Semantic Awareness for Automatic Image Interpretation

THÈSE N^{O} 5635 (2013)

PRÉSENTÉE LE 1ER MARS 2013

À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS

GROUPE IMAGES ET REPRÉSENTATION VISUELLE

PROGRAMME DOCTORAL EN INFORMATIQUE, COMMUNICATIONS ET INFORMATION

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Albrecht Johannes LINDNER

acceptée sur proposition du jury:

Prof. P. Dillenbourg, président du jury
Prof. S. Süsstrunk, directrice de thèse
Prof. J. Allebach, rapporteur
Prof. R. Hersch, rapporteur
Prof. P. Le Callet, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2013

Acknowledgements

I have had a wonderful time at EPFL, which would not have been possible without all the many people who accompanied me on my way. In the following I would like to thank them for all their support that I enjoyed in both my professional and personal life.

First, I thank my supervisor Prof. Sabine Süsstrunk who sketched out a visionary topic for my thesis and set me on the right track at the beginning and whenever it was necessary. Besides her qualities as an academic supervisor she was also a great mentor. My start at EPFL was not linear as it took me a while to find a project I felt comfortable with. The door she opened was a great opportunity and was the starting point for this thesis. It is with immense gratitude that I acknowledge her faith she had in me to offer a position in her lab and all the support I got from her in the subsequent years.

The other important person who guided me in this thesis is Dr. Nicolas Bonnier from Océ whom I know from my time in Paris when he supervised me during my Master project. He established the connection between Sabine and me and spoke up for me at Océ to grant the PhD funding. During my PhD I enjoyed seeing him for a week every other month in Paris and he often gave me good advice. I am very thankful for this and I am glad that we transitioned from our professional relationship to friendship.

At this point I also want to acknowledge Océ for the funding and in particular Christophe Leynadier, the leader of the research and development division onsite in Paris.

It was an honor to have a jury of distinguished professors on my doctoral committee: Prof. Jan P. Allebach, Prof. Pierre Dillenbourg, Prof. Roger Hersch, and Prof. Patrick Le Callet. I thank them for their time reading my thesis, the interesting discussion and all the valuable feedback they gave.

I further want to acknowledge my colleagues' contributions. Dr. Radhakrishna Achanta was always a good reference to discuss new ideas that he easily grasped and developed. He often gave me good advice and I value his opinion. Dr. Appu Shaji pointed out a simple yet effective way to greatly improve the scalability of the statistical framework. This allowed for a significant

leap forward. I further had many fruitful discussions about statistics with Dr. Jayakrishnan Unnikrishnan that helped to deepen my understanding. I always enjoyed a coffee break with Dominic Rüfenacht and appreciated the interesting discussions we had about technical and non-technical topics.

I thank Kristyn Falkenstern for the paper we wrote together in which we joined our work. I also thank the students that did semester projects with me: Mehmet Candemir, Gökhan Yildirim (good to have you as a colleague now), Yves Lange, Bryan Zhi Li, Anaëlle Maillard, and Pierre-Antoine Sondag. These projects were a great experience for me and contributed in one way or the other to this thesis, especially Bryan's work on color naming.

A great thank you goes to the secretaries that keep the group functioning in the background and always have a smile on the face: Jacqueline Aeberhard and Virginie Rebetez.

On a personal side I first want to thank my girlfriend Paola for her patience when I had to work longer, for the good moments we had together and for the love we share.

Finally and most important I want to thank my parents who did a remarkable effort to provide me a good education. It is thanks to their enduring support that I could follow and develop my interests and ultimately complete this thesis. I thus dedicate this thesis to my parents.

Lausanne, January 14, 2013

Abstract

Finding relations between image semantics and image characteristics is a problem of long standing in computer vision and related fields. Despite persistent efforts and significant advances in the field, today's computers are still strikingly unable to achieve the same complex understanding of semantic image content as human users do with ease. This is a problem when large sets of images have to be interpreted or somehow processed by algorithms. This problem becomes increasingly urgent with the rapid proliferation of digital image content due to the massive spreading of digital imaging devices such as smartphone cameras.

This thesis develops a statistical framework to relate image keywords to image characteristics and is based on a large database of annotated images. The design of the framework respects two equally important properties. First, the output of the framework, i.e. a relatedness measure, is compact and easy-to-use for subsequent applications. We achieve this by using a simple, yet effective significance test. It measures a given keyword's impact on a given image characteristic, which results in a significance value that serves as input for successive applications. Second, the framework is of very low complexity in order to scale well to large datasets. The test can be implemented very efficiently so that the statistical framework easily scales to millions of images and thousands of keywords

The first application we present is semantic image enhancement. The enhancement framework takes two independent inputs, which are an image and a keyword, i.e. a semantic concept. The algorithm then re-renders the image to match the semantic concept. We implement this framework for four different tasks: tone-mapping, color enhancement, color transfer and depth-of-field adaptation. Unlike conventional image enhancement algorithms, our proposed approach is able to re-render a single input image for different semantic concepts, producing different image versions at the output to reflect the image context.

We evaluate the proposed semantic image enhancement with two psychophysical experiments. The first experiment comprises almost 30'000 image comparisons of the original and the enhanced images. Due to the large scale, we crowdsourced the experiment on Amazon Mechanical Turk. The majority of the

enhanced images was proven to be significantly better than the original images. The second experiment contains images that were enhanced for two different keywords. We compare our proposed algorithm against histogram equalization, Photoshop auto-contrast and the original. Our proposed method outperforms the others by a factor of at least 2.5.

The second application is color naming, which aims at relating color values to color names and vice versa. Whereas conventional color naming depends on psychophysical experiments, we are able to solve this task fully automatically using the significance values. We first demonstrate the usefulness of our approach with an example of 50 color names and then extend it to the estimation of memory colors and color values for arbitrary semantic expressions. In a second study, we use a list of over 900 English color names and translate it to 9 other European and Asian languages. We estimate color values for these over 9000 color names and analyze the results from a language and color science point of view.

Overall, we present a statistical framework that relates image keywords to image characteristics and apply it to two common imaging applications that benefit from a semantic understanding of the images. Further we outline the applicability of the framework to other applications and areas.

Keywords: semantic image understanding, image enhancement, data mining, statistics, large scale, crowd-source, color naming.

Zusammenfassung

Relationen zwischen Bildsemantik und Bildcharacteristika zu finden ist ein seit langem anhaltendes Problem des maschinellen Sehens und verwandten Gebieten. Trotz beharrlicher Anstrengung und beachtlichem Fortschritt sind heutige Computer immer noch auffallend unfähig ein komplexes semantisches Bildverstehen zu erreichen das an dasjenige von Menschen heranreicht. Dies ist ein Problem wenn große Bildmengen interpretiert oder anderweitig bearbeitet werden müssen. Dieses Problem wird wegen der massiven Verbreitung von digitalen Kameras zunehmend dringend.

Diese Promotion entwickelt ein statistisches System um Schlüsselworte mit Bildcharakteristika zu verknüpfen und fußt auf einer großen Datenbank annotierter Bilder. Das Design des Systems respektiert zwei gleich bedeutende Eigenschaften. Erstens ist das Ergebnis des Systems, eine Relationsmessung, kompakt und für nachfolgende Applikationen einfach zu nutzen. Dies erreichen wir mit einem einfachen, aber effektiven Signifikanztest. Er misst die Beeinflussung einer gegebenen Bildcharakteristik durch ein Schlüsselwort und resultiert in einem Signifikanzwert. Zweitens ist die Komplexität des Systems klein, um eine einfache Skalierung zu großen Datenbanken zu ermöglichen. Der Test kann sehr effizient implementiert werden sodass das System sehr einfach zu Millionen von Bildern und Tausenden von Schlüsselwörtern skaliert.

Die erste Applikation die wir vorstellen ist semantische Bildverbesserung. Das Verbesserungssystem hat zwei unabhängige Eingänge: ein Bild und ein Schlüsselwort (semantischer Kontext). Der Algorithmus rendert das Bild um es dem semantischem Kontext anzupassen. Wir implementieren das System für vier verschiedene Anwendungen: Tonwertkorrektur, Farbverbesserung, Farbtransfer und Adaptierung von Tiefenunschärfe. Im Gegensatz zu konventionellen Bildverbesserungsalgorithmen ist unser vorgeschlagene Ansatz in der Lage ein Bild für verschiedene semantische Kontexte zu rendern. Dies resultiert in verschiedenen Versionen des Bildes die den jeweiligen semantischen Kontext wiedergeben.

Wir evaluieren das semantische Bildverbesserungssystem mit zwei psychophysischen Experimenten. Das erste Experiment umfasst nahezu 30'000 Bilderver-

gleiche zwischen der originalen und der verbesserten Version. Wir benutzten Crowdsourcing auf Amazon Mechanical Turk für das Experiment aufgrund der hohen Anzahl an Vergleichen. Der Großteil der verbesserten Bilder wurde für signifikant besser befunden als die originalen Bilder. Das zweite Experiment beinhaltet Bilder die für zwei verschiedenen Kontexte verbessert wurden. Wir vergleichen unsere Methode mit Histogrammegalisierung, Photoshop autocontrast und dem Original. Unsere Methode überragt die anderen um einen Faktor von mindestens 2.5.

Die zweite Anwendung is Frabnahmengebung wobei Farbnahmen mit Farbwerten verknüpft werden. Im Gegensatz zu konventionellen Methoden die auf psychophysischen Experimenten beruhen ist unser Ansatz voll automatisch aufgrund der Signifikanzwerte. Wir demonstrieren die Nützlichkeit zuerst anhand von 50 Farbnahmen und erweitern dann zur Schätzung von *memory colors* und Farbwerten willkürlicher semantischer Ausdrücke. In einer zweiten Studie übersetzen wir eine Liste von über 900 englischen Farbnahmen in neun andere europäische und asiatische Sprachen. Wir bestimmen dann Farbwerte für diese über 9000 Farbnahmen und analysieren die Ergebnisse aus einer linguistischen und einer farbwissenschaftlichen Perspektive.

Insgesamt präsentieren wir ein statistisches Systemn das Schlüsselwörter und Charakteristiken von Bilder verknüpft und wenden es auf zwei weit verbreitete bildbezogene Applikationen an die von einem semantischem Verstehen des Bildinhalts profitieren. Desweiteren umreißen wir die Ausweitung auf andere Anwendungen und Gebiete.

Schlüsselwörter: semantisches Bildverstehen, Bildverbesserung, data-mining, Statistik, Skalierung, crowd-source, Farbnahmengebung.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

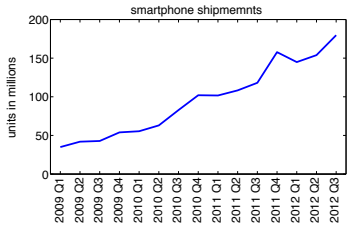
Digital image and video capturing devices are omnipresent in modern life. One example for the widespread distribution of such devices is the thriving smartphone market as shown in Figure 1.1(a). It makes cameras more accessible than ever before, because people carry their phones with them most of the day. An estimate of the total number of photos taken per year is shown in Figure 1.1(b). The curve is growing exponentially ever since photography started in the 1820s and there is no sign of saturation, yet.

More and more of these images are stored in online databases of tremendous scale. Flickr, an online image sharing community, reports the number of uploads in the last minute on their web page, which usually varies between two and three thousand[1]. This extrapolates to an order of magnitude of 1 billion images per year. Facebook, a social networking service, reports that 250 million photos were uploaded every day during the last three months of 2011 [24]. In total, Facebook stores "more than 100 petabytes (100 quadrillion bytes) of photos and videos" on their servers.

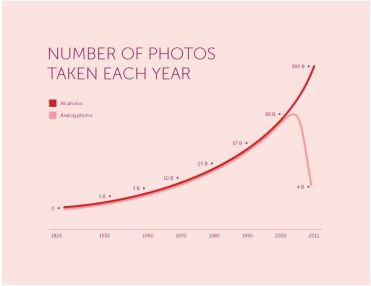
The vast amount of images and videos drives the development of new technologies to handle the data in a more automatic fashion. It is desirable to build computer systems that assist users to archive, query, retrieve, edit, interpret, or understand multimedia data. But this turns out to be an extremely difficult challenge.

Let us consider that a photographic artist is given the left image reproduced in Figure 1.2 with the request to "smoothen the woman's skin and add some appealing gloss to it, sharpen her cheek- and jawbone to make her appear thinner and more strict, make her hair stand up as a prolongation of her viewing direction, remove the fold on her dress and exchange the background with some smooth glow that directs the viewer's focus to the center." The artist is able

[1] http://www.flickr.com/photos/



(a)



(b)

Figure 1.1: Left: Shipments of the top 5 smartphone vendors per annual quarter. The thriving market exemplifies the omnipresence of image and video capturing devices in our daily life. The peaks in the fourth quarter every year are due to Christmas sales. Source: "IDC Worldwide Quarterly Mobile Phone Tracker" reports [38]. Right: An estimate of the number of photos taken each year. Figure reproduced from Jonathan Good [31].

to understand the message and alter the image accordingly as shown on the right hand side. On the contrary, this task as a whole exceeds today's computer systems' capabilities by far, even though the one or the other sub-task (i.e. skin smoothing) can be done automatically.

Figure 1.2: Image before and after a Photoshop® touch-up by a graphic artist. Today's algorithms are far from achieving such complex image transformations automatically. Photo attribution: Patrick Rigon.

This gap between humans' and computers' understanding of objects is referred to as the "semantic gap" and illustrates today's computers striking inabil-

ity to achieve the same complex semantic understanding of multimedia content as human users do. It is due to the fact that a given object is described by humans in a natural language and by computers in a numeric language. Linking these two worlds, i.e. bridging the semantic gap, is the ultimate goal of research in computer vision and related domains. This is challenging because human language is a vast space as attested by the authors of the Oxford English Dictionary [71]:

> This suggests that there are, at the very least, a quarter of a million distinct English words, excluding inflections, and words from technical and regional vocabulary not covered by the OED, or words not yet added to the published dictionary, of which perhaps 20 per cent are no longer in current use. If distinct senses were counted, the total would probably approach three quarters of a million.

Finding links between these hundreds of thousands of words and billions of images is an undertaking of tremendous scale and demands novel algorithms specifically designed for keeping computational costs within reasonable bounds.

Even if a system that links the digital and the human languages can be built, there are further challenges to make it useful in our daily life. It is acceptable if the learning of the links takes some hours to days on a powerful desktop or server computer, but an application for end users has to be light-weight and provide real-time feedback on the user's input. It is thus not possible to store a large image database on the user's device and derive an appropriate action from it every time the user inputs a semantic expression. Instead, the links have to be pre-computed and stored in a compact form that makes them easy and instantaneously accessible to subsequent applications.

1.1 Thesis Goals

The goals of this thesis can be summarized in two parts:

1. Develop a framework that links semantic expressions to image characteristics, i.e. bridges the semantic gap. Its design respects two equally important properties: it has to provide an efficient interface to make the semantic links accessible to subsequent applications and it has to easily scale to large vocabularies and image databases.

2. Adopt the previously learned links for common image-related applications in order to achieve a semantic awareness of the scenes. This added semantic component improves the applications over the state-of-the-art.

1.1.1 First goal: Bridging the semantic gap

We achieve the first goal with a statistical framework that is based on a large database of annotated images. It uses a light-weight statistical test to determine a keyword's significance for a given image characteristic. The estimated significance values can be positive or negative and indicate whether a characteristic is dominantly present or absent in an image. An example is reproduced in Figure 1.3 that shows how the semantic expression *flower* is linked to the spatial distribution of lightness in images. The positive values in the middle indicate that *flower* images are generally brighter in the middle. The negative values in the surrounding indicate that *flower* images tend to be darker along the image borders, especially at the top.

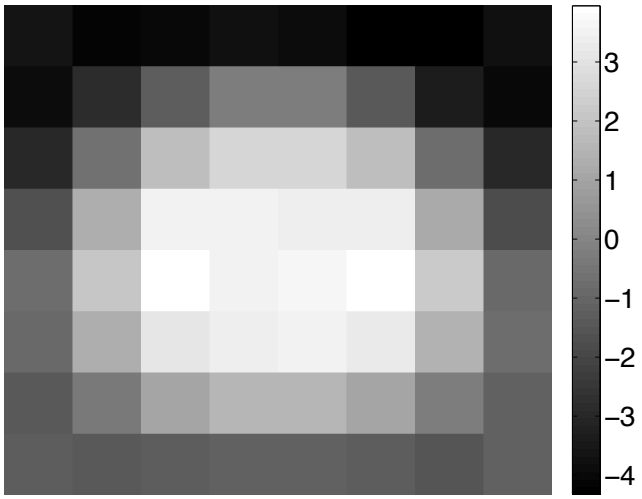


Figure 1.3: The statistical framework is able to link semantic expressions with image characteristics. This example shows the result for the keyword *flower* and a coarse lightness layout descriptor. The bright (positive) values indicate that *flower* images tend to be brighter in the middle than a non-*flower* image. The dark (negative) values indicate regions where *flower* images are generally darker.

1.1.2 Second goal: Applications

The first application, semantic image enhancement, has the goal to re-render an image for a given semantic context. Figure 1.4 shows two examples of semantic image enhancement. Our algorithm takes as inputs the images on the left together with their keywords *autumn* and *macro*, respectively. It then enhances the images in order to emphasize the associated semantic context; in the first

case the colors are enhanced and in the latter the depth-of-field. It is important to note that the method we developed handles arbitrary keywords of an unlimited vocabulary.

(a) input image (b) enhanced for *autumn*

(c) input image (d) enhanced for *macro*

Figure 1.4: Examples for semantic image enhancement. Top: the input image's color are adapted to the semantic context *autumn*. Bottom: the input image's depth-of-field is reinforced to better match the semantic context *macro*. Photo attributions left: * *starrynight1* (Flickr) and right: Zhuo and Sim [114].

The task in color naming, our second application, is to find a color name given a color value or vice versa. Traditionally, color naming is done with psychophysical experiments in which observers have to name different color patches. The statistical framework allows us to discard any human intervention and solely

rely on annotated images from the world wide web. Figure 1.5 shows a distribution in CIELAB color space that estimates how much each color is related to the color name *periwinkle blue*. The estimated color values in sRGB color space for *periwinkle blue* are 139, 150, 209. The scalability of the automatic framework allows us to estimated more than 9000 colors in 10 different languages.

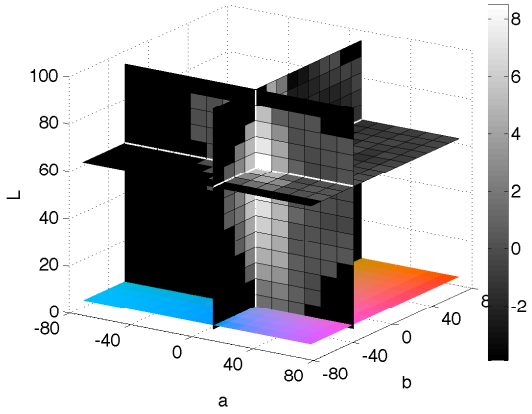


Figure 1.5: A distribution in CIELAB color space for the color name *periwinkle blue* computed automatically with annotated images from the world wide web. The estimated color is at the crossing of the three orthogonal planes and the corresponding sRGB values are 139, 150, 209.

1.2 Thesis outline

This thesis is structured as follows:

- Chapter 2: State-of-the-art
  We discuss relevant work of the related fields of this thesis.

- Chapter 3: Linking Words with Characteristics
  This chapter introduces the scalable statistical framework that is used for the subsequent applications. We use in this chapter the MIR Flickr database with 1 Million annotated Flickr images [36].

- Image enhancement applications:

  - Chapter 4: Semantic Tone-Mapping
    Our first semantic image enhancement application is tone-mapping. We present the complete enhancement pipeline in detail and prove its superiority over the state-of-the-art with psychophysical experiments.
  - Chapter 5: Additional Semantic Re-rendering Algorithms
    This chapter introduces two other semantic re-rendering algorithms. They are based on the same generic framework presented in the previous chapter, but adapt an image's color or depth-of-field to a semantic expression, respectively.

- Color naming applications:

  - Chapter 6: Color Naming
    In this chapter we explain how the statistical framework can be used for automatic color naming. We demonstrate its functioning with an example of 50 English color names and memory colors.
  - Chapter 7: A Large-Scale Multi-Lingual Color Thesaurus
    Finally we extend the color naming to over 9000 color names in 10 different languages. We discuss the accuracy of the estimations and perform further advanced analysis of the data. We also present the color thesaurus web page.

- Chapter 8: Conclusions
  Reflections on this thesis and future work.

1.3 Contributions

We present a highly scalable statistical framework that computes a keyword's impact on image characteristics. The performance is demonstrated on a database

with millions of images and thousands of keywords as well as with images downloaded from Google Image Search. We implemented characteristics for color, local image structure, global spatial layout and characteristics computed in the Fourier domain.

We propose a semantic image enhancement pipeline that re-renders an image with respect to a semantic context. We implement the semantic image enhancement for tone-mapping, color enhancement, color transfer and depth-of-field adaptation.

We apply the scalable statistical framework also to color naming, where color names have to be matched to color values. Unlike traditional psychophysical experiments, our approach solves the task completely automatically with images downloaded from Google Image Search. We do estimations for over 9000 color names in 10 European and Asian languages to demonstrate the performance of our method.

Chapter 2

State-of-the-Art

This thesis covers a variety of research fields that are introduced in this state-of-the-art overview. We start with a discussion on semantic and numeric image descriptors in Section 2.1. The following Section 2.2 introduces hypothesis testing, which forms the mathematical basis of the statistical framework presented in this thesis. The statistical framework employs statistical tests to infer a keyword's impact on a characteristic. This field is referred to as data-mining and is presented in Section 2.3.

The fist application we present image enhancement. We introduce the state-of-the-art of image enhancement in Section 2.4. As we conducted psychophysical experiments to validate our framework we present this topic in Section 2.5, including a discussion on crowd-sourced psychophysical experiments on the world wide web. The second application is color naming and we present the corresponding state-of-the-art in Section 2.6. Memory colors, a closely related field, is discussed in Section 2.7.

2.1 Image descriptions

There are two fundamentally different ways to describe an image: a semantic description done by a human being and a numeric description generated by a computer. The conceptual difference between them is generally denoted as the "semantic gap" [88]. It describes the difficulty in linking the two worlds in order to realize computer programs with a semantic understanding of image content. As this thesis deals with bridging the gap between these two types of descriptors, we discuss both of them.

2.1.1 Semantic description: image keywords

Semantic image descriptions, i.e. image metadata, are standardized by the International Press Telecommunications Council (IPTC) [82]. The "IPTC header" contains different fields such as title, scene code, description, and keywords, which can contain additional semantic information for an image. Keywords are defined by the IPTC as follows [82]:

> Keywords to express the subject of the image. Keywords may be free text and don't have to be taken from a controlled vocabulary.

The uncontrolled vocabulary enables users to freely express their thoughts when looking at an image. This can provide more information than a field with a controlled vocabulary such as the scene code [82]. Two example images with annotations from the MIR Flickr database [35, 36] are reproduced in Figure 2.1. Note that the annotation in Figure 2.1(b) contains mixed English and Spanish keywords.

(a) *chicago*, *hancock*, *cloud*, *skyscraper*

(b) *luces*, *lights*, *car*, *coche*, *choes*, *cars*, *a5*, *dirección*, *madrid*, *alcorcón*, *explore*, *long-exposure*, *luz*, *light*

Figure 2.1: Two example images with their annotations from the MIR Flickr database [35, 36]. As the vocabulary is uncontrolled, users can use languages other than English. Photo attributions left: Martin Griffith and right: David Cornejo.

In the context of this thesis, we focus on image keywords because they are abundantly available on the internet for free. Online image sharing communities stimulate social tagging, which provides a rich resource for semantic research. For example, Flickr makes its database accessible via a public API, where images can be downloaded together with their annotations and other metadata (see Section 2.3).

There are other sources for semantic image descriptions such as the image filename or text in the local surrounding of the image in a compound document as is often the case e.g. on web pages, in magazines, or in this very thesis. In fu-

ture work, the methods can potentially be extended to handle entire paragraphs of text.

Image semantics is a rapidly growing research field. The computer vision community regularly competes on open databases to measure and compare their algorithms that aim at detecting an image's semantic content. Examples are the *Pascal Visual Object Classes Challenge* [3], the *Image Cross Language Evaluation Forum* [1] or the *ImageNet Large Scale Visual Recognition Challenge* [2, 20]. The databases that are provided for these challenges contain manually annotated images for algorithm training and testing. It is worth pointing out that the last-cited challenge, with its 14,197,122 images in 21,841 classes[1], is significantly larger than the others.

The manual annotation of large images databases is a tedious task because it is monotonic and repetitive. LabelMe [84] is a project that facilitates image labeling with adequate graphical user interfaces and the option to incorporate it into Amazon Mechanical Turk, a crowd-sourcing internet marketplace for human labour (see also Sec. 2.5). An interesting approach is the ESP game [100] that turns a labeling task into an online game.

We do not use images from computer vision challenges because they are specifically designed for this task. Instead we use images from the world wide web or Flickr, an online image sharing web page for amateurs and professionals. This assures us that the developed methods work with everyday images.

2.1.2 Numeric description: image characteristics

A numeric image description is a vector that describes certain image characteristics and is extracted from an image by an algorithm. There are numerous descriptors that are designed for different purposes. Descriptors can be designed for single keypoints such as SIFT [54] or linear binary patterns [56], two- and three-dimensional shapes such as the MPEG-7 visual shape descriptors [10], or entire images such as SIFT codebooks [97] or the MPEG-7 color layout descriptor [15].

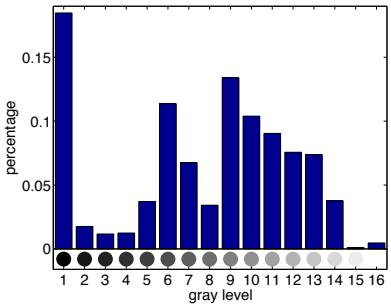
In the context of this thesis, we focus on low-level image characteristics due to the two applications we study: automatic color naming and semantic image enhancement. The first is based on simple color histograms and the latter is based on characteristics that can be altered using common image processing algorithms.

Two example descriptors that we implemented are shown in Figure 2.2, which are gray-level histograms and lightness layout descriptors. The layout descriptor stems from a regular 8×8 grid superposed over the image independent of its size or aspect ratio. Computing each grid cell's average value of the

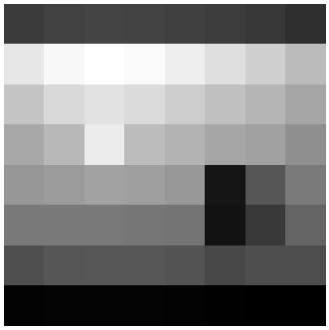
[1] state: October 2012

respective characteristic, leads to a 64-dimensional layout descriptor. Note that a regular 8×8 grid is also used in the MPEG-7 standard for the color layout descriptor [15]. The advantage of this gridding is that it guarantees invariance to an image's resolution and aspect ratio.

(a) example image



(b) gray-level histogram



(c) spatial lightness layout

Figure 2.2: Example image (top) and its gray-level characteristics (bottom). For the spatial layout characteristic, an 8 × 8 grid is superposed on the image and the average lightness value per grid cell is computed. Photo attribution: Atilla Kefeli.

Numeric image descriptors are denoted with different terms in the literature such as "descriptor", "feature", or "characteristic". In this thesis we use the term "characteristic", because it covers two contexts in which it is used: image description and image processing (see Chapters 4 and 5). In the first context we say "We compute characteristic x of the image" and in the latter "We change characteristic x of the image".

In a recent study Deselaers and Ferrari demonstrated that there is a direct relation between numeric and semantic image descriptors [21]. They computed image descriptors for all images in ImageNet [2] and made an interesting observation: the more two images are semantically related, the more their visual descriptors are similar. We observe the same tendency, which leads to statisti-

cally significant patterns that we exploit for different applications presented in Chapters 4, 5 and 6.

2.2 Statistical hypothesis testing

Our framework that relates image keywords to image characteristics in Chapter 3 uses statistical tests. A statistical hypothesis test verifies whether a result is statistically significant, i.e. whether it is unlikely that an outcome happened by chance, given a significance threshold. A hypothesis test consists of the following steps:

1. Formulate a null hypothesis $\mathcal{H}_0$ and an alternative hypothesis $\mathcal{H}_1$.
2. Make statistical assumptions about the observed process, e.g. the shape of the probability density function.
3. Choose an appropriate test statistic T and derive its expected distribution under the null hypothesis.
4. Choose a significance level α, a threshold below which the null hypothesis is rejected.
5. Given the observed values, compute the observed test statistic T_{obs}.
6. The significance level α defines an acceptance interval I_{acc} for the test statistic. If the observed statistic falls into the interval $T_{obs} \in I_{acc}$ the null hypothesis is accepted, otherwise the alternative hypothesis is accepted.

Statistical tests can be split into two basic groups: parametric and non-parametric. Parametric tests make assumptions on the underlying distributions of the observed variables, e.g. Gaussian or exponential distributions. Non-parametric tests do not require knowing the type of distribution beforehand. As the statistical framework presented in this thesis is designed to function with any possible image characteristics, it is not possible to assume a specific distribution. Therefore, this overview focuses on non-parametric tests.

2.2.1 Non-parametric tests

A non-parametric statistical test is a special hypothesis test, where the data does not need to follow a particular distribution. We need a statistical test that compares two sets of random drawings and assess whether their underlying probability distributions are similar or not (see Chapter 3). Thus, this overview focuses on three commonly used tests with this property, namely the Mann-Whitney-Wilcoxon, the Kolmogorov-Smirnov, and the Chi-square tests. The comparison at the end of this section discusses the differences between the three tests.

Mann-Whitney-Wilcoxon test

The MWW test was first presented by Wilcoxon in 1945 [106] and two years later discussed by Mann and Whitney [57] on a more solid mathematical basis. The test assesses whether one of two random variables is stochastically larger than the other, i.e. whether their medians differ.

Let X_1 and X_2 be sets of drawings from unknown distributions, respectively. The MWW test to assess whether the two underlying random variables are identical is done in three steps:

1. The elements of the two sets X_1 and X_2 are concatenated. If X_1 an X_2 have cardinalities n_1 and n_2, respectively, the joint set has cardinality $n_1 + n_2$.

2. The elements in the joint set are sorted in increasing order. The smallest (first) element has rank 1, the largest (last) element has rank $n_1 + n_2$.

3. The ranksum is the sum of the ranks from all those elements that came from the set X_1. Wilcoxon denoted this statistic with T.

As an example, let us consider the two sets $X_1 = \{5.2, -2.2\}$ and $X_2 = \{9, 3.0, 5.9\}$, then the ranksum is computed as follows:

1. $X_1 \cup X_2$: 5.2, -2.2, 9, 3.0, 5.9

2. sort: $\overset{1}{-2.2}$, $\overset{2}{3.0}$, $\overset{3}{5.2}$, $\overset{4}{5.9}$, $\overset{5}{9}$ (positional indexes stacked on top of the values)

3. ranksum: $T = 1 + 3 = 4$

Under the null hypothesis, i.e., when both sets are drawn from the same distribution, the mean and variance of the statistic T are [106, 57]:

$$\mu_T = \frac{n_1(n_1 + n_2 + 1)}{2} \tag{2.1a}$$

$$\sigma_T^2 = \frac{n_1 n_2(n_1 + n_2 + 1)}{12} \tag{2.1b}$$

The mean and variance can be used to normalize the statistic, yielding the standard z value:

$$z = \frac{T - \mu_T}{\sigma_T} \tag{2.2}$$

In the above example we obtain $z = \frac{4-6}{\sqrt{3}} = -1.15$. The z value is positive (negative) if the median of the first distribution is larger (smaller) than the one from the second distribution. If the medians are equal, the z value is equal to zero. This can be seen in the graphs of the third column of Figure 2.3.

Kolmogorov-Smirnov test

The two-sample Kolmogorov-Smirnov test assesses whether two probability distributions differ or not [45, 89, 27]. It is sensitive to location and shape.

Given two drawings X_1 and X_2, the empirical cumulative distributions functions are $F_1(x)$ and $F_2(x)$, respectively. Then the test statistic is computed as:

$$D_{n1,n2} = \sup_x \left| F_1(x) - F_2(x) \right| \tag{2.3}$$

which is the maximum difference between the two cumulative distribution functions along the horizontal x-axis. The cardinalities of X_1 and X_2 are n_1, n_2, respectively. The statistic $D_{n1,n2}$ can be normalized using precomputed tables [89].

Chi-square test

The Pearson's Chi-square test assesses whether an observed random variable with distribution O follows an expected distribution E [77].

Let O_i and E_i be the relative frequency of bin i under the observed and expected probability function, respectively. Then the Chi-square test is:

$$X^2 = \sum_{i=1\ldots n} \frac{(O_i - E_i)^2}{E_i} \tag{2.4}$$

Under the null hypothesis, i.e., when the observations are indeed drawn under distribution E, X^2 follows a χ^2-distribution.

Comparison

Figure 2.3 qualitatively shows for different input distributions (columns 1 and 2) the behavior of the three presented tests (columns 3-5). The first distribution is always rectangular. The second distribution:

- has same shape, but different median (1st row)
- has equal median, but different shape (2nd row)
- has equal median, but different shape (3rd row)
- is identical to the first distribution (4th row)

If the test statistic is zero, the respective graph is marked with a dashed frame. One sees that the Mann-Whitney-Wilcoxon test is only unequal to zero in the first case where the medians are different. The Kolmogorov-Smirnov test measures the difference in shape for the example in row two. However, it barely

measures the difference in shape in the third row since the cumulative distribution functions are very similar. The test statistic is close to zero. The Chi-square test also measures the difference in the third row since it sums up the squared differences in every single bin.

In the context of semantic image enhancement we only decrease or increase certain characteristics in an image; we do not alter the shape of their distribution (see Chapter 4). In this context it is thus a disadvantage to use a test that is sensitive to shape changes, and we favor the MWW test over the other tests. However, it is possible that an application different from the ones presented in Chapters 4 and 5 can benefit from a sensitivity to shape changes. The statistical test has to be chosen to match the desired properties for the application in mind.

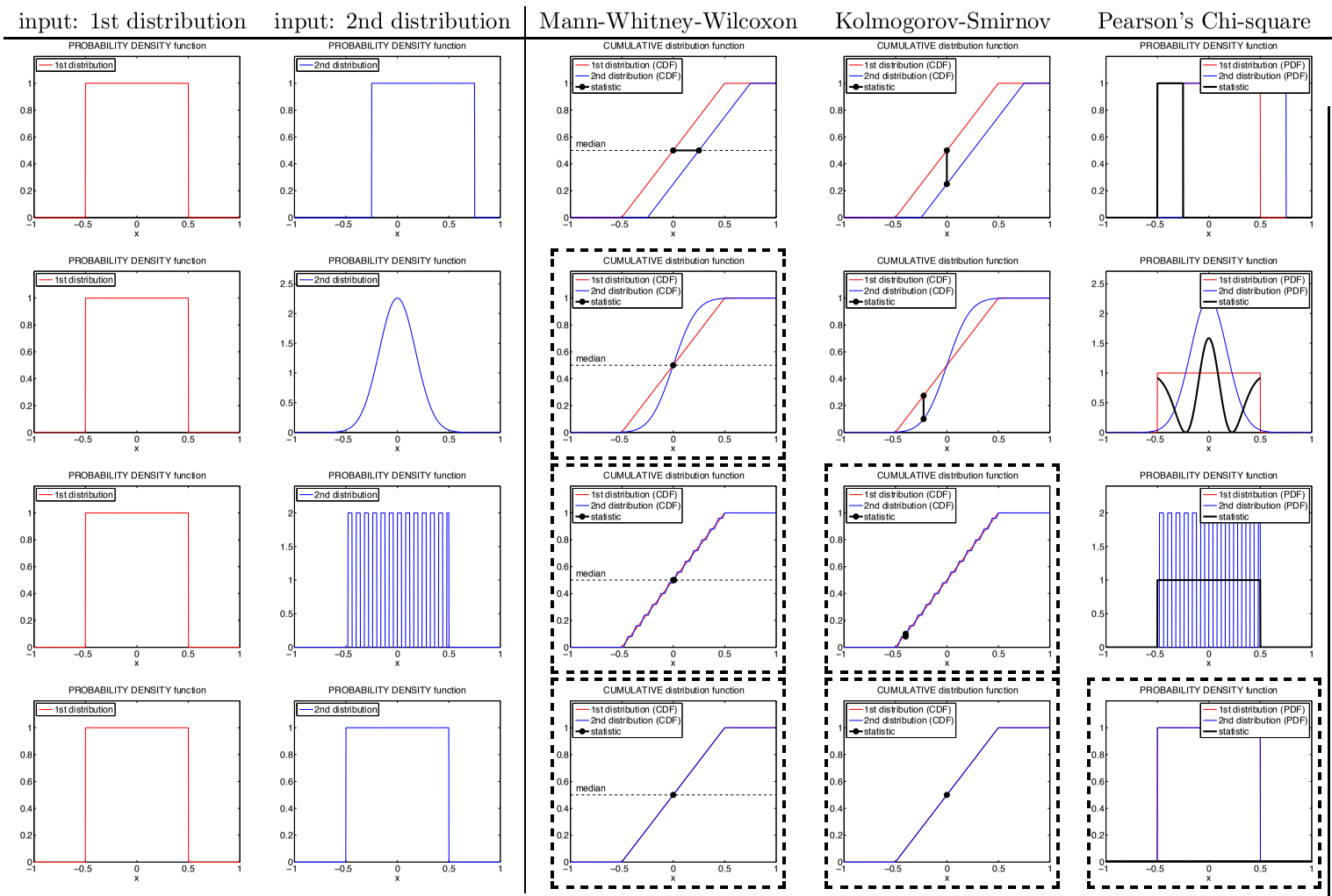


Figure 2.3: Comparison of different non-parametric tests. Every row shows two input distributions (first two columns) and their statistical analysis using different hypothesis tests (columns three to five). The dashed frame indicates that the test statistic is zero.

2.3 Data- and Image-mining

We are using a statistical test to find patterns of image characteristics for different keywords. This field is called data-mining and aims at finding patterns in large datasets using statistical or machine learning techniques. It is also known as knowledge discovery, which emphasizes the goal to generate previously unknown information. The field emerged with the availability of abundant data that can be collected and stored as computer storage became cheaper. Data-mining has applications in many diverse fields such as education [83], finance [22], fraud detection [75, 44], marketing [67] and so forth [107].

One key aspect of data-mining is the high quantity of data, which compensates for a possible lack of quality. This is known as the "wisdom of crowds" [90], which describes that a crowd of non-experts can be more knowledgeable than a few experts. This is of great importance for data from the world wide web where numerous non-experts contribute content. The data from each single person might be unreliable, but the plentifulness makes it possible to extract meaningful knowledge. A well known algorithm that uses this principle is Google's PageRank algorithm [12].

Image-mining refers to data-mining in the context of images. The required data can come from online image sharing communities, which are rich sources of images along with semantic context (see below). Application areas are, for example, image annotation [99], tag relevance estimation [93] (Figure 2.4), concept modeling [7, 50, 48] or automatic image interpretation [53, 51, 52]. For further reading there are two extensive survey articles on data-mining algorithms for classification tasks [108] and semantic image interpretation using associated social data [87].

Large databases of images can be acquired from the world wide web using Google Image Search[2] or a social photo sharing web page such as Flickr[3]. Alternatively, existing databases can be used such as MIR Flickr [35, 36] or The Flux [92]. Other potential sources for large-scale image collections are Facebook[4] or Google's Picasa[5]. However, the last two do not provide an open access for research purposes. The two main sources we used are Flickr and Google Image Search.

Images from Flickr can be accessed through the Flickr API[6] with e.g. a Python script or a program written in C. The API offers functions to query the Flickr database for images with specific criteria such as the presence of a keyword or the time interval in which the image was taken. One can then download the

[2] http://images.google.com/
[3] http://www.flickr.com/
[4] http://www.facebook.com/
[5] http://picasaweb.google.com/
[6] http://www.flickr.com/services/api/

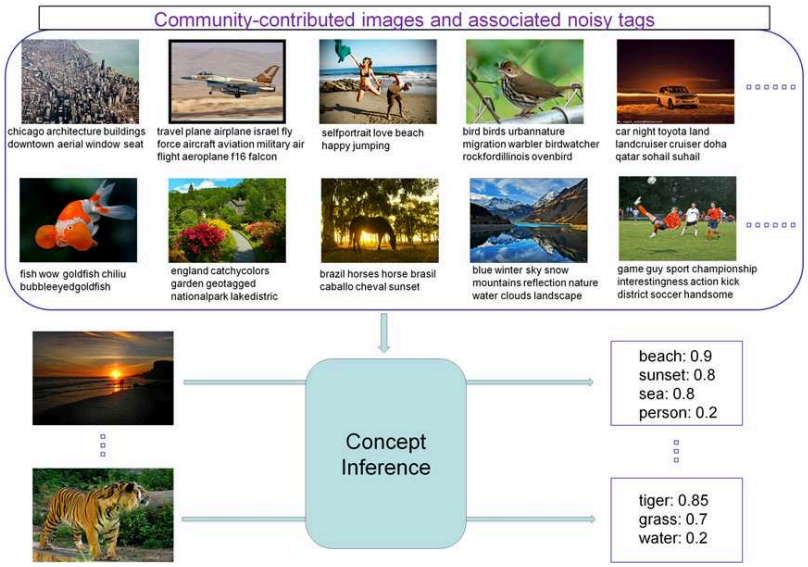


Figure 2.4: Tang et al. present a framework that infers semantic concepts from community-contributed images with noisy tags. The number behind the inferred concepts are an estimation of the tag's relevance (the Figure is reproduced from Tang et al. [93]). Methods that are robust against noisy input are important for image-mining on databases downloaded from potentially unreliable sources in the world wide web.

resulting images in different sizes along with their complete annotations, i.e. all tagged keywords. Two examples are reproduced in Figure 2.1.

Images associated with a specific keyword can be downloaded from Google Image Search with URL search query parameters [32]. It is possible to either download the small thumbnail images provided by Google in the search result overview or to follow the links to the original images on the internet. Downloading the thumbnails is significantly faster not only due to the small file size but also because the thumbnails are hosted by a Google server with high bandwidth. Additionally, it can happen that Google's index is out of date, and the image is not available any more on the original web page, but still listed in the search result.

There are two main differences between images from Flickr and Google that are important for this thesis. First, Google requires a keyword search query whereas on Flickr it is possible to download images and their annotations by sending a blank search query. It is thus possible to build a keyword vocabulary using Flickr as opposed to Google, where all keywords have to be known

beforehand. Second, images from Google Image Search can only be associated with a single keyword, the one from the query, whereas Flickr provides a list of all keywords for each image.

We use Flickr images for semantic image enhancement (Chapters 4 and 5) and Google images for color naming (Chapters 6 and 7).

2.4 Image enhancement

Image enhancement is a well studied topic in academia and industry. This section gives an overview of different approaches that are relevant for the context of this thesis (Chapters 4 and 5). This overview categorizes image enhancement into enhancement based on expert rules, enhancement derived from example images, enhancement based on classification, and artistic image enhancement.

2.4.1 Enhancement based on expert rules

This group of algorithms relies on a set of rules (defined by a human expert) that an enhanced image should satisfy. An input image is then modified so that it better respects these rules. These methods can work on a single image without any other input.

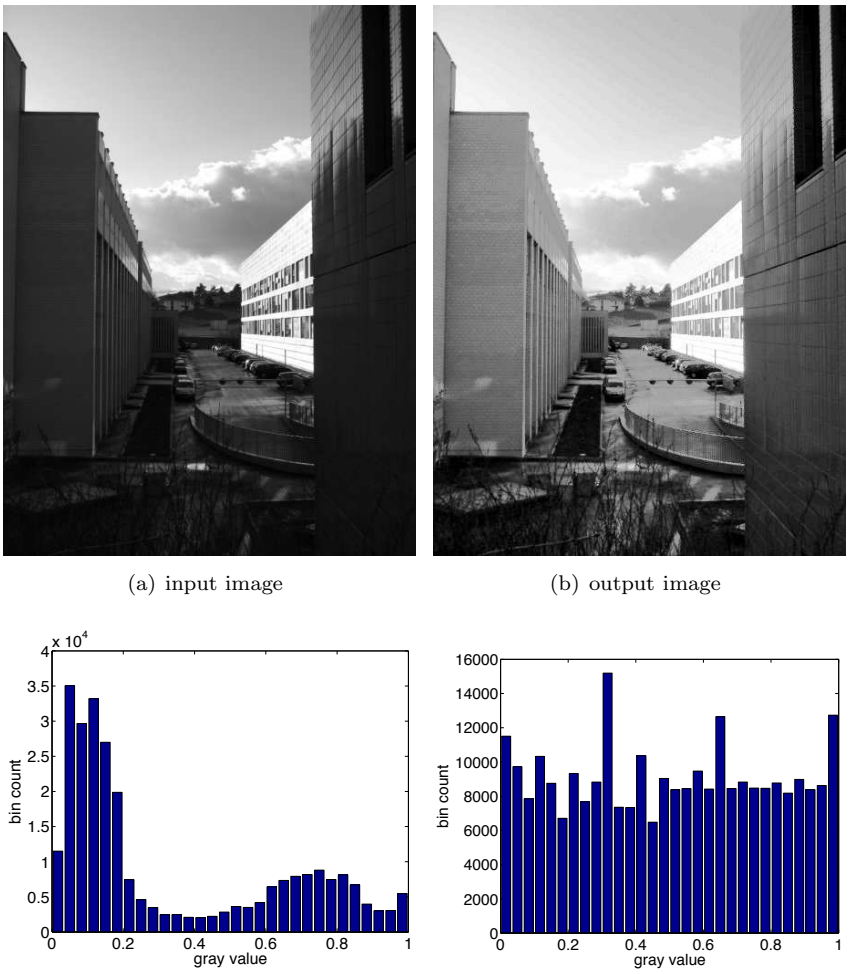
A simple example is the rule that in an image's histogram, the bin counts should be more or less equal. This so-called histogram equalization process improves an image's contrast and has been known for a few decades [37]. Figure 2.5 shows an example case with histograms. Another example is unsharp masking, where the input image is convolved with a high-pass filter and added back to the original image in order to make it look sharper [78].

More recent and sophisticated examples are methods that increase region saliency from Fredembach [28] or adjust color harmony in an image from Cohen-Or et al. [17] and Sauvaget and Boyer [86]. Wang et al. [103] and Murray et al. [62] present methods to adjust an image's color composition with predefined color themes, such as "nostalgic" or "spicy". However, their approaches are limited as the color themes are manually defined. On the contrary, our approach presented in Chapters 4 and 5 can interpret any semantic expression at the input and deduce an appropriate image processing from it.

2.4.2 Enhancement derived from example images

Example-based algorithms adjust the characteristics of an input image with those of one or more example images. Depending on the example images, different enhancements can be achieved.

Reinhard et al. [79] propose a system that transfers the colors from an example image to an input image. This is done by changing for each color



(a) input image (b) output image

(c) histogram of input image (d) histogram of output image

Figure 2.5: Example of histogram equalization, a well-known image enhancement method that is based on an expert rule: an image looks good if the bin counts of its histogram are almost equal [37].

channel separately the mean and variance of the input image to match those from the example image. Kang et al. [41] develop a method where a user creates personal example images in a previous step. The parameters from the example set are then used to personalize the enhancement of a new input image.

Wang et al. [104] present a framework to map colors and gradients. They use an example set of registered image pairs of scenes taken with a low-end and a high-end camera. The mappings from the low to the high-end images are

then applied to an input image. This can enhance images taken with a low-end camera as shown in Figure 2.6.

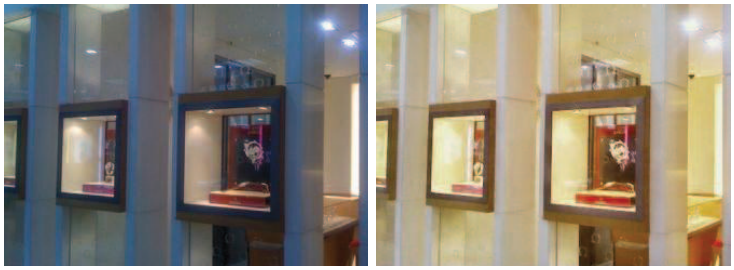


Figure 2.6: Example of image enhancement with rules that are derived from example images. The input image (left) is taken by an *iPhone 3G* and the output image (right) is processed to mimic the color and tone style of a *Canon EOS 5D Mark II*. This is achieved by learning a mapping from many image pairs that show scenes taken with both cameras, respectively. Images reproduced from Wang et al. [104].

Yet another example is pursued in a research project of Harvard's GVI lab and Adobe Systems Inc. [33]. They propose a method to query a large database for similar images based on SIFT [54] and GIST [70] descriptors. The input image is then processed in order to look more like the similar images from the database. The application areas the authors refer to are "restoring natural appearance to images taken with a camera that suffer from common artifacts; and enhancing the realism of computer-generated (CG) images."

2.4.3 Enhancement based on classification

Algorithms of this group depend on a manual or automatic classification of an image (or regions of it) into a fixed set of image categories. The image processing is then optimized for each class.

Such systems are omnipresent in the form of "modes" in e.g. digital printers and cameras. On printers, the user's classification of a document into "draft" or "presentation" invokes different algorithms to process it for printing. On cameras, scenes modes such as "portrait", "kids and pets" or "foliage" imply certain characteristics of a scene that the camera can account for by choosing parameters for the scene capture and the image processing.

Figure 2.7 shows two images of the same scene taken with a *Canon PowerShot S100* in "automatic" and in "foliage" mode, respectively. It is visible that the photo in "foliage" mode has more vivid colors due to the camera's internal processing for this image class.

(a) "automatic" mode (b) "foliage" mode

Figure 2.7: Illustration for class dependent image enhancement. Two images of the same scene taken with a *Canon PowerShot S100* with two different modes: "automatic" and "foliage". Photos of the latter class are rendered with more vivid colors.

The drawback of such scene-based processing is that it does not scale well due to practical limitations. Cameras and printers usually do not have more than 10 to 20 scene modes for two reasons. First, it is an unreasonable demand to the user to search through a list of possibly hundreds of modes and choose the right one. Second, each scene mode's algorithm has to be manually implemented, which is a laborious work.

Automatic systems have been proposed by Moser and Schroeder [60] and Ciocca et al. [16]. They use common classes, such as "sky", "skin", or "vegetation". Both approaches are adapted to image semantics. However, only seven and three semantic concepts are distinguished, respectively.

We present an approach to handle an arbitrary number of keywords, i.e. scenes, in a single framework. This frees the user from searching for the right scene mode from a limited list and allows the system to automatically realize a specific processing algorithm for each keyword.

2.4.4 Artistic image enhancement methods

A somewhat different group of image enhancement algorithms create artistic effects. An example of this is defocus magnification, where the goal is to additionally blur out-of-focus regions so that the object in focus is more accentuated [6, 114]. These algorithms first compute a defocus map [65] and then intentionally blur the image according to the estimated defocus level as shown in Figure 2.8.

Other enhancement methods of this category are, for example, painterly re-rendering or time-lapse fusion. In painterly re-rendering, an image is processed

(a) input image (b) depth map

(c) output image

Figure 2.8: Example of defocus magnification reproduced from Bae and Durand [6]. The depth map indicates the etimated depth of each pixel in the input image, which is used to increase the out-of-focus blur.

in order to appear as an artist's painting [112, 113]. Time-lapse fusion is a technique to simulate arbitrarily long exposure times by merging several images of the same scene into one single photo as shown in Figure 2.9 [23].

Figure 2.9: Example images for time-lapse fusion. Many images of the same scene are merged in order to realize arbitrarily long exposure times creating artistic photographic effects. Images reproduced from Estrada [23].

Artistic image enhancement is a large field and goes beyond the scope of

this thesis because we intend to retain an image's naturalness. However, we use the defocus magnification techniques and implement a semantic adaptation of out-of-focus blur as presented in Section 5.2

2.5 Psychophysical experiments

Psychophysics is the scientific study of the relation between stimulus and sensation [29]. One of the earliest works on this field are the studies of E. Weber on tactile senses and weight perception in the 1830s [105]. The term *psychophysics* was then coined by the German experimental psychologist G. Fechner [25]. Inspired by Weber's work, his goal was to build a theory that could link matter and mind, i.e. connecting the public real world and a person's private impression of it.

Within this large field we focus on visual stimuli of still images to evaluate the performance of the framework for semantic image enhancement (see Chapter 4). This is necessary in order to prove the algorithm's performance. A psychophysical experiment has to fulfill three main conditions:

1. Enough observers/stimuli of enough diversity in order to deliver statistically significant results.

2. A controlled environment in order to make the results reproducible and minimize external spurious effects.

3. Clear instructions that are the same for all observers.

Images can be judged independently on an absolute scale or relative to each other in a comparative experiment. Scoring an image on an absolute scale, i.e. from 1 (*worst*) to 10 (*best*), has inherent difficulties. For instance a 5 for one person might be a 7 for another person, or it might not be clear how bad (good) and image needs to be to be rated as 1 (10). Even though it is possible to train the observers beforehand and to normalize their responses to some extent, a comparative experiment avoids these drawbacks.

One of the first psychometric models for paired comparison tests has been proposed by Thurstone [95]. The goal is to derive the mean quality difference between two samples A and B from the number of times A (B) has been preferred over B (A). An alternative model is the Bradley-Terry-Luce model [11, 55]. However, both models produce very similar results [96].

Standards and recommendations to conduct psychophysical studies are given by different organisation such as the International Organization for Standardization (ISO) [4] or the International Telecommunication Union (ITU) [72, 13]. These standards define the laboratory environment, the requirements for observers, the analysis of the results, and so forth. For instance the number of

observers is *"at least 10 (and preferably 20)"* (ISO) or *"at least 15"* (ITU). The regulations aim at guaranteeing both repeatability and significance of the experiment. One has to bear in mind that these are lower bounds, i.e. it is possible that a specific psychophysical study needs more observers to be statistically significant.

The drawbacks of traditional psychophysical experiments in a controlled laboratory environment are the high costs of human labor and equipment, especially for larger studies. These high costs can be avoided with crowd-sourcing, an alternative approach for psychophysical experiments that became popular during the last years [81, 47, 14, 80, 43, 42].

Crowd-sourcing is a process where a task is divided into small units of work that are distributed to many people. The responses of all workers are then aggregated to gain a complete picture of the global task. There are many different online platforms for crowd-sourcing, such as Amazon Mechanical Turk (AMT)[7], microWorkers[8], clickworker[9], and shortTASK[10]. These services provide infrastructure to design the tasks in the form of an html web page, distribute the tasks to qualified workers, acquire the results and pay the workers.

The quality of psychophysical experiments accomplished using online crowd-sourcing is a widely discussed topic [81, 14, 80, 43, 73]. Ribeiro et al. [81] and Keimel et al. [43] explicitly compare results from traditional and crowd-sourced experiments for audio and video quality assessment, respectively. These and all other studies that were found for this state-of-the-art all attested that the quality of crowd-sourcing is comparable to experiments in a standardized laboratory environment.

The reported good quality of crowd-sourced experiments might be astonishing at first sight because the 2nd condition (controlled environment) is violated. However, crowd-sourcing makes it easy to increase the number of observers (1st condition) to compensate for this.

In addition to this, the evaluation of the images in a web-based experiment is a more realistic scenario because people look more and more at soft-copies of images on screens instead of printed hard-copies. Finally, a comparative study is more robust to varying viewing conditions, because both images are displayed next to each other simultaneously. A possibly wrongly calibrated monitor or bright surrounding light thus always affects both images.

An inherent regulating mechanism that enforces high quality is the fact that workers can be punished. A requester on AMT can reject a worker's result without any explanation and approval of a third party. In that case a worker is punished in two different ways. First, AMT does not pay the worker. Sec-

[7] https://www.mturk.com/
[8] http://microworkers.com/
[9] http://www.clickworker.com/
[10] http://shorttask.com/

ond, the worker's success rate (i.e. reputation) drops, which is often used by requesters as a criteria to grant access to their tasks. On the other side, the workers do not have an official lobby to demand their rights, but workers can use forums to exchange their experiences.

An interesting observation is that increased payment does increase the quantity but not the quality of the work [58]. Ribeiro et al. [81] also found that increased payment increases the quantity of work. They further conclude that the throughput can be additionally increased by awarding bonuses, clear instructions and a well designed user interface. In our experiment we decided to pay one US cent per comparison, which attracted enough observers to carry out the experiment.

2.6 Color naming

In color naming, the well known study of Berlin and Kay [9] proposes that a language has, depending on its stage, two to eleven basic color terms. The simplest language distinguishes only black and white. As a language evolves, new color terms are added in a strict chronological order: red, green, yellow, blue and so forth. Thus a language of a higher stage contains all color terms of the previous stages. Fully evolved languages all have at least the same eleven basic color terms. As this study suggests, color naming is a research subject in many fields such as linguistics, psychology, and ethnology.

Despite the importance of the different aspects of color naming, we focus on the acquisition of a numeric model for a given color expression (Chapters 6 and sec:thesaurus). This is usually very labour intensive since the responses of many observers have to be gathered in order to achieve statistical significance. Recent publications used web-based approaches to crowd-source the task to a large public [59, 63]. Moroney's color naming experiment [59] still continues online and the color names and their corresponding RGB values are accessible [66]. A screenshot of Moroney's web page is reproduced in Figure 2.10.

In a recent study, van de Weijer et al. [98, 30] use images from Google Image Search to learn a generative model for colors. The authors use a modified PLSA based model with a Dirichlet prior and enforced uni-dimensionality. The method performs well, but requires a retraining of the entire statistical model if a new color term is added. In our framework presented in Chapter 7 it is possible to add a new color term without affecting previous estimations.

2.7 Memory colors

Memory colors are colors that everybody knows by heart such as sky blue, vegetation green and skin tones. We introduce this topic here because we also

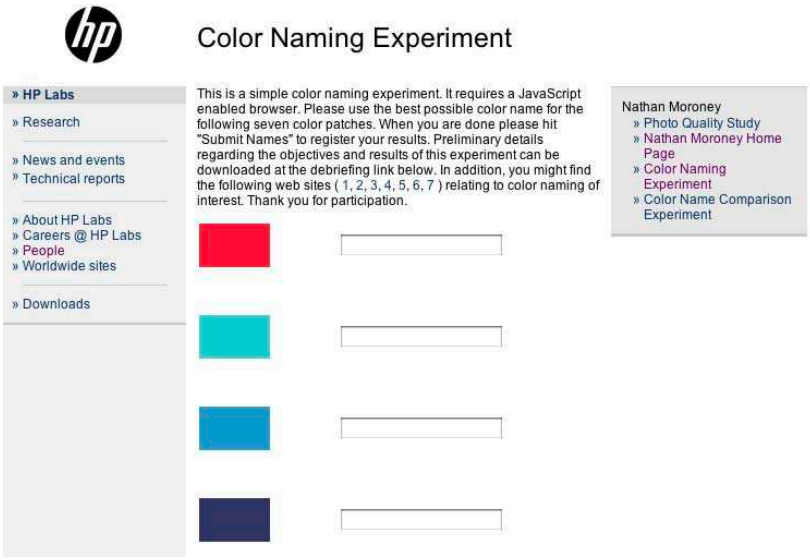


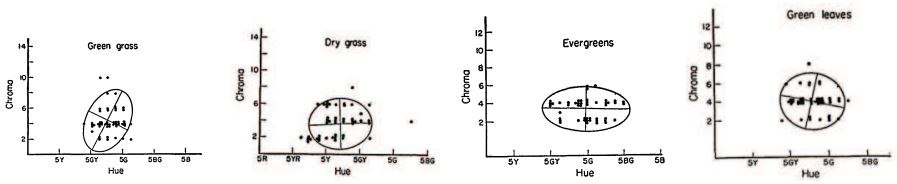
Figure 2.10: Screenshot of the online color naming experiment from Nathan Moroney [66]. The users are asked to type in the names of the displayed colors and submit their answer. The results from many users are aggregated to estimate a color name's color values.

apply the automatic color naming framework to memory colors in Section 6.2.1. At the beginning of research in this area, memory colors were mostly discussed from a psychologist's point of view [34, 5]. Adams discusses in his article [5] the appearance of grass, snow, coal, gold, and blood under different illumination conditions. However, the lack of adequate color spaces limited research and applications in this field.

The invention of the Munsell Color System [61, 68] allowed to describe memory colors in a perceptual color space. In 1960, Bartleson defined ten different memory colors in the Munsell hue and chroma plane using 50 observers [8]. The categories had subtle nuances such as "green grass", "dry grass", "evergreens", and "green leaves" as shown in Figure 2.11.

Memory colors are important to assess different qualitative aspects in image reproduction. Yendrikhovskij et al. showed that a deviation from the memory color prototype is perceived as unnatural [110]. Taplin et al. demonstrated that if a color shift is unavoidable, observers agree on a preferred hue angle of the shift [94].

The active tuning of memory colors in image reproduction systems is a common application in industry. Park et al. proposed a method to adjust skin colors for a more preferred image rendering [74]. You and Chien presented a



(a) green grass (b) dry grass (c) evergreens (d) green leaves

Figure 2.11: Variability ellipses for different memory colors in Munsell color space. Figures reproduced from Bartleson [8].

framework to enhance blue sky [111]. Other work focuses on segmenting memory colors in images [64, 40, 85]. The extracted maps can be used for further image processing.

The statistical framework (Chapter 3) also provides an automatic solution to determine memory colors as shown in Section 6.2.1.

2.8 Chapter summary

In this state-of-the-art summary we reviewed the literature for all relevant fields of this thesis. In the field of image descriptions (Sec. 2.1) we covered both the semantic and the numeric side. For the latter we focused on low-level descriptors due to the target applications of this thesis. In statistical data-mining (Sec. 2.2 and 2.3) we justified the use of the non-parametric Mann-Whitney-Wilcoxon test. Further, we reviewed image enhancement algorithms (Sec. 2.4) and pointed out how they can benefit from a semantic component. Psychophysical experiments (Sec. 2.5) are an important aspect of our work and we especially focused our review on crowd-sourced experiments that scale better than conventional experiments. Finally, we reviewed literature on color naming (Sec. 2.6) and the related field of memory colors (Sec. 2.7). Both fields usually demand psychophysical experiments that can be avoided with the techniques presented in this thesis.

Chapter 3

Linking Words with Characteristics

This chapter presents the statistical framework that links image keywords with image characteristics using data-mining techniques. This is the core pf the subsequent applications in Chapters 4 to 7. For data-mining to be effective, an abundance of data has to be available. In this thesis we focus on two data sources, namely Flickr and Google Image Search. Keywords in the context of Flickr images stem from the annotations of the photographer and the Flickr community, and in the context of Google Image Search they stem from the search query.

The statistical framework, the core of the learning method, is discussed in Section 3.1. A method to compare the impact of different keywords is explained in Section 3.2 and examples are given in Section 3.3. All examples in this chapter are based on the MIR Flickr database with 1 million annotated high-quality images [36].

3.1 Statistical framework

This section presents the statistical framework in four steps: the application of the statistical test to annotated image data (Sec. 3.1.1), an intuitive interpretation of the statistical test (Sec. 3.1.2) and the computational efficiency (Sec. 3.1.3).

We assume that all images of the MIR Flickr database [36] are encoded in sRGB color space. The images are in jpg file format [91] and the longer side of the images is 500 pixels long.

3.1.1 Measuring a keyword's impact on a characteristic

Our database consists of image/annotation pairs $(I_i, A_i) \in I_{db}$. An annotation is an ordered set of one or more keywords $A_i = \{w_1, w_2, \ldots\}$. Given a keyword w, the database can be split into two subsets $I_w = \{I_i | w \in A_i\}$ and $I_{\overline{w}} = \{I_i | w \notin A_i\}$ that contain all images annotated with keyword w and all remaining images, respectively. The keyword subset I_w is usually significantly smaller than $I_{\overline{w}}$. Clearly, $I_w \cap I_{\overline{w}} = \emptyset$ and $I_w \cup I_{\overline{w}} = I_{db}$.

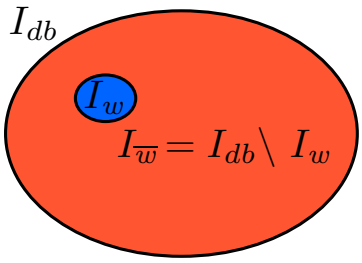


Figure 3.1: The complete database of image/annotation pairs $(I_i, A_i) \in I_{db}$ is split into two subsets: a smaller subset $I_w = \{I_i | w \in A_i\}$ containing all images that are annotated with keyword w and a larger subset $I_{\overline{w}} = \{I_i | w \notin A_i\} = I_{db} \setminus I_w$ containing all other images that are not annotated with w.

For an image I, a characteristic $C \in \mathbb{R}$ can be computed from it. This can be anything we want to characterize in an image. Examples are the percentage of pixels that have a certain gray-level or the output of Gabor filters. The set $\mathcal{C}_w^j = \{C_i^j | I_i \in I_w\}$ is defined as the collection of the characteristic j of all images annotated with keyword w. The set $\mathcal{C}_{\overline{w}}^j$ is analogously defined as the collection of the characteristic j from images in $I_{\overline{w}}$.

In order to assess how a keyword influences a characteristic j, the values in the sets $\mathcal{C}_w^j$ and $\mathcal{C}_{\overline{w}}^j$ have to be compared against each other. The task is to determine how the values of the two sets differ.

There are many ways to compare two distributions to each other such as the Kullback-Leibler divergence [46], the earth mover's distance [49] or statistical methods that compare the empirical distribution functions of two random variables. We prefer statistical methods over the others, because significance is a well known mathematical concept. Statistical significance measures whether an observation is a systematic pattern rather than just chance.

In the general case, the values do not follow a known distribution. Hence, we use methods from non-parametric statistics. A commonly used test is the Mann-Whitney-Wilcoxon (**MWW**) ranksum test (see Section 2.2.1 or [106, 57]), which assesses whether two observations have equally large values, i.e., by how much their medians differ. There are other non-parametric tests such as the

Kolmogorov-Smirnov test or the Chi-square test that additionally assess whether two distributions have different shapes. More details on non-parametric tests are given in Section 2.2 or Walpole [102].

For the application presented in this thesis the absolute value of a characteristic is important, not the shape of its distribution. Thus we use the **MWW**-test (see Section 2.2 for more details). The test statistic T, its expected mean μ_T and variance σ_T^2 are computed according to the algorithm described in Section 2.2.1 and lead to the significance value (repetition of Eq. 2.2 and 2.1):

$$z = \frac{T - \mu_T}{\sigma_T} \tag{3.1a}$$

$$\mu_T = \frac{n_w(n_w + n_{\overline{w}} + 1)}{2} \tag{3.1b}$$

$$\sigma_T^2 = \frac{n_w n_{\overline{w}}(n_w + n_{\overline{w}} + 1)}{12} \tag{3.1c}$$

where T is the rank sum of $\mathcal{C}_w^j$'s indexes in the sorted list of the joint set $\mathcal{C}_w^j \cup \mathcal{C}_{\overline{w}}^j$ and n_w, $n_{\overline{w}}$ are the cardinalities of $\mathcal{C}_w^j$ and $\mathcal{C}_{\overline{w}}^j$, respectively.

3.1.2 Interpretation of the z value

The z value is a useful measure to assess the relationship between keywords and low-level image features. The higher its magnitude, the more the corresponding characteristic is important for the keyword, and vice versa.

To give a better intuition for the z value, we consider an example where the tested image characteristic is a 16 bin gray-level histogram. For each of the equidistant bins, we calculate z_{night}^j from the two sets $\mathcal{C}_{night}^j$ and $\mathcal{C}_{\overline{night}}^j$, where $j = 1 \ldots 16$.

Figure 3.2 shows the distributions of all pairs, along with their corresponding z_{night}^j values. It is clearly visible that images annotated with *night* have more dark pixels ($z > 0$ for low gray-levels) and less bright pixels ($z < 0$ for high gray-levels). The z values smoothly vary between -130 and 124. The difference between $\mathcal{C}_{night}^j$ and $\mathcal{C}_{\overline{night}}^j$ is less significant for z values close to zero, which is the case for $j = 5$ (a medium gray-level). Overall though, an image's "*night*ness" is strongly related to its gray-level distribution.

Figure 3.3 shows the same plots but for the keyword *statue*. The two distributions are much more similar, the z values are closer to zero. This tells us that an image's gray-level distribution and its "*statue*ness" are not related.

We can thus introduce a simple ranking criterion for a given characteristic and keyword, which is the difference between the maximum and the minimum z-value as indicated in Figure 3.2. According to the examples depicted in Figures 3.2 and 3.3, we obtain $\Delta z_{night}^{\text{gray-level hist}} = 124.3 - (-130.0) = 254.3$ and $\Delta z_{statue}^{\text{gray-level hist}} = 6.5 - (-1.2) = 7.7$.

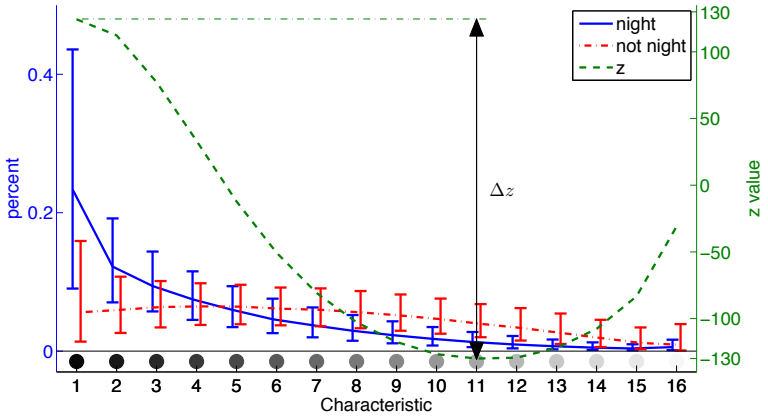


Figure 3.2: Left axis: The 16 characteristics of the sets $\mathcal{C}^j_{night}$ and $\mathcal{C}^j_{\overline{night}}$ measuring the percentage of image pixels falling into each bin. Each characteristic is represented with its median and its 25% and 75% quantiles. The markers at the bottom indicate the mean gray-level of each characteristic. For visualization purposes, the two curves have a small horizontal offset. Right axis: The corresponding z values indicate that images annotated with $night$ contain more dark ($z > 0$) and less bright ($z < 0$) pixels than the other images not annotated with $night$.

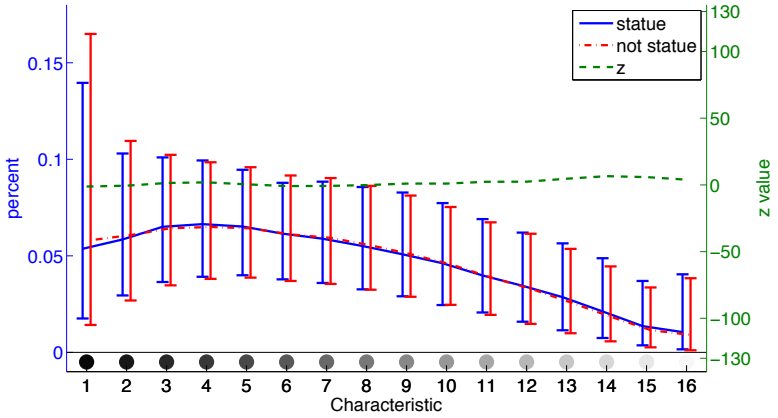


Figure 3.3: Same plot as in Figure 3.2 but for keyword $statue$. The distributions are more similar and the z values are closer to zero.

3.1.3 Computational efficiency

The one million example images and their characteristics are always the same. They are just differently split into the two subsets $\mathcal{C}_w^j$ and $\mathcal{C}_{\overline{w}}^j$ for every keyword w. This means that the characteristics have to be computed and sorted only once. Then, to compute the ranksum statistic for a keyword, we need only to sum the corresponding elements' ranks in the pre-sorted list. Computing this indexed sum takes 35.9ms for 16 z values (e.g. gray-level histogram) on a MacBook Pro (2.5GHz Core 2 Duo). The code is written in Matlab and the core functions are implemented as mex-files. The main bottleneck of our current implementation is the query for a given keyword, as we parse text files with regular expressions in Matlab. This takes 50s per keyword, but we are confident that a standard MySQL implementation will reduce this time significantly.

3.2 Comparing z values from Different Keywords and Characteristics

The z values can be computed for many keywords and characteristics. We use all keywords that occur in at least 500 images of the MIR Flickr database, 2858 in total. Additionally to the gray-levels, we compute other image characteristics: lightness, chroma, hue angle (all three in CIELAB space [39]), linear binary patterns [56], responses of high-pass and Gabor filters (image details), and frequency distributions in the Fourier domain. They are either summarized in a 16-bin histogram or in a 64-dimensional layout descriptor as shown in Figure 2.2(c). More details on the characteristics are in Appendix A.

As the z value depends on the number of associated images, a simple comparison of z values from different keywords is not possible. Section 3.2.1 explains how they can still be compared and Section 3.2.2 gives an overview comparison of 50 selected keywords.

3.2.1 Dependency on N_w

The z value depends on the number of images per keyword N_w as can be seen in Equations 3.1. This is an inherent property of any statistical test: more samples increase credibility and thus result in a higher significance value.

If the significance values from keywords with different numbers of samples have to be compared it is necessary to introduce a reference sample size N_w^*. All the variables from the statistical test can then be converted to this reference

sample size as follows (see Appendix D for details):

$$T^* = \frac{N_w^*}{N_w} \cdot T \tag{3.2a}$$

$$\mu_T^* = \frac{N_w^*}{N_w} \cdot \mu_T \tag{3.2b}$$

$$\sigma_T^{*2} = \frac{N_w^* N_{\overline{w}}^*}{N_w N_{\overline{w}}} \cdot \sigma_T^2 \tag{3.2c}$$

$$z^* = \sqrt{\frac{N_w^* N_{\overline{w}}^*}{N_w N_{\overline{w}}}} \cdot z \tag{3.2d}$$

The better comparability can be demonstrated with the keywords *bw*, *blackandwhite* and *blackwhite* that all represent the same semantic concept. The standard significance values are:

$$\Delta z_{\text{bw}}^{\text{chroma}} = 502.1 \quad \Delta z_{\text{blackandwhite}}^{\text{chroma}} = 379.0 \quad \Delta z_{\text{blackwhite}}^{\text{chroma}} = 230.1$$

The unequal values are a consequence of the different sample sizes

$$N_{bw} = 30294 \quad N_{blackandwhite} = 17092 \quad N_{blackwhite} = 6157$$

The compensated values are:

$$\Delta z_{\text{bw}}^{*\text{chroma}} = 63.5 \quad \Delta z_{\text{blackandwhite}}^{*\text{chroma}} = 64.3 \quad \Delta z_{\text{blackwhite}}^{*\text{chroma}} = 65.4$$

All three values are approximately equal, which is in accordance with the fact that they express the same semantic concept.

Figure 3.4 shows a scatter plot for all keywords w that occur at least 500 times in the database ($N_w \geq 500$) and 14 different descriptors; N_w is on the horizontal and Δz_w on the vertical axis. The three high peak values (marked with a large red cross) come from the previously discussed keywords *bw*, *blackandwhite* and *blackwhite*, respectively. The green root functions indicate equal $\Delta z_w{}^*$ values for a reference sample size of $N_w^* = 500$.

3.2.2 Comparison of 50 selected keywords and 14 characteristics

Figure 3.5 shows Δz_w^* values for different combinations of characteristics and 50 selected keywords w[1]. The scores are intuitively clear; *night* relates strongly to the gray-level histogram as the respective images tend to be very dark. *Blue* and *flower* have strong correspondence with hue and chroma characteristics. Spatial

[1] A longer table is reproduced in Appendix B. The full table for all 2858 keywords is provided on the research page: http://ivrg.epfl.ch/SemanticEnhancement.html.

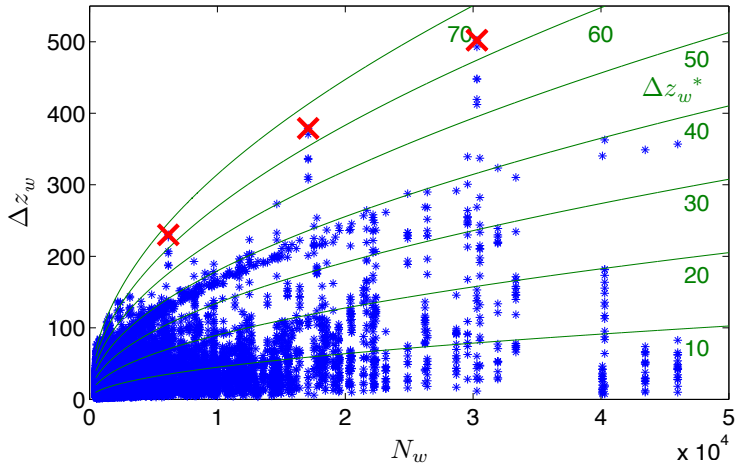


Figure 3.4: Number of images per keyword N_w versus significance Δz_w for different keywords w and descriptors. The three peak values (marked with a large red cross) are from keywords *bw*, *blackandwhite* and *blackwhite*, respectively. Even though they stand for the same semantic concept, their significance values are not the same due to the dependency on N_w. The green lines indicate constant ${\Delta z_w}^*$ values, which compensate for this dependency. Consequently, the ${\Delta z_w}^*$ values for the three keywords are approximately equal: $\Delta z_{bw}^{\text{chroma}*} = 63.5$, ${\Delta z_{blackandwhite}^{\text{chroma}}}^* = 64.3$ and ${\Delta z_{blackwhite}^{\text{chroma}}}^* = 65.4$.

layouts are significant for the keywords *sunrise* and *sunset* as they have a distinct spatial distribution of colors. The keywords *macro*, *flower* and *bokeh* strongly relate to high frequency content as these images often have a blurred background. However, there are also keywords that do not show strong correspondence with the tested characteristics, e.g. *happy* or *day*. Thus, our framework allows us to explicitly test if a given keyword has a predominant corresponding image characteristic or not.

This is important for image applications in general, as the absence of a significant characteristic implies that a given algorithm will not affect these images. For instance, in our image enhancement application of Chapters 4 and 5, the algorithm will not try to automatically improve images based on characteristics that are not relevant for a given keyword.

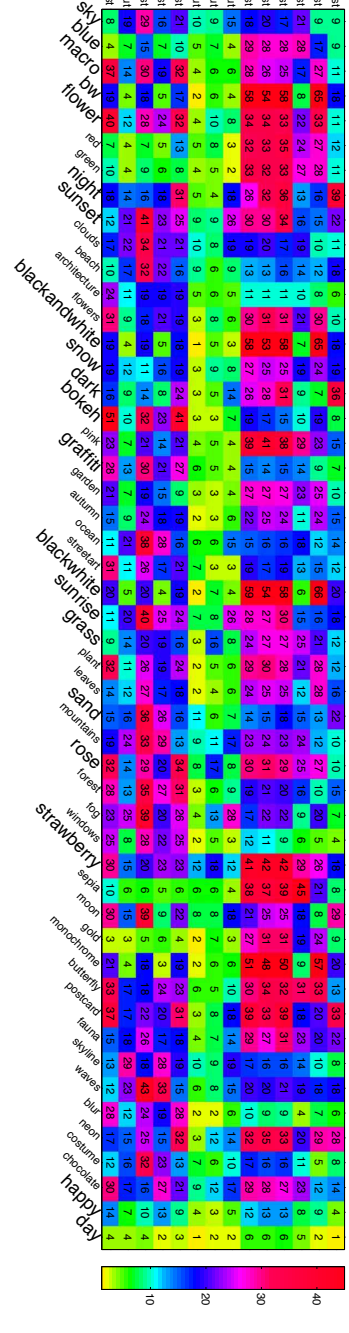


Figure 3.5: Δz^* values for 50 keywords and 14 characteristics. Note how the different keywords correspond to different characteristics. For instance, *bw* strongly equates with the chroma histogram (absence of high chromatic colors), *sunset* has a distinct spatial lightness layout (bright center and dark surrounding), and *graffiti* strongly relates to an image's high frequency content and linear binary patterns. *Day* and *happy* have very weak correspondence to any of the tested characteristics. Keywords that are referred to in the thesis have a larger font size. See the Appendixes A and B for more details on the characteristics and more significance values, respectively. The full table for all 2858 keywords is provided on the research page: http://ivrg.epfl.ch/SemanticEnhancement.html.

3.3 Examples

This section shows example significance distributions for different keywords and characteristics. Characteristics based on global histograms and spatial layouts are presented in Section 3.3.1 and 3.3.2, respectively. Details about the characteristics are listed in Appendix A.

3.3.1 Global histogram characteristics

Figure 3.6 shows z^* values for chroma, hue angle and linear binary pattern [56] histograms and keywords *red*, *green*, *blue*, and *flower*, respectively. The chroma characteristic of all four keywords are very similar, because all of them indicate the presence of saturated colors ($z^* < 0$ in low-chroma bins 1 to 4 and increasingly higher z^* values further on). However, the three color names have very discriminative hue angle histograms. *Flower* has a particular histogram of linear binary patterns (significantly more obtuse angles in bins 6 to 13 and less acute angles). The z_{blue}^{chroma} values decrease for very high chroma values due to frequent tagging of mid-saturated sky blue.

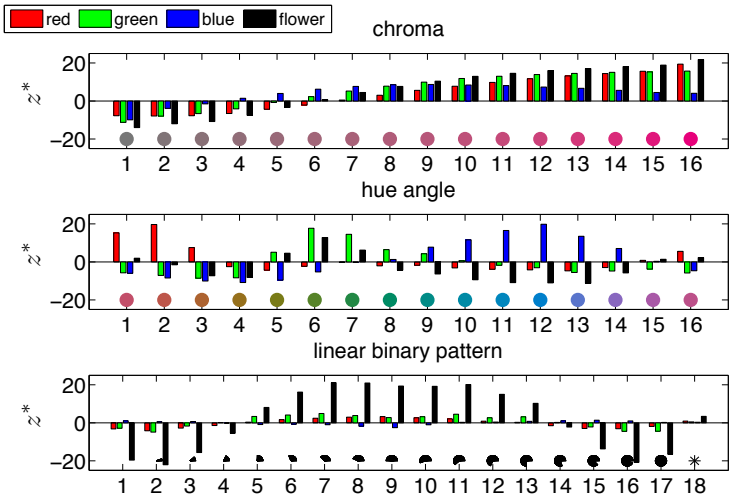
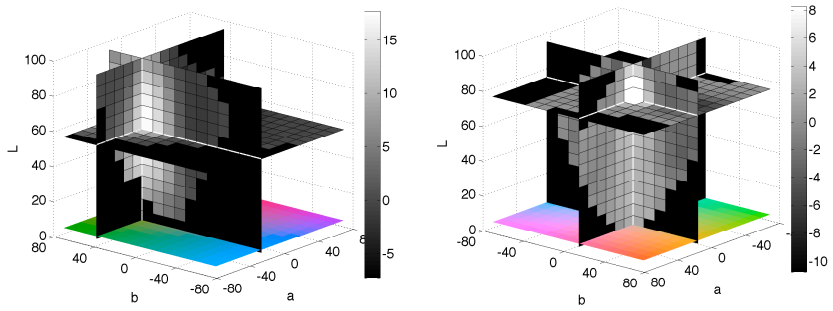


Figure 3.6: z^* values for chroma, hue angle and linear binary pattern histograms for keywords *red*, *green*, *blue*, and *flower*. Note that all 4 keywords have very similar chroma characteristics. The color names are best discriminated in their hue angle distributions. *Flower* has distinct linear binary patterns (angles increasing from acute to obtuse in bin 1-17 and miscellaneous patterns in bin 18).

An example of z^* values for a 3-dimensional CIELAB histogram is given in Figure 3.7 for *grass* and *skin*, respectively. The three orthogonal planes show cross sections of the distributions and intersect in the maximum. The z^* values are encoded with a gray-level heat map as indicated by the vertical bar. The color plane at the figure's bottom shows the histogram bin colors for the horizontal plane with constant L value that goes through the maximum z^* value. We can see that the maximum is at a green color in CIELAB space for *grass* and at a pale flesh tone for *skin*. The z^* values attenuate with increasing distance from the maximum.



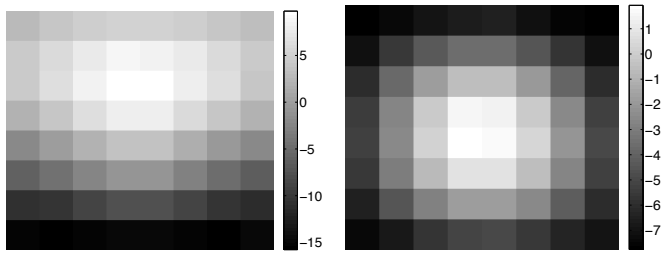
(a) *grass* (b) *skin*

Figure 3.7: The z^* value distributions in a 3-dimensional heat map for *grass* and *skin*, respectively. Each maximum is located at the crossing of the three orthogonal planes. The homogeneous dark areas at the plane borders are out-of-gamut values. At the bottom, we show the histogram bin colors for the constant L plane through the maximum value for a better orientation in CIELAB space.

3.3.2 Spatial layout characteristics

To compute the spatial layout, we superpose a regular grid with 8×8 rectangular grid cells. The values of the respective characteristic are averaged in each grid cell. This makes the characteristic independent of the image's size or aspect ratio. The approach is inspired by the MPEG-7 color layout descriptor [15].

Figure 3.8 shows significance values for the lightness layout characteristic of keywords *sunset* and *light* and an example image for each keyword, respectively. It is visible that *sunset* images are, with respect to other images, significantly brighter in the upper middle and significantly darker towards the borders, especially at the bottom. Images annotated with *light* are only slightly brighter than average in the center as the z^* values are positive, but close to zero. The surrounding is significantly darker.



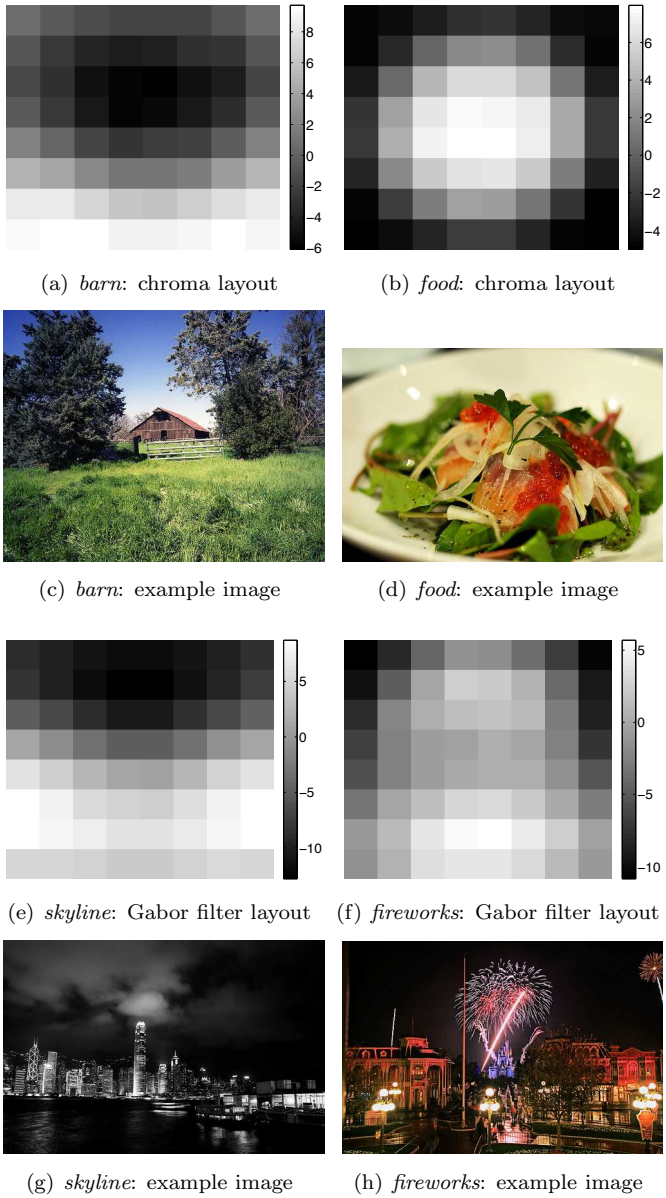
(a) *sunset*: lightness layout (b) *light*: lightness layout

(c) *sunset*: example image (d) *light*: example image

Figure 3.8: Top row: z^* values for spatial lightness layout in images annotated with *sunset* and *light*, respectively. *Sunset* images are significantly darker at the bottom and significantly brighter in the upper center. *Light* images are darker towards the borders and slightly brighter in the center. Bottom row: example image for both keywords. Photo attributions left: James Gentry and right: Håkan Dahlström.

Figure 3.9 is similar to Figure 3.8 but for chroma layout and Gabor filter layout characteristics. The images show that in general:

- *barn* images have a significantly desaturated center (the barn) and a significantly saturated foreground (grass or other nature scenes).
- *food* images have a center with significantly high saturation (the food) and a surrounding with significantly low saturation (a plate or a table).
- *skyline* images are significantly less structured in the top part (the sky) and significantly more structured in the bottom part (the city).
- *firework* images have two blobs of significantly high structure in the bottom and upper middle (the fireworks and illuminated objects at the bottom), whereas the two top corners contain significantly less structure.



(a) *barn*: chroma layout (b) *food*: chroma layout

(c) *barn*: example image (d) *food*: example image

(e) *skyline*: Gabor filter layout (f) *fireworks*: Gabor filter layout

(g) *skyline*: example image (h) *fireworks*: example image

Figure 3.9: Similar to Figure 3.8 but with different characteristics and keywords as indicated in the sub-captions. The significance distributions of the Gabor filter layout are computed with horizontal, vertical and diagonal Gabor filter of two different scales. The heatmap shows the 64 significance values for each spatial grid cell and thus represents the presence of structure (non-flat regions) without a preference for a specific direction or scale. Photo attributions in order: *refractionless (restricted)* (Flickr), Masaaki Komori, *wenzday01* (Flickr), and Joe Penniston.

3.4 Chapter summary

This chapter introduced in Section 3.1 a statistical framework that links arbitrary image keywords with arbitrary image characteristics. The framework is based on a large image database that is split into two subsets for a given keyword; images annotated with the keyword in a smaller subset, and all the other images in a larger subset. A statistical test is then used to determine whether a given characteristic is significantly smaller or larger in the first subset in comparison to the second subset.

The framework is computationally very efficient, because the characteristics can be computed and sorted offline and then used for all keywords. The framework easily scales to millions of images and thousands of keywords.

The significance of a statistical test depends on the number of samples. It is thus not possible to directly compare the significance values of two keywords with a different number of images. We showed in Section 3.2 how the significance values can be normalized to a reference sample size in order to compare different keywords.

We illustrated in Section 3.3 examples of significance distributions for histograms of colors and linear binary patterns, as well as for spatial layouts of lightness, chroma and Gabor filter responses, respectively. The significance values show expected patterns for the different semantic concepts.

Chapter 4

Semantic Tone-Mapping

For the first re-rendering application, a gray-level tone-mapping curve is computed that accounts for the image's semantic context. It is a global operation that maps an input pixel's gray-level to a new gray-level in the output image and thus alters the image's gray-level distribution.

In this chapter we first introduce the basic principle of semantic image re-rendering in Section 4.1. Then we discuss in Section 4.2 how and by how much an image's gray-levels have to be changed for a given input image and an associated semantic context. Section 4.3 proposes a method to build a tone-mapping function that realizes this required change. Finally, we present two psychophysical experiments that demonstrate the framework's superiority over state-of-the-art methods in Section 4.4.

4.1 Basic principle of semantic re-rendering

To semantically re-render an image for a specific semantic concept, its characteristic needs to be changed according to two components: semantic context and image content. Figure 4.1 illustrates this with the example of an image and the semantic context *gold*. The image component in the re-rendering workflow is based on the input image's pixels and represents its characteristics (e.g. a color histogram). The semantic component is based on an associated keyword and represents the significance of certain characteristics for the given semantic context. This is realized with the significance values from the statistical framework presented in Chapter 3.

The proposed re-rendering framework is flexible and can be realized for any characteristic by implementing an adequate semantic processing unit that fuses the image and semantic components (see Fig. 4.1). In this thesis we demonstrate semantic re-rendering with three different characteristics:

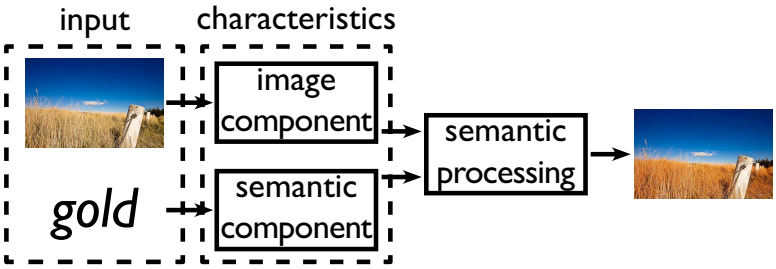


Figure 4.1: Illustration of the semantic re-rendering workflow. The framework takes as inputs an image and an associated keyword. In the first step the input image's characteristic and the keyword's significance distribution (see Chapter 3) are computed. The image and semantic cues are fused in the semantic processing step. The output image is enhanced with respect to the semantic concept. Input image from Meredith Farmer.

1. Tone mapping with gray-level histograms (this chapter).
2. Color enhancement with RGB histograms (Section 5.1, Chapter 5).
3. Change the depth-of-field with Fourier domain histograms (Section 5.2, Chapter 5).

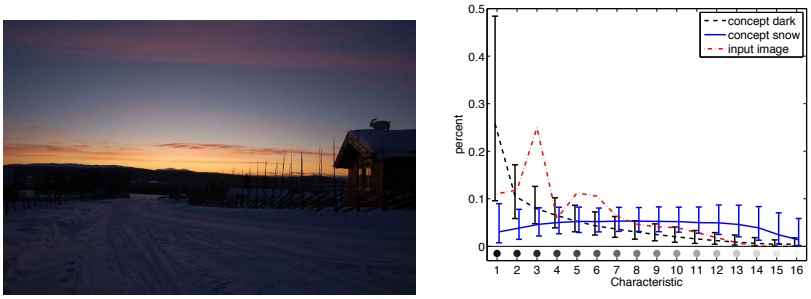
4.2 Assessing a Characteristic's Required Change

To re-render an image for a specific semantic concept, its characteristic needs to be changed according to two components: semantic context and image content. Hence we define two conditions that need to be fulfilled in order to alter the gray-level distribution:

1. The characteristic is significant for the semantic concept (i.e., high Δz as shown in Fig. 3.2)
2. The characteristic in the present image is too low or too high for the given concept.

Consequently, an image will not be altered if the characteristic is not influenced by the keyword or if the image is already a good example for it. We explain the implementation with the example of the image in Figure 4.2(a).

The first component is the significance distribution of the semantic concept and is assessed via the z value from Equation 2.2. If the z value is positive (negative), the value of the corresponding characteristic has to be increased (decreased) in order to emphasize the semantic concept. We assume a linear



(a) input image

(b) characteristics for *snow*, *dark* and the input image

Figure 4.2: Left: Example image that is used to explain the semantic tone-mapping framework. Right: Image's gray-level characteristic and the gray-level distributions of all images annotated with *snow* and *dark*, respectively. The curve shows the median and 25%/75% quantiles. Photo attribution: Marius Pedersen

relationship between the z values and the strength of the image processing; meaning that if the z value's absolute value is k times higher, the processing is k times stronger.

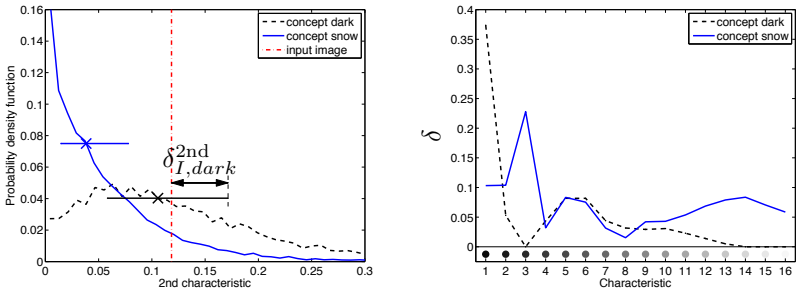
The second component is image dependent. We assess how well the given image already fulfills the desired characteristics for its semantic concept. We compare the image's characteristics to the characteristics of all images with the same keyword. Therefore, we compute the difference to a quantile:

$$\delta_{I,w}^{j} = \begin{cases} \max\left[0, Q_{1-p}\left(\mathcal{C}_{w}^{j}\right) - C_{I}^{j}\right] & \text{if } z_{w}^{j} \geq 0 \\ \max\left[0, C_{I}^{j} - Q_{p}\left(\mathcal{C}_{w}^{j}\right)\right] & \text{if } z_{w}^{j} < 0 \end{cases} \tag{4.1}$$

where $\delta_{I,w}^{j}$ signifies the difference measure for input image I with keyword w under characteristic j, C_{I}^{j} is image I's characteristic j, $Q_{p}(\cdot)$ measures a set's p-quantile and $\mathcal{C}_{w}^{j}$ are all characteristics j of images annotated with w.

If we use the 50% quantile $Q_{0.5}$ to compute the difference in Equation 4.1, the second condition is already fulfilled ($\delta = 0$) if the input image's characteristic is average for its semantic concept. If, however, we want to emphasize the significant characteristics more, a lower quantile has to be chosen. We found that a 25%-quantile is a good tradeoff between a desired enhancement and an extreme overshooting, which would happen for quantiles in the order of 5%.

The computation of the δ values is illustrated in Figure 4.3(a). The plot shows the probability density distributions of the second gray-level bin (almost black) for all images annotated with *snow* and *dark*, respectively. The horizontal error bars indicate the 25% and 75% quantiles and are the same as the vertical



(a) details for 2nd bin of Fig. 4.2(b)

(b) δ values for *dark*, *snow*

Figure 4.3: Computation of the δ value from Equation 4.1. Left: Probability density functions in the 2^{nd} bin of the characteristic with 25% and 75% quantiles (see Figure 4.2(b)). The $\delta^{2nd}_{I,dark}$ value is indicated with the arrow and measures how much the 2^{nd} bin of the gray-level characteristic has to be altered in order to increase the concept *dark* in the input image I (Fig. 4.2(a)). Right: δ values for all 16 bins of the semantic concepts *dark* and *snow*, respectively.

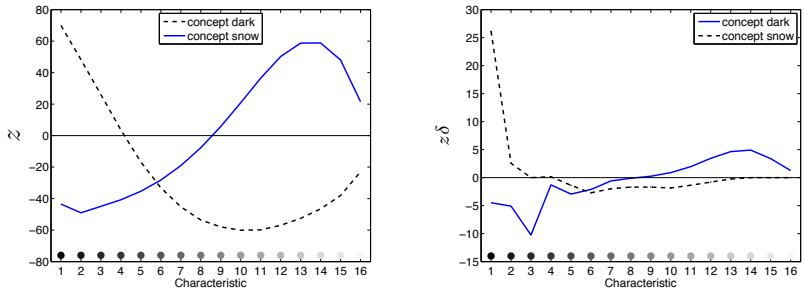
error bars in Figure 4.2(b). The 2nd bin of the input image's characteristic is indicated with a vertical red line. To increase the semantic concept *dark* in the input image I, the value of its characteristic's 2nd bin has to be increased as indicated by the δ value. The δ values for the complete gray-level characteristic is shown in Figure 4.3(b) for semantic concepts *dark* and *snow*, respectively.

Similarly to the dependency on the z values, we implement a linear relationship between the δ values and the strength of the enhancement. Thus, the image processing has to be proportional to the product of z and δ values. Figure 4.4(a) shows the z values for the semantic concepts *dark* and *snow* and Figure 4.4(b) shows the products of the z and δ values. If the product is positive (negative) for a specific bin, i.e. a gray-level, the image needs more (less) pixel values in that bin.

4.3 Building a Tone-Mapping Function

We use the z value from Equation 2.2 and the δ value from Equation 4.1 to determine a tone-mapping of an image's gray-levels. We compute a pixel's gray level as the average of its R, G and B values. The tone-mapping then changes the pixel's gray-level by multiplying the R, G and B values with the same factor. According to our previous assumptions, the change a processing introduces to an image has to be proportional to the product $z\delta$.

In the case of a tone-mapping function, the strength is given by its slope.



(a) z values for *dark*, *snow*

(b) $z\delta$ values for *dark*, *snow*

Figure 4.4: Left: z values for semantic concepts *dark* and *snow*. Right: Product of z and δ values. The product $z\delta$ indicates for which gray-levels the image's characteristic has to be increased ($z\delta > 0$) or decreased ($z\delta < 0$).

If at gray-level g, the slope is $m(g)$, the pixels in the interval around g are redistributed to a gray-level interval of $m(g)$ times the size. This holds for $m > 1$ (decreasing density) as for $m < 1$ (increasing density). A slope equal to one is the identity transform. As the $z\delta$ value indicates how strongly a characteristic has to be altered, the slope is:

$$m = \begin{cases} 1/\left(1 + Sz\delta\right) & \text{if } z\delta \geq 0 \\ 1 + S|z\delta| & \text{if } z\delta < 0 \end{cases} \tag{4.2}$$

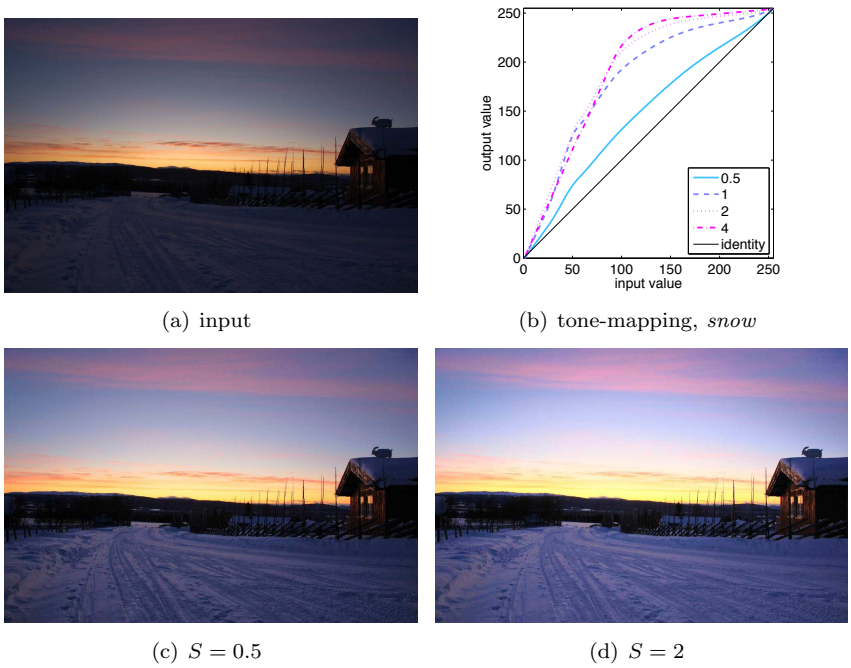
where S is a proportionality constant that controls the overall strength of the tone-mapping. According to the equation, the slope is $m = 1$ if the z or δ values are zero.

Extreme slope values are not desirable. A very steep mapping increases quantization artifacts and noise in homogeneous areas, and a very flat mapping reduces local contrast. Thus, the slope is cropped to a range $[1/m_{max} \;\; m_{max}]$. This is an inherent problem for any tone-mapping applications [76] and not specific to this approach. We used $m_{max} = 5$, which is a good compromise between limiting extreme tone-mappings and allowing visible changes.

The slope values from Equation 4.2 are linearly interpolated for 256 values in the interval [0 255] by using the representative mean gray-level of each characteristic. Because these values specify the slope, they are the derivative of the tone-mapping function. An integration thus yields the desired function.

Due to the continuity of the slope values, the mapping function is continuous and differentiable. This guarantees a certain smoothness constraint that is beneficial for non-invasive processing. In a final step, we scale the mapping function to the interval [0 255] in order to maintain the image's black and white points.

The graph in Figure 4.5 shows tone-mapping functions for different proportionality constants S for keyword *snow*. The smaller the S is, the closer the mapping function is to the identity transform, which is depicted by the thin black line. Higher S values lead to a more extreme mapping. The two images at the bottom show the output for $S = 0.5$ and $S = 2$.



(a) input

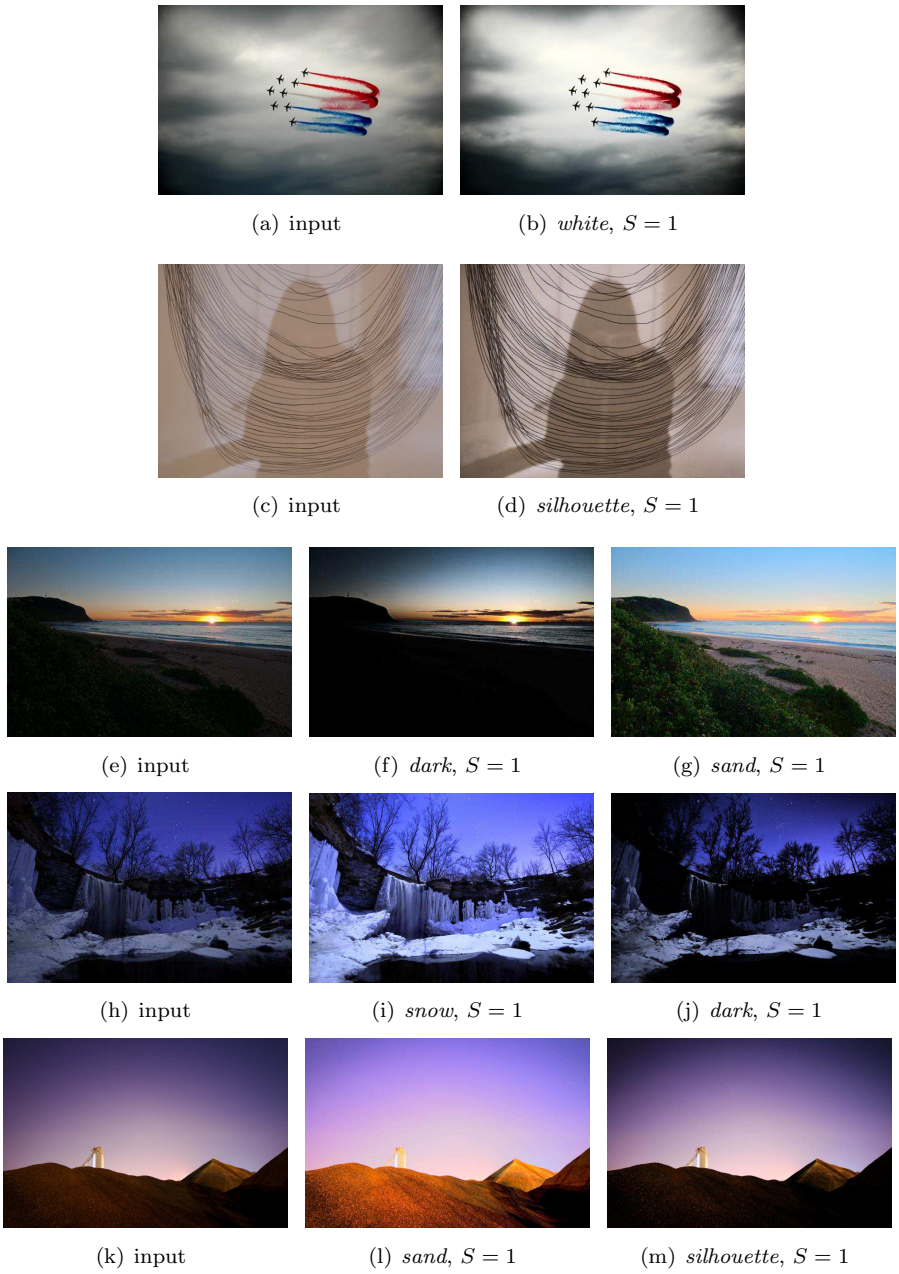
(b) tone-mapping, *snow*

(c) $S = 0.5$

(d) $S = 2$

Figure 4.5: Top: Input image and tone-mapping function to increase the semantic concept *snow* derived from the $z\delta$ values in Figure 4.4(b) ($S \in \{0.5\ 1\ 2\ 4\}$). Bottom: Two output images for $S = 0.5$ and $S = 2$, respectively. Input image from Marius Pedersen.

Figure 4.6 shows more example images that were semantically tone-mapped for different keywords. A key novelty of our semantic image re-rendering is the ability to adapt an image to an arbitrary semantic concept. This allows us to compute two different output images from the very same image by just altering the associated keyword as illustrated in three bottom rows in Figure 4.6. Each of the the output images is better for its semantic concept than the input image. Psychophysical experiments that show this and other properties are presented in the following section. The effect of the δ value can be observed in Figures 4.6(e) to 4.6(g). As the image is already low-key, the processing is less strong to re-render for the semantic concept *dark* than for *sand*.



(a) input (b) *white*, $S = 1$

(c) input (d) *silhouette*, $S = 1$

(e) input (f) *dark*, $S = 1$ (g) *sand*, $S = 1$

(h) input (i) *snow*, $S = 1$ (j) *dark*, $S = 1$

(k) input (l) *sand*, $S = 1$ (m) *silhouette*, $S = 1$

Figure 4.6: Example images of semantic tone-mapping. Images in the bottom three rows are tone-mapped for two different semantic concepts. The adaption of an image to an arbitrary semantic concept is a key novelty of our framework. More examples are in Appendix C or http://ivrg.epfl.ch/SemanticEnhancement.html. Photo attributions from top to bottom: Martin Filliau, Patricia Glave, John Campbell, Matthew Kuhns, and Dave Rizzolo.

4.4 Psychophysical Experiments

We evaluate the semantic gray-level enhancement with two psychophysical experiments. The first experiment shows that the semantically enhanced version is better than the original image. In the second experiment, we demonstrate that our algorithm outperforms other gray-level enhancement algorithms.

4.4.1 Proposed method versus original image

For the first experiment, we choose eight keywords with relatively high z values because low z values intentionally generate tone-mapping curves close to identity (see Eq. 4.2). The keywords w and their corresponding Δz^*_w values (in brackets) are *white* (14), *dark* (36), *sand* (22), *snow* (19), *contrast* (18), *silhouette* (27), *portrait* (11), and *light* (20), respectively.

For each keyword we selected 30 images from Flickr that have been annotated with the respective keyword, and we semantically re-rendered them with four different parameters $S \in \{0.5, 1, 2, 4\}$. Thus, we tested 960 images in total.

We set up a large-scale experiment using Amazon Mechanical Turk, where we showed the original and the enhanced image next to each other together with the corresponding keyword in the title as shown in Figure 4.7. We asked 30 observers to select the image that best matches the keyword and paid them 1 cent per comparison.

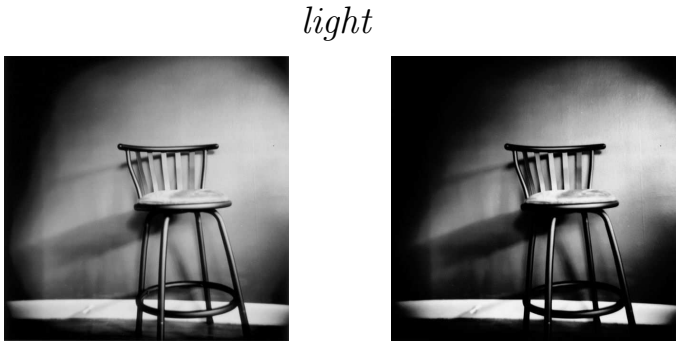
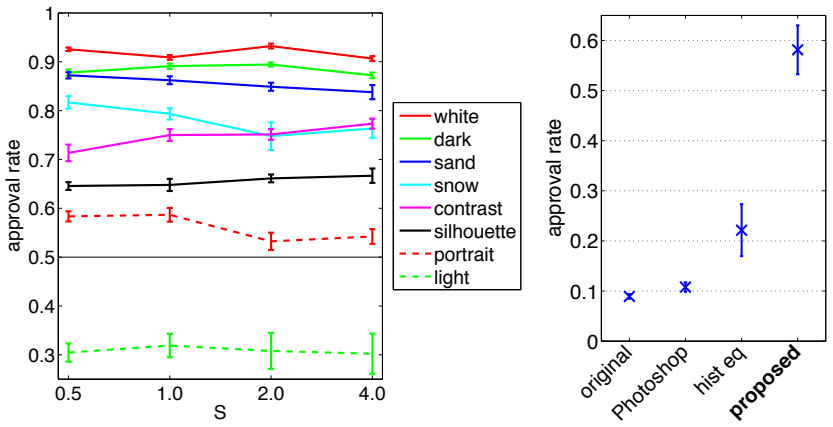


Figure 4.7: Setup of the first psychophysical experiment. The observers saw the original and the enhanced image next to each other and the keyword at the top. The position of the original and enhanced images was switched at random. They were asked to select the image that best matches the keyword. The enhanced image is on the right in this example. Photo attribution: Wesley Furgiuele.

Figure 4.8(a) shows the results of the psychophysical experiment. The S parameter is plotted on the horizontal axis and the approval rate for the en-

hanced image on the vertical axis. The approval values for all parameters S and all keywords, except one, are above 50%. Overall, the enhanced images are preferred and images in the *white* category have the highest rate (93%). This is not surprising as it is directly related to the gray-level characteristics.



(a) 1st experiment (b) 2nd experiment

Figure 4.8: Results from two psychophysical experiments. Left: Approval rates from 30 observers for different S values. Images of all except one semantic concepts are enhanced with success rates of up to 93%. The approval rate for *light* jumped to 62% in another experiment where we invited only artists. Right: Approval rate from 40 observers comparing the proposed method against other contrast enhancement methods (histogram equalization and Photoshop's auto contrast function). The proposed method scores more than 2.5 times better than the 2nd best. The error bars in both figures show the variances across different images.

The approval rate for images with *light* is surprisingly low and the variances are relatively large, which is due to the fact that there are two interpretations for this semantic term (see also Fig. 4.7):

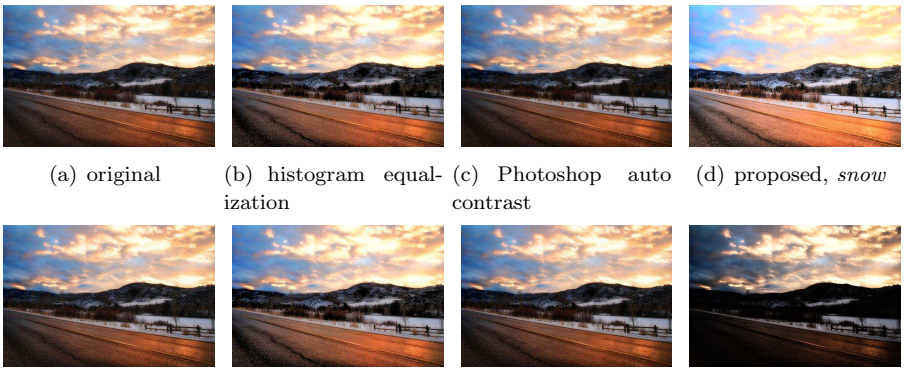
1. The image is bright in general.

2. The image shows a light source that is visually important due to the dark surrounding.

We reason that photographers and artists have rather the second point of view and carried out another experiment. We invited 20 photographers to judge the 30 images with keyword *light* ($S = 1$) and the resulting approval rate significantly jumped to 62% in favor of our algorithm.

4.4.2 Proposed method versus other state-of-the-art methods

The second psychophysical experiment compares our semantic image enhancement against other image only based contrast enhancements. We used four versions of each image, which were the original and three enhanced versions from Photoshop's auto contrast, Matlab's histogram equalization and our semantic framework for $S = 1$, respectively. In order to show the benefit of our semantically adaptive image enhancement we selected all images of the previous experiment that were annotated with at least two of the eight keywords. Figure 4.9 illustrated the experiment. The observers saw four images at one time, placed in a row, entitled with a keyword and had to decide which images best matched the keyword. As the other algorithms are not able to adapt an image to a semantic concept, our proposed version is the only one that changes. In total there are 29 such cases that were tested in the experiment.



(a) original (b) histogram equalization (c) Photoshop auto contrast (d) proposed, *snow*

(e) original (f) histogram equalization (g) Photoshop auto contrast (h) proposed, *dark*

Figure 4.9: Top: Four versions of the same image where our proposed algorithm re-rendered the image for *snow*. Bottom: the same, but for *dark*. The observers in the second psychophysical experiment saw one such row of four images along with the semantic concept and had to decide which image best matched the keyword. The arrangement of the images was again randomized from trial to trial. Photo attribution: Jim Nix.

Figure 4.8(b) visualizes the results from 40 observers: 58.1% voted for our version, 22.2% for the histogram equalization, 10.8% for Photoshop's auto contrast and 8.9% for the photographer's original. The variances across the different images is indicated with vertical error bars. We see that our semantic enhancement has significantly higher approval rates and scores on average more than 2.5 times better than the 2nd best method. This is because our semantic enhancement is the only method able to adapt to an image's semantic context.

4.5 Chapter summary

In this chapter we introduced the semantic re-rendering framework at the example of tone-mapping. The framework takes as inputs an image with an associated keyword and then determines separately an image and a semantic component. The image component stems from the image's characteristics and the semantic component is derived from the z values for the given keyword. Both components are fused into a semantic processing, which re-renderes the input image for the given concept.

Chapter 5

Additional Semantic Re-rendering Algorithms

The semantic gray-level tone-mapping from the previous chapter can easily be extended to semantic color enhancement, color transfer and out-of-focus adaptation as shown in this chapter in Sections 5.1, 5.1.1 and 5.2, respectively.

5.1 Semantic color enhancement

On the same lines as the gray-level tone-mapping, we can implement a semantic color transfer. As before, this requires two components that adapt to the image keyword and to the image pixels, respectively. The goal here is to emphasize the colors in an image that are related to a given semantic concept.

However, it is important not to apply a global color shift to the entire image as this would look unnatural in certain image regions, such as a human face or a blue sky. Therefore, the image dependent component (δ in the gray-level case) has to be spatially varying in the color case.

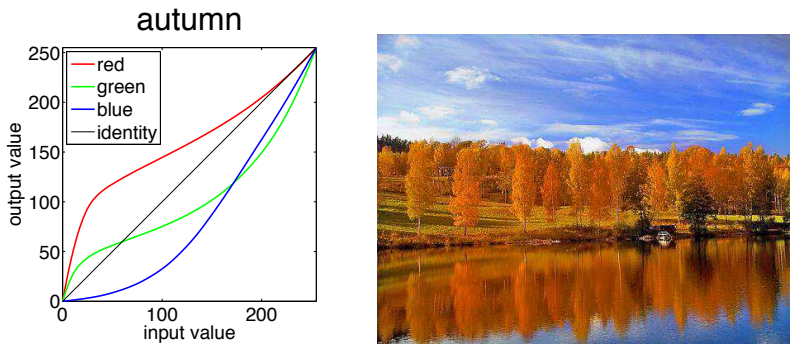
This requirement is accounted for with a spatial weight map ω that encodes how much each pixel belongs to the semantic concept as shown in Figure 5.1(b). The map is simply the z value for each pixel color $\mathrm{col}(p)$ at position p in the image under the given semantic concept w. To assure smooth transitions, the map is blurred with a Gaussian blurring kernel with a sigma σ of 1% of the image diagonal. Further, the 5% and 95% quantiles ($Q_{0.95}$ and $Q_{0.05}$) are linearly mapped to 0 and 1, respectively, to remove potential outliers.

$$\tilde{\omega}(p) = g_\sigma * z_w\big(\mathrm{col}(p)\big), \qquad \forall p \in \text{image plane} \tag{5.1}$$

$$\omega(p) = \min\left(1, \frac{\max\big(0, \tilde{\omega}(p) - Q_{0.05}\big)}{Q_{0.95} - Q_{0.05}}\right) \tag{5.2}$$

(a) input

(b) weight map ω, *autumn*



(c) tone-mapping curves

(d) output, *autumn*

Figure 5.1: Top: input image and associated weight map for the semantic concept of *autumn*. The bright regions in the map indicate regions that belong to the concept in the input image. Bottom row: Tone-mapping curves for the three color channels and the final output image. The semantic concept is emphasized only in those regions that already belong to it in the input image; other regions remain unprocessed. Photo attribution: Stefan Perneborg.

The semantic component is again based on z values, but this time with an $8 \times 8 \times 8$ histogram in sRGB color space. The tone-mapping curve is derived as before with Equation 4.2 and for each color channel separately. The only difference is that the δ value is omitted as this is accounted for by the weight map ω (Fig. 5.1(b)). The three tone-mapping curves derived for the semantic concept of *autumn* are reproduced in Figure 5.1(c).

We apply the derived tone-mapping on each color channel of the input image I_{in} resulting in a globally processed image I_{tmp}. As explained before, the final output has to show processed pixels only in those regions that belong to the semantic concept. The output image I_{out} is thus a linear combination of the input image and the intermediary globally processed image I_{tmp}, and the weights

are taken from the weight map ω:

$$I_{\text{out}} = (1 - \omega) \cdot I_{\text{in}} + \omega \cdot I_{\text{tmp}} \tag{5.3}$$

The resulting output image I_{out} is reproduced in Figure 5.1(d). Note that the image does not have a global color cast, but the semantic concept is emphasized only in image regions that are already part of it in the input image. Figure 5.2 shows more examples for the semantic concepts of *grass*, *strawberry*, *gold*, and *sunset*.

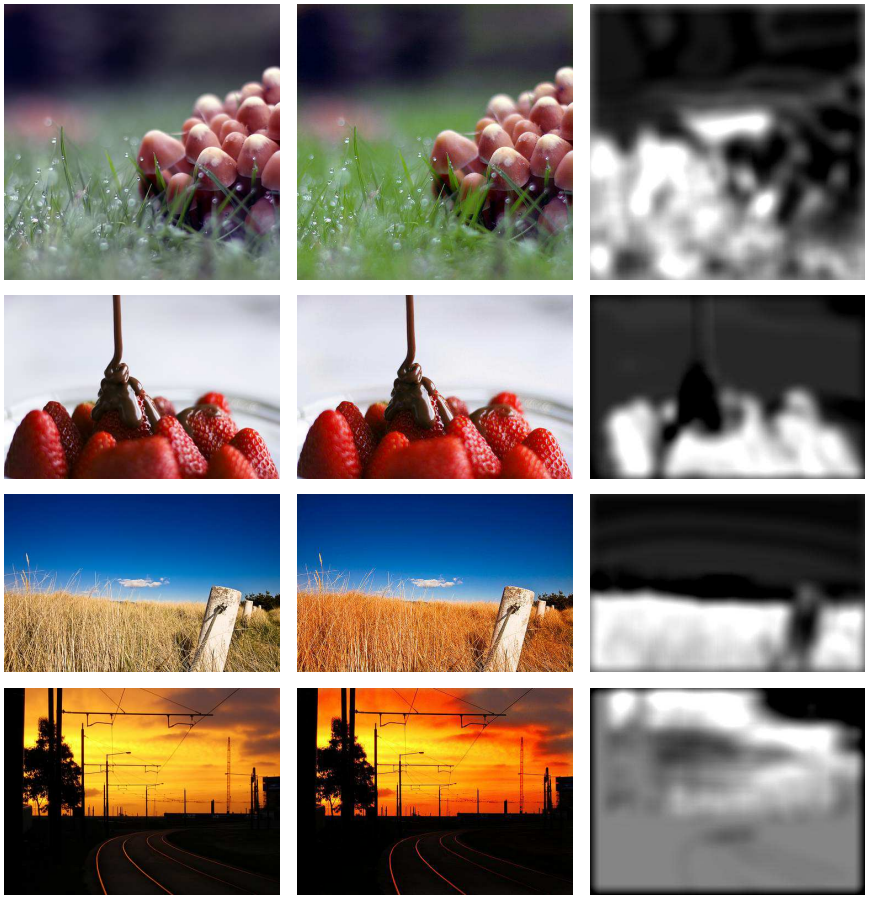
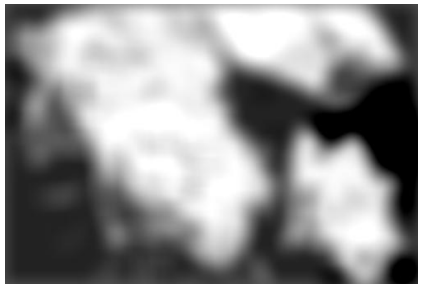


Figure 5.2: More examples of the semantic color transfer for keywords *grass*, *strawberry*, *gold* and *sunset*. The three columns show the input image, output image and weight map, respectively. Photo attributions from top to bottom: Roland Polczer, José Eduardo Deboni, Meredith Farmer, and Gopal Vijayaraghavan.

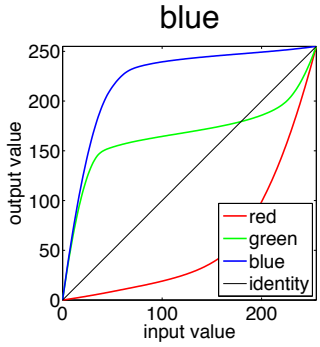
5.1.1 Semantic color transfer

Additionally, our algorithm for the semantic color enhancement is able to handle different semantic concepts for the tone-mapping curves and the weight map, respectively. Hence, it can be used for semantic color transfer. Figures 5.3(a) and 5.3(b) show an image of a rose and the associate weight map for the concept of *rose*. However, the tone-mapping we apply stems from the keyword *blue*, as shown in Figure 5.3(c). Figure 5.3(d) shows the output image in which the roses are colored in blue. This is similar to other color transfer methods [103, 62]. Note, however, that our method handles an arbitrary semantic expression.

(a) input



(b) weight map, *rose*



(c) tone-mapping curves

(d) output, *rose* → *blue*

Figure 5.3: Color transfer using two different semantic concepts. Top: input image and weight map for the semantic concept of *rose*. Bottom: tone-mapping curves for semantic concept *blue* and final output image. The roses are colored in blue. Photo attribution: Philipp Klinger.

5.1.2 Failure cases

The color enhancement workflow is more difficult than the semantic tone mapping presented in the previous chapter. The tone-mapping is a global operation,

which proved to be considerably robust in our experiments. However, the semantic color enhancement and color transfer require a local processing that we realized with weight maps. The weight maps are in the current implementation the main reason why the proposed algorithms fails in some cases.

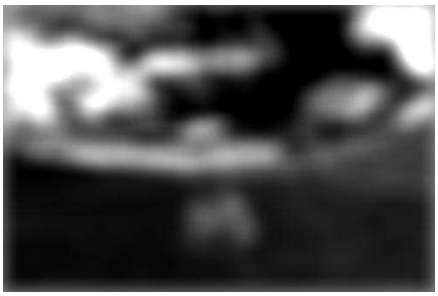
Figure 5.4 shows a failure case for the semantic concept of *sky*. The algorithm re-colored the cloud in the upper right corner in blue, which looks unnatural. The reason for the failure is an erroneous weight map as shown in Figure 5.4(c). Our detection method to find regions that correspond to a semantic concept is based on colors only. In this case the very dark clouds are classified as part of *sky* because it correlates also with dark grays in the MIR Flickr database.

Figure 5.5 shows a different failure case for the semantic concept of *strawberry*. The reason for the failure here is, that the keyword in the image's annotation is not present in the image. In this case the algorithm has difficulties to determine the corresponding image regions, which results in an unpleasing red shift more or less everywhere in the image.

This is due to the rescaling of the weight map to the interval [0 1] as indicated in Equation 5.2. This issue could be avoided by aborting the re-rendering if the values in the intermediary weight map $\tilde{\omega}$ fall below a significance threshold. However, the choice of a good threshold is not trivial.

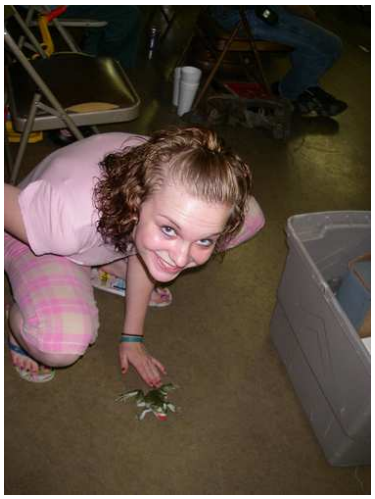
It is thus desirable to develop alternative methods to compute more robust weight maps. This can be realized by more advanced computer vision algorithms that use other characteristics than just color, but this is out of the scope of this thesis.

(a) input (b) output, *sky*



(c) weight map, *sky*

Figure 5.4: Failure case for the semantic concept of *sky*. The cloud in the top right corner has been mistaken for *sky* and re-colored in blue. The reason is an erroneous weight map based on color information. Other computer vision techniques can improve detection results, but this is out of the scope of this thesis. Photo attribution: Mani Babbar.



(a) input

(b) output, *strawberry*

(c) weight map, *strawberry*

Figure 5.5: Failure case for the semantic concept of *strawberry*, which occurs in the annotation but not in the image. Consequently, the enhancement for *strawberry* produces an unpleasing result. Photo attribution: Robert Batina.

5.2 Semantic depth-of-field adaptation

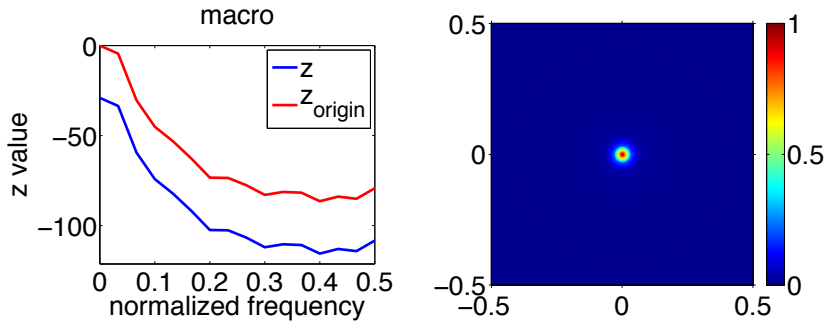
Defocus magnification is important in cases where a photographer intends an artistic blur of the background in order to accentuate the object in focus. In order to demonstrate the versatility of the presented statistical framework, we show how the significance values can be used in this context.

To account for the semantics, we compute z values describing the spatial frequency content in the Fourier domain. We do not distinguish between different orientations and thus obtain a radially averaged one-dimensional descriptor with 16 bins. The first bin describes the DC component and the lowest frequencies and the following bins describe increasing frequencies, respectively. The example plot in Figure 5.6(a) shows that the keyword *macro* relates to an absence of high frequencies, as indicated by the negative z values.

As we do not want to alter the brightness of the image, we shift the curve up with an additive constant so that the first z value (representing the DC component) is equal to zero. These shifted values are denoted z_{origin} as their graph starts at the origin. We then compute the necessary change in the frequency domain similar to Equation 4.2:

$$F = \begin{cases} 1/\left(1 + S \cdot |z_{\text{origin}}|\right) & \text{if } z_{\text{origin}} < 0 \\ 1 + S \cdot z_{\text{origin}} & \text{if } z_{\text{origin}} \geq 0 \end{cases} \tag{5.4}$$

where S is a proportionality constant that controls the overall strength, and F is the filter in the Fourier domain. In order to multiply it with the Fourier transform of an image we generate a radially symmetric version with a simple linear interpolation as shown in Figure 5.6(b) for $S = 1$.



(a) significance values (b) filter F

Figure 5.6: Left: z and z_{origin} values for semantic concept *macro*. The negative values indicate an absence of high frequencies. Right: Corresponding multiplier in the Fourier domain computed with Eq. 5.4 and $S = 1$; it has a strong low-pass behavior.

Similar to our two previous image enhancement examples, we not only implement a semantic component, but also an adaption to the input image itself. In this case, we need a map indicating regions with only low frequency content as it is done in defocus estimation [114]. Figures 5.7(a) and 5.7(b) show an image and its corresponding defocus map reproduced from Zhuo and Sim [114].

We compute an intermediary image I_{tmp} using the input image I_{in} and the filter F:

$$I_{\text{tmp}} = \mathscr{F}^{-1}\Big(\mathscr{F}(I_{\text{in}}) \cdot F\Big) \tag{5.5}$$

where $\mathscr{F}(\cdot)$ and $\mathscr{F}^{-1}(\cdot)$ denote the Fourier transform and its inverse, respectively.

We again use a linear weighting of the images I_{in} and I_{tmp} (Eq. 5.3), where the weights are taken from the defocus map. The final output is shown in Figure 5.7(c). Note that the background is more blurred than in the input image, whereas the boy remains in focus. Figure 5.8 show another example for the semantic concept of *flower* that has a similar characteristic as *macro*.

(a) input (b) weight map, *macro*

(c) output, *macro*

Figure 5.7: Example for the semantic concept *macro*. Top row: input and associated weight map. Bottom: Output image; note that the background is more blurred whereas the boy remains in focus. Input image and weight map reproduced from Zhuo and Sim [114].

(a) input

(b) weight map, *macro*

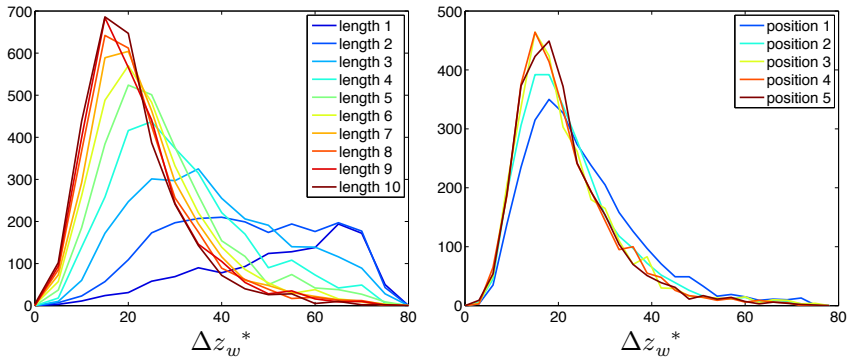
(c) output, *macro*

Figure 5.8: Same as Figure 5.7, but for the semantic concept of *flower*. Input image and weight map reproduced from Zhuo and Sim [114].

5.3 Improvements and extensions for future semantic image re-rendering

The re-rendering applications in this and the previous chapter process an image according to one single keyword. However, images are often annotated with more than one keyword and it is yet unclear how the proposed methods can be extended to include multiple keywords.

We observed that a keyword's significance depends on the total number of keywords in the annotation string as well as its absolute position in the annotation string as illustrated in Figure 5.9. The histograms in the left plot show that a keyword is more significant the fewer keywords are in the annotation. The right plot indicates that a keyword is slightly more significant if it is at the beginning of the annotation. However, this is effect is not as strong and holds only for the first position. The dependency on the annotation length is stronger because the annotation length is a stricter criterion; e.g. a keyword in an annotation of length two has to be at the first or second position.



(a) dependency on annotation length (b) dependency on position in annotation

Figure 5.9: Histogram of $\Delta z_w{}^*$ values for 2858 keywords and the RGB characteristic used for the semantic color enhancement. A keyword's significance depends on two factors: the total length of the annotation string, i.e. number of keywords, (left) and its position within the annotation string (right).

This dependency can be used to weigh the influences of multiple keywords a priori. Keywords that occur in short annotations or at the very beginning of an annotation get more influence on the semantic re-rendering than keywords in longer annotations or keywords further back.

5.4 Chapter summary

In this chapter we presented three additional semantic re-rendering applications, which are color enhancement in Section 5.1, color transfer in Section 5.1.1 and depth-of-field adaptation in Section 5.2. The basic principle of the workflow is the same as for semantic tone-mapping (see Fig. 4.1 in Chapter 4), but the semantic component, the image component and their fusion are implemented differently. In the case of color enhancement we use RGB histogram characteristics to determine a 3-channel tone-mapping and spatial weight maps to confine the color enhancement to relevant regions. In the case of depth-of-field adaptation we use a Fourier domain characteristic to determine an appropriate low-pass filter and a defocus estimation as weight map.

Chapter 6

Color Naming

The goal of color naming is to associate a semantic expression, usually a color name, with its corresponding color value. This chapter describes how the statistical framework from Chapter 3 can be used to accomplish this task automatically without any psychophysical experiments. Section 6.1 discusses traditional color naming, where the semantic expressions are color names. Section 6.2 then extends the discussion to semantic expressions other than color names.

6.1 Traditional color naming

Traditionally, color naming focusses on color names such as *red* or *yellow*. This section demonstrates how the statistical framework automatizes this task on an example dataset of 50 color names with publicly available ground truth. The results are discussed with respect to color accuracy and more technical details of the estimation process.

6.1.1 Dataset

We use the 50 most common[1] color names from Nathan Moroney's color naming experiment [66]. N. Moroney implemented the experiment as a web site where observers see a uniform color patch with an adjacent text box to type in the color's name. A dense sampling of the color space and an aggregation of multiple observer's responses then leads to color value estimations of the different color names. The color name's sRGB and CIELAB values can be downloaded from the web page and form the ground truth for this experiment. The color patches in Figure 6.2 show the 50 color names along with their estimated color value (see section below).

[1] status on October 20, 2011

We downloaded for each color name 200 JPEG images using Flickr's API [26]. The search query was simply the color name itself. The downloaded images are assumed to be in sRGB color space.

6.1.2 Determine a color names's color values

The significance values z_w^j of the statistical framework introduced in Section 3.1.1 are a measure of association between a semantic expression w and an image characteristic j. In the context of color naming, the characteristic is a 3-dimensional CIELAB histogram with $15 \times 15 \times 15$ bins in the ranges $0 \leq L \leq 100$, $-80 \leq a \leq 80$ and $-80 \leq b \leq 80$, respectively. The choice of histogram is subject to two compromises. First, the number of bins has to be a reasonable compromise between precision and memory footprint. We used 15^3 bins as discussed in Section 6.1.4. Second, the histogram intervals along the chroma axes are a compromise between too many out-of-gamut bins (large interval) and not enough bins in regions where the gamut is large (small interval). Values outside the range on the a and b axis are clipped to the closest bin.

The significance values of all histogram bins j are computed for a given color name. We then find the bin j^* with maximum z value. The color name's estimated color values are the center CIELAB values of the maximum bin.

Figure 6.1 shows the $z_{magenta}^j$ values in a 3-dimensional heat map. The three orthogonal planes are defined by $L = L_{j^*} = 63.3$, $a = a_{j^*} = 42.7$ and $b = b_{j^*} = -21.3$ and their intersection is in the bin center with maximum significance value z^{j^*}=20.0. For better orientation the bottom plane shows the bin centers' colors for $L = L_{j^*}$.

Figure 6.2 shows an overview of the estimated color values for all 50 color names. One might argue about the one or the other color, but the estimated colors are quite good overall. It is difficult to find clear errors with a brief visual judgment. A more objective evaluation of the estimations' accuracies is given in the following section.

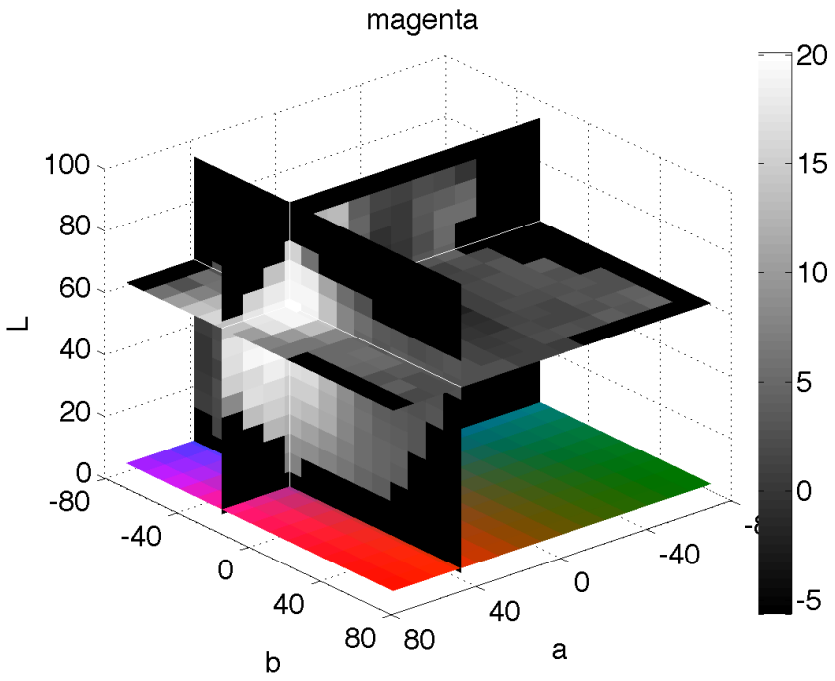


Figure 6.1: The $z^j_{magenta}$ values in a 3-dimensional heat map. The maximum is $z^{j^*}_{magenta}$= 20.0 and is at the crossing of the three orthogonal planes. The homogeneous dark areas along the plane borders are out of gamut values. For better orientation in the ab-plane we show on the plot's floor a plane indicating the colors for the bins with $L = L_{j^*}$.

Figure 6.2: 50 semantic terms with their associated color patches.

6.1.3 Accuracy

We compare our color estimations against the ground truth data from N. Moroney's web site using the widely used ΔE distance measure. Figure 6.3 shows the ΔE distances of all 50 color names. The distance distribution shows that the two estimations are relatively close to each other with a few outliers. It is worth to point out that due to the binning in CIELAB histograms we introduce an inherent quantization error. A color within a bin can have a distance to its center of up to ΔE= 8.2, which is indicated by the dashed red line.

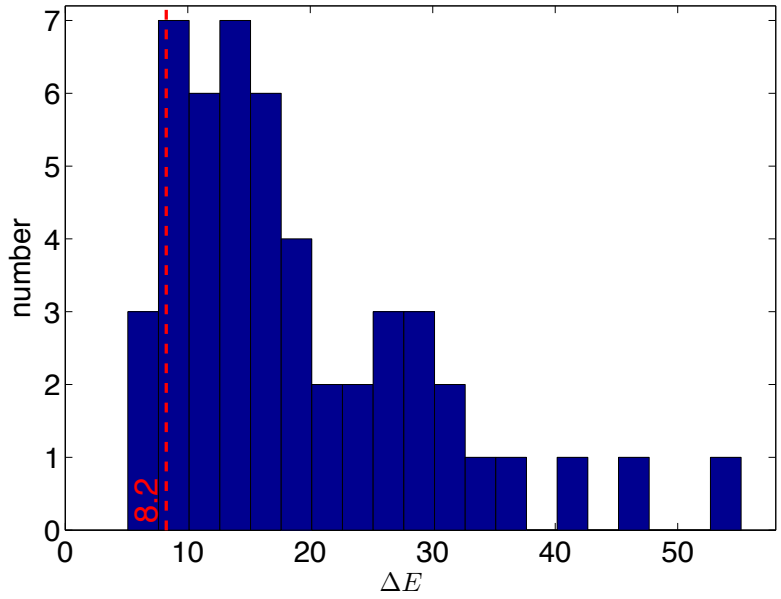


Figure 6.3: The ΔE distance in CIELAB color space when comparing the color values with maximum z value with the values from Moroney's database. The distance a color within a bin can deviate from its center is up to ΔE= 8.2.

The outliers with highest ΔE distances are: *puce* (ΔE= 55.2), *royal blue* (ΔE= 45.3) and *lime* (ΔE= 42.1). While these distances seem large, it is worth to investigate each case in order to gain a better understanding of the framework's functioning

Puce

Our estimate is 27, 13, 10 while Moroney's values are 171, 134, 55. Even though this color name is rarely used and opinions about its correct tint di-

verge (a very complete online color database [18] reports 204, 136, 153), our estimate is clearly too dark.

The reason for this is, that the term *puce* has two other meanings: *Puce Moment*, a music group and *puce* as the french translation of *microchip*. It turns out that *puce* is more often used to refer to the band or the microchip than to the color. The images from the band's concerts have, like most live stage acts, a black background with the band members illuminated in the foreground. The same dominance of black is present in images showing microchips. Thus, black is over-proportionally present in images with keyword *puce* and thus our framework finds this association.

The true origin of the color name *puce* is different. It comes from the French word for flea, puce, possibly a reference to the 16-19th century source of the carmine dye colour that was extracted from Mexican scale insects (resembling fleas).

We see that the reason for the large deviation is semantic ambiguity. This can be seen as a positive and a negative point. If the task was to find the exact values for the color *puce* it is better to do a color naming experiment since human observers do understand the semantic ambiguity. However, if the task is to find what the semantic expression means for the majority of images, our estimate is better.

Royal blue

There exist at least two kinds of royal blue: a traditional royal blue 0, 35, 102 [18] and a modern royal blue as defined by the Word Wide Web Consortium (W3C) 65,105,225 [101]. Society's perception must have changed from the darker version to the brighter version over time. Our and Moroney's estimates are 19, 49, 107 and 39, 41, 212, which are closer to the original and the modern version, respectively.

The reason why the statistical framework ranks the traditional royal blue first is due to the "Royal Blue Coach Services", an English coach operator from 1880 to 1986. Their coaches were varnished in traditional royal blue; which is obvious when considering the early founding year. The coaches seem to have a very active fan community that preserves them for nostalgic reasons. They also post many pictures online so that the analysis ranks this color first.

Again, our estimate is different from what one would expect at first sight but it is not wrong. The distance between the color *traditional royal blue* and our estimate of the semantic expression *royal blue* is $\Delta E = 12.6$. The distance between Moroney's estimate and the W3C's definition of *royal blue* is much higher: $\Delta E = 40.5$.

Lime

There is no straight forward explanation why the estimate 186, 204, 124 is not bright and saturated enough. Moroney's estimate for this color name is better: 106, 239, 59. The best explanation to give is that an estimation is only correct with a certain probability.

For a given semantic expression (i.e. color name) our system computes a z value for all possible color values. So far we considered only the color value with the highest z value, but for a deeper insight also the other z values need to be analyzed.

In order to consider all z values we rank for a given color name the color estimates with decreasing z value. We computed for the best 1000 color estimates the ΔE distance to Moroney's value. The values for the first rank are thus the ones shown in the histogram in Figure 6.3 (the histogram shows the color estimate with highest z values, i.e. 1st rank). The following 999 distances are the deviations for the less significant colors.

The results for all the 50 color names are summarized in Figure 6.4. It shows the rank on the logarithmic horizontal axis and the ΔE distance on the vertical axis. The deviations continuously grow for increasing ranks and become more prone to noise. The graph illustrates that it is not possible to guarantee a specific error. But it shows that, from a probabilistic viewpoint, colors that are ranked first are better estimates.

6.1.4 Dependency on number of bins

The only parameter our framework depends on is the number of bins in the histogram. To show its effect on the results, we compute the ΔE distances between ours and Moroney's estimates for $2^3, 3^3, 4^3, \ldots, 32^3$ histograms bins. Figure 6.5 shows the median and 25% and 75% quantiles of the ΔE distance as a function of the number of bins. The additional red curve is the maximum quantization error, which is the distance between the bin center and bin corner. Please note that the horizontal axis is not linear, but cubic.

It is visible that the error is high for very small bin numbers and then decreases for higher number of bins. The plot also shows that the error stops improving for approximately 12^3 or more bins. Our choice of 15^3 bins is thus on the safe side, but not excessively high.

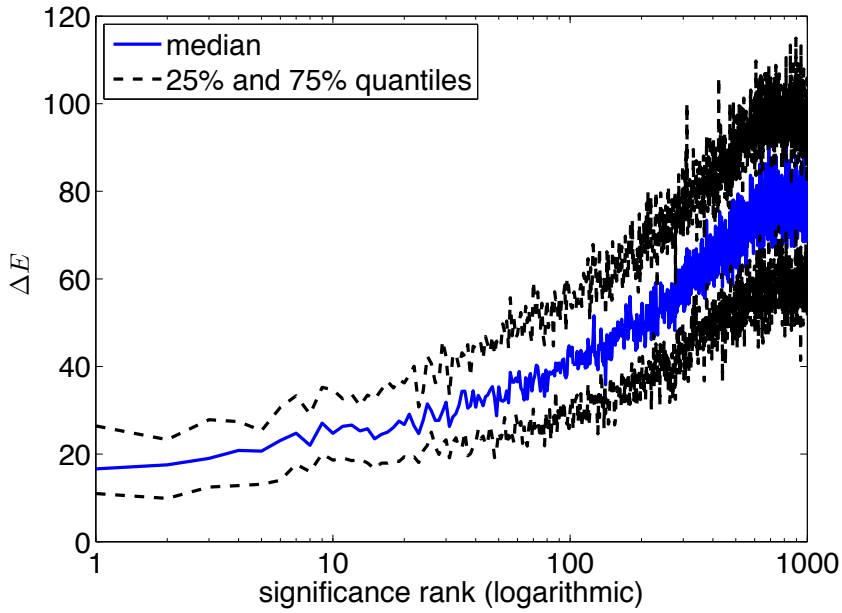


Figure 6.4: ΔE distances between Moroney's 50 color names and our estimations. We compare not only our best estimate (color value with the highest z value) but the first 1000 estimates (sorted by decreasing z values). This significance rank is plotted along the logarithmic horizontal axis. It is clearly visible that color estimates on the first ranks have smaller errors.

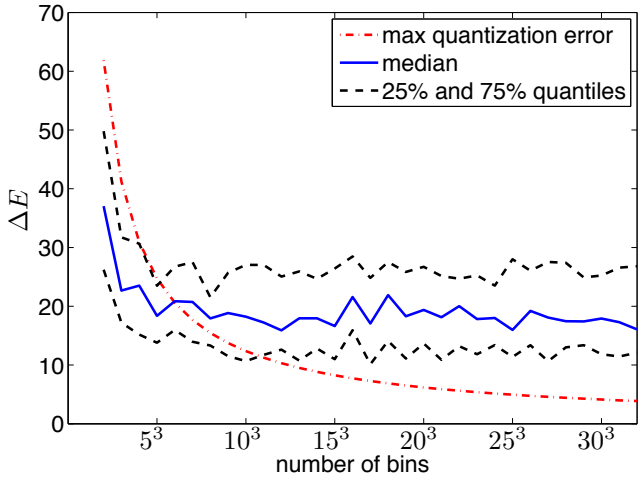


Figure 6.5: Median and 25% and 75% quantils of ΔE error between ours and Moroney's estimates as a function of the number of bins. Please note that the horizontal axis is not linear, but cubic. The red curve shows the maximum quantization error, the distance between a histogram bin's center and corner.

6.2 Other semantic expressions than color names

In this section we discuss the estimation of color values for semantic expressions that are not color names. We first show results for memory colors in Section 6.2.1 and then for arbitrary semantic expressions in Section 6.2.2.

6.2.1 Memory Colors

We use the three basic memory colors, which are *vegetation*, *skin*, and *sky*. We then chose other additional keywords that modify the tint of the memory colors in a distinct way. We combined *vegetation* with *wet*, *dry*, *leaves*, *bush*, further *skin* with *caucasian*, *tan*, *bright*, and *dark*, and finally *sky* with *sunny*, *rain*, *overcast*, and *sunset*. We then downloaded 500 images for each combination.

The rows in Figure 6.6 show the output of our statistical analysis for the different combinations of memory colors. It is clearly visible how the shade of a memory color varies with the specific context; e.g. tanned skin is darker than caucasian skin. The variations of a memory color can be very extreme, such as for sky under different environmental conditions.

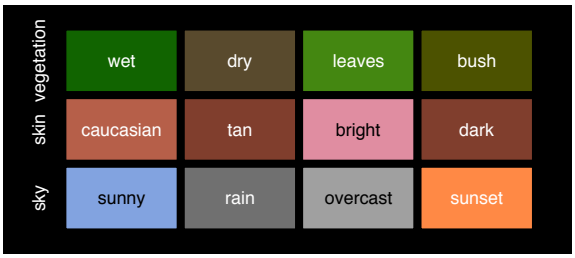


Figure 6.6: Example memory colors from our automatic algorithm. The three basic categories (vegetation, skin and sky) are further refined by the additional keywords indicated in the center of each patch.

To give an intuition of the z value distribution and how it changes for different semantic expressions we show in more details the cases *sky+sunny* and *sky+sunset*. In order to show the z values on a plane we computed also color histograms on the ab-plane, discarding the luminance information.

Figure 6.7 shows the bin centers' colors in the ab-plane and the corresponding z value distribution as a heat map. One sees how the expression *sunset* causes the z values to rise in the orange and red regions of the histograms. For *sunny* the highest z values are as expected in the blue region.

We compare our memory color values with Yendrikhovskij's values [110]. Figures 6.8(a) to 6.8(c) each show his ellipses for *vegetation*, *skin* and *sky* in the

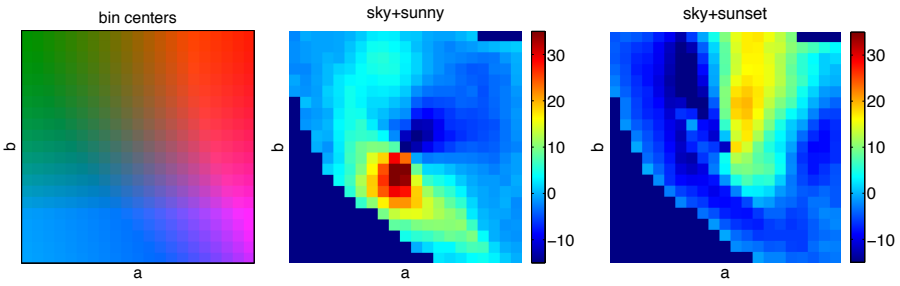
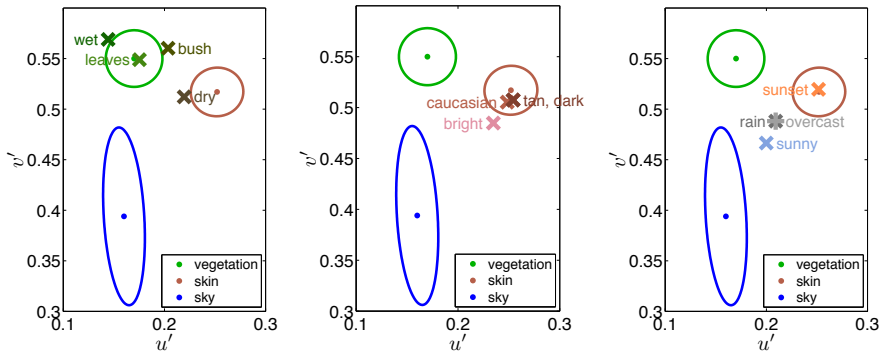


Figure 6.7: The map on the left shows the colors of the bin centers. In the middle and on the right are the z value distributions for *sky+sunny* and *sky+sunset*, respectively. The dark blue homogeneous areas are out of gamut values.

$u'v'$ plane, respectively. For clarity we do not show the whole z distribution for each keyword combination, but only the value with maximum z value. They are plotted as labeled cross-marks in the respective color.

Our values lie within or relatively close to the ellipsis for *vegetation* and *skin*. Our estimations for *sky* differ more from Yendrikhovskij's ellipse. The reason is that sky drastically changes under the different weather conditions we used for this experiment.



(a) vegetation (b) skin (c) sky

Figure 6.8: All three subfigures show the ellipses from Yendrikhovskij [110] for *vegetation*, *skin* and *sky*. Our results are visualized with crossmarks: (a) variations of *vegetation*, (b) variations of *skin*, (c) variations of *sky*. For clarity only the color estimates with maximum z values are shown, not the complete z distribution.

The distinction between different varieties of a memory color is significant. This is crucial for high quality image rendering since images with wrong memory colors appear unnatural [110]. There is not a single vegetation green in the world, but it visibly changes across landscapes and human observers expect to see it the way they know it. The same holds for skin tones and sky blues.

6.2.2 Arbitrary semantic expressions

In the next experiment we do not limit the semantic expressions to memory colors or color names. We downloaded for 20 randomly chosen semantic expressions 200 images each. Figure 6.9 illustrates the semantic expressions[2] and their associated colors. Even though one might want to argue about the correct tint of the one or the other example, they are all reasonable estimates and demonstrate that our approach is not limited to color names only.

Figure 6.9: 20 arbitrary semantic expressions along with their estimated color values. The Figure demonstrates that our approach is not limited to color names only.

6.2.3 Association strength

We finally show that, apart from assessing an associated color value, the z values can be used to estimate the association strength. The higher the z value the more an association is significant. Thus, semantic expressions that are very meaningful in terms of color have a higher z value.

We compare the maximum z values from the color names (Section 6.1) and the arbitrary semantic expressions (Section 6.2.2). Figure 6.10 shows the max-

[2] *granny smith* is a green kind of apple and *lakers* is a basketball team from the United States with a violet and yellow outfit.

imum z values of both sets in a histogram plot. The highest z values are solely from color names. This is not surprising since color names have a stronger link to colors by definition. The highest values stem from *red* (28.8), *yellow* (26.3) and *purple* (26.0). Among the arbitrary semantic expressions the highest z_{max} values are obtained for *granny smith* (14.9), *ferrari* (14.4) and *smurf* (13.6).

For the sake of completeness we downloaded also 200 images with keywords for which we expect low z_{max} values. The results are: *poster* (4.5), *painting* (4.0) and *boredom* (2.5). The reason why the z values are low is straight forward. None of these semantic expressions can be associated with a specific color, even though *poster* and *painting* might be colorful in general.

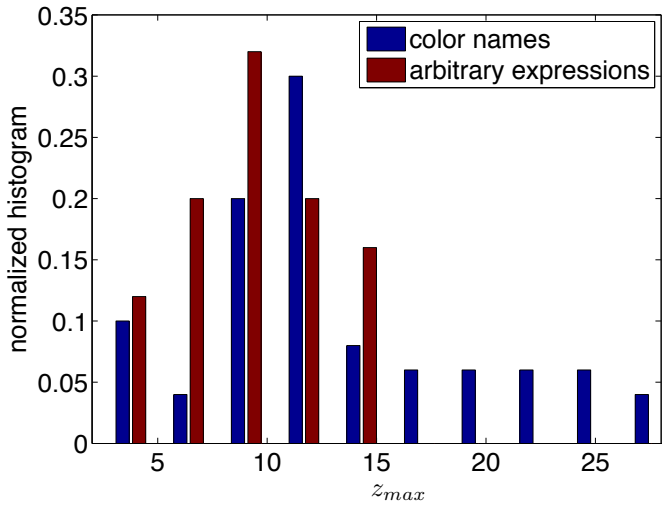


Figure 6.10: Histogram of the maximal z values for the color names (Section 6.1) and the arbitrary semantic expressions (Section 6.2.2). The color names have higher z_{max} values and are thus stronger associated with color.

We see that the statistical framework allows to relate a color value to any semantic expression. Moreover, the significance value indicates how relevant the color value is for the expression.

6.3 Chapter summary

This chapter presented the application of our statistical framework to color naming. We started with 50 English color names in Section 6.1 and discussed the accuracy of the estimations. We especially pointed out the problem of semantic ambiguity that we observe for the color names *puce* and *royal blue*. The color

estimations were then extended in Section 6.2 to other semantic expressions than color names. We first showed results for the three memory colors *sky*, *skin* and *vegetation* and compared them against ground truth from Yendrikhovskij [110]. Then we picked 20 arbitrary semantic expressions such as *dolphin* or *pretzel* and demonstrated that the statistical framework also handles these cases.

Chapter 7

A Large-Scale Multi-Lingual Color Thesaurus

In this chapter we extend the color naming experiment from the previous chapter to over 9000 color names in ten languages. We explain how we acquired the list of color names and then build the database in Section 7.1. In Section 7.2 we show how we estimate a color value given a color name's distribution. Then we discuss the accuracy of the estimations in Section 7.3. We highlight language related imprecisions that are important due to the usage of ten different languages. The large amount of estimated colors allows us to do a more advanced statistical analysis of the estimations, which is presented in Section 7.4. Section 7.5 introduces an interactive web page that makes the color estimations easily accessible. Finally, we discuss the topic of automatic color naming in Section 7.6.

7.1 Building a Database

We took the 950 English color names that were derived in the XKCD Color Survey [19] and translated them into nine other languages, namely Chinese, French, German, Italian, Japanese, Korean, Portuguese, Russian, and Spanish, respectively. Each translation has been done by a native speaker with a good level of English.

In some cases the translation of a color name is difficult, because the destination language does not have this precise color name, or because two varieties of a color name in English translate to the same expression in the destination language. Examples are the four color names *burple*[1], *purpleish blue*, *purpley*

[1] A combination of *blue* and *purple*

blue, and *violet blue*, which all translate to the same expression in Chinese.

We download for all color names and all languages 100 images each, using Google Image Search. To guarantee that we acquire only images from the present language we use the **cr** (country restrict) and **lr** (language restrict) fields as defined in Google's Custom Search API [32]. This is important for color names such as *rose* that have the same spelling in English and French. A simple query for *rose* would therefore lead to an undesired mixed search result from both languages. The search query is the "color name" in quotes plus the word color in the respective language. Two example queries are *"cloudy blue"+color* and *"bleu nuageux"+couleur* for English and French, respectively.

A complete set for one language comprises $100 \times 950 = 95\,000$ images, which has a download time in the order of one day. This process can run in the background as it does not require significant computational power. We assume that the downloaded JPEG images are encoded in sRGB color space.

7.2 Color value estimation

We perform two steps to determine for a given color name its estimated color values $L^{\text{est}}, a^{\text{est}}, b^{\text{est}}$. First, we find the maximum bin of the z value distribution. As the bin centers are quantized we do an interpolation step in the neighborhood of the maximum bin. We compute a weighted mean over the 27 bin neighborhood $\mathcal{N}$ in 3-dimensional CIELAB space, where the weights are given by the z values: $L^{\text{est}} = \sum_{i \in \mathcal{N}} z_i L_i \Big/ \sum_{i \in \mathcal{N}} z_i$, where L_i is the L value that corresponds to bin i. The a^{est} and b^{est} values are computed accordingly.

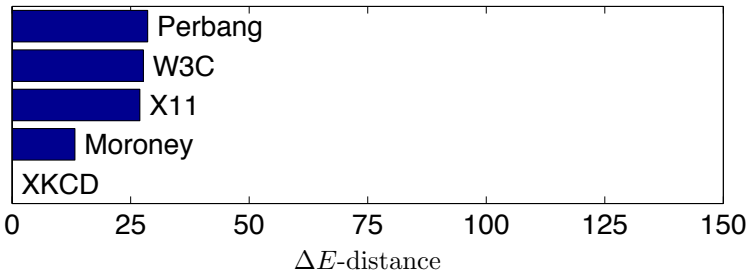
7.3 Accuracy analysis

A good way to measure the accuracy of an estimation is to compare it against ground truth. However, this is difficult in color naming due to the lack of reliable ground truth data. In fact, it is almost impossible to create reliable ground truth data, because color naming involves natural language, which is too vague for a strictly quantitative validation as explained in the following section.

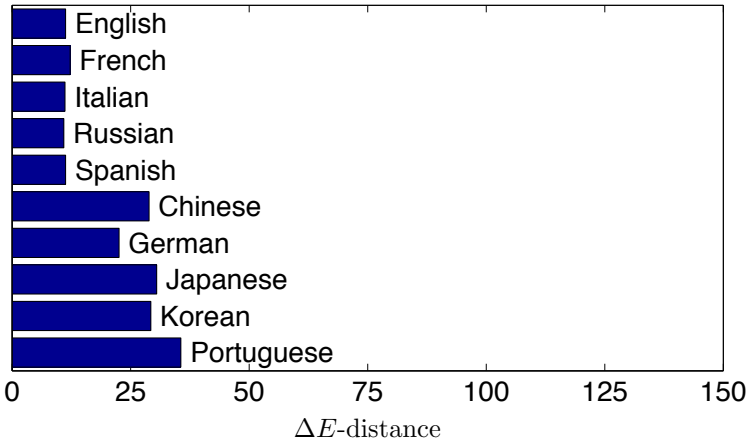
7.3.1 Language-related imprecisions

Let us consider the color name *maroon* whose sRGB values are given in several color databases: 64, 35, 39 (Perbang, an online color database [18]), 128, 0, 0 (W3C's CSS Color Module Level 3 [101]), 176, 48, 96 (X11 Color Names [109]), 140, 28, 61 (Moroney's web-based experiment [59]), and 101, 00, 33 (XKCD Color Survey [19]). Our estimate for English is 100, 32, 40. It is difficult to decide with certainty which of the color values represents the true *maroon*.

Figure 7.1(a) shows the ΔE distances between the color values for *maroon* from the XKCD database and the other databases. The distances between arbitrary pairs of databases are even larger; the maximum is for Perbang's and W3C's values: ΔE=49. For a better visual comparison, the horizontal axes in Figures 7.1 and 7.2 have the same scale.



(a) distance between XKCD and different databases



(b) distance between XKCD and ten different languages

Figure 7.1: Top: ΔE distances between the color value for *maroon* from the XKCD database and the values from other databases. Bottom: ΔE distance between the XKCD value for *maroon* and our estimations for all languages. The horizontal axes have the same scale as the ones in Figure 7.2 for a better visual comparison.

We argue that a discussion about the true color value of *maroon*, and any other color name, is strongly influenced by opinions/tastes and can not be taken as a fact. Consequently, a performance evaluation such as measuring the widely used ΔE distance in CIELAB space between our estimates and a ground "truth" has to be carefully interpreted (see Fig. 7.1).

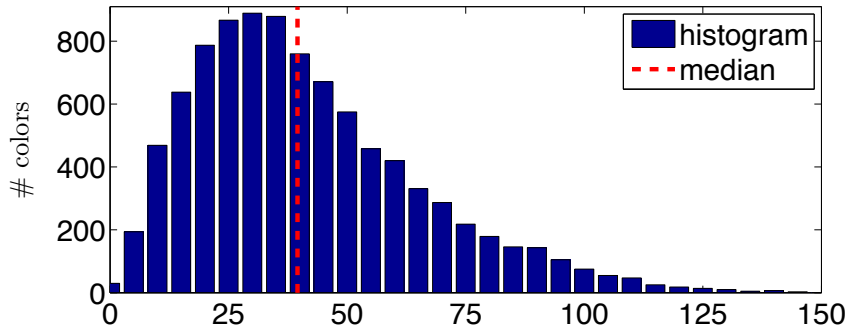
It is also non-trivial to compare results from translations of a single color name into different languages. Our French translation for *maroon* is *bordeaux* and we estimate it as 83, 20, 30. If we translate the French expression back to English we could also say *bordeaux red* or *dark red*, which makes the French estimation justifiable. The German translation is *kastanienbraun*, which literally means *chestnut brown*. Hence, our estimation has a brown tint 70, 29, 27. The Italian translation is *rosso bordeaux*, which means *reddish bordeaux* and our estimation is accordingly more reddish 101, 33, 41. For Portuguese we have *castanho* (*chestnut*) and obtain 73, 54, 41. The Chinese translation is 栗色 (*chestnut* + *color*) 63, 33, 25, the Korean is 적갈색 (*reddish brown* + *color*) 39, 0, 0, and the Russian is бордовый (wine red) 85, 19, 31. The Japanese color name is the same as the Chinese, because the translator could not find a corresponding expression and thus used the Chinese vocabulary; a common practice in Japan. Nevertheless, we estimate a different value as we use Google's language and country restrictions: 96, 62, 48.

7.3.2 Overall accuracy

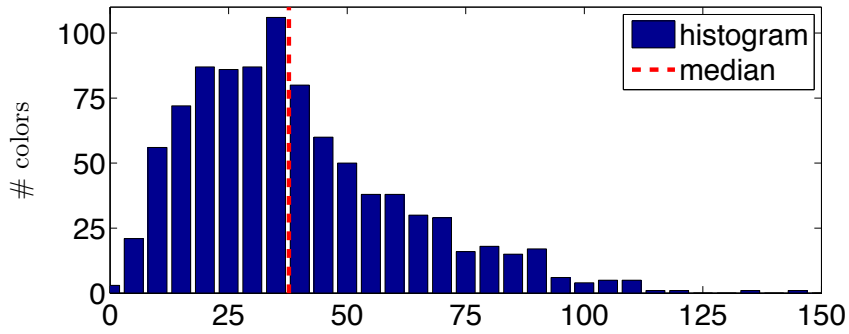
The ΔE distances between the XKCD value and our estimations for all languages are plotted in Figure 7.1(b). We can split the languages into two groups. First, languages in which *maroon* has been translated to some type of *red* (top 5 in Fig. 7.1(b)). In these cases the ΔE distances are lower than for any database (see Fig. 7.1(a)). In the other group of languages the translation is related to *chestnut* and *brown*. In these cases the estimations are more brownish and the ΔE distances are higher.

Figure 7.2(a) shows the ΔE distances for all color names in all languages between the estimated values and the XKCD value for English. Considering the large distances for a single color name between different databases (e.g. up to ΔE=49 for *maroon*), the estimations are in a reasonable range. Figure 7.2(b) shows the ΔE distances for only the English color names. As the color names come from the same language there are no additional deviations due to the translation. Consequently, the ΔE distances are smaller than in the global set.

Figure 7.3 shows color patches for 50 color names in ten languages. The patches are sorted by increasing hue angle of the English color estimation. We see that these example estimations are correct within expected variations due to language and translation imprecisions.



(a) ΔE-distance: all languages



(b) ΔE-distance: English

Figure 7.2: ΔE distances between the English XKCD color values and our estimations for all languages (top) and only the English terms (bottom). The distances are in a reasonable range considering that color values are subject to vaguenesses of human languages and deviations from translations. This is demonstrated in the text at the example of the color *maroon* and in Figure 7.1. The dashed red lines indicate the medians of the distributions.

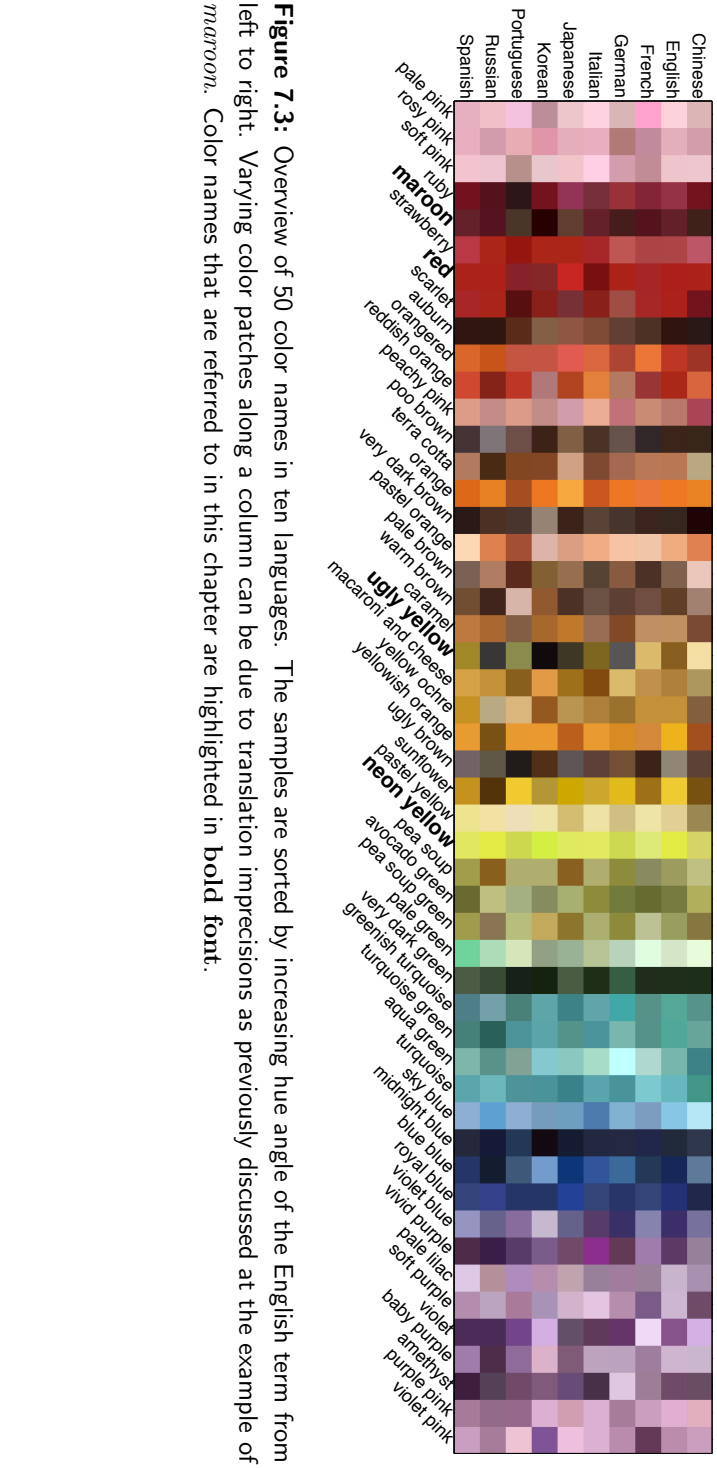


Figure 7.3: Overview of 50 color names in ten languages. The samples are sorted by increasing hue angle of the English term from left to right. Varying color patches along a column can be due to translation imprecisions as previously discussed at the example of *maroon*. Color names that are referred to in this chapter are highlighted in **bold font**.

7.3.3 Failure cases

We show two failure cases in Figure 7.4 in order to discuss the limitations of the statistical approach. Korean is a single outlier among all estimations for *raspberry*. The Korean expression for this color is 나무딸기 where the first two characters mean *wood* and the last two *strawberries*. The image results in Korean show raspberries in the woods with a significant amount of green leaves so that green is the most dominant color. The color name *greenish tan* produces ambiguous results. For some of the languages, the framework estimates rather green colors and for others rather tanned skin colors. An interesting case is the German translation *grünlich hellbraun* (*greenish light brown*), which is due to the fact that the German expression for *tanned* literally means "*browned*". However, *greenish light brown* is an expression that is so rarely used that even Google Image search can not provide search results for this query. In this case the term *light brown* dominates the modifier *greenish*.

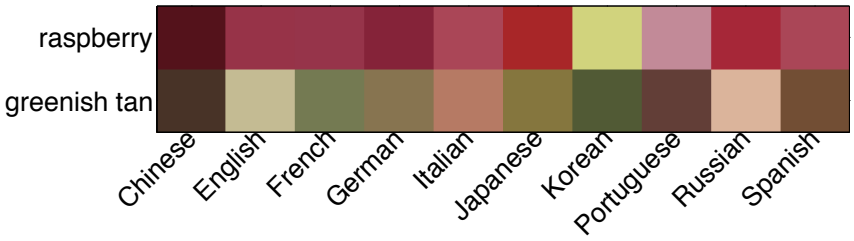


Figure 7.4: Two failure cases. *Raspberry* fails, because the Korean images with raspberries contain a significant amount of leaves. Hence the estimations is a green color. *Greenish tan* is ambiguous and leads to greenish colors in English, French, and Korean and to skin colors in the other languages.

We can see that imprecisions of natural languages limit the precision of the statistical framework in cases where there is semantic ambiguity or where a semantic concept is difficult to express in the given language. However, the difficulty to translate certain colors names to other languages is a general problem of language and not a drawback of the automatic estimation.

Figure 7.5 shows for all ten languages the color with the lowest ΔE distance to the XKCD ground truth in the top row and the ones with the highest ΔE distance in the bottom row. Each patch shows the estimated color on the left and the ground truth on the right. It is interesting to observe that all 10 colors with the lowest ΔEdistance have relatively low saturation. Further research is necessary to assess whether this is just chance or a consistent behavior of the statistical estimation.

Figure 7.5: Top: 10 colors with the lowest ΔE distance to the XKCD ground truth. Bottom: 10 colors with the highest ΔE distance to the XKCD ground truth. All patches show the estimated color on the left and the ground truth color on the right hand side.

7.4 Advanced analysis

The abundance of data allows a more advanced analysis of the estimated significance distributions. In this section we demonstrate two properties: first, higher z values implicate a higher accuracy of the estimated color and second, color names have more variance along the lightness axis than along the two chromatic plane axes in CIELAB space.

7.4.1 Higher significance implicates higher accuracy

Let $\mathcal{L} = \{\text{Chinese, English, French, German, Italian, Japanese, Korean, Portuguese, Russian, Spanish}\}$ be the set of all languages, $\hat{z}_{l,w}$ the maximum z value of the significance distribution of color name w and language $l \in \mathcal{L}$, and $\mathbf{c}^{\text{est}}_{l,w} = (L^{\text{est}}, a^{\text{est}}, b^{\text{est}})^T$ the estimated color triplet in CIELAB space. We then compute for each color name w the average maximum z value over all languages l, denoted $\overline{z}_w$, and the average ΔE distance between any two estimations of different languages $l_1 \neq l_2$, denoted $\overline{\Delta E}_w$:

$$\overline{z}_w = \frac{1}{|\mathcal{L}|} \sum_{l \in \mathcal{L}} \hat{z}_{l,w} \tag{7.1}$$

$$\overline{\Delta E}_w = \frac{1}{|\mathcal{L}|(|\mathcal{L}|-1)} \sum_{l_1 \in \mathcal{L}} \sum_{l_2 \in \mathcal{L} \setminus \{l_1\}} ||\mathbf{c}^{\text{est}}_{l_1,w} - \mathbf{c}^{\text{est}}_{l_2,w}||_2 \tag{7.2}$$

where $|\cdot|$ signifies the cardinality operator and $||\cdot||_2$ the Euclidean distance, respectively.

$\overline{\Delta E}_w$ can be visualized as the average deviation of a color name for different languages. For example the deviations for *neon yellow* are smaller than for *ugly yellow* as can be seen in Figure 7.3, which is reflected in the corresponding values: $\overline{\Delta E}_{neon\ yellow} = 11.5$ and $\overline{\Delta E}_{ugly\ yellow} = 42.1$, respectively. $\overline{\Delta E}_w$ can be high due to estimation errors or translation difficulties as previously discussed for *maroon*.

Figure 7.6 shows the mean, 25% and 75% quantiles of the $\overline{\Delta E}_w$ values as a function of the corresponding $\overline{z}_w$ value. It is visible that the deviation decreases for increasing average significance. The average significance values for the above example are $\overline{z}_{neon\ yellow} = 8.7$ and $\overline{z}_{ugly\ yellow} = 5.0$, which is in accordance with the overall trend. It is important to remember that the z value is a function of the number of images per keyword as explained in Section 3.2.1. Because the estimations in this chapter are done with 100 images per keyword, the values are lower than in the previous chapters.

We conclude that estimations become better for higher significance values. This is the case when the translated color names are well defined and the related images all have a single dominant color.

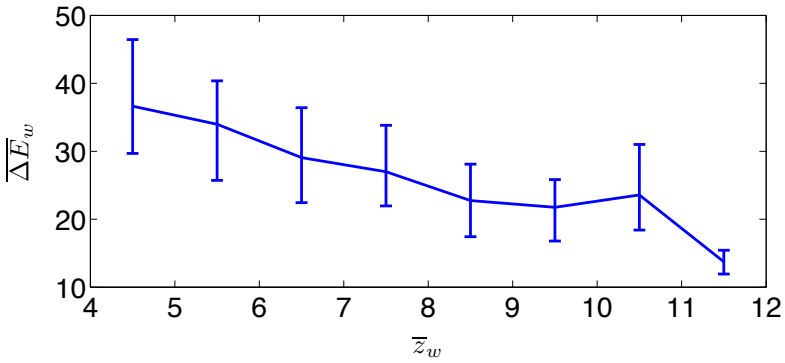


Figure 7.6: $\overline{\Delta E}_w$ (mean, 25% and 75% quantiles) as a function of $\overline{z}_w$. The deviation of color name estimations for different languages decreases on average with increasing significance.

7.4.2 Tints of a color stretch mainly along the L axis

So far we only considered the maximum bin of the significance distribution and its direct neighbors to estimate a color's CIELAB values. However, the distribution itself contains more information that can be exploited for a deeper insight.

The significance distribution has a blob around the maximum bin and its values decrease with increasing distance from the center, as can be seen in Figure 6.1. We compute the 2nd derivative at the estimated color $\mathbf{c}^{\text{est}}$ to determine how quickly the significance values decrease:

$$\left.\frac{\delta^2 z(\mathbf{c})}{\delta L^2}\right|_{\mathbf{c}=\mathbf{c}^{\text{est}}} \approx \left.\frac{z(L^{\text{est}}-\Delta L)-2z(L^{\text{est}})+z(L^{\text{est}}+\Delta L)}{\Delta L^2}\right|_{a=a^{\text{est}},b=b^{\text{est}}} \tag{7.3}$$

where ΔL is the distance between two neighboring bins along the L axis. The equation is analogous for the a and b directions.

The second derivative is always negative in this case, because the z distribution has a maximum at $\mathbf{c}^{\text{est}}$. Therefore, the plot in Figure 7.7 shows its absolute value, i.e. curvature, for convenience. It is visible that the curvature along the L axis is smaller than along the a and b axes.

A similar result is obtained when one fits a Gaussian curve to the z values around the estimated color $\mathbf{c}^{\text{est}}$ in CIELAB-space. We use a symmetric $5\times5\times5$ neighborhood around the center bin and fit a least-squares Gaussian function

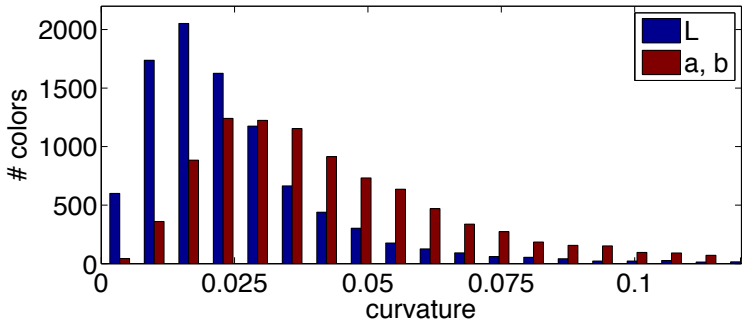


Figure 7.7: Histogram of the absolute value of the 2nd derivative, i.e. curvature, at the maximum turning point of the z distribution. It is visible that the curvature is smaller in the direction of the L axis than for the a and b axes. This means that color names are more independent of small lightness changes than changes in the chromatic plane.

to the significance values:

$$g(\mathbf{c}) = A \cdot \exp\left[-\frac{1}{2}\left(\frac{(L-L^{\text{est}})^2}{\sigma_L^2} + \frac{(a-a^{\text{est}})^2}{\sigma_{a,b}^2} + \frac{(b-b^{\text{est}})^2}{\sigma_{a,b}^2}\right)\right] \tag{7.4}$$

where $\mathbf{c} = (L, a, b)^T$ is a position in CIELAB, σ_L the standard deviation in L direction and $\sigma_{a,b}$ the standard deviation in the a and b directions, respectively. The histogram in Figure 7.8 shows that the spread in the L direction is approximately twice as large as in the a and b directions.

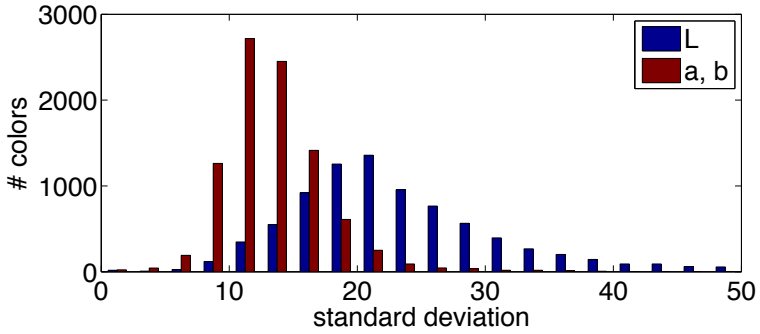


Figure 7.8: Histogram of the standard deviations of the Gaussian curve around the color centers. A color name's spread in CIELAB is approximately twice as large in the L direction than in the two chromatic directions.

This is an intuitive result when looking at basic color names such as *red* or *green*, because they are hue names and allow for more variation along the lightness axis. Our large scale analysis shows that this is not restricted to basic color names but a general trend for all the 9000 color names studied.

7.5 Web page

To make the color estimations easily accessible we designed a color thesaurus web page[2] on which people can explore the colors. It is possible to navigate through color space along the lightness, chroma and hue angle axes or to find similar color in other languages. Further it is possible to search colors by name or to pick them from a color wheel.

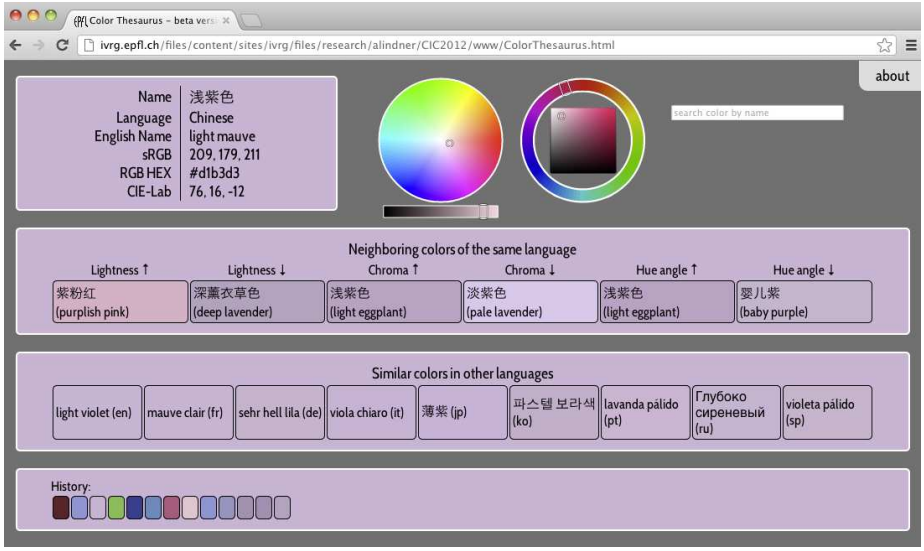


Figure 7.9: Screenshot of the interactive color thesaurus web page. Users can browse through the color space and across languages to explore colors. Colors can also be search by name or picked from a color wheel.

7.6 Discussion

The large-scale multi-lingual color thesaurus demonstrates the strength of a fully automatic approach. To our knowledge, this is the only color thesaurus of that scale in terms of color names and languages. The practicability of the

[2] http://colorthesaurus.epfl.ch/

estimated values is significantly increased with the interactive web page as it allows to wander in different directions in color space and find similar color in other languages.

We encountered language related difficulties during the preparation of the database, and the analysis of the results. One problem is that color names might have a second meaning and that second context is more frequently used than the color context. This is the case for *puce* as shown in Section 6.1.3. Another example for this is the Korean color name for raspberry, 나무딸기. As the first two characters mean *wood* and the last two *strawberries* the acquired images contain forst scenes which cause an erroneous green estimation. A different type of difficulty is that some color names can not be translated directly to other languages. One examples is *maroon*, which is discussed in Section 7.3.1.

It would be interesting to compute the color estimations a second time in the more modern CIECAM02 color space instead of CIELAB. It is possible that the precision of the estimations increases, but this comes at the cost of a computationally heavier color space transformation.

7.7 Chapter summary

In this chapter we presented a large-scale multi-lingual color thesaurus. Section 7.1 presented the acquisition of images for over 9000 color names in 10 languages that are used to estimate their color values as described in Section 7.2. We then discussed the estimated values in terms of language and colorimetric accuracy in Section 7.3. Section 7.4 presented results of a more advanced analysis demonstrating that higher significance correlates with higher accuracy and that tints of a color mainly stretch along the lightness axis. The web site presented in Section 7.5 makes the estimated color values easily accessible to the public.

Chapter 8

Conclusions

8.1 Thesis summary

This thesis had two main goals:

1. Develop a statistical framework that relates image keywords to image characteristic, i.e. bridges the semantic gap. Two important requirements for the framework are: easy-to-use output for subsequent applications and scalability to very large datasets.

2. Demonstrate that imaging applications benefit from a semantic awareness of a scene, which is provided by the statistical framework.

The statistical framework that we presented in Chapter 3 uses a significance test to determine whether a characteristic is dominantly present or absent for images annotated with a given keyword. The measure is expressed as a standardized z value that is positive for a dominant presence and negative for a dominant absence of a characteristic. We use a non-parametric test so that the framework generalizes well to any type of characteristic without a priori knowledge of the underlying distribution. We choose the Mann-Whitney-Wilcoxon test because it is not sensitive to the shape of a probability distribution, but only the median. This is favorable for the applications presented in this thesis. However, other applications might benefit from other tests.

Significance values can be computed for a variety of characteristics such as gray-level and color histograms or spatial layout of structure or linear binary patterns. The result in all cases is a significance distribution, which is a compact summary of a keyword's impact on image characteristics and can be used by subsequent applications.

Large parts of the computations can be done offline and only once per characteristic. The pre-computed intermediary steps can then be loaded and used

for any keyword. This reduces the complexity to estimate the significance of the characteristic for a given keyword to just n_w summation operations, where n_w is the number of images annotated with the keyword w and typically in the order of a few hundred to ten-thousand. This design thus easily scales to millions of images and thousands of keywords.

We proved the usefulness of the significance values for two applications, which are semantic image enhancement and automatic color naming. We demonstrated that both applications benefit from the semantic awareness computed by our framework.

Our semantic image enhancement framework takes two independent inputs, which are an image and a keyword as explained in Chapters 4 and 5. The image is then re-rendered in order to match the semantic context of the keyword. We implemented a semantic tone-mapping, color enhancement, color transfer and depth-of-field adaptation.

The semantic tone-mapping in Chapter 4 was evaluated with two psychophysical experiments. The first one was crowd-sourced on Amazon Mechanical Turk and comprised almost 30'000 pairwise image comparisons of the original and the enhanced images. The observers significantly preferred our images for all except one keyword with approval rates of up to 90%. The one exception was the keyword *light* due to the ambiguity of the word. In a subsequent study we invited only artists to judge the light images and the approval rate doubled from 30% to 60%.

The second psychophysical experiment focused on images that can be enhanced for two conflicting keywords, meaning that e.g. the one implies a brightening and the other a darkening of the image. We compared our proposed algorithm against histogram equalization, Photoshop auto-contrast and the original. Our enhanced version outperformed all other versions by a factor of 2.5 or more.

The second application, color naming, benefits a lot from the semantic awareness provided by our framework. Traditionally, color naming is done with a psychophysical experiment where users have to type in the color names for different color patches. Our method allows to relate color names with color values fully automatically.

Chapter 6 presented automatic color naming using significance distributions in CIELAB color space for 50 color names from Moroney's web-based experiment [59]. Further we extended color naming to memory colors and arbitrary semantic expressions such as *chocolate* or *dolphin*.

We took color naming another step further in Chapter 7. We translated a list of over 900 English color names to nine other Asian and European languages. We automatically downloaded images from the world wide web for these over 9000 color names using Google Image Search. We then discussed the estimations from a language and a color science point of view. Language considerations were

important to analyze the results, because there are color names that do not exist in other languages such as the English color *maroon*. An advanced analysis of the complete significance distributions then showed that a higher z value correlates with a better precision of the estimation. We further demonstrated that colors names stretch more along the lightness axis than along the chromatic a and b axes in CIELAB color space.

8.2 Reflections and future research

The complexity of the statistical framework is reduced to a point where approximately 10'000 or less summations are sufficient to relate a keyword to a characteristic. It is not impossible, but at least extremely difficult to further decrease this complexity. A further improvement of the current implementation should thus focus on two technical issues. First, minimize read/write operations on the hard drive as much as possible. And second, migrate to a database implementation that allows for faster query and access of data.

An interesting research topic could be to extend the framework to multi-dimensional hypothesis tests as the current framework has only a one-dimensional view of the data for each significance test. This approach would very likely increase the complexity of the test significantly. In this case it would be worth to develop methods to decrease the complexity again.

The semantic image enhancement offers a large field of possibilities for future research. One possibility is to develop semantic re-rendering algorithms for more characteristics. One example is motion blur, which can be reinforced for images with keywords like *speed* or *jump*. This requires a directional non-isotropic blur. Another example is sharpness and contrast of corners and edges for images with keywords like *street sign*, *architecture* or *fence*.

It is also possible to design re-rendering workflows for devices with specific properties such as a small gamut or a low-contrast. Keywords that indicate structure can invoke a gamut compression rather than gamut clipping in order to preserve details at the expense of saturation. And vice versa, keywords that indicate flat surfaces can lead to a gamut clipping that preserves saturation rather than details.

Further, it is also imaginable to develop semantic re-rendering for observers with color vision deficiencies such as red-green color blindness. For example if the keywords indicate the presence of a *red ball* on *green grass*, the red ball can be automatically brightened in order to make it more distinguishable from the surrounding green grass.

The psychophysical evaluation of semantically enhanced images can also be extended with further experiments and analysis. For instance one can perform a deeper investigation of the results from the pairwise comparisons using a

Bradley-Terry-Luce model to gain more insight into the success and failure cases. An additional experiment could compare the automatically enhanced images against images that were manually corrected by Photoshop power users.

To be more robust in a real-world scenario the semantic image enhancement has to deal with multiple keywords per image. An easy approach for this scenario is to just compute the average of all keywords' significance distributions and apply this to the image. In the best case the keywords signify similar semantic concepts (e.g. *grass, green, nature*) and the average causes a similar output as any single keyword. If however the keywords describe conflicting semantic concepts they can cancel each other out resulting in no visual impact or undesired effects.

A more sophisticated approach could be to estimate each keywords importance for the input image. Meaningless or wrong keywords in the annotation can then have less influence on the processing. This can also be helpful for images that have machine-generated keywords that might not be as relevant as keywords from a real person.

Closely related to this is the computation of the weight maps. The current weight maps use only a single characteristic: color or defocus estimation. It is desirable to develop more robust region labeling techniques to determine the relevant regions based on a multitude of features. Labeling image regions is a large research field and it is possible that there are methods that can be adopted to this framework.

Another challenge are keywords with a double meaning such as *light*. *Light* can either mean that the image is just bright or that the image id darkened making a light source stand out more. We observed this conflict between two different groups of observers from Amazon Mechanical Turk and invited artists, respectively. To solve this conflict the image has to be re-rendered not only to match its semantic context, but also the user's taste.

A large market for automatic image enhancement based on semantic context are social and image sharing platforms such as Picasa, Flickr or Facebook. There are users of these services that just upload the images from their camera or smartphone to a folder without enhancing the visual quality. However, there are plenty of sources for related semantic information such as title, keywords, image name, postings from friends or other users, or even camera metadata such as GPS coordinates. All this can be used to determine a favorable re-rendering for everybody's images.

Semantic re-rendering can also be implemented in printers or photocopy machines. In this case it is possible to either re-render an image before feeding it into the print engine or to determine optimal parameters of the print engine such as rendering intent, black point compensation, print speed or ink droplet size. The semantic input can come from keywords in the image's file header or from

surrounding text in a composed document with mixed text and image content. Photocopy machines would need an additional optical character recognition step to extract the relevant information.

Yet another area for semantic image re-rendering are movies. In this case the semantic input can come from subtitles. The algorithm would then re-render each frame according to the semantic context derived from the subtitles.

Our enhancement framework is very light-weight because it uses pre-computed significance values and the current implementations (tone-mapping, color and depth-of-field) do not require heavy computations. The framework can thus be implemented even on handheld devices or embedded systems where battery life and computing power is a scarce resource.

The color thesaurus presented in Chapter 7 covers a considerable amount of color names and languages. This can fuel research in the field of language and culture of colors. It is for example possible to search for similar properties among Asian color names that are different among European color names.

The color value estimation can also be extended to entire paragraphs, texts or books, i.e. find a palette of 5 colors for Shakespeare's *Romeo and Juliet* . Color palettes are important for designers that need a few colors for a template and layout of a page. The scalable statistical framework could be used to learn significance distributions for all words of a language. This can then help to automatically adapt the color design of webpages or other documents with written text.

The statistical framework can not only be applied to image enhancement and color naming, but to many types of computer vision related problems. As the framework determines the relevance of characteristics, i.e. features or descriptors, it can add value to image retrieval, classification and other tasks that require an automatic image understanding.

Finally, the semantic input can be broadened to not only text, but also speech, gestures, brain activity, skin conductivity or other biological signals. This can help to gather more relevant information and ultimately result in computing systems that adapt to a users mood.

Appendix A

Characteristics

This chapter gives a complete overview of all implemented characterstics that have been used to create Figure 3.5, Table B.1 and the supplementary material http://ivrg.epfl.ch/SemanticEnhancement.html. We give for each characteristics its name, short id, dimensionality and description.

Graylevel histogram
short id: glH
dimensions: 16
description: The image is converted to graysacle by averaging the RGB values of each pixel. The graylevel values in the range [0 255] are then summarized in a histogram with 16 equidistant bins. The histogram is normalized so that its elements sum to 1.

Chroma histogram
short id: chH
dimensions: 16
description: The image is converted to CIELAB color space. Using the a and b channels the chroma radius is determined. The chroma value is always non-negative but the maximum depends varies for different hue angles and the color space before the conversion. The chroma values are summarized in a histogram with 16 equidistant bins in the range [0 50]. Chroma values greater than 50 are added to the last bin. The interval [0 50] is a compromise between a too large interval with most of the high chroma bins empty and a too small interval with too many chroma values being clipped to the last bin. The histogram is normalized so that its elements sum to 1.

Hue anlge histogram
short id: haH
dimensions: 16

description: The image is converted to CIELAB color space. Using the a and b channels the hue angle is determined in the interval $[0^\circ \ \ 360^\circ)$. The hue angle values are then summarized in a histogram with 16 equidistant bins. Pixels that have a chroma value of less than 1 are excluded from the histogram, because they can not be distinguished from the closest shade of neutral gray ($\Delta E < 1$). The histogram is normalized so that its elements sum to 1.

RGB histogram
short id: rgbH
dimensions: $8^3 = 512$
description: The image is opened in sRGB color space and uint8 encoding with values in $\{0, 1, 2, \ldots, 255\}$. The pixel values are summarized in a 3-dimen-sional histogram with 8 equidistant bins along each axis. The histogram is normalized so that its elements sum to 1.

CIELAB histogram
short id: labH
dimensions: $8^3 = 512$
description: The image is converted to CIELAB color space. The 3-dimensional histogram has 8 equidistant bins along each axis in the intervals $[0 \ \ 100]$ (lightness axis) and $[-80 \ \ 80]$ (a and b axes). Values outside the interval $[-80 \ \ 80]$ are added to the closest bin. The interval is a compromise between a too large interval with too many empty bins towards the interval borders and a too small interval with too many values being clipped to the closest bin. The histogram is normalized so that its elements sum to 1.

CIELCH histogram
short id: lchH
dimensions: $8^3 = 512$
description: The image is converted to CIELCH color space. The 3-dimensional histogram has 8 equidistant bins along each axis in the intervals $[0 \ \ 100]$ (lightness axis), $[0 \ \ 50]$ (chroma axis), and $[0 \ \ 360]$ (hue angle axis). Chroma values greater than 50 are added to the last bin. The interval $[0 \ \ 50]$ is a compromise between a too large interval with most of the high chroma bins empty and a too small interval with too many chroma values being clipped to the last bin. The histogram is normalized so that its elements sum to 1.

Lightness layout
short id: liL
dimensions: $8^2 = 64$
description: The image is converted to CIELAB color space. The L channel is subsampled using a coarse 8×8 grid and averaging all values within each cell. The grid is independent of the image's size and aspect ratio. This gives a coarse representation of how lightness is distributed in the image. The 8×8 feature

array is then scaled to the interval [0 1]. This scaling makes the feature robust against overall lightness changes.

Chroma layout
short id: chL
dimensions: $8^2 = 64$
description: The image is converted to CIELAB color space. Using the a and b channels the chroma radius is determined. The chroma channel is subsampled using a coarse 8×8 grid and averaging all values within each cell. The grid is independent of the image's size and aspect ratio. This gives a coarse representation of how chromaticity is distributed in the image. The 8×8 feature array is then scaled to the interval [0 1]. This scaling makes the feature robust against overall chromaticity changes.

Hue angle layout
short id: haL
dimensions: $8^2 = 64$
description: The image is converted to CIELAB color space. Using the a and b channels the hue angle is determined. The hue channel is subsampled using a coarse 8×8 grid and averaging all values within each cell[1]. The grid is independent of the image's size and aspect ratio. This gives a coarse representation of how hues are distributed in the image. The 8×8 feature array is then scaled to the interval [0 1]. This scaling makes the feature robust against overall hue changes.

Details histogram
short id: deH
dimensions: $3 * 16 = 48$
description: The image is converted to CIELAB color space and only the L channel is retained. The lightness channel is then blurred with a Gaussian blurring kernel with a variance equal to 10% of the image diagonal and subtracted from the non-blurred lightness channel. Then the absolute value of the difference is computed. The details histogram is composed of three separate histograms that each have 16 equidistant bins in the interval [0 40]. The values from the high-pass channel are binned into the three histograms according to the lightness value at the same pixel position. Pixel positions with a lightness values in the interval [0 33.3] belong to the first, the interval (33.3 66.6] to the second and the interval (66.6 100] to the third histogram, respectively. Each histogram is scaled to the interval [0 1]. the three histograms represent details in the shadow, mid-tones and highlight regions, respectively.

[1] The averaging of hue angles is done in the following way: every pixel is represented by a vector of unit length and respective hue angle. All these vectors are concatenated to a long vector whose hue angle is defined to be the average hue angle of all pixels.

Frequency histogram
short id: frH
dimensions: $21^2 = 441$
description: The image is converted to CIELAB color space and transformed to Fourier domain. The absolute value of the Fourier domain representation is then resized to 21×21 pixels. The size is uneven in order to have a single center value that collects the DC component and the very low frequencies.

Gabor filter histogram
short id: gabH
dimensions: $4 * 2 * 16 = 128$
description: The image is converted to CIELAB color space. The L channel is filtered with different Gabor filters that change in angle and size. We use four angles (0°, 45°, 90°and 135°) and two sizes (21×21 and 41×41 pixels). The frequency is chosen to have a main psitive bump in the middle and two negative bumps next to it. An example is given in Figure A.1.

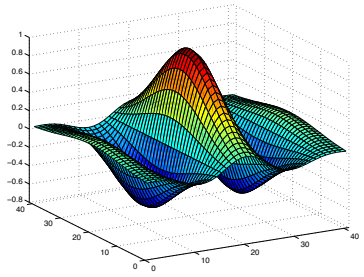


Figure A.1: Example Gabor filter with size 41×41 and angle 0°.

In total eigth histograms of the filters' responses are computed. Each histogram has 16 equidistant bins in the range $[-16\ \ 16]$. Values outside the range are added to the closest bin at the border.

Gabor filter layout
short id: gabL
dimensions: $4 * 2 * 8^2 = 512$
description: The image is converted to CIELAB color space. The L channel is filtered with different Gabor filters that change in angle and size. We use four angles (0°, 45°, 90°and 135°) and two sizes (21×21 and 41×41 pixels).

Each filter output is subsampled using a coarse 8×8 grid and averaging all values within each cell. The 8×8 feature array is then scaled to the interval $[0\ \ 1]$. This scaling makes the feature robust against overall lighntess changes.

Linear binary pattern histogram
short id: lbpH
dimensions: 18
description: The image is converted to CIELAB color space. For each pixel its Linear Binary Pattern (LPB) is computed using the L channel [69, 56]. LBPs describe a corner type within a 5×5 neighborhood around a pixel position. It varies from acute to obtuse angles. A LBP is a concatenation of binary values indicating whether the center value h_0 is greater than the circular neighbors $h_1 \dots h_{16}$ as shown in Figure A.2. In total there are 16 corner types plus two special cases. First, the center value is larger/smaller than all 16 neighbors. Second, there is no corner because the neighbor values are sometimes larger and sometimes smaller. See Ojala et al. for a deeper discussion [69, Sec. 2.4]. The frequencies of the different patterns are thus counted in an 18 dimensional histogram. The histogram is normalized so that its elements sum to 1.

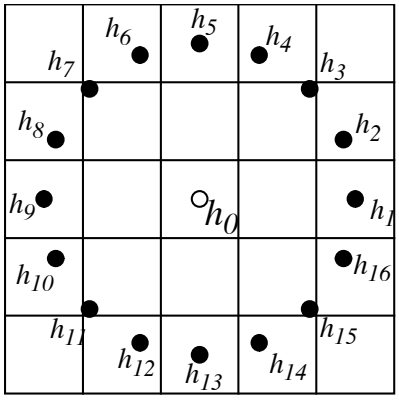


Figure A.2: The circularly symmetric neighbor set of 16 pixels in a 5×5 neighborhood used for linear binary patterns. Figure reproduced from Ojala et al. [69].

Appendix B

Overview of Δz^*_w values

This Section contains three tables for the 200 most frequently used keywords, the 200 most significant keywords and the 14 characteristics ranked by their significance, respectively. The full tabel for all 2858 keywords can be found here:
http://ivrg.epfl.ch/SemanticEnhancement.html

B.1 The 200 most frequently used keywords

Table B.1 shows for 14 characteristics and the 200 most frequently used keywords in the MIR Flickr database [36] the Δz^* values computed with Equation 3.2. The keywords are sorted by decreasing database frequency, shich is given in brackets behind each keyword. Appendix A lists details for the computation of the different characteristics.

Table B.1: Δz^* values for the 200 most frequent keywords.

keyword (database frequency)	graylevel hist	chroma hist	hue angle hist	RGB hist	CIE-Lab hist	CIE-Lch hist	lightness layout	chroma layout	hue angle layout	details hist	frequency hist	gabor filter hist	gabor filter layout	linear binary pattern hist
nikon (46007)	5	3	3	5	7	7	3	1	2	6	4	4	5	9
canon (43456)	6	2	3	6	7	7	3	1	1	6	2	4	3	8
nature (40291)	11	17	15	18	20	19	2	2	1	10	10	15	7	13
2008 (40115)	1	2	3	4	5	5	2	1	1	4	2	5	3	3
sky (33343)	9	9	21	17	20	18	15	9	10	21	16	29	19	8
blue (31934)	9	17	28	28	28	29	4	7	5	10	7	15	7	4
macro (30529)	11	27	17	25	26	28	6	6	4	32	19	30	14	37
bw (30294)	18	65	8	58	54	58	4	6	2	17	5	18	4	19
flower (29567)	11	33	22	33	34	34	8	10	4	32	24	28	12	40
water (28688)	6	6	13	13	12	12	8	5	2	5	20	20	10	8
red (26418)	12	27	24	35	33	32	3	8	5	13	5	7	4	7

Continued on next page

Table B.1 – *Continued from previous page*

keyword (database frequency)	glH	chH	haH	rgbH	labH	lchH	liL	chL	haL	deH	frH	gabH	gabL	lbpH
portrait (26243)	11	14	13	15	15	16	6	3	6	16	23	19	11	18
green (24887)	11	28	27	33	32	33	2	5	4	8	6	9	4	10
art (23170)	4	3	4	7	7	6	2	2	2	10	8	9	6	14
hdr (22394)	19	13	10	20	22	18	5	7	8	28	31	31	18	8
california (22278)	3	3	8	8	8	7	3	3	3	8	7	11	8	6
light (22176)	20	11	5	21	18	16	10	3	2	15	3	12	4	7
night (22168)	39	16	13	36	32	26	18	4	5	31	18	16	14	18
sunset (21676)	22	15	16	34	30	30	26	9	9	25	23	41	21	12
white (21433)	14	23	7	23	21	21	5	4	2	9	9	9	3	5
film (20498)	16	15	8	19	19	18	10	3	2	7	11	9	7	7
clouds (20407)	11	10	19	17	20	19	19	8	10	21	21	34	22	17
usa (20323)	3	4	6	8	6	7	3	3	3	9	7	9	7	6
abigfave (19546)	5	8	5	9	9	9	4	2	2	5	6	7	9	5
winter (19487)	11	15	13	16	16	17	3	5	2	14	14	9	9	13
geotagged (19226)	2	3	8	7	7	8	4	3	4	11	13	12	8	11
street (19051)	7	14	5	14	15	14	5	3	2	20	11	26	11	15
beach (18552)	18	12	14	16	13	13	9	6	9	16	22	32	17	10
people (18441)	9	15	5	13	12	14	4	2	3	9	11	22	7	7
landscape (18187)	7	7	17	16	18	17	16	10	9	14	29	26	22	15
architecture (18187)	6	8	10	11	11	11	5	6	5	19	19	19	11	24
city (17971)	6	4	7	8	9	8	4	3	3	19	15	19	10	12
flowers (17781)	10	30	21	31	31	30	6	8	3	19	21	18	9	31
yellow (17556)	14	35	20	37	37	36	7	10	1	12	8	12	5	13
blackandwhite (17092)	18	65	7	58	53	58	3	5	1	18	5	19	4	19
snow (16438)	19	24	19	25	25	27	8	9	3	19	16	11	12	19
tree (16171)	2	6	7	11	12	12	4	3	3	20	19	21	7	18
anawesomeshot (15575)	6	9	5	10	10	10	4	2	2	6	9	9	9	5
girl (15558)	10	7	11	14	13	14	8	3	6	11	19	13	10	15
black (15489)	18	28	5	26	26	27	3	4	2	9	6	14	4	13
color (15837)	5	19	5	16	16	15	4	2	1	4	3	2	4	3
cat (15292)	6	11	17	17	17	18	12	7	4	20	18	19	11	15
sea (15049)	14	13	18	16	15	15	14	6	6	14	28	37	20	13
explore (14983)	5	4	3	5	6	6	5	1	2	4	3	4	4	3
travel (14754)	2	5	7	7	7	7	5	4	4	9	12	12	9	10
urban (14751)	6	5	6	9	9	9	2	2	2	19	14	19	6	15
aplusphoto (14677)	6	7	5	9	10	10	5	3	2	8	8	9	11	7
france (14670)	2	3	6	6	6	6	3	2	2	9	8	7	6	9
bokeh (14646)	8	19	10	15	17	19	7	3	3	41	23	32	10	51
goldstaraward (14456)	7	12	5	12	12	12	5	3	2	5	6	5	10	7
london (14417)	5	7	6	9	9	10	3	2	1	12	9	15	6	10
spain (14293)	4	4	6	7	8	8	4	3	3	10	10	11	12	11
japan (14242)	6	2	3	4	4	4	2	1	1	9	5	13	5	5
trees (13919)	6	5	10	13	16	15	10	6	4	26	25	25	13	28
reflection (13590)	6	6	8	10	10	9	6	3	3	7	13	9	8	5
italy (13573)	2	7	5	8	9	8	5	2	2	11	11	10	7	12
2009 (13165)	2	2	2	2	3	3	1	1	1	3	3	4	3	5
sanfrancisco (13004)	4	4	9	9	10	9	3	3	1	11	7	14	6	9
europe (12855)	2	1	6	6	7	7	5	2	3	11	10	11	8	9
bird (12654)	16	8	13	21	16	18	8	6	3	19	11	23	11	12
blueribbonwinner (12612)	6	9	6	10	11	10	4	2	2	6	6	6	10	6
pink (12505)	15	23	29	38	41	39	4	5	4	21	14	21	7	23
spring (12296)	12	17	14	21	22	19	3	4	1	8	8	13	5	17
sun (12089)	8	9	8	17	15	15	16	6	5	15	14	24	13	8
eos (12001)	6	3	5	7	7	8	2	1	1	5	2	4	3	11
woman (11952)	10	11	11	12	13	13	8	3	5	8	17	14	10	10
selfportrait (11928)	13	7	15	15	14	13	4	4	5	21	23	25	12	11
animal (11790)	15	6	10	19	19	19	10	6	2	16	11	18	9	15
germany (11723)	4	3	4	6	6	7	3	3	2	12	11	11	6	9
espana (12246)	4	4	6	7	8	8	5	3	4	12	10	14	14	11
orange (11709)	18	33	20	38	38	36	6	10	3	16	4	16	6	11
diamondclassphotographer (11608)	4	8	4	9	10	9	5	3	2	6	6	6	9	5
me (11691)	8	5	12	10	11	10	4	3	4	17	20	20	10	10
graffiti (11378)	7	9	14	15	14	15	4	5	6	27	21	30	13	28
nyc (11368)	6	7	7	10	10	11	3	2	2	15	9	21	6	11
canada (11230)	2	2	8	7	7	7	3	2	1	8	9	6	6	8
platinumphoto (10965)	6	9	4	10	10	10	6	2	2	5	7	7	10	5
dog (10938)	8	11	10	15	14	15	11	6	2	20	17	17	9	15
uk (10942)	3	3	6	7	7	7	5	5	2	9	10	10	7	8
england (10723)	5	4	8	8	8	8	6	6	2	13	11	12	7	11
park (10715)	6	7	10	12	12	11	4	5	2	12	12	16	8	10
photo (10808)	1	2	3	3	3	4	2	2	1	3	2	2	4	2
building (10483)	5	7	12	11	12	12	7	6	5	20	19	20	13	21
365days (10465)	12	5	15	14	13	13	4	4	4	21	21	24	12	12

Continued on next page

Table B.1 – *Continued from previous page*

keyword (database frequency)	glH	chH	haH	rgbH	labH	lchH	lL	chL	haL	deH	frH	gabH	gabL	lbpH
car (10401)	8	10	12	14	12	14	11	5	3	16	18	31	16	12
soe (10022)	7	11	5	11	12	11	5	1	2	4	8	8	9	4
italia (10029)	3	6	5	8	8	8	6	3	3	10	10	10	9	11
old (9974)	2	8	10	11	11	10	4	4	2	21	11	21	8	17
olympus (9803)	3	4	6	8	8	8	2	3	1	4	5	6	5	4
naturesfinest (9728)	13	21	15	21	23	22	3	2	2	13	11	18	8	17
photography (9741)	4	3	4	6	6	7	2	1	1	3	2	2	3	3
newyork (9732)	6	9	5	11	11	11	3	2	2	15	9	20	5	12
texture (9715)	15	11	13	14	13	14	11	4	4	19	20	18	12	27
food (9707)	12	21	19	32	30	29	8	13	6	18	24	17	14	31
d80 (9700)	4	2	4	5	5	6	2	1	2	4	4	4	4	10
paris (9569)	9	11	4	10	10	10	2	2	1	14	6	17	5	9
australia (9566)	5	8	4	7	8	7	2	2	3	3	6	8	5	6
river (9564)	4	7	12	12	11	11	10	5	4	15	25	18	15	19
supershot (9407)	4	8	5	10	10	9	4	3	2	6	5	6	9	6
theunforgettablepictures (9375)	5	11	4	11	12	12	6	3	2	6	6	6	11	7
theperfectphotographer (9307)	4	8	5	11	11	10	6	3	2	8	8	6	11	6
photoshop (9298)	7	4	3	7	9	10	6	3	1	8	5	6	4	6
garden (9269)	10	25	23	27	27	27	4	3	3	9	15	19	7	21
vintage (9014)	14	7	11	15	14	14	9	3	2	13	8	8	15	18
cute (8970)	13	5	13	18	15	18	12	3	6	19	20	19	11	19
impressedbeauty (8963)	7	9	8	11	13	12	3	4	2	5	5	8	12	10
d300 (8771)	5	3	4	8	10	10	4	2	2	8	5	6	6	14
colors (8593)	8	26	7	24	24	22	3	4	1	7	6	8	3	5
christmas (8564)	13	18	17	25	24	23	7	5	2	13	11	16	8	8
summer (8543)	6	10	8	11	12	12	3	4	3	5	6	8	6	5
sigma (8542)	6	5	6	9	10	10	4	3	4	7	9	5	7	9
365 (8512)	11	4	12	12	11	12	4	4	3	18	16	18	9	13
autumn (8353)	15	24	11	24	25	22	6	3	2	19	18	24	9	15
2007 (8334)	3	3	3	5	6	6	1	1	1	5	4	4	3	5
tokyo (8294)	12	5	7	11	11	9	4	1	2	15	7	17	7	9
d40 (8245)	9	5	4	9	12	12	6	1	2	10	7	6	5	8
50mm (8256)	10	4	6	12	9	11	5	2	2	21	17	17	7	29
lights (8232)	34	21	14	33	32	28	15	6	5	29	14	14	12	12
bridge (8204)	7	7	8	9	9	9	8	5	6	16	20	16	16	15
flickr (8750)	5	5	2	4	6	6	3	1	2	4	2	5	2	2
dof (8169)	8	15	10	12	14	17	6	2	3	35	19	26	10	48
church (8165)	6	6	7	11	12	12	7	4	6	20	19	23	15	18
portugal (8053)	8	7	10	11	11	11	4	2	4	15	12	11	9	15
flickrdiamond (8003)	5	8	3	9	10	9	5	3	3	5	7	7	10	6
man (7961)	7	18	4	14	15	15	3	3	3	5	10	13	6	6
digital (7915)	4	5	2	5	5	5	2	1	1	4	4	3	2	4
creativecommons (7725)	7	5	3	6	6	6	5	3	2	7	11	8	5	10
square (7665)	7	8	3	11	12	12	5	2	1	5	7	8	4	5
sony (7603)	3	7	7	9	10	9	4	2	1	6	5	4	5	12
beautiful (7569)	4	11	6	10	9	10	4	2	4	5	5	7	5	13
pentax (7553)	7	7	5	8	8	10	3	5	2	11	4	7	8	12
lake (7543)	6	10	19	17	17	17	13	6	4	10	27	26	17	12
unitedstates (7522)	11	10	9	14	11	14	5	3	2	11	5	12	8	8
wall (7493)	10	4	3	9	7	7	4	4	3	18	16	9	8	30
brasil (7437)	2	7	6	9	9	9	2	1	3	4	3	3	5	2
window (7436)	8	6	5	8	9	9	3	2	4	12	16	17	7	15
ocean (7435)	14	12	18	16	16	15	15	6	6	16	28	38	21	11
love (7336)	5	4	9	8	7	8	4	3	2	11	11	11	5	11
house (7302)	4	3	4	8	8	8	7	6	4	16	15	17	10	17
streetart (7195)	12	15	13	19	17	19	3	3	7	21	17	26	11	31
music (7161)	25	9	13	24	20	20	7	6	3	19	15	19	9	14
birds (7148)	15	7	14	19	14	16	7	7	3	18	10	22	9	8
colour (7144)	4	18	6	16	16	16	2	3	1	4	2	3	3	3
closeup (7131)	10	19	14	17	17	20	5	3	3	21	16	25	10	29
taiwan (7114)	9	9	7	13	13	12	5	3	3	6	11	9	6	17
abstract (7042)	5	14	7	15	15	15	6	3	5	10	10	11	10	13
deutschland (6656)	5	5	4	7	7	7	4	3	3	14	13	14	7	12
abandoned (6595)	11	12	5	16	16	15	4	4	2	29	21	27	10	25
self (6539)	13	7	13	15	14	14	4	3	5	20	21	24	11	11
40d (6517)	7	3	6	6	7	9	3	2	2	7	3	8	5	11
dark (6463)	36	7	9	31	28	26	14	5	3	24	8	14	9	16
fall (6437)	16	24	9	23	24	22	6	2	3	20	18	24	9	14
shadow (6376)	12	7	7	16	17	16	5	3	2	9	6	12	7	15
fun (6364)	3	6	6	10	10	9	5	2	2	2	6	7	5	4
newyorkcity (6338)	6	9	7	12	12	12	3	2	2	16	10	22	5	13
fuji (6311)	8	3	9	12	9	11	9	2	1	6	6	7	5	5
brazil (6306)	4	5	5	9	10	9	2	1	3	6	3	6	5	2

Continued on next page

Table B.1 – *Continued from previous page*

keyword (database frequency)	glH	chH	haH	rgbH	labH	lchH	liL	chL	haL	deH	frH	gabH	gabL	lbpH
museum (6289)	4	3	8	8	8	8	3	2	2	6	5	6	7	11
sign (6230)	7	8	11	12	11	12	2	3	2	11	14	19	9	22
texas (6167)	4	8	3	7	7	6	4	2	2	5	4	2	5	5
blackwhite (6157)	20	66	6	58	54	59	4	7	2	19	4	20	5	20
home (6152)	2	5	9	9	8	9	3	3	1	9	7	8	5	7
india (6128)	5	6	6	8	8	9	3	4	2	3	3	2	5	4
project365 (6086)	6	6	5	8	7	7	5	3	3	14	12	12	9	16
glass (6072)	8	3	10	14	13	13	4	2	4	10	6	11	5	7
sunrise (6030)	18	16	15	30	27	28	26	8	7	24	25	40	20	11
wildlife (6022)	21	14	16	26	25	24	14	8	3	16	15	18	12	12
grass (6012)	12	21	25	27	27	24	8	16	3	16	19	20	14	9
plant (5985)	12	28	21	28	30	29	6	5	2	24	19	26	11	32
400d (5947)	11	6	6	13	15	16	4	2	2	10	2	8	5	13
asia (5907)	5	2	6	7	7	7	4	2	2	10	4	14	6	5
cielo (5874)	10	9	22	19	21	20	16	7	10	24	15	32	22	8
sculpture (5833)	5	10	5	12	11	13	5	3	1	8	6	4	8	14
longexposure (5816)	32	17	15	29	28	25	16	2	8	27	21	17	15	10
nikkor (5799)	7	3	3	9	9	10	5	1	1	8	4	7	4	14
barcelona (5758)	8	5	5	8	6	9	4	4	1	11	10	14	11	7
beauty (5740)	7	6	9	13	12	12	5	2	5	8	14	11	7	16
life (6111)	8	3	5	9	9	10	3	3	2	7	7	6	4	8
train (5705)	9	10	6	12	12	12	12	6	5	18	16	26	12	14
animals (5696)	14	8	13	18	18	18	11	7	2	14	12	17	10	14
leaves (5650)	16	28	12	25	25	24	6	4	2	18	17	27	12	14
35mm (5653)	15	12	10	18	17	17	11	5	1	8	7	11	6	12
purple (5620)	10	29	35	36	40	35	4	3	4	21	14	16	8	24
face (5522)	9	10	13	14	14	13	8	4	6	16	20	22	13	15
vacation (5506)	2	8	8	8	9	10	4	3	3	7	10	9	9	8
ice (5485)	14	18	19	22	19	22	3	7	2	11	15	14	8	12
florida (5414)	2	4	6	6	6	5	3	3	2	7	6	9	5	3
road (5428)	5	3	7	8	9	8	8	5	7	10	19	14	14	14
boat (5354)	7	8	14	13	12	11	12	4	3	10	28	22	14	18
decay (5349)	9	9	5	12	13	14	3	2	2	32	22	29	8	23
rain (5349)	8	3	10	12	10	11	7	5	2	9	6	9	7	6
d200 (5340)	7	9	8	10	11	12	2	5	4	5	8	4	9	11
lomo (5322)	15	9	16	19	19	19	23	6	2	14	14	12	12	12
sand (5305)	22	13	15	18	15	14	7	6	11	16	26	36	16	15
mountain (5295)	7	10	23	21	22	22	14	10	8	14	27	27	23	18
manhattan (5258)	7	5	8	9	11	12	3	2	2	17	10	23	5	12
golddragon (5195)	7	10	4	12	13	12	5	4	3	8	7	9	13	8
lightroom (5202)	13	5	6	14	18	18	11	2	2	11	3	9	7	16
new (5201)	3	5	4	6	7	7	2	2	1	10	5	12	4	6
party (5157)	14	5	14	16	17	15	8	5	5	15	16	18	9	12

B.2 The 200 most significant keywords

Table B.2 shows the 200 keywords with the highest average significance values for the given 14 characteristics.

Table B.2: Δz^* values for the 200 most significant keywords.

keyword (database frequency)	graylevel hist	chroma hist	hue angle hist	RGB hist	CIE-Lab hist	CIE-Lch hist	lightness layout	chroma layout	hue angle layout	details hist	frequency hist	gabor filter hist	gabor filter layout	linear binary pattern hist
poladroid (622)	53	28	40	58	61	64	49	47	12	34	36	33	63	28
bainnewsservice (567)	41	73	1	67	63	71	18	10	1	47	31	48	52	43
goldenbeauty (530)	57	39	14	46	42	43	19	24	9	56	23	61	61	66
postcardcollection (790)	55	28	33	61	58	60	31	21	6	44	36	37	35	54
cleverampcreativecaptures (548)	56	38	13	45	40	42	19	25	9	55	23	60	60	64

Continued on next page

Table B.2 – *Continued from previous page*

keyword (database frequency)	glH	chH	haH	rgbH	labH	lchH	lhL	chL	haL	deH	frH	gabH	gabL	lbpH
bravissimo (574)	56	39	13	45	41	42	20	24	9	55	22	60	59	63
vintagevalentinesday (621)	51	25	36	59	59	59	31	24	6	44	33	36	34	50
vintagevalentinesdaypostcard (620)	51	25	36	59	59	59	31	24	6	44	33	36	34	50
vintageholiday (1411)	52	28	34	59	58	59	30	22	6	42	33	35	33	51
mybestphoto (573)	55	38	12	44	40	42	20	24	9	54	22	59	58	62
vintagepostcard (1571)	52	29	32	59	56	58	30	20	6	42	34	35	32	53
anythingdigital (601)	53	38	11	43	39	41	19	24	9	53	21	58	58	61
1910s (811)	40	65	3	63	61	64	20	8	1	44	28	41	47	42
remolar (536)	50	20	32	54	48	43	36	13	11	46	33	40	54	46
deltallobregat (561)	50	20	30	54	47	43	36	13	10	46	33	39	53	45
anita (629)	53	36	12	43	38	39	18	22	7	51	21	55	54	58
exclusive (648)	50	36	10	40	36	39	17	23	8	50	20	57	53	55
mudpeople (521)	46	39	25	47	50	56	20	16	13	30	25	38	31	44
mudfest (511)	46	39	25	47	50	56	21	15	12	30	25	38	31	44
boryeong (578)	46	39	25	47	50	56	21	15	12	29	24	38	31	43
daecheon (558)	46	39	25	47	49	56	21	16	12	29	24	38	31	43
sx70 (1670)	38	19	7	38	44	35	38	36	11	45	49	39	52	21
damncool (691)	47	32	10	38	34	36	17	21	7	45	19	50	50	53
600 (616)	40	21	11	37	46	38	39	37	14	38	39	34	44	20
ephemera (1180)	40	20	30	47	47	48	26	14	5	43	28	27	28	48
mudwrestling (613)	46	37	20	45	48	53	20	14	12	31	22	38	29	36
pistil (552)	36	33	21	42	44	46	11	4	8	53	24	57	24	44
hamster (593)	36	24	30	42	43	47	23	8	12	44	19	43	23	40
powerhousemuseum (563)	27	65	2	60	56	63	22	7	1	34	12	27	26	29
desdibuix (1085)	41	22	21	40	33	44	30	24	5	25	29	28	42	45
bohigas (1082)	41	22	21	40	33	44	30	24	5	25	29	28	42	45
miquelbohigascostabella (1080)	41	22	20	40	33	44	30	24	5	25	29	28	42	45
hibiscus (548)	15	44	30	45	46	48	15	17	16	42	24	29	15	38
instant (561)	36	19	13	37	45	38	35	28	9	35	40	30	42	15
dariosanches (536)	34	32	34	48	49	48	15	9	5	35	25	39	21	27
dariosan (545)	34	32	34	48	48	47	15	9	5	35	25	39	21	27
fireworks (1283)	44	21	26	42	42	38	27	14	4	49	33	20	25	32
greatflowermacros (521)	17	35	27	34	37	37	12	29	5	41	29	29	41	43
insectes (912)	24	40	41	44	44	44	13	9	7	30	25	31	25	36
automobiles (618)	39	32	7	33	29	33	15	18	9	41	18	52	46	41
canoneos400ddigital (983)	39	20	21	39	32	43	29	22	5	25	26	27	39	42
libraryofcongress (4034)	23	42	6	38	41	48	11	10	4	33	33	30	46	43
difocus (529)	35	7	28	44	37	40	22	11	17	31	36	39	25	36
insecte (845)	23	40	41	42	44	40	14	11	7	31	21	34	21	38
polaroid (4856)	34	19	9	36	41	32	34	28	7	37	41	32	39	17
motorsport (584)	36	31	9	31	27	31	15	18	9	34	20	51	45	46
stamen (1007)	25	38	16	38	40	42	9	7	7	44	25	46	18	43
excellentsflowers (950)	15	43	29	44	43	41	15	16	8	34	31	22	15	40
pollen (732)	25	41	18	41	42	44	8	6	5	41	23	43	18	41
muddy (606)	38	30	19	39	41	46	18	14	11	26	20	34	26	34
throughtheviewfinder (589)	24	11	19	29	36	37	32	12	3	42	45	31	42	33
insecta (862)	33	33	34	41	36	37	21	8	8	27	21	35	26	35
polaroid600 (665)	26	15	13	30	34	32	29	25	18	40	46	31	36	20
taxonomy:kingdom=animalia (533)	34	30	36	43	38	42	27	6	10	22	25	32	26	22
insectos (1021)	25	34	40	40	37	41	18	5	8	27	22	31	25	38
collections (733)	41	30	10	32	30	29	14	18	7	38	15	39	43	44
lightpainting (847)	52	25	21	48	45	41	17	10	8	43	19	22	16	21
flickrfloresclosеupmacros (515)	16	42	27	40	40	41	16	15	8	35	28	25	16	37
highspeed (607)	39	30	8	31	28	29	16	19	7	36	15	40	43	44
lightstream (515)	30	31	28	39	40	38	26	9	17	30	24	20	31	20
arthropoda (902)	32	30	33	39	34	36	23	6	8	26	21	34	26	34
fofurasfelinas (528)	23	8	28	31	32	34	21	9	8	47	42	39	26	34
taiwanese (973)	31	3	28	38	32	38	21	9	17	31	35	37	25	34
ahqmacro (610)	20	45	32	43	44	42	10	9	6	27	23	35	25	16
tulip (1599)	14	38	27	37	37	37	5	16	6	40	26	38	15	40
flowerotica (1242)	11	39	23	36	37	39	11	20	6	38	25	27	24	40
mimamorflowers (953)	14	43	27	42	42	41	14	15	7	30	30	20	14	37
sp90 (653)	12	30	27	31	33	34	14	23	5	39	25	26	36	40
kaleidoscope (508)	13	37	16	32	33	30	20	18	13	40	28	52	24	18
auniverseofflowers (553)	16	41	29	40	41	39	13	11	7	35	24	27	13	37
fantasticflower (1278)	9	38	27	38	38	37	13	20	7	36	27	20	21	41
giuss95 (696)	23	31	29	37	34	33	12	12	6	23	35	37	26	34
flowerscolors (1127)	11	39	26	37	38	38	12	20	6	35	26	20	22	41
flowerwatcher (818)	8	38	27	40	40	39	11	20	7	34	28	19	19	41
tamronsp90mmf28dimacro (686)	13	30	26	30	33	33	12	24	4	39	25	25	37	40
showmeyourqualitypixels (600)	38	23	11	28	26	27	14	19	7	36	15	40	43	44
trix (732)	20	66	5	59	55	60	6	7	2	22	7	25	9	27
t4lagree (511)	33	13	18	28	27	25	20	16	7	31	34	48	25	44

Continued on next page

Table B.2 – *Continued from previous page*

keyword (database frequency)	glH	chH	haH	rgbH	labH	lchH	liL	chL	haL	deH	frH	gabH	gabL	lbpH
macroflowerlovers (1249)	9	37	24	35	36	37	11	19	5	38	26	26	22	43
sunflower (553)	25	44	35	47	47	48	10	16	6	12	19	14	18	26
astronomy (518)	36	18	17	35	28	27	30	8	8	33	15	53	19	40
boeing (768)	19	25	34	33	28	33	21	12	8	25	31	42	32	23
macros (769)	22	31	17	28	30	30	19	22	8	38	25	33	24	36
bento (720)	14	33	15	39	38	34	17	17	11	24	28	34	23	36
bigcat (580)	25	24	28	35	38	36	11	15	6	35	24	40	19	27
insects (2490)	23	36	33	37	37	37	16	6	7	25	21	31	24	29
blancinegre (793)	20	67	8	59	54	59	5	8	2	21	7	18	7	26
tamron90 (729)	12	29	25	29	32	32	13	21	4	39	23	27	35	40
thorgalsen (719)	18	18	26	24	33	34	20	17	7	31	23	54	36	20
vodcars (719)	18	18	26	24	33	34	20	17	7	31	23	54	36	20
vod (719)	18	18	26	24	33	34	20	17	7	31	23	54	36	20
supercars (548)	16	20	27	25	32	33	23	17	7	29	22	53	35	20
masterphotos (1151)	12	34	26	34	36	35	12	19	5	36	25	26	20	37
upclose (554)	24	34	36	39	38	40	12	11	4	20	25	26	22	26
noiretblanc (2209)	18	65	6	58	54	59	6	6	2	21	6	22	9	24
bud (798)	15	30	24	32	36	34	10	7	5	39	21	43	19	38
southkorea (720)	34	29	20	36	35	41	15	10	9	22	20	28	23	31
awesomeblossoms (1934)	10	37	26	37	38	37	11	17	5	33	25	22	17	37
petals (1772)	17	36	17	30	33	35	9	13	6	38	26	34	14	44
petal (1231)	20	32	17	33	35	36	8	8	5	38	23	43	16	38
strawberry (697)	18	26	29	42	42	41	12	18	12	23	23	20	15	30
exotics (593)	16	19	26	23	32	32	21	15	7	28	23	53	34	21
passaro (773)	29	22	29	40	40	40	12	8	5	31	19	33	18	23
lepidoptera (761)	20	32	33	36	35	37	10	7	7	22	26	23	21	39
christmaslights (501)	36	32	22	42	42	39	19	11	3	35	18	21	14	14
meiji (765)	16	29	34	38	40	41	11	9	4	35	11	34	20	25
recipe (905)	19	24	27	38	37	35	13	21	8	25	26	23	16	34
wildflower (660)	18	34	32	36	40	37	11	7	6	29	20	26	15	35
surfer (546)	31	21	34	36	29	34	11	10	3	19	27	48	19	24
blancoynegro (2404)	18	62	10	57	52	57	3	5	2	21	5	22	10	21
macromarvels (2243)	14	36	29	36	37	37	9	11	5	34	18	28	20	31
vegetarian (520)	14	29	21	39	38	36	12	18	9	22	29	21	19	38
blackwhite (6157)	20	66	6	58	54	59	4	7	2	19	4	20	5	20
biancoenero (1552)	17	66	8	58	54	60	2	5	2	20	5	22	7	18
period (641)	16	28	34	37	40	41	9	9	4	37	10	34	19	26
a16 (593)	30	3	25	34	28	35	21	8	14	33	29	34	24	26
tomato (585)	15	32	24	43	41	40	10	21	10	22	23	18	14	31
1940s (576)	26	27	11	26	28	31	17	12	9	31	28	22	40	36
largeformat (553)	25	22	9	29	34	31	17	10	10	35	29	22	39	32
tulips (922)	15	38	27	39	39	36	5	15	7	26	23	26	14	33
candle (750)	34	28	27	39	39	37	17	14	3	33	16	28	14	14
nightphotography (549)	40	23	13	37	37	31	22	7	8	40	25	21	19	20
ilford (1366)	18	66	5	60	55	58	11	7	2	9	11	11	10	19
byn (570)	16	63	9	57	52	57	5	11	2	16	5	18	12	19
bee (1554)	15	36	25	34	36	38	9	3	5	29	25	28	18	40
nb (1375)	17	62	6	55	51	56	6	7	2	20	6	23	9	21
blume (1100)	13	35	25	35	36	36	8	12	5	33	25	27	13	38
illumination (522)	39	26	13	41	40	35	19	12	5	38	19	25	14	15
wolfgangstaudt (912)	22	26	20	32	28	31	10	11	17	36	31	31	25	19
supercar (720)	27	21	10	22	21	25	16	18	9	34	18	54	37	27
succulent (620)	18	37	29	35	36	37	7	8	5	21	25	23	16	42
bwartaward (549)	18	63	7	56	52	56	6	6	2	19	6	19	8	21
postcard (3103)	33	20	18	39	33	39	18	8	3	31	20	22	17	37
korean (873)	31	32	14	35	37	38	15	10	8	20	21	27	19	30
66111 (625)	21	25	21	31	27	31	10	11	16	39	30	31	25	19
bw (30294)	18	65	8	58	54	58	4	6	2	17	5	18	4	19
valentinesday (1173)	28	13	29	40	42	41	19	16	5	20	16	23	21	23
dragonfly (829)	21	27	29	32	34	32	15	6	7	30	13	40	24	26
insecto (673)	16	30	29	35	36	35	15	7	6	31	17	28	22	29
blute (518)	12	31	24	31	34	35	9	12	6	37	26	28	11	40
orchid (1076)	9	27	24	31	34	36	9	12	4	33	30	32	15	39
stainedglass (509)	26	18	9	29	32	26	14	11	6	50	28	47	15	24
butterflies (872)	20	35	29	34	37	35	10	12	6	24	23	24	21	24
nocturna (707)	41	18	14	37	34	28	20	7	9	34	26	21	23	22
blackandwhite (17092)	18	65	7	58	53	58	3	5	1	18	5	19	4	19
tyskland (564)	11	22	14	24	24	26	22	8	7	44	35	42	21	33
lightning (517)	28	10	27	30	27	27	23	5	14	41	12	44	27	18
fiore (1086)	15	33	26	33	35	34	11	9	4	31	25	23	13	40
sweets (774)	20	16	21	33	32	31	14	15	8	31	28	27	17	39
predator (594)	23	30	30	33	34	32	12	15	4	24	24	27	21	22
felt (792)	27	16	16	34	35	35	17	14	9	33	23	33	21	17

Continued on next page

Table B.2 – *Continued from previous page*

keyword (database frequency)	glH	chH	haH	rgbH	labH	lchH	lhL	chL	haL	deH	frH	gabH	gabL	lbpH
cavalli (516)	16	25	23	35	34	32	26	15	12	17	21	28	24	22
makro (699)	14	30	20	28	30	33	11	10	5	37	19	36	16	40
dessert (1299)	15	17	22	33	32	32	11	15	8	31	32	27	16	37
goldenphotographer (849)	32	20	10	24	22	23	11	17	6	34	15	38	39	37
fuzzy (582)	25	20	19	28	33	34	15	7	8	34	20	37	16	32
americameridionale (502)	13	30	24	39	38	37	22	15	12	11	20	20	25	21
deutsche (501)	11	22	14	22	23	24	22	6	6	45	34	44	19	34
flower (29567)	11	33	22	33	34	34	8	10	4	32	24	28	12	40
nightshot (2209)	39	19	13	36	35	29	20	8	7	38	24	18	19	20
valentine (1166)	28	12	26	39	40	39	17	14	6	20	15	24	21	24
americadelsud (505)	13	30	24	39	38	37	21	15	12	11	20	20	24	21
poppy (724)	21	37	22	35	35	37	6	13	7	25	14	30	13	29
prints (526)	19	30	37	39	37	38	12	8	4	30	8	24	15	23
fleur (3021)	10	32	26	34	35	35	7	12	4	32	21	22	16	37
colouratart (705)	13	26	23	28	30	31	13	24	3	28	19	21	33	31
bugs (669)	18	35	29	32	34	32	13	3	6	24	18	29	21	29
buds (517)	17	33	19	33	34	33	6	8	5	28	30	23	12	42
ameriquedusud (508)	14	29	24	38	38	36	22	15	12	11	20	20	24	20
insect (4744)	13	32	31	32	34	33	13	3	5	25	19	31	20	31
cheese (712)	16	25	24	37	36	32	11	14	8	24	25	21	16	32
curious (623)	23	29	32	33	35	33	11	14	4	20	21	25	19	22
sonnenuntergang (675)	23	13	18	34	29	29	28	11	11	27	22	41	23	11
santafedelaveracruz (518)	26	30	24	36	37	32	13	4	6	29	18	30	22	13
embroidery (859)	32	15	7	32	30	29	18	12	9	32	25	25	24	29
native (761)	25	29	29	31	33	32	13	10	4	24	22	27	17	23
wildflowers (604)	21	35	30	38	42	38	6	11	4	15	17	25	16	21
atardecer (2393)	26	11	17	31	28	28	27	12	9	27	23	43	24	11
macrophotosnolimits (1198)	15	34	27	32	33	34	11	4	5	31	18	30	21	22
dragoncon (910)	25	11	24	27	29	30	15	6	12	25	29	44	23	17
nacht (909)	40	18	13	36	34	28	19	7	8	35	23	18	17	21
gt (715)	21	22	13	24	18	25	18	16	7	31	19	49	32	22
4x5 (586)	26	21	9	30	33	29	17	8	9	30	26	20	34	25
animalia (880)	28	21	26	34	31	33	22	5	7	19	19	27	22	22
amanecer (621)	22	12	13	29	26	24	29	12	8	27	28	44	28	14
rodent (548)	25	13	23	34	34	33	17	7	5	27	20	33	21	23
rose (4253)	10	27	25	29	31	30	8	17	8	34	20	29	14	32
monochrome (3155)	20	57	9	50	48	51	6	6	2	19	3	18	4	21
hbw (1469)	14	25	13	20	22	25	9	4	4	48	23	38	13	56
lily (1086)	14	30	19	31	31	33	12	10	7	30	24	28	12	33
sunset (21676)	22	15	16	34	30	30	26	9	9	25	23	41	21	12
macromix (955)	13	42	25	37	39	36	10	3	6	23	22	21	17	19
gig (1138)	38	18	22	37	29	28	14	8	4	28	21	25	19	21
crossprocessing (1131)	25	26	25	39	42	40	22	7	2	19	12	20	12	20
t4l (973)	27	12	17	21	23	21	19	10	5	31	29	35	20	41
surfing (849)	29	21	31	34	27	30	10	7	3	18	24	40	15	22
stem (507)	9	27	19	33	33	34	7	9	5	31	20	42	17	25
neon (2569)	28	29	20	33	35	32	14	12	3	32	15	25	15	17
crossprocessed (1634)	24	25	27	38	41	39	24	10	2	15	12	18	13	22
tramonto (1533)	23	11	16	32	28	28	28	11	7	28	23	43	21	11

B.3 The characteristics ranked by significance

The following table ranks the characteristics by their average significance value for all characteristics.

gabor filter hist	17.8
CIE-Lch hist	17.4
CIE-Lab hist	17.2
RGB hist	17.1
details hist	15.9
linear binary pattern hist	15.1
frequency hist	14.0
gabor filter layout	12.8
chroma hist	11.9
hue angle hist	11.3
graylevel hist	11.2
lightness layout	8.6
chroma layout	6.3
hue angle layout	4.6

Table B.3: Significance for 14 descriptors averaged over the 2858 most frequently used keywords.

Appendix C

Tone-Mapping Examples

In the following we show two semantic tone-mapping examples for eight different keywords (see also Chapter 4). The complete psychophysical experiment on Amazon Mechanical Turk comprises 30 images per keyword. The full browsable collection with all images can be found here:
http://ivrg.epfl.ch/SemanticEnhancement.html.

input | *white*, $S = 1$

approval = 94% – photo attribution: David Rout.

input | *white*, $S = 1$

approval = 89% – photo attribution: Marcia Peterson.

input | *dark*, $S = 1$

approval = 89% – photo attribution: Judy Olesen.

input | *dark*, $S = 1$

approval = 83% – photo attribution: Andrew Connell.

input

sand, $S = 1$

approval = 76% – photo attribution: Njambi Ndiba.

input

sand, $S = 1$

approval = 88% – photo attribution: Stanislav Miticky.

input | *snow*, $S = 1$

approval = 65% – photo attribution: Femkje Stroop.

input | *snow*, $S = 1$

approval = 83% – photo attribution: Marco Imber.

input | *contrast*, $S = 1$

approval = 79% – photo attribution: Nick Rooney.

input *contrast*, $S = 1$

approval = 74% – photo attribution: Lee Sarter.

input *silhouette*, $S = 1$

approval = 74% – photo attribution: Paul Salort.

input *silhouette*, $S = 1$

approval = 77% – photo attribution: Lindsay Bell.

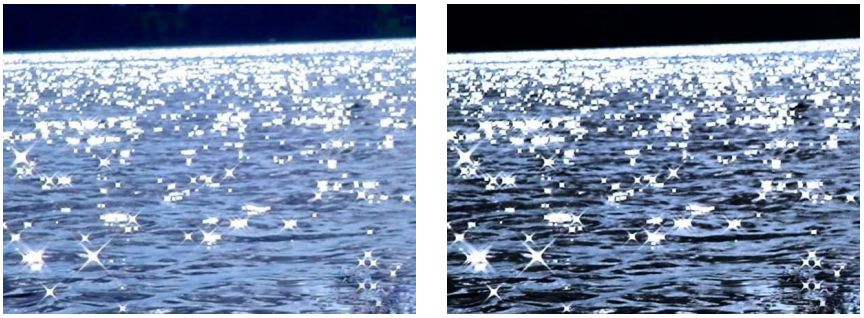
input | *portrait*, $S = 1$

approval = 72% – photo attribution: Alonso Manuel.

input | *portrait*, $S = 1$

approval = 48% – photo attribution: Scott Halford.

input | *light*, $S = 1$



approval = 19% (artists: 77%) – photo attribution: Marjut Sajadi.

input

light, $S = 1$

approval = 28% (artists: 85%) – photo attribution: *pondhoppers* (Flickr).

Appendix D

Derivation for z^* values

The significance value z of a statistical test depends on the number of samples observed: the more samples the more significant the result. It is thus not possible to directly compare the significance values from two tests with different sample sizes. The equivalent to "statistical significance" but without the dependence on the sample size is called "effect size".

To our knowledge there is no measure of effect size for our given scenario that could be implemented as efficiently as the **MWW** test for a large number of consecutive tests on the same data (see Section 3.1.3). Nevertheless, we can compare the significance values for different keywords by computing, based on a given test result, how significant the test result would have been if it had been done with a different sample size.

Let $X_1 = \{x_1^1, \dots x_{n_1}^1\}$ and $X_2 = \{x_1^2, \dots x_{n_2}^2\}$ be two sets with cardinalities n_1 and n_2, respectively. To compute the ranksum statistic T, the values in the joint set $X_1 \cup X_2$ have to be sorted. The values $x_1^1, \dots x_{n_1}^1$ then have assigned rank indexes $r_1, \dots r_{n_1}$ with $r_i \in \{1, 2, \dots n_1 + n_2\}$. The rank indexes of the second set are not considered.

The ranksum statistic T of the **MWW** test is the sum of the rank indexes r_i of the first set's elements [106, 57]:

$$T = \sum_{i=1}^{n_1} r_i \tag{D.1}$$

and the expected mean and variance of the statistic T are:

$$\mu_T = \frac{n_1(n_1 + n_2 + 1)}{2} \tag{D.2a}$$

$$\sigma_T^2 = \frac{n_1 n_2(n_1 + n_2 + 1)}{12} \tag{D.2b}$$

In order to investigate the influence of the set cardinality we have a closer look at the test statistic. The expected value of T is the sum of the expected values of the rank indexes r_i:

$$E[T] = E\left[\sum_{i=1}^{n_1} r_i\right] = \sum_{i=1}^{n_1} E\left[r_i\right] \tag{D.3}$$

As the expectation of the rank indexes does not depend on their order, it can be considered as a constant R that solely depends on the underlying distributions of the values in both sets and the cardinality of the joint set $N = n_1 + n_2$. We thus obtain

$$E[T] = \sum_{i=1}^{n_1} R = n_1 \cdot R \tag{D.4}$$

We now consider the case where the sets have varying cardinalities, but the total number of values $N = n_1 + n_2$ is constant and $n_1 << N$. If the underlying distributions are fixed, then the ranksum T is expected to depend linearly on the first set's cardinality $|X_1| = n_1$.

We can thus compute the expected ranksum statistic T^* that would have been obtained in a hypothetical test where the cardinality of the first set is n_1^* instead of n_1 (and $n_2^* = N - n_1^*$ instead of n_2 for the second set) as:

$$T^* = \frac{n_1^*}{n_1} T \tag{D.5}$$

The expected mean and variance of the hypothetical test are:

$$\mu_T^* = \frac{n_1^*(n_1^* + n_2^* + 1)}{2} = \frac{n_1^*(n_1 + n_2 + 1)}{2} = \frac{n_1^*}{n_1}\mu_T \tag{D.6a}$$

$$\sigma_T^{*2} = \frac{n_1^* n_2^*(n_1^* + n_2^* + 1)}{12} = \frac{n_1^* n_2^*(n_1 + n_2 + 1)}{12} \tag{D.6b}$$

$$= \frac{n_1^* n_2^*}{n_1 n_2}\sigma_T^2 \tag{D.6c}$$

and the normalized test statistic:

$$z^* = \frac{T^* - \mu_T^*}{\sigma_T^*} = \frac{\frac{n_1^*}{n_1}(T - \mu_T)}{\sqrt{\frac{n_1^* n_2^*}{n_1 n_2}}\sigma_T} = \sqrt{\frac{n_1^* n_2}{n_1 n_2^*}} \cdot z \tag{D.7}$$

This equation makes it possible to normalize the significance values of many tests with different sample sizes to a reference sample size n_1^*. As this eliminates the dependency on the sample size, one can directly compare the z^* values of different tests.

Bibliography

[1] "Image cross language evaluation forum," http://www.imageclef.org/, last checked Jan. 2013. (Cited on page 11.)

[2] "ImageNet large scale visual recognition challenge," http://www.image-net.org/, last checked Jan. 2013. (Cited on pages 11 and 12.)

[3] "Pascal visual object classes," http://pascallin.ecs.soton.ac.uk/ challenges/VOC/, last checked Jan. 2013. (Cited on page 11.)

[4] I. 20462-1, *Photography - psychophysical experimental methods to estimate image quality - part 1: Overview of psychophysical elements*, International Organization for Standardization, 2004. (Cited on page 25.)

[5] G. K. Adams, "An Experimental Study of Memory Color and Related Phenomena," *The American Journal of Psychology*, vol. 34, no. 3, pp. 359–407, July 1923. (Cited on page 28.)

[6] S. Bae and F. Durand, "Defocus magnification," *Eurographics*, vol. 26, no. 3, pp. 571–579, 2007. (Cited on pages 23 and 24.)

[7] K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan, "Matching words and pictures," *Journal of Machine Learning Research*, vol. 3, pp. 1107–1135, February 2003. (Cited on page 18.)

[8] C. J. Bartleson, "Memory Colors of Familiar Objects," *Journal of the Optical Society of America*, vol. 50, no. 1, pp. 73–77, 1960. (Cited on pages 28 and 29.)

[9] B. Berlin and P. Kay, *Basic Color Terms: Their Universality and Evolution*. University of California Press, 1969. (Cited on page 27.)

[10] M. Bober, F. Preteux, and W. Y. Kim, "MPEG-7 visual shape descriptors," Mitsubishi Electric Information Technology Centre Europe, Tech. Rep. VIL01D112, 2001. (Cited on page 11.)

[11] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. the method of paired comparisons," *Biometrika*, vol. 39, no. 3/4, pp. 324–345, 1952. (Cited on page 25.)

[12] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer Networks and ISDN Systems*, vol. 30, no. 1-7, pp. 107–117, 1998. (Cited on page 18.)

[13] R. BT.500-13, *Methodology for the subjective assessment of the quality of television pictures*, International Telecommunication Union, 2012. (Cited on page 25.)

[14] K.-T. Chen, C.-C. Wu, Y.-C. Chang, and C.-L. Lei, "A crowdsourceable qoe evaluation framework for multimedia content," in Proc. *ACM International Conference on Multimedia*, 2009, pp. 491–500. (Cited on page 26.)

[15] L. Cieplinski, "MPEG-7 Color Descriptors and Their Applications," in Proc. *Computer Analysis of Images and Patterns*, vol. 2124, 2001, pp. 11–20. (Cited on pages 11, 12, and 40.)

[16] G. Ciocca, C. Cusano, F. Gasparini, and R. Schettini, "Content aware image enhancement," in Proc. *Artificial Intelligence and Human-Oriented Computing*, vol. 4733/2007, Rome, 2007, pp. 686–697. (Cited on page 23.)

[17] D. Cohen-Or, O. Sorkine, R. Gal, T. Leyvand, and Y.-Q. Xu, "Color harmonization," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 624–630, 2006. (Cited on page 20.)

[18] Color Database, http://www.perbang.dk, last checked Jan. 2013. (Cited on pages 74 and 84.)

[19] X. color survey, http://blog.xkcd.com/2010/05/03/color-survey-results/, last checked Jan. 2013. (Cited on pages 83 and 84.)

[20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: a large-scale hierarchical image database," in Proc. *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255. (Cited on page 11.)

[21] T. Deselaers and V. Ferrari, "Visual and semantic similarity in imagenet," in Proc. *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1777–1784. (Cited on page 12.)

[22] D. Enke and S. Thawornwong, "The use of datamining and neural networks for forecasting stock market returns," *Expert Systems with Applications*, vol. 27, no. 4, pp. 927–940, 2005. (Cited on page 18.)

[23] F. J. Estrada, "Time-lapse image fusion," in Proc. *Colour and Photometry in Computer Vision Workshop at ECCV*, 2012, p. 441. (Cited on page 24.)

[24] Facebook, Inc., "Form s-1 registration statement under the securities act of 1933," *United States Securities and Exchange Commission*, Registration No. 333, February 1, 2012. (Cited on page 1.)

[25] G. T. Fechner, *Elemente der Psychophysik*. Breitkopf und Härtel, 1860. (Cited on page 25.)

[26] Flickr API, http://www.flickr.com/services/api/, last checked Jan. 2013. (Cited on page 70.)

[27] J. Frank J. Massey, "The Kolmogorov-Smirnov Test for Goodness of Fit," *Journal of the American Statistical Association*, vol. 46, pp. 68–78, 1951. (Cited on page 15.)

[28] C. Fredembach, "Saliency as compact regions for local image enhancement," in Proc. *IS&T Color and Imaging Conference*, 2011, pp. 14–18. (Cited on page 20.)

[29] G. A. Gescheider, *Psychophysics: The Fundamentals*. Lawrence Erlbaum Assoc Inc, 1997, vol. 3. (Cited on page 25.)

[30] T. Gevers, A. Gijsenij, J. van de Weijer, and J.-M. Geusebroek, *Color in Computer Vision: Fundamentals and Applications*. John Wiley & Sons, 2012. (Cited on page 27.)

[31] J. Good, "How many photos have ever been taken?" http://blog.1000memories.com/94-number-of-photos-ever-taken-digital-and-analog-in-shoebox, last checked Jan. 2013. (Cited on page 2.)

[32] Google Custom Search API, https://developers.google.com/custom-search/docs/xml_results, last checked Jan. 2013. (Cited on pages 19 and 84.)

[33] Harvard University, Graphics, Vision & Interaction, "Content-specific image enhancement using large image collections," http://gvi.seas.harvard.edu/node/276, last checked Sept. 2012. (Cited on page 22.)

[34] K. E. K. Hering, *Zur Lehre vom Lichtsinne*. G. A. Agoston, 1878. (Cited on page 28.)

[35] M. J. Huiskes and M. S. Lew, "The MIR flickr retrieval evaluation," in Proc. *ACM International Conference on Multimedia Information Retrieval*, 2008. (Cited on pages 10 and 18.)

[36] M. J. Huiskes, B. Thomee, and M. S. Lew, "New trends and ideas in visual concept detection: The mir flickr retrieval evaluation initiative," in Proc. *ACM International Conference on Multimedia*, 2010, pp. 527–536. (Cited on pages 7, 10, 18, 31, and 109.)

[37] R. Hummel, "Image Enhancement by Histogram Transformation," *Computer Graphics and Image Processing*, vol. 6, no. 2, pp. 184–195, 1977. (Cited on pages 20 and 21.)

[38] IDC Corporate USA, http://www.idc.com, last checked Jan. 2013. (Cited on page 2.)

[39] ISO 11664-4:2008(E)/CIE S 014-4/E:2007, *CIE Colorimetry - Part 4: 1976 L*a*b* Colour Space*, 2007. (Cited on page 35.)

[40] M. Jaber, E. Saber, and F. Sahin, "Extraction of Memory Colors Using Bayesian Networks," in Proc. *IEEE International Conference on System of Systems Engineering*, 2009, pp. 1–6. (Cited on page 29.)

[41] S. B. Kang, A. Kapoor, and D. Lischinski, "Personalization of image enhancement," in Proc. *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 1799–1806. (Cited on page 21.)

[42] C. Keimel, J. Habigt, and K. Diepold, "Challenges in crowd-based video quality assessment," in Proc. *International Conference on Quality of Multimedia Experience*, 2012, pp. 13– 18. (Cited on page 26.)

[43] C. Keimel, J. Habigt, C. Horch, and K. Diepold, "Qualitycrowd – a framework for crowd-based quality evaluation," in Proc. *Picture Coding Symposium*, 2012, pp. 245–248. (Cited on page 26.)

[44] E. Kirkos, C. Spathis, and Y. Manolopoulos, "Data mining techniques for the detection of fraudulent financial statements," *Expert Systems with Applications*, vol. 32, pp. 995–1003, 2007. (Cited on page 18.)

[45] A. Kolmogorov, "Sulla determinazione empirica di una legge di distributione," *Giornale dell' Istituto Italiano degli Attuari*, pp. 83–91, 1933. (Cited on page 15.)

[46] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951. (Cited on page 32.)

[47] S. A. Kunath and S. H. Weinberger, "The wisdom of the crowd's ear: Speech accent rating and annotation with amazon mechanical turk," in Proc. *NAACL HLT Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, 2010, pp. 168–171. (Cited on page 26.)

[48] Q. V. Le, M. Ranzato, R. Monga, M. Devin, K. Chen, G. S. Corrado, J. Dean, and A. Y. Ng, "Building high-level features using large scale unsupervised learning," in Proc. *International Conference in Machine Learning*, vol. arXiv:1112.6209, 2012. (Cited on page 18.)

[49] E. Levina and P. Bickel, "The Earth Mover's Distance is the Mallows Distance: Some Insights from Statistics," in Proc. *IEEE International Conference on Computer Vision*, vol. 2, 2001, pp. 251–256. (Cited on page 32.)

[50] J. Li and J. Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, 2003. (Cited on page 18.)

[51] A. Lindner, N. Bonnier, and S. Süsstrunk, "What is the color of chocolate? - extracting color values of semantic expressions," in Proc. *Conference on Color in Graphics, Imaging and Vision*, 2012. (Cited on page 18.)

[52] A. Lindner, B. Z. Li, N. Bonnier, and S. Süsstrunk, "A large-scale multilingual color thesaurus," in Proc. *IS&T Color and Imaging Conference*, 2012, pp. 30–35. (Cited on page 18.)

[53] A. Lindner, A. Shaji, N. Bonnier, and S. Süsstrunk, "Joint statistical analysis of images and keywords with applications in semantic image enhancement," *ACM International Conference on Multimedia*, 2012. (Cited on page 18.)

[54] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004. (Cited on pages 11 and 22.)

[55] R. D. Luce, *Individual choice behavior; a theoretical analysis.* Wiley, 1959. (Cited on page 25.)

[56] T. Mäenpää, "The Local Binary Pattern Approach To Texture Analysis – Extensions and Applications," Ph.D. dissertation, University of Oulu, 2003. (Cited on pages 11, 35, 39, and 107.)

[57] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947. (Cited on pages 14, 32, and 125.)

[58] W. Mason and D. J. Watts, "Financial incentives and the "performance of crowds"," in Proc. *ACM SIGKDD Workshop on Human Computation*, 2009, pp. 77–85. (Cited on page 27.)

[59] N. Moroney, "Unconstrained web-based color naming experiment," in Proc. *IS&T/SPIE Symposium on Electronic Imaging*, vol. 5008, Color Imaging VIII: Processing, Hardcopy, and Applications, 2003, pp. 36–46. (Cited on pages 27, 84, and 98.)

[60] S. Moser and M. Schroeder, "Usage of DSC meta tags in a general automatic image enhancement system," in Proc. *IS&T/SPIE Symposium on Electronic Imaging*, vol. 4669, San Jose, CA, USA, January 2002, pp. 259–267. (Cited on page 23.)

[61] A. H. Munsell, *Munsell Book of Color*. The Munsell Color Company, 1929. (Cited on page 28.)

[62] N. Murray, S. Skaff, and L. Marchesotti, "Towards automatic concept transfer," in Proc. *ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering*, 2011, pp. 167–176. (Cited on pages 20 and 60.)

[63] D. Mylonas, L. MacDonald, and S. Wuerger, "Towards an Online Color Naming Model," in Proc. *IS&T Color and Imaging Conference*, 2010, pp. 140–144. (Cited on page 27.)

[64] F. Naccari, S. Battiato, A. R. Bruna, A. Capra, and A.Castorina, "Natural scenes classification for color enhancement," *IEEE Transactions on Consumer Electronics*, vol. 51, no. 1, pp. 234 – 239, 2005. (Cited on page 29.)

[65] V. Namboodiri, "Recovery of relative depth from a single observation using an uncalibrated (real-aperture) camera," in Proc. *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–6. (Cited on page 23.)

[66] Nathan Moroney, http://www.hpl.hp.com/personal/ Nathan_Moroney/, last checked Jan. 2013. (Cited on pages 27, 28, and 69.)

[67] E. W. Ngai, L. Xiu, and D. Chau, "Application of datamining techniques in customer relationship management: A literature review and classification," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2592–2602, 2009. (Cited on page 18.)

[68] D. Nickerson, "History of the Munsell Color System and Its Scientific Application," *Journal of the Optical Society of America*, vol. 30, no. 12, pp. 575–586, December 1940. (Cited on page 28.)

[69] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Gray scale and rotation invariant texture classification with local binary patterns," in Proc. *ECCV*, vol. 1842/2000, 2000, pp. 404–420. (Cited on page 107.)

[70] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001. (Cited on page 22.)

[71] Oxford English Dictionary, http://oxforddictionaries.com/words/how-many-words-are-there-in-the-english-language, last checked Oct. 2012. (Cited on page 3.)

[72] R. P.910, *Subjective video quality assessment methods for multimedia applications*, International Telecommunication Union, 1999. (Cited on page 25.)

[73] G. Paolacci, J. Chandler, and P. G. Ipeirotis, "Running experiments on amazon mechanical turk," *Judgment and Decision Making*, vol. 5, no. 5, pp. 411–419, 2010. (Cited on page 26.)

[74] D.-S. Park, Y. Kwak, H. Ok, and C.-Y. Kim, "Preferred skin color reproduction on the display," *Journal of Electronic Imaging*, vol. 15, no. 4, 2006. (Cited on page 28.)

[75] C. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," School of Business Systems, Monash University," arXiv:1009.6119, 2010. (Cited on page 18.)

[76] 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," *Journal of Computer Vision, Graphics, and Image Processing*, vol. 39, pp. 355–368, 1987. (Cited on page 49.)

[77] R. L. Plackett, "Karl pearson and the chi-squared test," *International Statistical Review*, vol. 51, no. 1, pp. 59–72, 1983. (Cited on page 15.)

[78] A. Polesel, G. Ramponi, and V. J. Mathews, "Image enhancement via adaptive unsharp masking," *IEEE Transactions in Image Processing*, vol. 9, no. 3, pp. 505–510, 2000. (Cited on page 20.)

[79] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Journal of Computer Graphics and Applications*, vol. 21, no. 5, pp. 2–9, 2001. (Cited on page 20.)

[80] F. Ribeiro, D. Florencio, and V. Nascimento, "Crowdsourcing subjective image quality evaluation," in Proc. *International Conference on Image Processing*, 2011, pp. 3097–3100. (Cited on page 26.)

[81] F. Ribeiro, D. Florencio, C. Zhang, and M. Seltzer, "CROWDMOS: An Approach for Crowdsourcing Mean Opinion Score Studies," in Proc. *International Conference on Acoustics, Speech, and Signal Processing*, 2011, pp. 2416 – 2419. (Cited on pages 26 and 27.)

[82] D. Riecks, *Photo Metadata*, International Press Telecommunications Council, Standard, 2010. (Cited on page 10.)

[83] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert Systems with Applications*, vol. 33, pp. 135–146, 2007. (Cited on page 18.)

[84] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: a database and web-based tool for image annotation," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 157–173, 2008. (Cited on page 11.)

[85] B. T. Ryu, J. Y. Yeom, C.-W. Kim, J.-Y. Ahn, D.-W. Kang, and H.-H. Shin, "Extraction of Memory Colors for Preferred Color Correction in Digital TVs," in Proc. *Color Imaging XIV: Displaying, Processing, Hardcopy, and Applications*, vol. 7241. IS&T/SPIE Symposium on Electronic Imaging, 2009. (Cited on page 29.)

[86] C. Sauvaget and V. Boyer, "Harmonic colorization using proportion contrast," in Proc. *ACM AFRIGRAPH*, 2010, pp. 63–69. (Cited on page 20.)

[87] N. Sawant, J. Li, and J. Z. Wang, "Automatic image semantic interpretation using social action and tagging data," in Proc. *Multimedia Tools and Applications*, vol. 51, 2011, pp. 213–246. (Cited on page 18.)

[88] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, 2000. (Cited on page 9.)

[89] N. Smirnov, "Table for estimating the goodness of fit of empirical distributions," *The Annals of Mathematical Statistics*, vol. 19, no. 2, pp. 279–281, 1948. (Cited on page 15.)

[90] J. Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. Doubleday; Anchor, 2004. (Cited on page 18.)

[91] R. T.81, *Information technology – Digital compression and coding of continuous-tone still images – Requirements and guidelines*, International Telecommunication Union, 1992. (Cited on page 31.)

[92] D. Tamburrino, P. Schönmann, P. Vandewalle, and S. Süsstrunk, "The Flux: Creating a Large Annotated Image Database," in Proc. *IS&T/SPIE Symposium on Electronic Imaging*, vol. 6808, 2008, pp. 28–30. (Cited on page 18.)

[93] J. Tang, S. Yan, R. Hong, G.-J. Qi, and T.-S. Chua, "Inferring semantic concepts from community-contributed images and noisy tags," in Proc. *ACM International Conference on Multimedia*, 2009, pp. 223–232. (Cited on pages 18 and 19.)

[94] L. Taplin and G. Johnson, "When Good Hues Go Bad," in Proc. *Conference on Color in Graphics, Imaging and Vision*, 2004, pp. 348–352. (Cited on page 28.)

[95] L. L. Thurstone, "A law of comparative judgment," *Psychological Review*, vol. 34, no. 4, 1927. (Cited on page 25.)

[96] K. Tsukida and M. R. Gupta, "How to analyze paired comparison data," Department of Electrical Engineering University of Washington, Tech. Rep. UWEETR-2011-0004, 2011. (Cited on page 25.)

[97] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1582–1596, 2010. (Cited on page 11.)

[98] J. van de Weijer, C. Schmid, J. Verbeek, and D. Larlus, "Learning color names for real-world applications," *IEEE Transactions in Image Processing*, vol. 18, no. 7, pp. 1512 – 1523, 2009. (Cited on page 27.)

[99] J. Verbeek, M. Guillaumin, T. Mensink, and C. Schmid, "Image annotation with tagprop on the mirflickr set," in Proc. *ACM International Conference on Multimedia*, 2010, pp. 537–546. (Cited on page 18.)

[100] L. von Ahn and L. Dabbish, "Labeling images with a computer game," in Proc. *ACM Conference on Human Factors in Computing Systems*, 2004, pp. 319–326. (Cited on page 11.)

[101] *CSS Color Module Level 3*, W3C Recommendation, World Wide Web Consortium, 7 June 2011. (Cited on pages 74 and 84.)

[102] R. Walpole, R. Myers, and S. Myers, *Probability and Statistics*. Prentice Hall International, 1998, vol. 6. (Cited on page 33.)

[103] B. Wang, Y. Yu, T.-T. Wong, C. Chen, and Y.-Q. Xu, "Data-driven image color theme enhancement," in Proc. *ACM SIGGRAPH*, vol. 29, 2010. (Cited on pages 20 and 60.)

[104] B. Wang, Y. Yu, and Y.-Q. Xu, "Example-based image color and tone style enhancement," in Proc. *ACM SIGGRAPH Asia*, vol. 30, 2011. (Cited on pages 21 and 22.)

[105] E. H. Weber, *Tastsinn und Gemeingefühl*. Rudolph Wagner's Handwörterbuch der Physiologie, 1846. (Cited on page 25.)

[106] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945. (Cited on pages 14, 32, and 125.)

[107] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2011, vol. 3. (Cited on page 18.)

[108] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1–37, 2008. (Cited on page 18.)

[109] X11 color names, http://cvsweb.xfree86.org/cvsweb/*checkout*/xc/programs/rgb/rgb.txt?rev=1.1, last checked Sept. 2012. (Cited on page 84.)

[110] S. Yendrikhovskij, "Color reproduction and the naturalness constraint," Ph.D. dissertation, Thesis, Technische Universiteit Eindhoven, 1998. (Cited on pages 28, 78, 79, 80, and 82.)

[111] J.-Y. You and S.-I. Chien, "Saturation Enhancement of Blue Sky for Increasing Preference of Scenery Images," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 2, pp. 762–767, 2008. (Cited on page 29.)

[112] K. Zeng, M. Zhao, C. Xiong, and S.-C. Zhu, "From image parsing to painterly rendering," in Proc. *ACM Transactions on Graphics*, vol. 29, no. 1, 2009, pp. 2:1–2:11. (Cited on page 24.)

[113] M. Zhao and S.-C. Zhu, "Customizing painterly rendering styles using stroke processes," in Proc. *ACM SIGGRAPH/Eurographics Symposium on Non-Photorealistic Animation*, 2011, pp. 137–146. (Cited on page 24.)

[114] S. Zhuo and T. Sim, "Defocus map estimation from a single image," *Pattern Recognition*, vol. 44, no. 9, pp. 1852–1858, 2011. (Cited on pages 5, 23, 65, and 66.)

Curriculum Vitae

Education

2008 – 2013: PhD research at *EPFL* (Switzerland) under the supervision of Prof. Sabine Süsstrunk:

- Topic: data mining, semantic image processing, color science, machine learning.
- Responsibilities: IT infrastructure, web presence, teaching assistant, project supervision.
- MERL best student paper award at IS&T Color and Imaging Conference 2012.
- Award for best teaching assistants in the IC faculty, 2012.
- Graduation date: March 1, 2013.

2006 – 2008: Master of Science in Image and Signal Processing at *Telecom ParisTech* (France).

2002 – 2006: Master of Science in Electrical Engineering and Information Technology at *University of Stuttgart* (Germany):

- Specialization in Microelectronics and Optical Systems.
- Achieved above the 98th percentile in the *Vordiplom* (exams after 2 years).
- "Anton- und Klara Röser Stiftung" Award for top students.
- "Robert-Bosch-Stiftung" scholarship.
- Prize "Continuing excellent performance in a weekly mathematics competition".

2001 A-levels:

- Prize for the best result in the maths and physics combination option.
- Prize from the DPG (German physics association) for very good achievements in physics.

Experiences

Internships

- 2011, 5 months, *Océ*, Paris: Data mining on large scale image databases; Matlab and MEX files.
- 2007, 9 months, *Océ*, Paris: Printer Transfer Functions and Gamut Mapping Algorithms; Matlab.
- 2002, 3 months, *Porsche*, Stuttgart: Basics in Electrical Engineering and Mechanics.

Projects

- 2007, 3 months, student research project: Multi-scale object extraction in high res. images; Matlab.
- 2006, 5 months, student research project: Carbon nanotubes as a semiconducting layer for thin film transistors. Assembly in clean room and large-scale analysis of measurements; Perl, Octave.
- 2006, 2 months, team work: Noise filter for medical ultrasonic 3D-images; C.
- 2005, 5 months, team work: Design and construction of a RISC-processor; VHDL.

Other

- 2004, 1.5 years: Teaching assistant for advanced mathematics at *University of Stuttgart*.

Self employment

2001 – present: Creation of props for magicians. Acts with these gadgets won international prizes: Grand Prix Winner Monte Carlo, Sarmoti Award Las Vegas, German Master of Magic.

Leisure

soccer, swimming, cycling, alpine, reading, do it yourself