SERF: Optimization of Socially Sourced Images using Psychovisual Enhancements

Aaron Koehl
Dept. of Physics, Computer Science, and Engineering
Christopher Newport University
Newport News, VA 23606, USA
aaron.koehl@cnu.edu

Haining Wang
Dept. of Electrical and Computer Engineering
University of Delaware
Newark, DE 19716, USA
hnw@udel.edu

ABSTRACT

Online communities and social networks are the most popular sites on the Internet, and have exploded with multimedia content in the last decade. Most web designers recognize that site images can be saved with lower fidelity to reduce bandwidth consumption and increase capacity, though many are reluctant to do so for aesthetic concerns. However, there are many images that site designers have little direct control over—socially sourced images. Many social networks automatically reduce the fidelity of uploaded images in order to conserve bandwidth. Social networks also contain a vast archive of images with popularity indicators, such as likes and shares, which recent work has correlated with psychovisual features within the images. In this paper, we investigate the trade-off between fidelity reduction and selected psychovisual enhancements. We demonstrate that even simple enhancements can be used to enable more aggressive optimization of socially sourced content, which has implications for static content delivery networks and image servers. Through user testing on real images, we validate the efficacy of our proposed approach.

CCS Concepts

•Information systems → Multimedia content creation; *Social networking sites;*

Keywords

psychovisual enhancements, image fidelity

1. INTRODUCTION

Psychophysics seeks to investigate the relationship between sensory stimuli and perceived sensations, usually associating the stimuli with some measurable quantity to determine points along a function frontier. These ideas underpin the development of lossy compression techniques in both the visual and aural media. By identifying stimuli that humans are sensitive to (and conversely, those they are tolerant to), we can inform the development of new compression techniques.

The point along the frontier at which humans become perceptually aware of some stimulus is known as the difference limen. In the case of lossy image compression, where high frequency luminance and chrominance information are algorithmically discarded, the point at which artifacts become perceptible relative to some reference image is called the "just noticeable difference" (JND) or psychovisual threshold.

Given the proliferation of image compression libraries and codecs, most libraries follow a "one size fits all" approach, for instance by using standard quantization matrices with limited parameterization. Though other parameters exist in these libraries, the most useful parameter is the quality factor. It is helpful to think of the useful psychovisual threshold as a range of quality factors, as the actual value is very turbulent and highly image-dependent. Since the psychovisual threshold is different for every image, many web developers are reluctant to optimize near the psychovisual threshold, instead adopting a "safe" range much higher than necessary out of convenience. Commonly, this means saving images with a quality factor between 85 and 95, with a corresponding file size two or three times larger than an acceptable file saved at lower fidelity.

Recent work has focused on autotuning the quality parameter to the psychovisual threshold, by experimentally determining a frontier against measurable computer vision features [15, 30]. The availability of good, efficient machine learning algorithms (such as SVM) has paved the way for these advances. Further, the availability of big data sets harvested from massive social media platforms allows extension into social research domains, such as the problem of predicting whether an image will be "liked" or "repinned" [4, 7, 26]. As we will show, the insights gained from these recent works are not limited to optimizing ad spend, but can also be applied to multimedia compression.

Machine learning has also allowed for automatic correction and enhancement of common image problems. Algorithms that have been present in high-end commercial photography software for years are now able to be informed by very large data sets. The adoption of auto-enhancements in Google Plus, Facebook, Etsy, and Instagram are highly visible examples demonstrating that image enhancements can be performed at scale [9, 13, 20].

In this paper, we challenge current notions of usable quality (quantization) parameters by positing the following: *Can we psychovisually enhance an image to make fidelity reduction more palatable?* We show that this is indeed the case,

MMSys'16, May 10-13, 2016, Klagenfurt, Austria

DOI: http://dx.doi.org/10.1145/2910017.2910609

Figure 1: SERF data set. Original images (left). Images on the right represent minimally processed psychovisual enhancements used in our study [31].

by exploiting recent research findings correlating psychovisual features to popularity [4, 26], and by applying these features to performance optimization.

The main contributions of this work are summarized as follows:

1. We introduce SERF (Scaling with Enhanced Reduced Fidelity), a technique for compressing images beyond the psychovisual threshold, and demonstrate its utility through user acceptance testing.

2. We characterize the emergent problem of social image growth and enable its optimization within the constraints of existing lossy web image formats and readily available image transformations.

3. We challenge existing notions of image quality settings for socially sourced images and provide new research directions for *situational performance optimization*, by leveraging recent advances in image characterization.

The rest of this paper is organized as follows. Section 2 surveys related work. Section 3 characterizes the landscape of images on the Internet and socially sourced images. Section 4 provides an overview of lossy compression techniques and recent advances in recompression. Section 5 introduces our technique and advances our hypothesis. Section 6 details our experiments and presents our evaluation results. Section 7 includes a discussion and future directions, and Section 8 concludes with a summary of our findings.

2. BACKGROUND AND RELATED WORK

Due to the increase in video and image traffic on the Internet combined with the surge in the use of social networks, especially for mobile devices, much research focus has been devoted to optimizing content distribution infrastructure. The focus is jointly on quality of experience (QoE) for the client and bandwidth management for the data center. The increase in volume has propelled advances in locality prediction [14] and content distribution [24, 33], while device constraints have led to infrastructure and protocol developments [17, 34, 35]. Other efforts have been placed in identifying new image formats [11] which are promising but unfairly hindered by their lack of adoption in consumer devices. There has also been considerable interest from the web research community in analyzing and predicting popularity based on accessible features residing within social image data sets. We survey recent related work in three broad areas.

Image Popularity and Metrics. Popularity prediction receives much attention due to its intrinsic economic value—if we can predict which images will receive the most “likes”, “shares”, or “pins”, companies can focus their efforts on producing and posting these types of images. Khosla et al. studied the idea of what makes an image popular by extracting both low level and high level computer vision features, and correlating them with normalized view counts extracted from Flickr. Using support vector regression, they developed a tool to predict an image’s popularity with a rank correlation of 0.81, specifically by combining psychovisual features with social cues [26].

The data analytics firm Curalate applies image-recognition algorithms to the study of social media trends, publishing results using large data sets from both Pinterest and Instagram. By extracting low level features from 8 million Instagram images, and correlating them with “likes”, the study found that features such as high lightness (vs. low), blueish images (vs. reddish), single dominant colors (vs. multiple), low saturation, and high texture (vs. smooth) all contribute to between 18% and 80% more likes [4].

Examining over 30 visual characteristics of 500,000 images on Pinterest and correlating them with “repins”, Curalate found that features such as medium lightness were repinned 20 times more frequently than dark images, and images with 50% saturation are repinned 10 times more often than desaturated images. Curiously, in contrast to Instagram, reddish-orange images (vs. blueish) received twice as many repins, and smooth images (vs. high texture) are repinned 17 times more [7].

While the primary focus of these studies is on digital marketing strategy and how to develop effective visual imagery on social networks, we see a tremendous value in applying these social studies to performance optimization, since many psychovisual enhancements can be machine-learned and al-

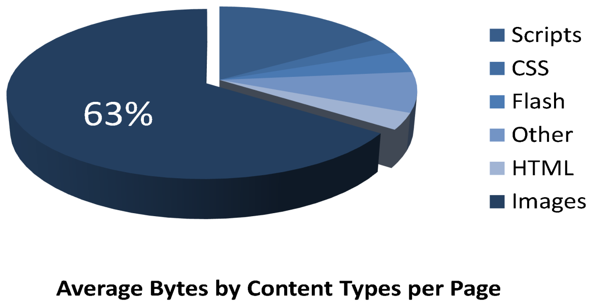

Figure 2: Images account for 63% of total byte transfer per page, on average.

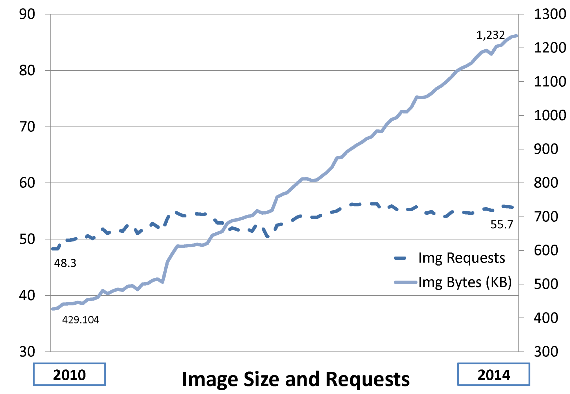

Figure 3: Image sizes have increased steadily, while requests have remained relatively constant.

gorithmically applied. These studies indicate the "what", and it will be worthwhile in social research to surmise and test the "why"; nevertheless, the studies provide useful quantitative and qualitative insights.

Image Compression. Image fidelity has long been studied in the context of performant multimedia. On the client side, this includes well-performing video codecs (divx, mp4) and image compression schemes (jpeg, png, gif) which are in general usage today. The GIF image format uses a reduced 8-bit palette of colors and employs (lossless) LZW entropy coding. GIF is appropriate for most interface graphics, but generally does not perform well for photographs, since a large amount of color information must be thrown away. However, dithered GIFs would provide an additional avenue of research, as this introduces perceptible artifacts which may be offset by psychovisual enhancements. Still, GIFs of photographic scenes are regularly outperformed by lossy compression. JPEG2000 employs wavelet transforms and suffers less from artifacts than JPEG, but the format is not yet as widely supported. Our technique currently relies on lossy compression and quantization for its efficacy.

More recent research includes Google's WebP format, which is perceptually similar to JPEG yet smaller [11], and is derived from the VP8 video format. WebP supports both lossy and lossless compression modes, and is based on block-prediction. We discuss pertinent tradeoffs between JPEG and WebP in our evaluation, but note that our technique can be applied regardless of format.

Psychovisual Enhancements. Due to the prevalence of cameras embedded in smart phone devices, there are several current research studies in providing systems for mobile image enhancement systems [27]. These systems focus on color enhancement, noise reduction, and adjustments for skin tones with a nonfunctional requirement of being easy to use.

Closely related are middleware systems which aim to deliver a faster multimedia experience, either by decreasing client side rendering latency, or by reducing file size to enhance server throughput. Historically, AOL (America Online) adopted the ART format as a highly compressed image format, still included today in AOL's TopSpeed web proxy service [5]. More recently, Facebook began adoption of Google's WebP format [11], which is supported in Google Chrome natively, and most other browsers via plugins.

In 2013, Google unveiled a feature called Auto Enhance, which is by default used on uploaded photos to the Google Plus social platform [20]. This machine-learning system algorithmically enhances exposure and colors, providing a psychovisually more stimulating user experience. Facebook unveiled a semi-automated procedure (magic wand) in December of 2014, and allows users the choice of modification level. While the source is not available for testing, it demonstrates that psychovisual enhancements can be performed at scale.

ImageMagick, GraphicsMagick and GD are also used by social networks to process large numbers of image uploads at scale [13]. We mention them here as they use a set of standard quantization coefficients so that our results can be reproduced.

3. IMAGE CHARACTERIZATION

HTTPArchive [23] provides timing and request data for web research, by crawling the top 1 million landing pages in Alexa [6] using both standard browser and mobile test frameworks. We present data harvested from the standard desktop crawl between 2010 and 2014 and give an interpretation below, as the motivation for our study.

Web pages have grown considerably in just the last four years in both the number of requests per page and the size of those requests. This, despite performance advice to the contrary to reduce requests, which enhances QoE by saving on DNS lookups and rendering time [32]. Fortunately, many graphics can be cached for future requests. Relative to other content types, static images account for 63% of bytes transferred as shown in Figure 2.

From 2010 to October 2014, aggregate image transfer sizes have increased by a factor of 3, while the number of images per landing page has remained relatively stable (within 10%), as noted in Figure 3. We attribute this to photographic trends in web design, availability of HD displays in mobile devices, and increased availability of broadband.

To put image trends in perspective, we include overall trends, as shown in Figure 4. Web fonts have dramatically increased in popularity, which are loaded on demand unlike traditionally installed system fonts. CSS requests have doubled over time, yet the average CSS bytes per request have remained constant. We attribute the increase in JavaScript size to the increase in the use of libraries, such as jQuery [3]. Both the number and the size of HTML requests have also increased, though this is partially attributed to a change in testing methodology to accommodate lazy loading[1].

3.1 Social Networks

Overall, web pages are becoming much larger (Figure 5) and this trend is expected to continue, placing higher demands on server infrastructure and network bandwidth. It is also important to note that these trends do not fully characterize additional content growth in social media. For example, the average adult Facebook user in the U.S. has spent 39 minutes per day on the site in 2014 [16], with comparatively little content downloaded from its landing page in the Alexa top 1M. Further, the distribution of content by file format, with jpeg consisting of 46%, gif at 24%, and png at 28%, is typical of promotional and navigational graphics found on the web—in particular, over half are lossless formats, which are especially suitable for visually regular images, rather than the photographs found on social networks.

OSNs have exploded with multimedia content in recent years, where jpeg is by far the most prevalent because it is the most commonly supported format generated by mobile devices and consumer cameras, and is designed for high-color photography. For example, on the author's online forum site running vBulletin [2], of one quarter million images collected since 2003, jpeg accounts for 96.5% of all socially sourced images. With lossy jpeg as the dominant format, *is there an opportunity to optimize for this?*

4. LOSSY COMPRESSION

We give an overview of lossy image compression using the jpeg format, to provide intuition regarding the quality factor. A standard raster image in the RGB color space is made up of a grid of pixels, each containing three 8-bit intensity values ranging from 0 to 255. By combining intensities of red, green, and blue, nearly any color in the visible spectrum can be represented. It is also well suited to photographs with many colors. Unfortunately, this representation is costly; a 1024×1024 image, at 24 bits per pixel would consume 3.1 MB.

The YCbCr color space contains three channels, luminance (Y), and two complementary channels of chrominance (Blue-Yellow and Red-Green). Luminance when rendered is similar to a grayscale image, and contains most of the details we perceive in a photo, e.g., we perceive the changes in intensity and brightness. The chrominance channels contain the color information within an image, but comparatively less detail. Conversion between the RGB and YCbCr color spaces is lossless.

The human visual system (HVS) is much more sensitive to luminance than chrominance, which provides one basis for lossy compression in the YCbCr color space. The jpeg encoding procedure consists of five steps: (1) Conversion from RGB to YCbCr, (2) Chrominance subsampling, (3)

[1]In 2012, HTTPArchive changed the test methodology to wait until network activity ceased before finishing a page crawl, which results in more requests per page—however, accommodating lazy loading seems not to have altered the image trends that predate this change by 2 years.

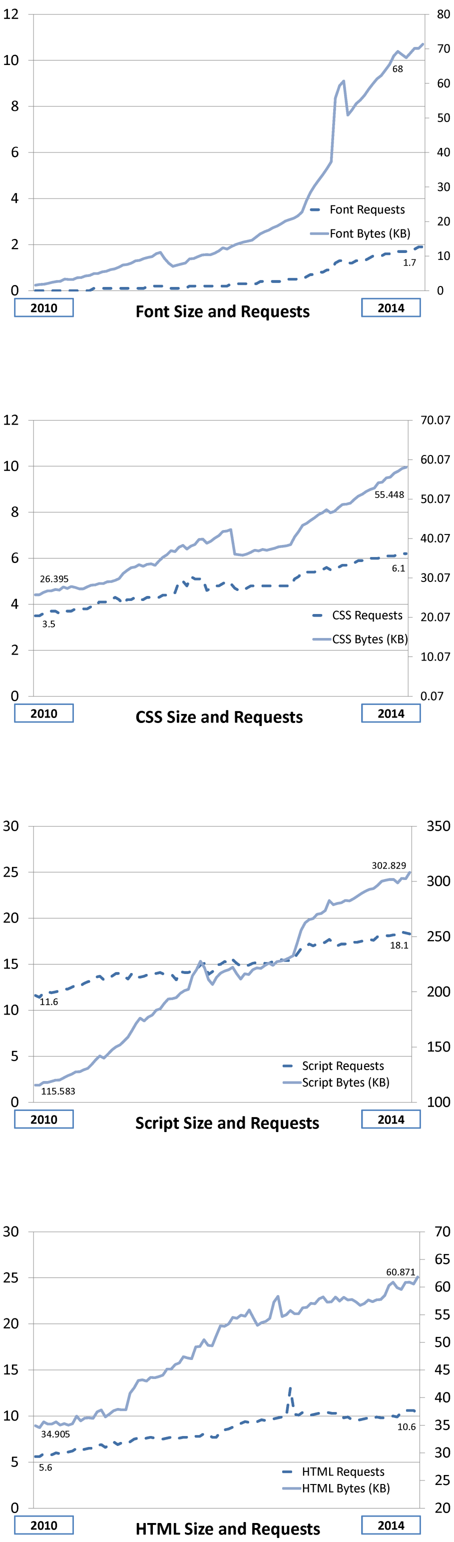

Figure 4: Overall landing page trends from the Alexa top 1 million for Fonts, CSS, JavaScript, and HTML show that web pages are getting larger and at a rapid pace.

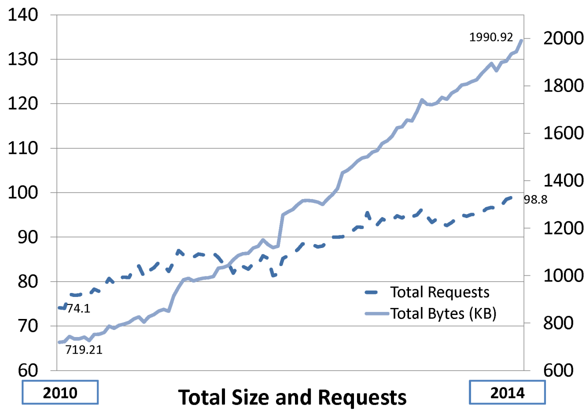

Figure 5: Aggregate web page size has increased considerably in four years.

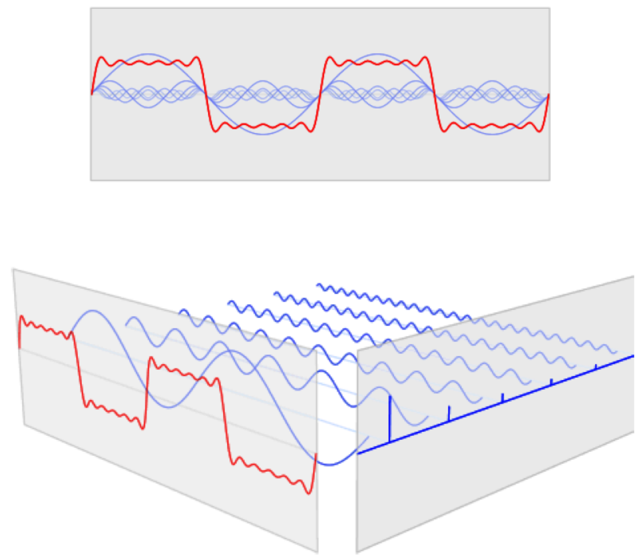

Figure 6: Discrete Cosine Transform (DCT) used in JPEG encoding. Lower frequency (large) waves are more important to the overall shape of the signal.

Discrete cosine transform (DCT), (4) Quantization, and (5) Entropy coding.

In chrominance subsampling, up to 75% of color information is discarded from the two color channels, for example, by replacing blocks of four pixels with a single averaged chrominance. The luminance channel is preserved. The remaining steps are performed individually for each channel.

The discrete cosine transform (DCT) converts a signal from the time domain into the frequency domain, by expressing a signal as a superposition of fixed frequency cosine waves with varying amplitudes. We can express a wave by storing the amplitudes within the frequency domain. Each image channel is partitioned (spatially) into 8x8 blocks (matrices) of floating point values. The DCT works by assigning a coefficient (amplitude) to each wave. Thus, given an input signal vector of 8 values in the time domain, DCT will output a vector of 8 amplitudes corresponding to 8 different frequencies. Jpeg encoding uses a 2-dimensional DCT, and can express any 8x8 block as a superposition of fixed 8x8 basis images. The choice of 8x8 blocks allows for a number of computational efficiencies but larger block sizes have recently been examined in literature [18].

An important observation is that the lower frequency (large) waves are the most critical to the overall shape of a given signal, as shown in Figure 6, while high frequency waves add details that may not be noticed or missed if discarded. Thus, jpeg encoding works by assigning a greater emphasis on preserving the coefficients for low frequency basis functions (large blocks), and selectively discarding coefficients for high frequency basis functions (detailed information).

To recap, each 8x8 block of pixels is expressed as an 8x8 matrix of coefficients, usually ranging from -1024 to 1024. This independent partitioning into blocks is one reason for square artifacts observed in low fidelity jpeg images. The next step is quantization, which is critical to the compression ratios achieved by lossy image formats.

The jpeg standard provides for a quantization table, an 8x8 matrix of divisors used to divide the blocks of coefficients obtained from the DCT step. The quantization matrix contains smaller values in the upper left corner, with progressively larger values toward the lower right quadrant. When the matrices are divided piecewise, the larger quantization divisors cause many of the DCT coefficients to become 0, effectively discarding the coefficients pertaining to high frequency information.

The coefficients are then collected from the quantized matrix in a zig-zag pattern, starting with the upper left corner and following the diagonals thereby placing most of the zeros at the end. The long strings of 0's are compressed using run-length encoding, followed by entropy coding using Huffman tables.

Importantly, the jpeg standard recommends a stock set of quantization tables determined experimentally through psychovisual tests. Choosing the optimal quantization table is a hard problem, and is often image-dependent. Different tables are suggested for luminance and chrominance channels. In practice, many software packages and cameras use customized tables—a detailed source of several hundred quantization tables can be found in [21].

4.1 Quality Factor

The quality parameter (often ranging from 0-100) is used to scale the coefficients of the quantization matrix, and the scale is typically nonlinear. The Independent Jpeg Group (libjpeg) uses the following scaling, where Q is the quality parameter:

$$ScaleFactor = \begin{cases} \frac{5000}{Q} & Quality < 50 \\ \\ 200 - 2Q & Quality \geq 50 \end{cases}$$

Each element M_i of the quantization matrix is then scaled as $New_i = M_i \cdot ScaleFactor + 50$. Certain quality settings may trigger other optimizations. For instance, Adobe PhotoShop at its highest quality settings will skip chrominance subsampling, and `cjpeg` at 100 will generate a quantization matrix of all 1's. Nonetheless, commercial and open source

Figure 7: An image saved at 10% quality jpeg, note the blocking artifacts and loss of gradient information in the sky [31].

software based on `libjpeg` will often use the default quantization settings and scaling described above. Web designers are often reluctant to save images below a quality factor of 85, and many routinely save at a (misguided) quality setting of 99. The recommended setting embedded in vBulletin uses an aggressive quality factor of 75 for uploaded images. Each time a jpeg is saved, its quality becomes more degraded. Determining the original quantization matrix is an active area of forensic study, which seeks to determine if a photo has been tampered with or in some cases to recover the quantization signature from the device with which the image was captured. Combining two images saved at different quality settings will often contain detectable anomalies [19, 22].

Recompression. Especially for high resolution photo archives placed on the web, recent work has focused on improving compression within the domain of existing image tools. Shoham et al. use experimentally acquired distributions to model and simulate the human visual system (HVS), autotuning the jpeg quality factor down to the psychovisual threshold [30]. This threshold varies depending on image features present in the analyzed image, and has been incorporated into a commercial tool, JPEGMini. For average observers, the processed images are perceptually unchanged. Otherwise, finding a low, perceptually lossless quality factor is a manual trial-and-error process, and for this reason web designers avoid it.

Figure 8: An image saved at 10% quality, in jpeg (left) and webP (right). WebP applies more smoothing yet maintains more consistent color (best judged at increased zoom levels) [31].

5. SERF IMAGES

The psychovisual threshold is the lowest quality setting whereby the original and compressed images are perceptually invariant. After using selected algorithmic enhancements, how much can we reduce the *effective psychovisual threshold*? In this section, we introduce SERF, a technique for Scaling with Enhanced Reduced Fidelity images.

Figure 7 shows two jpeg images saved at a quality factor of 10, using `libjpeg-turbo`. The loss of high frequency information in the trees is difficult to perceive, but the low frequency sky has lost the gradient and has introduced many 8x8 blocking artifacts. In general, the baseline jpeg codec does a good job of maintaining texture in high frequency areas, such as fur, but suffers from artifacts, loss of color information, loss of gradients, and blocking around edges.

5.1 Google WebP Codec

WebP provides a similar quality (quantization) setting as jpeg, and is a format designed for uploading true color photographs on the web with a focus on performance. Adopted by Opera and Google, and currently being piloted by Facebook, support for the WebP format is growing. The WebP format employs block prediction, and noticeably employs more smoothing at lower quality factors, as shown in Figure 8. While the coloration is better preserved in the webp image, the loss of texture in the fur is obvious; yet at a quality factor of 10 it is clearly preferable to baseline jpeg. We adopt webp as the lossy format for our study due to its imminence and potential, and without loss of generality. We employ it with default parameters, though there are dozens of parameters available to control the strength of the various algorithms employed, including smoothing, sharpness, and targets for peak signal-to-noise ratio (PSNR) which can be readily explored in future works.

5.2 Psychovisual Enhancements

The two most well-known psychovisual enhancements made to photographs are adjustments to brightness and contrast. Increasing the brightness linearly raises the intensity of each pixel value by a fixed amount, which can cause clipping in

Figure 9: Original (left), and enhanced (right). Increasing vibrance and enhancing shadows are two effective psychovisual enhancements for socially-sourced images [31].

areas that are already bright, and cause darker areas (e.g., blacks) to become lighter gray. Similarly, increasing the contrast raises the intensity on a biased scale, such that darker colors are increased by a smaller amount (or not at all), while higher intensity values are increased by a greater amount, widening the range between darker and lighter colors.

Empirically noting that socially-sourced images most often suffer from either poor exposure or poor color balance, we focus our research efforts on two psychovisual enhancements directly related to brightness, contrast, and color. Further studying other combinations of enhancements and validating their efficacy are subjects of future work. For our purposes, we explore two psychovisual features which work jointly to (1) yield a strong effect when modified, (2) be identifiable via machine learning, and (3) satisfy the constraint of being algorithmically efficient: Saturation and Shadows.

5.2.1 Enhancing Saturation

Saturation is a parameter representing the overall intensity of colors within a scene, analogous to the brightness of the chrominance channels. In the same way that contrast is a biased approach to increasing pixel intensity (brightness), we use a scaled approach to increasing saturation commonly called vibrance. Increasing the vibrance of a scene will boost desaturated colors more than colors at high intensity, to prevent color clipping at high values.

5.2.2 Enhancing Shadows

Many images contain luminance information which is obscured by relatively low exposure. Problems with underexposure can often be rectified by enhancing shadows or stretching midtones, or by increasing contrast and brightness, or by adjusting an image's gamma. Given the prevalence of poor lighting and its effect on consumer devices, many socially sourced images benefit from increasing their perceived brightness, lifting detail otherwise lost to dark shadows. In situations where an image is too dark, image data may be lost to the shadows entirely; however, there usually remains some image information which will still benefit from shadow enhancement. Figure 9 demonstrates two representative images captured with indoor and outdoor lighting which suffer from underexposure, to which we have applied psychovisual enhancements to vibrance and shadows. These enhanced images form the basis for our proposed performance optimization.

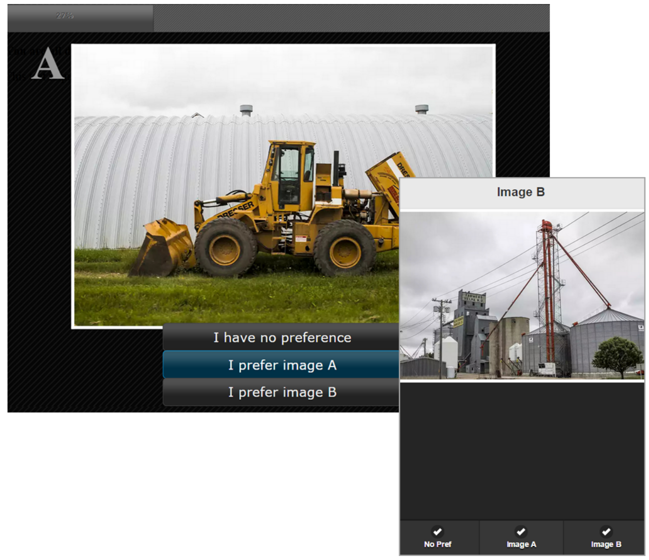

Figure 10: Double stimulus forced choice comparison employed in our user acceptance test interface. A desktop interface was implemented using jQuery-UI, and a mobile interface via jQuery-mobile. Clicking the image toggles the displayed choice.

5.3 Fidelity Reduction of Enhancements

Our ultimate goal is to improve the performance of the static image server by increasing acceptable compression. Recent work attempts to reduce image fidelity down to the psychovisual threshold, which is the level of quantization in an image where the human visual system begins to reject introduction of artifacts. Previously, this reduction is a manual effort requiring trial and error, leading most web operators to use generic quality settings in a safe range high above the limen.

We hypothesize that it is possible to reduce the psychovisual threshold by selectively introducing psychovisual enhancements. Thus, for all images in this study, we focus only on boosting saturation (vibrance) and lightening shadows—two enhancements that are readily available for implementation and amenable to automation with machine learning computer vision systems [25].

Simultaneous reduction in image fidelity and introduction of psychovisual enhancements frames a multi-objective tradeoff for a user: to what extent can the quality factor be acceptably decreased by first applying psychovisual enhancements? Put differently, does the average user prefer a high fidelity image of potentially poor psychovisual quality, or a psychovisually enhanced image degraded with compression artifacts?

Given the explosive growth of social networks, their reliance on lossy image compression, and recent studies into image popularity [4, 7, 14, 26], being able to further optimize and scale the static image server or static content distribution network (CDN) is of paramount interest to social

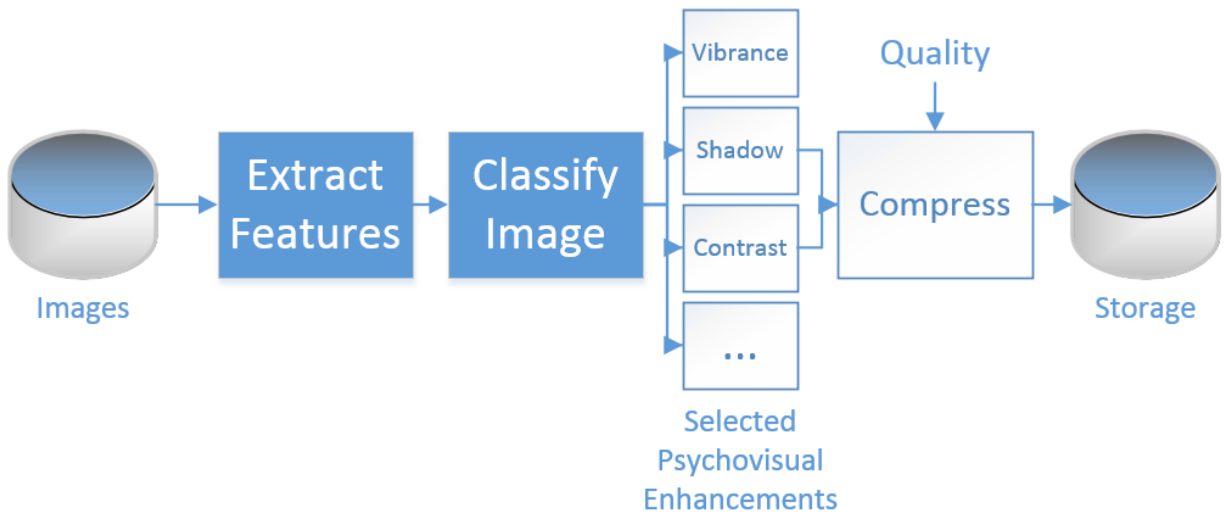

Figure 11: An overview of the proposed SERF pipeline.

networks.

Figure 11 presents an architectural overview for a SERF-enabled site. Stored images can be analyzed for simple computer vision features and classified. For instance, human faces may be enhanced differently to preserve skin tones. Depending on the classification, selected psychovisual enhancements are applied, such as saturation, brightness, contrast, vibrance, and midtone/shadow enhancement. The enhanced images are then compressed with a supplied target quality setting—below the psychovisual threshold—and stored. Situational optimization will determine whether a user's image request is served from the original archive or an enhanced archive [25]. Based on our evaluation below, we anticipate most users will have images served from the compressed storage, yielding a considerable gain for network operators.

6. EVALUATION

We evaluate the effectiveness of our approach through user acceptance testing on study participants (n=90), using the method of constant stimuli. In order to quantify users' preferences with the least amount of bias, our test subjects have no knowledge of the image modifications or treatments applied to our test images or any prior knowledge of the study. Treatments are applied randomly and with random intensity, and are displayed in random order. Study participants included ages from 19 to 74, with 75% in the 20 to 30 age group. No pre-screening of subjects was performed for amateur photographers or those with reduced visual acuity.

We adopt a modified double stimulus forced-choice comparison espoused in [28], as the method is both simple and has the least sample variance of four common image quality assessment techniques. Using a web browser, we design a system which allows a user to toggle between two images. One of the images is a high quality reference image (original), and the other image is a SERF version of the same image, which contains psychovisual enhancements yet is saved at one of six low quality factors. Since we use large images (as small images hide degradation in quality), two images cannot fit on the screen at the same time. Thus, only one image is displayed to a user at a time, but the user is able to toggle between the images freely before making their judgment. When the user toggles, the image quickly fades from one image to the other, which can help the user see artifacts introduced during the compression step.

To control for sampling bias, we randomize the order of the reference image and the SERF image, as shown in Figure 10, and allow the user to cycle between the two images "A" and "B" as many times as he or she pleases, with no restrictions on time. With the stimulus still visible, the user is then asked to indicate which image is preferable, or if no preference is discernible, the user can choose "No Preference". In cases where no image is preferable to the other, it is advantageous to choose the smaller of the two images in the interest of performance. We also allow users to conduct the test in a variety of realistic viewing conditions, rather than a specialized, darkened test lab. Mobile devices used in the test were all owned by participants in the user study, and included a mix of modern smart phones with HD displays but no tablets.

6.1 Source Images

Our data set consists of original full-color photographs which were losslessly saved in the PNG format. These images were scaled to a maximum dimension of 1000 pixels, and converted to WebP at a quality setting of 95, to be used as the unmodified reference images in our study. A second set of SERF images was created as well. These pictures were minimally enhanced by increasing vibrance and lightening shadows, and again saved losslessly. Any enhancements were performed on the whole histogram and not on individual areas of the image. Each of these psychovisually enhanced (SERF) images were then saved in WebP format at six low quality settings of 10, 20, 30, 40, 50 and 60. Picture subject matter was only lightly screened so as not to include any humans in the scenes in order to control for subliminal preferences and vagaries of skin tone adjustment, which could be studied in the future. The 15 images in the data set were sufficiently varied including a mix of close-up and distance subjects, indoor and outdoor scenes, and images

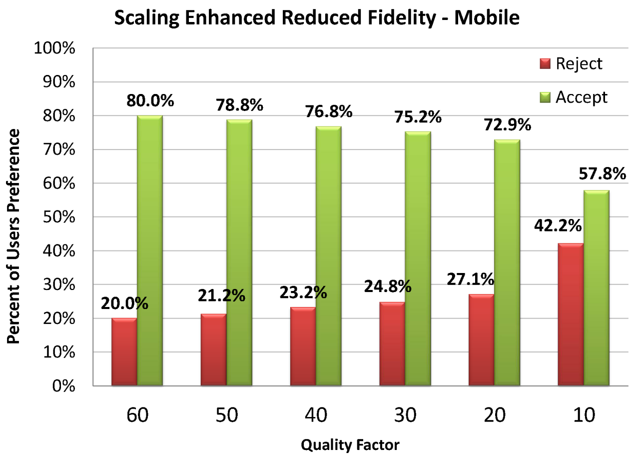

Figure 12: SERF images tested on smart phones. 7% difference between quality 60 and quality 20, and a remarkable 58% acceptance at quality 10.

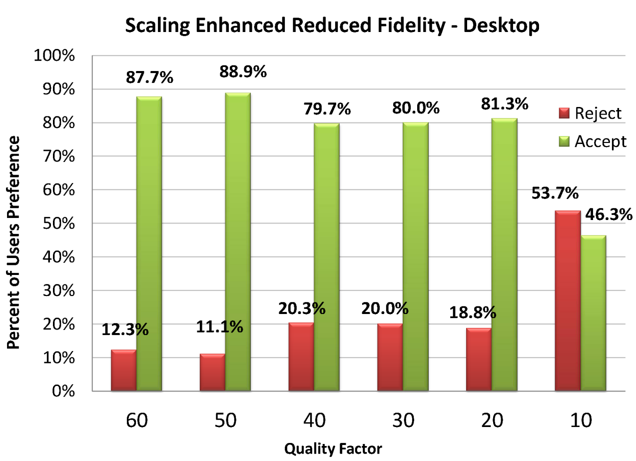

Figure 13: SERF images tested on desktops. 6.4% difference between quality 60 and quality 20.

in both portrait and landscape modes. Each image contained at least one maximal horizontal or vertical dimension of 1000 pixels. To ensure all mobile user test participants could take the test without a WebP plugin, we converted all of the WebP images back to lossless PNG, maintaining the blurring, artifacts, and color degradation introduced during WebP compression. In practice, this won't be necessary as more mobile devices support this format. Our full set of test images, originals and enhancements are available online and are shown in Figure 1. All enhancements were limited to the algorithmically simple, reproducible enhancements described above, in lieu of any localized or artistic manipulations. The reference images were also saved at six reduced fidelity settings for a baseline test.

6.2 Results

Figures 12 and 13 show our experimental user acceptance results on the mobile devices and desktop machines, respectively. Remarkably, for mobile devices, nearly 58% of users accepted the psychovisual enhancements even when fidelity was degraded to a quality factor of 10! At a quality of 60,

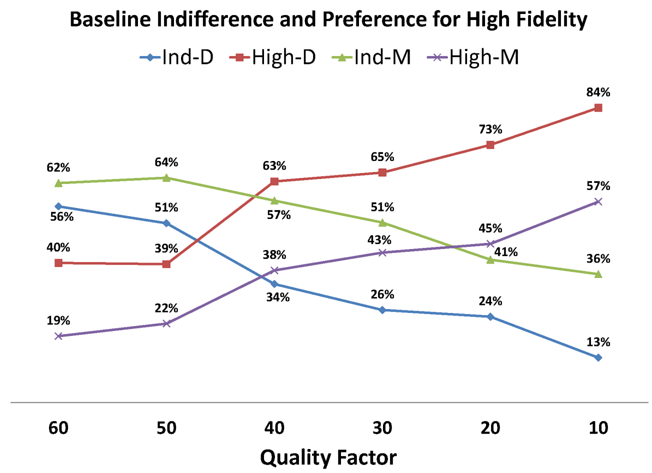

Figure 14: Test with unenhanced images. Users are more indifferent to image fidelity at higher quality factors. As fidelity decreases, users indicate a preference for the higher fidelity image when given the choice.

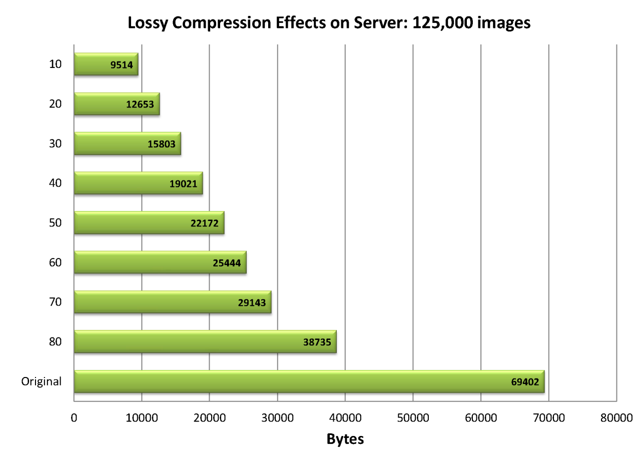

Figure 15: Impact study: conversion of 125,000 random images from an online community operated by the author.

80% of mobile users and 87% of desktop users preferred the psychovisual enhancements. As expected, user acceptance and quality factors are negatively correlated. However, for both desktop and mobile devices, there is only a 7% difference in acceptance between a quality factor of 60 and a quality factor of 20, indicating a strong preference for psychovisually enhanced images, even when those images are degraded by saving at a very low quality setting. Thus, subtle psychovisual enhancements serve as an enabler for increased or more aggressive lossy image compression. We attribute the more pronounced drop off at 10% due to the non-linearity of the quality factor on the degree of quantization.

In our experiments, subjects compared high-fidelity unenhanced reference images to low-fidelity enhanced images, mediated by compression strength. We are successful if the images are not rejected due to introduced compression arti-

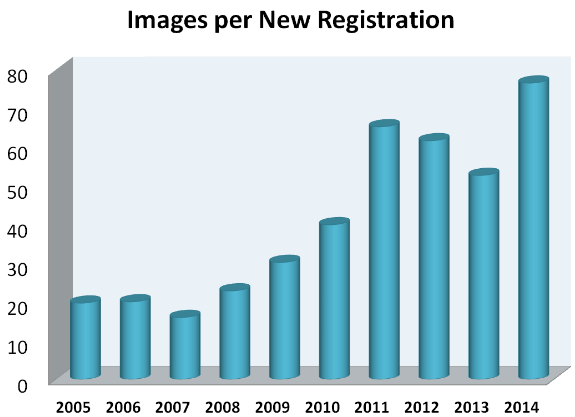

Figure 16: Rate of uploaded images has increased significantly on an online forum operated by the author (93K users), a global trend attributed to the convenience of mobile devices and the availability of embedded cameras.

facts. By performing the test at multiple quality levels, we observe that compression artifacts become more significant at lower quality. Thus, we also conducted two baseline experiments on the original (unenhanced) images, seeking an overall positive correlation between the level of indifference to image fidelity and quality factor. The purpose is to validate that the following assumptions hold for our test set: as quality levels decrease, subjects will express a) increasingly stronger preferences for higher fidelity images, and b) decreasing levels of indifference to the two images. Figure 14 shows the percent of unenhanced stimuli at each quality level which users were indifferent to reduced fidelity (Ind), or exhibited a preference for the high fidelity image (High), for both mobile (M) and desktop (D) platforms. These trends validate our assumption that users are sensitive to image fidelity without psychovisual enhancements, displaying a preference for the less compressed images. Overall, while users are sensitive to artifacts introduced during compression, psychovisual enhancements provide a powerful countereffect to temper these baseline trends.

To put our results in perspective, we also conduct a measurement study with 125,000 jpeg images randomly sourced from a site the author operates. Figure 15 shows the impact of conversion to WebP for quality factors in the "safe" range ($q \geq 80$), followed by conversion with the lower quality factors targeted by our study. The average uploaded image size is just under 70K. Quality factors of uploaded jpeg images cannot reliably be determined as they are not stored with the files, but empirically the quality of images on the site is high. Conversion at our highest levels of acceptability ($q = 60$) represents a bandwidth savings of more than a factor of two, to an average file size of just under 26K. A dedicated, static image server could potentially double or triple its capacity with this technique.

7. DISCUSSION

Online communities and social networks are the most popular sites on the Internet, and in the last decade have exploded with multimedia content. Facebook.com, the second busiest site in terms of network traffic, currently has 1.28 billion monthly active users, and 609 million mobile daily active users [8]. Internet Brands, owner of the vBulletin forum software and a number of online companies, serves 100 million users daily with its products [2].

Uploading images, photos, memes, and videos to supplement concepts ranging from exciting to mundane is now typical on many online social networks. In 2009, Facebook had accumulated over 260 billion images [12], and in September 2013 reported an average of 350 million photos uploaded daily [1]. A 2014 analysis of 70K pages suggests that photos compose more than half of daily new posts on the social platform [10]. Figure 16 shows an increase in the rate of new photos uploaded to an online forum operated by the author, a global trend exacerbated by the convenience of mobile embedded cameras and a culture of sharing.

Most web designers recognize that static site images can be saved with lower fidelity by employing lossy compression to reduce bandwidth consumption and decrease loading times, though many are reluctant to do so for aesthetic reasons. Socially sourced images, such as those dynamically uploaded to online communities, news sites, and social networks are rarely under a site operator's direct inspection or control. These images are often produced using mobile devices, with very little enhancement, and often captured in less than optimal lighting conditions. As a result, they tend to be of an observably lower quality than professional media. However, we assert that this provides a new avenue for study and optimization.

Further optimizing the static image server or CDN is an increasingly difficult problem, as the emergence of social media and mobile devices as a dominant platform provides an ever-changing engineering domain. The focus is jointly on quality of experience (QoE) for the client and bandwidth management for the data center. Improvements have been made in caching and locality prediction, as well as operating system improvements. Other efforts have been placed on identifying new image formats and compression techniques, which are promising but currently hindered by their lack of adoption in consumer devices [11]. We propose an optimization which can complement this arsenal of research advances within the constraints of existing infrastructure and image formats.

The cache potential of socially-sourced images is often short as well; they are viewed many times shortly after uploading but are then relegated to the long tail of infrequently accessed images [12]. Fortunately, lossy compression can have a significant positive impact with even a cold cache by reducing the file size. For instance, Facebook will reduce the fidelity of jpeg images upon resizing or when greater than 100K [9]. For large photo archives, recent work has focused on the technique of recompression, increasing compression until artifacts become human perceivable [15, 30].

The performance community has recently advanced the idea of situational performance optimization [29] as a necessary paradigm going forward to cope with variable user preferences in the web landscape. Only a decade ago, web technologies were much more limited, and aside from browser incompatibilities and catering to broadband vs. dial-up, fewer engineering trade-offs were necessary. Today, the landscape also includes dozens of mobile devices, browsers and frameworks, substantial client libraries, apps and web services, variable client networks, and locality concerns, creating a smorgasbord of potential conflicts. Additionally, social media's dominance creates an impetus for optimization at a much grander scale. Of course, situational optimization does come at additional cost, and relies on collecting metrics within every system and the willingness to dynamically classify each user, choosing the appropriate optimizations.

Facebook has been criticized by high-end photographers for compressing their images to a lower fidelity; the provision of the "High Resolution" check box provided to the user when uploading photos is an example of situational optimization. Google Plus was similarly criticized for auto-enhancing images. However, for the data center the performance mantra is to optimize the critical path for the average transaction. If applicable, we suggest that networks enable users to turn off optimizations—as our results show there will likely be enough users on average who either prefer or are indifferent to optimizations which can have a great impact on a data center. User QoE is sometimes at odds with the need to reduce bandwidth, processing, and storage requirements in large data centers. These results show promise: audiophiles did not hinder the progress of the lossy MP3, nor did photographers completely hinder the spread of jpeg as the dominant format on the web.

7.1 Future Directions

The SERF technique can be extended to identify and accomodate more sophisticated psychovisual enhancements based on additional popularity metrics, such as lightness, darkness, texture, and dominant color. Beyond psychovisual enhancements, there may be classes of images which compress better than others based on their content, and recent deep learning and neural networks show promise in automating the identification of these classes at scale. Though we chose a wide-range of subjects in our test set, identifying if there are specific classes of images that benefit most from our technique is an open problem. Similarly, it will be interesting to quantify the strength of any psychovisual enhancement in isolation and study its effects for a particular image. We suspect a diminishing return—too much contrast or saturation may counter its positive effects. Finding this performance knee and identifying optimal parameters is an open problem. Further, we intend to extend our work to include photos of people, which we suspect will require a different approach to image enhancements. We intend to evaluate a prototype on a community of 90K+ users to gather large scale feedback. In addition, we will prototype a machine learning system for fine-grained tuning of fidelity parameters, image classification, and user classification for situational optimization.

8. CONCLUSION

We introduced SERF, a technique for enabling data centers to scale their static image servers by compressing images below the psychovisual threshold. The introduction of psychovisual enhancements allows users to accept a perceived reduction in fidelity offset by a stronger perceived increase in quality, where the brightness and shadow enhancement outweigh the degradation in quality introduced by extreme lossy compression. We also demonstrate the potential impact of our approach by showing that even at the highest quality levels considered by our study, we can still achieve a factor of two savings in average bandwidth. Through user acceptance testing on an array of images, our results show that image fidelity can be reduced when psychovisual enhancements are applied. Remarkably, there is only a 7% difference between a quality of 60 and a quality of 20 for both mobile and desktop devices. This opens a new, exciting avenue for multimedia performance optimization through social and psychovisual research.

Acknowledgments

We would like to thank Prof. Alan Skees, MFA, for contributing the images used in our study. We would also like to thank the anonymous reviewers for their insightful comments. This work was partially supported by the ARO Grant W911NF-15-1-0287. Opinions, findings, conclusions and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agency.

9. REFERENCES

[1] Facebook. www.facebook.com.
[2] Internet brands: vbulletin forum software. www.vbulletin.com.
[3] jquery, the write less, do more, javascript library. www.jquery.com.
[4] Curalate insights: Analysis of instagram. http://blog.curalate.com/image-qualities-that-drive-likes-on-instagram, Nov. 2013.
[5] Wikipedia: Aol art image format. http://en.wikipedia.org/wiki/ART_image_file_format, Dec. 2013.
[6] Alexa top sites. http://www.alexa.com/topsites, 2014.
[7] Curalate insights: Analysis of pinterest. http://blog.curalate.com/the-perfect-pinterest-pin how-to-optimize-your-images-for-maximum engagement-on-pinterest, Jun. 2014.
[8] Facebook newsroom. http://newsroom.fb.com/company-info, May 2014.
[9] Facebook support: How can i make sure my photos display in the highest quality? https://www.facebook.com/help/www/266520536764594, Nov. 2014.
[10] Quintly analytics: Which facebook post type rules the news feed? http://www.quintly.com/blog/2014/03/data-analysis-which-facebook-post-type-rules-the-news-feed/, Mar. 2014.
[11] Webp: A new image format for the web. https://developers.google.com/speed/webp, March 2014.
[12] D. Beaver, S. Kumar, H. C. Li, J. Sobel, and P. Vajgel. Finding a needle in haystack: Facebook's photo storage. In *USENIX OSDI*, 2010.
[13] M. Brittain. Batch processing millions of images at etsy. https://codeascraft.com/2010/07/09/batch-processing-millions-of-images, Jul. 2010.
[14] J. Cheng, L. A. Adamic, P. A. Dow, J. M. Kleinberg, and J. Leskovec. Can cascades be predicted? In *23rd International World Wide Web Conference, WWW '14, Seoul, Republic of Korea, April 7-11, 2014*, pages 925–936, 2014.
[15] S. Ding, F. Huang, Z. Xie, Y. Wu, and L. Ma. A novel customized recompression framework for massive internet images. In *Proceedings of the First Int'l Conf. on Computational Visual Media*, CVM'12, pages 9–16, Berlin, Heidelberg, 2012. Springer-Verlag.

[16] eMarketer Inc. emarketer: Facebook us ad revenues. http://www.emarketer.com/Article/Facebooks-US-Ad-Revenues-Outpace-Users-Average-Daily-Time-Spent-on-Site/1011215, Sep. 2014.

[17] J. Erman, V. Gopalakrishnan, R. Jana, and K. K. Ramakrishnan. Towards a spdy'ier mobile web? In *Proceedings of the Ninth ACM Conference on Emerging Networking Experiments and Technologies*, CoNEXT '13, 2013.

[18] F. Ernawan, N. A. Abu, and N. Suryana. A psychovisual threshold for generating quantization process in tchebichef moment image compression. *JCP*, 9(3):702–710, 2014.

[19] H. Farid. Exposing digital forgeries from jpeg ghosts. *IEEE Transactions on Information Forensics and Security*, 4(1), Mar. 2009.

[20] I. Google. Google auto enhance. http://gplusproject.appspot.com/auto-enhance/index.html, May 2013.

[21] C. Hass. Jpeg quantization tables. http://www.impulseadventure.com/photo/jpeg-quantization.html, 2008.

[22] C. Hass. Jpegsnoop - file decoding utility. http://www.impulseadventure.com/photo/jpeg-snoop.html, 2014.

[23] HTTPArchive. Tracking how the web is built. http://httparchive.org, Oct. 2014.

[24] C. Huang, A. Wang, J. Li, and K. W. Ross. Measuring and evaluating large-scale cdns paper. In *Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement*, IMC '08, pages 15–29. ACM, 2008.

[25] S. B. Kang, A. Kapoor, and D. Lischinski. Personalization of image enhancement. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1799–1806, June 2010.

[26] A. Khosla, A. D. Sarma, and R. Hamid. What makes an image popular? In *23rd International World Wide Web Conference, WWW '14, Seoul, Republic of Korea, April 7-11, 2014*, pages 867–876, 2014.

[27] M.-C. Lee and S.-B. Cho. Interactive differential evolution for image enhancement application in smart phone. In *IEEE Congress on Evolutionary Computation*, pages 1–6, 2012.

[28] R. K. Mantiuk, A. Tomaszewska, and R. Mantiuk. Comparison of four subjective methods for image quality assessment. *Comp. Graph. Forum*, 31(8), Dec. 2012.

[29] G. Podjarny. Situational performance optimization - the next frontier. http://calendar.perfplanet.com/2012/situational-performance-optimization-the-next-frontier/, Dec. 2012.

[30] T. Shoham, D. Gill, and S. Carmel. A novel perceptual image quality measure for block based image compression. In *Proceedings of SPIE*, Image Quality and System Performance VIII. SPIE, 2011.

[31] A. Skees. Vacation photos. Private collection, ©2014 Alan Skees, MFA.

[32] S. Stefanov. *Web Performance Daybook*. Oreilly and Associate Series. O'Reilly Media, Incorporated, 2012.

[33] S. Traverso, K. Huguenin, I. Trestian, V. Erramilli, N. Laoutaris, and K. Papagiannaki. Tailgate: handling long-tail content with a little help from friends. In *Proceedings of the 21st World Wide Web Conference 2012, WWW 2012, Lyon, France, April 16-20, 2012*, pages 151–160, 2012.

[34] X. S. Wang, A. Balasubramanian, A. Krishnamurthy, and D. Wetherall. How speedy is spdy? In *Proceedings of the 11th USENIX Symposium on Networked Systems Design and Implementation, NSDI 2014, Seattle, WA, USA, April 2-4, 2014*, pages 387–399, 2014.

[35] M. P. Wittie, V. Pejovic, L. Deek, K. C. Almeroth, and B. Y. Zhao. Exploiting locality of interest in online social networks. In *CoNEXT*, page 25. ACM, 2010.