

# Aesthetic Photo Enhancement using Machine Learning and Case-Based Reasoning

Joachim Folz<sup>1</sup> Christian Schulze<sup>1</sup> Damian Borth<sup>1</sup> Andreas Dengel<sup>1,2</sup>

<sup>1</sup> German Research Center for Artificial Intelligence (DFKI), Germany    <sup>2</sup> University of Kaiserslautern, Germany  
{joachim.folz, christian.schulze, damian.borth, andreas.dengel}@dfki.de

## ABSTRACT

Broad availability of camera devices allows users to easily create, upload, and share photos on the Internet. However, users not only want to share their photos in the very moment they acquire them, but also ask for tools to enhance the aesthetics of a photo before upload as seen by the popularity of services such as Instagram. This paper presents a semi-automatic assistant system for aesthetic photo enhancement. Our system employs a combination of machine learning and case-based reasoning techniques to provide a set of operations (contrast, brightness, color, and gamma) customized for each photo individually. The inference is based on scenery concept detection to identify enhancement potential in photos and a database of sample pictures edited by desktop publishing experts to achieve a certain look and feel. Capabilities of the presented system for instant photo enhancements were confirmed in a user study with twelve subjects indicating a clear preference over a traditional photo enhancement system, which required more time to handle and provided less satisfying results. Additionally, we demonstrate the benefit of our system in an online demo<sup>1</sup>.

## Categories and Subject Descriptors

I.4.9 [Image Processing and Computer Vision]: Applications; I.2.6 [Artificial Intelligence]: Learning—*Knowledge acquisition*

## Keywords

Aesthetics; Enhancement Potential Detection; Photographic Problem Detection; Image Processing; Photo Enhancement; Desktop Publishing; Case-Based Reasoning; Machine Learning; Knowledge Acquisition

<sup>1</sup>Access demo at <http://www.madm.eu/demos/rsip>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

ASM'15, October 30, 2015, Brisbane, Australia.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-3750-2/15/10 ...\$15.00.

DOI: <http://dx.doi.org/10.1145/2813524.2813531>.

## 1. INTRODUCTION

Since digital photography has almost completely replaced its analog predecessor, photo editing via image processing offers opportunities for billions of amateur photographers to reflect different aesthetic intentions, i.e., to convey a specific emotion, look, or style through their works. The success of services like Instagram<sup>2</sup> has shown a high demand for easy to use solutions for image editing. While existing tools such as Adobe Photoshop<sup>3</sup> or The GIMP<sup>4</sup> offer great flexibility to experienced users, their complexity (or price) can prevent novices from achieving their goals. Our proposed solution to this dilemma is two-tiered: First, the set of applicable operations is intelligently filtered to those that can tap into existing enhancement potential (EP). For instance, if a photo has an undesirable amount of contrast—or any other photometric property—value and there is a known enhancement operation that can change its contrast to a more desirable value, we say there is EP. Second, suggest ready-to-use sets of operations extracted from expert knowledge that serve a given editing intention.

Both approaches are employed to achieve this: EP detectors are trained on tailored data sets. Their output is used to decide whether the corresponding enhancement is available and to suggest initial correction parameters. Further, expert photos are retrieved from a knowledge base of example images edited by desktop publishing (DTP) professionals using Adobe Photoshop. From this knowledge base we retrieve candidates that are photometric similar or show the same kind of scenery, e.g., landscape. Operations applied by the expert are extracted from retrieved candidates and replayed on user photos.

In the following, we describe our solutions for detecting EP in more detail, how similarity between photos is defined and sceneries are detected. Finally, we analyze user behavior and reception in a case study with twelve users. We especially focus on the preferences with regards to the available enhancement sources as shown via their usage during the study and the subjective opinions of the subjects.

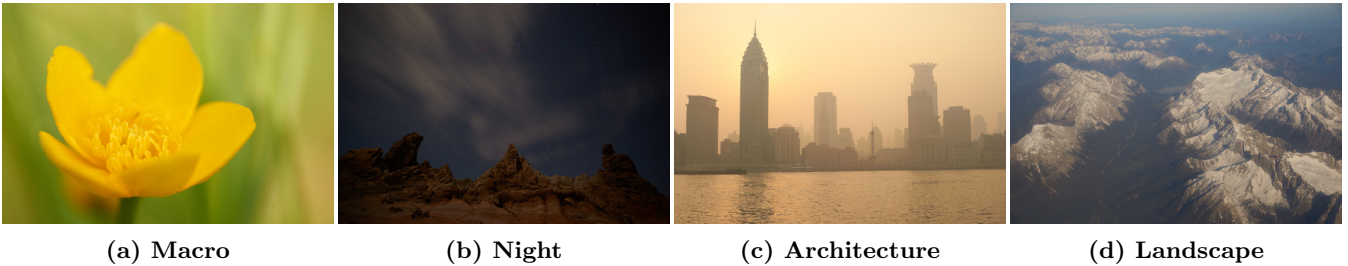
## 2. RELATED WORK

A simple method for contrast enhancement often found in photo editing tools is linear min-max histogram equalization, which simply stretches lightness values to span the entire available range. Though this method can be effective,

<sup>2</sup><http://instagram.com>

<sup>3</sup><http://adobe.com/Photoshop>

<sup>4</sup><http://gimp.org>



**Figure 1: Four problematic source photos representing the covered scenery concepts, which were sent to DTP experts for editing. Photos ©Ricoh Company Ltd., Tokyo Japan**

results are often unsatisfactory as single bright or dark pixels can prevent effective enhancement. Extreme enhancements are applied if the original lightness range is small. Improved variants aim to compensate for some of these problems, but introduce several parameters [6, 9]. Other examples for more sophisticated methods employ, e.g., the Retinex theory of color vision [4, 5, 7]. Satisfying results for both contrast and color correction are achieved in many cases, though careful tuning of the parameters is required to avoid visible artifacts like bright halos between areas of different lightness.

Highly parametrized enhancements are however not compatible with the primary design goal of our system: offering a simple, yet flexible and powerful approach to photo editing. Instead we use enhancements for contrast, brightness and color based on linear interpolation and extrapolation for our EP approach [2]. They take a single parameter and are used, e.g., by the Pillow imaging library<sup>5</sup>. The methods can be described as instantiations of one generic formula

$$I' = (1 - \alpha)I_m + \alpha I = I_m + \alpha(I - I_m) \quad (1)$$

with original image  $I$ , strength parameter  $\alpha$ , enhanced result  $I'$ , and reference image  $I_m$  for each operation. This reference image, also called degenerate image, corresponds to the lowest amount of contrast, color, or brightness possible for the source image. An enhancement is achieved by blending  $I$  and  $I_m$  as defined above and either interpolating for  $\alpha < 1$  to attenuate or extrapolating for  $\alpha > 1$  to intensify. For contrast and brightness  $I_m$  represents a uniform color: the mean value and 0 respectively. Color can be changed using the luminance image.

Regarding our machine learning approach to detect EP, measures for the perceived properties of photos are required. The contrast image from the well-known the Tamura texture features is a good way to describe contrast [10]. Color is estimated from saturation information taken from  $HSV$  color space. Luminance—useful for brightness detection—can be gathered directly from the  $Y$  channel of  $YCbCr$  space as it is used for the JPEG, the most likely file format for uploads by normal users.

### 3. ENHANCEMENT SOURCES

#### 3.1 Enhancement Potential Detection

To limit the number of available options to a useful minimum, we only present operations that can tap into existing EP. For example, the option to modify the contrast is only available if EP is detected. Detectors are represented by

<sup>5</sup><https://python-pillow.github.io/>

SVMs that are trained on tailored data sets for each type of enhancement. We use the standard libSVM library following a procedure recommended by Hsu et al. [3].

*Contrast and saturation* data sets consist of randomly selected photos from Flickr. First, 507 are manually labeled as “potential” and “no potential”. These are then automatically re-segmented by a Bayesian classifier to reduce confusion and finally used to bootstrap a much larger data set of 30,000 images.

We test different variants of smoothed histograms (8 to 256 bins) of the Tamura contrast for feature representation:

- Global: standard smoothed histogram.
- Saliency weighted: instead of a fixed value each pixel contributes its weight, in this case the saliency value as computed by the Boolean map saliency model [11].
- Saliency masked: only those pixels with saliency greater zero contribute.
- Interest point detectors: detect Harris and DoG interest points and use resulting patches to compute histograms.

In our tests, the saliency weighted variant performs better than the alternatives for both contrast (85.3% MAP) and saturation (82.35% MAP), though global and saliency masked are comparable. Interest point based features give considerably worse performance at a maximum of 80.4% and 74.2% respectively.

These enhancements require a single strength parameter. We calculate an initial value to present to the user from the classification confidence of the utilized detector. If the classifier returns the maximum possible confidence of 1, the highest amount of EP is assumed. A confidence of 0.5 corresponds to indecision and the enhancement is not recommended. For a confidence value  $c$  we can calculate the strength value as

$$s_{con} = s_{sat} = 2(c - 0.5). \quad (2)$$

Low confidence values therefore generate negative EP and, in turn, also negative strength for the anticipated correction, i.e., a highly saturated photo may result in a suggestion to decrease the color intensity.

For *gamma correction* a different approach is necessary. Training data is instead collected from DPChallenge<sup>6</sup>, a website that regularly hosts themed challenges for professional and semi-professional photographers since 2002. Their archive lists a total of 2,116 completed challenges

<sup>6</sup><http://dpchallenge.com>

at the time of writing. Its competitive nature, diverse themes and strict voting rules turn it into a reliable source for aesthetic photo analysis [8]. Unfortunately they are not suitable for training contrast or color properties as-is, since challenge themes can take the form of “low/high contrast” and “monochrome”. Upon inspection, no challenges specifically targeting gamma values were found.

Hence, DPChallenge was crawled on 16<sup>th</sup> Feb 2015 for the three highest and lowest scoring photos of every archived challenge with at least 100 valid (not disqualified) entries. This lower bound is enforced to ensure a sufficient community participation for more accurate average ratings. Additionally, each photo must decode without errors, be at least 240k pixels in size, and not a uniform color.

Assuming the gamma of the highest scoring photos to be in a desirable range, a data set for training can automatically be created. For every picture we draw a random intermediate value  $\gamma'$  from the uniform distribution  $\mathcal{U}(0.5, 1.5)$ . In case  $\gamma' = 1$  another value is drawn. The real gamma value  $\gamma$  is derived from the following empirically determined function

$$\gamma = \frac{1}{\gamma'^2} \quad (3)$$

and applied to the photo. It results in almost linear changes of perceived brightness, with respect to  $\gamma'$ . Hence, this representation is used within the proposed system. If  $\gamma' > 1$ , the photo is labeled as “gamma too high” or “gamma too low” if  $\gamma' < 1$ .

Using a data set of 900 images of training and test samples respectively, we achieved 79.06% MAP for gamma correction detection on luminance histograms (8 to 256 bins). The confidence  $c$  of the “gamma too low” class is transformed back to  $\gamma'$  values for use in the system as follows:

$$\gamma' = \frac{1}{3}c + 1. \quad (4)$$

*Brightness* is handled via a manually tuned model targeting an average luminance. Its suggested strength values  $s$  are obtained from the difference of saliency weighted mean luminance values  $l_m$  and a target luminance  $l_t = 160$ , such that

$$s_{bri} = 0.005(l_m - l_t). \quad (5)$$

All deployed detectors share the property that its enhancement is suppressed, if the suggested strength is weaker than  $\pm 5\%$  in order to ignore very low EP.

### 3.2 Expert Database

Highly similar photos are expected to require very similar operations for accomplishing the same editing intention. For example, two photos sharing the same color profile need the same color balance correction for a “cool” look. Following this argumentation, we selected a set of 22 problematic photos, representing a scenery of either architecture, landscape, macro, or night. Due to privacy concerns portraits were omitted. See Figure 1 for examples displaying low exposure and/or contrast. Next, we randomly select four sets of eight images and send them to four different DTP experts for editing with Adobe Photoshop. Our goal is to let the experts work as freely as possible so we can learn their different interpretations of editing intentions. We therefore asked the experts to process each photo with the intention to invoke a “cool”, “warm”, “technically correct”, and “natu-

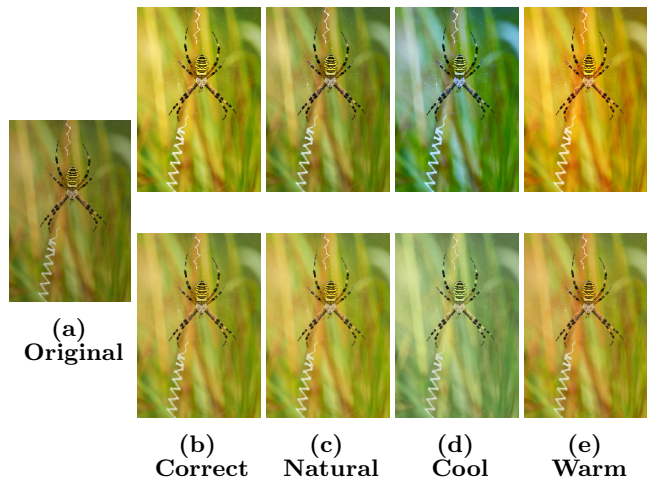


Figure 2: Example of an original photo and two different expert interpretations (top and bottom) of four intentions. Photo ©Ricoh Company Ltd., Tokyo Japan.



Figure 3: Example for image retrieval from the expert data base. (a) © © © Anne Arnould [1]. (b) and (c) ©Ricoh Company Ltd., Tokyo Japan.

ral” look, because they can be explained without complex instructions. Moreover, any detailed description of these looks was specifically omitted to let the experts follow their personal intuition. This results in a total of 128 professionally edited samples representing our data base. By that, different interpretations of the same intentions were achieved, as is illustrated in the example in Figure 2.

Assuming that there exists a photometrically similar entry in our knowledge base, it can now be utilized to enhance user photos exactly like experts would. To find suitable samples, we consider four photometric properties of a photo: contrast (via Tamura texture features), brightness, saturation and hue profile. They are represented by the aforementioned saliency weighted histograms. The retrieval is implemented via a nearest neighbor search on the individual histograms using late fusion with uniform weights.

In addition to selecting photometrically similar photos, we assume that there exists also a correlation between the visible scenery and its corresponding expert operations. To evaluate this hypothesis, the system allows the detection of scenery types in a user photo and delivers an according selection of entries from our knowledge base. Each type of scenery is handled by a specific set of nearest-neighbor

classifiers using color histograms, color correlograms, and tiny color images as feature representations with late fusion.

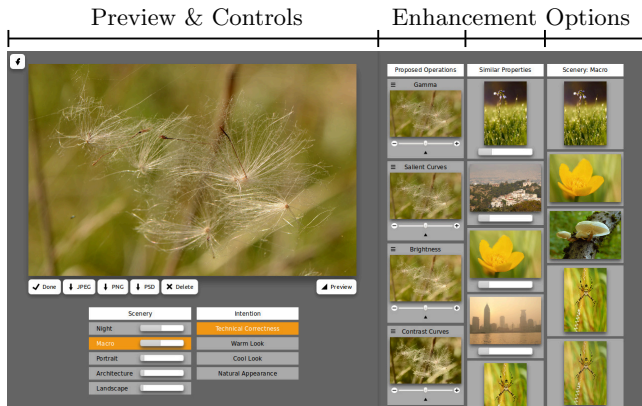
Figure 3 shows an example of a possible retrieval result for a given user image, where the query image (3a) represents a macro photo with green grass in the background. The most similar data base photo (3b) also shows a macro image with grass. With respect to the displayed concept, our data base also provides a macro photo (3c) showing a matching background.

Once candidate photos are available, we can start extracting the list of operations applied by the experts from the original Photoshop file. This is achieved by parameter parsing of the contained adjustment layers. Our system is capable of replicating the set of operations applied by the four DTP experts and reapply them to user photos.

## 4. SYSTEM DESIGN

The system follows a client-server architecture, where the back-end provides the previously described detection, retrieval, and storage services, while the user interface and workflow management are managed by the front-end, realized as a web application running in any modern browser. After selecting a photo for processing, the client requests the following information:

- detected EPs
- detected scenery
- photometrically similar photos
- photos showing the same scenery
- user selected sceneries
- user selected intentions
- preview version of the photo
- thumbnails for enhancement and expert photos



**Figure 4: Screenshot of the RSIP system in enhancement mode showing preview & controls along with enhancement options. Enhanced photo by Arnould [1].**

Once all required information is available, the system enters enhancement mode as shown in Figure 4. The left side is dedicated to a large preview of the current result and controls for saving, downloading, and changing scenery and intention. Enhancement options are organized in columns

on the right. Our bank of EP detectors delivers a list of applicable enhancements with suggested initial strength. They are shown in the leftmost column and are initially disabled. Users can then enable individual enhancements at their discretion, change the strength by up to  $\pm 25\%$  of the full range, and reorder them. The other columns contain the list of photometrically similar photos ordered from lowest to highest distance (center), and photos showing the same type of scenery (right). A simplified workflow of a user uploading and enhancing a photo consists of the following steps:

- 1: Upload & select photo
- 2: Enable or adjust enhancement, or select knowledge base photo
- 3: Check preview
- 4: If satisfied save, else goto 2

Note that selection of enhancements and expert photos is exclusive by design: The client prevents simultaneous selections of both enhancement operations and knowledge base photos. Additionally, only a single retrieved photo can be selected at a time. These limitations are imposed by the client – not the back-end – to prevent confusion about the order of applied operations. We may in the future allow users to first select an expert photo and subsequently apply additional operations.

Every time the user selects or adjusts an enhancement or changes the correction order, the preview is updated and a new EP detection is performed given the current state.

## 5. EVALUATION

### 5.1 Study Setup

A case study with twelve subjects was conducted to evaluate the proposed system. All subjects are university students, both female and male, of ages 21 to 27, with no or minor experience in the field. Each subject is asked to enhance the same set of provided photos. The client is displayed in fullscreen mode on a Dell UltraSharp U3014 monitor (30in) on factory settings. The test hardware is placed in a specially prepared, sound dampened and darkened room at Ghose Lab at TU Kaiserslautern.

The procedure for each subject is as follows:

- Brief introduction to the system, including: different screens, general enhancement workflow, photo selection, changing intentions, saving etc.
- Introduction to EP detectors option
- Part one: use EP detectors for 15 minutes
- Answer questionnaire for part one
- Introduction to expert knowledge option
- Part two: use expert knowledge for 15 minutes
- Answer questionnaire for part two

The statements in both parts of the questionnaire aim to determine subject satisfaction with respect to usability and achieved result. For that purpose the Likert scale (1-7) is being utilized. Furthermore, to gather objective data, all subject-performed actions are logged by the system and linked to each individual. Thereby, the total experiment time per subject is limited to one hour.



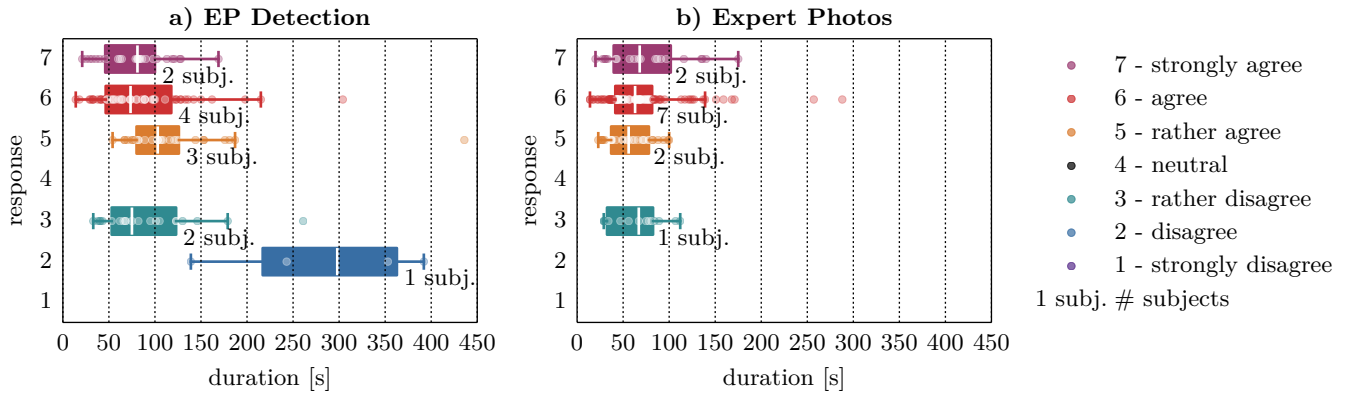


Figure 5: Satisfaction with respect to processing duration of individual edit sessions when using either approach, (a) EP detection or (b) expert knowledge, exclusively. The majority of subjects agrees to being satisfied with the achieved result. Subjects spending less time for adjusting enhancements with the EP detection approach are more satisfied, while no correlation is found for the expert knowledge method.

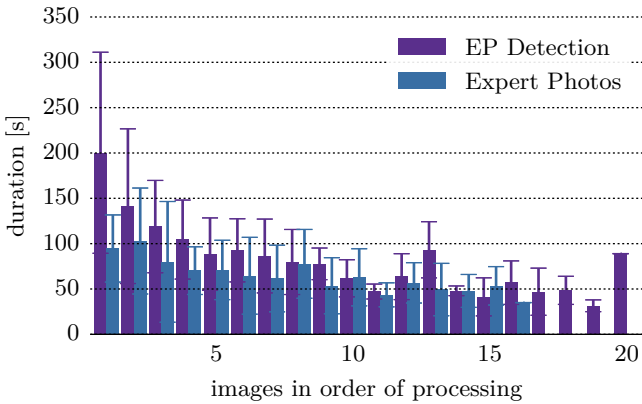


Figure 6: Progression of time taken by subjects to edit photos using either EP detectors or expert photos as operation source.

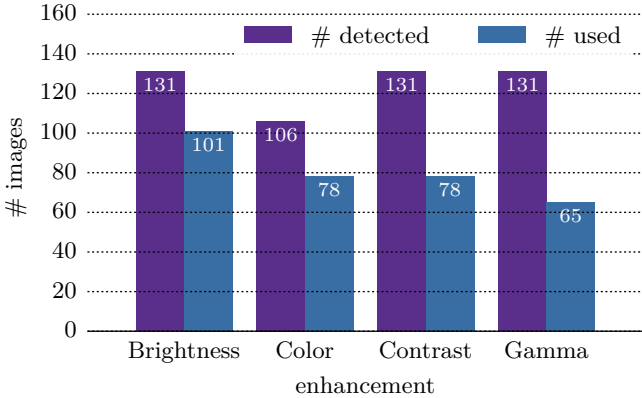


Figure 7: Utilization of available enhancement operations.

## 5.2 Results

When designing the system described before, our first and foremost goal was to create a satisfying experience for its users. Therefore, both parts of the questionnaire contain a

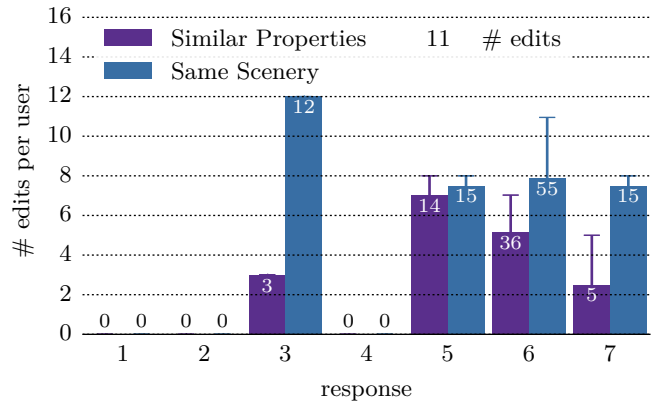


Figure 8: Responses to the statement “I am completely satisfied with the results” and the corresponding utilization of the two expert photo sources. Bars show count per user, numbers give raw count.

statement “I am completely satisfied with the achieved results”. Responses to these statements, along with time taken to edit photos, are illustrated in Figure 5. The majority of subjects agreed with the statement, resulting in a median Likert score of 5.5 for the EP approach and 6 for expert knowledge. Using EP detection, subjects tend to give more positive responses if they spent less time on adjustments. however, no correlation could be found when utilizing expert knowledge, partially because no subject being truly dissatisfied with the achieved result. The learning rate was found to be fast for both approaches as compared in Figure 6. Please note that subjects had to learn both, using the system in general, as well as the EP method. Still, after three to five processed images only small differences can be observed and the strongest learning effects occurs until ten. We can see that, even though there are more degrees of freedom, EP detection can be utilized at the same speed as reusing expert knowledge. All proposed enhancements were detected multiple times and also utilized by the study subjects. Figure 7 shows the number of times each operations was available and used in the final result. Image brightness was used most often, followed by color and contrast. Even though brightness

was detected less often. Finally, gamma correction was used the least. Comparing the two expert photo sources, similar properties and same scenery, Figure 8 displays the statistics of their utilization with respect to satisfaction. While both were utilized, more satisfied subjects tend to utilize same scenery more frequently.

## 6. DISCUSSION & FUTURE WORK

We aimed to create an intelligent and easy-to-use system for photo editing. The conducted user study has demonstrated a high learning rate for both provided approaches when using the system. Most learning effects are achieved after five trials, while after ten only small improvements in edit duration can be observed between both methods. The study also revealed that both implemented approaches for image enhancement, namely EP detection and reuse of DTP expert knowledge, provide satisfactory results for the majority of subjects, although they were more satisfied when using the latter. Given a choice between the two provided knowledge sources, subjects preferred scenery detection over similar properties. We believe that they found it easy to identify the scenery, but could not precisely judge the relevant photometric properties contrast, brightness, color, and hue. Subjects then tended to reject similar photos based on differing scenery alone. However, samples retrieved using similar properties were still chosen for a significant part of edits, thereby justifying the existence of both sources. Future systems may combine both ideas simultaneously, i.e., searching for similar entries in the set of same scenery photos. This, however, requires a significantly larger knowledge base than we were able to gather so far. Additionally, it could be shown that user satisfaction correlates with the time needed for processing an image when using EP detection. Subjects that take longer to process a photo appear to be less satisfied. With respect to the individual methods, it was demonstrated that for EP detection all enhancements have been used. Brightness was applied most frequently, followed by contrast and color, and finally gamma. We believe that we can employ the data we gathered in the case study to create more accurate EP detectors and suggest more satisfying enhancement values. One way of achieving this would be to use the described DPChallenge data set for more of the training. This requires a thorough categorization of challenge themes first to ignore entries that introduce bias. For instance, challenges that require a particular style or technique of photography, or abstract content, which is outside the scope of this work.

## 7. ACKNOWLEDGMENTS

This work is partially funded by the German Federal Ministry for Education and Research (BMBF) projects “HySociaTea” (grant no. 01IW14001) and “Multimedia Opinion Mining” (grant no. 01WI15002). We thank our partner Ricoh for their expertise and financing for the joint “RicoH Semantic Image Processing” (RSIP) project that contributed to the creation of the described system and expert knowledge base. Our thanks also go to Ghose Lab at TU Kaiserslautern<sup>7</sup> for their help with the design of our questionnaire and the opportunity to conduct our case study in their laboratories.

<sup>7</sup><http://www.sowi.uni-kl.de/psychology-of-perception>

## 8. REFERENCES

- [1] A. Arnould. Trapped in a spider web ... Copyright Information: Creative Commons Attribution-ShareAlike 2.0 Generic (CC BY-SA 2.0), see <https://creativecommons.org/licenses/by-sa/2.0/legalcode>, August 2014. URL [https://www.flickr.com/photos/anne\\_arnould/15135738995/](https://www.flickr.com/photos/anne_arnould/15135738995/). Photograph.
- [2] P. Haeberli and D. Voorhies. Image processing by linear interpolation and extrapolation. *IRIS Universe Magazine*, 28:8–9, 1994.
- [3] C.-W. Hsu, C.-C. Chang, and C.-J. Lin. A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University, 2003. URL <http://www.csie.ntu.edu.tw/~cjlin/papers.html>.
- [4] I.-S. Jang, T.-H. Lee, H.-G. Ha, and Y.-H. Ha. Adaptive color enhancement based on multi-scaled retinex using local contrast of the input image. In *Optomechatronic Technologies (ISOT), 2010 International Symposium on*, pages 1–6, Oct 2010. doi: 10.1109/ISOT.2010.5687343.
- [5] I.-S. Jang, W.-J. Kyung, T.-H. Lee, and Y.-H. Ha. Local contrast enhancement based on adaptive multiscale retinex using intensity distribution of input image. *Journal of Imaging Science and Technology*, 55(4):40502–1, 2011.
- [6] Y. Jin, L. M. Fayad, and A. F. Laine. Contrast enhancement by multiscale adaptive histogram equalization. In *International Symposium on Optical Science and Technology*, pages 206–213. International Society for Optics and Photonics, 2001.
- [7] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *Image Processing, IEEE Transactions on*, 6(7):965–976, 1997.
- [8] D. Joshi, R. Datta, E. Fedorovskaya, Q.-T. Luong, J. Wang, J. Li, and J. Luo. Aesthetics and emotions in images. *Signal Processing Magazine, IEEE*, 28(5): 94–115, Sept 2011. ISSN 1053-5888. doi: 10.1109/MSP.2011.941851.
- [9] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987.
- [10] H. Tamura, S. Mori, and T. Yamawaki. Textural features corresponding to visual perception. *Systems, Man and Cybernetics, IEEE Transactions on*, 8(6): 460–473, 1978.
- [11] J. Zhang and S. Sclaroff. Saliency detection: A boolean map approach. In *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, 2013.