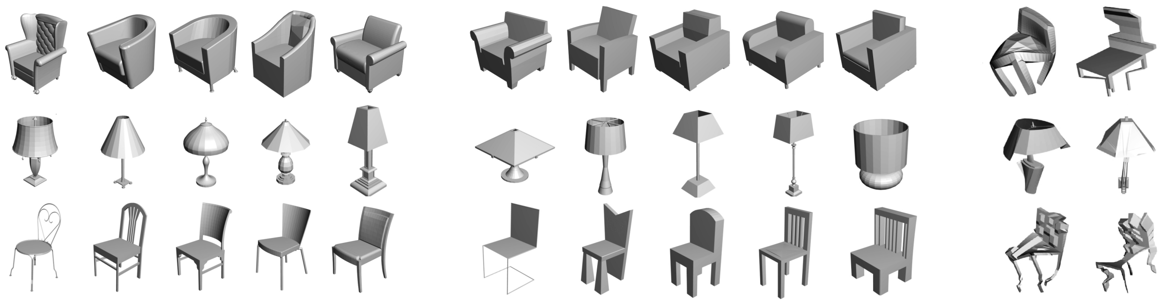

Figure 6. We rank the set of test shapes (i.e. not part of training shapes) and illustrate the aesthetics based ranking for top and bottom five shapes in this figure. The last two shapes (i.e. with least score) in each category are not part of the test set. They are intentionally distorted to look ugly. These ugly shapes are also employed during data collection process as part of control questions.

ure 7). The difference between bounding box and intrinsic volume is that the bounding box volume is the volume of smallest box that encloses the given shape and intrinsic volume is the inner volume occupied by the shape, which gives an estimate of how thick or thin a shape is perceived. As shown in the Figure 7, these features alone do not correlate well with the shape aesthetics.

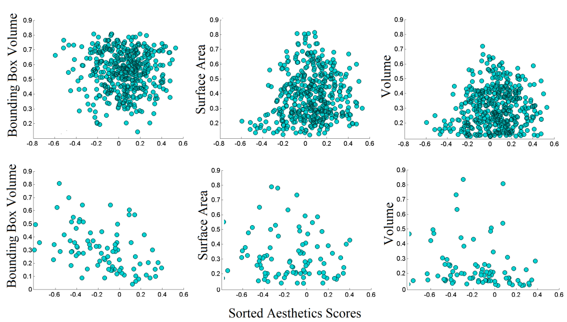

Figure 7. Left to right, plots of bounding box volume, surface area, and intrinsic volume against sorted aesthetics scores along x-axis. First row is for the club chairs and second row is for table lamps.

Curvature and Structural Symmetry. We manually divide the available dining chair models into two groups. The first group contains 52 models with planar surfaces and the second group contains 215 models with non-planar or curved surfaces. We compute the mean of aesthetics scores for each group, which is -0.0612 (with std=0.1612) and 0.0536 (with std=0.1511) for first and second groups respectively. On performing the two-sample t-test (assuming equal variance), we observe a significant effect (t=4.1023; $p<0.0001$), implying a difference in the population means between the two groups.

As described for curvature above, for table lamps, we manually form two groups of available shapes based on the rotational symmetry along y-axis (up-axis). The first group contains 68 models (symmetric group) and the second group contains 10 models (non-symmetric). The mean of aesthetics score for first and second groups are 0.2212 (with std=0.1918) and -0.0616 (with std=0.2113) respectively. We found that the population means between the two groups are different with t-test suggesting a significant effect (with values t=3.3972; $p<0.001$ and assuming equal variances).

Aesthetics Duality. The principle of "form follows function" signifies that the form of an object (e.g. a chair) should follow its intended function or use. Based on this principle, we argue that our aesthetic judgments about day-to-day objects may be influenced by both their form and functionality. In this section, we look at the perceptual aspect of the functionality of a 3D shape and it's potential influence on the perceived visual aesthetics. Specifically, we use the term "functional aesthetics" to denote the degree to which a shape or an object visually appears to serve its intended function. We hypothesize that while responding to shape aesthetics tasks (as described in Section 3), human participants not only consider form but also the perceptual functionality of a shape. To test this, we collect perceptual functional aesthetics response data by showing the same pairs of 3D shapes for dining chairs as used for collecting shape aesthetics data (Section 3). Specifically, this time the participants are asked to choose which shape they think is functionally more aesthetic or ergonomic to

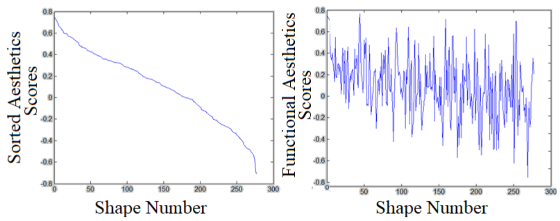

Figure 8. Plotting ”shape aesthetics” and ‘functional aesthetics”. On the left, we plot sorted (high-to-low) shape aesthetics scores (y-axis) for 277 models (x-axis), and on the right, we have “functional aesthetics” scores in the same order of shapes as in the first plot.

Figure 9. Comparing the two rankings produced using the paired data collected by asking “which shape is functionally more aesthetic or ergonomic” (i.e. to learn “functional aesthetics”) and “which shape is more attractive, aesthetic or beautiful” (i.e. to learn “visual aesthetics”).

use. With the collected data, we train a ranking neural network in the same fashion as done for shape aesthetics to predict “functional aesthetics scores”. These learned scores are used to rank the shapes based on their functional aesthetics.

On comparing the functional aesthetics ranking with the shape aesthetics ranking using Spearman’s rank correlation, we get a value of 0.1912. This signifies that there is a weak association between the two types of aesthetics rankings. To visualize this aspect, we also plot the two predicted scores in Figure 8. For comparison, the visual ranking for functional aesthetics for top 5 and bottom 5 models is shown in Figure 9. We can easily notice the qualitative differences in the shape features for both rankings. In the first row, we show the top five and bottom five functionally aesthetic shapes, and similarly in the second row, shapes are listed based on their visual aesthetics. We can notice that the first five shapes in the first row look more ergonomic than the bottom five shapes and there are differences in shape features between the two rows.

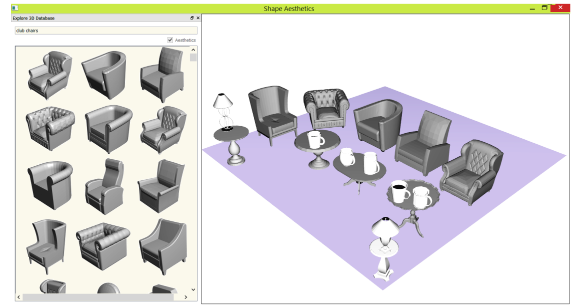

Figure 10. Aesthetics-based search and scene Composition. On the left panel, our tool can displays the available 3D shapes in random or ranked order (i.e ranking based on scores computed using our approach). From this panel, top shapes can be interactively selected to compose 3D scenes on the right panel.

Applications

We demonstrate our learned aesthetics measure by means of applications in search, ranking, visualization, and scene composition.

Search and Scene Composition

To demonstrate the use of learned aesthetics measure in interactive scene composition applications, we implement a tool (Figure 10) that allows ranking and interactive selection of shapes on the basis of their aesthetics. The user can first rank the shapes and then browse through the results to interactive select shapes to compose scenes with. This approach may allow them to compose more aesthetic scenes for different applications in computer graphics.

Prediction and Ranking

Our approach can be used with new shapes to predict their aesthetics scores. These scores can then be used to rank the shapes from high to low aesthetics. This application is particularly useful as recent online 3D shape repositories are expanding exponentially without having the means to help users find the kind of shapes they are looking for. In addition to the rankings shown in Figure 5, we present the rankings for remaining categories of shapes in Figure 11.

Visualization

Our approach can be used to create aesthetics based visualizations of large 3D shape datasets. One such visualization is presented in Figure 12.

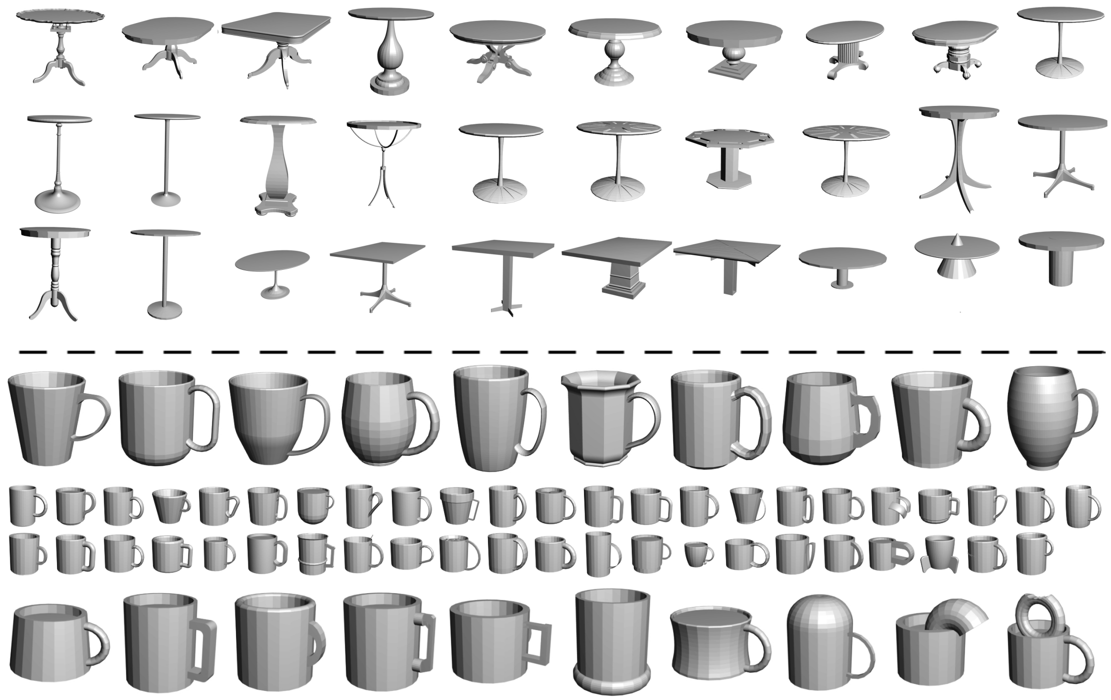

Figure 11. We presents ranking results for the remaining categories. Tables and mugs ranked (from top to bottom and left to right in each row) according to our aesthetics measure (please zoom in to see shape details).

In this visualization, each shape is displayed in a size proportional to it's aesthetics score. To compute the position of each shape on the two dimensional grid, we use t-SNE [22] algorithm. This approach allows for easy visualization of aesthetics related aspect of 3D shapes in a large dataset. If we zoom in (Figure 12), we can visually observe the clusters of shapes that have differences in shape features such as curvature and symmetry. The overall idea is to visualize a large dataset of 3D shapes in one image based on their aesthetics.

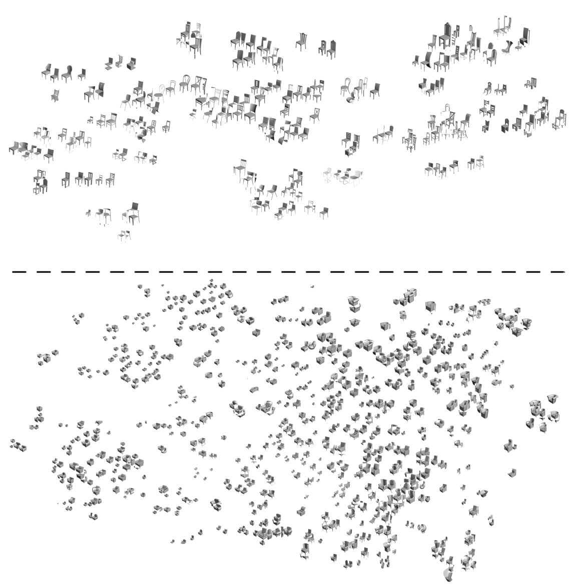

Figure 12. Aesthetics-based visualization (please zoom in for details) created by varying shape size based the aesthetics scores predicted by our approach.

Conclusion and Limitations

In this paper, we explore 3D shape aesthetics without using any pre-existing aesthetics features (e.g. curvature) or any handcrafted notions (e.g. rule of thirds). Our method leverages human aesthetics preference data in the form of forced pair comparisons to automatically learn a measure of 3D shape aesthetics. We demonstrate that a convolutional neural network taking multiple-views of 3D shapes as input is able to make reliable predictions of relative aesthetics scores, which are well aligned with human perception of aesthetics as shown by our results. Specifically, our predictive deep ranking formulation can be trained to come up with an overall aesthetics ranking from paired aesthetics ranking data. The aesthetics measure proposed in this paper is not a formal measure as such, it is learned via large human aesthetics preference data. We build tools to show how our method can be used in applications such as scene composition and search. As with any data-driven method, one limitation of our approach is that it only works with shape categories it is trained on i.e it can't make predictions about the unseen shape categories.

We see several exciting directions in which to extend and improve this work. First, in addition to the applications mentioned above, our approach can be used in an interactive modeling tool to predict the aesthetic value of a 3D shape under creation at different time intervals, however, a system that also suggests potential edits to guide more aesthetic shape creation would be more useful. Second, we prefer to pair shapes from the same categories (e.g. a chair is paired with a chair) for collecting aesthetics preference data, human perception of aesthetics may transcend shape categories. Thus, we speculate that learning a model of 3D shape aesthetics using human preferences on shapes from different categories would throw more light on the problem of 3D shape aesthetics. The first step in this direction would be to see how good humans are at comparing aesthetics of shapes of different categories. Third, our aesthetics measure takes into consideration only the shape or form of an object. However, aesthetics of an object can be influenced by other perceptual attributes such as the color, texture, and material. These features can also be included in the study of 3D shape aesthetics. Fourth, another promising direction to look into would be to study the shape aesthetics at part level. Finally, in this work, our focus was to model visual perceptual aesthetics of day-to-day use shapes by considering opinions of novices and non-experts, who could be the potential users of those shapes. For example, when we browse through furniture models online with an intention to purchase them, we dont take the opinion of someone trained on aesthetics, we just buy what looks nice or beautiful to us. The future works may take into consideration the opinions of artists who are trained on aesthetics.

ACKNOWLEDGMENT

The authors would like to thank the volunteers on Facebook for providing their responses to shape aesthetics study. They would also like to thank Microsoft Research for supporting this research.

REFERENCES

1. C. H. Séquin, "CAD Tools for Aesthetic Engineering," *Computer-Aided Design*, vol. 37, no. 7, pp. 737–750, Jun. 2005.
2. S. Bergen and B. J. Ross, "Aesthetic 3D Model Evolution," *International Conference on Evolutionary and Biologically Inspired Music, Sound, Art and Design*, pp. 11–22, 2012.
3. K. T. Miura and R. U. Gobithaasan, "Aesthetic Curves and Surfaces in Computer Aided Geometric Design," *International Journal of Automation Technology*, vol. 8, no. 3, pp. 304–316, 2014.
4. E. Garces, A. Agarwala, D. Gutierrez, and A. Hertzmann, "A Similarity Measure for Illustration Style," *ACM Trans. Graph.*, vol. 33, no. 4, pp. 93:1–93:9, Jul. 2014.
5. T. Liu, A. Hertzmann, W. Li, and T. Funkhouser, "Style Compatibility for 3D Furniture Models," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 85:1–85:9, Jul. 2015.
6. M. Lau, K. Dev, W. Shi, J. Dorsey, and H. Rushmeier, "Tactile Mesh Saliency," *ACM Trans. Graph.*, vol. 35, no. 3, Jul. 2016.
7. H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view Convolutional Neural Networks for 3D Shape Recognition," *ICCV*, 2015.
8. A. X. Chang, T. A. Funkhouser, L. J. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu, "ShapeNet: An Information-Rich 3D Model Repository," *CoRR*, 2015.
9. F. Giannini and M. Monti, "CAD Tools Based On Aesthetic Properties," *Eurographics Italian Chapter Conference*, 2002.
10. B. Pham and J. Zhang, "A fuzzy shape specification system to support design for aesthetics," *Soft Computing in Measurement and Information Acquisition*, pp. 39–50, 2003.
11. K. Fujita, T. Nakayama, and S. Akagi, "Integrated Product Design Methodology For Aesthetics, Functions And Geometry With Feature-Based Modeling And Constraint Management," *International Conference on Engineering Design*, pp. 1753–1756, 1999.
12. P. Locher and C. Nodine, "The perceptual value of symmetry," *Computers & mathematics with applications*, vol. 17, no. 4-6, pp. 475–484, 1989.

13. J. Friedenberg and M. Bertamini, “Aesthetic preference for polygon shape,” *Empirical Studies of the Arts*, vol. 33, no. 2, pp. 144–160, 2015.
14. S. Wannarumon, “An Aesthetics Driven Approach to Jewelry Design,” *Computer-Aided Design and Applications*, vol. 7, no. 4, pp. 489–503, 2010.
15. Y.-n. Lu and C.-h. Ho, “Impact of curvature of product shape on aesthetic judgments,” in *KEER2014. Proceedings of the 5th Kanesi Engineering and Emotion Research; International Conference; Linköping; Sweden; June 11-13*, no. 100. Linköping University Electronic Press, 2014, pp. 743–754.
16. R. Datta, D. Joshi, J. Li, and J. Z. Wang, “Studying Aesthetics in Photographic Images Using a Computational Approach,” *ECCV*, pp. 288–301, 2006.
17. L. Marchesotti, N. Murray, and F. Perronnin, “Discovering beautiful attributes for aesthetic image analysis,” *International journal of computer vision*, vol. 113, no. 3, pp. 246–266, 2015.
18. Y. Eisenthal, G. Dror, and E. Ruppin, “Facial Attractiveness: Beauty and the Machine,” *Neural Computation*, vol. 18, no. 1, pp. 119–142, Jan. 2006.
19. X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang, “Rating image aesthetics using deep learning,” *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp. 2021–2034, 2015.
20. J. Hu, J. Lu, and Y. P. Tan, “Discriminative Deep Metric Learning for Face Verification in the Wild,” *CVPR*, pp. 1875–1882, 2014.
21. J. Redi and I. Povoa, “Crowdsourcing for Rating Image Aesthetic Appeal: Better a Paid or a Volunteer Crowd?” *ACM Workshop on Crowdsourcing for Multimedia*, pp. 25–30, 2014.
22. L. van der Maaten and G. E. Hinton, “Visualizing High-Dimensional Data Using t-SNE,” *J. Machine Learning Research*, vol. 9, 2008.

Kapil Dev is a lecturer in the school of Mathematics, Computer Science and Engineering, Liverpool Hope University, Liverpool, UK. His research interests include nonphotorealistic rendering, computer graphics, human-computer interaction, and machine learning. Dev has a PhD in computer science from Lancaster University, UK. Contact him at kapil.saini@hotmail.com.

Manfred Lau is an assistant professor in the School of Creative Media, City University of Hong Kong, Hong Kong. His research interests are in computer graphics, HCI, and digital fabrication. He has a PhD in computer science from Carnegie Mellon University. Contact him at manfred.lau@gmail.com.