Supplementary Material: Image Visual Realism: From Human Perception to Machine Computation

Shaojing Fan, Tian-Tsong Ng, Bryan L. Koenig, Jonathan S. Herberg, Ming Jiang, *Member, IEEE,* Zhiqi Shen, Qi Zhao, *Member, IEEE*

1 Image collection

Image type & image source: Our dataset included photos taken by both amateurs and experts. Some of the amateur photos were from the authors' personal collections, some were from the INRIA Holidays dataset [1]. Professional photos were downloaded from a graphic design website (www.nipic.com), and from collections of three professional photographers. We also included some photographic images from the Columbia dataset [2].

Most CG images in our database were collected from several popular computer graphics forums (www.forums.cgsociety.org, www.forums.3dtotal.com). Some were from the homepages of renowned CG artists. We also made screenshots from computer games and CG movies. The CG images varied in type, including 3D (model-based rendered) and 2D (hand painted). They were rendered using various tools, such as Maya, 3DsMax, Zbrush, Photoshop.

For a quantitative and comprehensive realism study, we also included matte painting (MP) images in our database. Digital matte painting is a developing technology in the computer graphics field inspired by the film industry. An MP image is composed of a base plate, which can be a photograph or moving footage, with CG images or animations superimposed on top of it [3]. Some of our MP images were downloaded from the most popular matte painting website (www.mattepainting.org). Others were from screenshots of various movies. We selected MP images that were CG for more than 1/3 of their area.

We did not include obviously CG images like cartoons. Furthermore, we excluded those with obviously visible artifactual defects. We also excluded images with unrealistic scenes, like spaceships flying in the city, or scenes of exotic spaces. All images were scaled and cropped about their centers to be 256×256 pixels.

Image content: We categorized each image into scenes based on SUN scene categories [4]. Even within the same type of scene, images varied in levels of lighting and color conditions. For example, some images of people were studio portraits, others were snapshots in natural surroundings. To ensure a semantic balance in our dataset, the number of images assigned to each scene category matched for photos and CG images.

2 Pilot study for psychophysics study I

Prior to carrying out large-scale experiment, we performed a pilot study to determine how many participants we needed to reliably estimate human judgments of visual realism. 166 workers from Mechanical Turk ($>$95% approval rate and $<$15% abandon rate in Amazon's system) made their judgements on 60 images of different realism levels. Thus each image was scored by 166 people. We calculated a *realism score* (ranges from 0 to 1) for each image as the proportion equal to the number of judgments indicating that the image is a photo over the total number of judgments for that image.

We next used bootstrapping to evaluate how reliable the judgments were for various numbers of participants. For multiple group sizes, we randomly split the participants into two equal-sized groups and calculated the Spearman's rank-order correlation (ρ) between the two groups' realism scores. We did so 25 times per group size. We also calculated the root mean square errors (RMSE) of each image's realism scores in the similar way, using the data of all 160 participants as ground truth. When the number of participants was over 30, ρ was close to 0.8 and RMSE was around 0.075 (Fig. 1), suggesting 30 judgments per image is sufficient to reliably estimate visual realism.

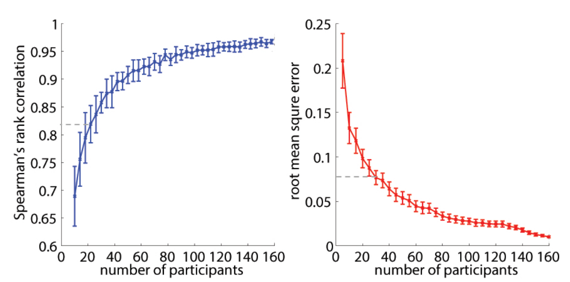

Fig. 1. Human performance analysis on pilot study. (a) Spearman's rank correlation between two random splits of participants as a function of participants size. (b) Root mean square error of image scores as a function of participants size. All results are averaged over 25 random splits.

Please look at the image on the left and answer the following questions:

1. Please rank the degree that the image appears to be a photograph versus a computer-generated image:
   (1) Definitely a photo
   (2) Probably a photo
   (3) Not clearly a photo or a computer-generated image
   (4) Probably a computer generated-image
   (5) Definitely a computer generated-image

2. Is the scene in the image familiar to you?
   (1) Yes, I think I have seen similar things/scenes before.
   (2) No, I have never seen similar things/scenes before.

3. Is this image attractive to you?
   (1) Very attractive
   (2) Moderately attractive
   (3) Neither attractive nor unattractive
   (4) Moderately unattractive
   (5) Very unattractive

4. How does the lighting effect appear to you?
   (1) Natural
   (2) Not clearly natural or unnatural.
   (3) Unnatural

5. Is the image colorful?
   (1) Very colorful
   (2) Moderately colorful
   (3) Not colorful

6. Does the color in the image appear natural?
   (1) Yes, the color appears natural.
   (2) No, the color appears strange or unusual.

7. Do the colors appear to go well together?
   (1) Yes, the colors go well together.
   (2) No, the colors seem strange/unusual together.

8. How sharp is the image?
   (1) Very sharp
   (2) Moderately sharp
   (3) Neither sharp nor blurry
   (4) Moderately blurry
   (5) Very blurry

9. What's the quality of the image?
   (1) Very high quality
   (2) Moderately high quality
   (3) Medium quality
   (4) Moderately low quality
   (5) Very low quality

10. Do you see shadows in the image?
   (1) Yes, there are obvious shadows in the image.
   (2) Yes, but the shadows are not obvious. They are only in a small region, are very light, or something similar.
   (3) No, there are no shadows in the image.

11. If you see shadows in the image, would you characterize them as sharp or soft?
   (1) Most shadows are sharp.
   (2) Some shadows are sharp, and some are soft.
   (3) Most shadows are soft.

12. Are there living things in the image?
   (1) Yes
   (2) No

13. Does the image show a naturally-occurring combination/arrangement of objects?
   (1) Yes, both combination and arrangement are natural.
   (2) Natural combination, but unnatural arrangement.
   (3) Natural arrangement, but unnatural combination.
   (4) No, both combination and arrangement are unnatural.

14. Are the objects in the scene look familiar to you?
   (1) Yes, all objects are familiar.
   (2) Some are familiar, some are unfamiliar.
   (3) No, all objects are unfamiliar.

15. Is the appearance of the objects natural?
   (1) Mostly natural
   (2) Some are natural, some are unnatural.
   (3) Mostly unnatural

16. Does the image contain fine details?
   (1) A lot of fine details
   (2) Some fine details
   (3) No fine details

17. Do the scene and objects look clean?
   (1) Yes
   (2) No

18. Is this image unusual or strange?
   (1) Highly unusual/strange
   (2) Moderately unusual/strange
   (3) Not at all unusual/strange

19. Does this image look like it is a photograph taken by a professional photographer?
   (1) Definitely yes
   (2) Probably yes
   (3) Not clearly yes or no.
   (4) Probably no
   (5) Definitely not

20. This is an image of:
   (1) Very neat space
   (2) Moderately neat space
   (3) Not clearly neat or cluttered space
   (4) Moderately cluttered space
   (5) Very cluttered space

21. This is an image of:
   (1) Very empty space
   (2) Moderately empty space
   (3) Not clearly empty or full space
   (4) Moderately full space
   (5) Very full space

1

22. Does the image appear to have objects of focus?
   (1) Definitely yes
   (2) Probably yes
   (3) Not clearly yes or no
   (4) Probably not
   (5) Definitely not

23. Does the perspective of the image appear natural?
   (1) Definitely natural.
   (2) Moderately natural.
   (3) Not clearly natural or unnatural
   (4) Moderately unnatural.
   (5) Definitely unnatural.

24. How exciting is this image?
   (1) Very exciting
   (2) Moderately exciting
   (3) Neither exciting nor boring
   (4) Moderately boring
   (5) Very boring

25. Is this image mysterious?
   (1) Very mysterious
   (2) Moderately mysterious
   (3) Not at all mysterious

26. How happy does this image make you?
   (1) Very happy
   (2) Moderately happy
   (3) Neither happy nor unhappy
   (4) Moderately unhappy
   (5) Very unhappy

27. Does this image make you sad?
   (1) Very sad
   (2) Moderately sad
   (3) Not at all sad

28. Is there a storyline in the picture?
   (1) Definitely yes
   (2) Probably yes
   (3) Probably not
   (4) Definitely not

29. Is the scene dynamic/energetic/in motion?
   (1) Very dynamic/energetic
   (2) Moderately dynamic/energetic
   (3) Not at all dynamic/energetic

30. Does the image appear to be a close-range shot or distant-view shot?
   (1) Very close range
   (2) Moderately close range
   (3) Between close and distant
   (4) Moderately distant view
   (5) Very distant view

If the image contains people, please answer the following questions, otherwise leave them BLANK.

31. Are the faces of any of the people visible?
   (1) Yes
   (2) No

32. Does any person appear to make eye-contact with a viewer of the image?
   (1) Yes
   (2) No

33. How many people are in the image?
   (1) One
   (2) Two
   (3) More than two, but not a crowd.
   (4) A large assembling of people

34. Do any persons who seem to be the focus of the image appear to be posing for the image?
   (1) Definitely yes
   (2) Probably yes
   (3) Definitely not

35. What are the persons who seem to be the main focus of the image doing? Check the box by each appropriate activity.
   (1) Standing, but not walking.
   (2) Sitting or lying down
   (3) Walking
   (4) Running
   (5) Riding/driving
   (6) Talking/Shouting
   (7) Other activities

If the faces of the persons in the main focus are visible, please answer the following questions, otherwise leave them BLANK.

36. Are the expressions of the persons who are the main focus of the image genuine?
   (1) Definitely genuine
   (2) Probably genuine
   (3) Perhaps genuine, perhaps fake.
   (4) Probably fake
   (5) Definitely fake

37. What are the expressions of the persons who are the main focus of the image? Check the box by each appropriate expression.
   (1) Happiness
   (2) Anger
   (3) Surprise
   (4) Sadness
   (5) Fear
   (6) Disgust
   (7) Contempt
   (8) Neutral (no expression)
   (9) Can't tell

38. How attractive are the persons who are the main focus of the image?
   (1) Very attractive
   (2) Moderately attractive
   (3) Not at all attractive

2

Fig. 2. Questionnaire for image attribute annotation on Amazon Mechanical Turk.

3 Questionnaire for psychophysics study II

Fig.2 shows the questions used for image attribute annotation on Amazon Mechanical Turk.

4 Empirically-based feature design

In this section, we describe in detail our feature design. The list of features is reported in Table 1.

4.1 Naturalness

We have shown that naturalness is informative of image realism. We focused on two features of naturalness that are computable using current computational methods, which are *natural semantics* and *natural color*.

Natural semantics: The idea was inspired from Datta and Wand [5], where they proposed a *familiarity* feature in the classification of image aesthetics. Similarly, we defined a measure for *semantic familiarity* using the content-based similarity measure commonly used in image retrieval. We used $10,000$ images from the SIMPLIcity dataset [6] as a pre-determined *anchor* database of images with common scenes and objects. We then computed the image similarity by using color, illumination and texture information [7], and performed a robust content-based matching with the anchor database. Primarily meant for image retrieval applications, we used it here to quantify familiarity. The familiarity measure was denoted by the distances of the top 50 matches.

In effect, these measures should yield higher values for uncommon images. Because of the strong correlation between visual realism and unusualness, it is intuitive that a higher value of familiarity corresponds to greater unusualness and hence we expect lower realism score.

Natural color: [8], [9] suggested that an image will look realistic if its color is consistent with the colors in human memory. We computed a *color familiarity* feature by employing the method in [10]. Similar to [10], each pixel in our image was classified as a color name which are learnt from real-world images. We densely sampled the feature with a grid spacing of 4 and learned a dictionary of size 256. We then applied 2-level spatial pyramid pooling to obtain the color descriptors. We finally got 5376 dimension feature on color familiarity.

Natural image statistics: [11] introduced several statistical models that represent the regularities inherent in natural images. High contrast local image patches which mainly correspond to the edge structures were studied and shown to display some regular patterns. This motivated us to use gradient information in modeling image naturalness. Let $I(x, y)$ denote the image intensity, we computed the surface gradient of the image intensity with a scaled constant α as Equation 1:

$$|grad(\alpha I)| = \sqrt{\frac{|\nabla I|^2}{\alpha_{-2} + |\nabla I|^2}} \quad (1)$$
$$where\ |\nabla I| = \sqrt{I_x^2 + I_y^2}$$

The constant α was to control the weight of emphasis on the low gradient region versus the high gradient region. We computed the gradient on R, G, B channels ($|grad(\alpha I)|_R$, $|grad(\alpha I)|_G$, $|grad(\alpha I)|_B$) at every pixel of an image with $\alpha = 0.25$. We used spatial pooling to reduce the dimension to 98 in the final algorithm.

TABLE 1
Features used in computational realism assessment.

Attribute	Feature	Dimension
Naturalness	Content-based similarity measure	50
	Color compatibility	48
	Color familiarity	5376
	Natural image statistics	98
Attraction	Saturation, hue, and illumination	6
	Contrast	3
	Edge distribution	1
	Blur	1
	Self geometric similarity	5376
Oddness	Local outlier factor	3
Face	Face detector	2

4.2 Attraction

To capture the aesthetics of an image, we propose several features that are commonly used by extensive works in the aesthetic evaluation area. We first employed basic first and second order HSV features. We then applied Ke's method [12] for extracting three aesthetic features, which are luminance, contrast, and edge distribution. We also used local self-similarity geometric patterns (SSIM [13]) to represent content symmetry, which is often regarded as a measure of aesthetics. The detailed descriptions are as follows.

Saturation, hue, and illumination We computed features defined in the HSV space. Saturation indicates chromatic purity. Pure colors in a photo tend to be more appealing than dull or impure ones [5]. We computed the average saturation $f_s = \frac{1}{XY}\Sigma_{x=0}^{X-1}\Sigma_{y=0}^{Y-1}I_S(x,y)$ as the saturation indicator. Hue and illumination were similarly computed by averaging over I_H and I_V separately. Although the interpretation of such features is not as clear as saturation, they were found to be predictive of image aesthetics [5], [12]. We also calculated their variances and got a six-dimension feature in total.

Contrast: We used the similar contrast quality measure as [12], except that we computed the gray-scale level histogram of each image on R, G, B channels separately, and measured the width of the middle 98% gray level mass on each channel.

Edge distribution: The spatial distribution of the high frequency edges of an image was computed to capture its *simplicity*. A uniform distribution of edges might indicate snapshots having cluttered backgrounds, while the opposite may indicate aesthetic photos that have well defined subjects and objects in focus [12]. Similar to [12], we applied a 3×3 Laplacian filter with $\alpha = 0.2$ to the R, G, B channels of an image separately and took the mean across the channels. We then normalized the Laplacian image sum to 1. We calculated the area of the bounding box that encloses the top 96.04% of the edge energy of the Laplacian image L by projecting it to the x and y axes independently, so that the area of the bounding box is denoted by $1 - w_x w_y$, with w_x and w_y being the box's normalized width and height.

Blur: The degree of blur of an image is a strong indication for its quality and aesthetics. A blurry photo of a scene is almost always worse than a sharp photo of the same scene [12]. For blur prediction, we estimated the maximum frequency of the image I_b by taking its two dimensional Fourier transform and counting the number of frequencies whose power was greater than some threshold θ. We then normalized it by the size of the image [12]. We set $\theta = 5$ in our algorithm.

Self-similarity pattern: Following [13] we calculated local "self-similarity (SSIM) descriptors" by computing a correlation surface from each pixel q in an image. We densely sampled the SSIM descriptors with a grid spacing of 4 and learned a dictionary of size 256.

5 Feature dimensions of comparing algorithms

We compared our algorithm with other state of the art methods. Table. 2 shows the feature dimensions of the comparing methods.

TABLE 2
Feature dimensions for comparing algorithms.

Category	Feature type	Dimension
Signal feature	Wavelet	216
	Geometry feature	196
	Camera noise	4
	Color compatibility	77
Object & scene feature	SIFT	1280
	GIST	512
	HOG2x2	2100
	LBP	1239

References

[1] H. Jegou, M. Douze, and C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search," in *ECCV 2008*, pp. 304–317, Springer, 2008.

[2] T.-T. Ng, S.-F. Chang, J. Hsu, and M. Pepeljugoski, "Columbia photographic images and photorealistic computer graphics dataset," *Columbia University, Advent Technical Report*, pp. 205–2004, 2005.

[3] C. Barron, "Matte painting in the digital age," in *ACM SIGGRAPH 98 Conference abstracts and applications*, p. 318, ACM, 1998.

[4] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, "Sun database: Large-scale scene recognition from abbey to zoo," in *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*, pp. 3485–3492, IEEE, 2010.

[5] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach," in *ECCV 2006*, pp. 288–301, Springer, 2006.

[6] J. Z. Wang, J. Li, and G. Wiederhold, "Simplicity: Semantics-sensitive integrated matching for picture libraries," *TPAMI*, 2001.

[7] Kirk, "Content based image retrieval." https://github.com/kirk86/ImageRetrieval, 2013.

[8] J. Lalonde and A. Efros, "Using color compatibility for assessing image realism," in *ICCV*, 2007.

[9] S. Y. Choi, M. Luo, M. Pointer, and P. Rhodes, "Investigation of large display color image appearance-lll: Modeling image naturalness," *JIST*, vol. 53, no. 3, pp. 31104–1, 2009.

[10] J. Van De Weijer, C. Schmid, and J. Verbeek, "Learning color names from real-world images," in *CVPR, 2007*, pp. 1–8, IEEE, 2007.

[11] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, "On advances in statistical modeling of natural images," *Journal of Mathematical Imaging and Vision*, vol. 18, no. 1, pp. 17–33, 2003.

[12] Y. Ke, X. Tang, and F. Jing, "The design of high-level features for photo quality assessment," in *CVPR*, vol. 1, pp. 419–426, IEEE, 2006.

[13] E. Shechtman and M. Irani, "Matching local self-similarities across images and videos," in *CVPR, 2007*.