Guided Selfies using Models of Portrait Aesthetics

Qifan Li, Daniel Vogel
Cheriton School of Computer Science, University of Waterloo
{qifanli, dvogel}@uwaterloo.ca

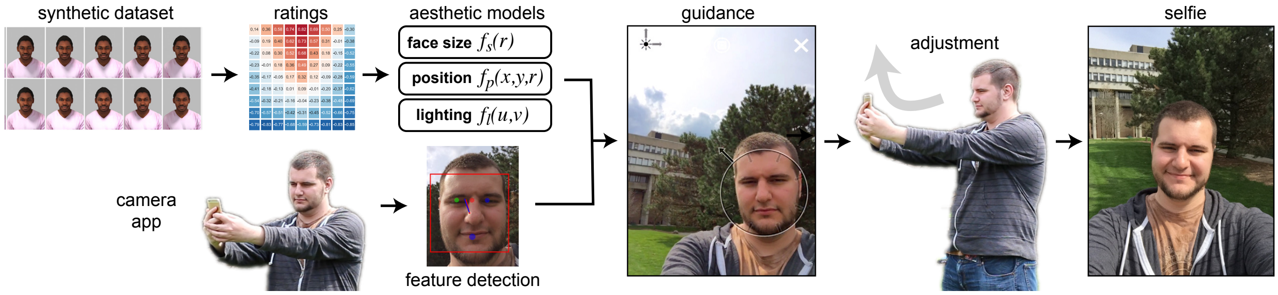


Figure 1: Methodology and techniques to enable interactive aesthetic guidance when taking a selfie on an unmodified smartphone.

ABSTRACT
We introduce techniques enabling interactive guidance for better self-portrait photos ("selfies") using a smartphone camera. Aesthetic quality is estimated using empirical models for three parameterized composition principles: face size, face position, and lighting direction. The models are built using 2,700 crowdworker assessments of highly-controlled synthetic selfies. These are generated by manipulating a virtual camera and lighting when rendering a realistic 3D model of a human to methodically explore the parameter space. A camera application uses the models to estimate the aesthetic quality of a live selfie preview based on parameters measured by computer vision. The photographer is guided towards a better selfie by directional hints overlaid on the live preview. A study shows the technique provides a 26% increase in aesthetic quality compared to a standard camera application.

ACM Classification Keywords
H.5.2 User Interfaces: Input devices and strategies

Author Keywords
mobile computing; computational photography;

INTRODUCTION
Smartphone self-portrait photographs, called "selfies", account for more than 30% of pictures taken by people aged 18 to 24 [15]. However, not everyone has photography skills, so these are often unsatisfying images. Visual aesthetics can be improved after a selfie is taken by editing and re-touching [8], but this requires extra effort, reduces realism, may degrade resolution, and aspects like face distortion and lighting direction are hard to correct. Using a "selfie stick" or a smartphone case with a built-in ring light [23, 24] can reduce distortion and improve lighting, but using accessories is not always convenient.

We introduce a methodology and techniques for interactive aesthetic guidance when taking a selfie on standard smartphone (Fig. 1). Instead of encoding rule-of-thumb photographic principles [26, 27, 33], we derive quantitative aesthetic models empirically. Rather than learning principles from a dataset of near-random images taken from the internet (e.g. [18, 10, 36, 9, 22, 38]), we generate highly-controlled synthetic self-portrait photographs using 3D rendering. This enables a methodical exploration and validation of underlying compositional features, in our case: face size, face position, and lighting direction.

Our dataset of synthetic selfies is used in 2,700 crowdworker assessments. These assessments are transformed into three quantitative models, each mapping a compositional feature configuration to an aesthetic score. We then describe a smartphone application to analyse the preview image with computer vision methods when taking a selfie. With this analysis, the configuration of each feature may be calculated and fed to our models to find the current aesthetic score and directions of improvement. Corresponding directional hints are then overlaid on the preview image, guiding the photographer to move the smartphone to improve one type of aesthetic quality. In a study, selfies taken with our system were rated 26% higher compared to those taken by a standard camera application. To make the problem tractable, our focus is on single-person selfies without a dominant background object. However, our system is designed so the photographer can selectively ignore some guidance to include other people or compensate for backgrounds like tourist landmarks. We make the following contributions:

- A systematic assessment of three features of selfie aesthetics: *face position*, *face size*, and *lighting direction*.
- Empirically-derived models to estimate the aesthetic score and *direction of improvement* for each feature.

DIS 2017, June 10-14, 2017, Edinburgh, United Kingdom.

DOI: http://dx.doi.org/10.1145/3064663.3064700

- A smartphone camera application to guide novice photographers to take better single person selfies using our models and a method to estimate dominant lighting.
- A controlled experiment validating the application's usability and capability to increase aesthetic quality.

BACKGROUND AND RELATED WORK

Although photography is an art form, rules-of-thumb have been proposed to make even casual photographs more aesthetically pleasing. Child [7] emphasizes that composition is important to attract and keep the viewer's attention and it complements the communication between the viewer and the photograph. One of the most common composition rules is the *rule-of-thirds* [31]. This is a 3×3 grid formed by dividing the image horizontally and vertically into nine equal portions. The grid line intersection points are called "power-points" which are commonly considered the best locations to place significant elements. For portrait photography, guides recommend how to use the rule-of-thirds to compose portrait styles such as "head-and-shoulders", "three-quarter", and "full-length" [17]. For example, placing the eyes in the top third for single person portraits [11]. A related compositional principle is the size of the subject [22, 25, 29]. In portraiture, some recommend filling the frame with the face [11], but face size is also dependent on the intended portrait style [17].

In addition to these spatial composition principles, Hurter [17] emphasizes the importance of lighting composition. Hurter argues light is the dominant factor to represent a three-dimensional reality in a two-dimensional space, "Just as a sculptor models clay to create the illusion of depth, so light models the shape of the face to give it depth and form." Different rules have been proposed for portraiture lighting. Some suggest a side-light so a pattern of light and shadow defines the face [7, 16], others recommend a frontal light to eliminate shadows and flatten the face [34].

Computational Aesthetics

Since such rules-of-thumb are imprecise and questionable, researchers have attempted to calculate the aesthetic quality of images using computational methods. One approach is to model aesthetic quality as a machine learning problem with an unconstrained number of features. A large corpus of images is rated by people to establish a ground truth for aesthetic quality. Many global and local features are extracted from the same images that were rated (typically image statistics related to colour, edges, contours, saliency, etc.). Then, machine learning algorithms are trained using the ratings and extracted features to automatically rate images for aesthetics. Most investigations use photographs acquired from an ad hoc online database [18, 10, 36, 9, 22, 38] since large amounts of data is needed for training. However, modelling using many features and with uncontrolled training images often spanning many classes (e.g landscapes, animals, groups, portraits, etc.) makes it difficult to find underlying aesthetic rules [25].

This general approach has also been applied to portrait photographs. Khan and Vogel [20] model aesthetic quality of individual portraits using a small set of features from photography tutorials, including face position and lighting. Males et al. [27] train on features they argue are used by professional photographers, including composition, face size, and features related to light, like contrast, lightness, and highlights. Redi et al. [33] use features inspired by photographic guides, including compositional rules. Chen et al. [6] develop a method to extract features of artistic lighting.

These works validates our focus on face position, face size, and lighting, but they still use highly variable training images from online databases. Although Redi et al.'s [33] results suggest portrait aesthetics are independent of age, gender, and ethnicity, other works suggest otherwise. Mazza et al. [29] find that dress and gender influence the perception of "head-shot" photographs and Xu et al. [37] found people rating photos for aesthetic quality often commented on non-compositional features like smile and mood. Also, Manovich et al.'s "Selfie City" [28] — a theoretic, artistic, and quantitative study of selfies — provides compelling evidence that age, gender, pose, and facial expression are linked to highly rated photographs.

This suggests that rating uncontrolled images to train models for compositional aesthetic quality is problematic. In addition to non-aesthetic confounds, precisely extracting high-level features like lighting direction is difficult for arbitrary images and many features change from image to image so understanding why a rater chose a certain image is difficult. Moreover, many of these works create features based on rules-of-thumb, but recent work suggests some rules, like the rule-of-thirds, may not improve aesthetic ratings [37, 2]. These factors all introduce errors and uncertainty into the aesthetic model.

Our highly controlled synthetic image corpus eliminates these potential confounds when rating aesthetics. In addition, our corpus isolates compositional features when rating, disentangling complex interactions resulting from comparative ratings when multiple features vary.

Realtime Aesthetic Guidance

The works above develop aesthetic models to evaluate existing images. Some explicitly use evaluated images as examples to teach amateur photographers methods to improve aesthetics [38], but they do not assist while taking a photograph. Our goal is to model aesthetics to create an interactive camera application with aesthetic guidance. This is complementary to guidance for low level features like exposure, luminance, and motion blur [30, 3].

Previous systems exist for realtime aesthetic guidance for photography and video. Ma et al. [26] develop an aesthetic model to suggest where to pose a person in a landscape photograph (visualized by a rectangle). However, the technique is not implemented on a smartphone or tested for interactive use. San pedro and Church [35] describe a smartphone camera application to guide photographers using 22 composition and 7 exposure features applicable to a wide class of images (proposed by Obrador et al. [31]). Feedback uses a musical composition mapped to feature scores, a small pilot study is inconclusive whether aesthetic quality is improved.

NudgeCam is a smartphone application to record product demonstration videos [4]. Text messages and a coloured box around the face indicate when there are problems with face size, face location, and scene brightness, tilt, and stability. A related system by Kumano et al. [21] uses icons to indicate when video motion is poor and to suggest when to zoom or pan the shot. These two systems encode existing videography rules-of-thumb, and neither is formally evaluated.

There are two direct precursors to our work. A commercial app called *Camera51* [1] "intelligently analyzes a scene and looks for lines, shapes, and people" [12] to provide guidance to a single, globally optimal composition. The underlying algorithms and aesthetic rules are unknown, and the user has no ability to selectively improve specific compositional principles while ignoring others.

Xu et al. [37] developed a realtime guidance system for portrait photographs. Face position and subject size are guided using a rule-based aesthetic model derived directly from the rule-of-thirds: the face should be on a grid line or power point, and the subject width should be one-third of the image. Their system requires a special laptop and external three-camera array. A study conducted outdoors evaluates whether the system improves aesthetics compared to a baseline system showing only a static rule-of-thirds grid. Unlike Xu et al., our system is specifically focused on self-portraits, it works on a standard smartphone, it considers lighting direction in addition to face position and size, our guidance is based on empirically-derived models derived from a synthetic dataset, and our final evaluation environment is highly-controlled.

DATASET AND AESTHETIC QUALITY RATINGS

Our goal is to help novice photographers take self-portraits that average people find more aesthetically pleasing. We accomplish this by rendering hundreds of synthetic portraits that systematically vary three key compositional features: face size, face position, and lighting direction. These portraits are used to gather aesthetic ratings in an online crowdsourcing experiment. We explain how these ratings are transformed into models to guide people when taking selfies in a later section.

3D Rendered Synthetic Selfie Dataset

All synthetic portraits are 3D renderings of realistic-looking 3D human models in front of a uniform 18% grey background. We use 3 female and 3 male models, with each gender pair having Caucasian, Asian, or Black ethnicity features. Models were selected to be consistent and "average" looking with neutral facial expressions, typical hair styles, no glasses, and similar poses. The arms of our 3D models are not poseable, so our synthetic selfies do not capture the subject though holding the smartphone. However, it is unlikely this small detail alters aesthetic ratings. By generating portraits across all six models, we eliminate confounds like background, gender, and smile when rated for aesthetic quality.

Renderings were generated using Blender (www.blender.com), an open-source 3D modelling program. Virtual camera properties were set to simulate the front camera of an iPhone 6 (focal length 45mm, field-of-view 54.2°, aspect ratio 3:4), and a parallel lighting source was configured to simulate the sun or dominant indoor light source. Ambient light was set to soften shadows and simulate typical lighting conditions outside or in brightly lit interiors. Using a Python API, the positions of the human model, virtual camera, and parallel light source were manipulated to explore each aesthetic feature as explained below.

Note that dataset generation and aesthetic quality ratings were interleaved as the three features were explored in sequence. First, portraits to explore face size were generated and assessed. These results established which face sizes to use when generating portraits to explore face position. After face position was assessed, the results were used to select face positions and sizes when exploring the lighting direction feature.

Face Size

Face size is an important feature because a face in a portrait should be large enough to be the centre of interest, but not so close that facial features distort [13]. We define *face size* as the ratio between the width of face — twice the distance between the eyes — and one-third of the image width. This ratio normalizes the feature and relates it to the rule-of-thirds guideline which suggests subjects should be sized to one cell in a rule-of-thirds grid. A face size of 1.0 matches this existing guideline.

To generate portraits exploring face size, we position the virtual camera and human model to render 19 images with face size ranging from 0.2 to 2.0 in steps of 0.1 (Fig. 2). The range of ratios is based on measuring real selfies taken as close as possible without face cropping and as far as possible using a standard "selfie stick." The face position and lighting direction are constant: face position is centred, and the light source shines straight onto the face.

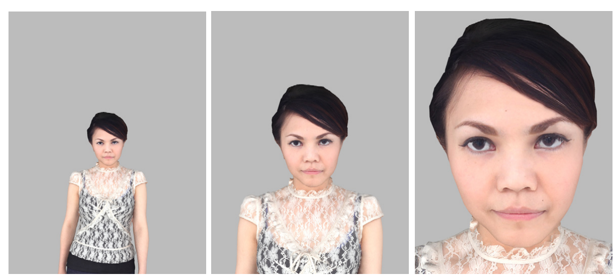


Figure 2: Example face size portraits: 0.4, 0.8, 2.0 using Asian female.

Face Position

The rule-of-thirds is commonly used to position subjects in photographs [7, 17] and previous research has employed it for aesthetic ratings [26, 9] and guidance [37]. We define *face position* as the 2D location of the centroid of the eyes and parametrize it using a subdivided 3×3 rule-of-thirds grid. Each grid cell is divided by 4 to construct a 12×12 grid with face position denoted as (x,y), where x and y are multiples of $\frac{1}{12}$ in terms of normalized image position. This way, face positions $(4,4)$, $(4,8)$, $(8,8)$, and $(8,4)$ correspond to rule-of-thirds "power points" assumed to be preferred positions [20].

Based on the distribution of face size aesthetic ratings (recall dataset generation and rating was interleaved), we explored face positions using 4 face sizes: 0.3, 0.5, 0.8, and 1.0. For each face size, 81 images were rendered with the virtual camera placed to create face positions ranging from (2,2) to (10,10) — a 9×9 subset of the 12×12 grid parametrization that avoids extremely cropped faces (Fig. 3). The face is kept oriented towards the camera by aligning the model's face direction (represented as the normal of a transparent plane) to the camera film plane. As in face size images, the light source shines straight onto the face.

Figure 3: Example face position portraits: (4,4), (6,6), (9,2) using Asian male with face size 0.8.

Lighting Direction

The dominant light source is assumed to be omnidirectional with parallel rays like the sun. We define *lighting direction* as the light source position in spherical coordinates azimuth and polar angle (θ, ϕ), with the origin at the centre of the head. For example, the light is directly in front of the face at $(0°, 90°)$, to the right side and half-way above at $(-90°, 45°)$, and directly above at $(0°, 0°)$.

To reduce the space to explore, we assume light should not come from below or behind the face. This restricts the range of θ to $[-120°, 120°]$ and ϕ to $[0°, 90°]$. With step sizes of 30° for θ and 11.25° for ϕ, each component has 9 settings creating 81 total combinations. We use aesthetic rating results from face size and face position to create 4 sets of lighting images: size 0.3, position (6,2); size 0.5, position (6,3); size 0.8, position (6,4); and size 1.0, position (6,4). The virtual camera position was fixed to maintain the face size and position as the light source partially orbited around the head position (Fig. 4).

Figure 4: Examples of lighting direction portraits: (-60°,67°), (0°,90°), (90°,90°) using Black female with face size 1.0, face position (6,4).

Aesthetic Rating

These sets of synthetic selfies are rated for aesthetic quality on Amazon Mechanical Turk (AMT). Workers picked a multiple best and worst images for each set of images portraying a single 3D human model for each compositional feature.

Participants

We recruited 2,700 AMT workers without any experience, age, or location criteria (though the majority live in the United States). Our objective is to get aesthetic ratings from "average people." Workers were paid between $0.10 and $0.30 per task (called a HIT on AMT). The full protocol was approved by a research ethics board.

Task and Implementation

The task is to view a set of synthetic portraits of a single model exploring a single compositional feature, and pick at least N good and N bad images. For example, pick 8 good and 8 bad images among 81 portraits of the Caucasian male model with different face positions. Note that comparing two images at a time would not be practical — there would be 3,240 comparisons in just one face size image set.

User interfaces for AMT tasks must be clear and efficient to encourage workers to complete the task correctly and honestly [32]. We iteratively developed our interface with these goals (Fig. 5). Thumbnails of images are arranged in a 9×9 grid (for face position and lighting direction), or a 1×19 row (for face size). Clicking on a thumbnail displays it as a full image and selects it (when the task begins, a random thumbnail is selected). Thumbnails can be selected and viewed quickly by dragging the mouse or using cursor keys.

Clicking on a rating button, or using a shortcut key, classifies the selected image and indicates the decision with a colour outline: green for 'good' (shortcut key '1'), blue for 'undecided' (key '2'), and red for 'bad' (key '3'). An undecided decision is automatically assigned when a thumbnail is selected for more than 2 seconds, but no good or bad rating provided. The worker can change their ratings at any time. The remaining number of good and bad ratings are shown in the rating buttons. Once the minimum number of ratings is completed, a submit button is enabled. The task was implemented as a web application using AngularJS.

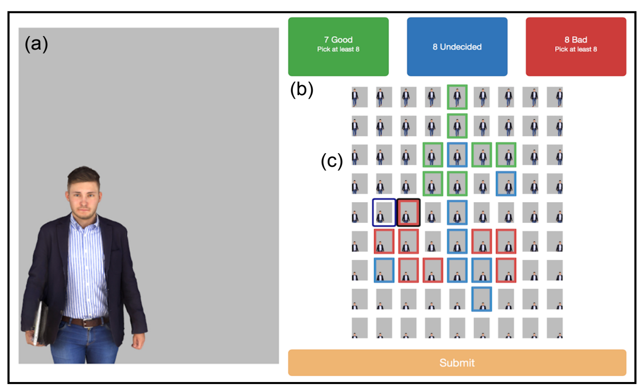


Figure 5: Aesthetic rating task interface (Caucasian male, rating face position): (a) large image view; (b) rating buttons; (c) thumbnails.

Design

Recall that we generated one set of images for face size, four sets of images for face position, and four sets of images for lighting direction, making nine sets in total. Each task rates images in one set for one human model: for face size, workers had to pick at least 3 good and 3 bad images; for face position and lighting direction, they had to pick at least 8 good and 8 bad images. With 9 sets and 6 human models there are

$9 * 6 = 54$ task variations. For each task variation, we recruited 50 workers, requiring 2,700 workers in total.

The procedure was a typical AMT HIT. When the HIT was accepted, the worker acknowledged a consent form and then viewed a video demonstrating the task interface. Instructions clarified that the goal is to "choose good and bad examples of images when considering them as selfies."

Results

To calculate the score of each image (representing a specific configuration for one of the principles being evaluated with a specific model), we summed all ratings using the following tally: +1 for each 'good', -1 for each 'bad', and 0 otherwise. Each image tally is then divided by the number of times that image was rated (regardless if 'good', 'bad', or 'undecided') to normalize scores to a range between -1 and 1.

To examine the consistency of scores and actual sample size for each image, we calculated the standard error of the mean (SEM) and percentage of tasks each image was rated. Rating scores were consistent, all SEMs are less than 0.05. Actual sample sizes were excellent for face size, with all images rated by at least 74% of workers, and good for face position and lighting direction with all images rated by at least 30% and 45% of workers respectively.

Face Size

Figure 6 illustrates the results. The preferred face sizes are 0.8 (score 0.33) and 0.5 (score 0.32). The score dips to 0.22 between those peaks and down to 0.13 for the smallest face size 0.2. However, all face sizes less than 1.4 have a positive rating. In contrast, scores for large face sizes fall sharply, decreasing to -0.6 for face size 2.0.

This suggests people prefer faces to be 50% of a rule-of-thirds-grid cell (16% of image width) or 80% to 90% of a rule-of-third grid cell (27% to 30% of image width). Note that faces very far from the camera, approximately less than 30% of a rule-of-thirds grid cell (10% of image width), are rated lower. Most pronounced are low ratings for faces very close to the camera, with negative scores when faces are larger than 130% of a rule-of-thirds grid cell (43% of the image width).

Face Position

Figure 6 illustrates the results. A mass of positive scores 0.7 or greater can be seen when the face is roughly horizontally centred. This positive mass moves vertically up the image as the face size becomes smaller. Positions near the bottom and sides are negative, most below -0.3, likely due to cropping.

Higher aesthetic ratings for a centred faces break from the accepted rule-of-thirds principle [9]. Rows and columns labelled 4 and 8 represent rule-of-thirds "power lines", and cells at positions $(4,4)$, $(8,4)$, $(4,8)$, and $(8,8)$ represent rule-of-thirds "power points." There is no indication of higher ratings for power-lines or power points. This adds empirical support to Xu et al.'s [37] informal observations.

Lighting Direction

The pattern of scores for the four sets of lighting direction scores appear very similar. Figure 6 illustrates the scores for the four sets and the aggregated results. A mass of positive scores can be seen when θ is between $-30°$ to $30°$ and ϕ is between $67.5°$ to $78.75°$, with a notable positive extension down to $\phi = 90°$ at $\theta = 0°$, and out to $\theta = \pm 60°$ at $\phi = 78.75°$.

Higher aesthetic ratings for lighting shining directly (or nearly directly) on the face reduces shadows and evenly lighting the entire face. This is exactly the result when using a smartphone cases with a built-in ring light [23, 24], but it contradicts some past work [20, 6] suggesting that lighting a face from one side is preferable.

Summary

These ratings suggest clear preferences for face size, face position, and lighting when taking selfies: faces should not be too big, faces should be centred and near the top without getting too close to the edge, and lighting should shine from the front. Although these may appear elementary, they have not been derived empirically before. Moreover, having matrices of scores will enable our guidance system to suggest local directions of improvement (i.e. "move your face right") which is more flexible for users than simply suggesting the globally preferred location (i.e. "move your face here").

AESTHETIC MODELS

In order to use the ratings for real-time guidance, we transform the matrices of scores into three models. Each model estimates an aesthetic score and direction of improvement for one compositional feature given the current state of compositional feature parameters. The main challenge is how to transform empirically known scores at discrete parameter settings into continuous functions that return a score for any measured face size, face position, and lighting direction.

The general form of each model is $(s, \boldsymbol{d}) = f(\omega_0, \omega_1, ..., \omega_n)$. It is a function f which accepts compositional feature parameters $\{\omega_0, \omega_1, ..., \omega_n\}$ to calculate a score s and a vector $\boldsymbol{d}$ describing the *direction* in compositional feature space that will improve the score.

The specific models are:

- Face Size: $(s_s, \boldsymbol{d_s}) = f_s(r)$

  The model uses an interpolated lookup into the 1×19 row matrix of face size scores. Given the detected face size ratio r, the score s_s is a linear interpolation between the two known scores at face sizes defining the interval where r lies. If r is outside the 1×19 row matrix, the s_s is extrapolated using the two closest known scores. The one-dimensional direction of improvement $\boldsymbol{d_s}$ is the direction towards the largest interval score.

- Face Position: $(s_p, \boldsymbol{d_p}) = f_p(x, y, r)$

  Recall there are four 8×8 matrices of face position scores for four face size ratios (0.3, 0.5, 0.8, 1.0). The model uses a two-step interpolated lookup into these four matrices. The detected face size ratio r is used to construct an 8×8 matrix by linearly interpolating corresponding elements of matrices forming the interval around r. If r is outside all matrices, elements are extrapolated. Then the score s_p is a bilinear interpolation (or extrapolation) of the elements

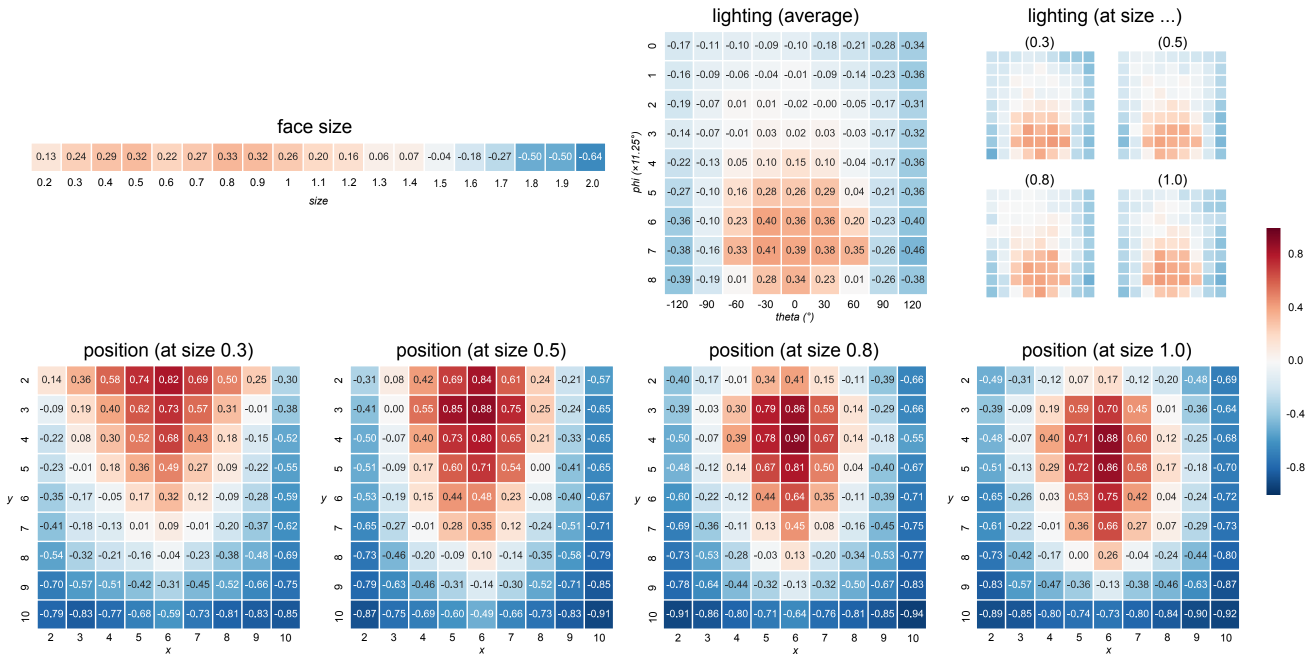


Figure 6: Visualizations of aesthetic ratings for size, position, and lighting. Note lighting is averaged across size given similar patterns for specific sizes.

surrounding the 2D interval where the detected position (x, y) resides in the interpolated matrix. The direction $\boldsymbol{d_p}$ is the 2D direction towards the surrounding element with the highest score.

- Lighting Direction: $(s_l, \boldsymbol{d_l}) = f_l(u, v)$

  Unlike face position or face size, we cannot directly measure 3D lighting directions Θ and Φ from a 2D image. Instead, the model uses (u, v), vector components to represent the direction and magnitude of the brightest patch of skin around the nose. We describe our algorithm to compute (u, v) later. These lighting direction components are transformed into the best estimate for Θ and Φ by finding the nearest neighbour to a set of canonical components (u^*, v^*) computed using the 3D human models and known Θ and Φ. With Θ and Φ, the corresponding score can be found in the single aggregated 8×8 lighting score matrix. The direction $\boldsymbol{d_l}$ is the 2D direction towards the surrounding matrix element with the highest score.

SMARTPHONE APPLICATION

We created a smartphone "app" that detects the current face size, face position, and lighting direction using computer vision techniques and passes these to the aesthetic models. The results are used to provide interactive guidance by suggesting how to move the smartphone to improve aesthetics. Our app runs on an iPhone 6 with iOS 9.3 using OpenCV 2.4.10 for computer vision.

Feature Detection and Calculation using Computer Vision

Each frame of the camera preview is captured and downsampled to 240×320 pixels. The native IOS `CIDetector` is used to find the bounding box of the face, mouth, and eyes. An OpenCV Haar classifier is used to detect the nose, but to improve speed and robustness, only the region bounded by the eyes and mouth is searched. Face and eye detection run at 12 FPS and nose detection at 8 FPS, acceptable for real time photo guidance. These detected features are used to compute model parameters as follows.

Face Size (r) and Face Position (x, y)

Recall face size and face position are expressed relative to a rule-of-thirds grid. The current face size (r) is the ratio of twice the distance between the detected eye positions to one-third of the width of the preview image. The current face position (x, y) is the centroid of the eyes expressed in twelfths of the preview image width and height. A low-pass filter [5] is applied to stabilize the face size and face position ratio across frames.

Lighting Direction (u, v)

To calculate the lighting direction, we examine the pattern of luminance on and around the nose. The protruding nose geometry produces local shadows and highlights that capture the global lighting information. Previous work used the pattern of light on the nose for 2D tracking [14].

The detected nose region-of-interest (ROI) is downsampled to 100×100 px and converted to HSV colour space. Eight 9×9 px patches are defined radially around the centre of the nose ROI, each with a corresponding patch direction vector $\boldsymbol{v_i}$ from nose ROI centre to patch centre (Figure 7a). The ratio of the median luminance of each patch l_i to the luminance at nose ROI centre l_c is used to scale each corresponding direction vector: $(l_i/l_c) \times \boldsymbol{v_i}$. For robustness, the steps above are repeated for patches at radii from 9 px to 45 px. The final lighting direction vector component (u, v) is the sum of the scaled vectors from all iterations. A low-pass filter [5] is also applied to stabilize (u, v) across frames.

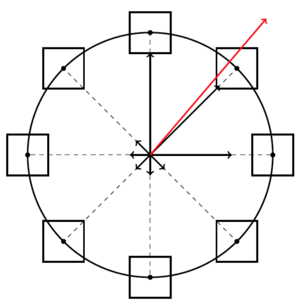


Figure 7: Lighting direction detection: (a) algorithm illustration; (b) test examples.

We informally tested this algorithm in two stages. First, we ran the algorithm on the synthetic images used for the lighting direction ratings. To prevent a confound from imperfect nose detection, the nose region was defined in Blender by rendering a rectangle. A comparison of the calculated lighting directions with known lighting directions for different values of Φ and Θ suggest the algorithm works well when the face is lit from the front ($\Theta \in [-90°, 90°]$), but becomes noisy when lit from behind ($\Theta \in [-120°, -90°] \cup [90°, 120°]$) or top ($\Phi = 0°$).

Second, we ran the algorithm on real photos (1 female and 2 males) taken at different orientations to the sun (examples in Figure 7b). The calculated lighting directions were consistent for all three sets with similar noisy measurements when lit from the top and behind. It also appeared to be robust to the presence of eyeglasses and facial hair.

In practice, we found it works well inside as well, long as there is a reasonably dominant light source. Diffuse lighting or very low light introduces some noise, but in those conditions light direction has no effect. The noisy estimates when lit from behind or above will likely guide people towards some better lighting position, albeit in a less consistent direction.

User Interface Design

The primary interface components are the guidance icons for the three compositional features. When a face is detected, a circle surrounding the face indicates tracking is working and guidance icons are overlaid for each feature.

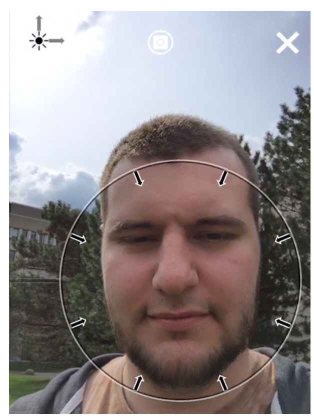
Face Size Guidance – The direction of small radial arrows indicates whether the face should be made larger (by moving the phone closer) or smaller (by moving the phone farther away) (e.g. Figure 8-a). The transparency of the arrows reflects the difference between the current face size score and the highest possible face size score. This means the arrows are very opaque when the score is poor, so face size adjustment appears more strongly suggested. If the arrows disappear, the score is near optimal.

Face Position Guidance – The direction of a large arrow emanating from the tracking circle indicates the direction to move the face. The direction is constrained to 8 cardinal compass directions for simplicity. Face movement is accomplished by slightly tilting the phone, for example, if the arrow points NW, then slightly tilting the phone SE moves the face up (N) and left (W). Like face size, arrow transparency reflects the relative difference between the current face position score and the best possible score.

Lighting Direction Guidance – Arrows are drawn around a sun icon located in the top-left corner. Horizontal arrows indicate the phone should be orbited around the head to adjust side-lighting (azimuth, Θ) and vertical arrows indicate an up or down tilt to adjust polar-angle (Φ), In practice, these motions are accomplished by fixing both the camera and the face and rotating or bending the body. For example, arrows pointing up and right indicate that the smartphone should be tilted up and orbited right to improve lighting (Figure 8-b). As before, the transparency of the arrows indicates how close the current lighting score is to the optimal score.

Once the face is positioned, the picture is taken by tapping a circular shutter button (or pressing either volume button for convenience). A debug mode, activated by double-tapping, is used for testing and demonstration. In this mode, the guidance visualization is augmented with tracked positions for the face, eyes, mouth, and nose, the estimated lighting direction vector, as well as the numeric scores for all three features.

Since guidance provides *directions of improvement* for each feature rather than a direct path to a single *global maximum*, the application indirectly handles selfies other than simple self-portraits. For example, when taking a self-portrait in front of a tourist landmark like a statue or building, the photographer can ignore the position guidance to keep that object in frame, but still optimize the size and lighting of the face.



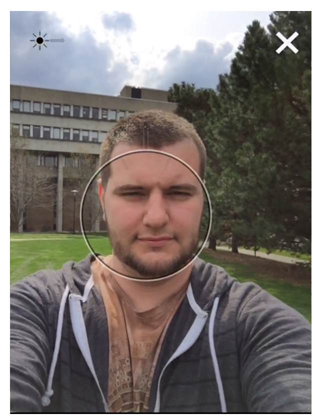


Figure 8: Guidance user interface: (left) inward arrows suggest making the face smaller and arrows around sun icon suggest rotating up and right to improve lighting; (right) transparent inward arrows indicate face size is optimal, lighting is near optimal with some right rotation possible, and position can be slightly improved indicated by nearly transparent upward arrow at the top of face circle.

EVALUATION

We evaluated our app from two perspectives: usability and effectiveness at improving selfie photograph aesthetics. In our study, participants took selfie photos without and with our app in an indoor controlled setting. Usability was evaluated by examining logged usage patterns and subjective ratings. Aesthetic effectiveness was evaluated by asking AMT workers to rate the best photos taken by each participant without and with our app.

Although Xu et al.'s [37] system was not designed for selfies, we considered using it as a baseline. Since it ran on a landscape-oriented "Ultrabook PC" with a 3-camera array, this would have required re-designing the visual guidance to work in portrait aspect ratio and creating a simplified version suitable for a smartphone with a single-camera. This would have significantly altered the goal and fidelity of their system, but more importantly reduced the comparison to one between their rule-of-thirds model and our data-driven model. Xu et al. already discuss how rule-of-thirds did not correlate with better photo aesthetics and our empirical results already provide further evidence of this. For these reasons, we used a standard camera app as our baseline.

Participants

We recruited 20 participants from a university campus (mean age 24.4, 11 female). Recruitment was limited to people who could view a smartphone screen without eye glasses (since the eye detection algorithm is less reliable with dark-rimmed eye wear). Our participants exhibited visible diversity in terms of skin pigments and facial features. Participants had a range of experience: 7 took selfies daily or weekly, 10 monthly or yearly, and 3 almost never. Only 1 participant had taken a course in photography.

Apparatus

Participants used the smartphone app described above, configured to run without and with visible guidance. Regardless whether guidance was shown, the app ran the full compositional feature analysis and computed the scores and directions of improvement. These were logged, even when invisible to the user. This ensured the refresh rate of the preview mode was the same regardless of guidance, and most importantly, provided quantitative data to test whether aesthetic ratings (as determined by our models) were improved using guidance. A *launch* button provided a timestamp that together with a shutter button or volume button timestamp, enabled calculation of picture-taking time.

Instead of conducting the study outdoors with natural sunlight, we built an indoor studio so that lighting and background are controlled and consistent (Fig. 9-a). A 2.7 m wide roll of grey backdrop paper was hung from the ceiling like a curtain to create a circular space 4m in diameter. A bright light simulating the sun was placed high at one end and 5 additional bulbs were spaced around the circle for ambient fill light. To keep participants at the centre of the circle, they sat on a swivel office chair throughout the study. Note this made lighting direction adjustments by swivelling somewhat more convenient than turning once body while standing.

Task

The task was to take ten selfie photos. First, five without guidance (BASELINE condition) and then, five *with guidance* from our app (GUIDED condition). The presentation order was fixed due to strong carry-over effects if GUIDANCE was performed first. After both parts were completed, the participant selected their best BASELINE photo and their best GUIDED photo separately.

Design and Protocol

At the beginning of the session, participants were asked to focus on three compositional principles, face size, face position, and lighting direction, but the optimum configuration for those principles was not conveyed. In each part, they were told to take the five best photos they can and that they would pick the best one afterwards. To avoid other factors like smile affecting their choice or aesthetic quality, we specifically asked participants to use the same neutral expression in all photos.

At the beginning of the second part using the GUIDED condition, participants were told they would now be using a camera application with guidance (they did not know this before). The meaning of the guidance icons were explained without any hint of what the best settings were for the three compositional principles. In fact, they were told they could follow or ignore the guidance to make the evaluation realistic — if they were told to always follow guidance, everyone would end with global optimum settings determined by the model.

Finally, A post-experiment questionnaire gathered subjective scores for GUIDED (only) regarding *Ease of Learning*, *Ease of Use*, *Accuracy of Guidance*, *Operation Speed*, and *Hand Fatigue*. All scores were on a continuous scale from 1 to 5, with 1 being worst and 5 being best.

This is primarily a within subjects design. Dependent measures for usability are photo-taking time and compositional feature scores for BASELINE and GUIDED, as well as subjective scores for GUIDED. The study took 30 mins on average. The dependent measure for aesthetic effectiveness is a score assigned by AMT workers to the best BASELINE and GUIDED photos.

Results

Given only two-level independent variables, and all dependent variables are continuous, interval, and ordinal, a t-test with .05 critical value is used for statistical tests.

Photo-Taking Time

The photo-taking time is the duration between pressing the "launch" button until the moment the photo is taken. The average time for taking photos with BASELINE was 20 s (SD 22), significantly lower than 33 s (SD 27) with GUIDED ($t(198) = -3.72$, $p < .001$). These long times may be an effect of the study setting. Regardless, the relative difference indicates following guidance has a time cost. Perhaps because more time is spent considering composition options before taking a photo. The average picture-taking time when participants took their single "best" photos supports this theory: the average time to take the best BASELINE photo was 28 s (SD 28) compared to 36 sec (SD 36) for GUIDED. A significant difference $t(38) = -0.821$, $p < .042$, but smaller in magnitude.

Subjective Ratings of User Experience

Since participants rated user experience qualities on a continuous scale from 1 to 5, we examine mean values (Table 1). All ratings are positive, ranging between 3.7 and 4.3. Participant comments indicate that *Ease of Use* and *Accuracy of Guidance* were negatively affected by flickering arrows for lighting guidance. The score for *Hand Fatigue* was affected by the need to hold the arm still to read and follow guidance.

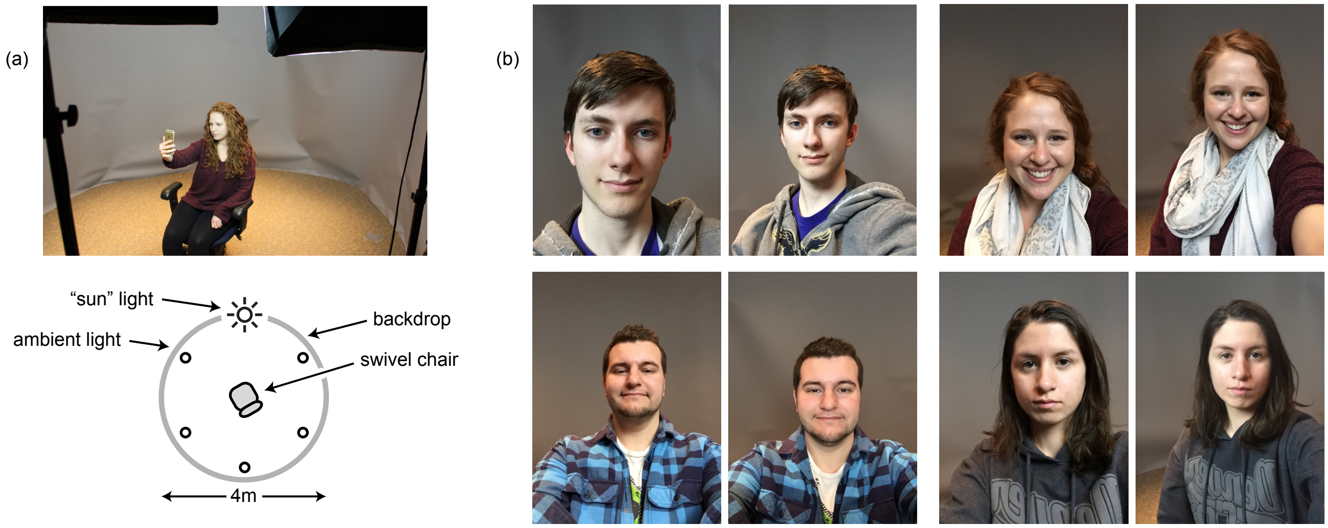


Figure 9: Evaluation: (a) controlled environment used for study; (b) examples of "best" photo pairs, left BASELINE, right GUIDED.

Ease of Learning	4.2 (±0.18)
Ease of Use	4.0 (±0.19)
Accuracy of Guidance	3.7 (±0.24)
Operation Speed	4.3 (±0.21)
Hand Fatigue	3.8 (±0.26)

Table 1: Mean user experience ratings (± SEM) for GUIDED

	BASELINE	GUIDED	t-test
Size	0.170 (±0.027)	0.245 (±0.023)	t(198)=-2.05, p<.05
Position	0.714 (±0.037)	0.723 (±0.022)	t(198)=-0.20, n.s.
Lighting	-0.019 (±0.012)	0.049 (±0.024)	t(198)=-2.46, p<.02

Table 2: Mean feature scores (± SEM) for BASELINE and GUIDED.

Note that although participants only rated GUIDED, there is an implied comparison to BASELINE, since GUIDED was completed second and BASELINE represents a real-world baseline familiar to participants.

Feature Scores

We compared mean feature scores for the best GUIDED and BASELINE photos selected by each participant. The scores for face size and lighting direction are significantly higher for GUIDED (values and t-tests in Table 2). This indicates that our app was successful in guiding participants to improve those two features according to our aesthetic models. This suggests that an increase in overall aesthetic quality, when the photos are rated in the following section, may be attributed to our app providing useful guidance. Similar face position feature scores is likely the result of people instinctively placing their face near the centre of the frame.

Aesthetic Effectiveness

To measure aesthetic effectiveness, we recruited 100 AMT workers to rate each pair of best photos taken by the participants. We did not specify any experience, age, or location criteria for the workers (though the majority live in the United States). Xu et al. [37] report ratings by AMT and expert photographers follow the same pattern, and our focus is on aesthetic attractiveness to the general population.

The task interface was modelled after the one used by Xu et al. A pair of photos is displayed with the best photos presented randomly on the left or right. The worker selected a rating between 0 and 100 for each photo using a slider. They were also instructed to consider the aesthetic quality of the face size, face position, and lighting direction, similar to the synthetic photo rankings described earlier. A text box was available for additional feedback.

We removed five participants from the set to be rated because they used very different expressions or poses in their pair of photos. For example, they were smiling in one and frowning in the other, or they turned their head to the side in one and looked straight at the camera in the other. The concern is that these pairs introduce a "pose and expression confound" that could alter a worker's aesthetic judgement. In all, 15 pairs of photos were rated by each worker, totalling 3,000 ratings.

Results

Table 3 summarizes mean ratings. Overall, selfie photographs taken with GUIDED have significantly higher ratings; $t(2998) = 17.37$, $p < .0001$. The average rating for GUIDED is 68.9 compared to 54.8 for BASELINE, an absolute difference of 14.1 — a 26% improvement in aesthetic quality. The low Standard Error of the Mean (SEM) and the relatively large absolute difference show this is a large and stable effect. The pair with the most improvement for GUIDED (the "highest pair") improved by 71% and the pair with the least improvement (the "lowest pair") decreased by -1%. The SEMs across all pairs is low, suggesting consistency among workers. The SEM of 2.4 for the "lowest pair" is among the highest across all pairs, suggesting this was the most difficult pair to judge.

In their additional feedback, workers often cited a central face position and a face lit with even, bright light as reasons for higher ratings. Some workers commented that photos appearing "washed out" were less attractive, and this was sometimes caused by frontal lighting encouraged by the guidance system. This could be remedied with more attention to the overall

	BASELINE	GUIDED	Improvement
All Pairs	54.8 (±0.6)	68.9 (±0.5)	26%
Highest Pair	45.4 (±1.6)	77.6 (±2.0)	71%
Lowest Pair	61.5 (±2.6)	61.1 (±2.4)	-1%

Table 3: Mean aesthetic ratings (± SEM) for BASELINE and GUIDED. "highest" is the pair with greatest mean improvement, "lowest" is the pair with least mean improvement.

lighting exposure of the face. Comments pertaining to face size were less consistent: some preferred a larger face, others a smaller one. This may be related to the two-peaked distribution for face size discovered by the synthetic selfie ratings. Overall, worker feedback supports our findings in the first rating experiment and provides evidence that workers were focusing on compositional principles for their assessment.

DISCUSSION

Our results show that our methodology to build highly-controlled aesthetic models, our computer vision algorithms, and our guidance interface, combine into a system that is effective at helping people take better selfies.

Limited Aesthetic Style

It is important to note that system encourages a symmetric and plain aesthetic style, no doubt due to the limited range of synthetic selfies used for aesthetic quality measurements. People also appreciate selfies with asymmetric compositions, dramatic lighting, or a tightly cropped face [19, 28]. Similarly, the neutral grey background may have affected how people rated the synthetic selfies. Generating and evaluating a wider range of synthetic selfies could diversify our system to guide people also to these more artistic types of selfies.

Implications for Human-Computer Interaction

Our aesthetic rating methodology relies on a fast, usable, and accurate interface for rating a large set of related images. We believe our solution for rating a synthetic dataset has applications in other fields such as rating options for visual design or ground truth dataset creation in machine learning.

Our system relies heavily on effective guidance visualizations packaged in a usable and responsive system. Using arrows for guidance was effective, but also makes the camera preview more cluttered and partially obscures the subject. An interesting future direction is to evaluate how subtle this style of feedback can be, while still being effective at communicating directions of improvement.

Learning about Aesthetic Preferences from the Models

When pondering what the models suggest, one may conclude that the recommendations are obvious: do not make the face too big or too small, put it near the top-centre, and make sure it is bright. Before passing judgement, recall that centering the face contradicts the commonly referenced rule-of-thirds and there are competing rules-of-thumb for lighting that suggest using side-lighting to create shadows. Moreover, stretching the arm to keep the size from being too large is not natural, but sometimes such awkward movements are necessary to produce a better quality photograph. After all, our goal was to discover, and empirically validate, what the optimum compositional rules are for self-portrait photographs as determined by "normal people."

Technical Limitations

Our system is restricted by computer vision capabilities. Face detection is constantly improving, but tracking is lost under extreme lighting or cropping. While our lighting estimation algorithm was successful at providing guidance, the lighting arrows sometimes flicker when the discretized model is at a threshold position. This creates oscillation between two different recommended directions. Adding anti-hysteresis methods and increasing the resolution of the lighting model (more scores at intermediate angles) would correct this small issue.

CONCLUSION

Our work contributes a systematic assessment of three basic features of selfie aesthetics using synthetic photographs and thousands of ratings. We transform these ratings to create interpolation-based models to estimate the aesthetic score and direction of improvement for each feature. With the help of a computer vision algorithm to estimate the dominant lighting direction in a person's face, we designed a smartphone camera application to guide novice photographers to take better selfies. A controlled experiment validated the usability of the application and its capability to increase aesthetic quality.

We see our three compositional features as an initial test of our methodology and guidance interface. Other compositional features like colour, texture, and balance could be included, as well as features like head tilt, facial expression, and background contrast. We think scaling our model to more features is possible. There is likely some inter-feature independence allowing some features to be evaluated independently. If not, the same "pipeline" approach that we used for position and lighting can be used: the results of one feature reduce the search space of the next. Note that although alternative machine-learning regression techniques can handle high dimensional data, they result in a black box preventing us from learning about aesthetics based on the final models.

Currently, if a photographer wishes to take a selfie in front of a background object (such as a tourist landmark) they have the option to ignore some of our system's guidance (such as position). An enhanced system could recognize salient background objects and alter the guidance accordingly. Models which consider the compositional relationship between the face and the object could be empirically generated using the same methodology used here. Synthetic selfies could be generated with a generic background object (perhaps a sphere) in different compositional positions relative to the person and the photograph frame. Finally, our methodology could be applied to photos of two or more people, specific classes of people (e.g babies, athletes), picture-taking settings (e.g. restaurant, beach), or entirely different subjects (e.g. cars, landscapes).

We hope our work demonstrates how a methodical and controlled study of aesthetics can lead to usable and useful applications which can increase aesthetic quality.

ACKNOWLEDGMENTS

This work was made possible by funding from NSERC Discovery Grant (#402467) and a Google Focused Award project on context-aware mobile computing.

REFERENCES

1. ArtInCam LTD. 2014. Camera51. (2014). https://www.camera51.com
2. Subhabrata Bhattacharya, Rahul Sukthankar, and Mubarak Shah. 2010. A Framework for Photo-quality Assessment and Enhancement Based on Visual Aesthetics. In *Proceedings of the International Conference on Multimedia (MM '10)*. ACM, New York, NY, USA, 271–280. DOI: http://dx.doi.org/10.1145/1873951.1873990
3. Stephen A. Brewster and Jody Johnston. 2008. Multimodal Interfaces for Camera Phones. In *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '08)*. ACM, New York, NY, USA, 387–390. DOI: http://dx.doi.org/10.1145/1409240.1409295
4. Scott Carter, John Adcock, John Doherty, and Stacy Branham. 2010. NudgeCam: Toward Targeted, Higher Quality Media Capture. In *Proceedings of the 18th ACM International Conference on Multimedia (MM '10)*. ACM, New York, NY, USA, 615–618. DOI: http://dx.doi.org/10.1145/1873951.1874034
5. Géry Casiez, Nicolas Roussel, and Daniel Vogel. 2012. 1€ filter: a simple speed-based low-pass filter for noisy input in interactive systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2527–2530.
6. Xiaowu Chen, Xin Jin, Hongyu Wu, and Qinping Zhao. 2015. Learning templates for artistic portrait lighting analysis. *IEEE Transactions on Image Processing* 24, 2 (2015), 608–618.
7. John Child. 2008. *Studio Photography: Essential Skills*. Focal Press.
8. Casio Computer Co. 2016. Life Style Digital Cameras. (2016). http://www.casio-intl.com/asia-mea/en/dc/lineup/ Hardware: Digital Cameras.
9. Ritendra Datta, Dhiraj Joshi, Jia Li, and JamesZ. Wang. 2006. Studying Aesthetics in Photographic Images Using a Computational Approach. In *Computer Vision – ECCV 2006*, Aleš Leonardis, Horst Bischof, and Axel Pinz (Eds.). Lecture Notes in Computer Science, Vol. 3953. Springer Berlin Heidelberg, 288–301. DOI: http://dx.doi.org/10.1007/11744078_23
10. S. Dhar, V. Ordonez, and T.L. Berg. 2011. High level describable attributes for predicting aesthetics and interestingness. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. 1657–1664. DOI: http://dx.doi.org/10.1109/CVPR.2011.5995467
11. Christina Dickson. 2009. 6 Tips for Perfect Composition in Portrait Photography. (Sept. 2009). http://digital-photography-school.com/6-tips-for-perfect-composition-in-portrait-photography/
12. Felix Esser. 2014. Camera51 Android app makes you a master photographer by guiding your framing. (Oct. 2014). http://www.digitaltrends.com/photography/camera51-android-app-makes-master-photographer-guiding-framing/
13. Ohad Fried, Eli Shechtman, Dan B. Goldman, and Adam Finkelstein. 2016. Perspective-aware Manipulation of Portrait Photos. *ACM Trans. Graph.* 35, 4, Article 128 (July 2016), 10 pages. DOI: http://dx.doi.org/10.1145/2897824.2925933
14. Dmitry O Gorodnichy and Gerhard Roth. 2004. Nouse 'use your nose as a mouse'perceptual vision technology for hands-free games and interfaces. *Image and Vision Computing* 22, 12 (2004), 931–942.
15. Melanie Hall. 2013. Family albums fade as the young put only themselves in picture. (June 2013). http://www.telegraph.co.uk/technology/
16. Darlene Hildebrandt. 2013. Lighting Ratios to Make or Break your Portrait. (2013). http://digital-photography-school.com/lighting-ratios-to-make-or-break-your-portrait/
17. Bill Hurter. 2007. *Portrait Photographer's Handbook* (3 ed.). Amherst Media.
18. Wei Jiang, A.C. Loui, and C.D. Cerosaletti. 2010. Automatic aesthetic value assessment in photographic images. In *Multimedia and Expo (ICME), 2010 IEEE International Conference on*. 920–925. DOI: http://dx.doi.org/10.1109/ICME.2010.5582588
19. Mahdi M Kalayeh, Misrak Seifu, Wesna LaLanne, and Mubarak Shah. 2015. How to take a good selfie?. In *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, 923–926.
20. Shehroz S. Khan and Daniel Vogel. 2012. Evaluating Visual Aesthetics in Photographic Portraiture. In *Proceedings of the Eighth Annual Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging (CAe '12)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 55–62. http://dl.acm.org/citation.cfm?id=2328888.2328898
21. Masahito Kumano, Kuniaki Uehara, and Yasuo Ariki. 2006. Online training-oriented video shooting navigation system based on real-time camerawork evaluation. In *Multimedia and Expo, 2006 IEEE International Conference on*. IEEE, 1281–1284.
22. Congcong Li and Tsuhan Chen. 2009. Aesthetic Visual Quality Assessment of Paintings. *Selected Topics in Signal Processing, IEEE Journal of* 3, 2 (April 2009), 236–252. DOI: http://dx.doi.org/10.1109/JSTSP.2009.2015077

23. LuMee LLC. 2016a. LuMee Phone Case. (2016). https://lumee.com/collections/all [Hardware: Phone Accessory].

24. Ty-Lite LLC. 2016b. Ty-Lite Phone Case. (2016). https://ty-lite.com [Hardware: Phone Accessory].

25. Wei Luo, Xiaogang Wang, and Xiaoou Tang. 2011. Content-based photo quality assessment. In *Computer Vision (ICCV), 2011 IEEE International Conference on*. 2206–2213. DOI: http://dx.doi.org/10.1109/ICCV.2011.6126498

26. Shuang Ma, Yangyu Fan, and Chang Wen Chen. 2014. Finding your spot: A photography suggestion system for placing human in the scene. In *2014 IEEE International Conference on Image Processing (ICIP)*. 556–560. DOI: http://dx.doi.org/10.1109/ICIP.2014.7025111

27. M. Males, A. Hedi, and M. Grgic. 2013. Aesthetic quality assessment of headshots. In *ELMAR, 2013 55th International Symposium*. 89–92.

28. Lev Manovich. 2014. SelfieCity. (2014). http://selfiecity.net

29. Filippo Mazza, Matthieu Perreira Da Silva, Patrick Le Callet, and IEJ Heynderickx. 2015. What do you think of my picture? Investigating factors of influence in profile images context perception. In *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 93940D.

30. Christopher McAdam, Craig Pinkerton, and Stephen A. Brewster. 2010. Novel Interfaces for Digital Cameras and Camera Phones. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '10)*. ACM, New York, NY, USA, 143–152. DOI: http://dx.doi.org/10.1145/1851600.1851625

31. P. Obrador, L. Schmidt-Hackenberg, and N. Oliver. 2010. The role of image composition in image aesthetics. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*. 3185–3188. DOI: http://dx.doi.org/10.1109/ICIP.2010.5654231

32. Bahareh Rahmanian and Joseph G Davis. 2014. User interface design for crowdsourcing systems. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*. ACM, 405–408.

33. Miriam Redi, Nikhil Rasiwasia, Gaurav Aggarwal, and Alejandro Jaimes. 2015. The Beauty of Capturing Faces: Rating the Quality of Digital Portraits. *CoRR* abs/1501.07304 (2015). http://arxiv.org/abs/1501.07304

34. Patrick Rice. 2005. *Professional Techniques for Black & White Digital Photography*. Amherst Media, Buffalo, NY.

35. Jose San pedro and Karen Church. 2013. The Sound of Light: Induced Synesthesia for Augmenting the Photography Experience. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. ACM, New York, NY, USA, 745–750. DOI: http://dx.doi.org/10.1145/2468356.2468489

36. Lai-Kuan Wong and Kok-Lim Low. 2009. Saliency-enhanced image aesthetics class prediction. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*. 997–1000. DOI: http://dx.doi.org/10.1109/ICIP.2009.5413825

37. Yan Xu, Joshua Ratcliff, James Scovell, Gheric Speiginer, and Ronald Azuma. 2015. Real-time Guidance Camera Interface to Enhance Photo Aesthetic Quality. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1183–1186. DOI: http://dx.doi.org/10.1145/2702123.2702418

38. Lei Yao, Poonam Suryanarayan, Mu Qiao, James Z Wang, and Jia Li. 2012. Oscar: On-site composition and aesthetics feedback through exemplars for photographers. *International Journal of Computer Vision* 96, 3 (2012), 353–383.