Low-level pixelated representations suffice for aesthetically pleasing contrast adjustment in photographs

Tandra Ghose¹, Yannik T. H. Schelske¹, Takeshi Suzuki², and Andreas Dengel³

¹Department of Psychology, University of Kaiserslautern, Germany
²Ricoh Institute of Technology, Japan
³Knowledge Management Department, DFKI Kaiserslautern, Germany

Today’s web-based automatic image enhancement algorithms decide to apply an enhancement operation by searching for “similar” images in an online database of images and then applying the same level of enhancement as the image in the database. Two key bottlenecks in these systems are the storage cost for images and the cost of the search. Based on the principles of computational aesthetics, we consider storing task-relevant aesthetic summaries, a set of features which are sufficient to predict the level at which an image enhancement operation should be performed, instead of the entire image. The empirical question, then, is to ensure that the reduced representation indeed maintains enough information so that the resulting operation is perceived to be aesthetically pleasing to humans.

We focus on the contrast adjustment operation, an important image enhancement primitive. We empirically study the efficacy of storing a pixelated summary of the 16 most representative colors of an image and performing contrast adjustments on this representation. We tested two variants of the pixelated image: a “mid-level pixelized version” that retained spatial relationships and allowed for region segmentation and grouping as in the original image and a “low-level pixelized-random version” which only retained the colors by randomly shuffling the 50 x 50 pixels. In an empirical study on 25 human subjects, we demonstrate that the preferred contrast for the low-level pixelized-random image is comparable to the original image even though it retains very few bits and no semantic information, thereby making it ideal for image matching and retrieval for automated contrast editing. In addition, we use an eye tracking study to show that users focus only on a small central portion of the low-level image, thus improving the performance of image search over commonly used computer vision algorithms to determine interesting key points.

Keywords: Aesthetic judgment, computational aesthetics, contrast adjustment, photo editing, eye movements

Corresponding author: ghose@sowi.uni-kl.de

Acknowledgements. This work was funded by a Marie Curie grant (CIG#293901) from the European Union awarded to Tandra Ghose, and joint project for DKFI-RICOH (Project RSIP-2 (A12.384, Kst 15540) for the Knowledge Management Group at DFKI, Kaiserslautern, directed by Andreas Dengel).

We thank Rupak Majumdar in help with editing the manuscript.
Highlights:

- Aesthetic judgment of most preferred contrast for a given image
- Current scenario – proliferation of digital photography, auto image enhancers
- Bottleneck– storing, retrieving, matching high resolution images for editing
- Empirically test image reduction technique for contrast adjustment task
- Establish comparable preference for original and low-resolution versions

Digital photography on mobile devices, social media, and image-based advertising has exploded in the modern Internet. There are almost a trillion images shared on Web-based image, and the 2016 Mary Meeker’s Internet Trends report claims over 3 billion digital images are uploaded every single day, with growing relevance to ecommerce. A central problem in the digital image deluge is to improve the quality of an image automatically. Indeed, digital image enhancement has a wide diversity of applications, ranging from sharing “pretty” pictures on social media to selecting the best image for digital advertising, and there is a desire for finding a universal “magic button” that gives rise to the “Aha-experience” for an enhanced image.

Automatic image enhancement tools for digital photographs promise “one-click” image quality improvements. In recent years, with the proliferation of digital photography, auto image enhancers have also gained in popularity: a Google search for these terms returns over 720,000 results, with a multitude of commercial or online offerings. In a typical use of auto image enhance, a user uses the enhancer and then spends time in deciding whether to keep or discard the “enhanced” version. This judgment is based on whether the enhanced photograph is “aesthetically pleasing,” an informal but intuitive and automatic concept for human users.

Developing good auto-enhancers is a central problem of computational aesthetics, which aims to construct quantitative models for human-like aesthetic preference for photographs. Aesthetics of photographs results from a complex interplay of objective, computable, low-level image features and subjective, psychological, semantics and emotions (Datta, Joshi, Li, & Wang, 2006). While it may be possible to develop a comprehensive theory of aesthetics in the future, current practice in auto-enhancement is to retrieve enhancements from existing corpora of images.

Concretely, automatic image enhancement procedures have a set of image processing operations, such as contrast improvement or color saturation, and they decide whether to apply an operation and by how much by querying existing database of previously enhanced images. A naïve approach would query the database of existing images for other images that are “similar” to the current image and then apply similar amounts of enhancement on the new image. Unfortunately, this naïve approach does not scale because of high storage...
requirements to store entire images and because a pixel-level image proximity search is computationally expensive.

For faster image matching and retrieval, computer vision models strive to adopt modularity and hierarchical processing as reported in the biological vision literature (Latto, 1995; Smeulders, Worring, Santini, Gupta, & Jain, 2000; Zeki, 1999). Thus, for each operation, one can store an “aesthetic signature” relevant to that operation: a set of image features and the level of enhancement (Aydın, Smolic, & Gross, 2015). Given a new image, the auto-enhancer extracts the signature and searches for images in the database with similar signature. A good choice of the signature ensures that if an enhancement was made to a different image with a similar signature to enhance its quality, then applying the same enhancement to the current image will enhance its aesthetic quality as well.

A fundamental problem then is the empirical identification of features relevant for each image enhancement operation. The features should be sufficient to identify the level at which enhancement should be applied, while ensuring significant reductions in storage and search costs. Moreover, to date, semantics-based image retrieval algorithms do not work efficiently unless a human annotated ground truth label is available (Datta, Joshi, Li, & Wang, 2008; Deselaers, Keysers, & Ney, 2008). Therefore, empirically finding features that are independent of image semantics is important for faster and more efficient image retrieval, matching, and transformation.

This study

In this study, we empirically investigate the problem of identifying task relevant features for a basic but very important image enhancement procedure, contrast adjustment. Our choice of contrast adjustment is motivated by the central role of contrast in visual perception (Geisler, Albrecht, & Crane, 2007; Ramachandran & Hirstein, 1999).

Contrast is a low-level image feature, and it is reasonable to expect a semantics-independent feature representation that is both storage efficient as well as sufficient, i.e., searching for other images “similar” in these features should provide accurate estimates for the amount of adjustment. In our study, we chose an algorithm for image reduction that has been developed inspired by pixel art. In pixel art, artists manipulate the digital image at the level of pixels to create aesthetically pleasing compositions (Goldberg & Flegal, 1982).

We empirically explore the efficacy of pixel art as a representation of images for contrast adjustment. In our first study, we collect behavioral data (“the most aesthetically pleasing level of contrast for an image”) for the original image (the “high-level”) where all semantic information is preserved), a pixelated version of the original image (the “mid-level”) where the most representative colors and their spatial relationships are preserved, and a pixelated-random version (the “low-level”) that only preserves the most representative colors but not their
spatial relationships. We made the study manageable by preselecting four image categories that are commonly uploaded to social media: landscape-with-water, landscape-without-water, portrait (or a macro with face as focal object), and non-human macro. In short, our results show that the low-level image, which only preserves the most representative colors but not their positioning, is sufficient for contrast adjustment and, perhaps surprisingly, performs better than the mid-level representation.

Moreover, we measure associated eye movements made by expert and novice observers on original, mid-level, and low-level images in an attempt to understand how users extract relevant features for contrast adjustment. Our main finding is that the high-level and mid-level images show global fixation patterns as users scan the entire image for semantic information. In contrast, users primarily fixated on the center of the low-level image in order to make their contrast preference. Our findings suggest that it is sufficient to sample a small portion of the saved pixelated image for image retrieval, in contrast to extracting all key points, as is currently done in search algorithms based on computer vision algorithms.

In the following paragraphs we provide general background on computational aesthetics, justify our choice of image enhancement process and our image reduction algorithm, and describe our data collection method and main results.

### Computational aesthetics

The field of computational aesthetics draws inspiration from empirical aesthetics (Fechner, 1876) and neuroaesthetics theories (Zeki, 1999), which support a modular structure for aesthetic sense. Fechner (1876) laid the foundation for this new scientific approach to aesthetic science, i.e., the systematic search for stimulus properties that are associated with attractiveness and beauty. According to the modular structure, aesthetic sense corresponds to specialized brain mechanisms, or modules, tuned to analyze different aspects of visual processing (Zeki, 1999). The better a visual feature resonates with the visual processing mechanism, the stronger the aesthetic response (Latto, 1995). Datta et al. (2006) provide a comprehensive overview of computational aesthetics for photographs. They explain how this field aims to identify patterns in photographic aesthetics, for example, wide appreciation of natural scenery, preference for saturation of primary colors —red (fruits, flowers, sunset), blue (sky, water), and green (grass, foliage)— and certain compositional rules, e.g., the rule of thirds. They provide a computational framework that predicts aesthetic or emotional response to a photograph based on various feature representations. Commonly used features include low level pixel-based features, such as color, contrast, and saliency map, as well as high level rules, semantics, and segmentation based features, e.g., rules for incorporating the rule-of-thirds and the golden-ratio (Bhattacharya, Sukthankar, & Shah, 2010), for training different types of machine learning algorithms. They acknowledge
the presence of two problems. Firstly, there is an “aesthetic gap”, that is, lack of coincidence between the features extracted from low-level visual data (the pixels in the image) and the aesthetic response aroused in a given observer in a particular situation. Secondly, there is the issue of efficient retrieval of images of comparable aesthetic rating. Finally, they provide feature plots of aesthetic ratings of more than 50,000 images from Web-based systems that allow for aesthetic rating of photographs, namely, Photo.net, DPChallenge, Terragalleria and ALIPR (Datta et al., 2006).

Other researchers have attempted to find “aesthetic signature” of images (Aydın et al., 2015) by empirically collecting subjective ratings from observers and then carefully defining aesthetic attributes such as sharpness, depth, clarity, tone, colorfulness, etc. for training their software. Li, Loui, and Chen (2010) proposed an automatic aesthetic photo quality assessment system for consumer photos with faces. Ke, Tang, and Jing (2006) captured perceptual differences in professional and non-professional photos and designed a system called ACQUINE to automatically return a photo’s aesthetic quality. They trained their software on the difference of features in professional and non-professional “snapshots” photographs, based on spatial distribution of edges (edges in professional photos are clustered near the center of the image where the focal object is usually located), color distribution (color palette is chosen differently by professionals and non-professionals), hue (professional photos look more colorful and vibrant but surprisingly have less number of unique hues) and two low-level features, contrast and brightness (professional photos have higher contrast than snapshots).

**Importance of contrast adjustment in photo editing**

Based on visual neuroscience, contrast appears to be a particularly important component of visual perception. Contrast is encoded by most, if not all, of the cells in the primary visual cortex (Geisler et al., 2007). Contrast has been proposed as one of the eight “laws” in the neurological theory of artistic experience (Ramachandran & Hirstein, 1999). They speculate extracting contrast leads to discarding redundant information and helps allocate attention to information rich regions of changes in luminance and edges or “regions of interest” for the neurons, and that there is an evolutionary advantage if the whole organism finds the same feature interesting or pleasing. They use the prevalence of camouflage devices in both prey and predators to illustrate the natural importance of the attention-grabbing effect of contrast.

High figure-ground contrast affects perceptual fluency or the ease of processing image-based properties, and images with high figure-ground contrast evoke more positive aesthetic ratings (Reber & Schwarz, 2002). The thriving cosmetic industry supports the hypothesis that there is an aesthetic appeal in facial contrast (Jones, Russell, & Ward, 2015). Facial contrast is a characteristic pattern of darker features and lighter skin (Sinha, 2002) and luminance difference
between facial features (Russell, 2009). Facial contrast is positively correlated with attractiveness rating of female faces and negatively with male faces (Nestor & Tarr, 2008). Facial contrast decreases with age and plays an important role in age perception (Porcheron, Mauger, & Russell, 2013).

Global image corrections, where the brightness and contrast of the photo is altered, have been important since the beginning of photography. A technique called the cyanotype, developed in the 1840s, increased the inherent contrast of an image by using hydrogen peroxide. Modern black and white printing paper is manufactured to adjust the contrast of negatives that may not have desired range of contrast (Rand, Broughton, & Quinenz-Fiedler, 2011).

Snapshots from point-and-shoot cameras suffer from underexposure or overexposure resulting in photos that are either too dark or too bright. This issue can be addressed by correcting the lighting globally for all pixels in the image. Simple contrast adjustment can greatly enhance the aesthetic appeal of an image by modifying the tones (range of the darkest to the lightest part of the image). It can either increase contrast by spreading the tones to either extremity (dark/light) or decrease contrast by moving more of the tones to the center. While improving contrast can make the image “pop” and make colors appear saturated, too much contrast makes the photo look unrealistic. Adobe Photoshop® reports giving amateur users a slider to adjust brightness and contrast often led to undesirable outputs, therefore they added an “intelligent” backend in their product Adobe Photoshop CS3.¹

It would be ideal if there is a knowledge base available for amateur users for quickly retrieving “similar images” and the associated contrast correction used by an expert photo-editor in order to make a recommendation of range of correction values to the amateur user or automatically enhance the given image for them. Here we report a study where contrast adjustment was done by expert and amateur users on high-resolution original image and low-resolution versions. We empirically propose an image reduction technique that resulted in preferred-contrast similar to the original image. Thus, we claim that the reduction technique preserves task-relevant features while reducing the computational load of image retrieval and similarity-match calculation.

**Justification for recording eye-tracking**

As mentioned above, aesthetic experience depends on low-level image based features and high-level influence from semantics, compositional rules, expectations, expertise level and subjective state of the observer. Eye movement is an ideal metric for aesthetic experience because it is influenced by both low-level and high-level features, as summarized below, and provides an insight to the discrete sampling process of features in space and time of observation that finally leads to an overall perception of an image.

Eye movement metrics, such as, saccades, fixations, scanpaths, dwell time, and heatmaps have been used to illustrate the relationship between artistic images and observer’s aesthetic response to it (Buswell, 1935; Locher, 2006; Locher, 2017; Rand, Broughton, & Quinenz-Fiedler, 2011).

¹ http://www.photoshopessentials.com/photo-editing/brightness-contrast/
Krupinski, Mello-Thoms, & Nodine, 2007; Massaro et al., 2012; Tatler, Wade, & Kaulard, 2007). Although there are individual differences in viewing pictures, there are enough similarities across observers. For example, in a given image, there are regions with higher density of fixations when fixations are pooled across all observers indicating the presence of some information-rich feature (Babcock, Lipps, & Pelz, 2002). Eye movements during aesthetic judgments of complex scenes and art are driven by saliency maps of low-level features (Itti & Koch, 2000), object based saliency (Einhäuser, Spain, & Perona, 2008), and top-down effects driven by semantics (Henderson, Weeks, & Hollingworth, 1999). A two-stage model has been proposed to describe the relationship between eye movements for exploring pictorial art and visual aesthetics (Locher et al., 2007). The first stage consists of rapid global exploration of the image to get the holistic impression consisting of the semantic and structural components or the “gist” of the image. It is followed the second stage in which local exploration of interesting pictorial features is done to generate aesthetic appreciation of the image or satisfy “cognitive curiosity” which we interpret as the process of collecting information required to perform a given task. It has been documented that there are task-related patterns of eye movement while viewing pictures (Buswell, 1935; Yarbus, 1967), art (Locher et al., 2007), and complex real-world scenes (Henderson, 2003).

It has been proposed that observers dwell longer on a stimulus that they find aesthetically pleasing. This hypothesis has been supported by longer gaze duration at the image that is chosen as more aesthetically pleasing in a 2AFC task for judging facial attractiveness (Shimojo, Simion, Shimojo, & Scheier, 2003) and in a 8AFC task for judging grayscale photographic art (Glaholt & Reingold, 2009). Holmes and Zanker (2012) investigated the task of choosing the most-preferred image in 2AFC, 4AFC, and 8AFC tasks for four image categories: objects, buildings, commercial products and shapes. They propose a metrics based on accumulated fixation duration scaled according to sustained interest as the “oculomotor signature” of aesthetic preference for photographs.

Eye movement analysis has also been used to investigate sentiment attached to images using adjective noun pairs (Al-Naser, Chanijani, Bukhari, Borth, & Dengel, 2015). They used eye tracking to infer the how attention is deployed to image regions during assessment of adjective noun pairs and difference between fixation patterns for objective vs. subjective and local vs. global labels.

Moreover, eye movement studies have consistently shown that expertise and level of sophistication of the observer for appreciating art affects the eye movement pattern while viewing and rating artwork (Nodine, Locher, & Krupinski, 1993; Zangemeister, Sherman, & Stark, 1995). These studies report that experts engage more in global scanning with greater amplitude and duration of fixation, while novices prefer to dwell longer on local aspects.

**Image reduction technique**

Our image reduction technique is based on pixel art. It is inspired by the aesthetics of counted thread embroidery, cross-stitch embroidery, mosaic
and bead arts, and Cubism that feature flattened perspective and emphasize experience from the Gestalt of small regions filled with intense colors. Pixel art has been extensively used for enhancing the clarity and usability of interfaces in the video gaming industry (Goldberg, & Flegal, 1982). Creating pixel art from a high-resolution image requires a tedious process where the artist selects “similar” pixels by hand and carefully chooses colors that would be appropriate for filling the mosaic so that the final image best depicts the subject of the original image. At the algorithmic level, the challenge lies in optimally segmenting the image for “similar” regions such that meaningful features are grouped together (e.g., eyes, nose, or mouth in a portrait) and filling the region with an optimum color from a color palette generated from the original image. A recently proposed fully automated algorithm addresses these issues and down-samples high-resolution images into low resolution output that matches the process performed by pixel artists better than the previous algorithms (Gerstner et al., 2013).

In the study, we collect behavioral data (“the most aesthetically pleasing level of contrast for an image) and associated eye movements made by expert and novice observers on original and reduced-images in an attempt to find task relevant features for contrast adjustment in image editing. Our results show that the low-level pixelized image, which only preserves the most representative colors but not their positioning, is sufficient for contrast adjustment and, perhaps surprisingly, performs better than the mid-level representation. Moreover, based on eye movement data we conclude that it is sufficient to sample a small portion of the low-level pixelized image for performing contrast adjustment comparable to the high resolution image with full semantic content and elicits global scanning.

Method

Participants

All participants had normal or corrected-to-normal (with soft contact lenses) vision and normal color vision. All participants gave informed consent for their participation in the study that was run in accord with the policies of the Ethics Committee for the Protection of Human Subjects at the Technical University of Kaiserslautern, Germany. All participants were naïve with respect to the goal of the experiment.

Group-1: Experts (Professional / Semi-professional): There were four participants in this group. Two were professional photographers from local photo studio and other two were students in virtual design at the University of Applied Sciences in Kaiserslautern. The professionals charged between 40 € to 80€ per hour. The virtual design students were paid 25€ each for their time.

Group-2: Novices: Twenty-one young adults affiliated to the Technical University of Kaiserslautern were each paid 6.00€ per hour or awarded course credit to participate in the experiment. The participants reported having little or no experience with image processing.
Methods – Three levels of image reduction were used that progressively reduced the semantic content of the image. Reduction was done by selecting 16 most representative colors for the entire image extracted by using k-means clustering. The image was divided into 50 X 50 pixel blocks and filled with most representative color for that block (“pixelized”). The blocks were randomly shuffled to remove cues to spatial distribution of the representative colors, grouping and region segmentation (“pixelized-random”). Each trial started at the lowest contrast level and ended when the participants chose the most aesthetically pleasing contrast. We used 25 images from four different categories: landscape with water, landscape without water, macro (non-human) and macro (portrait). Twenty different levels were rendered for each of the 300 images (4 X 25 X 3 variants – original, pixelized, pixelized-random).

Figure 1. Methods – Three levels of image reduction were used that progressively reduced the semantic content of the image. Reduction was done by selecting 16 most representative colors for the entire image extracted by using k-means clustering. The image was divided into 50 X 50 pixel blocks and filled with most representative color for that block (“pixelized”). The blocks were randomly shuffled to remove cues to spatial distribution of the representative colors, grouping and region segmentation (“pixelized-random”). Each trial started at the lowest contrast level and ended when the participants chose the most aesthetically pleasing contrast. We used 25 images from four different categories: landscape with water, landscape without water, macro (non-human) and macro (portrait). Twenty different levels were rendered for each of the 300 images (4 X 25 X 3 variants – original, pixelized, pixelized-random).

Stimuli

We obtained 200 images that were edited to look “most aesthetically pleasing” by a professional photographer who was not a participant in this study. These images from four different categories (Landscape with Water, Landscape without Water, Macro (with non-human focal object), and Portrait or macro with face as focal object; 50 images each) and were converted to a lossless file format (PNG). However, the darkest and the lightest points were not the same for every image. Therefore, the contrast was normalized such that for each image the darkest color was mapped to black and the lightest color was mapped to white, colors in between were mapped accordingly. We used Python ImageOps library method...
“Autocontrast.” The size of images was normalized to 1250 x 1000 pixel by scaling the shortest side to 1250 pixel and then cropping the image. The authors manually removed images where the previous operations created unwanted artifacts, such as where the focal object was cropped or resizing artifacts spoiled the image or some previous manipulation (borders) were detected. The final chosen set consisted of 100 “original” images with 25 in each of the preceding categories.

A reduced pixelized version was created from each original image by first quantizing the color space of each image individually (16 representative colors per image using k-means clustering in RGB color space with k = 16). Then the images were divided into squares of size 50 x 50 pixels and each square was filled with a representative color using the nearest neighbor method, i.e. the square was filled with the color of the central pixel of this square; these images are called “pixelized” images. Then another variant “pixelized-random” was created by randomly shuffling the spatial distribution of the 50 x 50 pixel blocks of the pixelized version such that the overall color distribution was retained but cues to semantics and region segmentation of the photograph were lost. During stimuli creation, the authors manually monitored and repeated the randomization process for images that produced unwanted artifacts of uniform blobs of colors in some region of the image.

The motivation for using these three different image reductions was to successively reduce top-down influences on aesthetic judgment and eye movement patterns.

For the “original” image, the semantics, meaning, scene/object layout, and color compositions were preserved. We hypothesize that any image processing task performed for the original version would show a combined effect of low-, mid– and high-level processing, top-down knowledge-based influence due to the presence of semantic, object, and context relevant information.

Each original image was quantized into uniformly colored blocks of 50 x 50 pixels (or 1.1° x 1.1° visual angle at our viewing distance) to create the “pixelized” version. This version was devoid of semantic and context relevant information. We hypothesize that image processing operations for the pixelized version would show combined effect of low– and mid-level processing. This version preserved the spatial relationship of patches of the original image allowing region segmentation and grouping by color similarity and gradually changing contrasts.

A “pixelized-random” version was created for each pixelized image by randomly placing the 50 x 50 blocks within the frame. This version was devoid of high– and mid-level cues. Only the colors from the previous steps were preserved although there spatial locations were randomized, thus removing cues to semantics and region segmentation. We hypothesize that image processing operations for the pixelized-random version would show effect of low-level processing based on image properties such as brightness, contrast and color.

Finally, twenty different levels of contrast were pre-rendered for each of the 300 images (100 original, 100 pixelized and 100 random-pixelized images). We used Michelson contrast defined as (contrast = (I_{MAX} - I_{MIN}) / (I_{MAX} + I_{MIN}), where I_{MAX} and I_{MIN} represent the highest and lowest luminance values, respectively. For each image, 10 images with a lower contrast and 9 images with a higher contrast were generated by blending the image I_o with a uniformly grey image I_g according to the formula I_{g'} * (1 - c) + I_o * c, where c corresponds to the contrast level in the range [0, 1.9] in 0.1 steps. The grey value of I_g was the mean grey value of I_o. For example, the lowest contrast level c = 0 produces a uniformly grey image; the contrast level c = 1.0 produces the original image, and the highest contrast level was c = 1.9.

Our aim was to identify the image reduction that preserves features relevant to contrast adjustment by comparing the data obtained for the two reduced-image variants to the original image.

2 http://effbot.org/imagingbook/imageops.htm#tag-ImageOps.autocontrast.
Apparatus

Stimuli were shown on a Mitsubishi Diamond Pro 2070SB monitor with a screen size of 0.406 x 0.305 m\(^2\) at a resolution of 1280 x 1024 pixel at a refresh rate of 85 Hz. The participant was positioned such that their eyes were centered on the screen and the viewing distances was 0.8 m. Therefore 1 degree of visual angle corresponded to 45 pixels. The participant’s head was stabilized using a chin and forehead rest.

The participant’s right eye was recorded using an SR Research Eyelink 1000 sampling at 1000 Hz. The sequencing and display of image as well as recording of response was controlled by Eyelink Experiment Builder® Software. At the beginning of the experiment, a 9-point calibration sequence was conducted for each participant; the calibration sequence was repeated when necessary. At the beginning of each trial, a drift correction was performed.

General Experimental Design

We used a mixed factorial design with expertise as the between-subject factor and [Image category x Image simplification x Levels of contrast] as the within-subject factors. There were four different image categories (landscape with water, landscape without water, portrait, and macro), three levels of image simplification (“original”, “pixelized”, “pixelized-random”), and 20 levels of contrast. The 300 images were divided into 3 blocks such that each subject only saw one variation of an image to avoid carry over effects in image editing. It is documented that short-term and long-term memory affects perceived color (Allred & Olkkonen, 2015; Bloj, Weiß, & Gegenfurtner, 2016).

Procedure

At the beginning of the experiment, an eye tracker calibration procedure was carried out, and a drift correction was performed before each new stimulus image. All images were shown at the center of the screen. At the beginning of a trial, an image was shown at its lowest contrast level, that is, the uniformly grey image. By pressing keys, the participant could then increase or decrease the contrast level in a stepwise manner and the image on the screen was replaced with the corresponding pre-generated lower or higher contrast version of the image. The participant could arbitrarily increase or decrease the contrast level of the image. Eye movements were recorded from stimulus display onset (i.e., as soon as the stimuli was presented with lowest contrast) until the stimulus was removed (i.e., the most aesthetically pleasing contrast level for the image was chosen by the participant). For each trial the eye movements, the corresponding (i.e., the currently displayed) contrast level, and the final decision (i.e., the personally preferred contrast level) were recorded. Each trial was untimed; the participants were able to complete one block, with 100 trials, in 45 to 60 minutes.

Task

The participants were instructed to find the most aesthetically pleasing version of the displayed image by using Key-A to decrease contrast and Key-L to increase contrast. They were instructed to end the trial by pressing Key-G, when they were satisfied with their selection of the most aesthetically pleasing image.

Results

The goal of behavioral data analysis was to determine if there is a difference in preferred contrast based on expertise (expert vs. novice) and image
reduction (original vs. pixelized; original vs. pixelized random). From the eye movement recordings we were interested in finding if there is any difference in the fixation pattern (mean number of fixations, average fixation time, spatial location of fixation for each image category and image reduction) depending on expertise (experts vs. novices) and between human and machine (OpenCV point-of-interest (POI) detector algorithm).

**Behavioral measure**

For every image, we recorded the contrast that was judged by the participants to be the most aesthetically pleasing. The statistical analysis was performed using Bayesian estimation that incorporates a Markov Chain Monte Carlo method to generate 100,000 credible parameter values given the data (Kruschke, 2013). This method yields distribution of likely means, standard deviations, and effect sizes for a given data. It also gives the mean difference between two groups and a 95% highest density interval (HDI) to indicate if this difference is statistically significant (the 95% HDI should be well above zero).

We used scripts for R (R Development Core Team, 2012) developed by Kruschke (2011), to implement MCMC methods to generate a set of 100,000 credible parameter values, histogram of credible parameter estimates, and the interval that contains 95% of the credible estimates.

Overall, we observed no significant differences due to expertise. Figure 2A shows the comparison of means for the experts (mean = 9.85) and the novices (mean = 9.73), difference of mean = –0.113, (95% HDI: [-0.358, 0.136] includes 0, thus the difference is not significant) for data collected for the four image categories and the three image variants. Therefore, we combine the data from two groups for all images and Figure 2B shows the combined data for all 25 participants for preferred contrast for the three levels of image reduction –the original image, the pixelized version, and the pixelized-random version.

For the different image categories, the difference in mean was significant for the following comparisons: Macro vs. Landscape without water, Portrait vs. Landscape with water, and Portrait vs. Landscape without water (see Table-1). As elaborated in the introduction, facial contrast, the balance between luminance of facial features, is critical for judging facial attractiveness. We hypothesize the observers were being conservative while increasing contrast in order to prevent the “washed-out” look (see Figure 3) that gives an unflattering appearance to the skin. The semantics of the portrait is important because this effect is observed for the original and pixelized version that clearly convey the face motif of the image but not in the pixelized-random version with semantic and region segmentation cues removed.

In order to find which image reduction ideally preserves the task relevant features for contrast adjustment we compared the data for the three image variants. We observed that the contrast preference is similar for original (mean = 9.54) and random-pixelized versions (mean = 9.44), difference of mean = 0.105,
(95% HDI: [-1.72, 0.389]) was not significant. A higher contrast was chosen for the pixelized version (mean = 10.3). The difference in mean was significant when compared with the original (difference = 0.716, 95% HDI: [0.447, 0.982]) and with the pixelized-random version (difference = 0.816, 95% HDI: [0.497, 1.13]). This result holds for all different image categories (see Table-1). The final results show that task-relevant information required for selecting the most aesthetically pleasing contrast is retained in the random-pixelized version generated by our method. Even though the pixelized version is “more similar” to the original image in its appearance and computationally less expensive than the original high resolution image, it is not ideal as a proxy for task relevant features for contrast adjustment. We hypothesize that in the pixelized version, similar color share boundaries and aesthetic judgment are driven by the need to improve the contrast of those regions (e.g. region of the hairline in Figure 2 and Figure 3), thereby leading to an overestimation of the most aesthetically pleasing contrast for that image variation. The computational load is the same for the pixelized and pixelized-random image, but only by empirical testing it is possible to decide image representations for building modules in computational databases that can lead to human-like performance for image enhancement operations.

In order to better understand the role of expertise, effect of semantics and image properties on contrast editing we further plot the data for each image variant in Figure 4. There were no significant differences in the group means for the pixelized and pixelized-random versions. However, the preferred contrast differed for the two groups of participants for the original images, experts (mean = 10.1) and the novices (mean = 9.4), difference of mean = – 0.707, (95% HDI: [-1.02, –0.403] is far from 0, thus the difference is significant).

The experts chose a contrast level that was similar to the contrast (mean = 10.1) of the original image (contrast 10.0) that was obtained from a professional photographer who did not participate in this study. From these results we were pleased to see some consensus among experts for the level of contrast that is aesthetically pleasing. Expertise in the domain of photo editing shows up as the ability to find the contrast that appeals to the masses. In a post-experimental interview, we observed that the novice observers found the original contrast of the image aesthetically pleasing when presented just by itself, however, the novices were not able to select which one was the original one when they were presented with variations that differed slightly in contrast setting. In general, the novices preferred a lower contrast (mean=9.4) compared to the original image. Our empirical data supports the results from computational aesthetics provided by the ACQUINE system (Ke et al., 2006) that non-professional photographs can be distinguished from professional ones because they have lower contrast. It appears that novices do not have the implicit knowledge to get the “Aha-experience” for contrast when capturing and editing photos.
Figure 2. **A. Experts vs. Novices** – There was no significant difference in preferred contrast for the two groups when combined over all image categories and image variants. **B. Most aesthetically pleasing contrast data** – For the two reduced image variants, the mean preferred contrast was comparable for the original and the pixelized-random version but not for the original and the pixelized version. This is surprising because the pixelized-random variant has all cues to semantics, grouping and region segmentation removed. For some image editing operations, such as, contrast adjustment it is better to remove mid-level cues to image segmentation, for reasons discussed in the text.
Table 1

**Most aesthetically pleasing contrast for each image category** – In each cell, in the first row and column, the three values indicate the most preferred contrast for original, pixelized, and pixelized-random image variants for the given category. The other cells show the difference of mean for the given image variant between the two categories (* represents that the difference was statistically significant).

<table>
<thead>
<tr>
<th></th>
<th>Landscape with water</th>
<th>Landscape w/o water</th>
<th>Macro</th>
<th>Portrait</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>original</strong></td>
<td>9.77</td>
<td>10.2</td>
<td>9.36</td>
<td>8.87</td>
</tr>
<tr>
<td><strong>pixelized</strong></td>
<td>10.5</td>
<td>10.8</td>
<td>9.77</td>
<td>9.77</td>
</tr>
<tr>
<td><strong>pix-random</strong></td>
<td>9.33</td>
<td>9.69</td>
<td>9.25</td>
<td>9.25</td>
</tr>
</tbody>
</table>

| Landscape w/o water | |                  |       |          |
|---------------------|----------------------|-------|----------|
| **Landscape with water** | |                  |       |          |
| 9.77                |                      |       |          |
| 10.5                |                      |       |          |
| 9.33                |                      |       |          |

<table>
<thead>
<tr>
<th>Landscape w/o water</th>
<th>Macro</th>
<th>Portrait</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.2</td>
<td>0.458</td>
<td></td>
</tr>
<tr>
<td>10.8</td>
<td>0.318</td>
<td></td>
</tr>
<tr>
<td>9.69</td>
<td>0.357</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Macro</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9.36</td>
<td>0.408</td>
<td>0.865*</td>
</tr>
<tr>
<td>10.2</td>
<td>0.274</td>
<td>0.597</td>
</tr>
<tr>
<td>9.77</td>
<td>0.422</td>
<td>0.072</td>
</tr>
</tbody>
</table>

| Portrait | |          |          |
|----------|----------------------|----------|
| 8.87     | 0.899* | 1.35*    | -0.469  |
| 9.77     | 0.727  | 1.05*    | -0.448  |
| 9.25     | 0.089  | 0.443    | -0.512  |
Figure 3. “Washed-out” Look of Portraits – As contrast level is increased beyond an optimum value, portraits show a “washed-out” look because it reduces texture of the skin and makes it look unflattering, as illustrated here by exemplars of highest contrast level (=20). However, other image categories seem resistant to such look and may seem more aesthetically pleasing at higher contrast due to increased vividness.
Figure 4. Experts vs. Novices Comparison for the three image variants – There was no significant difference in preferred contrast between the two groups for the pixelized and pixelized-random variants (B and C). However, there was significant difference in preferred contrast for the original images (A) indicating importance of semantics in eliciting expert knowledge in photo editing. The results imply that the original version is indispensable for fine tuning the final contrast in order to get the “Aha experience” generated by expert photo editors. Nevertheless, computer-based similarity-match for contrast editing on reduced pixelized-random version can help decrease the number of possible photo editing options to be considered for final editing.
Eye movement measurement analysis

Fixations shorter than 80 ms and longer than 800 ms were excluded. These lower and upper cut-offs are chosen to eliminate noise due to wrong classification of the eye movements by the eye tracker (short fixations belong either to a decelerating saccade or are too short to extract information anyway) or to disregard unnatural eye movement behavior (depending on the task the average fixation takes around 300 ms). Furthermore only fixations which are at least 1° visual angle within the image are considered in order to guarantee that only the foveal field of vision covered the image content.

Figure 5. Eye movement analysis – Individual differences – shows that there are huge individual differences in the number of fixations made and the time spent for a single fixation. The durations were normalized by total time of trial for a given image for every participant. The pattern of spatial distribution of fixation and mean fixation duration of a fixation does not differ between the experts and the novices.
A total of 16673 fixations were recorded, of which 3551 fixations were excluded according to the above criterion. Each fixation was weighted by the normalized duration of fixation. The duration of fixation is important because looking at a point for 300ms may indicate that it is more important than one looked at for 80ms. The durations were normalized by total time of trial for a given image for every participant. The fixations were normalized by duration of the trial because as different people have different looking preferences. Some people prefer to look at things for longer duration compared to others. The total duration of trial also can depend on the preference of subjects for a certain image. So we used a normalizing procedure of dividing the fixation duration at a point (x, y) by the total duration of that trial. The sum of normalized fixation for all subjects at a point (x, y) is represented in the heatmaps.

We found huge individual differences in the number of fixations made and the time spent for a single fixation. The pattern of spatial distribution of fixation and mean fixation duration of a fixation does not differ between the experts and the novices in any statistically significant way (see Figure 5). In our data we found no evidence of difference in fixation pattern of experts and novices (Nodine et al., 1993; Zangemeister et al., 1995) that claim that experts look longer at few global patterns while novices make multiple fixations of shorter duration and collect more local information. We hypothesize that this observed difference is because our task, “selecting most aesthetically pleasing contrast” was different from the previous studies exploring “aesthetic evaluation by generating verbal reaction”.

The distribution of fixation duration at preferred contrast shows that the subjects looked at the preferred contrast for the longest duration followed by the next higher and next lower contrast for all three image variants, original, pixelized, and pixelized-random (see Figure 6). It points to a strategy that people look back and forth between the most preferred and the next-most preferred contrast before deciding on the one they feel is the most aesthetically pleasing one. Another way of interpreting the data could be that people prefer to gaze at the most-aesthetically preferred image longer as reported in previous studies (Holmes & Zanker, 2012; Shimojo et al., 2003). The significant outcome of this analysis is the longest duration of fixation is correlated to the preferred contrast chosen in the image.
Figure 6. Fixation duration is highest at preferred contrast for a given image – The distribution of fixation duration at preferred contrast shows that the participants looked at the preferred contrast for the longest duration followed by the next higher and next lower contrast for all three image variants, original, pixelized, and pixelized-random.
In Figure 7 the heatmaps of fixation are plotted for the different image variants in rows (original, pixelized, pixelized-random respectively) and the different image categories in columns (landscape with water, landscape without water, macro and portrait). Note that the fixations are collapsed over different contrast levels. It can be seen that a different distribution pattern along the category dimension is visible for the original images and in traces for the pixelized images. In particular for landscape images the fixations are located mostly in the lower half of the image, suggesting that the sky (dominant concept in the upper half of landscape images) is of lower relevance for contrast adjustment. With fixations relatively similarly distributed around the center of the screen, images belonging to the portrait and macro category evoke the same eye movement patterns. This is expected since the portrait category is assumed to be only a special case of the macro category (“macro image of human face”). The bigger spread of fixations around the center of the screen in case of the portrait category is explained by the fact that the focal object of the portrait category (i.e., the face) covers in general more of the image than the focal object of the macro category. The significant outcome of this analysis is that fewer centralized fixations are required by humans when they perform
contrast adjustments on reduced pixelized-random images. Again, this result can significantly improve the efficiency of automatic image processing algorithms by incorporating human-like strategies for sampling reduced images. These findings are reinforced in the histograms plotted in Figure 8.

**Figure 8. Eye movement analysis** – The histograms of mean fixation distance from the center and number of fixation for each image variant show that for the pixelized-random image the task-relevant information can be gleaned by a relatively fewer number of fixations close to the center. With increased the semantic and spatial region segmentation information the fixations are more global and more fixations are required for acquiring the task-relevant features from the image.
Figure 9. Comparison of feature selection by experts, novices and computer algorithm. There was no significant difference in the pattern of spatial distribution of fixations on the images for the two groups of human observers (experts vs. novices). However, the spatial distribution of fixations made by OpenCV point-of-interest detector scans the entire image even when the semantic information is reduced. The difference between the human observer and computer algorithm is most pronounced for the pixelized-random version of the images. This information can help reduce the number of keypoints chosen for certain image processing operations by the backend of the automated image enhancement systems.
Figure 8 shows that the mean distance of fixation from the center of the image is maximum for the original image and decreases progressively as the high-level and mid-level cues are removed. This finding supports the two-stage process of gaze behavior suggested in exploration of art (Locher et al., 2007) because global explorations are hypothesized to be made for gleaning semantic information in the first stage and then local explorations, mostly close to the center of the image is performed in the second stage. Since contrast adjustment is dependent on lightness and darkness of pixels in an image it can be performed without extensively scanning the scene globally as long as most task relevant features are present in a small neighborhood. The pixelized-random version is an ideal manipulation that retains all task relevant features of image lightness and presents the most representative colors in small neighborhood. This drastically reduces the number of fixations that is required in every trial for gleaning the important task-relevant features (the right column in Figure 8). Most fixations are clustered around the center of the image due to center-bias that has been well documented previously. Central fixation bias exists irrespective of the image feature distribution in the image or the task (free viewing/visual search; Tatler, 2007). There tends to be a reliable bias toward having more features and objects in the center of natural scenes (e.g., Parkhurst & Niebur, 2003; Reinegal & Zador, 1999; Tatler, 2007). This bias is in part because photographers tend to place objects of interest at the center of the viewfinder. Thus, if fixations and features correlate, a centrally biased distribution of features in scenes would result in the observed central biases in human fixation distributions.

There was no significant difference in the pattern of spatial distribution of fixations on the images for the two groups of human observers (experts vs. novices). OpenCV is a popular computer algorithm for implementing local scale invariant features or SIFT keypoints. The program returns the location of keypoints that capture the characteristics of a given image for image matching. We computed the SIFT keypoint for each of 300 images used in our experiment and the difference in distance between the eye fixations made by human observers on that image. As shown in Figure 9, the difference between the location of human fixations and the keypoints detected by the computer algorithm progressively increases from original to pixelized version and is the greatest for the pixelized-random version. The spatial distribution of fixations made by OpenCV point-of-interest detector scans the entire image even when the semantic information is reduced. The difference between the human observer and computer algorithm is most pronounced for the pixelized-random version of the images. This information can help reduce the number of keypoints chosen for certain image processing operations by the backend of the automated image enhancement systems.

3 http://docs.opencv.org/3.1.0/da/df5/tutorial_py_sift_intro.html

PSIHOLOGIJA, 2017, Vol. 50(3), 239–270
Discussion

In this study, our goal was to empirically test different versions of low-resolution images that can elicit contrast adjustment responses comparable to high-resolution original images obtained from a professional photographer. We implemented a computer algorithm for image reduction that is inspired by pixel art. We generated two low-resolution versions with the same computational load: a pixelized “mid-level” version that preserved spatial distribution of the most representative colors and cues to grouping and region segmentation, and a pixelized-random “low-level” version had these cues removed by spatial shuffling of these colors. Our main finding is that the low-level image is sufficient for contrast adjustment and performs better than the mid-level representation for four image categories that are commonly uploaded to social media: landscape-with-water, landscape-without-water, portrait, and non-human macro. Based on our results, we suggest that only by empirical testing it may be possible to identify image representations, which retain task-relevant features for building modules for aesthetic judgments in computational databases that can lead to human-like performance for image enhancement operations. At the first glance, the appearance of the mid-level representation is more similar to the original image but careful execution of empirical studies is required for identifying the image representation that retains the task-relevant features.

The implication of this finding for computational aesthetics is that, for a given image processing operation, it is possible to find reduced image representations, in order to modularize the process of storing and retrieving task-relevant features of a given image-plus-image-enhancement combination. It is not necessary to store the entire high resolution original image but just the “correct” reduced version. This would reduce the requirements for storage memory network load in web based image processing tools, a usual bottleneck in such systems. Reduction of an image to the task relevant information would also increase the retrieval precision by decreasing the noise (introduced by task irrelevant information). Analysis of the eye movement data provides an insight how people glean information differently over the three image variants and could provide information to reduce the number of sampling points an image processing operation.

In our study, we collect behavioral data (“the most aesthetically pleasing level of contrast for an image) and associated eye movements made by expert and novice observers on original and reduced-images in an attempt to find task relevant features for contrast adjustment in image editing. Our main finding is that the high-level and mid-level images show global fixation patterns as users scan the entire image for semantic information. In contrast, users primarily fixated on the center of the low-level image in order to make their contrast preference. Our findings suggest that it is sufficient to sample a small portion of the saved pixelated-random image for image retrieval, in contrast to extracting all key points, as is currently done in search algorithms based on computer vision algorithms. Another direction of future research would to use different
image features from the fixations done by users while performing the image editing task and train classifiers to distinguish between experts and novices. If it is possible to computationally model the expertise level, then we can further narrow the search for task-relevant features for a given image editing operation. While the pixelized-random version leads to excellent results for adjusting the contrast to the “most aesthetically pleasing” level, it may not suffice for a similar color adjustment task because color perception depends on context and memory of scenes (Allred & Olkkonen, 2015; Bloj et al., 2016), an issue we plan to empirically test in near future.

In summary, we report behavioral and eye movement data to find reduced image representation for aesthetically pleasing adjustment of contrast in photographs and suggest possibilities for modularizing the process of storing and retrieving task-relevant features for the backend of computational aesthetics. This would reduce the requirements for storage memory network load in web based image processing tools, a usual bottleneck in such systems as well as time required for identifying keypoints for calculating image similarity by using more human-like centrally biased fixation patterns on the reduced images.

References


PHOTOEDITING FOR MOST AESTHETICALLY PLEASING CONTRAST 
ON REDUCED IMAGES


© 2017 by the Serbian Psychological Association

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution ShareAlike 4.0 International license

PSIHOOLOGIJA, 2017, Vol. 50(3), 239–270
Preparation of Stimuli

Code to Normalize Contrast

```python
import os

import Image
import ImageOps

def autocontrast(directory='portrait'):
    # maximizes image contrast, remaps colors such that darkest pixel becomes black and lightest pixel becomes white
    for file_name in os.listdir(directory):
        image = Image.open(os.path.join(directory, file_name))
        autocontrast_image = ImageOps.autocontrast(image)
        autocontrast_image.save(os.path.join(directory, 'normalized_'+ file_name.replace('.jpg', '.png')))

if __name__ == '__main__':
    autocontrast()
```

Code for creating color-reduced versions

```python
# reduce color space from original images to a quantized color space version, palette size specifies how many colors remain in the image
import os
import scipy.misc
import scipy.cluster.vq

def autocontrast(directory='portrait'):
    # maximizes image contrast, remaps colors such that darkest pixel becomes black and lightest pixel becomes white
    for file_name in os.listdir(directory):
        image = Image.open(os.path.join(directory, file_name))
        autocontrast_image = ImageOps.autocontrast(image)
        autocontrast_image.save(os.path.join(directory, 'normalized_'+ file_name.replace('.jpg', '.png')))

if __name__ == '__main__':
    autocontrast()
```

```python
import os

import Image
import ImageOps

def autocontrast(directory='portrait'):
    # maximizes image contrast, remaps colors such that darkest pixel becomes black and lightest pixel becomes white
    for file_name in os.listdir(directory):
        image = Image.open(os.path.join(directory, file_name))
        autocontrast_image = ImageOps.autocontrast(image)
        autocontrast_image.save(os.path.join(directory, 'normalized_'+ file_name.replace('.jpg', '.png')))

if __name__ == '__main__':
    autocontrast()
```

```python
# reduce color space from original images to a quantized color space version, palette size specifies how many colors remain in the image
import os
import scipy.misc
import scipy.cluster.vq

palette_size = 16

for image_name in os.listdir('base'):
    image = scipy.misc.imread(os.path.join('base', image_name))
    colors = image.reshape(-1, 3)
    cluster_centers, cluster_assignments = scipy.cluster.vq.kmeans2(colors, palette_size)
    for i in range(len(cluster_centers)):
```
import os
import collections
import numpy
import scipy.misc

cell_size = 50

for image_name in os.listdir('color_reduced'):
    image = scipy.misc.imread(os.path.join('color_reduced', image_name))
    resized_image = scipy.misc.resize(image, (image.shape[0] / cell_size, image.shape[1] / cell_size), interp='nearest')
    pixelated_image = scipy.misc.resize(resized_image, (resized_image.shape[0] * cell_size, resized_image.shape[1] * cell_size), interp='nearest')
    scipy.misc.imsave(os.path.join('pixelized', 'pixelized_' + image_name.replace('.jpg', '.png')), pixelated_image)
    random_image = resized_image.reshape(-1, 3)
    numpy.random.shuffle(random_image)
    random_image = random_image.reshape((image.shape[0] / cell_size, image.shape[1] / cell_size, 3))
    random_pixelated_image = scipy.misc.resize(random_image, (random_image.shape[0] * cell_size, random_image.shape[1] * cell_size), interp='nearest')
    scipy.misc.imsave(os.path.join('pixelized_random', 'pixelized_random_' + image_name.replace('.jpg', '.png')), random_pixelated_image)

import os
import Image
import ImageEnhance

directory = 'base'

# generate different contrast versions of the images provided in the respective folder
import os
import Image
import ImageEnhance

#directory = 'base'
directory = ‘pixelized’

directory = ‘pixelized_random’
for image_name in os.listdir(directory):
    image = Image.open(os.path.join(directory, image_name))
    enhancer = ImageEnhance.Contrast(image)
    for i in range(20):
        contrast = i / 10.
        contrast_image = enhancer.enhance(contrast)
        contrast_image.save(os.path.join('contrast_' + directory, str(i) + ‘.’ + image_name))

Eyelink ® 1000 eye tracker with Experiment Builder Software.

The sequencing and display of image as well as recording of response was controlled by Eyelink Experiment Builder ® Software. The screenshots below show (1) the overall program for controlling the experiment (2) the loop the displays image, updates images based on user response, and records the final choice of the user while recording eye-movements at every millisecond.

(1) the overall program for controlling the experiment
(2) the loop displays image, updates images based on user response, and records the final choice of the user while recording eye-movements at every millisecond.