A perceptual metric for photo retouching

Eric Kee and Hany Farid[1]

Department of Computer Science, Dartmouth College, Hanover, NH 03755

In recent years, advertisers and magazine editors have been widely criticized for taking digital photo retouching to an extreme. Impossibly thin, tall, and wrinkle- and blemish-free models are routinely splashed onto billboards, advertisements, and magazine covers. The ubiquity of these unrealistic and highly idealized images has been linked to eating disorders and body image dissatisfaction in men, women, and children. In response, several countries have considered legislating the labeling of retouched photos. We describe a quantitative and perceptually meaningful metric of photo retouching. Photographs are rated on the degree to which they have been digitally altered by explicitly modeling and estimating geometric and photometric changes. This metric correlates well with perceptual judgments of photo retouching and can be used to objectively judge by how much a retouched photo has strayed from reality.

photo manipulation | photo analysis

Advertisers and fashion and fitness magazines have always been in the business of creating a fantasy of sorts for their readers. Magazine covers and advertisements routinely depict impossibly beautiful and flawless models with perfect physiques. These photos, however, are often the result of digital photo retouching. Shown in Fig. 1 are three recent examples of photo retouching in which the models were digitally altered*, in some cases almost beyond recognition.

Retouched photos are ubiquitous and have created an idealized and unrealistic representation of physical beauty. A significant literature has established a link between these images and men's and women's satisfaction with their physical appearance (1–8). Such concerns for public health has led the American Medical Association (AMA) to recently adopt a policy to "discourage the altering of photographs in a manner that could promote unrealistic expectations of appropriate body image."[†] Concern for public health and for the general issue of truth in advertising has also led the United Kingdom to consider legislation that would require digitally altered photos to be labeled.[‡] Perhaps not surprisingly, advertisers and publishers have resisted any such legislation.

A rating system that simply labels an image as digitally altered or not would have limited efficacy because it would not distinguish between common modifications such as cropping and color adjustment and modifications that dramatically alter a person's appearance. We propose that the interests of advertisers, publishers, and consumers may be protected by providing a perceptually meaningful rating of the amount by which a person's appearance has been digitally altered. When published alongside a photo, such a rating can inform consumers of how much a photo has strayed from reality, and can also inform photo editors of exaggerated and perhaps unintended alterations to a person's appearance.

Popular photo-editing software, such as Adobe Photoshop, allows photo editors to easily alter the appearance of a person. These alterations may affect the geometry of the subject and may include slimming of legs, hips, and arms, elongating the neck, improving posture, enlarging the eyes, or making faces more symmetric. Other photometric alterations affect skin tone and texture. These changes may include smoothing, sharpening, or other operations that remove or reduce wrinkles, cellulite, blemishes, freckles, and dark circles under the eyes. A combination of geometric and photometric manipulations allows photo retouchers to subtly or dramatically alter a person's appearance.

We have developed a metric that quantifies the perceptual impact of geometric and photometric modifications by modeling common photo retouching techniques. Geometric changes are modeled with a dense locally-linear, but globally smooth, motion field. Photometric changes are modeled with a locally-linear filter and a generic measure of local image similarity [SSIM (9)]. These model parameters are automatically estimated from the original and retouched photos as described in *Materials and Methods*. Shown in Fig. 2, from left to right, are an original and a retouched photo and a visualization of the measured geometric and photometric modifications.

The extent of photo manipulation is quantified with eight summary statistics extracted from these models. The amount of geometric modification is quantified with four statistics: the mean and standard deviation of the motion magnitude computed separately over the subject's face and body. The amount of photometric modification is quantified with four statistics. The first two statistics are the mean and standard deviation of the spatial extent of local smoothing or sharpening filters. The second two statistics are the mean and standard deviation of the similarity metric SSIM.

We show that these summary statistics combine to yield a metric that correlates well with perceptual ratings of photo alteration. This metric can be used to automatically rate the amount by which a photo was retouched.

Results

A diverse set of 468 original and retouched photos was collected from a variety of on-line sources. Human observers were asked to rank the amount of photo alteration on a scale of 1 (very similar) to 5 (very different). Given an original and retouched photo, we estimate the geometric and photometric modifications and extract eight summary statistics that embody the extent of photo retouching. Observer ratings were correlated against the summary statistics using nonlinear support vector regression (SVR). See *Materials and Methods* for complete details.

Shown in Fig. 3 is the correlation between the mean of 50 observer ratings per image and our metric. Each data point corresponds to one of 468 images rated on a scale of 1 to 5. The predicted rating for each image was determined by training an SVR on 467 images using a leave-one-out cross-validation methodology. The R-value is 0.80, the mean/median absolute prediction error is 0.30/0.24 with a standard deviation of 0.24 and a max absolute error of 1.19. The absolute prediction error is below
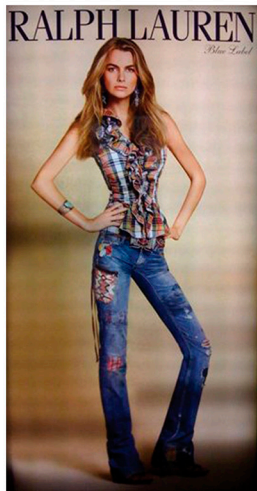
Fig. 1. A magazine cover featuring Faith Hill and the unretouched photo (top). An advertisement for an Olay beauty product featuring the model Twiggy and a contemporaneous photo of Twiggy (middle). This ad was banned in the United Kingdom by the Advertising Standards Authority because it considered the postproduction retouching of the original ad misleading to consumers. An ad featuring the model Filippa Hamilton and a contemporaneous ad of Hamilton (bottom). Images from: (Top) http://jezebel.com/278919/heres-our-winner-redbook-shatters-our-faith-in-well-not-publishing-but-maybe-god; (Middle) http://www.guardian.co.uk/media/2009/dec/16/twiggys-olay-ad-banned-airbrushing; (Bottom) http://www.mdig.com.br/index.php?itemid=8244. Sites accessed November 8, 2011.

0.5 for 81.4% of the images, and below 0.75 and 1.0 for 94.4% and 99.1% of the images, respectively.

Each observer rated 70 pairs of before/after images. The intraclass reliability is 0.97, showing that the mean observer rating is consistent.[§] Each observer rated a random set of five images three separate times, the presentations of which were uniformly distributed throughout the duration of the experiment. The mean/median within observer standard deviation is 0.34/0.31, showing that observers are relatively consistent in their individual ratings.

To determine which of our eight summary statistics were most critical for predicting observer ratings, we trained and tested 255 SVRs, one for each possible subset of size 1 to 8. The best performing SVR with one statistic consisted of the mean of the geometric facial distortion (statistic 1 as described in subsection *Perceptual Distortion*), which yielded an *R*-value of 0.58. The best performing SVR with two statistics consisted of the standard deviation of the geometric body distortion and the standard deviation of the photometric SSIM (statistics 4 and 6), which yielded an *R*-value of 0.69. And, the best performing SVR with three statistics consisted of adding the standard deviation of the geometric facial distortion to the previous SVR (statistics 4, 5, and 6), which yielded an *R*-value of 0.76. The best performing SVR of size 6 had an *R*-value of 0.80, equal to that of the full set of size 8. This subset of size 6 consisted of the statistics 1, 2, 4, 6, 7, and 8 as described in subsection *Perceptual Distortion*. Although six statistics are sufficiently powerful, they are extracted from each component of the geometric and photometric models. Therefore, there is little cost in using all eight statistics in terms of computational complexity or in terms of training the SVR.

The results presented above employed a nonlinear regression technique (SVR) to predict observer ratings. We also tested a linear SVR to validate the use of a nonlinear SVR over a simpler linear SVR. The *R*-value for the linear SVR is 0.72, as compared to 0.80 for the nonlinear SVR. The mean absolute prediction error is 0.34 with a standard deviation of 0.27 as compared to 0.30 and 0.24 for the nonlinear SVR. The max absolute error jumps from 1.19 to 1.93. Overall, the nonlinear SVR affords a considerably better prediction of observer ratings as compared to a linear SVR.

We also compared our metric against two standard image similarity metrics. A metric based only on the mean and standard deviation of a standard application of SSIM yields an *R*-value of 0.52 as compared to our approach that had an *R*-value of 0.80. A metric based on only the mean squared error between the before and after image performed much worse with a *R*-value of only 0.30. Standard image similarity metrics perform poorly because they do not compensate for, or measure, large-scale geometric distortions.

Shown in Fig. 4 are representative images with minimal (top) and maximal (bottom) prediction error. The over- and underestimations illustrate some of the limitations of our model. The perceptual distortion in the first two images (lower) is overestimated because there is a large photometric difference for the young boy (removal of blemishes) and a large geometric difference for the young woman (change in shape and position of the head), but neither of these differences correspond to a large perceptual difference in appearance. On the other hand, the perceptual distortion in the next three images is underestimated. The change to the symmetry of the young man's face, the addition of make-up to the woman, and the addition of teeth to the man are each relatively small from a photometric and geometric perspective but yield a large perceptual difference in appearance. Even with these limitations, we can reasonably measure perceptual distortion over a diverse range of photo alterations and content.

[§]The intraclass reliability (10) is computed as $\sigma_t^2/(\sigma_t^2 + \sigma_w^2/n)$, where the between-image variance is $n\sigma_t^2 + \sigma_w^2$, the within-image variance is σ_w^2, and n is the number of ratings per image.
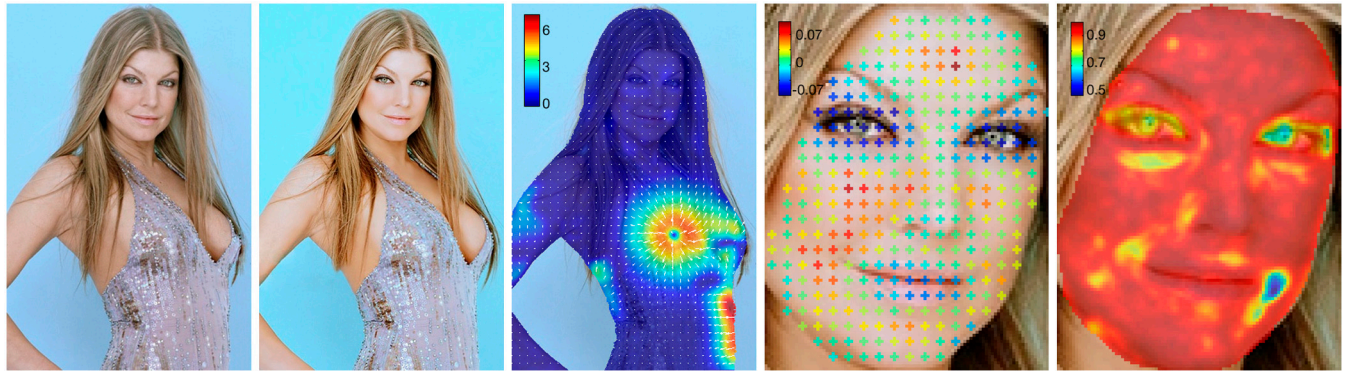
Fig. 2. Shown from left to right is an original and modified photo, the measured geometric distortions depicted as a vector field (the superimposed color-coding corresponds to vector magnitude), and the measured photometric distortions (linear filter in which a positive value denotes blurring and a negative value denotes sharpening, and image similarity in which smaller values denote larger image differences). First two images from: http://www.antesydespues .com.ar/en/20-famosas-sin-photoshop/, accessed October 26, 2011.

Discussion

Thanks to the magic of digital retouching, impossibly thin, tall, and wrinkle-free models routinely grace advertisements and magazine covers with the legitimate goal of selling a product to consumers.[¶] On the other hand, an overwhelming body of literature has established a link between idealized and unattainable images of physical beauty and serious health and body image issues for men, women, and children. Such concerns have led the AMA to discourage photographic alterations that promote unrealistic expectations of body image. It is our hope that a perceptually relevant metric of photo retouching can help find a balance between these competing interests.

We have developed a quantitative and perceptually meaningful metric to rate a photo on the amount of digital retouching. This metric correlates well with observer ratings of photo retouching. Providing a rating of photo retouching alongside a published photo can inform the public of the extent to which photos have strayed from reality (although it remains to be seen if this rating can mediate the adverse effects of being inundated with unrealistic body images). Such a rating may also provide incentive for publishers and models to reduce some of the more extreme forms of digital retouching that are common today. This measure can also help photo retouchers and editors because, even when an original and retouched photo are available, it can be difficult to see and quantify the extent of photo alterations [e.g., (11)].

The industry-wide deployment of a system to rate and label published photos will require buy-in and feedback from publishers, professional photo retouchers, and body-image and health experts. A large-scale rating system would have to quickly provide a rating to publishers so as to not interfere with publication schedules. The core computational component of our system is fully automatic, however a user currently annotates the hair/head, face, and body. When deploying an industry-wide rating system, this annotation could either be done automatically or with fairly minimal user assistance. As with any technology of this nature there is the inevitable cat and mouse game that will ensue, so it will be important to periodically review and refine the core technology to account for possible countermeasures and new photo-editing techniques that emerge. And finally, because no technology is perfect, one might provide publishers with the ability to appeal a rating.

[¶]While the judicious use of make-up and lighting can significantly alter the appearance of a model, subsequent digital retouching can create highly idealized and unobtainable body images that no amount of make-up or lighting can produce. We focus on this latter charade because we consider it to be more significant.

Materials and Methods

Geometric. The geometric transformation between local regions in the before and after images is modeled with a 6-parameter affine model. The luminance transformation is modeled with a 2-parameter model embodying brightness and contrast. This 8-parameter model is given by:

$$cf_a(x,y) + b = f_b(m_1 x + m_2 y + t_x, m_3 x + m_4 y + t_y), \quad [1]$$

where f_b and f_a are the local regions of the before and after images, c and b are the contrast and brightness terms, m_i are the terms of the 2×2 affine matrix, and t_x and t_y are the translation terms. The luminance terms on the left-hand side are incorporated only so that the geometric transformation can be estimated in the presence of luminance differences between the before and after images. A quadratic error function in these parameters is defined by approximating the right-hand side of Eq. 1 with a first-order truncated Taylor series expansion. This error function is then minimized using standard least-squares optimization. Because these geometric parameters are estimated locally throughout the image, the resulting global transformation can lead to unwanted discontinuities. A global penalty on large motions and a smoothness constraint are imposed by penalizing the local model parameters proportional to their magnitude and the magnitude of their local gradient. The addition of this smoothness constraint requires an iterative minimization which is boot-strapped with the result of the least-squares optimization. (See ref. 12 for complete details). This optimization is embedded within a coarse-to-fine differential architecture (13) in order to contend with
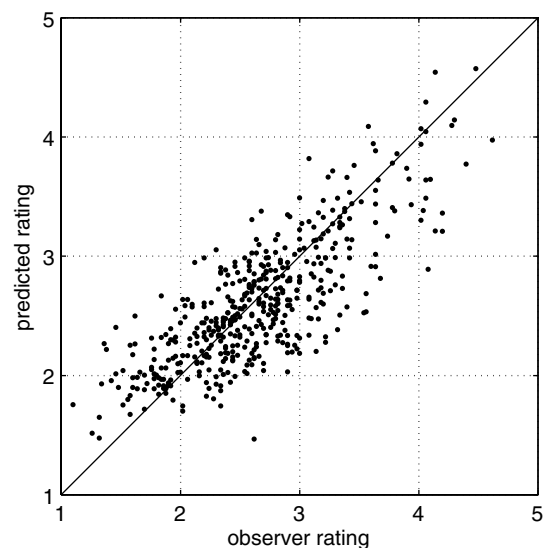

Fig. 3. A nonlinear SVR was used to correlate observer ratings of photo retouching to our geometric and photometric measurements of photo retouching. Shown here are the results from a leave-one-out cross-validation.
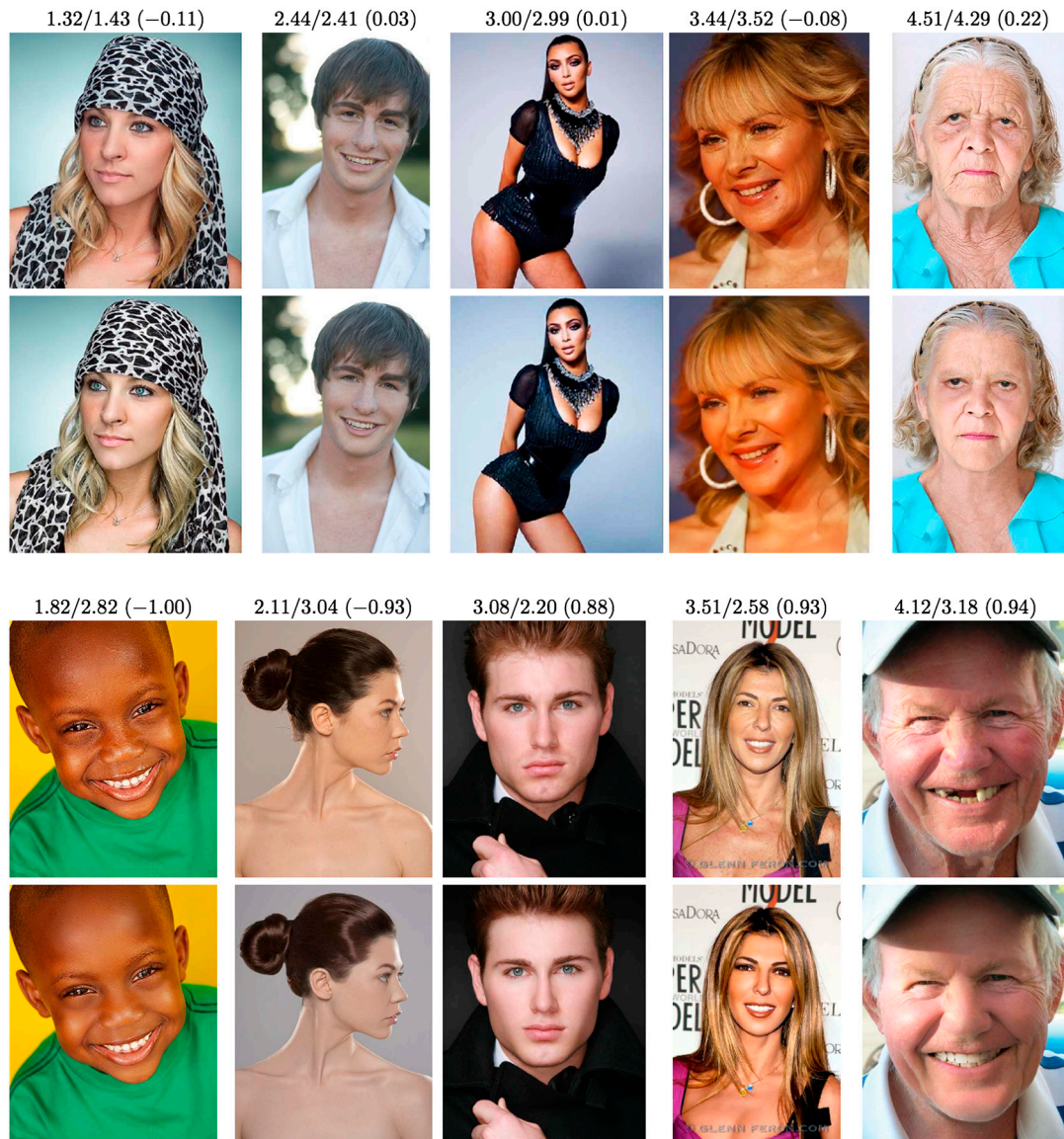
Fig. 4. Representative examples of minimal (top) and maximal (bottom) prediction errors. Shown are the before (upper) and after (lower) images and the numeric value specifies the observer rating/predicted rating and (their difference). Images from: (Top): (First) http://www.flickr.com/photos/reginapagles/5559285896/in/photostream; (Second) http://th05.deviantart.net/fs48/300W/i/2009/198/4/9/Beauty_retouch_9_by_hidden_silly.jpg; (Third) http://www.befter.net/user/ThalesRC/beft/kim-kardashian-before-and-after-photoshop/; (Fourth) http://www.antesydespues.com.ar/20-famosas-sin-photoshop/; (Fifth) http://www.flickr.com/photos/pauloarrivabene/2503732663/in/set-72157605134336657. (Bottom): (First) http://www.portraitprofessional.com/gallery/children/?p=0; (Second) http://jw-pixx.deviantart.com/gallery/?offset=48#/d2b4o49; (Third) http://hidden-silly.deviantart.com/art/Beauty-retouch-42-153708869; (Fourth) before image http://www.buddytv.com/articles/project-runway/profile/nina-garcia.aspx, after image http://www.glennferon.com/portfolio1/portfolio30.html#; (Fifth) before image http://www.stockvault.net/photo/104139/toothless-lew, after image http://www.photoretoucherpro.com/. Sites accessed October 26, 2011.

both large- and small-scale geometric changes. A model of missing data is also incorporated that contends with the case when portions of the after image have been entirely removed or added relative to the before image. (See ref. 14 for complete details). Once estimated, the geometric transformation is represented as a dense two-dimensional (2D) vector field:

$$\vec{v}(x,y) = \begin{pmatrix} m_1 x + m_2 y + t_x - x \\ m_3 x + m_4 y + t_y - y \end{pmatrix}. \qquad [2]$$

This estimation is performed only on the luminance channel of a color image. The before and after images are initially histogram equalized to minimize any overall differences in brightness and contrast. The background in each image is replaced with white noise in order to minimize any spurious geometric distortion. This geometric model embodies the basic manipulation afforded by the Photoshop liquify tool used by photo retouchers to alter the global or local shape of a person.

Photometric. Basic photometric modifications between local regions in the after image and the geometrically aligned before image are modeled with a 9×9 linear filter, h, given by:

$$f_a(x,y) = h(x,y) \star \tilde{f}_b(x,y), \qquad [3]$$

where $\star$ is the convolution operator, and $\tilde{f}_b$ is the geometrically aligned before image region, Eq. **1**. The filter h is estimated locally using a conjugate gradient descent optimization with a Tikhonov regularization. The regularization is used to enforce symmetry (i.e., zero-phase) on the estimated filter h. This estimation is performed only on the luminance channel of a color image.

Photometric modifications that are not captured by Eq. **3**, are measured with the similarity measure SSIM (9). This measure embodies contrast and structural modifications as follows:

$$C(x,y) = c(x,y)^\beta s(x,y)^\gamma, \qquad [4]$$

Kee and Farid

Fig. 5. Representative examples of photo editing [before (top row) and after (bottom row)]. Images from: (First) http://www.flickr.com/photos/manekineko/4526217217/; (Second) http://merengala.blogspot.com/2010/12/sin-photoshop-y-con-photoshop.html; (Third) before image http://www.flickr.com/photos/carreon/3509720855/in/set-72157603947306951, after image http://www.photoretoucherpro.com/; (Fourth) http://www.flickr.com/photos/manekineko/4645113101/; (Fifth) before image http://models.com/v-magazine/v-size-2.html, after image http://www.befter.net/user/Scarione/beft/fat-model-befter-edited-by-thescarione/. Sites accessed October 26, 2011.

where

$$c(x,y) = \frac{2\sigma_a\sigma_b + C_2}{\sigma_a^2 + \sigma_b^2 + C_2} \quad \text{and} \quad s(x,y) = \frac{\sigma_{ab} + C_3}{\sigma_a\sigma_b + C_3}, \quad [5]$$

and where μ_a, μ_b and σ_a, σ_b are the means and standard deviations of the image regions f_a and $\tilde{f}_b$, and σ_{ab} is the covariance of f_a and $\tilde{f}_b$. The various constants are $\beta = 1$, $\gamma = 1$, $C_2 = (0.03)^2$, and $C_3 = C_2/2$. Note that in this implementation of SSIM the brightness term is excluded because it did not impact observers' judgments. For the same reason, SSIM is computed only on the luminance channel of a color image. This photometric model embodies basic blurring, sharpening, and special effects afforded by various Photoshop filters.

Perceptual Distortion. The amount of photo distortion is quantified from eight summary statistics that are extracted from the geometric and photometric models described above and shown in Fig. 2. These statistics consist of four geometric and four photometric measurements: (1-2) the mean and standard deviation of the magnitude of the estimated vector field $\vec{v}(x,y)$, Eq. **2**, projected onto the gradient vector of the underlying luminance channel. This projection emphasizes geometric distortions that are orthogonal to image features, which are more perceptually salient. These two statistics are computed only over the face region, which quantify geometric facial distortion. (3-4) the mean and standard deviation of the magnitude of the estimated vector field $\vec{v}(x,y)$, Eq. **2**, projected onto the gradient vector and computed over the body region. These projected vectors are weighted based on specific body regions. The bust/waist/thigh regions are weighted by a factor of 2, the head/hair regions are weighted by a factor of 1/2, and the remaining body regions have unit weight (a full range of weights were explored and the final results are not critically dependent on these specific values). These two statistics quantify geometric body distortion, and are computed separately from the facial distortion because observers weight facial and body distortions differently. These four geometric statistics do not include global translation because the before and after images are initially aligned; (5-6) the mean and standard deviation of the SSIM, Eq. **4**, computed over the face region. These statistics quantify photometric modifications not captured by the linear filters; and (7-8) a measure D of the frequency response of the linear filters h, Eq. **3**:

$$D = \sum_{\omega} |\tilde{F}_b(\omega)|\omega - \sum_{\omega} |H(\omega)\tilde{F}_b(\omega)|\omega, \quad [6]$$

where $H(\omega)$ and $\tilde{F}_b(\omega)$ are unit-sum normalized one-dimensional (1D) frequency responses of the filter h and the local region $\tilde{f}_b$ which are computed

by integrating their 2D Fourier transforms across orientation. The parameter D is positive when h is a blurring filter, negative when h is a sharpening filter, and is tailored to our analysis of people in which filtering is commonly used to remove or enhance facial features. The mean and standard deviation of D, computed over the face region, are the final two statistics.

In summary, there are a total of eight summary statistics. The first four geometric statistics are the mean and standard deviation of the estimated vector field computed separately over the face and body. The second four photometric statistics are the mean and standard deviation of SSIM and the frequency response of the linear filters.

Before/After. A collection of 468 before/after images were collected from a variety of on-line resources, primarily the websites of photo retouchers showcasing their services. These images spanned the range from minor to radical amounts of retouching. Shown in Fig. 5, from left to right, are representative examples with increasing amounts of photo retouching.

Perceptual Ratings. A group of 390 observers was recruited through Amazon's Mechanical Turk. This crowd sourcing utility has become popular among social scientists as a way to quickly collect large amounts of data from human observers around the world (15). Observers were initially shown a representative set of 20 before/after images in order to help them gauge the range of distortions they could expect to see. Observers were then shown 70 pairs of before/after images and asked to rate how different the person looked between the images on a scale of 1 to 5. A score of 1 means "very similar" and a score of 5 means "very different." This yielded a total of 50 ratings per each of 468 images. The presentation of images was self-timed and observers could manually toggle between the before and after images as many times as they chose (observers are better able to see the modification when toggling rather than viewing side-by-side). In order to measure the consistency of observer responses each observer rated a random set of five images three times each. The presentation of these images was evenly distributed throughout the trial. Each observer was payed $3 for their participation and a typical session lasted 30 min. Given the uncontrolled nature of the data collection, some data filtering was necessary. Approximately 9.5% of observers were excluded because they frequently toggled only once between the before and after image and they responded with high variance on the repeated trials. (Dataset S1)

Support Vector Regression. Support vector regression (16) was used to estimate a mapping between user ratings and eight summary statistics extracted from the geometric and photometric models of photo retouching (each statistic was individually scaled into the range [−1,1]). Specifically, a nu-SVR with a Gaussian radial basis kernel was employed (17). A leave-

one-out cross-validation was performed in which the SVR was trained on 467 of 468 image ratings and tested on the remaining image. This training and testing was repeated 468 times in which each image was individually tested. The SVR has two primary degrees of freedom: (i) the scalar γ specifies the spatial extent of the kernel function; and (ii) the scalar c specifies the penalty applied to deviations of each data point from the regression func-

tion. These parameters were selected by performing a dense 2D grid search to maximize the correlation coefficient of each training set.

1. Myers PN, Jr, Biocca FA (1992) The elastic body image: The effect of television advertising and programming on body image distortions in young women. *J Comm* 42:108–133.
2. Lavine H, Sweeney D, Wagner SH (1999) Depicting women as sex objects in television advertising: effects on body dissatisfaction. *Pers Soc Psychol B* 25:1049–1058.
3. Agliata D, Tantleff-Dunn S (2004) The impact of media exposure on males' body image. *J Soc Clin Psychol* 23:7–22.
4. Wykes M, Gunter B (2005) *The Media & Body Image* (Sage Publications, London, England).
5. Derenne JL, Beresin EV (2006) Body image, media, and eating disorders. *Acad Psychiatr* 30:257–261.
6. Grogan S (2007) *Body Image: Understanding Body Dissatisfaction in Men, Women and Children* (Routledge, New York, NY).
7. Dittmar H (2009) How do 'body perfect' ideals in the media have a negative impact on body image and behaviors? Factors and processes related to self and identity. *J Soc Clin Psychol* 28:1–8.
8. Smeesters D, Mussweiler T, Mandel N (2011) The effects of thin and heavy media images on overweight and underweight consumers: social comparison processes and behavioral implications. *J Consum Res* 36:930–949.
9. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. *IEEE T Image Process* 13:600–612.
10. Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86:420–428.
11. Simons DJ, Rensink RA (2005) Change blindness: past, present, and future. *Trends Cogn Sci* 9:16–20.
12. Periaswamy S, Farid H (2003) Elastic registration in the presence of intensity variations. *IEEE T Med Imaging* 22:865–874.
13. Simoncelli EP (1999) *Handbook of Computer Vision and Applications* (Academic Press, San Diego, CA), pp 397–420.
14. Periaswamy S, Farid H (2006) Medical image registration with partial data. *Med Image Anal* 10:452–464.
15. Paolacci G, Chandler J, Ipeirotis PG (2010) Running experiments on Amazon mechanical turk. *Judgm Decis Mak* 5:411–419.
16. Vapnik V (1998) *Statistical Learning Theory* (John Wiley and Sons, New York, NY).
17. Chang CC, Lin CJ (2001) *LIBSVM: a library for support vector machines*, Software available at http://www.csie.ntu.edu.tw/cjlin/libsvm.