# Image Quality Assessment Using the SSIM and the Just Noticeable Difference Paradigm

Jeremy R. Flynn, Steve Ward, Julian Abich IV, and David Poole

University of Central Florida, 4000 Central Florida Boulevard, Orlando, FL, USA
{jflynn,jabich}@ist.ucf.edu, steve@iqatest.com,
mr.david.a.poole@gmail.com

**Abstract.** The structural similarity index (SSIM) has been shown to be a superior objective image quality metric. A web-based pilot experiment was conducted with the goal of quantifying, through the use of a sample of human participants, a trend in SSIM values showing when the human visual system can begin to perceive distortions applied to reference images. The just noticeable difference paradigm was used to determine the point at which at least 50% of participants were unable to discern between compressed and uncompressed grayscale images. For four images, this point was at an SSIM value of 96, while for two images it was at 92, for an average of 95. These results suggest that, despite the wide differences in the type of image used, the point at which a human observer cannot determine that compression has been used hovers around an SSIM value of 95.

**Keywords:** Applied cognitive psychology, Designing for pleasure of use, Display design, Formal error prediction techniques, Human error, Human Factors / System Integration, Psychophysics for display design.

## 1 Introduction

The Internet is rich with images and media consumption is at an all-time high. According to a Pew report on online usage of photos and videos, 56% of the internet users sampled either created and uploaded photos to the internet or took existing images and reposted them to image sharing websites [1]. Websites that cater to this behavior are wildly popular. Tumblr.com has a blogging service where users primarily post images and videos, and has ranked the 36th most visited website in the world, followed closely by Pinterest.com, an online pin board that essentially has a wall of images from all over the Internet, which has ranked 38th [2]. Imgur.com is ranked 97th globally [2], and its only function is for users to share and display uploaded images. In an average month, there are over 61 million photos uploaded, 33 billion image views, and over 4 petabytes of bandwidth used by Imgur alone [3]. With such a large amount of traffic, it becomes important to optimize bandwidth usage and load times which requires the compression of images. Imgur's policy is to automatically compress, resize, and adjust the quality of images that are otherwise too large in an effort to make them more easily viewable online and to save space [4], but this may

D. Harris (Ed.): EPCE/HCII 2013, Part I, LNAI 8019, pp. 23–30, 2013.

noticeably decrease the image quality. The objective of our work was to reveal whether an image quality index can be used to determine the point in which human observers cannot tell the difference between compressed and uncompressed images. This metric could then be applied to all compressed visual media, but here the focus is online image databases due to the potential impact in this domain. Online image database services could use the metric as a part of an automated image adjustment procedure to ensure that image compression does not noticeably detract from perceptual quality.

Images are not stored as raw source signals, instead they are compressed into a format. According to Shen and Kuo, the quality of the compressed image depends on the data source, coding bit rates, and the compression algorithm [5]. For lossy compression, which includes JPEG, the researchers state that there is a trade-off between lower bit rates at the cost of increased distortion in image quality. JPEG is an acronym for Joint Photographic Experts Group and is formally defined by a joint ISO/ITU-T standard, ISO/IEC IS 10918-1 or the ITU-T Recommendation T.81 [6]. Raw digital images compressed in the JPEG format are ubiquitous on the internet, in presentations, and in documentation. JPEG images, even at the lowest compression, are smaller in storage size than many other types of image formats [7].

Small storage size is important when dealing with large servers which contain, in some cases, millions, or even billions, of images. In this situation, it is advantageous to minimize image file sizes while maintaining sufficient image quality, such that an average human observer cannot perceive a distortion or loss of image quality due to compression of the original image. Since the JPEG format is very common on the internet and with digital imagery, it was chosen as the type of distortion to be applied to the reference images used in this study.

Having a large sample of subjects available to quickly and efficiently determine the quality of an image, or determine when an image reaches a level of distortion that is detectable, is not practical or feasible. To address this need, there are a range of different methods of analyzing images compression as it relates to the perceptual capabilities of the human visual system (HVS), ranging from mathematical algorithms to complex models that seek to analyze and quantify elements of an image based on features pertinent to the HVS, such as contrast, masking, and summation [8]. In light of the multiple different methods available, the most useful would be an analysis metric that could quickly quantify the image. Traditional methods to achieve this include the mean squared error (MSE) or peak signal-to-noise ratio (PSNR). The usage of MSE, for example, may be problematic as it does not always provide an adequate evaluation of image quality as it would be perceived by the HVS [9]. Rather than a simple measurement of error between signals, an algorithm that accounts for structural similarity between images would better model how the HVS perceives distortion. The structural similarity approach depends on the assumptions that natural images are highly structured and the HVS is suited for extracting structural information from scenes. It then follows that an accurate approximation of perceived image quality should be the measurement of structural similarity. Research has shown that an algorithm based on a structural similarity approach, such as the Structural SIMilarity (SSIM) index, more closely resembles the way humans perceive structural distortions in an image and thus assess image quality [9].

The purpose of this study was to collect and analyze subjective responses from human participants to determine if the point at which a compressed image is noticeably different from the uncompressed original aligns with a particular SSIM value, which can then be used to predict the point at which the average human can begin to perceive distortion due to compression in an image.

# 2 Method

## 2.1 Reference Images

We initially planned on using images from a common image database such as the University of Southern California's SIPI database [10]. However, we discovered that many of these images lacked the necessary requirements that were desired in a set of reference images, which includes high quality and a variety of subjects. Additionally, most of the images in that database were under some form of copyright protection or the copyright status was unknown. It was decided that the reference images used in the web-based survey would be of high quality and be free of any copyright issues.

The six images chosen to be reference images in this study were selected from the Wikimedia Commons [11] website because they all met the criteria of being high quality and of varied subject matter. They were all freely licensed under the Creative Commons Attribution-Share Alike 2.5 Generic license [12]. The images exhibited various characteristics, including a public domain image of Albert Einstein, a landscape of the Arnisee region in Switzerland, a bald eagle, a complex pattern of cracks in desiccated sewage, an apple, and a windmill.

ImageMagick [13], an open-source image processing utility, was used to convert the original color reference images to grayscale, resize them to have maximum dimensions of 384 pixels, and apply the various degrees of JPEG compression. This was done in an effort to systematically control how all of the images were processed. The value of 384 pixels was chosen due to the limitations of small screen resolutions that could possibly be used by some of the participants. Octave [14], an open-source numerical computation tool, was used to calculate the SSIM values for all the distorted images. Figure 1 shows how various degrees of distortion affect image quality, as measured by the SSIM, where lower values translate to lower image quality.

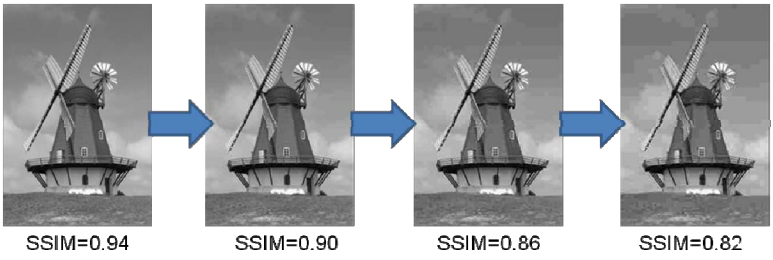


**Fig. 1.** Increasing degrees of distortion and associated SSIM values

### 2.2 Participants

A total of 30 participants, comprised of 24 females and 6 males, completed the online experiment and submitted results. They were recruited through word-of-mouth requests and online postings. No compensation was given for participation in this study. Participants were also informed of their ability to withdraw from the study at any time, in which case none of their biographical data or results would be submitted. The mean age of the participants was 34 years ($SD$ = 15).

### 2.3 Biographical Questionnaire

If participants agreed to participate in the experiment, they were directed to a biographical questionnaire. The questionnaire collected biographical data such as the participant's age, sex, primary language, quality of vision (normal or corrected-to-normal), experience with photo editing or image processing, and computer and video game proficiency.

### 2.4 Image Comparisons

After completing the questionnaire, participants were presented with an instructional page that described the task they would be performing. A practice image comparison session was then given to familiarize participants with the task. The image used was of the Giza Necropolis, which was also selected from the Wikimedia Commons and processed in the same manner as the six reference images. This particular image was not used as a part of the actual experimental task. Participants were presented with the following instructions: "These images are different. One of them has severe distortion. Severe distortions are noticeable by their blockiness. A distorted image may appear on either the left side or the right side. These messages will not be shown during the actual experiment. They are only instructional." Participants were presented with two buttons to click, "Identical" and "Different." They were then given feedback about the decision they made in the practice session, but it was made known that no feedback would be given to participants during the actual task.

For each of the 6 reference images, 10 degrees of JPEG compression were applied. The reference image and its distorted versions comprised an image set. To each image, different levels of JPEG compression were applied until the image reached the following specific SSIM values: 82, 84, 86, 88, 90, 92, 94, 96, 98, 99.9. This resulted in 10 different versions of each of the six images in addition to the original reference with varying levels of distortion as quantified by the SSIM.

Within an image set, the reference image was compared to itself 10 times and compared once per distorted version of the reference. Consequently, in each image set, 20 image comparisons were made. The sets were presented in a random order, as were the distorted and reference images in each set. The reference image was placed randomly on either the right or left side of its counter image. Every comparison was made one at a time.

On each page, participants were presented with the two images. Participants were also given the following instructions: “If you can perceive a difference in the images, select Different. If you cannot perceive a difference in the images, or you are unsure if they are different, select Identical.” One of the images was the original reference while the other was the same image with some level of distortion applied. Participants then clicked one of the two buttons, “Identical” or “Different.”

After a selection of “Identical” or “Different”, the images disappeared for a brief inter-stimulus interval to reduce any visual artifacts, and the buttons were disabled to prevent accidental double-clicking. The next pair of images appeared 300ms after the previous pair had disappeared. The buttons were enabled 500ms after the new image pair appeared. While the buttons were disabled, no selection could be made. All the images were pre-loaded to avoid any delay in the image presentation.

Participants were given the opportunity to take as many breaks as they liked. They could work at their own pace and were not restricted to complete the task in a particular amount of time. The client-side code was written in JavaScript. When the last image comparison was made, the result data and the biographical information were serialized from a JavaScript object into a string using the JavaScript Object Notation (JSON) library. The JSON-formatted string was submitted automatically without user interaction. The server-side code that processed and stored the results was written in PHP. Results could be downloaded for further analysis in a spreadsheet.

## 3 Results

Based on the answers to the biographical questionnaire, 28 participants reported having normal or corrected-to-normal vision, with only 2 reporting they did not. Sixty-seven percent of the participants had some type of prior image processing or photo editing experience, while 33% had no such experience. On average, participants spent about 32 hours per week using a computer for various tasks. Twenty-six participants reported spending little to no time playing video games. Four participants played video games more than 20 hours per week, which raised the average of video game use to about 7 hours per week ($SD$ = 15.75).

Seventy-seven percent of the participants felt they were above average in computer proficiency, 13% felt they were average, and 10% felt they were below average. Thirty-two percent of the participants felt they were above average in video game proficiency, 46% felt they were average, and 22% felt they were below average.

The data were scanned in an effort to remove those participants who may not have been completing the task, but rather were simply clicking buttons. As noted previously, for each of the six images the participant was presented with an image set ten times where both images were the uncompressed, original reference image. If they responded more than 50% of the time that the identical images were different, this indicates that they may have not been actually completing the task. We conducted the analysis in both the original and cleaned datasets, and while some of the averages were different between the two, ultimately the results were the same. We decided to retain the cleaned data as it is more accurate.

The just noticeable difference (JND) paradigm was utilized in this study. The JND is the point in which half of the participants report perceiving a difference between two stimuli. Eckert and Bradley [8], citing the previous work of Watson et al., suggested that utilizing JND is an effective method of determining the point at which an individual is able to perceive the visual difference between compressed and uncompressed images. Therefore, we examined the data across SSIM values for each image to determine the point at which at least 50% of participants were unable to discern between compressed and uncompressed images. See Figure 2 for a visual representation of the percentage of participants who perceived the compressed and uncompressed image as being identical by SSIM value for each of the six images used in this study, and see Table 1 for the actual values with the JND SSIM value highlighted.

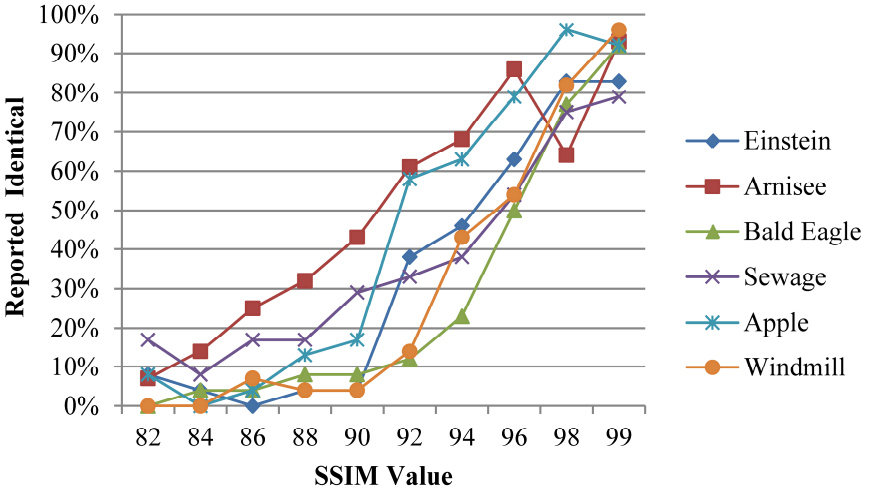


**Fig. 2.** Percentage of participants that reported that the uncompressed and compressed images were identical by SSIM value for all six images

**Table 1.** Percentage of participants that reported that the uncompressed and compressed images were identical by SSIM value for all six images with JND point highlighted

| SSIM | 82 | 84 | 86 | 88 | 90 | 92 | 94 | 96 | 98 | 99 |
|---|---|---|---|---|---|---|---|---|---|---|
| Einstein | 8% | 4% | 0% | 4% | 4% | 38% | 46% | 63% | 83% | 83% |
| Arnisee | 7% | 14% | 25% | 32% | 43% | 61% | 68% | 86% | 64% | 93% |
| Eagle | 0% | 4% | 4% | 8% | 8% | 12% | 23% | 50% | 77% | 92% |
| Sewage | 17% | 8% | 17% | 17% | 29% | 33% | 38% | 54% | 75% | 79% |
| Apple | 8% | 0% | 4% | 13% | 17% | 58% | 63% | 79% | 96% | 92% |
| Windmill | 0% | 0% | 7% | 4% | 4% | 14% | 43% | 54% | 82% | 96% |

For four images, the JND was at an SSIM value of 96, while for two images it was at 92, for an average of 95. These results suggest that the point at which a human observer cannot determine that compression has been used hovers around an SSIM value of 95.

## 4 Discussion

Using the just noticeable difference paradigm, we examined the data across SSIM values for each image to determine the point at which at least 50% of participants were unable to discern between compressed and uncompressed grayscale images. Our results suggest that, despite the wide differences in the type of image used, the point at which a human observer cannot determine that compression has been used hovers around an SSIM value of 95. This is useful since the SSIM can be used to analyze images after compression to predict whether the decrease in quality will be perceptible by the user.

It is important to note that two images (the landscape of the Arnisee region and the apple) reached the JND at an SSIM value of 92, while participants reported the JND for the other four images at 96. This may indicate that the content of the image itself affects the JND point from either a bottom-up or a top-down processing perspective. Regarding bottom-up visual perception, the HVS processes visual information by analyzing structures, which is why techniques such as SSIM are so apt at predicting visual perception of distorted images [15]. Future research efforts could focus on how different types of features that are perceived by the HVS, as represented in a wide variety of images, affect the JND as quantified by the SSIM. This would allow for parsing the components of human vision against which the SSIM algorithm can be tested. An alternative method would be to examine the top-down approach of visual perception and examine how the content, meaning the actual subject, of images affect the JND. For example, previous research has suggested that some images, such as faces, are processed in different areas of the brain as compared to other objects [16]. In the present study, the face of Albert Einstein was not perceived any differently with respect to the JND point from an image of an eagle, a windmill, or the complex pattern of cracks in desiccated sewage. Replicating this study that combines both processing perspectives with a wide variety of images that focus on familiar and unfamiliar faces and objects, as well as images that manipulate the type and complexity of HVS features, may reveal a different pattern of results.

This study had some limitations. The results only pertain to grayscale images. One potential avenue of future research involves replicating this study using the same images displayed in full color to determine if the JND point changes. This is especially pertinent as it seems that the majority of images on the internet are in color, not grayscale.

Another limitation of this study was the sample size. The purpose of this work was to develop a method of quantifying image compression perception based on SSIM utilizing the JND paradigm. As a pilot experiment, it revealed that this method does yield meaningful results. Replication using these methods on a large scale by companies that deal in image hosting services would allow for a far-greater sample size. A company could embed the survey within their website through a pop-up message that offers the user a chance to complete a survey--this is essentially a crowd-sourcing technique for data collection and a method already employed by some companies to gather customer feedback [17]. This would provide the company with information based on their own image set as how to best automate the compression of their images

based on SSIM, utilizing the JND paradigm or even selecting their own criterion (e.g., a website specializing in art may wish to determine the SSIM value at which 90% of users cannot discern between compressed and uncompressed images). Hopefully these steps provide additional evidence and guidance for the ways that the SSIM index value can be used to determine optimal image compression.

## 5 Author's Note

The experiment address is `http://iqatest.com`. The source code may be downloaded from Google Code at `https://code.google.com/p/iqatest`. It is released under the GNU General Public License v3, copyrighted 2010 Steve Ward.

## References

1. Rainie, L., Brenner, J., Purcell, K.: Photos and videos as social currency online (2012), `http://pewinternet.org/Reports/2012/Online-Pictures.aspx`
2. Alexa Top Sites, `http://www.alexa.com/topsites`
3. Imgur Site Stats, `http://imgur.com/stats/month`
4. Imgur FAQ, `http://imgur.com/faq#quality`
5. Shen, M.Y., Kuo, C.C.J.: Review of Postprocessing Techniques for Compression Artifact Removal. J. Vis. Commun. Image Represent. 9, 2–14 (1998)
6. Skodras, A., Christopoulos, C., Ebrahimi, T.: The JPEG 2000 Still Image Compression Standard. IEEE Signal Process. Mag. 18, 36–58 (2001)
7. Richardson, G.: Jpeg Compression (2003), `http://photo.net/learn/jpeg/`
8. Eckert, M.P., Bradley, A.P.: Perceptual Quality Metrics Applied to Still Image Compression. Signal Process. 70, 177–200 (1998)
9. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Trans. Image Process. 13, 600–612 (2004)
10. University of Southern California's SIPI Image Database, `http://sipi.usc.edu`
11. Wikimedia Commons, `https://commons.wikimedia.org`
12. CC BY-SA 2.5 Generic License, `https://creativecommons.org/licenses/by-sa/2.5/`
13. ImageMagick, `http://www.imagemagick.org`
14. GNU Octave, `https://www.gnu.org/software/octave/`
15. Wang, Z., Bovik, A.C., Lu, L.: Why is Image Quality Assessment So Difficult? In: IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. IV-3313–IV-3316. IEEE Press, New York (2002)
16. Kanwisher, N., McDermott, J., Chun, M.M.: The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. J. Neurosci. 17, 4302–4311 (1997)
17. Chan, S.: That popup survey tool for Fresh & New feedback (2011), `http://www.freshandnew.org/2011/01/that-popup-survey-tool-for-fresh-new-feedback/`