Models and Analysis of Visual Discomfort Measures for Stereoscopic Images

Thesis · March 2015

DOI: 10.13140/RG.2.1.4642.9921

1 author:

Werner Zellinger
Johannes Kepler University Linz
7 PUBLICATIONS **6** CITATIONS

SEE PROFILE

JOHANNES KEPLER
UNIVERSITÄT LINZ | JKU

TNF

Technisch-Naturwissenschaftliche
Fakultät

Models and Analysis of Visual Discomfort Measures for Stereoscopic Images

MASTERARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Masterstudium

Computermathematik

Eingereicht von:
Werner Reisner, BSc.

Angefertigt am:
Department of Knowledge-Based Mathematical Systems

Beurteilung:
Univ.-Prof. Dr. Erich Peter Klement

Mitwirkung:
Dr. Mag. Bernhard A. Moser
Assoz.-Prof.in Dr.in Mag.a Susanne Saminger-Platz

Linz, March, 2015

Acknowledgements

Diese Arbeit wäre nicht möglich gewesen ohne die Hilfe von Freunden und Kollegen. Bei sechs davon möchte ich mich hier gerne bedanken:

Danke an Univ.-Prof. Dr. Erich Peter Klement für die Betreuung dieser Arbeit.

Danke an Assoz.-Prof.in Dr.in Susanne Saminger-Platz für die Zweitbetreuung dieser Arbeit.

Danke an Dr. Thomas Natschläger und Dipl.-Ing. Johannes Himmelbauer für die Hilfe bei statistischen Fragen.

Danke an meinen Mentor in allen wissenschaftlichen und beruflichen Dingen Dr. Bernhard Moser und

Danke meiner Verlobten und einfühlsamen Freundin Marion.

Abstract

The term "visual discomfort" refers to a subjective sensation of discomfort that accompanies watching stereoscopic image- or video contents. Eye strain, headache and nausea are only some symptoms covered by this definition. It is generally agreed that visual discomfort is a key aspect of the quality of experience when watching 3D content and therefore an important part influencing the overall acceptance of 3D technology. Re-rendering of 3D contents can prevent visual discomfort and is based on an accurate prediction of that phenomenon.
Consequently designing visual comfort measures, which predict visual discomfort, is a major research-activity of 3D technology production and 3D display systems.

In literature, various approaches modelling the extent of visual discomfort based on image data analysis, can be found. All these approaches are based on the extraction of image features as input for machine learning models. Therefore, the correct selection of these image features influences the accuracy of visual comfort measures. Although lots of research has been done in this direction, state-of-the-art measures can still be improved.

This work addresses the quality of state-of-the-art visual discomfort prediction measures and the right choice of the used image features. Particularly, the computational efficiency of such models is investigated. It turns out, that a novel approach based on a texture image feature improves state-of-the-art computational models for measuring visual comfort in terms of accuracy and, above all, time complexity. This result is underpinned by statistical tests on public available databases.

Contents

List of Figures

1. Introduction

In literature the term *visual discomfort* refers to a subjective sensation of discomfort that accompanies watching stereoscopic image- or video streams [25] (more precise definition later). Eye strain, headache and nausea are only some symptoms covered by the definition of visual discomfort [40]. Results of clinical and subjective assessments show that bad quality 3D content can cause permanent damage to the visual system of children [11] and, visual discomfort turns out to be a key aspect of the overall quality of experience when watching 3D content [23, 28]. As a consequence, visual discomfort has direct impact on the overall acceptance of 3D technology [32], particularly in the context of 3D film production and 3D display systems [47, 48, 16, 20].

To minimize visual discomfort, significant work is required during post-production of the 3D production workflow. This post-production step is called depth grading and is the step where a stereographer tries to adjust 3D content to ensure immersive, yet comfortable experience. In order to adapt the 3D content to the viewing system or even automate the depth grading process, visual discomfort measures are required.

These measures try to predict the level of visual discomfort which accompanies 3D content. In literature, various approaches modelling the extent of visual discomfort based on image data analysis, can be found. All these approaches are based on the extraction of image features from a depth map storing the depth information of the stereoscopic images. Based on the extracted features, machine learning functions are used as a mapping from the feature space onto a range of visual discomfort scales.

Although, powerful machine learning functions exist, the quality of visual discomfort measures is primarily based on an accurate choice of the used features. Thus, most researchers try to model new and powerful image features, as they use only linear or piecewise-linear regression functions. This is also because, interpretability is an important property of visual discomfort measures, which is only given by comprehensible feature aggregation.

There are authors who prefer low-level features like depth range or mean depth of pixels, mainly based on first order statistics [24, 30, 3] that are derived from a depth map called *disparity map*. Recently, Sohn et al. [38] proposed the application of higher level image analysis techniques like segmentation to describe objects in an stereoscopic image. They use features like object thickness and relative difference of mean disparity of objects. The latter approach allows a substantial improvement of the prediction accuracy of visual discomfort measures compared to measures based on first order statistics only.

Unfortunately these higher level features need substantially more computation time to evaluate than approaches based on low level features. This fact can be problematic since visual discomfort measures often are part of computational models in the 3D post-production [29,

39, 41].

The central questions of this work are about the optimal combination of image features in terms of balancing both, prediction accuracy and time complexity.

This is done by analysing image features of some representative state-of-the-art models and the used features. The analysis shows, that using second order statistics is a new approach in the context of visual discomfort measures. Furthermore, a second order statistical feature from the field of texture analysis is proposed. It is called Haralick Contrast (*HC*) and it is used to model the contrast of a disparity map. It turns out that this feature allows substantial improvements in terms of prediction accuracy and runtime of state-of-the-art visual discomfort measures.

In more detail, the experimental evaluations of this work address the validation of the following four claims with respect to the approaches of [24, 30, 3, 21, 38]:

Claim 1 (Prediction Accuracy)
The expected prediction accuracy, which can be achieved by combinations including *HC*, is significantly higher than for combinations without *HC*.

Claim 2 (Feature Selection)
Taking the eight features under consideration, inclusive *HC*, into account, a total number of four features is appropriate to predict visual discomfort.

Claim 3 (Time Complexity)
HC allows substantial time complexity improvement without significant loss of prediction accuracy, compared to the state-of-the-art approaches under consideration.

Claim 4 (Further Improvement)
The prediction accuracy which can be achieved by the best combination without *HC* can be improved by *HC*.

These claims are underpinned by statistically tests on two publicly available databases [19, 10].

2. Problem Description and Structure of This Work

This work is partially supported by the Software Competence Center Hagenberg© (SCCH), which tries to optimize a special step in the 3D content production workflow. This post-production step is called depth grading and is the step where a stereographer tries to adjust the 3D content to ensure immersive, yet comfortable, experience. In order to optimize this step, a new and automated 3D re-rendering software is under development.

The central problem of this software is about an accurate re-rendering of stereoscopic images in order to minimize the effects of visual discomfort. Visual discomfort refers to a subjective sensation of discomfort that often accompanies watching stereoscopic image- or video contents, which will be defined in more detail later (section 3.1)

Minimization of visual discomfort requires an accurate prediction of this phenomenon. Thus, the development of visual discomfort measures has become a major research-activity of 3D technology production and 3D display systems.

2.1. Problem Description

Unfortunately, state-of-the-art prediction models often show high evaluation time, which can cause problems if they are used in optimization processes. This work should deal with this problems. More precisely formulated:

Goal
This work should answer the question, if a visual discomfort measure can be designed, that achieves the same prediction accuracy as state-of-the-art models, while it needs comparably lower runtime.

The answer to this question is Yes, which is verified in this work by the development of a visual discomfort measure which has the required properties. The approach is based on an accurate analysis and characterization of some representative state-of-the-art measures. The analysis of some shortcomings of these measures leads to the proposition of a new feature in that field. The relation between this feature, called Haralick Contrast (HC), and visual discomfort is discussed. Afterwords, the advantages of the HC feature compared to other state-of-the-art image features, used for visual discomfort prediction, are outlined. This work concludes with experiments on two publicly available databases, which underpin arising claims of this work.

The work is structured as follows:

2.2. Structure of this Work

Chapter 3 provides a comprehensive introduction to the concept of visual discomfort, its influencing factors and possible consequences of that phenomenon.

The basic structure of visual discomfort measures is analysed in chapter 4, which consists of four main parts: an outline of the basic structure of state-of-the-art visual discomfort measures (section 4.1, a characterization of state-of-the-art image features (section 4.2), a detailed discussion of some state-of-the-art approaches (section 4.3) and, a problem analysis of this approaches (section 4.4).

Chapter 5 is intended to give a detailed discussion of the new approach of this work, mainly based on the introduction of a feature from texture analysis. Section 5.1 gives a motivation for the approach from three research fields. Section 5.2 describes the new approach in detail. The chapter concludes with a summary of the proposed approach.

Chapter 6, after a small review on experimental design, based on the goal of this work, provides experiments, discusses several properties of state-of-the-art models and compares the results with the approach of this work. The experimental analysis should provide evidence regarding the following aspects: prediction accuracy (section 6.3), feature selection (6.4), and, time complexity versus prediction accuracy (section 6.5).

Appendix A is intended to give the basics for all algorithms, which are addressed.

3. Visual Discomfort

This chapter provides a comprehensive introduction to the concept of visual discomfort. After a short definition of this term, some consequences of that phenomenon are outlined. The main part of this chapter consists of the outline of influencing factors of visual discomfort.

3.1. Basic Definition

In the literature, the terms *visual fatigue* and *visual comfort* have been used interchangeably to describe the discomfort that might accompany the use of 3D imaging technologies [40]. However, Lambooij et al. [25] suggested a distinction between these two terms giving the only existing formal definition of them.

Definition 1 (Visual Discomfort and Visual Fatigue)
The term *visual fatigue* refers to a decrease in performance of the visual system produced by a physiological change. *Visual discomfort* refers to the subjective sensation of discomfort that accompanies the physiological change.

Therefore, visual discomfort can be measured by asking viewers of stereoscopic contents to report he level of perceived discomfort that occurs while watching. This leads to the fact that models, which predict visual discomfort, are based on research coming from the analysis of subjective assessment data.

People participating in these subjective assessments reported symptoms like eye strain, headache and nausea when asked about visual discomfort [40].

3.2. Consequences of Visual Discomfort

Research shows, that visual discomfort is a key aspect of the overall quality of experience that someone has when watching stereoscopic 3D content [23, 28]. Consequently it has direct impact on the acceptance of 3D technology including 3D movies and 3D display systems and has become a major research activity in that field [47, 48, 16, 20].

But more, intensive watching of 3D content accompanied by a high level of visual discomfort can seriously harm the visual system of people. Especially, the visual system of children, which is still growing, can be damaged [11].

Consequently, minimizing the viewers discomfort is a major research activity in the field of 3D technology, which has to take the full extent of the viewers watching experience into account.

3.3. Factors Influencing Visual Discomfort

To model the experience of visual discomfort it is important to identify the main factors negatively influencing visual discomfort. As outlined in chapter 1, this information is used to develop image features, which form the basis of visual discomfort measures.

Due to Tam et al. [40] the most relevant factors influencing visual discomfort can be grouped into five categories: accommodation-vergence conflict, parallax distribution, binocular mismatch, depth inconsistencies and cognitive inconsistencies.

3.3.1. Accommodation-Vergence Conflict

When watching a 3D object our visual system uses two different physiological processes to focus the object. The first process is called *accommodation process* and refers to the adaptation of the eyes lenses in order to sharp the object on the retina. The second process is called *vergence process* and it refers to the adaptation of the relative angular constellation of the eyes in order to direct the eyes at the same object.

Accommodation and vergence are normally yoked when viewing objects in a natural scene. However, the normal interaction between these two processes can be disrupted when viewing stereoscopic images as described in [40]. When looking at stereoscopic displays the responses created from these processes can cause conflicts in the visual perception. One reason for this accommodation-vergence conflict can be excessive parallax [30, 46] and it is generally assumed that, to minimize this conflict, the disparities of a stereoscopic image should be limited by so-called *comfort limits*.

The parallax of an object is the difference in the apparent position of an object viewed along two different lines of sight, i.e. position difference in the right and left eyes view. Accordingly, the word *disparity* is defined as the absolute pixel difference of two corresponding pixels in a left and right view of a stereoscopic image.

Thus, the influence of the accommodation-vergence conflict can be modelled by image features like range and maximum, extracted from the *disparity map*, which stores the disparities of all pixels.

3.3.2. Parallax Distribution

The distribution of the parallax of an object is given by the spatial arrangement of the pixels disparities. Intuitively spoken, the more image details in different depth levels, the more exists competition for visual attention at various potential objects of interest. Such ambiguity, concerning visual attention, can lead to discomfort in the visual perception.

Parallax distribution can be modelled by image features describing characteristics like spatial frequency or disparity gradient. It has been shown that, when modelled accurately, such features can have a high correlation to visual discomfort [38].

3.3.3. Binocular Mismatch and Depth Inconsistencies

The term binocular mismatch refers to pixel mismatches between the left and right view of stereoscopic images caused by distortions of the images. In [22], Kooi and Toet study the effects of distorting transformations on stereoscopic images. The results show little impact of

transformations like rotation, magnification and key stone distortions, in contrast to blur and vertical offset transformations. This transformations can be caused e.g. by lossy compression or transition of stereoscopic content and they are a possible source of visual discomfort [40].

3.3.4. Perceptual and Cognitive Inconsistencies

Perceptual and cognitive inconsistencies are caused by a mismatch between our cognition of the real world and insufficient appearance induced by a 3D display system. For example, cognitive confusion might occur when objects are only partially visible at the boarder of a display system. This effect is called *edge violation* [34]. For example, an object (e.g. hand with five fingers) which is supposed to be in front of the screen, is only partially visible (e.g. hand with four fingers) which confuses our cognition of the object (e.g. hand). A possible solution of the problem of edge violation is the use of a floating window, which consists of a virtual border perceptually located closer to the viewer than the object [40].

4. State-of-the-Art for Visual Discomfort Prediction

As outlined in section 3.1 visual discomfort can be measured using subjective assessments as [19] and [10]. Surprisingly, there are no standard methods for the measurement of visual comfort for stereoscopic images [40]. The International Telecommunications Union (ITU) gives recommendations on subjective methods for stereoscopic imaging [17], but these recommendations only consider picture and depth quality.

In light of this deficiency, most researchers analysing visual discomfort use modified versions of these recommendations. This results in different scales and questionnaires which makes it hard to compare state-of-the art visual discomfort measures. Due to this fact, in our experiments, we use two different publicly available databases storing stereoscopic images and subjective assessment data (see chapter 6).

Apart from that, this subjective assessment data is used to train visual discomfort measures which predict the level of discomfort that people will suffer.

4.1. Visual Discomfort Measures

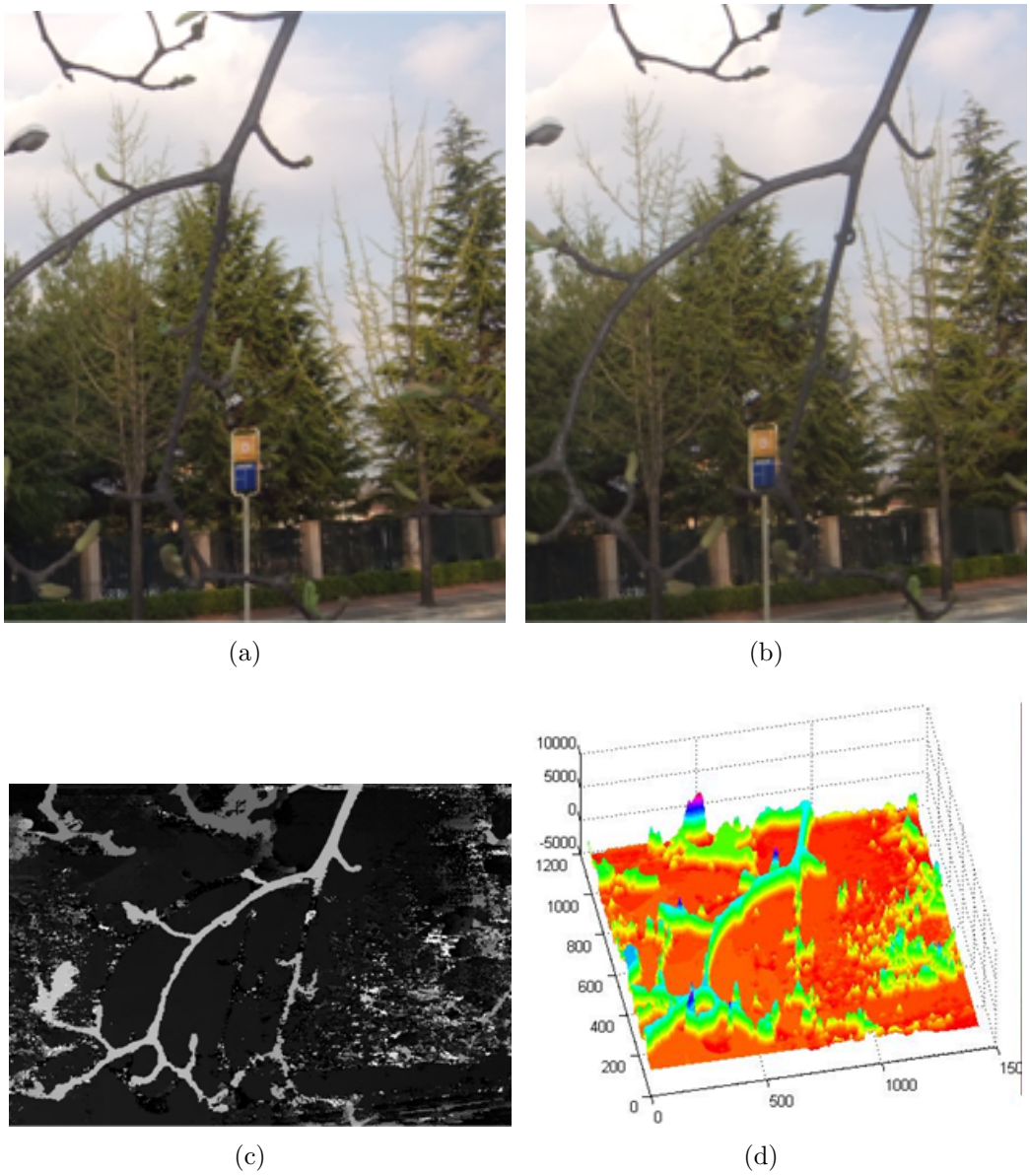
In general, visual discomfort measures are computational models for predicting visual discomfort. The input of a visual discomfort measure is a stereoscopic image normally consisting of two separate views of the same scene. The measures consist of three main components: a disparity map generation component, a feature extraction component and a machine learning function.

4.1.1. Components of Visual Discomfort Measures

The disparity map generation component extracts the depth information which is stored in the left and right view of the stereoscopic image. This component is followed by a feature extraction part and a machine learning function which maps the image features onto a range of visual discomfort scales.

Disparity Map Generation

A stereoscopic image consists of a right and a left view (fig. 4.1a and fig. 4.1b)) which are often stored as two separate files. To extract the depth information of the two streams, disparity generation algorithms are used (for instance see [14]).



(a) (b)

(c) (d)

Figure 4.1. *(a) Part of a right and (b) left view of a stereoscopic image. The disparity map of the whole image displayed as (c) grey image and (d) 3D projection.*

As defined in section 3.3.1, the disparity of a pixel in the left respectively right view is defined as the distance (measured in pixels) to the corresponding pixel in the right respectively left view. Therefore a disparity map is always related to one of this views. As an example of a possible output of the component see figures 4.1c and 4.1d.

Disparity generation algorithms normally have many parameters which can be tuned and they should work with various input images which differ in many properties like colour, luminance and resolution.

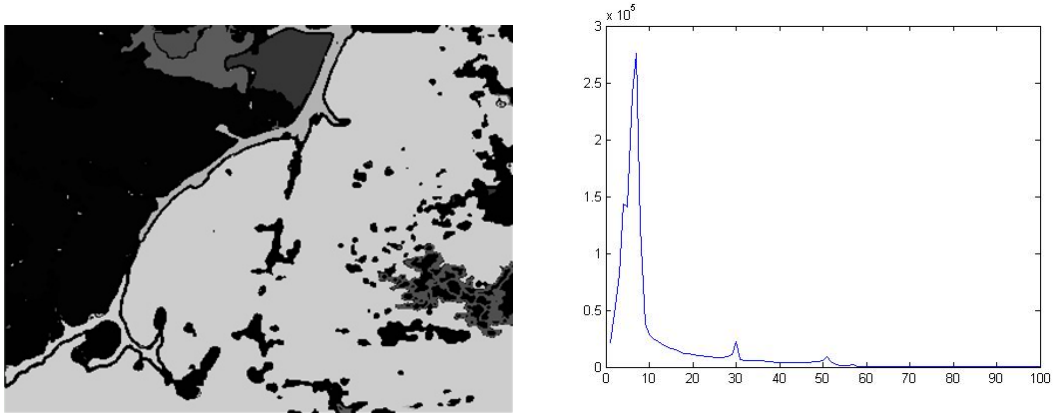
Another important fact is that some regions visible in one view can be occluded in the other view. Thus, some pixels in the disparity map are not defined. These pixels are then interpolated using so-called *hole filling algorithms*.

Due to the various input images and the occluded regions, the resulting disparity maps often include artefacts and noise. Therefore, the connected feature extraction component should be

able to deal with such erroneous disparity maps.

Feature Extraction Component

The feature extraction component is maybe the most important part of a visual discomfort measure. This component extracts image features like mean, maximum or object size from the disparity map. Figure 4.2a shows a segmented disparity map, from which object specific features can be extracted and figure 4.2b shows a histogram of a disparity map, from which features like mean or variance can be extracted.



(a) (b)

Figure 4.2. *(a) Segmented disparity map, where equal coloured pixels belong to the same object, (b) Grey level histogram of normalized disparity values.*

The used image features should have different properties like, high correlation to the perceived visual discomfort which accompanies the stereoscopic image, they should be stable against noise and artefacts and they should be fast to evaluate. Section 4.1.2 addresses these requirements in detail.

The output of the component is a vector of image features, where each represent a certain property of the image.

Machine learning Function

The extracted image features are used as input for a machine learning function. This function maps the features onto a range of scales.

Most researcher use linear [24] or piecewise linear [38] functions, which are trained using subjective assessment data.

4.1.2. Properties of Visual Discomfort Measures

In the previous sections it was argued that the superior of the visual discomfort measures lie on the prediction accuracy. But also other properties are important like time complexity, interpretability, robustness and the possibility to easily adapt the model parameters to various input.

Most state-of-the-art visual discomfort measures are based on the extraction of image features as input for linear [24] or piecewise linear [38] machine learning functions, who show the expected properties. In the following these properties are discussed in detail.

Interpretability

Visual discomfort measures are used to objectively assess image data. The results of the prediction process is often used to identify image properties or image regions which affect visual discomfort, see e.g. [39]. Therefore, the output of the measure should be directly related to some image properties. This will be especially the case, if the used machine learning function aggregates the image features in a comprehensible way and if the used features have a clear interpretation.

Thus, most researchers prefer simple mathematical models as machine learning functions and use features which are directly related to the perceived visual discomfort.

Simplicity of Parameter Adaptation

Stereoscopic images can differ in various properties like colour, luminance, resolution and blur. Thus, the scope of the applicability of visual discomfort measures is also important. This includes also the simplicity of adapting its model parameters to images of various different properties.

Most researchers use simple and parameter free machine learning functions [24, 30, 3, 21, 38]. Thus, the parameters of the measures are the one of the disparity generation component and the feature extraction component.

Since this work is not intended to analyse disparity generation algorithms, there will not be a focus on these parameters. Once the parameters of the disparity generation module are set, the major emphasises of the parameter adaptation lie on the feature extraction parameters, which will be discussed in detail.

Prediction Accuracy

As noted above, prediction accuracy is the most important property of visual discomfort measures. This property depends on the correlation of the used image features and the way how they are aggregated.

Thus, the prediction accuracy can be measured by the correlation between the combination of the used features and subjective assessed visual discomfort scores. This fact will be used later in the experimental part, to rate the quality of visual discomfort measures.

Robustness

As noted above, images can differ in various properties. This phenomenon often causes erroneous disparity (see section 4.1.1 disparity generation). In that case, machine learning functions only produce accurate results, if the feature extraction process is robust against noise. Therefore, the robustness of the visual discomfort measure is directly related to the way how features are designed.

Computational Efficiency

Visual discomfort measures are used for computational optimization in the rendering process of 3D contents [29, 39, 41]. This makes computational efficiency (i.e. runtime) a critical issue.

Some optimization procedures, including the one of the SCCH make it possible to work without the disparity generation module, because of existing disparity maps from pre-process steps. Therefore, the major runtime is required for the feature extraction process. Consequently, minimizing the runtime of the feature extraction component is an important goal of this work.

Summary

In this section some important properties of visual discomfort measures are briefly discussed. This shows that an accurate model of visual discomfort needs well designed image features with certain properties and a fast and relatively comprehensible aggregation of them. Table 4.1 summarizes the most important properties of visual discomfort measures.

Property	**Description**
Interpretability	Comprehensiveness of results
Simplicity of Parameter Adaptation	Simplicity of model parameter adaptation to images with different properties
Prediction Accuracy	Accuracy of visual discomfort scores
Robustness	Similar prediction results for images with different properties
Computational Efficiency	Low time complexity of evaluation process

Table 4.1. *Important properties of visual discomfort measures*

4.2. Categorization of Image Features Used for Visual Discomfort Prediction

To identify the differences of state-of-the-art image features used for visual discomfort prediction, we need a representative categorization of the features. This section is intended to give such a characterization. Note that the categorization of state-of-the-art image features used for visual discomfort prediction is a basic result of this work.

Different feature categories are shown in figure 4.3. The pyramid shows different levels of image features used for visual discomfort assessment. As we traverse the pyramid from the bottom up, we get increasingly higher information representation by different categories of image features.

At the very lowest level one deals with a large number of individual pixels. With this raw data, one may perform some one-dimensional sorting process of the individual pixel intensities. Using this low-level process one gets a *first-order statistic* of pixel values, i.e. distribution of a pixel value (e.g. intensity).

The *colour histogram* of an image (figure 4.2b) is an example of a first-order statistic, as it shows the distribution of the pixels colour (0-100 in the figure). First order statistics like the

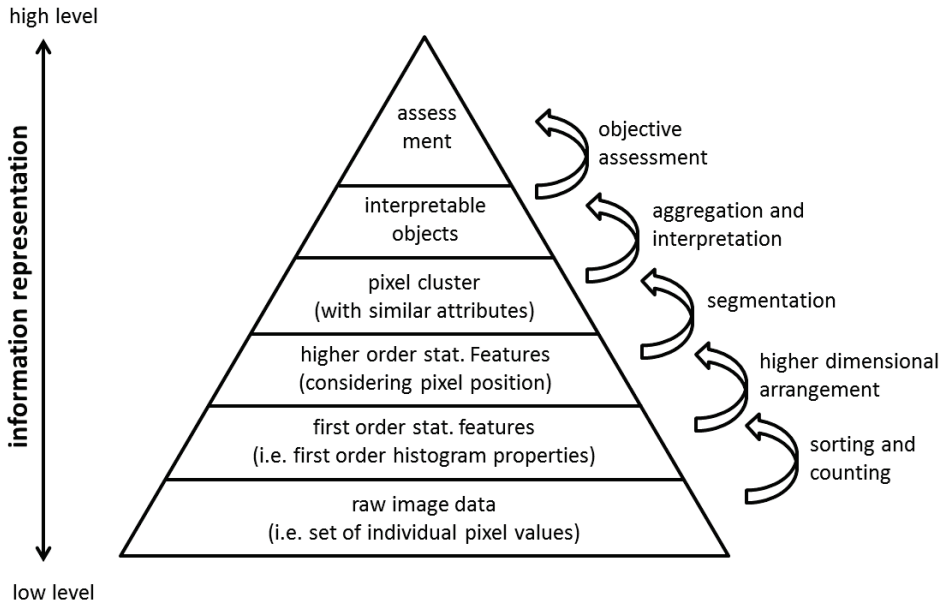


Figure 4.3. *Hierarchical feature extraction pyramid.*

colour histogram can be used for fast extraction of first-order statistical features like mean, variance or mean of the 10% greatest pixel values. Many visual discomfort measures [24, 30, 3] make use of such features derived from the disparity map.

By using higher dimensional arrangement or counting processes, one gets higher order statistics. Let us define a *k-order statistic* of an image as the joint probability distribution of pixels w.r.t. k values. For example, one can define a second-order statistic of a colour image as the joint probability distribution of the red value and the green value of pixels. This second-order statistic can be defined as matrix $\{a_{ij}\}_{i,j=0,\dots,255}$, where the red and green colour is coded by the values $0, \dots, 255$. Thus, the value a_{ij} is the number of pixels which have the red colour i and the green colour j, divided by the total number of pixels. As an other example of a second-order statistic the grey level co-occurrence matrix of a grey image will be defined in section 5.2. This matrix represents the joint probability distribution of the pixel intensities of two neighbouring pixels w.r.t. a special neighbourhood structure.

Processes like segmentation or clustering are higher level image processing steps which result in clusters of pixels often called *segments*.

By aggregation and meaningful interpretation someone can represent the scene, shown by image, as constellation of objects. This objects can be identified by real world objects who interact with each other.

The top of the pyramid is the objective assessment, in our case an objective assessment score of visual discomfort. This assessment is done by a machine learning function, which maps the image features onto a range of scales.

With this categorization in mind, state-of-the-art visual discomfort measures can be classified. The categories of the used image features help in the next steps to analyse properties and arising problems of the proposed features.

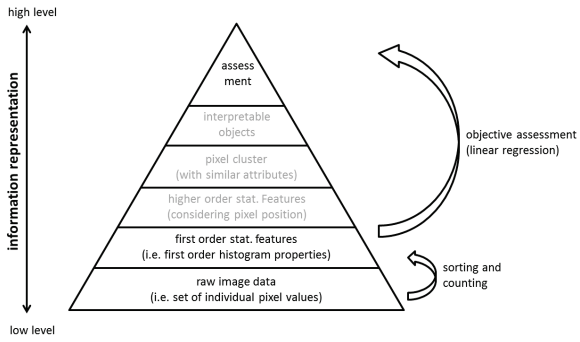
4.3. State-of-the-Art Image Features Used for Visual Discomfort Prediction

State-of-the-art visual discomfort measures can be categorized based on the used image features. In the last five years many researchers created computational models of visual discomfort by using image features representing special image properties. Most of them used first-order statistical features as defined in section 4.2. Apart from that, in 2013, at the beginning of this work, some researchers used higher level image analysis techniques [38, 19]. The latter approaches allow substantial improvement of the prediction accuracy compared to other approaches. However, these approaches have some shortcomings like high runtime, resulting in problems (see section 2), which will be discussed and solved later.

In the following we will compare some representative visual discomfort measures based on the used image features. We use the categories of section 4.2 and compare them with respect to the feature properties of section 4.1.2. This is done by short arguments concerning these properties, especially interpretability and simplicity of parameter adaptation. More sophisticated tests on the quality of the other desirable properties are done in chapter 6.

4.3.1. Choi et al. (2010)

in [3] rely on first-order statistical features to predict the visual discomfort associated with stereoscopic movies. As machine learning component, they use linear regression between their features and visual discomfort scores. They suggest to use mean and variance of the disparity map to model the mean deviation from the comfort zone and the parallax distribution (see section 3.3) respectively. The mean disparity *Mean* is calculated as the mean of all pixels in the disparity map $\mathcal{D}$. The disparity variance *Var* is calculated as the standard variance of all disparities.



$$Mean(\mathcal{D}) = \frac{1}{\#\mathcal{D}} \sum_{x,y} \mathcal{D}(x,y)$$

$$Var(\mathcal{D}) = \frac{1}{\#\mathcal{D}} \sum_{x,y} (\mathcal{D}(x,y) - Mean(\mathcal{D}))^2$$

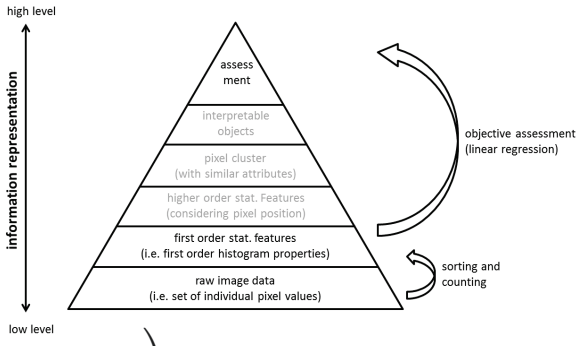
Where $\#D$ represents the cardinality of $\mathcal{D}$, i.e. the total number of pixels.

In combination with motion features their prediction results for stereoscopic movies show the statistical significance (see appendix A.5.2 F-test) of effects of the modelled characteristics. However, the correlation for the combination of the two features with visual discomfort for stereoscopic movies, was rather low (Pearson Correlation Coefficient *CC* of 0.708, *CC* is defined in appendix A.2). Experiments in chapter 6 also show rather low correlation of the single features with visual discomfort. Apart from that, *Mean* and *Var* are parameter free

features which are fast to compute. In addition to that, these features are known to be very robust against artefacts and noise. Table 4.3 summarizes these arguments.

4.3.2. Lambooij et al. (2011)

in [24] suggest to use mean and range of disparities, which are both first-order statistical features. Their *Mean* feature is chosen similarly to Choi et al. [3]. The $Range_\delta$ of a disparity map is calculated by the difference between the sum of the $\delta\%$ greatest disparities and the sum of the $\delta\%$ smallest disparities. Thus,



$$Range_\delta(\mathcal{D}) = \frac{1}{\#\mathcal{U}_\delta(\mathcal{D})}\left(\sum_{(x,y)\in\mathcal{U}_\delta(\mathcal{D})}\mathcal{D}(x,y) - \sum_{(x,y)\in\mathcal{L}_\delta(\mathcal{D})}\mathcal{D}(x,y)\right)$$

where $\mathcal{U}_\delta(\mathcal{D})$ is the set of indices of the $\delta\%$ greatest disparities and $\mathcal{L}_\delta(\mathcal{D})$ is the set of indices of the $\delta\%$ smallest disparities. $\#S$ is again the cardinality of S (i.e. the number of elements in S), therefore $\#\mathcal{U}_\delta(\mathcal{D})$ is the number of indices (x, y) in $\mathcal{U}_\delta(\mathcal{D})$ (i.e. $\delta\%$ of the total number of disparities in $\mathcal{D}$).

Lambooij et al. suggest to use delta as 10, but note that this is a heuristic value which is not underpinned by statistical evaluations.

In image processing it is a common way to calculate a maximum (minimum) feature as mean of some greatest (smallest) values. This is due to the fact, that images, especially disparity maps, are often noisy. By calculating an image feature in that way, one can overcome this drawback.

The experiments in chapter 6 show a high correlation of the $Range_{10}$ feature with visual discomfort. Since one can always take δ as 10 as suggested, there is a parameter free way to extract the feature. Using the described approach to overcome the problem of noise yields a feature which is robust against noise. Since, the range feature is used to model the range of the disparities, it is directly related to properties causing the accommodation-vergence conflict as described in section 3.3. Unfortunately the calculation of the 10% greatest and smallest values requires a sorting process which makes it slightly slower to evaluate than *Mean* and *Var*. This argumentative analysis of properties of Lambooij et al.s work [24] is summarized in table 4.3.

4.3.3. Kim et al. (2011)

in [21], aim at modelling parallax distribution and the influence of the accommodation - vergence conflict (see section 3.3). They aim at extracting range and maximum from the distribution of disparities similarly to Lambooij et al.. In addition, they model the influence of spatial frequency on visual discomfort. Spatial frequency is a characteristic of the parallax distribution and is measured in cycles per degree (cpd), i.e. the number of repetitions of a periodic pattern within a width of one degree. Because measuring the cpd is not possible in natural images they obtain a spatial frequency component by applying the Sobel-Operator to

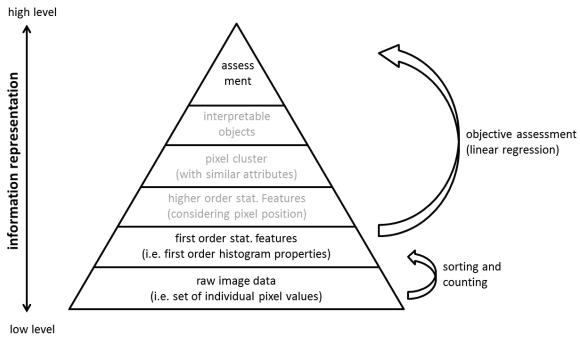
the disparity map.

Spacial frequency is modelled by adding new multiplicative factors to the maximum and range feature of Lambooij et al.. This factors are similar to maximum and range itself, but the summed disparities are weighted by gradient information induced by the application of the Sobel-Operator (appendix A.1) on the disparity map.

The maximum feature is defined by

$$Max_\delta(\mathcal{D}) = \frac{1}{\#\mathcal{U}_\delta(\mathcal{D})} \sum_{(x,y)\in\mathcal{U}_\delta(\mathcal{D})} \mathcal{D}(x,y)$$

Where $\mathcal{U}_\delta(\mathcal{D})$ is again the set of indices of the $\delta\%$ greatest disparities. The features $Max_\delta Sobel$ and $Range_\delta Sobel$ are defined by



$$\begin{aligned} Max_\delta Sobel(\mathcal{D}) &= Max_\delta(\mathcal{D}) + \lambda \cdot MaxGradient_\delta(\mathcal{D}) \\ Range_\delta Sobel(\mathcal{D}) &= Range_\delta(\mathcal{D}) + \lambda \cdot RangeGradient_\delta(\mathcal{D}) \end{aligned}$$

Where $RangeGradient_\delta(\mathcal{D})$ and $MaxGradient_\delta(\mathcal{D})$ are similar to $Range_\delta(\mathcal{D})$ and $Max_\delta(\mathcal{D})$, but the disparities are weighted by a gradient information between the disparities and their neighbours as noted above. The gradient information is induced by the application of the Sobel-Operator on the disparity map.

$$\begin{aligned} MaxGradient_\delta(\mathcal{D}) &= \frac{1}{\#\mathcal{U}_\delta(\mathcal{D})} \sum_{(x,y)\in\mathcal{U}_\delta(\mathcal{D})} \mathcal{D}(x,y)\cdot\mathcal{S}_\mathcal{D}(x,y) \\ GradientRange_\delta(\mathcal{D}) &= \frac{1}{\#\mathcal{U}_\delta(\mathcal{D})} \sum_{(x,y)\in\mathcal{U}_\delta(\mathcal{D})} \mathcal{D}(x,y)\cdot\mathcal{S}_\mathcal{D}(x,y) \\ &\quad - \frac{1}{\#\mathcal{L}_\delta(\mathcal{D})} \sum_{(x,y)\in\mathcal{L}_\delta(\mathcal{D})} \mathcal{D}(x,y)\cdot\mathcal{S}_\mathcal{D}(x,y) \end{aligned}$$

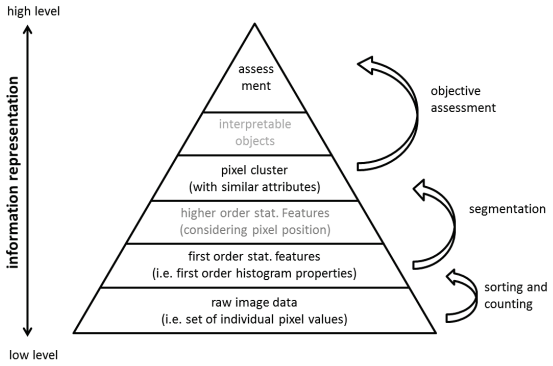
with the image $\mathcal{S}(\mathcal{D})$, which is the result of the application of he Sobel-Operator on the disparity map. Note that δ and λ are heuristic values for $Max_\delta Sobel$ and $Range_\delta Sobel$ and Kim et al. do not give any suggestions in [21] about its values. Because of heuristic observations and comparability, I chose this values as $\delta = 5$ for $Max_\delta Sobel$ and $\delta = 10$ for $Range_\delta Sobel$ and λ always as 1.

The maximum feature Max_5 has similar properties as $Range_{10}$, including the interpretation as model for the influence of the disparity range on the accommodation-vergence conflict. The computational complexity is nearly the same and the feature is parameter free. The only difference is that my experiments (chapter 6) show slightly worse prediction results for Max_5 when extracted from noisy images. This is probably because the set $\mathcal{U}_5(\mathcal{D})$ is too small to overcome the problem of very much noise. Of course someone could set this value greater than five, but in that case the correlation of the feature with visual discomfort is very low on the database of [19].

Kim et al. use $Max_{\delta}Sobel$ and $Range_{\delta}Sobel$ to model the spatial frequency of a stereoscopic image. They argue that a gradient image represent a spatial frequency component of a stereoscopic image. Note that spatial frequency is a characteristic of the parallax distribution of an image, which gives a clear interpretation for these features. Experiments (chapter 6) show that the correlation of the two features with visual discomfort is higher than the one of Max_5 and $Range_{10}$ respectively. This indicates, that $Max_{\delta}Sobel$ and $Range_{\delta}Sobel$ model the phenomenon of visual discomfort slightly better with respect to the correlation to that. In addition the features $Max_{\delta}Sobel$ and $Range_{\delta}Sobel$ give no evidence to assume sensitivity against noise in my experiments. Unfortunately the required application of the Sobel-Operator takes some time. It is important to note that the application of the Sobel-Operator induces pixel neighbour information. This approach goes towards second-order statistical features, which will be discussed in section 4.4. The arguments are summarized in table 4.3 again.

4.3.4. Sohn et al. (2013)

in [38], aim at modelling object specific properties of a scene. Their concept is based on the consideration of a scene, shown by an image, as constellation of different objects. As they refer to spatial properties of objects, the modelled image properties fall into the category of parallax distribution as factor influencing visual discomfort (section 3.3).



They introduce a concept of disparity gradient as relative disparity difference between the location of objects. In addition they aim at modelling the *stimulus width* of an object by proposing a disparity based feature, which takes the width of nearby objects into account.

As noted in the introduction, they use higher level feature extraction methods. By segmentation and grouping steps an image is splitted into parts, which are called *objects*. After that they extract the mean width of all objects, i.e. the mean length of rows included in the object (image segment). Then, for every object in the image, this value is divided by the mean disparity of the object, i.e. the mean of all disparities included in the object. They also use a logarithmic function because of empirical observation. This observations show, that approach with a logarithmic function linearizes (i.e. causes linear correlation) the relation between the factor and visual discomfort scores from subjective assessments.

More formally defined: Let o be an object (segment), i.e. a set of image pixels, and let $\mathcal{O}$ be the set of all objects in the disparity map. Let $\mathcal{R}(o)$ be the set of all horizontal lines (i.e. pixel rows) in the object $o \in \mathcal{O}$. Then the object thickness feature $OT(\mathcal{D})$ is defined by

$$OT(\mathcal{D}) \;=\; \min_{o \in \mathcal{O}} \ln\left(\alpha + \frac{\omega(o)}{\gamma(o)}\right)$$

with the mean object width $\omega(o)$ and the mean object disparity $\gamma(o)$ defined by

$$\omega(o) = \frac{1}{\#\mathcal{R}(o)} \sum_{r \in \mathcal{R}(o)} \#r$$
$$\gamma(o) = \frac{1}{\#o} \sum_{(x,y) \in o} \mathcal{D}(x,y)$$

where $|.|$ is the absolute value and alpha is chosen such that $\frac{\omega(o)}{\gamma(o)}$ is always positive. Note that in [38] the fraction $\frac{\omega}{\gamma}$ was multiplied by a factor α, to ensure that this feature is always positive in their experiments. Instead of that, I decided to rather add α to the fraction because of computational reasons. But note that this change only causes a slightly different scaling of the feature. Also note that in the original definition of the OT feature, the mean disparity $\gamma(o)$ is the mean of the absolute values of the disparities. Due to that, an object has the same object thickness feature if it is flipped on the screen, i.e. it has the same mean disparity $\gamma(o)$. This is no problem at all, if all disparities of a stereoscopic image are positive like in Sohn et al.s experiments, but there is a problem for images with negative disparities like in the experiments in chapter 6.

Of course these modifications do not change any results of the experiments of Sohn et al., which I review in chapter 6.

Now let us define the mean relative disparity feature $RD(\mathcal{D})$. Let $\mathcal{N}(o)$ be the set of neighbour objects of o, i.e. all objects which have neighbour pixels in the object o, except o itself, where neighbour pixels of (x,y) are are all pixels which have maximum euclidean distance $\sqrt{2}$ from (x,y). Then,

$$\mathcal{N}(o) = \{\tilde{o} \mid \exists(\tilde{x},\tilde{y}) \in \tilde{o}, \exists(x,y) \in o : \|(\tilde{x}-x, \tilde{y}-y)\|_2 \leq \sqrt{2}\}$$

Then

$$RD(\mathcal{D}) = \max_{o \in \mathcal{O}} \frac{1}{\#\mathcal{N}(o)} \sum_{\tilde{o} \in \mathcal{N}(o)} |\gamma(\tilde{o}) - \gamma(o)|$$

with the mean object disparity $\gamma(o)$ as defined above.

As noted in the introduction, this approach allows a substantial improvement of the prediction accuracy when using the proposed higher level features in combination with other features described in this chapter. Unfortunately, OT and RD show low correlation with visual discomfort in the experiments in chapter 6 and [38]. The extraction of the features is based on a segmentation of the disparity map. Sohn et al. recommend the mean-shift-segmentation algorithm [4]. Since this algorithm has many parameters it is not easy to adapt the parameters to new input images with different properties. This makes the features sensitive to noisy disparity maps when the parameters are not tuned properly. There is a good interpretation of the features, but the evaluation time is very high compared to first and higher order statistical features.

4.3.5. Summary

The work of four research groups is outlined and categorized above and the presented features are summarized in table 4.2. The question arises why some of the features should be used

when focusing on their shortcoming. Although some of the features have their shortcomings, the real quality of state-of-the-art measures is based on combinations of features. Therefore we want to analyse some shortcomings, which still arise even if combinations of features are used, in the next section.

Feature Notation	**Description**	**Literature**
Mean, *Var*	standard mean and variance of disparity map	Choi et al. [3]
$Range_{10}$	difference of 10% greatest and smallest disparities	Lambooij et al. [24]
Max_5	sum of 5% maximal disparity values	Kim et al. [21]
Max_5Sobel, $Range_{10}Sobel$	similar to $Range_{10}$ and Max_5 except additional factors, which are based on gradient information	Kim et al. [21]
RD, *OT*	mean relative disparity and width of nearby objects	Sohn et al. [38]

Table 4.2. *Survey of disparity-based features*

Properties	*Mean*	*Var*	$Range_{10}$	Max_5	Max_5-*Sobel*	$Range_{10}$-*Sobel*	*RD*	*OT*
Interpretability	2	2	3	3	3	3	3	3
Simplicity of Parameter Adaptation	3	3	3	3	3	3	1	1
Correlation with Visual Discomfort	1	2	3	3	3	3	1	1
Robustness w.r.t. Noise and Artefacts	3	3	3	2	3	3	2	2
Computational Efficiency	3	3	2	2	2	2	1	1

Table 4.3. *Desirable properties of image features used for visual discomfort measurement. Rating:* 1 - *neutral,* 2 - *good,* 3 - *very good.*

4.4. Problem Analysis

4.4.1. Problems of State-of-the-Art Visual Discomfort Measures

In the last section 4.3, some shortcomings of state-of-the-art image features used for visual discomfort prediction, are observed. Before I suggest some possible improvements based on this analysis, let us discuss the discovered shortcomings, which can not be eliminated by using combinations of features.

Prediction Accuracy

Sohn et al. [38] analysed the prediction accuracy of some image features included in our consideration. Their results show that the prediction accuracy of measures, using combinations of first order statistics only, is limited. Particularly they show that the prediction accuracy of

such measures can be improved using the object dependent features OT and RD. Beneath the evidence of this shortcoming of first order stat. features, these experiments give hints about possible improvements which will be discussed in section 4.4.2.

Time Complexity

Although the combination of first order statistical features with object dependent features can improve prediction accuracy, the runtime complexity is substantially higher (see chapter 6). This is a big shortcoming since visual discomfort measures are often used to analyse stereoscopic contents in real time as argued in section 4.1.2. Note that the application of the Sobel-Operator used to compute $Range_{10}Sobel$ and Max_5Sobel induces also time intensive computations.

4.4.2. Possible Improvements of State-of-the-Art Visual Discomfort Measures

The analysis of the four visual discomfort measures presented in section 4.3 and their categorization presented in section 4.2 can be summarized by the following four facts.

Properties modelled by object dependent features can not be modelled using first order statistics since pixel neighbourhood is considered, which can be modelled using second order statistics.

The fact that object dependent features have high runtime complexity, points out that features should be based on a lower information representation. Higher order statistical features have a lower information representation as outlined.

The weighting of first order statistics by disparity differences shows an improvement of the correlation of first order features with visual discomfort [21]. This also hints towards the design of second order statistical features.

Kim et al. noted in the conclusion of [21] that the extension of their work by a disparity contrast model could lead to better prediction accuracy.

Based on this four facts I decided to try to improve state-of-the-art visual discomfort measures by the modelling of a disparity-based contrast feature, based on a second order statistic, which will be outlined in the next section.

5. Improving Visual Discomfort Prediction by Disparity-Based Contrast

In the following sections, a new model of disparity-contrast will be proposed. The approach, which is new in the field of visual discomfort prediction, is based on the introduction of the Haralick Contrast Feature *HC*.

The motivation of this approach consists of three parts: a motivation from the state-of-the-art analysis in section 4, a motivation from the field of texture analysis and a motivation from psychophysics.

After the motivation, the feature will be proposed and mathematical simplifications are performed, which give the possibility of a fast computation of the feature. The proposition of the feature is closed by experiments on the optimal parameter settings.

5.1. Motivation

The approach of this work is motivated by three research fields: psychophysics, visual discomfort prediction and texture analysis.

Motivation from Psychophysics

Features, modelling object dependent properties, are known to have a high correlation with visual discomfort when combined with lower level information as for example first order statistical features extracted from the disparity map. My approach relies on the hypothesis that bottom-up features in the human visual sensorial system are relevant for visual discomfort. A bottom-up process in psychophysics is characterized by an absence of higher level information in sensory processing such as the contextual knowledge [43]. A crucial bottom-up aspect turns out to refer to the so-called *conspicuity area*, which, with single eye fixation, captures the spatial region around a center of gaze, where the target can be resolved from its background [42]. The human visual target conspicuity is measured by a psychophysical procedure and has been analysed for a range of static targets in static scenes [7, 9]. The investigations show that the conspicuity area is small if the target (object of interest) is surrounded by high spatial variability. Therefore distinctness of image details is not only influenced by natural characteristics like the parallax distribution (see section 3.3) but also by artefacts in the disparity maps. I propose to model the distinctness of image details by a disparity-contrast model.

Motivation from the State-Of-The-Art Analysis

The state-of-the-art analysis of section 4.3, especially the summary in section 4.4.2 hints towards the model of disparity-contrast as a higher order statistical feature. From the characterization in section 4.2 we know that higher order statistical features can be seen as intermediate

state between first order statistical features and object dependent features concerning the level of information representation. Thus, higher order statistical features can be used to balance both, prediction accuracy and evaluation time.

For computational efficiency reasons, we ask for the group of the most simple higher order statistical features. These features are the second order statistical features which are features extracted from a joint probability distribution of pixels with respect to the values. I use the joint probability distribution of disparity-pairs with respect to the two disparities values. This probability distribution is known as grey level co-occurrence matrix, and it comes from the field of texture analysis.

Motivation from Texture Analysis

In literature one can find various concepts for characterizing contrast, see, e.g. [35, 15, 42, 33, 2]. Motivated by observations in 4.3, I ask for a contrast model which makes use of pixel neighbourhood. Because pixel neighbourhood can not be modelled using first order statistics of pixels intensities, this leads to the motivation of a second order statistical features. Particularly, the motivation comes from experiments on texture characterization. For example, Julesz [18] claimed in his famous experiments on human visual perception of textures that for a large class of textures "no texture pair can be discriminated if they agree in their second-order statistics". Even if counterexamples have been found to this conjecture, the importance of second order statistics is generally agreed to be certain. Thus, the major statistical method used in texture characterization is the one based on the joint probability distribution of pixel-pairs [27]. From this probability distribution second order statistical features can be extracted, known as *Haralick Features*. One of this features, *Haralick Contrast* (*HC*), models the contrast of a texture and it is widely used in in the field of texture characterization. From this research field it is known, that the *HC* feature has computational interesting properties including high robustness, good interpretability and fast computation time [27].

In the following section 5.2 the extraction of the *HC* feature from the disparity map of an stereoscopic image is proposed. Furthermore, experiments on an accurate parameter setting of the feature are performed as part 5.2.2 of the section.

5.2. Disparity-Based Contrast

In the last section the introduction of the *Haralick Contrast* feature *HC* was motivated. In this section the extraction of this feature from the disparity map is proposed. After some mathematical simplifications, the optimal parameter setting for the *HC* feature, when used for visual discomfort prediction, is discussed in section 5.2.2.

5.2.1. Modelling Approach

The joint probability distribution of pixel-pairs of an image is given by the second order histogram known as grey level co-occurrence matrix. It is defined as a function

$$h_{\delta,\Theta} \; : \; \{1,\ldots,n\} \times \{1,\ldots,n\} \to [0,1]$$

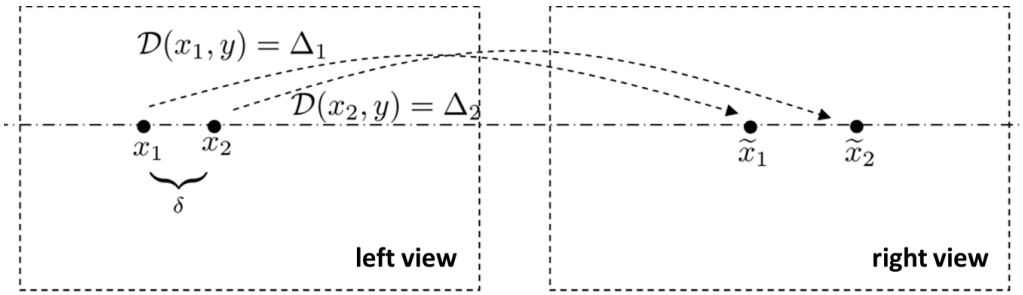


Figure 5.1. *Left and right view of a stereoscopic image,* (x_1, y) *and* (x_2, y) *are two pixels horizontally shifted by* δ*,* $(\tilde{x}_1, y)$ *and* $(\tilde{x}_2, y)$ *are the corresponding pixels in the right view,* $\Delta_1 = \mathcal{D}(x_1, y)$ *and* $\Delta_2 = \mathcal{D}(x_2, y)$ *encode the pixels disparities shown as horizontally shift between the left and right view of the image.*

where n denotes the number of grey-values. An entry, $h_{\delta,\Theta}(i, j)$ in the co-occurrence matrix represents the joint probability that a pair of pixels with in-between distance δ and direction Θ has the grey values i and j respectively. This yields a square matrix of dimension equal to the number of pixels grey values n in the image. Figure 5.3 gives an illustrating example of a stereoscopic image, the corresponding left disparity map, and the grey level co-occurrence matrix $h_{2,0^\circ}$ of the disparity map.

Now let us apply the co-occurrence matrix on a disparity map $\mathcal{D}$. Let us assume that the stereoscopic image was created under perfect conditions, i.e. all corresponding pixels appear horizontally in the two views of the image. Thus, the disparity value $\mathcal{D}(x, y)$ at the pixel position (x, y) encodes the distance Δ by which the pixel (x, y) in the left image has to be shifted horizontally in order to match the corresponding pixel in the right image. Figure 5.1 gives an illustrating example of these notations.

As a result, the grey level co-occurrence matrix $h_{\delta,0}$ with $\Theta := 0$ yields a special case of the Haralick Contrast Feature *HC*, [12], given by

$$HC_{\mathcal{D}}(\delta) = \sum_{\Delta_1, \Delta_2} (\Delta_1 - \Delta_2)^2 h_{\delta,0}(\Delta_1, \Delta_2) \tag{5.1}$$

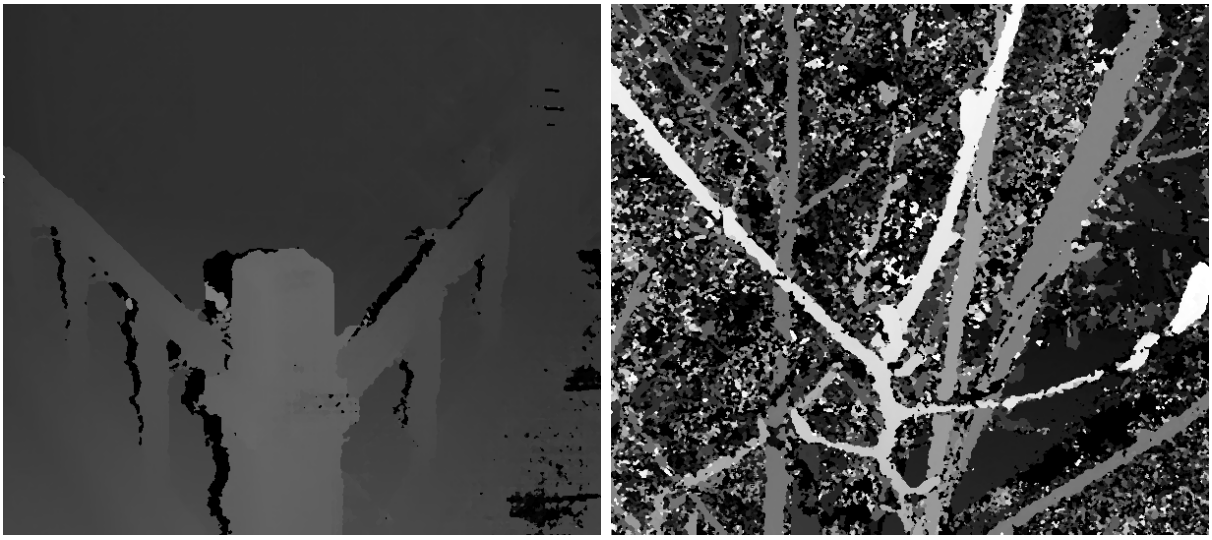
where Δ_1 and Δ_2 are the disparity values of the pixels (x_1, y) and (x_2, y) with distance $\|x_1 - x_2\|_2 = \delta$. The entry $h_{\delta,0}(\Delta_1, \Delta_2)$ of the co-occurrence matrix can be estimated by

$$h_{\delta,0}(\Delta_1, \Delta_2) = \frac{\#_\delta\{\Delta_1, \Delta_2\}}{\sum_{\Delta_x, \Delta_y \in} \#_\delta\{\Delta_x, \Delta_y\}} \tag{5.2}$$

where $\#_\delta\{\Delta_1, \Delta_2\}$ is the number of pixel pairs with values Δ_1 and Δ_2 and in-between distance δ. That is

$$\begin{aligned} &\#_\delta\{\Delta_1, \Delta_2\} = \\ &= \#\{((x_1, y), (x_2, y)) \in (\{1, \ldots, N\} \times \{1, \ldots, M\})^2 | \\ &\qquad \Delta_1 = \mathcal{D}(x_1, y) \wedge \Delta_2 = (x_2, y) \wedge \|x_1 - x_2\|_2 = \delta\} \end{aligned} \tag{5.3}$$

where N and M denote the number of the disparity rows and columns respectively and $\#S$ the cardinality of the set S. Note that using equations 5.2 and 5.3, the *HC* feature can be



(a) (b)

Figure 5.2. *Examples of disparity maps of stereoscopic images from [19] created by an algorithm of [1]. Figure (a), a detail of a railing, shows low HC and a low level of visual discomfort, while Figure (b), a crown of a tree, shows high HC and a high level of visual discomfort.*

simplified to

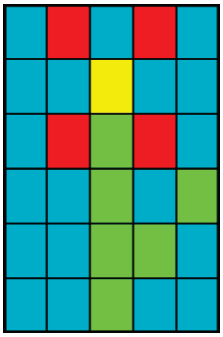
$$HC_{\mathcal{D}}(\delta) = \sum_{\Delta_1,\Delta_2} (\Delta_1 - \Delta_2)^2 h_{\delta,0}(\Delta_1, \Delta_2) = \\ = \frac{1}{NM} \sum_{x,y} (\mathcal{D}(x,y) - \mathcal{D}(x+\delta, y))^2 \tag{5.4}$$

Note that the *HC* feature is high, if the stereoscopic image has high spatial complexity, see e.g. figure 5.2. In the next section we will discuss the optimal parameter setting of the parameter δ.

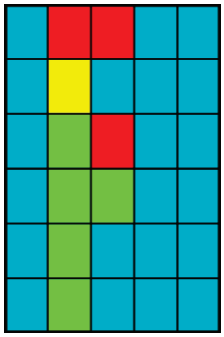
5.2.2. Parameter Setting

Although the *HC* feature could be simplified in the last section, the optimal choice of the parameter δ is still missing. In this section we will perform some experiments for an accurate setting of that parameter. The experiments are performed on two publicly available databases [19, 10]. The databases store subjective assessment data about visual discomfort and subjective image quality of stereoscopic images. As noted in chapter 4, this data is used for supervised learning to train and test visual discomfort measures. Apart from that, we will use this data to analyse the correlation of the *HC* feature with visual discomfort scores for different settings of the parameter δ.

First of all let us point out that the quality of the correlation between an image feature and visual discomfort can be measured using the Person Product-Moment Correlation Coefficient *CC* (see appendix A.2). The *CC*, for a set of images, can be calculated between the extracted feature values and the level of visual discomfort associated with the images. The level of visual discomfort is mostly experimentally determined using subjective assessment data, where the



(a)



(b)

0	0	-1	1	0
0	-1	1	1	0
0	-1	1	1	0
0	-1	1	-1	2
0	-1	-1	2	0
0	-1	1	0	0

(c)

i\j	*-1*	*0*	*1*	*2*
-1	*1*	*2*	*2*	*1*
0	*1*	*0*	*5*	*0*
1	*0*	*5*	*0*	*1*
2	*0*	*0*	*0*	*0*

(d)

Figure 5.3. *Stereoscopic image of a simplified flower: (a) left view (b) right view (c) left disparity map, where* -1 *identifies occluded pixels, (d) grey level co-occurrence matrix* $h_{2,0^\circ}$ *of the disparity map.*

mean score of the ratings of enough, i.e. more than 15, [17], subjects is stored. Such a mean score is also called *mean opinion score* (*MOS*).

In the two databases [19, 10] (referred to as *LausanneDB* and *KaistDB* respectively), the *MOS* of 54 respectively 120 stereoscopic images are stored. To extract the *HC* feature, I first computed disparity maps for all images. This was done by means of the *OpenCV* implementation, [1], of the semi-global block matching algorithm [14] (see e.g. figure 5.4).

To determine a reasonable choice for the parameter δ, I performed sensitivity analysis based on linear regression analysis between the *MOS* and *HC* extracted from the disparity maps of the images. Then, the *CC* between the images and the *MOS* were computed. The result of both databases is shown in figure 5.5.

The figure shows the correlation between the extracted *HC*s and the mean opinion scores of the images of the databases *KaistDB* (solid) and *LausanneDB* (dashed). The horizontal axis shows the value of the distance parameter δ in percent of image width. The vertical axis shows the *CC* between the extracted *HC* values and the *MOS* of the stereoscopic images.

The result shows a high correlation between the *HC* feature and visual discomfort after an accurate setting of the parameter δ. One can see that the correlation is lower on the *LausanneDB*. This could be due to the fact that the computed disparity maps, of the images of the database, are more noisy than the ones of *KaistDB*. Another reason could be that the questionnaire which yields the *MOS* of the *LausanneDB* are slightly different than for the *KaistDB*. Apart from that, the correlation on both databases is high, if the parameter δ is set in the interval $[5; 20]$. This shows that the feature is not very sensitive to parameter changes for different input data in that interval.

Thus, in all my experimental evaluations, I chose the parameter δ as 10 percent of image width.

The question arises, why the *CC* in this interval is higher than for other values, when analysed for both databases. This could be because of the smoothness of the disparity maps (smoothing operations are performed within the disparity algorithm, see [14]). Another reason could be that the human visual perception is more sensitive in discriminating stereoscopic image details in a certain range [31, 36]. Although it is very interesting to find possible answers to this question, this work is not intended to find new results in psychophysics, but the goal of this section is an accurate tuning of the δ parameter for predicting visual discomfort.

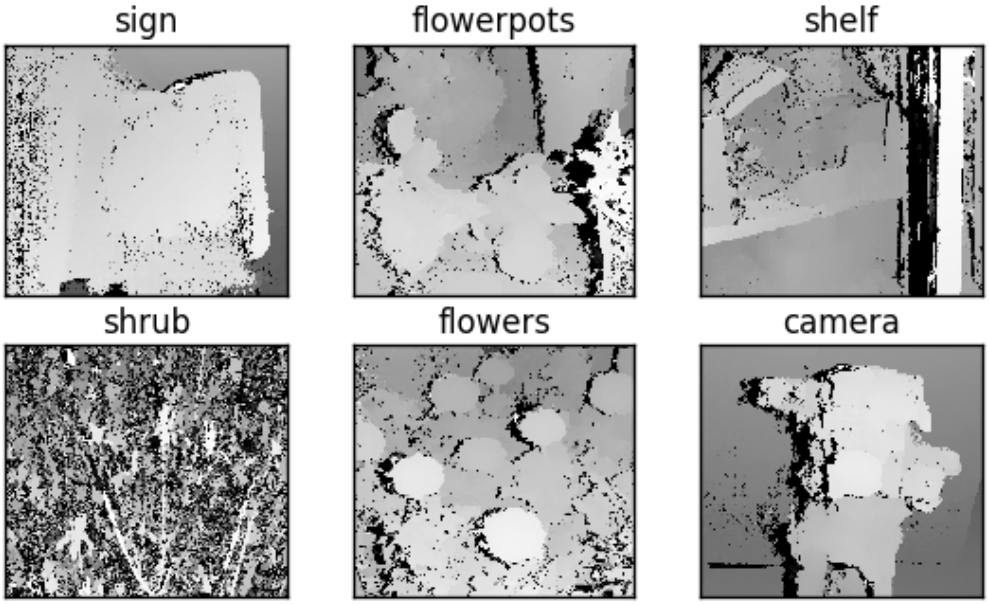
5.3. Summary

In this chapter I proposed the Haralick Contrast Feature as a model of disparity-based contrast. This was motivated from three fields namely psychophysics, visual discomfort prediction and texture analysis. I showed, that the feature can be simplified, which gives the possibility for fast evaluation of the feature. The chapter concludes with experiments on the optimal parameter setting of the feature, which shows low sensitivity against parameter changes. Another outcome of the experiments is a high correlation between the feature and visual discomfort. These facts are summarized in table 5.1, which concludes the summary of feature properties analysed in section 4.3. Note that the prediction accuracy and time complexity compared to other features will be discussed in more detail in chapter 6.

Properties	*HC*
Interpretability	3
Simplicity of Parameter Adaptation	3
Correlation withVisual Discomfort	3
Robustness w.r.t. Noise and Artefacts	3
Computational Efficiency	3

Table 5.1. *Conclusion of table 4.3. Desirable properties of image features used for visual discomfort measurement. Rating:* 1 - *neutral,* 2 - *good,* 3 - *very good.*

(a)



(b)

Figure 5.4. *Example of stereoscopic images and corresponding disparity maps, (a) left views, (b) disparity maps*

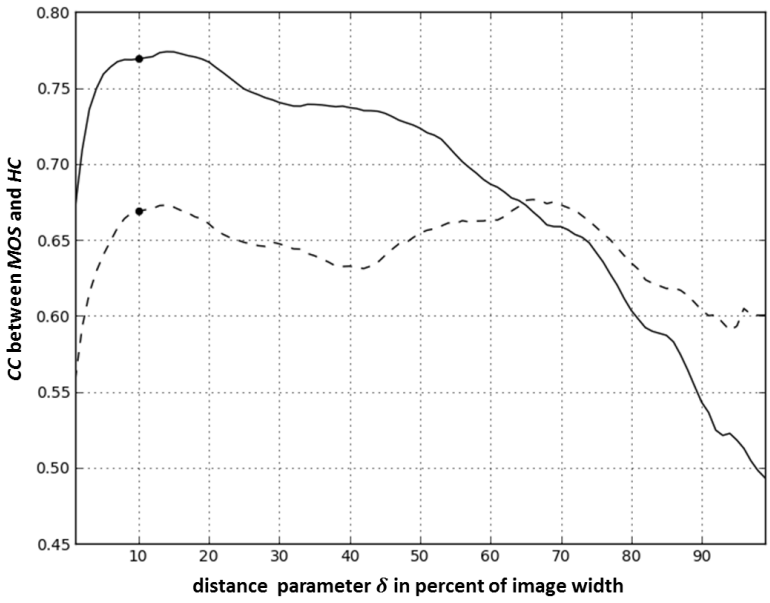


Figure 5.5. *Sensitivity analysis for the parameter* δ*. The graph shows the* CC *between the* MOS *and* HC *for the databases* KaistDB *(solid) and* LausanneDB *(dashed). The marked points indicate the corresponding correlation coefficients for* δ *equal to* 10% *of the image width.*

6. Experiments

This chapter is intended to achieve the goal of this work, which is given as a question. First this question is shortly reviewed and the answer is given as four claims. After that the experimental setup is discussed. The rest of this chapter consists of experiments which underpin the four claims. The chapter concludes with a summary about the observations of the experiments.

6.1. Goal of this Work

The goal of this work is given by a question: Is it possible to design a visual discomfort measure, which achieves the same prediction accuracy as state-of-the-art models, while it needs comparably lower runtime?

The answer to this question is Yes, and a measure having the desirable properties can be modelled using the *HC* feature proposed in section 5.2. More precisely formulated I will give the answer to this question and more, by four claims with respect to the approaches of [24, 30, 3, 21, 38] and the one of this work:

Claim 1 (Prediction Accuracy)
The expected prediction accuracy, which can be achieved by combinations including *HC*, is significantly higher than for combinations without *HC*.

Claim 2 (Feature Selection)
Taking all the features under consideration, inclusive *HC*, into account, a total number of four features is appropriate to predict visual discomfort.

Claim 3 (Time Complexity)
HC allows substantial time complexity improvement without significant loss of prediction accuracy, compared to state-of-the-art approaches under consideration.

Claim 4 (Further Improvement)
The prediction accuracy, which can be achieved by the best combination without *HC* can be improved by *HC*.

In the next sections, these claims are underpinned by statistical tests on two publicly available databases [19, 10]. The experimental analysis aims at providing evidence regarding the following aspects: prediction accuracy (section 6.3), feature selection (section 6.4) and runtime versus prediction accuracy (section 6.5). Concerning prediction accuracy, the potential of improving the overall prediction accuracy will be checked by taking various feature combinations into account. The section about feature selection is tackled by means of sensitivity analysis of the expected accuracy of visual discomfort prediction depending on combinations of features in order to determine the optimal choice of the number of features. Finally regarding run-

time versus prediction accuracy it will be looked at the Pareto fronts when taking prediction accuracy on the one hand, and computation time on the other hand into account.

Before that let me outline the experimental setup based on two publicly available databases.

6.2. Experimental Setup

The experimental analysis relies on two publicly available databases *KaistDB* [19] and *LausanneDB* [10]. Both databases store the information of subjective assessments following the guidelines of [17]. These guidelines, from the International Telecommunication Union (ITU), are about an optimal setup for the subjective assessments of visual quality. In both presented assessments, stereoscopic images of size 1080×1920 where shown by a stereoscopic display, three times the image width away from the rating subjects.

In the assessment of *KaistDB*, 19 subjects rated the level of visual discomfort associated with 120 stereoscopic images. They were asked to give a rating on a five point grading scale. The level of visual discomfort associated with a single image was then stored as *mean opinion score* (*MOS*) of all the 19 subjects.

LausanneDB consists of *MOS* values of 54 images resulting from the assessment of 17 subjects, where the subjective quality scores are obtained using an adapted single stimulus quality scale method. This method consists of rating stereoscopic images on a continuous scale by a single rating.

Note that, since the subjects were asked different questionnaires about a subjective feeling, which does not have overall accepted definition, results on both databases can be very different. Therefore, the same experimental setup can lead to different experimental results, which is a common phenomenon in objective video quality rating. This phenomenon is also strengthened by the fact that the only one definition of visual discomfort [25] was given four years before the beginning of this work.

The disparity maps used in the experiments are computed by the SCCH by means of the *OpenCV* implementation, [1], of the semi-global block matching algorithm [14]. Note that the analysis of disparity generation algorithms exceeds the intuition of this work and only the results are important. Examples of some disparity maps, displayed as normalized grey images, can be seen in figure 5.4.

To model the relation between features and visual discomfort, I employ *M5P* regression trees [37, 5] (see appendix A.3 for details of the algorithm). These machine learning functions combine a decision tree with linear regression functions in the leaves, which gives the possibility of compact and relatively comprehensible results. These functions are used because of two reasons. The first reason is that the resulting models are well interpretable and the second reason is about comparability. Since, Sohn et al. [38] also use *M5P* regression trees in their work, some of the results of this work can directly be compared with the results of [38].

To quantify prediction accuracy I rely on the Pearson Product-Moment Correlation Coefficient (*CC*), which is discussed in appendix A.2. The *CC*, evaluated between *MOS* of visual discomfort and the output of visual discomfort measures, gives a measure of the prediction accuracy.

Overfitting is a common phenomenon in computer vision. It occurs when a statistical model describes random error or noise instead of the underlying relationship. This happens, e.g. when a model is excessively complex, such as having too many parameters relative to the number of observations. This is the case especially if too much features are combined to a

function for predicting visual discomfort. To overcome this problem cross-validation can be performed, which is discussed in appendix A.4.

To check statistical significance, two statistical tests are performed. The non-parametric Mann-Whitney-U-Test [26] (MWU-test) and a one tailed F-test [8] (F-test). The MWU-test is used for testing the mean values of sets for being different with statistical significance level of 10^{-3}. The F-test is used for testing, if one sample has statistical significant different *CC* than a other.

To measure runtime, I rely on the *OpenCV* implementations [1] of the features. Since, for some features, there does not exist publicly available implementations, I implemented these features in *Python* [44] by using as much runtime optimized functions of [1] as possible. The runtime was then measured by the mean runtime of several computations using all images from both databases. For the object-based approach of [38] only the most time consuming part was considered, which is identified to be the pyramid-based mean-shift segmentation algorithm [4]. I repeated the evaluations 10 times on the 174 images of the two databases (*KaistDB*, *LausanneDB*) on a DELL OptiPlex 990 (the images were resized to 540×960). This results in a mean runtime of 1740 evaluations per feature.

The next section is devoted to the question whether the *HC* feature can be used to improve the prediction accuracy of state-of-the-art visual discomfort measures with respect to the approaches of [24, 30, 3, 21, 38], summarized in table 4.2. The result will be claim 1.

6.3. Prediction Accuracy

In the following it will be analysed, if the *HC* feature can be used to improve prediction accuracy of visual discomfort measures using combinations of features described in section 4.3. The natural question arises, if the most simple visual discomfort measures, which make use of single features only, can be improved by adding the Haralick Contrast Feature *HC* to the single feature. The first experiment is devoted to this question. It is analysed, whether a combination of the Haralick Feature *HC* with a single feature, proposed in [24, 30, 3, 21, 38], shows higher prediction accuracy than the visual discomfort measure based on the single feature and without *HC*. Note that it is not always true that a visual discomfort measure based on two features shows higher prediction accuracy than the measure based on one of these features only, since overfitting might occur.

Table 6.1 shows the result, which can be summarized by the following claim with respect to the features proposed in section 4.3.

Claim 5 (Single Feature Prediction Accuracy)
The prediction accuracy of visual discomfort measures based on a single feature can substantially be improved by adding the Haralick Contrast Feature *HC*.

The measures consist of a feature extraction part and a *M5P* regression tree (see appendix A.3) as machine learning function. To overcome the problem of overfitting, leave-one-out cross-validation (see appendix A.4) was performed, which results in one prediction score for every image in the database. The prediction accuracy of the visual discomfort measures was then measured by *CC* between the prediction results for all images of the databases and the *MOS* of the subjective ratings. Figure 6.1 illustrates this experimental approach.

The resulting *CC*s in table 6.1 show the prediction accuracy of the visual discomfort measure

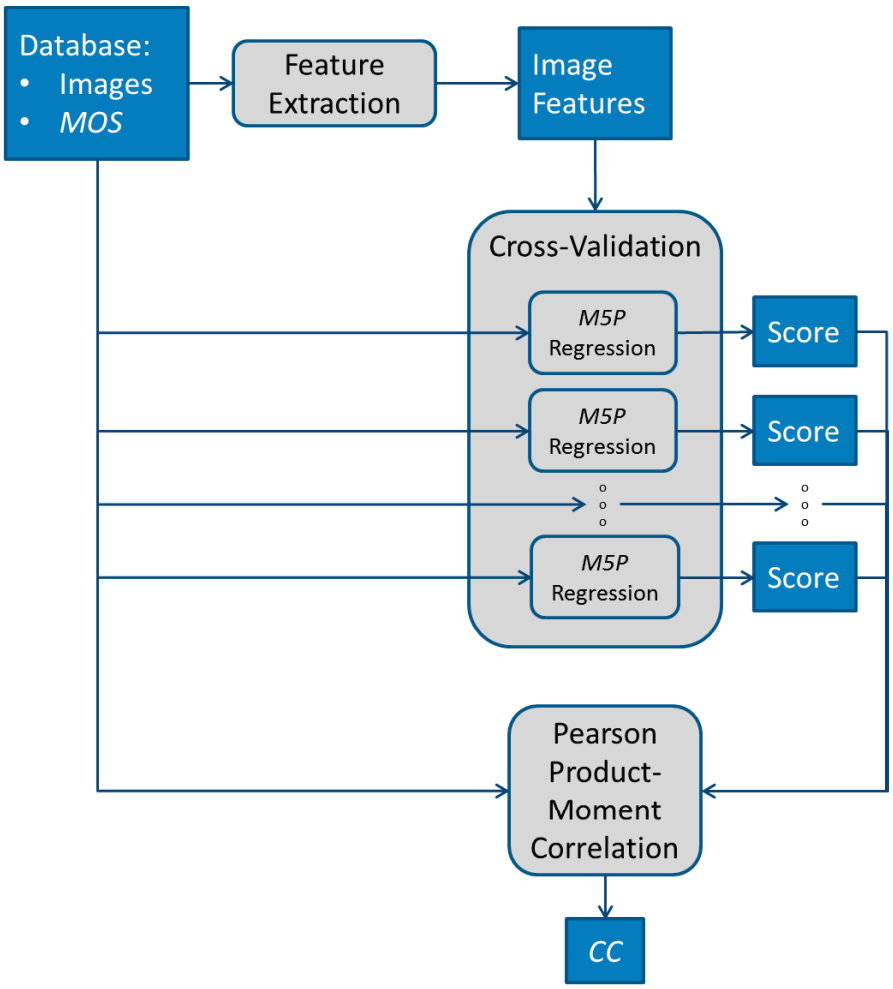


Figure 6.1. *Experimental approach*

based on different features and the combination with *HC*.

The first column shows the *CC*s of visual discomfort measures based on a single feature only, evaluated on the *KaistDB*. One can see that the prediction accuracy of measures, based on the features *RD* and *OT*, is very low when used as single features only. The other features show a *CC* over 74%, which is not low for measures predicting subjective feelings. The highest *CC* is derived by a measure based on the $Range_{10}Sobel$ feature, which is printed in bolt.

The second column shows the *CC*s of visual discomfort measures, based on a single feature only, evaluated on the *LausanneDB*. On that database, all except two measures show low prediction accuracy, i.e. $CC \leq 60\%$. These two features are the variance of disparity Var and the *HC* feature. Note that the measure based on the *HC* feature shows the highest prediction accuracy, which is another small result of this work. For the reason of the lower prediction accuracy of visual discomfort measures on the *LausanneDB*, I refer to section 5.2.2.

The computation of the *CC*s of single features confirm the property analysis of the single features in section 4.3 and tables 4.3 and 5.1.

The cells of the last two columns of the table show the prediction accuracy of a visual discomfort measure based on the feature, to which the row corresponds to, and the *HC* feature. The columns show that the combined use of the *HC* feature with one of the other features

always improves prediction accuracy compared to the measures based on the single features only. The improvement in *CC* ranges from 0.7% to 50.1%, which underpins claim 5.

Note that the results in this table are consistent with the tables in the work of Sohn et al. [38], i.e. the first column of table 6.1 shows very similar *CC*s as the tables of [38]. This indicates that there are no errors in the implementation of the features, as well as in the machine learning part (*M5P* regression), the cross-validation part and the *CC* computations.

Single Feature	**CC of Single Feature (*KaistDB*)**	**CC of Single Feature (*LausanneDB*)**	**CC of Single Feature + HC (*KaistDB*)**	**CC of Single Feature + HC (*LausanneDB*)**
RD	0.2465	0.3539	0.7478 (+ **0.5013**)	0.6972 (+ **0.3432**)
OT	0.3807	0.1024	0.7627 (+ **0.3820**)	0.7019 (+ **0.5994**)
Mean	0.2995	0.5822	0.7603 (+ **0.4608**)	0.7243 (+ **0.1420**)
Var	0.7449	0.6901	0.7698 (+ **0.0249**)	0.6972 (+ **0.0070**)
Range_{10}	0.7806	0.5248	0.7939 (+ **0.0133**)	0.7244 (+ **0.1997**)
Max_5	0.7707	0.4746	0.8047 (+ **0.0340**)	0.7312 (+ **0.2566**)
Max_5Sobel	0.7761	0.5089	0.8036 (+ **0.0275**)	0.7327 (+ **0.2238**)
Range_{10}Sobel	0.7841	0.5437	0.7933 (+ **0.0092**)	0.7168 (+ **0.1732**)
HC	0.7593	0.6972	0.7593 (+ 0.0000)	0.6972 (+ 0.0000)

Table 6.1. *Prediction performance of single features and the corresponding combination with HC, based on the ground truth data given by the databases* KaistDB *[19] and* LausanneDB *[10]. Improvements are marked in bolt.*

The next experiment is devoted to the question whether the *HC* feature can be used to improve prediction accuracy of visual discomfort measures using different combinations of features with potentially more than one feature. The 8 features under consideration (see table 4.2), together with the proposed *HC* feature, yield in total $\sum_{k=1}^{9} \binom{9}{k} = 911$ possible feature combinations, where $\binom{x}{y}$ denotes the binomial coefficient "x choose y". For all of these 911 combinations a *M5P* regression tree was trained using leave-one-out cross validation. This yield two correlation coefficients for each combination, one for each database. After that the results for all combinations without *HC* were combined into one set. The remaining combinations, i.e. combinations including *HC*, wre also combined into one set. Then the mean *CC* for both sets of combinations was computed. The result for *KaistDB* (*LausanneDB*) shows, that the mean *CC* 0.81 (0.79) of the set of combinations including *HC* is higher than the mean *CC* 0.79 (0.61) of the set of combinations without *HC*. The MWU-test shows the statistical significance of the results (KaistDB: $U = 25823.0, p\text{-value} = 4.419 \cdot 10^{-5}$, LausanneDB: $U = 6124.0, p\text{-value} = 2.8 \cdot 10^{-70}$), which underpins claim 1.

Again note that an improvement of prediction accuracy is not always achieved by adding features to the visual discomfort measures. Moreover, the prediction accuracy can become worse by this process, which will be discussed in the next section. Also note that a statistically significant difference of two values (e.g. mean *CC*) is a much stronger result than simply a difference (see appendix A.5).

6.4. Feature Selection

In the following experiments the impact of the number of image features, of a visual discomfort measure, is analysed. I often noticed the important fact that a high parameter complexity of visual discomfort measures, i.e. too many parameters relative to the number of training samples, can lead to a decrease of prediction performance, compared to the same measure with a lower number of parameters. This effect is known as *overfitting* and its impact on the prediction accuracy increases with the number of features. This is because the number of parameters of the machine learning function (e.g. *M5P regression tree*) increases with the number of features, which have to be combined. To restrict the number of features is therefore a regularization measure that helps to improve the behaviour of the machine learning model [45].

Therefore we analyse the prediction accuracy measured in *CC* of the best combination for every number of features, i.e. the combination with the highest *CC* for every feature number. Figure 6.2 shows the result.

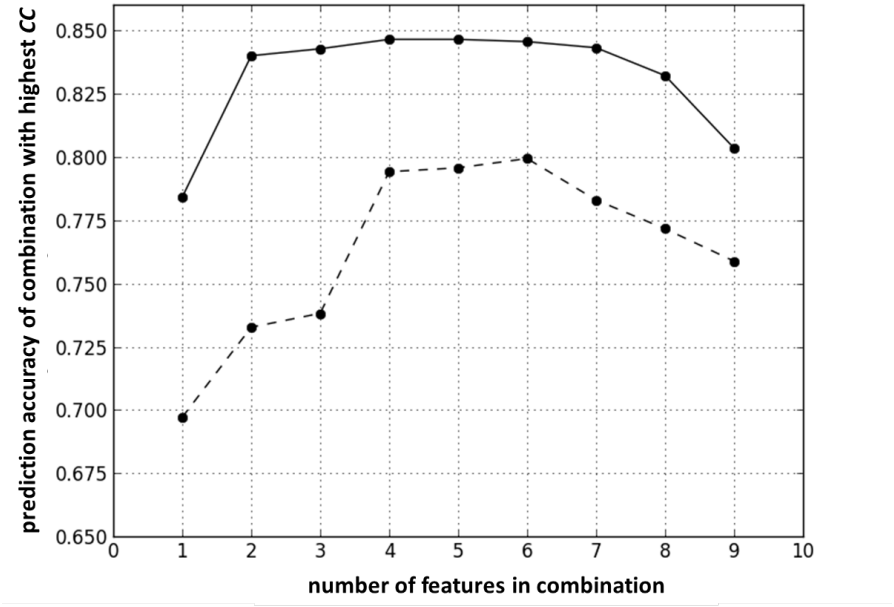


Figure 6.2. *Analysis of prediction accuracy, measured by* CC*, of the best (i.e. highest* CC*) combination with feature number* $n \in \{1, \ldots, 9\}$ *based on ground truth data from* KaisDB *(solid), [19], and* LausanneDB *(dashed), [10].*

The experimental results show that, taking the nine features under consideration into account, a feature number of four is appropriate to predict visual discomfort. Although the experiment on the *LausanneDB* shows two combinations with a higher feature number having higher *CC*, a statistical F-test indicates statistical equivalence (F-tests: $Z \leq 0.072$,p-value ≥ 0.47). These results can be summarized by claim 2.

6.5. Pareto-Front

The goal of this work was given as a question and it is about the balance of prediction accuracy and runtime of visual discomfort measures. Thus, I'm interested in feature combinations that

are characterized by the property that *it is impossible to reduce runtime, by exchanging image features, without deteriorating prediction accuracy, and vice versa.* Combinations with this property are called *Pareto-optimal* solutions and they lie on the so-called *Pareto front*, see e.g. [6]. Thus, I compute the Pareto front, with respect to prediction accuracy and runtime, of all combinations with at most four features.

The Pareto front obtained, using ground truth data from *KaistDB*, is shown in figure 6.3a. Every point in the figure corresponds to a combination with at most four features, where the corresponding runtime (*CC*) is shown by the horizontal (vertical) axis. Three clusters can be identified: combinations without object dependent features, *OT* and *RD*, having evaluation runtime lower than one second, combinations with one object dependent feature (evaluation time between three and four seconds) and combinations including *OT* and *RD*. The figure shows that combinations including the two object dependent features don't have higher prediction accuracy than other combinations. It also shows some combinations, with the *OT* feature and without the *RD* feature, having higher prediction accuracy than combinations without that feature. This fact is the main result of Sohn et al.'s work [38], when not considering results of the *HC* feature. The dashed rectangles in figure 6.3a indicate figures 6.3b and 6.3c, which scale up the enclosed parts in figure 6.3a.

Combinations in figures 6.3b and 6.3c, marked by '×' indicate combinations including the *HC* feature. Note that all combinations in figure 6.3b are without object-dependent features *OT* and *RD*. This figure also shows that many combinations on the Pareto front are with the *HC* feature including the one with the highest prediction accuracy. This phenomenon can also be seen in figure 6.3c, which shows the Pareto front including combinations with the highest prediction accuracy.

Before discussing the combinations, lying on the Pareto front in detail let us consider the results of the same experiment performed on the publicly available data of *LausanneDB*.

The Pareto front obtained using ground truth data from *LausanneDB* is shown in figure 6.4a. This figure shows that combinations with object dependent features don't have higher prediction accuracy than other combinations. The dashed rectangle in figure 6.3a indicates figure 6.3b, which scales up the enclosed part, including the Pareto front. Since the Pareto front of this experiment consists of combinations with runtime lower than one second, all the combinations lying on the Pareto front are without object dependent features.

Combinations in figures 6.4b, marked by '×', again indicate combinations including the *HC* feature. Note that all except two combinations, lying on the Pareto front, are with the *HC* feature.

Now let us consider the Pareto front of the experiments on *KaistDB* and *LausanneDB*, in detail. Tables 6.2 and 6.3 show the pareto optimal solutions of both experiments. Note that in the *KaistDB* experiment, the *HC* feature is part of over 50% of the combinations and, in the *LausanneDB*, the *HC* feature is part of all except two combinations. In both experiments the *HC* feature is part of the combination with the highest *CC*, which underpins claim 4.

The tables also indicate that the *HC* feature can not only be used to improve prediction accuracy, it also gives the opportunity to achieve substantial time complexity improvement. For example, let us consider the combination consisting of Var, Max_5Sobel, *HC* and $Mean$. This combination has a more than 58 times lower runtime than combinations with the object-dependent features proposed by Sohn et al. in [38]. In addition to that, there does not exist any combination out of 511 without *HC*, which shows a statistical improvement of prediction accuracy (F-test, KaistDB: $Z \in [0.0005; 0.6328], p\text{-value} \in [0.2634; 0.4998]$, LausanneDB: no better combination without *HC*). Note that this result holds for both databases *KaistDB* and *LausanneDB*. This indicates that the Haralick Contrast feature, extracted from the dispar-

ity map, with an appropriate parameter setting of δ, offers a reasonable trade-off between prediction accuracy and time complexity. This underpins claim 3 and gives the answer to the question, whether it is possible to design a visual discomfort measure, which achieves the same prediction accuracy as state-of-the-art models, while it needs comparably lower runtime. Thus, all arising problems of this work are solved and the goal of this work is achieved.

Combination on Pareto front	**Evaluation Time in Sec.**	**CC**
Mean	0.0004	0.300
Var	0.002	0.745
HC	0.006	0.759
HC, Mean	0.007	0.760
Var, HC	0.008	0.770
Max_5	0.016	0.771
Range_{10}	0.016	0.781
Max_5, Mean	0.017	0.796
Var, Max_5, Mean	0.018	0.803
Max_5, HC	0.022	0.805
Max_5, HC, Mean	0.023	0.814
Var, Max_5, HC, Mean	0.024	0.816
Max_5Sobel, HC, Mean	0.057	0.817
Var, Max_5Sobel, HC, Mean	0.059	0.817
Var, OT, HC, Mean	3.425	0.824
Max_5, OT	3.432	0.838
Max_5, OT, Mean	3.433	0.839
Var, Max_5, OT	3.434	0.842
Var, OT, Max_5, HC	3.440	0.846
OT, Range_{10}, Max_5, HC	3.455	0.846

Table 6.2. *Pareto optimal solutions of combinations with at most four features of experiment on KaistDB [19]*

Combination on Pareto front	**Evaluation Time in Sec.**	**CC**
Mean	0.0004	0.582
Var	0.002	0.690
HC	0.006	0.697
HC, Mean	0.007	0.724
Max5, HC	0.022	0.731
Var, Max_5, HC	0.024	0.733
Range_{10}, Max_5, HC, Mean	0.039	0.778
Var, Range_{10}, Max_5, HC	0.040	0.787
Range_{10}Sobel, Max_5, HC, Mean	0.072	0.794

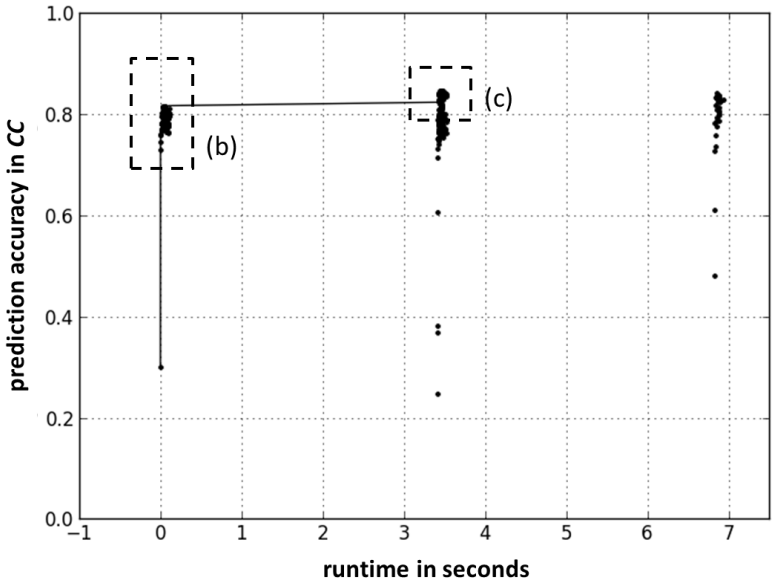
Table 6.3. *Pareto optimal solutions of combinations with at most four features of experiment on LausanneDB [10]*

6.6. Summary

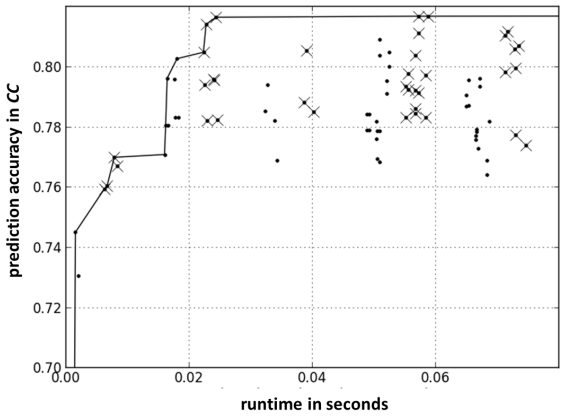
In this section experiments are done to underpin the claims given in the introduction. The experiments are performed concerning three aspects: prediction accuracy, feature selection and prediction accuracy versus runtime. The results show that the Haralick Contrast feature *HC*,

which is firstly considered by this work in the field of visual discomfort prediction, offers a reasonable trade-off between prediction accuracy and time complexity. Moreover it can be used to improve the prediction accuracy of state-of-the-art visual discomfort measures with respect to the approaches of [24, 30, 3, 21, 38].

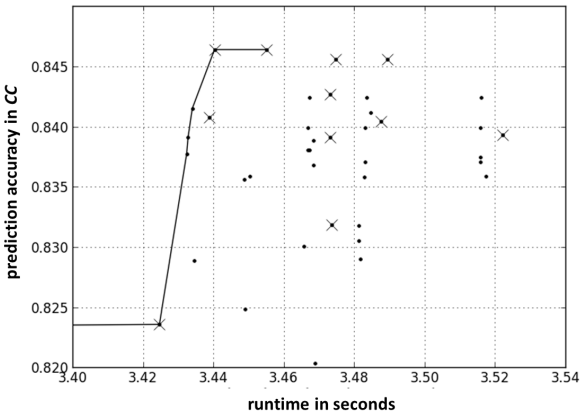
The goal of this work, described in section 2, was achieved by experiments on two publicly available databases [19, 10].



(a)

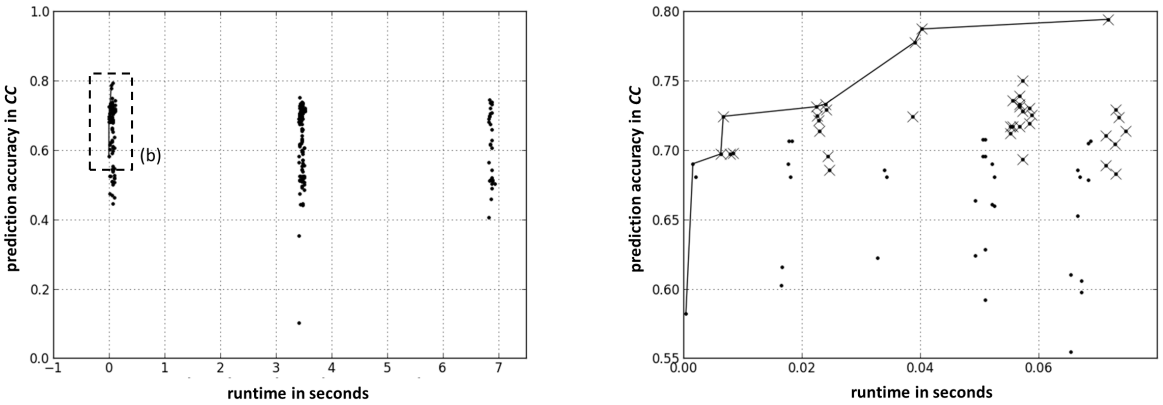


(b)



(c)

Figure 6.3. *Pareto front (solid line) of combinations with at most four features, considering prediction accuracy and runtime, based on ground truth data of* KaistDB *[19]. (b) and (c) scale up the details of (a) indicated by the dashed rectangles in (a). Combinations in (b) and (c) marked by '×' indicate combinations including the* HC *feature.*



(a) (b)

Figure 6.4. *Pareto front (solid line) of combinations with at most four features, considering prediction accuracy and runtime based on ground truth data of* LausanneDB *[10]. (b) scales up the details of (a) indicated by the dashed rectangle. Combinations in (b) marked by '×' indicate combinations including the* HC *feature.*

7. Conclusion and Outlook

In this work I addressed state-of-the-art computational models for predicting visual discomfort. Starting with an accurate characterization and analysis of state-of-the-art visual discomfort measures, a second order statistical feature was proposed. Derived from the grey-level co-occurrence matrix, commonly used in texture analysis, I came up with a computational efficient contrast feature based on the disparity map, the Haralick disparity contrast. Finally experimental analysis showed that this feature allows to improve prediction accuracy and runtime of state-of-the-art visual discomfort measures, when combined with other features. It remains future research to integrate the model into an automated re-rendering software for stereoscopic videos to minimize the effect of visual discomfort.

A. Computer Vision Preliminaries

A.1. Sobel Operator

The Sobel-Operator of an image uses two 3×3 kernels, which are convolved with the image to calculate approximations of derivatives. For a more formal definition, we first need to define the commonly used two dimensional convolution.

Definition 2 (2D Discrete Convolution)
Let f and h be two images with size (N_f, M_f) and (N_h, M_h) respectively. Then, the two-dimensional convolution $g(x, y) = f * h$ of the images is given by

$$g(i,j) = \sum_{n=0}^{N_f-1} \sum_{m=0}^{M_f-1} f(n,m)h(i-n, j-m)$$

where $0 \leq i \leq N_f + N_h - 1$ and $0 \leq j \leq M_f + M_h - 1$.

Now, let us define the *horizontal* and *vertical derivative approximations*.

Definition 3 (Horizontal and Vertical Derivative Approximation)
Let f be an image, then, the *horizontal* and *vertical derivative approximations* G_x and G_y of the image f are defined by

$$G_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} * f \qquad G_y = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} * f$$

Using that, one can define the Sobel-Operator by

Definition 4 (Sobel-Operator)
Let f be an image with size (N_f, M_f) and G_x, G_y be the *horizontal* respectively *vertical* derivative approximations of f. Then, the Sobel-Operator $S(f)$ of the image f is defined by

$$S(f)(i,j) = \sqrt{G_x(i,j)^2 + G_y(i,j)^2}$$

where $0 \leq i \leq N_f$ and $0 \leq j \leq M_f$.

Many image processing applications make use of the Sobel-Operator, like for example edge detection algorithms. For a typical output see figure A.1.

(a) (b)

Figure A.1. *Example of the application of the Sobel-Operator. (a) Original image, (b) Image after the Application of the Sobel-Operator.*

A.2. Pearson Product-Moment Correlation Coefficient

In statistics, the *Pearson Product-Moment Correlation Coefficient* (CC), is a measure of linear correlation between two populations. It gives a value between -1 and 1, where 1 is total positive correlation, 0 is no correlation and -1 is total negative correlation. It is defined by

Definition 5 (Pearson Product-Moment Correlation Coefficient)
Let $X = \{x_0, \ldots, x_N\}$ and $Y = \{y_0, \ldots, y_N\}$ be two populations, i.e. sets of samples x_i and y_i of the same cardinality $N+1$, and let μ_X and μ_Y be the mean values of X and Y respectively. Then, the Pearson Product-Moment Correlation Coefficient $CC(X,Y)$ of X and Y is defined by

$$CC(X,Y) = \frac{\sum_{i=0}^{N}(x_i - \mu_X)(y_i - \mu_Y)}{\sqrt{\sum_{i=0}^{N}(x_i - \mu_X)^2}\sqrt{\sum_{i=0}^{N}(y_i - \mu_Y)^2}}$$

Properties

The Pearson Product-Moment Correlation Coefficient $CC(X,Y)$ has the following properties

1. Strongest Correlation and Weakest Correlation:
   Correlations equal to -1 and 1 represent samples, lying on a line.

2. Symmetry:
   The Pearson Product-Moment Correlation Coefficient is symmetric, i.e. $CC(X,Y) = CC(Y,X)$.

3. Invariants To Translation and Scaling:
   The CC is invariant to translation and scaling of the two variables, i.e. if X and Y are two sets and $a, b, c, d \in \mathbb{R}$, $b, d \geq 0$, then $CC(X, Y) = CC(a + bX, c + dY)$.

Geometric Interpretation

Let $X = \{x_0, \ldots, x_N\}$ and $Y = \{y_0, \ldots, y_N\}$ two sets of populations, i.e. sets of samples x_i and y_i, of the same cardinality $N + 1$, and let μ_X and μ_Y be the mean values of X and Y respectively, then

$$\begin{aligned} CC(X,Y) &= \frac{\sum_{i=0}^{N}(x_i - \mu_X)(y_i - \mu_Y)}{\sqrt{\sum_{i=0}^{N}(x_i - \mu_X)^2}\sqrt{\sum_{i=0}^{N-1}(y_i - \mu_Y)^2}} = \\ &= \frac{\left\langle X - \mu_X \cdot \vec{1}, Y - \mu_Y \cdot \vec{1} \right\rangle}{\left\| X - \mu_X \vec{1} \right\|_2 \cdot \left\| Y - \mu_Y \vec{1} \right\|_2} \end{aligned} \tag{A.1}$$

with the $(N+1)$-dimensional vector $\vec{1}$, where all entries are ones. Let l_1 be the linear regression line of the set $X - \mu_X$, such that the mean squared error $\sqrt{\sum_{i=0}^{N}(x_i - l_1(x_i))^2}$ is minimized. Let l_2 be the regression line of $Y - \mu_Y$ then, by using equation A.1, we get

$$CC(X, Y) = \cos(l_1, l_2)$$

where cos refers to the cosine of the enclosed angle between l_1 and l_2. Thus, one can interpret the Pearson Product-Moment Correlation Coefficient CC as the cosine of the enclosed angle between the two regression lines, minimizing the mean squared error of the normalized sets.

A.3. M5P Regression Trees

In this section the *M5P regression algorithm* [37] is described. Regression trees are supervised machine learning models, which combine conventional decision trees with the possibility of regression functions in the leafs. This gives the possibility to divide the input values range into smaller categories, for which separate regression models are built. The resulting prediction model is a global non-continuous function for all input values.

For example let us consider multi-dimensional linear regression. The result of this supervised learning technique is a global continuous linear function, holding over the entire data-space. Thus, for data with lots of features interacting with each other in non-linear ways, assembling the linear regression function can be very difficult and the interpretation of the result can be very vague. To overcome these shortcomings the feature space can be sub-divided into smaller regions where the interactions have better interpretability. If the interactions of multiple features still remain vague one can again sub-divide this regions, i.e. she can perform *recursive partitioning*. This process can be repeated until the regions contain values, which show "clear" linear relations (the meaning of "clear" will be discussed later). Thus, the result contains a simple linear regression function for every region, which gives a global, well interpretable, and piece-wise linear regression model.
Let us consider a learning problem occurred due to an experiment of this work.

Given: *MOS* values for the 120 stereoscopic images of the *KaistDB* [19], a six-dimensional feature vector for every image consisting of the extracted features $Mean$, Var, Max_5, $Range_{10}$, $Range_{10}Sobel$ and HC from the corresponding disparity maps of the images

Goal: Regression tree function

A possible output of the *M5P* regression algorithm is displayed in figure A.2. The result splits the feature space into three parts: one part where the HC feature shows a value greater than 0.49, another part where HC is smaller or equal to 0.49, and Var is smaller or equal to 0.135 and the rest of the space. For each of the three regions, there exist a multi-dimensional linear regression function, for example for the part, where $HC(D_I) > 0.49$, the function has the form

$$\begin{aligned} VDC(I) = &- 0.1134 \cdot Mean(D_I) - 0.3861 \cdot Var(D_I) + 1.887 \cdot Range_{10}(D_I) \\ &- 0.8 \cdot Max_5(D_I) - 1.108 \cdot HC(D_I) + 1.954 \cdot Range_{10}Sobel(D_I) + 0.811 \end{aligned}$$

and it gives a visual discomfort value $VDC(I)$ for every image I. Note that the $Range_{10}Sobel$ feature is only used in this leaf of the tree. This indicates, that the use of the $Range_{10}Sobel$ feature, in this example, only makes sense, if the HC is high (> 0.49). Note that such interpretations are, in that way, not possible in multi-dimensional linear regression.

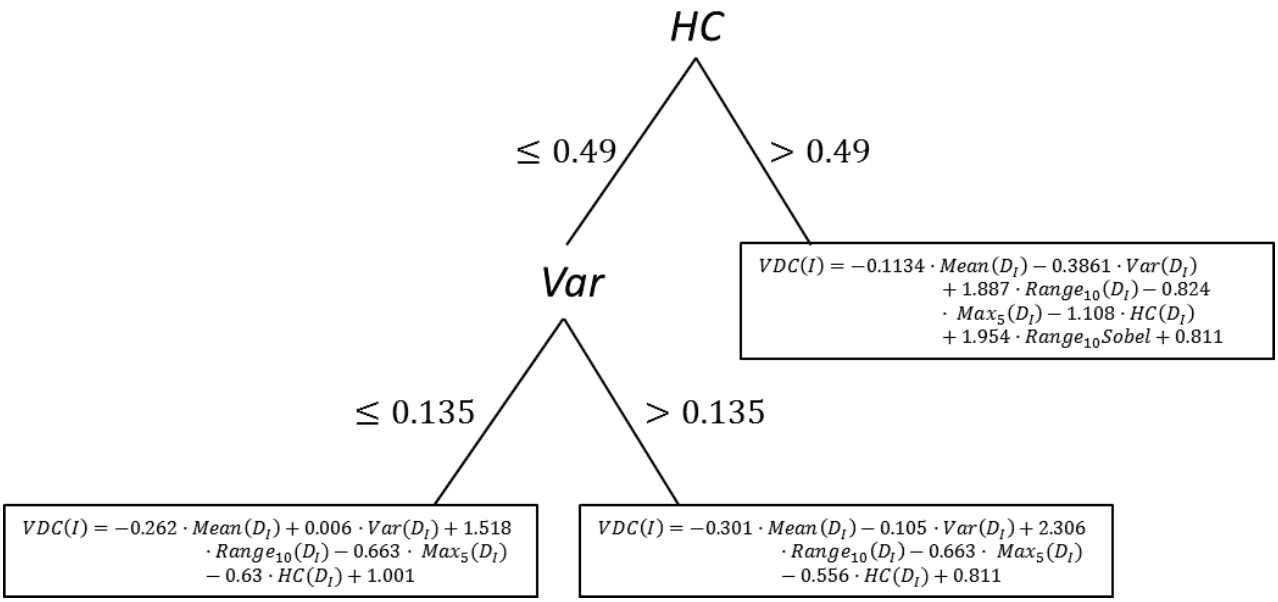


Figure A.2. *Example of a* M5P *regression tree, predicting visual discomfort* $VDC(I)$ *of an image* I *based on features extracted from the corresponding disparity map* D_I*.*

M5P Regression Tree Algorithm

Since this work is not intended to give the full theory of regression trees, only the major steps of the $M5P$ regression algorithm are proposed, as in [37].

Let $T = \{x_0, \ldots, x_N\}$ be a set of training samples with $x_i = \{f_{i,0}, \ldots, f_{i,n}\}$, feature values $f_{i,j}$ and $Y = \{y_0, \ldots, y_N\}$ a set of objective values y_i. Then the first step of the algorithm is

to compute the standard deviation sd_X of X.

$$sd_X = \sqrt{\sum_{i=0}^{N} (f_i - \mu_X)^2}$$

with the mean value μ_X of X. Unless the set X contains very few samples or the values vary only slightly, X and respectively Y are split according to the outcome of a test. Let X_j be the determined j-th subset of X, which is defined by the i-th outcome of the test with M possible outcomes. Thus, for every possible outcome of the test, a subset $X_j \subset X$ is defined. Now one can treat the standard deviation sd_{Xj} of X_j, as a measure of error. Then the expected error-reduction of the outcome X_j can be formulated as

$$\Delta error = sd_X - \sum_{i}^{M} \frac{\#X_j}{\#X} sd_{Xj}$$

with the cardinality $\#X_j$ of the set X_j. After that, the performed splitting strategy is chosen by the test, which maximises the $\Delta error$. This splitting strategy is then performed recursively on the outcomes $X_0, \ldots, X_M$ of the chosen test.

Since this process often causes over-elaborated structures, the sub-trees have to be pruned back, e.g. by replacing a sub-tree by leaf. The major innovation comes into play, after the tree has been built. A detailed formulation of these steps is precluded by this work, but the main ideas are as follows:

- **Error Estimation:**
  The algorithm often has to compute the accuracy of a model on unseen variable values x_i, i.e. when pruning some sub-trees back. A *residual* of a model, for a special variable value x_i, is defined as the difference between the objective value y_i of x_i and the predicted value $\tilde{x_i}$, of the model for the value x_i. The error of a model, for a set of variable values X_j, is first estimated by the average residual of all values $x_i \in X_j$. Since, this often over-determines the error, this value is multiplied by $\frac{\#X_j+\nu}{\#X_j-\nu}$, where ν gives the number of parameters of the model. This multiplied factor increases the error for models with many parameters when applied to small sets X_j of variable values x_i.

- **Linear Models:**
  The algorithm builds linear regression models for the leafs, in every iteration, by standard linear regression, minimizing the mean squared error. In addition the algorithm restricts the linear regression models to variables that are referenced by test or linear regression models somewhere in the sub-tree at the node [37]. This has the effect that, when comparing the accuracy of a sub-tree model with a linear regression model, these two types of models have the same parameters.

- **Pruning:**
  The algorithm estimates the above-mentioned error for every sub-tree and compares it with the simplified above-mentioned linear model. If the simplified linear model has the smaller error, the sub-tree is pruned to a leave.

These three major parts of the algorithm make it very interesting for a high number of features and a comparably low number of training samples, as it is the case in this work.

A.4. Cross-Validation

Given a dataset $X = \{x_0, \ldots, x_n\}$, the goal of cross-validation is to find a partition of the set X in two sets X_T, X_V with $X_T \cap X_V = X$, in order to *train* a model f with unknown parameters on the *training set* X_T and to *test* the model on the *test set* X_V. The training of the model is performed by parameter estimation using the training set X_T. Once, the parameters are estimated, the model is then tested using the test set X_V in order to validate the results of the training phase.

This process is used to estimate how well the model will generalize to an independent data set, for example in practice. Thereby the training phase is used to optimize the model parameters as well as possible. Then the test set X_V is used to validate the results of the test phase using quality measures based on this set. A typical result of the test phase is that the model does not fit to the test set X_V as well as to the training set X_T. This can happen when the model parameters are over-estimated in the training phase, which is called *over-fitting*. Over-fitting is particularly likely to happen in practice when the size of practical usable training data is small or when the number of model parameters is large. Therefore the test phase of the cross-validation process can also be seen as a test on how likely the model will generalize to unseen data when it is trained using the parameter estimation algorithm of the training phase.

Normally cross-validation is performed many times for different partitions of the data set X and the results of the validation phases are aggregated to an overall error measure. This error measure can then be used to rate how well the model and the parameter estimation process will generalize in practice.

A special case of cross validation is leave-one-out cross-validation. There the cross-validation process is performed multiple times on the training sets $X \setminus \{x_0\}, \ldots, X \setminus \{x_n\}$ and the test sets $\{x_0\}, \ldots, \{x_n\}$ respectively. Finally the, in the test phases, estimated errors are aggregated to one overall error measure. The advantage of leave-one-out cross-validation over other cross-validations is that all, except one, samples $x_i \in X$ can be used to train the model. Therefore this kind of cross-validation normally is used when the data set X is rather small, as in the experiments of this work (see e.g. sample size of [10]). Another advantage of this method is that the test set X_V consists only of one sample. Therefore the error estimation process is very sensitive to outliers. This outliers can store a high level on information in small datasets. The leave-one-out cross-validation process is shown in figure A.3.

A.5. Statistical Tests

In the following I want to give a short overview of statistical test procedures, cited from [13]. But before that, it is necessary to give some definitions.

Definition 6 (Statistical Test)
A statistical test is a method of deducing properties of an underlying distribution T, called test-statistic, by analysis of data, in order to verify a statistical hypothesis.

Definition 7 (Statistical Hypothesis)
A statistical hypothesis is a scientific hypothesis that is testable on the basis of observing a process that is modelled via a set of random variables.

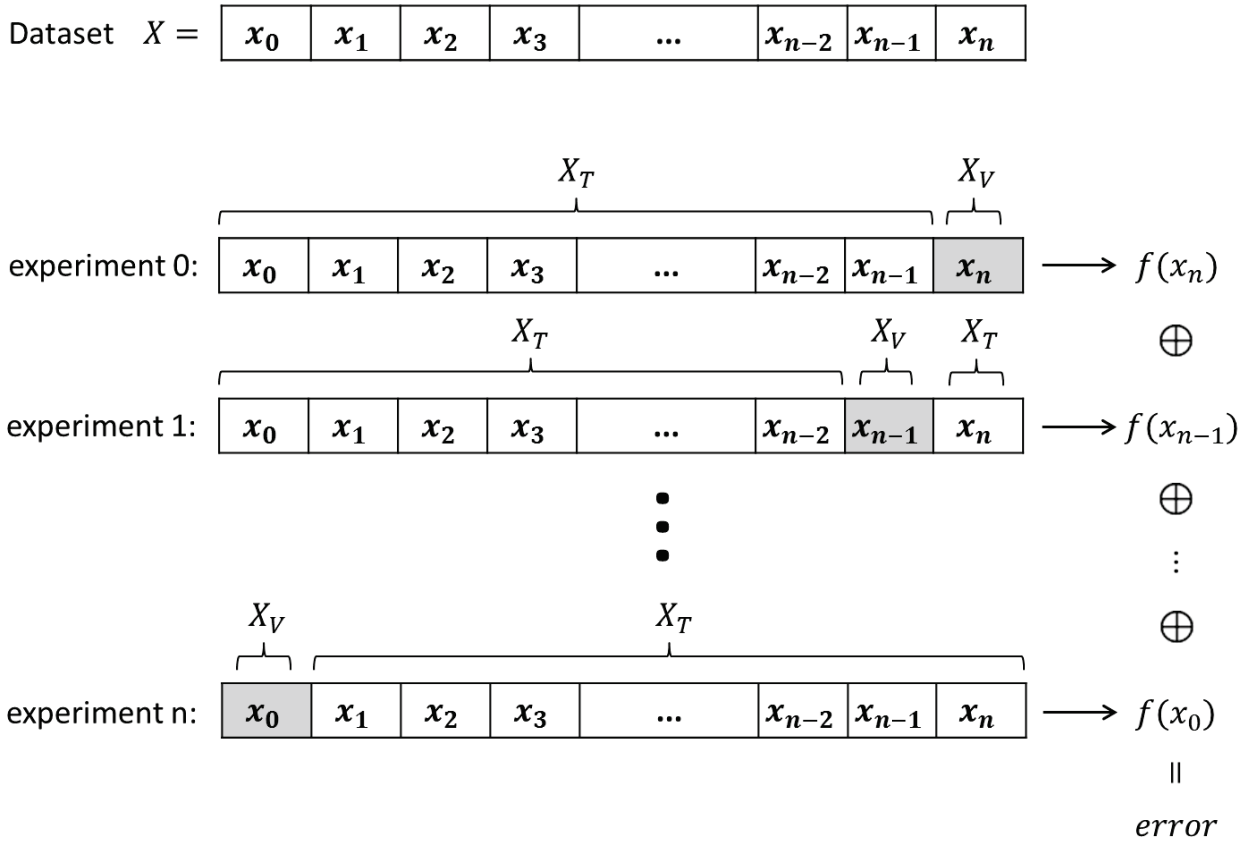


Figure A.3. *Leave-one-out cross-validation. In every experiment, the model f is trained using the sets $X\backslash\{x_0\}, \ldots, X\backslash\{x_n\}$ and tested, using the sets $\{x_0\}, \ldots, \{x_n\}$ respectively. The results $f(x_0), \ldots, f(x_n)$ of every iteration are then aggregated, shown by $\otimes$, to an overall error.*

Furthermore, a result of a statistical test can be *statistically significant*.

Definition 8 (Statistically Significant Result)
A test result of a statistical test is called statistically significant, if it has been predicted as unlikely by chance alone, according to a threshold probability, the so-called significance level.

The goal of a statistical test is to verify a statistical hypothesis H_1 by rejecting the so-called null-hypothesis H_0 :"H_1 does not hold" (H_1 is called alternative hypothesis of H_0). This is done by determining how likely it would be, for a given set of observations $X = \{x_0, \ldots, x_n\}$, to occur if H_0 was true. If there is a high probability $P(H_0$ holds) that X would not occur, if H_0 holds, then the hypothesis H_0 can be rejected and H_1 can be seen as verified. Note that the probability of making an incorrect decision is *not* the probability that H_0 holds, nor that H_1 is false, but often in practice it can be used to *indicate* that H_0 is true.

A statistical test procedure, given a set of observations $X = \{x_0, \ldots, x_n\}$, can be summarized by the following steps:

1. Formulation of the null-hypothesis H_0 and the alternative hypothesis H_1.
2. Choose a statistical test based on a test-statistic T.
3. Choose a confidence interval K based on a significance-level α, where K consists of

values of T, which hold under H_0 with probability smaller than α.

4. Compute from the observations X the observed value t_X of T.
5. Reject H_0, if $t_X \in K$.

It can be shown that the following process is equivalent to the previous one:

1. Formulation of the null-hypothesis H_0 and the alternative hypothesis H_1.
2. Choose a statistical test based on the test-statistic T.
3. Choose a significance level α.
4. Compute from the observations X the observed value t_X of T.
5. Calculate the $p - value$ p that is the probability, under H_0, of sampling a test statistic as least as extreme as t_X.
6. Reject H_0, if and only if $p < \alpha$.

In this form, a statistical test requires the calculation of the so-called p-value p, which is the probability of sampling a test-statistic as least as extreme as the corresponding value t_X of T, given X, under H_0. The p-value is defined as

Definition 9 (p-value)
The p-value p is the probability of sampling a value t of T "as least as extreme as" a given value t_X of T, given observations X, under a hypothesis H_0. Where, "as least as extreme" means the probability of either $\{T \geq t_X\}$ (right-tail-event), $\{T \leq t_X\}$ (left-tail-event), or, the smaller of $\{T \geq t_X\}$ and $\{T \leq t_X\}$ (double-sided-event).

$$p = \begin{cases} P(T \geq t_X | H_0) & \text{right-tail-event} \\ P(T \leq t_X | H_0) & \text{left-tail-event} \\ 2 \cdot \min\left(P(T \geq t_X | H_0), P(T \leq t_X | H_0)\right) & \text{double-sided-event} \end{cases}$$

Examples of statistical tests are the non parametric Mann-Whitney U test and the one tailed F-test.

A.5.1. Non Parametric Mann-Whitney U Test

The non parametric Mann-Whitney U test [26] is a test of the null-hypothesis that two populations are from the same distribution. Especially is tested, if a particular population tends to have larger values than an other. The test assumptions and formal statements of the hypothesis are the following. Let X and Y be two random variables with distribution functions F_X and F_Y respectively, which are shifted with parameter a, i.e. $F_Y(x) = F_X(x - a)$. Let $X = \{x_0, \ldots, x_n\}$ and $Y = \{y_0, \ldots, y_m\}$ be independent, then the Mann-Whitney U test is based on the hypothesis

$$H_0 : a = 0 \quad \text{and} \quad H_1 : a \neq 0$$

Moreover the test is based on the calculation of the test-statistic U, which distribution under H_0 is known. The test-statistic is defined as

Definition 10 (Mann-Whitney U Test Statistic)
The Mann-Whitney U test statistic of two sets of observations $X = \{x_0, \ldots, x_n\}$ and $Y = \{y_0, \ldots, y_m\}$ is defined as

$$U = \sum_{i=0}^{n} \sum_{j=0}^{m} [x_i = y_j]_?$$

where $[P]_? = 1$ iff the clause P is true, and, $[P]_? = 0$ if P is false.

With this definition the normal way of statistical tests, as outlined above, can be used as algorithm for the Mann-Whitney U test that is

1. Set $H_0 : a = 0$ and $H_1 : a \neq 0$, where a is assumed to be the shift between the distribution functions $f_Y(x) = f_X(x-a)$ of the observations $X = \{x_0, \ldots, x_n\}$ and $Y = \{y_0, \ldots, y_m\}$.
2. Set up the test-statistic U as defined above.
3. Choose a significance level α.
4. Calculate the p-value p, as defined above.
5. Reject H_0 iff $p < \alpha$.

This test is especially interesting, because it implies the rejection (acceptance) of the following hypothesis

$$H_0 : \mu_X = \mu_Y \quad \text{and} \quad H_1 : \mu_X \neq \mu_Y$$

with the mean values μ_X and μ_Y, of X and Y respectively. Note that this implication is used in this work to verify claim 1.

A.5.2. One Tailed F-Test

The test is based on the Fisher-Z-transformation [8] and it tests the null-hypothesis that two samples have statistically different Pearson Product-Moment Correlation Coefficient (CC). The Fisher-Z-transformation is defined as

Definition 11
Let $CC(X, Y)$ be the Pearson Product-Moment Correlation Coefficient of two observations $X = \{x_0, \ldots, x_N\}$ and $Y = \{y_0, \ldots, y_N\}$. Then the Fisher-Z-transformation z is defined as

$$z = \frac{1}{2} \ln \left(\frac{1 + CC(X,Y)}{1 - CC(X,Y)} \right)$$

where ln(.) denotes the natural logarithm.

If (X, Y) has approximately bivariate normal distribution and if all pairs (x_i, y_i) are independent for $i, j \in \{0, \ldots, N\}$, then z is approximately normal distributed with mean $\ln\left(\frac{1+CC(X,Y)}{1-CC(X,Y)}\right)$ and standard deviation $\frac{1}{\sqrt{N-3}}$.

This gives us the following steps for a statistical test based on a p-value. Let the three observations $X = \{x_0, \ldots, x_N\}$, $Y = \{y_0, \ldots, y_N\}$ and $G = \{g_0, \ldots, g_N\}$ be given. The goal is to compare the two Pearson Product-Moment Correlation Coefficients $CC(X, G)$ and $CC(Y, G)$ for being statistically significantly different, especially that $CC(X, G)$ is statistically significantly greater than $CC(Y, G)$.

1. Set the null-hypothesis for the correlation coefficients $\rho_X := CC(X, G)$ and $\rho_Y := CC(Y, G)$, $H_0 : \rho_X - \rho_Y = 0$ and the alternative hypothesis $H_1 : \rho_X - \rho_Y > 0$.

2. Set a significance-level α

3. Compute the Z-transformations z_X and z_Y, of ρ_X and ρ_Y respectively, as

$$z_X = \frac{1}{2} \ln \left(\frac{1 + \rho_X}{1 - \rho_X} \right) \quad \text{and} \quad z_Y = \frac{1}{2} \ln \left(\frac{1 + \rho_Y}{1 - \rho_Y} \right)$$

4. Note that the sum $X+Y$ of two independent normal distributed variables X and Y, with means μ_X and μ_Y and standard deviations sd_X and sd_Y respectively, is again normal distributed with mean $\mu_X + \mu_Y$ and standard deviation $sd_X + sd_Y$. This implies, that the difference $z_X - z_Y$ is normal distributed with mean

$$\frac{1}{2} \ln \left(\frac{1 + \rho_X}{1 - \rho_X} \right) - \frac{1}{2} \ln \left(\frac{1 + \rho_Y}{1 - \rho_Y} \right)$$

   and standard deviation 0 (the number of sample pairs (x_i, g_i) equals the number of pairs (y_i, g_i) and is N).

   Therefore, the p-value can be computed by

$$p = 1 - f(z_X - z_Y)$$

   where f is the probability density function of the normal distribution with mean and standard deviation computed as above.

5. Reject H_0 iff $p < \alpha$.

Bibliography

[1] Gary Bradski. "OpenCV-Library". In: *Dr. Dobb's Journal of Software Tools* (2000).

[2] Ran Carmi and Laurent Itti. "Visual causes versus correlates of attentional selection in dynamic scenes". In: *Vision research* 46.26 (2006), pp. 4333–4345.

[3] Jaeseob Choi et al. "Visual fatigue evaluation and enhancement for 2D-plus-depth video". In: *Proc. IEEE ICIP* (2010), pp. 2981–2984.

[4] Dorin Comaniciu and Peter Meer. "Mean shift: A robust approach toward feature space analysis". In: *IEEE Trans. Pattern Analysis and Machine Intelligence* 24.5 (2002), pp. 603–619.

[5] Dubravko Culibrk et al. "Salient motion features for video quality assessment". In: *IEEE Trans. Image Processing* 20.4 (2011), pp. 948–958.

[6] Matthias Ehrgott. *Multicriteria optimization*. Vol. 2. Springer, 2005.

[7] Frank L Engel. "Visual conspicuity, visual search and fixation tendencies of the eye". In: *Vision research* 17.1 (1977), pp. 95–108.

[8] Ronald Aylmer Fisher et al. "On the "Probable Error" of a Coefficient of Correlation Deduced from a Small Sample." In: *Metron* 1 (1921), pp. 3–32.

[9] Wilson S Geisler and Kee-Lee Chou. "Separation of low-level and high-level factors in complex tasks: visual search". In: *Psychological review* 102.2 (1995), p. 356.

[10] Lutz Goldmann, Francesca De Simone, and Touradj Ebrahimi. "Impact of acquisition distortions on the quality of stereoscopic images". In: *Proceedings of the International Workshop on Video Processing and Quality Metrics for Consumer Electronics* (2010), pp. 1–6. URL: `http://mmspg.epfl.ch/3diqa`.

[11] David A Goss and Huifang Zhai. "Clinical and laboratory investigations of the relationship of accommodation and convergence function with refractive error". In: *Documenta ophthalmologica* 86.4 (1994), pp. 349–380.

[12] Robert M Haralick, Karthikeyan Shanmugam, and Its´ Hak Dinstein. "Textural features for image classification". In: *IEEE Trans. Systems, Man and Cybernetics* 6 (1973).

[13] Joachim Hartung, Bärbel Elpelt, and Karl-Heinz Klösener. *Statistik: Lehr-und Handbuch der angewandten Statistik*. Walter de Gruyter, 2009.

[14] Heiko Hirschmuller. "Stereo processing by semiglobal matching and mutual information". In: *IEEE Trans. Pattern Analysis and Machine Intelligence* 30.2 (2008), pp. 328–341.

[15] Aapo Hyvärinen, Jarmo Hurri, and Patrik O Hoyer. *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision.* Vol. 39. Springer, 2009.

[16] Wijnand IJsselsteijn et al. "Perceived depth and the feeling of presence in 3DTV". In: *Displays* 18.4 (1998), pp. 207–214.

[17] ITU-R. "Methodology for the subjective assessment of the quality of television pictures". In: *Tech. Rep. BT.500-11* (2002).

[18] Bela Julesz. "Experiments in the visual perception of texture". In: *Scientific American* 232 (1975).

[19] Yong Ju Jung et al. "Predicting visual discomfort of stereoscopic images using human attention model". In: *IEEE Trans. Circuits and Systems for Video Technology* (2013). URL: http://ivylab.kaist.ac.kr/demo/3DVCA/3DVCA.htm.

[20] Ronald G Kaptein et al. "Performance evaluation of 3D-TV systems". In: *Electronic Imaging* (2008), pp. 680819–11.

[21] Donghyun Kim and Kwanghoon Sohn. "Visual Fatigue Prediction for Stereoscopic Image". In: *IEEE Trans. Circuits and Systems for Video Technology* 21.2 (2011), pp. 231–236.

[22] Frank L Kooi and Alexander Toet. "Visual comfort of binocular and 3D displays". In: *Displays* 25 (2004), pp. 99–108.

[23] Valentin Kulyk et al. "3D video quality assessment with multi-scale subjective method". In: *Fifth International Workshop on Quality of Multimedia Experience* (2013), pp. 106–111.

[24] Marc Lambooij, W A IJsselsteijn, and I Heynderickx. "Visual discomfort of 3D TV: Assessment methods and modeling". In: *Displays* 32.4 (2011), pp. 209–218.

[25] Marc Lambooij et al. "Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review". In: *Journal of Imaging Science and Technology* 53.3 (2009), pp. 30201–1.

[26] Henry B Mann and Donald R Whitney. "On a test of whether one of two random variables is stochastically larger than the other". In: *The Annals of Mathematical Statistics* 18.1 (1947), pp. 50–60.

[27] Andrzej Materka and Michal Strzelecki. "Texture analysis methods–a review". In: *Technical University of Lodz, Institute of Electronics, COST B11 report, Brussels* (1998), pp. 9–11.

[28] Urvoy Matthieu, Barkowsky Marcus, and Le Callet Patrick. "How visual fatigue and discomfort impact 3D-TV quality of experience: a comprehensive review of technological, psychophysical, and psychological factors". In: *Annals of Telecommunications* 68.11-12 (2013), pp. 641–655.

[29] Benoit Michel. *Digital Stereoscopy, Scene to Screen 3D Production Workflow*. Ed. by Gabriellle Leyden. Stereoscopy News, 2012. ISBN: 978-1-48015709-5.

[30] Yuji Nojiri et al. "Parallax distribution and visual comfort on stereoscopic HDTV". In: *Proc. IBC*. 2006, pp. 373–380.

[31] Farley Norman and James Todd. "Stereoscopic discrimination of interval and ordinal depth relations on smooth surfaces and in empty space". In: *Perception-London* 27.3 (1998), pp. 257–272.

[32] Levent Onural. *3D video technologies: An overview of research trends*. 2011.

[33] Derrick Parkhurst, Klinton Law, and Ernst Niebur. "Modeling the role of salience in the allocation of overt visual attention". In: *Vision research* 42.1 (2002), pp. 107–123.

[34] Robert Patterson and Aris Silzars. "Immersive stereo displays, intuitive reasoning, and cognitive engineering". In: *Journal of the Society for Information Display* 17.5 (2009), pp. 443–448.

[35] Eli Peli. "Contrast in complex images". In: *JOSA A* 7.10 (1990), pp. 2032–2040.

[36] Yury Petrov and Andrew Glennerster. "Disparity with respect to a local reference plane as a dominant cue for stereoscopic depth relief". In: *Vision research* 46.26 (2006), pp. 4321–4332.

[37] John R. Quinlan. "Learning with continuous classes". In: 92 (1992), pp. 343–348.

[38] Hosik Sohn et al. "Predicting Visual Discomfort Using Object Size and Disparity Information in Stereoscopic Images". In: *IEEE Trans. Broadcasting* 59.1 (2013), pp. 28–37.

[39] Hosik Sohn et al. "Visual comfort amelioration technique for stereoscopic image: Disparity remapping to mitigate global and local discomfort causes". In: *IEEE Trans. Circuits and Systems for Video Technology* (2013).

[40] Wa James Tam et al. "Stereoscopic 3D-TV: Visual Comfort". In: *IEEE Trans. Broadcasting* 57.2 (2011), pp. 335–346.

[41] Cédric Thébault et al. "Automatic depth grading tool to successfully adapt stereoscopic 3D content to digital cinema and home viewing environments". In: *IST/SPIE Electronic Imaging* (2013), pp. 86480X–86480X.

[42] Alexander Toet. "Computational versus psychophysical bottom-up image saliency: A comparative evaluation study". In: *IEEE Trans. Pattern Analysis and Machine Intelligence* 33.11 (2011), pp. 2131–2146.

[43] Anne M Treisman and Garry Gelade. "A feature-integration theory of attention". In: *Cognitive psychology* 12.1 (1980), pp. 97–136.

[44] Guido Van Rossum and Fred L Drake Jr. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995. URL: `http://www.python.org`.

[45] Vladimir Naumovich Vapnik. *Statistical learning theory*. Vol. 2. Wiley New York, 1998.

[46] Matthias Wöpking. "Viewing comfort with stereoscopic pictures: An experimental study on the subjective effects of disparity magnitude and depth of focus". In: *Journal of the Society for Information Display* 3.3 (1995), pp. 101–103.

[47] Sumio Yano. "Experimental stereoscopic high-definition television". In: *Displays* 12.2 (1991), pp. 58–64.

[48] Sumio Yano and Ichiro Yuyama. "Stereoscopic HDTV: Experimental system and psychological effects". In: *SMPTE journal* 100.1 (1991), pp. 14–18.

Statutory Declaration

Ich erkläre an Eides statt, dass ich die vorliegende Masterarbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe. Die vorliegende Masterarbeit ist mit dem elektronisch übermittelten Textdokument identisch.

Linz, 10.03.2015